NILPOTENT Lie Algebras of Vectorfields

and

Local Controllability of Nonlinear Systems

by

Matthias Kawski

Vordiplom, Universitaet Hamburg, 1982

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

1986
This thesis for the Doctor of Philosophy degree by

Matthias Kawski

has been approved for the

Department of

Mathematics

by

[Signature]

Henry G. Hermes

Robert W. Easton

Date Nov 26, 1986
Kawski, Matthias (Ph.D., Mathematics)

Nilpotent Lie algebras of vectorfields and local controllability of nonlinear systems

Thesis directed by Professor Henry G. Hermes

In chapter I it is shown that vectorfields generating a nilpotent Lie algebra are always associated to a dilation relative to suitable local coordinates.

In chapter II a necessary condition for small time local controllability (STLC) is proven. To further narrow the gap between necessary and sufficient conditions for STLC several particular control systems are discussed and shown to be STLC. Finally, a new class of control variations is introduced and illustrated in an example.
CONTENTS

CHAPTER

I. NILPOTENT LIE ALGEBRAS OF VECTORFIELDS .................. 1
   1.1 Introduction, Notation and Definitions ...................... 1
   1.2 Statement of the Theorem and Side Results .................. 4
   1.3 Technical Lemmata and Development of the Tools to
       Prove the Theorem ........................................ 6
   1.4 Proof of the Theorem ...................................... 22

II. LOCAL CONTROLLABILITY OF NONLINEAR SYSTEMS

................................................................. 25
   2.1 Introduction Including Statement of Old Results .......... 25
   2.2 Discussion of the Old Results, Motivation for and
       Statement of the New Results ................................ 28
   2.3 Technical lemmata and provision of the tools for proving
       the theorem and discussing the examples .................. 36
       2.3.1 Scaling lemma .......................................... 36
       2.3.2 Independence of [X^0, X^1] (0) from X^1 (0) .......... 39
       2.3.3 Chen- and Lie series and change of basis ............ 40
       2.3.4 Estimation of derivatives (X^1 \Phi) (0) ............... 56
       2.3.5 Some integral inequalities ........................... 59
   2.4 Proof of Theorem 1 ....................................... 62
2.5 Examples ........................................................................................................ 72

2.5.1 A new good bracket of type (2k,3) .................................................. 72

2.5.2 Balancing bad brackets of different types .................. 77

2.5.3 Balancing bad brackets of the same type ...................... 82

2.5.4 A new class of control variations ......................... 86

2.6. Conclusion ................................................................................................. 92

BIBLIOGRAPHY .................................................................................................. 94
FIGURES

1. A control variation ................................................................. 29
2. A control variation ................................................................. 31
3. Weight assignment to types of Lie products .............................. 32
4. A control variation ................................................................. 74
5. A control variation ................................................................. 79
6. A control variation ................................................................. 80
7. A control variation ................................................................. 83
8. A control variation ................................................................. 85
9. Projections of $x(t,u)$ ............................................................. 85
CHAPTER I

NILPOTENT LIE ALGEBRAS OF VECTORFIELDS

1.1 Introduction, Notation and Definitions

Let $X^1, \cdots, X^k$ be real analytic vectorfields on a real analytic n-dimensional manifold $M^n$. $L = L(X^1, \cdots, X^k)$ is the Lie algebra generated by the vectorfields $X^1, \cdots, X^k$, where $[V; W] = VW - WV$ is the Lie product of the vectorfields $V$ and $W$. We often use $(ad_V W) = [V; W]$ and $(ad^{k+1}_V W) = [V; (ad^k V; W)]$.

The central descending series of $L$ is defined by $L^{(1)} = L$ and inductively $L^{(k+1)} = [L; L^{(k)}]$. (Caution: Sometimes in the literature one uses $L^{(0)}$ for $L^{(1)}$.) The Lie algebra $L$ is nilpotent if for some smallest integer $\rho \geq 1$, $L^{(\rho)} = \{0\}$.

Vectorfields generating a nilpotent Lie algebra are very useful for practical purposes as they, for example, allow one to compute solutions of ordinary differential equations explicitly by decomposing the equation into a finite number of possibly much less difficult equations.

Specifically in control systems of the form

$$\dot{x} = X^0(x) + X^1(x), \ x(0) = 0,$$

(1)

with an integrable bounded control function $u$, one has a natural decomposition of the right side. When $L(X^0, X^1)$ is nilpotent, then it is possible to write the
solution \( x(\cdot, u) \) of (1) as a composition of a finite number of solutions of ordinary differential equations.

It is relatively simple to give examples of vector fields generating a nilpotent Lie algebra if one uses dilations. We will show that in the proper local coordinates, every finite collection of vector fields generating a nilpotent Lie algebra is associated to a dilation.

We next give a brief review of dilations and filtrations/gradations on the Lie algebra of vector fields with polynomial coefficients, where we try to follow as far as possible the notation as introduced in [14].

For a fixed choice of coordinates \( x = (x_1, \ldots, x_n) \) on \( \mathbb{R}^n \) and a nondecreasing sequence of positive integers \( 1 = r_1 \leq r_2 \leq \cdots \leq r_n \) define a one-parameter family of dilations \( \delta_t : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n \) by \( \delta_t(x) = (t^{r_1}x_1, \ldots, t^{r_n}x_n) \).

A polynomial \( p = p(x) \) is homogeneous of degree \( m \) w.r.t. the dilation \( \delta_t \) if \( p(\delta_t(x)) = t^m p(x) \). Let \( H_m \) be the set of all polynomials homogeneous of degree \( m \) w.r.t. \( \delta_t \), and set \( H_m = \{0\} \) if \( m < 0 \).

**Example:** If \( \delta_t(x_1, x_2, x_3) = (tx_1, t^2x_2, t^6x_3) \) is a dilation on \( \mathbb{R}^3 \), then \( p(x) = 7x_3 - x_2x_1^4 + 3x_1^6 \) is homogeneous of degree 6 w.r.t. \( \delta_t \), i.e. \( p \in H_6 \).

A vector field \( X(x) = \sum_{i=1}^{n} a_i(x) \frac{\partial}{\partial x_i} \) with polynomial coefficients \( a_i(x) \) is homogeneous of degree \(-k \in \mathbb{Z} \) if \( XH_m \subseteq H_{m-k} \) for all \( m \geq 0 \); this is equivalent to \( a_i(x) \in H_{r_i-k} \) for \( i = 1, \ldots, n \). Let \( \mathfrak{n}_{-k} \) be the set of all vector fields homogeneous of degree \(-k \), and set \( \mathfrak{n}_{-k} = \{0\} \) if \( k > r_n \).
Example: With the same dilation as above
\[ X(x) = 3 \frac{\partial}{\partial x_1} + 7x_1 \frac{\partial}{\partial x_2} + (3x_1^5 + x_1x_2^2) \frac{\partial}{\partial x_3} \in \mathfrak{n}_{-1}, \text{ i.e. } X \text{ is homogeneous of degree -1 w.r.t. } \delta. \]

One readily verifies \( H_m \cdot H_n \subseteq H_{m+n} \) if \( m, n \geq 0 \), and \( \mathfrak{n}_k \cdot \mathfrak{n}_l \subseteq \mathfrak{n}_{k+l} \) for all \( k, l \in \mathbb{Z} \): If \( X \in \mathfrak{n}_{-k}, Y \in \mathfrak{n}_{-l} \), then
\[ [X; Y]H_m = YXH_m - XYH_m \subseteq YH_{m-k} + XH_{m-l} \subseteq H_{m-k-l}. \]

The spaces \( H_m \) give a gradation on the algebra \( P \) of polynomials in \((x_1, \cdots, x_n)\): \( P = \bigoplus_{m \geq 0} H_m \). Similarly the spaces \( \mathfrak{n}_k \) give a gradation on the Lie-algebra \( \mathfrak{n} \) of vectorfields with polynomial coefficients, \( \mathfrak{n} = \bigoplus_{k \geq -r_n} \mathfrak{n}_k \).

Defining \( P_m = \sum_{k=0}^{m} H_k \) and \( \mathfrak{n}_{-k} = \sum_{l=-r_n}^{-k} \mathfrak{n}_l \), clearly \( P_{m+1} \supseteq P_m, m \geq 0 \), and \( \mathfrak{n}_{-k+1} \supseteq \mathfrak{n}_{-k}, -k \geq -r_n \), and one obtains the filtrations
\[ P = \bigcup_{k=0}^{\infty} P_k \text{ and } \mathfrak{n} = \bigcup_{l=-r_n}^{-1} \mathfrak{n}_l, \mathfrak{n}_{-1} = \bigcup_{l=-r_n}^{-1} \mathfrak{n}_l. \]

In the following we will call a polynomial \( p(x) \in P_m \) "of degree \( m \)" and a vectorfield \( X \in \mathfrak{n}_{-k} \) "of degree \( -k \)."

If \( X^1, \cdots, X^k \subseteq \mathfrak{n}_{-1} \) are vectorfields with polynomial coefficients, then clearly \( L(X^1, \cdots, X^k) = L \subseteq \mathfrak{n}_{-1} \) and \( L^{(i)} \subseteq \mathfrak{n}_{-i} \) for \( i \geq 1 \). But \( \mathfrak{n}_{-k} = \{0\} \) for \( k > r_n \), and thus \( X^1, \cdots, X^k \) generate a nilpotent Lie algebra.

Usually we will use greek letters \( \alpha = (\alpha_1, \cdots, \alpha_k), \nu = (\nu_1, \cdots, \nu_n) \) to denote multi-indices with \( \alpha_i, \nu_i \in \mathbb{Z}_0^+ \). As usual
\[ |\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_k, \]
\[ \alpha! = \alpha_1! \cdots \alpha_k!, \]
\[ x^\alpha = x_1^{\alpha_1} \cdots x_k^{\alpha_k}, \]
\[ D^\alpha = \frac{\partial |\alpha|}{\partial x_1^{\alpha_1} \cdots \partial x_k^{\alpha_k}}, \text{ and} \]
\[ r^\alpha = r_1 \alpha_1 + \cdots + r_k \alpha_k, \]

where we allow \( 1 \leq k \leq n \), even though \( r = (r_1, \cdots, r_n) \) is a fixed \( n \)-vector throughout this work.

For analytic vectorfields \( Y^1, \cdots, Y^k, Z \) and \( f : M \rightarrow \mathbb{R} \) analytic we define
\[ (\text{ad}^\alpha Y^i; Z) = (\text{ad}^{\alpha_1} Y^1; (\cdots; (\text{ad}^{\alpha_k} Y^k; Z) \cdots)) \]
and
\[ Y^\alpha f = (Y^1)^{\alpha_1}(Y^2)^{\alpha_2} \cdots (Y^k)^{\alpha_k} f \]
denoting an \( |\alpha| \)-th order partial derivative of \( f \).

### 1.2 Statement of the Theorem and Side Results

As already mentioned in the introduction, vectorfields \( X^1, \cdots, X^k \) which are of degree (at most) - 1 w.r.t. some dilation \( \delta_t \), generate a nilpotent Lie algebra \( L = L(X^1, \cdots, X^k) \).

We show that (in the proper local coordinates) the converse is also true.

**Theorem:** Let \( X^1, \cdots, X^k \) be real analytic vectorfields on the real analytic \( n \)-dimensional manifold \( M \) which generate the nilpotent Lie algebra \( L = L(X^1, \cdots, X^k) \). If \( p \in M \) is such that \( \dim L(p) = n \), then there are local coordinates \((x_1, \cdots, x_n)\) in a neighborhood of \( p \) and a dilation \( \delta_t(x) = (t^{r_1}x_1, \cdots, t^{r_n}x_n) \) such that relative to these coordinates \( X^1, \cdots, X^k \) have polynomial coefficients and are of degree -1 w.r.t. the dilation \( \delta_t \).
This immediately leads to: If \( X^i(x) = \sum_{j=1}^{n} a_{ij}(x) \frac{\partial}{\partial x_j}, i = 1, \cdots n; \) then \( a_{ij}(x) \) is a polynomial of degree \( r_j - 1 \) w.r.t. \( \delta_i \) and depends on \( x_1, \cdots x_{j-1} \) only. In particular \( a_{i1} = \text{const.} \) for \( i = 1, \cdots k, j = 1, \cdots n. \)

An immediate corollary to the theorem is the known result:

**Corollary:** If \( G \) is a nilpotent Lie group locally acting transitively on a real analytic manifold \( M^n \), i.e. \( \dim g(p) = n \), then there are local coordinates on \( M \) such that the action of \( G \) relative to these coordinates is polynomial.

At this point we state two other interesting facts as lemmata. Later they are needed in the proof of the theorem.

It is well-known that \( \dim L^{(j)} \) is strictly decreasing for \( j = 1, 2, \cdots \rho \) if \( L \) is nilpotent. However \( \dim L^{(j)}(p) \) need not be strictly decreasing — at least, it cannot increase since \( L^{(j)}(p) \succeq L^{(k)}(p) \) for \( j \leq k \). But we need the following lemma which also holds under milder hypotheses than requiring analyticity.

**Lemma 1.** If \( L = L(X^1, \cdots X^n) \) is a nilpotent Lie algebra of vectorfields on \( M^n \) with \( \dim L(p) = n \), then \( \dim L^{(2)}(p) < n. \)

Recall that a distribution \( \Delta \) is called regular at \( p \) if there is a neighborhood \( U \) of \( p \) such that for all \( i \geq 0 \) the \( i^{\text{th}} \) derived distribution \( \Delta^{(i)} \) is of constant dimension throughout \( U \). (The \( i^{\text{th}} \) derived distribution is the span of all Lie products of at most \( i \) vectorfields in \( \Delta \).) We need a similar property, but note that the \( L^{(i)} \) are decreasing, whereas the \( \Delta^{(i)} \) are increasing.
Lemma 2. If $L$ is a nilpotent Lie algebra of analytic vector fields on $M^n$ and $p \in M$ is such that $\dim L(p) = n$, then there is a neighborhood $U$ of $p$ such that for all $i = 1, 2, \cdots, n$ $\dim L^{(i)}(q)$ is constant for $q \in U$.

Finally a large portion of the proof of the theorem is based on

Lemma 3. If $Z^i = \frac{\partial}{\partial x_i} + \sum_{j=i+1}^n a_{ij}(x) \frac{\partial}{\partial x_j}$, $i = 1, 2, \cdots, n$ are vector fields of degree $-r_i$ w.r.t. the dilation $\delta_i$ (i.e. $a_{ij}(x)$ are polynomials of degree $r_j - r_i$ w.r.t. $\delta_i$) and $\psi: \mathbb{R}^n \to \mathbb{R}$ is analytic and satisfies $(Z^i \psi)(0) = 0$ for all $\alpha = (\alpha_1, \cdots, \alpha_n)$ with $r \cdot \alpha > s$, then $\psi \in H_s$, i.e. $\psi$ is a polynomial of degree $s$ w.r.t. $\delta_i$.

If $a_{ij} \equiv 0$, $i = 1, \cdots, n$, $j = i+1, \cdots, n$, one has $Z^i = D^i = \frac{\partial}{\partial x_i}$ and the lemma is immediate. But since the $Z^i$ may be noncommuting, the proof of lemma 3 requires some work.

Remark: In general it is not possible to achieve that the vector fields $X^i$ in the theorem are homogeneous of degree $-1$ (or $-d_i$), or equivalently that the polynomials $a_{ij}(x)$ are homogeneous w.r.t. some dilation, since this would immediately lead to a gradation of the Lie algebra $L(X^1, \cdots, X^n)$, which in turn would allow one to construct dilating automorphisms of $L = L(X^1, \cdots, X^n)$. However, there are examples of nilpotent Lie algebras in the literature, which do not admit any dilating automorphisms [24].

1.3 Proofs of the lemmata and technical claims.

We start by proving lemma 1. This enables us to construct local
coordinates and a dilation on $M^n$, which in turn we will use to prove lemma 2. Finally, we will verify several technical claims and prove lemma 3.

**Proof of Lemma 1.** Since $L$ is generated by a finite number of vectorfields and is nilpotent, $L$ and each ideal $L^{(i)}$ is finite dimensional.

Choose a basis $\{Z^1, \ldots , Z^n\}$ for $L$ such that $\{Z^1(p), \ldots , Z^n(p)\}$ are linearly independent and $\{Z^{n+1}, \ldots , Z^n\}$ vanish at $p$. We will show, if $\dim L^{(2)}(p) = n$, then $\dim L^{(i)}(p) = n$ for all $i \geq 1$, thus contradicting nilpotency of $L$.

Choose vectorfields $\{W^{2,1}, \ldots , W^{2,n}\} \in L^{(2)}(p)$ such that $W^{2,i}(p) = Z^i(p)$, $i = 1, 2, \ldots , n$. Thus there are $a_{ijk}, b_{ij} \in \mathbb{R}$ such that

$$W^{2,i} = Z^i - \sum_{j=n+1}^n b_{ij} Z^j \quad \text{and} \quad W^{2,i} = \sum_{1 \leq j < k \leq s} a_{ijk} [Z^j, Z^k], \quad i = 1, 2, \ldots , n.$$ 

Combining these two relations gives

$$Z^i = \sum_{1 \leq j < k \leq s} a_{ijk} [Z^j, Z^k] + \sum_{j=n+1}^n b_{ij} Z^j, \quad i = 1, 2, \ldots , n. \quad (2)$$

Now substitute for each $Z^\nu$, $1 \leq \nu \leq n$ on the right side of (2) the right side of the $\nu$th equation of (2). The result is $n$ equations of the form

$$Z^i = W^{3,i} \bmod \mathcal{K}_p, \quad i = 1, 2, \ldots , n, \quad (3)$$

where $W^{3,i} \in L^{(3)}$ and $\mathcal{K}_p = \text{span}\{Z^{n+1}, \ldots , Z^n\}$ is the isotropy subalgebra of $L$ at $p$.

Repeated substitution of (2) for all $Z^\nu$, $\nu = 1, 2, \ldots , n$, on the right side of (3) leads to equations of the form
\[ Z^i = W^{h,i} \mod \mathcal{H}_p, \quad i = 1, 2, \ldots, n, \quad h \geq 2 \quad (4) \]

with \( W^{h,i} \in L^{(h)} \) and \( \mathcal{H}_p \) as above. Evaluating (4) at \( p \) shows \( \dim L^{(h)} = n \), all \( h \geq 2 \), thus finishing the proof of lemma 1.

We proceed by defining a family of dilations and local coordinates in a neighborhood of \( p \). For \( i = 1, \ldots, n \) define

\[ r_i = \max\{ j \in \mathbb{Z}^+: \dim L^{(j)}(p) = (n+1) - i \} \text{ and } \]

\[ s_i = \min\{ (n+1) \cup \{ j \in \mathbb{Z}^+: r_j > r_i \} \}. \quad (5) \]

From \( \dim L^{(1)}(p) = n \) and lemma 1, \( r_1 = 1 \); and \( r_n = \rho - 1 \), where \( \rho \in \mathbb{Z}^+ \) is such that \( L^{(\rho)} = \{ 0 \} \) and \( L^{(\rho-1)} \neq \{ 0 \} \).

Choose vectorfields \( Y^1, \ldots, Y^n \in L = L(X^1, \ldots, X^n) \) such that \( Y^i \in L^{(r_i)} \), \( i = 1, 2, \ldots, n \) and \( Y^1(p), \ldots, Y^n(p) \) are linearly independent. Then \( Y^1(q), \ldots, Y^n(q) \) are linearly independent for \( q \) sufficiently close to \( p \) and hence there is a neighborhood \( U = U^1 \) of \( 0 \in \mathbb{R}^n \) such that the inverse of the map \( \gamma: U \to M^n \),

\[ \gamma(x) = (\exp x_1 Y^n) \circ \cdots \circ (\exp x_1 Y^1)(p) \quad (6) \]

defines local coordinates on the neighborhood \( \gamma(U) \) of \( p \in M^n \).

From now on we will work entirely in this neighborhood \( \gamma(U) \) of \( p \), and identify the point \( q \in \gamma(U) \) with its coordinates \( \gamma^{-1}(q) \in \mathbb{R}^n \), and similarly identify \( U \) with \( \gamma(U) \), if this will cause no confusion.

Furthermore we will have to restrict our work several more times (but a finite number of times) to a neighborhood \( U = U^i \) of \( p \). To avoid unnecessary confusing notation we will call all these neighborhoods \( U \), but
understand that $U$ stands for the intersection $U = \bigcap_i U^i$, on which all our statements hold.

Finally define a one-parameter family of dilations $\delta: [0,1] \times U \to U$ by

$$\delta_t(x) = (t^{r_1}x_1, \ldots, t^{r_n}x_n) \text{ with } r_i \text{ as above}$$

and let $r = (r_1, \ldots, r_n) \in (\mathbb{Z}^+)^n$. (If necessary, shrink $U$ s.t. $\delta_t(U) \subseteq U$ for all $t \in [0,1]$.)

**Proof of Lemma 2.** With $Y^1, \ldots, Y^n$, $r$, and local coordinates defined as above, there are uniquely determined analytic functions $\alpha_{t^h}^{\delta_h}: U \to \mathbb{R}$; $f, g, h = 1, 2, \ldots, n$, such that

$$[Y^f, Y^g](x) = \sum_{h=1}^n \alpha_{t^h}^{\delta_h}(x) Y^h(x); \ f, g, h = 1, \ldots, n.$$  \hspace{1cm} (8)

For $g = 1, \ldots, n$, let $A^g$ be the $n \times n$-matrix $A^g = (\alpha_{t^i}^{\delta_h})_{h=1,2,\ldots,n}$ and for $g = 1, \ldots, n, i = 1, \ldots, n$ let $A^{i,g}$ be the $i \times i$-matrix $A^{i,g} = (\alpha_{t^i}^{\delta_h})_{h=1,2,\ldots,i}$. By skew-symmetry of the Lie product $[\cdot, \cdot]$, we have

$$\alpha_{t^i}^{\delta_h} + \alpha_{t^i}^{\delta_h} = 0, \ f, g, h = 1, \ldots, n.$$  

For a point $x = (x_1, \ldots, x_n)$ in $U$, i.e. sufficiently close to zero, define $n$ smooth curves $\xi^i, i = 1, \ldots, n$ by

$$\xi^i(0) = 0, \ \xi^{i+1}(0) = \xi^i(x_i), \ \xi^i(s) = (\exp s Y^i)(\xi^i(0)), \ i = 1, 2, \ldots, n;$$

in particular $\xi^n(x_n) = x$. (Since we will work with one fixed $x$ at a time, we suppress the dependence $\xi(\cdot) = \xi_x(\cdot)$.)

We need a very simple form of Gronwall's lemma: If $f: \mathbb{R} \to \mathbb{R}$ is smooth, $f(0) = 0$ and $|f(t)| \leq K \int_0^t |f(s)| \, ds$ for all $t \geq 0$, some constant
K \geq 0$, then $f(\cdot) \equiv 0$.

We will show by induction on $\lambda$, $\lambda$ decreasing, that if $L^{(\lambda)}(0) = \text{span}\{Y^{\mu+1}(0), \cdots, Y^n(0)\}$, then $L^{(\lambda)}(x) \subseteq \text{span}\{Y^{\mu+1}(x), \cdots, Y^n(x)\}$ for $x \in U$, the other inclusion being trivial.

Starting with $\lambda = \rho$, we have $L^{(\rho)} = 0$.

Next suppose $L^{(\lambda+1)}(x) = \text{span}\{Y^{\mu+1}(x), \cdots, Y^n(x)\}$ for all $x \in U$ and $L^{(\lambda)}(0) = \text{span}\{Y^{\mu+1}(0), \cdots, Y^n(0)\}$. We claim

$$\alpha^g_{h, \lambda} \equiv \alpha^f_{g, \lambda} \text{ for } f = \mu+1, \mu+2, \cdots, n, \ h = 1, 2, \cdots, \nu, \text{ all } g \quad (10)$$

as a consequence of the induction hypothesis. (We have $Y^f \in L^{(\lambda)}$ for $f = \mu+1, \mu+2, \cdots, n$, and thus for any $g = 1, 2, \cdots, n$, $[Y^f; Y^g] \in L^{(\lambda+1)}$, which by induction hypothesis is a $C^\omega$-model of $\{Y^{\mu+1}, \cdots, Y^n\}$, thus giving (10).)

Let $W \in L^{(\lambda)}$ w.l.o.g. be such that $W(0) = 0$. (If $W(0) \neq 0$, then $W(0) = \sum_{j=\mu+1}^n w_j Y^j(0)$, and we consider $W' = W - \sum_{j=\mu+1}^n w_j Y^j$.)

There are analytic functions $\beta_j: U \to \mathbb{R}$ such that $W(x) = \sum_{j=1}^n \beta_j(x) Y^j(x)$. We have $\beta_j(0)$ and will show $\beta_j \equiv 0$, $j = 1, 2, \cdots, \mu$.

Compute for $i = 1, \cdots, n$

$$[W; Y^i] = \sum_{j=1}^n (Y^i \cdot \beta_j) Y^j + \beta_j [Y^i; Y^i]$$

$$= \sum_{j=1}^n (Y^i \cdot \beta_j) Y^j + \sum_{j=1}^n \sum_{k=1}^n \alpha^j_{k, \lambda} \beta_j Y^k$$

$$= \sum_{j=1}^n \left\{ (Y^i \cdot \beta_j) + \sum_{k=1}^n \alpha^j_{k, \lambda} \beta_j \right\} Y^j, \ i = 1, \cdots, n. \quad (11)$$

Since $[W; Y^i] \in L^{(\lambda+1)}$ for $i = 1, 2, \cdots, n$, there are analytic functions $d_{ij}: U \to \mathbb{R}$ such that $[W; Y^i] = \sum_{j=\mu+1}^n d_{ij} Y^j$, $i = 1, \cdots, n$ (by the induction
hypothesis).

Combining this with the last equation gives

\[ 0 = \sum_{j=1}^{\nu} \left\{ (Y^i \beta_j) + \sum_{k=1}^{n} \alpha_{k,j}^i \beta_k \right\} Y^i + \sum_{j=\nu+1}^{n} \left\{ (Y^i \beta_j) - d_{ij} + \sum_{k=1}^{n} \alpha_{k,j}^i \beta_k \right\} Y^i. \]  \hfill (12)

Using that \( \{Y^1, \ldots, Y^m\} \) are linearly independent (over \( C^\infty(U) \)), and that \( \alpha_{k,j}^i = 0 \) if \( k \geq \mu + 1 \) and \( j \leq \nu \), all \( i \), (note \( \nu \geq \mu \)), one obtains

\[ (Y^i \beta_j) = -\sum_{k=1}^{\mu} \alpha_{k,j}^i \beta_k, \quad i = 1,2, \ldots, n, \quad j = 1,2, \ldots, \mu, \ldots, \nu \]

If we write this in vector form, \( b = (\beta_1, \ldots, \beta_n) \) and \( b^\mu = (\beta_1, \ldots, \beta_\mu) \), then we obtain

\[ Y^i b^\mu = b^\mu A^i_{\mu}, \quad i = 1,2, \ldots, n. \]  \hfill (13)

By induction on \( i \) we show \( b^\mu(\xi^i(0)) = 0, \quad i = 1,2, \ldots, n+1 \).

\( b^\mu(\xi^1(0)) = b^\mu(0) = (0,0, \ldots, 0) = 0 \) starts the induction. Now suppose \( b^\mu(\xi^i(0)) = 0 \) and we show \( b^\mu(\xi^{i+1}(0)) = 0 \). For \( 0 < t < x_i \) compute

\[ b^\mu(\xi^i(t)) = b^\mu(\xi^i(0)) + \int_0^t (Y^i b^\mu)(\xi^i(s))ds \]

\[ = \int_0^t (b^\mu A^i_{\mu})(\xi^i(s))ds \]  \hfill (14)

Introducing norms, (e.g., \( \|b^\mu\| = \sum_{i=1}^{\mu} |\beta_i| \), etc.), there is a constant \( K < \infty \) such that \( \|A(\cdot)\| < K \) on \( U \), and

\[ |b^\mu(\xi^i(t))| \leq \int_0^t K |b^\mu(\xi^i(s))|ds, \quad t \in [0, x_i] \]

and thus \( b^\mu(\xi^i(t)) \equiv 0, \quad 0 < t < x_i \) and finally \( b^\mu(\xi^i(x_i)) = b^\mu(\xi^{i+1}(0)) = 0, \)
hence \( \beta_j(x) = 0 \) for \( j = 1, 2, \cdots \mu \) and \( W(x) = \sum_{j=\mu+1}^{n} \beta_j(x) Y_j^i(x), \) finishing the proof of lemma 2.

We continue by introducing the distributions \( \Delta^i \) on \( U, i = 1, \cdots n, \)
\( \Delta^i(x) = \text{span}\{ Y^{a+1-i}(x), \cdots Y^a(x) \}. \) We claim:

**Claim 0.** \( L^{(l)}(x) = \Delta^{n-l+1}(x), \) where \( l \) is such that \( r_{l-1} < j \leq r_l, \)
\( j = 1, 2, \cdots p-1. \)

**Proof:** We only have to consider \( r_{l-1} < r_l. \) By choice of the vectorfields \( Y^i, \)
\( L^{(r_l)}(0) = \text{span}\{ Y^i(0), \cdots Y^n(0) \} = \Delta^{n-l+1}(0). \) From \( j \leq r_l, \) \( L^{(j)} \supseteq L^{(r_l)}. \) From
\( r_{l-1} < j, \) \( \dim L^{(j)}(0) = \dim L^{(r_l)}(0), \) and by lemma 2, the dimensions of \( L^{(l)} \)
evaluated at a point in \( U \) are constant, thus

\[
L^{(l)}(x) = \Delta^{n-l+1}(x) \quad \text{for all} \quad x \in U.
\]

**Claim 1.** Relative to the coordinates \( (x_1, \cdots x_n), \) the vectorfields \( Y^i \) are of the form:

\[
Y^i(x) = \frac{\partial}{\partial x_i} + \sum_{j=\mu+1}^{n} c_{ij}(x) \frac{\partial}{\partial x_j}, \quad i = 1, \cdots n
\]

with analytic functions \( c_{ij} \) vanishing at \( x = 0. \) (For notational convenience let
\( c_{ij} \equiv 0 \) for \( i = 1, \cdots n; j = i+1, \cdots, s_{i+1}-1. \))

**Proof.** (By induction on \( i, \) \( i \) decreasing). \( Y^n \equiv \frac{\partial}{\partial x_n} \) is immediate.

Next suppose claim 1 is true for \( l = i+1, i+2, \cdots n. \) Compute
\[
\frac{\partial}{\partial x_i} = (\exp x_n Y^n) \ast \cdots \ast \circ (\exp x_{i+1} Y^{i+1}) \ast Y^i \circ (\exp x_i Y^i) \circ \cdots \circ (\exp x_1 Y^1)(0)
\]

\[
= Y^i + \sum_{|\nu| \geq 0} \frac{x^{\nu}}{\nu!} (\text{ad}^\nu x Y^n, \cdots ; (\text{ad}^i x Y_{i+1}, Y^i)) \cdots )(x)
\]  

(16)

where \( \nu = (\nu_{i+1}, \cdots, \nu_n) \). Each iterated bracket in the sum contains a factor \([Y^j, Y^i] \) for some \( i \geq i+1 \) and therefore lies in \( L^{(r_{i+1} + 1)} \) which is a \( C^\omega \)-module of \( \{Y^{s_{i+1}}, \cdots, Y^n \} \).

Note that since \( L \) is nilpotent, the sum (16) is finite, and using the induction hypothesis, claim 1 is immediate.

Also immediate from equation (16) is

**Claim 2.** Each \( c_{ij}(x) \) in (15) is a sum of terms of the form \( x_l c_{ij}^l(x) \) with \( i < l \leq n \) and \( c_{ij}^l \) analytic (when identifying \( c_{ij}(x) \) with its Taylor series about \( x = 0 \)).

Observe that by claim 1, \( \left\{ \frac{\partial}{\partial x_l}, \cdots, \frac{\partial}{\partial x_n} \right\} \) is an abelian basis for \( \Delta^{n-i+1}, l = 1, 2, \cdots, n \). (Each \( \Delta^l \) is involutive.)

**Claim 3.** Each of the functions \( c_{ij}(x) \) depends on \( x_1, \cdots, x_i \) only, where \( l \) is such that \( s_l \leq j \) (i.e. in particular \( l < j \)), \( i = 1, 2, \cdots, n \).

**Proof.** (By induction on \( l \), \( l \) decreasing)

Since \( Y^n \equiv \frac{\partial}{\partial x_n} \in L^{(r-1)}, [Y^n, Y^i] = 0, i = 1, \cdots, n \), and therefore

\[
0 \equiv [Y^i, \frac{\partial}{\partial x_n}] = \sum_{n=s_{i+1}}^n \frac{\partial c_{ij}}{\partial x_n} \frac{\partial}{\partial x_j},
\]

and hence \( \frac{\partial c_{ij}}{\partial x_n} \equiv 0; i = 1, 2, \cdots, n; j = s_{i+1}, \cdots, n. \)
Next suppose $\frac{\partial c_{ij}}{\partial x_h} \equiv 0$ for all $h > l$ if $s_h > j$. To show $\frac{\partial c_{ij}}{\partial x_l} \equiv 0$ if $s_l > j$, $i = 1, \cdots, n$, $j = s_{l+1}, \cdots, n$, we only have to investigate $l < j$ and from $[Y^l, Y^l] \in L^{(r_l+1)} \subseteq \Delta^{n-s_l+1}$

$$[Y^l, Y^l] = \sum_{j=s_{l+1}}^{s_l-1} \frac{\partial c_{ij}}{\partial x_l} \frac{\partial}{\partial x_j} - \sum_{h=s_{l+1}}^{n} \sum_{j=s_{l+1}}^{s_l-1} c_{ih} \frac{\partial c_{ij}}{\partial x_h} \frac{\partial}{\partial x_j} \mod \Delta^{n-s_l+1}$$

By the induction hypothesis, the double sum vanishes, giving

$$\frac{\partial c_{ij}}{\partial x_l} \equiv 0, \; i = 1, \cdots, n, \; j \leq s_l - 1$$

and thus $c_{ij} = c_{ij}(x_1, \cdots, x_h)$ with $s_h \leq j$.

**Proof of Lemma 3.** We will show that $(Z^g \psi)(0) = 0$ for all $\alpha \cdot r > s$ implies $(D^s \psi)(0) = 0$ for all $\alpha \cdot r > s$, the lemma then being immediate from the Taylor series expansion of $\psi$ about 0. Let $m_s = \{\alpha = (\alpha_1, \cdots, \alpha_n) : \alpha \cdot r > s\}$ and define an order relation on all multi-indices $\alpha \in (\mathbb{Z}_+^n)$ by

$$\alpha > \beta \iff \begin{cases} (|\alpha| > |\beta|) \\ \text{or} \\ (|\alpha| = |\beta| \text{ and there is an } i \in \{1, \cdots, n\} \\ \text{such that } \alpha_i > \beta_i \text{ and } \alpha_j = \beta_j \text{ for all } j < i) \end{cases}$$

(We do not work with lexicographical ordering, instead we introduced 
"$|\alpha| > |\beta| \text{ or } \cdots"$ in order to keep the sections $S_\alpha$: $\{\beta \in m_s : \beta < \alpha\}$ finite.)

Let $m_0 = \min\{m \in \mathbb{Z}_+^n : mr_n > s\}$ and if $n_0$ is the smallest integer such that $r_{n_0} = r_n$, define $\alpha^0 \in m_s$ by $\alpha^0_{n_0} = m_0$ and $\alpha^0_j = 0$ for all $j \neq n_0$. Then $\alpha^0$ is the smallest element in $m_s$.

We prove the lemma by induction on $\alpha$, starting with $\alpha^0$. Since
\[ r_n = r_n, \quad Z^n = \frac{\partial}{\partial x_n} \quad \text{and thus} \quad 0 = (Z^\alpha \psi)(0) \quad \text{immediately gives} \quad (D^\alpha \psi)(0) = 0. \]

Next suppose \((D^\beta \psi)(0) = 0\) for all \(\beta \in \mathcal{M}_\alpha\) with \(\beta < \alpha\) has already been shown and consider \(Z^\alpha \psi\). To do further analysis, write each \(Z^i\) occurring in \(Z^\alpha\) in the form
\[
Z^i = \frac{\partial}{\partial x_i} + \sum_{j=i+1}^n a_{ij} \frac{\partial}{\partial x_j}
\]
and use the linearity of the differential operators to write \((Z^\alpha \psi)\) as a sum of terms of the form
\[
(a_g a_{f_1} \frac{\partial}{\partial x_{f_1}} (a_g a_{f_2} \frac{\partial}{\partial x_{f_2}} (\cdots \frac{\partial}{\partial x_{f_{|\mu|}}} (a_g a_{f_{|\mu|}} \frac{\partial}{\partial x_{f_{|\mu|}}}) \cdots ))
\]
(17)
(with \(\alpha_i \equiv 1, \quad i = 1, \cdots n,\) for notational ease), and apply Leibniz' rules repeatedly until no derivative acts on any product anymore. The result is
\[
(Z^\alpha \psi) = \sum_{\mu} p_\mu(x)(D^\mu \psi)
\]
(18)
where the \(p_\mu(x)\) are sums of products of partial derivatives of the coefficients \(a_{gf}(x)\).

We claim \(p_\alpha \equiv 1, \mu\) ranges over all multi-indices \(\mu \leq \alpha\) and \(p_\mu \equiv 0\) if \(\mu \notin \mathcal{M}_\alpha\), i.e., \(\mu < \alpha^0\).

To verify this claim observe that the order \(|\mu|\) of the derivative \((D^\mu \psi)\) cannot be larger than the total number \(|\alpha|\) of \(Z^i\) applied to \(f\), hence \(|\mu| \leq |\alpha|\). Next observe that only if \(g_i = f_i\) for \(i = 1, 2, \cdots |\alpha|\) in (17), one may get \(\mu = \alpha\) in (18), hence \(p_\alpha \equiv 1\). Now fix one of those \(\mu\) in (18) with \(|\mu| = |\alpha|\), but \(\mu \neq \alpha\). Let \(k = \min\{j: \mu_j \neq \alpha_j\}\). Then \(\mu_1 = \alpha_1\) implies \(g_i = f_i = 1\) for \(i = 1, 2, \cdots \alpha_1\), and inductively \(\mu_j = \alpha_j\) for \(j < k\) implies
\( g_i = f_i \) for \( i = 1, 2, \ldots, \alpha_1 + \alpha_2 + \ldots + \alpha_{k-1} \). But now \( \alpha_k \neq \mu_k \) is only possible if \( \mu_k < \alpha_k \) and thus \( \mu < \alpha \).

Finally we investigate those \( \mu \) with \( \mu < \alpha^0 \), i.e. \( \mu \notin \mathcal{M}_\alpha \). We consider one fixed summand resulting from applying Leibniz's rule repeatedly to one fixed term of form (17). Let \( I \subseteq \{1, 2, \ldots, |\alpha|\} \) be the set of all indices \( i \), such that by Leibniz' rule \( \frac{\partial}{\partial x_{f_i}} \) acts on \( \psi \) only, and similarly \( J = \{1, 2, \ldots, |\alpha|\} \setminus I \), the set of all indices \( i \) such that \( \frac{\partial}{\partial x_{f_i}} \) acts on some \( \alpha_{s_k} t_{s_k}, i < h \leq |\alpha| \).

Using \( \alpha \in \mathcal{M}_s, \mu \notin \mathcal{M}_s \) or equivalently \( \alpha \cdot r > s \geq \mu \cdot r \) gives

\[
\alpha \cdot r = \sum_{i=1}^{\alpha} r_{s_i} > s \geq \mu \cdot r = \sum_{i \in I} r_{f_i} \quad \text{and thus}
\]

\[
\sum_{i \in J} r_{f_i} = \sum_{i=1}^{\alpha} r_{f_i} - \sum_{i \in I} r_{f_i} > \sum_{i=1}^{\alpha} (r_{f_i} - r_{s_i}),
\]

which means the total degree \( \sum_{i \in J} r_{f_i} \) of the derivatives acting on the coefficients \( \alpha_{s_k} t_{s_k} \) is higher than the total degree \( \sum_{i=1}^{\alpha} (r_{f_i} - r_{s_i}) \) of the product \( \sum_{h=1}^{k} \alpha_{s_h} t_{s_h} \) and thus we may conclude \( p_\mu = 0 \) if \( \mu \cdot r \leq s < \alpha^0 \cdot r \leq \alpha \cdot r \).

**Caution:** Here we have \( \mu \cdot r < \alpha^0 \cdot r \leq \alpha \cdot r \). This need not imply \( \mu < \alpha \) or \( \mu < \alpha^0 \) as illustrated in the following example.

If \( r = (1, 2, 6), s = 3 \), then the only multi-indices not in \( \mathcal{M}_3 \) are \((0,0,0), (0,1,0), (1,0,0) \) and \((2,0,0)\). \( \alpha^0 = (0,0,1) \) and for all \( \mu \in \mathcal{M}_3 \setminus \{(0,0,0)\} \) \( \mu \cdot r \leq s \leq \mu \cdot \alpha^0 \), but \( \mu > \alpha^0 \). (We need the ordering \( \alpha < \beta \) on the multi-indices only for this induction; whereas the degrees \( \alpha \cdot r \) w.r.t. \( \delta_i \) have geometric meaning.)
To conclude the proof of lemma 3, use $p_\alpha \equiv 1, p_\mu \equiv 0$ for $\mu \not\in \mathcal{M}_g$ and that the sum only ranges over $\mu$ with $\mu < \alpha$, to obtain

$$0 = (Z^\alpha \psi)(0) = (D^\alpha \psi)(0) + \sum_{\mu \in \mathcal{M}_g, \mu < \alpha} p_\mu (D^\mu \psi)(0).$$

By the induction hypothesis the sum is zero and thus $(D^\alpha \psi)(0) = 0$.

We continue with

**Claim 4.** Each of the coefficients $c_{ij}(x)$ in (15) is a polynomial in $(x_1, \cdots, x_n)$ of degree $r_j - r_i$ w.r.t. $\delta_i$.

**Proof:** We will show $(D^\alpha c_{ij})(0) = 0$ for all $\alpha = (\alpha_1, \cdots, \alpha_n)$ with $\alpha \cdot r > r_j - r_i$. The claim then is immediate from the Taylor series expansion of $c_{ij}$ about $x = 0$.

Almost the only fact we can use apart from the previous claims and lemmata is that

$$(ad^\alpha Y^\alpha, Y^i) = 0 \mod \Delta^{r_\alpha - q_i + 1}, \text{ if } r \cdot \alpha + r_i > r_j.$$

The problem, though, is that $(ad^\alpha Y^\alpha, Y^i)$ consists of differences of products of derivatives of many coefficients $c_{ij}$. But by a "tricky" use of the "triangular" structure of the matrix $C = (c_{ij})$ using the claims (1) through (3) we will be able to single out one specific derivative $Y^\alpha \left( \frac{\partial c_{ij}}{\partial x_k} \right)$ at a time and from

$$Y^\alpha \left( \frac{\partial c_{ij}}{\partial x_k} \right)(0) = 0 \text{ if } \alpha \cdot r + r_k > r_i - r_j$$

by virtue of lemma 3, we may conclude that $\frac{\partial c_{ij}}{\partial x_k}$ is of degree $r_i - r_j - r_k$ and finally that $c_{ij}$ is of degree $r_i - r_j$.

We will work through all $\frac{\partial c_{ij}}{\partial x_k}$ by three nested inductions.
We identify $c_{ij}(x)$ with its Taylor series about $x = 0$. The first induction is on $j$, $j$ increasing. The smallest $j$ occurring is $j = s_{i+1} = s_2$.

Letting $j = s_2$, then for $i = 1, 2, \cdots j-2$, $c_{ij}(x)$ depends on $(x_1, \cdots, x_{j-1})$ only (by claim 3) and each nonvanishing summand of $c_{ij}(x)$ must contain a factor $x_k$ with $i < k \leq j-1$ (by claim 2). Therefore it suffices to show that $\frac{\partial c_{ij}}{\partial x_k}(x)$ is a polynomial of degree $r_j - r_i - r_k$ for all $k$ s.t. $i < k \leq j-1$.

The second induction is on $k$, $k$ decreasing. We start with $k = j-1 = s_2-1$, and since $s_{k+1} = s_j > j$ $Y^{i-1} = \frac{\partial}{\partial x_{j-1}} \mod \Delta^{a-j}$, and we compute

$$[Y^i, Y^{i-1}] = \left[ \frac{\partial c_{ij}}{\partial x_{j-1}} \right] \frac{\partial}{\partial x_j} \mod \Delta^{a-j}$$

and

$$(a_d a \cdot [Y^i, Y^{i-1}]) = (-1)^{i} \cdot D^{\alpha} \left[ \frac{\partial c_{ij}}{\partial x_{j-1}} \right] \frac{\partial}{\partial x_j} \mod \Delta^{a-j}$$

where $\alpha = (\alpha_1, \cdots, \alpha_{j-1})$. (Note: there are no nonconstant $c_{gh}$ with $h < j$.)

For all $\alpha = (\alpha_1, \cdots, \alpha_{j-1})$ with $\alpha \cdot r + r_i + r_{j-1} > r_j$, this expression vanishes modulo $\Delta^{a-j}$, and thus $\frac{\partial c_{ij}}{\partial x_{j-1}}$ is a polynomial of degree $r_j - r_i - r_{j-1}$.

To continue the second induction, suppose $\frac{\partial c_{ij}}{\partial x_l}$ is a polynomial of degree $r_j - r_i - r_l$ for all $l$ such that $k < l < j-1$, and we consider $\frac{\partial c_{ij}}{\partial x_k}$.

By claim 2, we may restrict our considerations to $i < k$ and compute:
\[ [Y^i; Y^k] = \left( \frac{\partial c_{ij}}{\partial x_k} - \frac{\partial c_{kj}}{\partial x_i} \right) \frac{\partial}{\partial x_j} \mod \Delta^{n-i} \]

and for \( \alpha = (\alpha_1, \cdots, \alpha_{j-1}) \) such that \( \alpha \cdot \tau + r_k + r_i > r_j \)

\[ 0 = (\text{ad}^\alpha Y^\alpha; [Y^i; Y^k]) \mod \Delta^{n-i} \]

\[ = (-1)^{|\alpha|} \left( \left( Y^\alpha \frac{\partial c_{ij}}{\partial x_k} - \frac{\partial c_{kj}}{\partial x_i} \right) \frac{\partial}{\partial x_j} \right) \mod \Delta^{n-i} \]

\[ = (-1)^{|\alpha|} \left( D^\alpha \frac{\partial c_{ij}}{\partial x_k} - D^\alpha \frac{\partial c_{kj}}{\partial x_i} \right) \frac{\partial}{\partial x_j} \]

(since there are no nonconstant \( c_{gh} \) with \( h < j \)).

If \( \alpha_l \neq 0 \) for some \( l > k \), then interchange the order of differentiation in the last term, i.e. choose \( \beta \) so that \( D^\alpha \frac{\partial}{\partial x_l} = D^\beta \frac{\partial}{\partial x_l} \) and get

\[ D^\alpha \frac{\partial c_{kl}}{\partial x_l} (0) = D^\beta \frac{\partial c_{kl}}{\partial x_k} (0) = 0 \] by induction (2) hypothesis since

\[ \beta \cdot \tau + r_l = \alpha \cdot \tau + r_i > r_j - r_k \]

If \( \alpha_l = 0 \) for all \( l > k \), then each not yet vanishing summand in \( D^\alpha \left( \frac{\partial c_{kl}}{\partial x_k} \right) \) still contains a factor \( x_l \) for some \( l > k \) and therefore vanishes at \( x = 0 \). In both cases we obtain \( D^\alpha \frac{\partial c_{kl}}{\partial x_k} = 0 \), thus finishing the second induction and the start of the first induction.

Before continuing induction (1), we introduce some notation.

Let \( j \in \{a_2, \cdots, a_{s-1} \} \) fixed and define for \( 1 \leq f \leq g \leq j \), \( s \in \mathbb{Z}^+ \),

\[ \mathcal{J}^f(\pm s) = \left\{ \sum_{l=f}^{g} p_l(x) \frac{\partial}{\partial x_l} : p_l \in H_{r_{l-s}} \right\} \]
and for \(1 \leq f \leq g \leq j-1\)

\[
\mathcal{F}\phi(-s) = \left\{ \phi(x) \frac{\partial}{\partial x_j} : \phi(x) = \phi(x_1, \ldots, x_{j-1}) \right\} \text{ is analytic, each summand of } \phi \text{ contains a factor } x_h \text{ for some } h \in \{f+1, \ldots, j-1\}
\]

and \(\frac{\partial \phi}{\partial x_h}(x)\) is a polynomial of degree \(r_j - r_h - s\) if \(h \in \{g+1, \ldots, j-1\}\).

(We assume \(\phi\) is expanded in its Taylor series about \(x = 0\).)

Then one easily verifies

\[
\begin{cases}
\mathcal{F}\phi(-s); \mathcal{F}\phi(-s') \subseteq \mathcal{F}_{\max(f+1)}(s-s') \\
\text{(as a consequence of } \frac{\partial p_l}{\partial x_h} \equiv 0 \text{ if } h \geq l) \text{ and} \\
[\mathcal{F}\phi(-s); \mathcal{F}\phi^h(-s')] \subseteq \mathcal{F}^h(s-s') \\
[\mathcal{F}\phi(-s); \mathcal{F}\phi_l(-s')] \subseteq \mathcal{F}^l(s-s') \\
[\mathcal{F}\phi(-s); \mathcal{F}^l(-s')] = 0, \\
\mathcal{F}\phi(-s)(0) = 0 \text{ for all } s > 0; \ f, g = 1, 2, \ldots, j-2 \\
\mathcal{F}\phi(-s) = 0 \text{ if } s > r_j, 1 \leq f \leq g \leq j \\
\mathcal{F}\phi(-s) \subseteq \mathcal{F}^h(-s) \text{ if } f \leq h.
\end{cases}
\] (19)

To continue induction (1) suppose \(c_{ij}(x)\) is a polynomial of degree \(r_i - r_l\) for all \(l < j\) and consider \(c_{ij}(x)\) with \(i\) such that \(s_i+1 \leq j\). By claim 3, \(c_{ij}(x)\) does not depend on \(x_h\) with \(h \geq j\) for all \(l \leq j\) and therefore we may and will do all the following calculations modulo \(\Delta^{a-j}\).

By induction (2') on \(k\), \(k\) decreasing, we show \(\frac{\partial c_{ij}}{\partial x_k}\) is a polynomial of degree \(r_j - r_i - r_k\) for \(i < k \leq j\). To save unnecessary work start with \(k = j\) and by claim (3), \(\frac{\partial c_{ij}}{\partial x_j} = 0\). Next suppose \(\frac{\partial c_{ij}}{\partial x_l}\) is a polynomial of degree \(r_j - r_i - r_l\) for all \(l > k\) (and \(i < l \leq j\)). By claim (2) we only have
to consider \( k > i \).

By induction (3) on \( \alpha \), the length \( |\alpha| \) of \( \alpha \) increasing we show

**Claim 5.** For each \( \alpha = (\alpha_1, \ldots, \alpha_{j-1}) \)

\[
(-)^{|\alpha|}([d^\alpha Y^\alpha, [Y^i, Y^k]]) = \left(Y^\alpha \frac{\partial c_{ij}}{\partial x_k} \right) \frac{\partial}{\partial x_j} \in \mathcal{S}_{k+1}^l(-s) + \mathcal{F}_k^k(-s)
\]

where \( s = r \cdot \alpha + r_k + r_i \).

Observe that \( Y^k = \left( \frac{\partial}{\partial x_k} + \sum_{l=k+1}^{j-1} c_{kl} \frac{\partial}{\partial x_l} \right) + c_{kj} \frac{\partial}{\partial x_j} \), where by induction (1) hypothesis the expression in parenthesis is in \( \mathcal{S}_{k+1}^l(-r_k) \) and by claim (2) and induction (2') hypothesis

\[
c_{kj} \frac{\partial}{\partial x_j} \in \mathcal{S}_k^k(-r_k) + \mathcal{F}_k^k(-r_k) \text{ and similarly}
\]

\[
Y^i - c_{ij} \frac{\partial}{\partial x_j} \in \mathcal{S}_j^j(-r_i),
\]

by induction (2') hypothesis

\[
\left( \sum_{l=k+1}^{j-1} c_{kl} \frac{\partial}{\partial x_l} \right) c_{ij} \frac{\partial}{\partial x_j} \in \mathcal{S}_k^k(-r_i-r_k) \text{ and thus}
\]

\[
[Y^i, Y^k] - \frac{\partial c_{ij}}{\partial x_k} \frac{\partial}{\partial x_j} \in \mathcal{S}_{k+1}^l(-r_i-r_k) + \mathcal{F}_k^k(-r_k)
\]

using the relations (19). (We get \( \mathcal{S}_{k+1}^l \) instead of \( \mathcal{S}_k^l \) because \( c_{kk} \equiv 1 = \text{const.} \). This starts induction (3) with \( |\alpha| = 0 \).

Next suppose claim 5 is true for all \( \beta \) with \( |\beta| < a_0 \) and fix \( \alpha = (\alpha_1, \ldots, \alpha_{j-1}) \) with \( |\alpha| = a_0 \), letting \( l = \min \{ h : \alpha_h \neq 0 \} \), and define the multi-index \( \beta = \beta(\alpha) \) by \( \beta_i = \alpha_i - 1 \) and \( \beta_h = \alpha_h \) for all \( h \neq l \). Then by
induction (3) hypothesis

\[ (-l^\beta)(\text{ad}^\beta Y^\beta_i [Y^i_i; Y^k]) - Y^\beta \frac{\partial c_{ij}}{\partial x_k} \frac{\partial}{\partial x_j} \in \mathcal{J}_{k+1}(-s) + \mathcal{F}_k(-s) \]

with \( s = r \cdot \beta + r_i + r_k \).

Furthermore \( (\text{ad}^\alpha Y^\alpha_i; \cdot) = [Y^l_i; (\text{ad}^\beta Y^\beta_i; \cdot)] \), (that is why we chose \( l \) to be the smallest integer such that \( \alpha_i \neq 0 \).)

Finally \( Y^l_i \in \mathcal{J}_l(-r_l) + \mathcal{F}_k(-r_l) \) and using the relations (19) we obtain:

\[ (-l^\alpha)(\text{ad}^\alpha Y^\alpha_i [Y^i_i; Y^k]) - \left( Y^\alpha \frac{\partial c_{ij}}{\partial x_k} \right) \frac{\partial}{\partial x_j} \]

\[ = -[Y^l_i; (-l^\beta)(\text{ad}^\beta Y^\beta_i [Y^i_i; Y^k])] - \left( Y^\beta \frac{\partial c_{ij}}{\partial x_k} \right) \frac{\partial}{\partial x_j} \]

\[ \in [\mathcal{J}_l(-r_l) + \mathcal{F}_k(-r_l); \mathcal{J}_{k+1}(-s) + \mathcal{F}_k(-s)] \]

\[ \subseteq \mathcal{J}_{k+1}(-s-r_l) + \mathcal{F}_k(-s-r_l) \]

with \( s = r \cdot \beta + r_i + r_k \), thus finishing induction (3) and the proof of claim 5.

To finish both induction (1) and induction (2') consider \( \alpha = (\alpha_1, \cdots, \alpha_{j-1}) \) such that \( \sigma = \alpha \cdot r + r_i + r_k > r_j \). Then \( \mathcal{J}_{k+1}(-\sigma) = 0 \), \( \mathcal{F}_k(-\sigma)(0) = 0 \), and \( (\text{ad}^\alpha Y^\alpha_i [Y^i_i; Y^k])(0) = 0 \mod \Delta^{n-j} \). Therefore \( Y^\alpha \frac{\partial c_{ij}}{\partial x_k} (0) = 0 \). By lemma 3, \( \frac{\partial c_{ij}}{\partial x_k} \) is a polynomial of degree \( r_j - r_i - r_k \) for \( i < k \leq j-1 \) and thus \( c_{ij} \) is a polynomial of degree \( r_j - r_i \).

1.4. Proof of the theorem:

We show that relative to the coordinates \( (x_1, \cdots, x_n) \) and the dilation
\( \delta_i(x) \) defined in (6) and (7), the fields \( X^i \) are of the form

\[
X^i(x) = \sum_{j=1}^{n} a_{ij}(x) \frac{\partial}{\partial x_j}, \quad 1 \leq i \leq \kappa,
\]

with \( a_{ij}(x) \) being polynomials of degree \( r_j - 1 \). Since the proof is exactly the same for each \( i \), we will suppress the index \( i \), and just write \( X \) for \( X^i \).

Since \( Y^1(0), \ldots, Y^n(0) \) (as defined in §1.3) are linearly independent there are analytic functions \( b_j, \ j = 1, \ldots, n \) such that

\[
X(x) = \sum_{j=1}^{n} b_j(x) Y^j(x)
\]

(20)

We show by induction on \( j, j \) increasing, that \( b_j \) is a polynomial in \( x \) of degree \( r_j - 1 \).

Start with \( j = 1 \) and compute for \( 1 \leq i \leq n \)

\[
[X, Y^i] = (Y^i b_1) Y^1 + \sum_{l>1} ((Y^i b_l) Y^l - b_l [Y^l, Y^i])
\]

\[
= (Y^1 b_1) Y^1 \mod \Delta^{n-2},
\]

giving \( Y^1 b_1 \equiv 0 \) for \( i = 1, 2, \ldots, n \) and thus \( b_1(x) = \text{const} \).

Next suppose \( b_j(x) \) is a polynomial of degree \( r_j - 1 \) for all \( l < j \).

For \( \alpha = (\alpha_1, \ldots, \alpha_n) \) such that \( \alpha \cdot r > r_j - 1 \) compute:

\[
0 = (\text{ad}^\alpha Y^\alpha, X) \mod \Delta^{n+1-s_j}
\]

\[
= \sum_{l=1}^{j-1} \left( \text{ad}^\alpha Y^\alpha; b_l \frac{\partial}{\partial x_l} \right) + \left( \text{ad}^\alpha Y^\alpha; b_j \frac{\partial}{\partial x_j} \right) \mod \Delta^{n+1-s_j}
\]

By the inductions hypotheses, the sum vanishes identically modulo \( \Delta^{n+1-s_j} \) and since \( \frac{\partial c_{k,l}}{\partial x_h} = 0 \) for \( h \geq l \),
\[
\left( \text{ad}^\alpha Y^\alpha, b_j \frac{\partial}{\partial x_j} \right) = (\alpha \cdot \Gamma)(Y^\alpha b_j) \frac{\partial}{\partial x_j} \mod \Delta^{n+1-s_j}.
\]

Now \((Y^\alpha b_j) \equiv 0\) for all \(\alpha \cdot r > r_j - 1\), and by lemma 3, \(b_j\) is a polynomial of degree \(r_j - 1\), thus finishing the induction.

Finally to express \(X^i\) in terms of the abelian basis \(\frac{\partial}{\partial x_1}, \cdots \frac{\partial}{\partial x_n}\), combine (15) and (20) and use claim 4:

\[
X^i(x) = \sum_{j=1}^{n} b_{ij}(x) Y^j(x) = \sum_{j=1}^{n} \sum_{l=1}^{n} b_{ij}(x) c_{jl}(x) \frac{\partial}{\partial x_l}
\]

\[
= \sum_{l=1}^{n} a_{il}(x) \frac{\partial}{\partial x_l}\quad \text{with} \quad a_{il}(x) = \sum_{j=1}^{n} b_{ij}(x) c_{jl}(x)
\]

and \(a_{il}\) is of degree \((r_j - 1) + (r_l - r_j) = r_l - 1\) and thus \(X^i(x)\) is of degree -1 w.r.t. the dilation \(\delta_i(x), i = 1, \cdots, \kappa.\)
CHAPTER II
LOCAL CONTROLLABILITY OF NONLINEAR SYSTEMS

2.1 Introduction

In this part we will consider single-input affine control systems of the form

\[ \dot{x} = X^0(x) + uX^1(x), \quad x(0) = 0, \quad X^0(0) = 0, \]  

(1)

where \( x \in \mathbb{R}^n \), \( X^0 \) and \( X^1 \) are real analytic vectorfields on \( \mathbb{R}^n \), and \( u \) is an admissible control, that is \( u: [0, t_0] \rightarrow [-\epsilon_0, \epsilon_0] \) is measurable, \( x(\cdot, u) \) denotes the solution of (1) with control \( u \). The attainable set \( G_{\epsilon_0}(t) \) for \( t \in [0; t_0] \) and control bound \( \epsilon_0 > 0 \) is defined as

\[ G_{\epsilon_0}(t) = \{ x(t,u): |u(\cdot)| \leq \epsilon_0 \}. \]

The system (1) is small time locally controllable (STLC) if for every \( t, \epsilon > 0 \) \( G_\epsilon(t) \) contains the uncontrolled rest solution \( x \equiv 0 \) in its interior.

For two vectorfields \( V, W \) define their Lie product \( [V; W] = (\text{ad} V; W) = WV - VW \) and inductively \( (\text{ad}^{k+1} V; X) = (\text{ad} V; (\text{ad}^k V; W)), \ k = 1, 2, \ldots; \ \text{ad}^0 V; W) = W \). \( L = L(X^1, X^0) \) is the Lie algebra generated by \( X^0 \) and \( X^1 \), for \( k, l \in \mathbb{Z}_0^+ \), \( L^{(k,l)}(X^1, X^0) \) the subspace of \( L \) spanned by all Lie monomials containing \( k \) factors \( X^1 \) and \( l \) factors \( X^0 \) and similarly \( \mathcal{J}^k(X^1, X^0) \) the subspace of \( L \) spanned by all Lie monomials.
containing \( k \) factors \( X^1 \). We will call a vectorfield \( V \) of type \((k,l)\) if \( V \in \mathcal{L}^{(k,l)}(X^1, X^0) \).

For any set \( \mathcal{C} \) of vectorfields, \( \mathcal{C}(0) = \{ V(0) \in \mathbb{R}^n : V \in \mathcal{C} \} \).

In the analytic category \( \dim \mathcal{L}(X^1, X^0)(0) = n \) is a necessary and sufficient condition that \( \text{int} G_\epsilon(t) \neq \emptyset \) for all \( \epsilon, t > 0 \) [29]. We assume this holds throughout.

The problem is to give necessary and sufficient conditions for STLC of (1) in terms of subsets of \( \mathcal{L}(X^1, X^0)(0) \). The presently known results in this direction are:

**Theorem 0.1.** A sufficient condition that (1) is STLC is that \( \dim \mathcal{A}(0) = n \).

Theorem 0.1 is often referred to as the **linear test**.

**Theorem 0.2.** A sufficient condition that (1) is STLC is that for each \( k \in \mathbb{Z}^+ \)

\[
\mathcal{A}^k(0) \subseteq \sum_{j=1}^{\frac{n}{2} - 1} \mathcal{A}^j(0).
\]

This was shown in [16, 19] in the case \( n = 2 \) and extended to arbitrary dimension in [32] and is usually referred to as the **Hermes condition**.

**Theorem 0.3.** A sufficient condition that (1) is STLC is that there is a weight \( \theta \in [0,1] \) such that whenever \( X^* \) is a product of \( k \) factors \( X^1 \) and \( l \) factors \( X^0 \) with \( k \) even and \( l \) odd, then \( X^*(0) \) can be written as a linear combination of Lie products \( X^{*j} \) with \( k_j \) factors \( X^1 \) and \( l_j \) factors \( X^0 \), evaluated at zero, such that for each \( j \)

\[
k_j + \theta l_j < k + \theta l.
\]
This theorem was proven in [33] in a form applying to multi-input systems. In §1 and §4 we will discuss theorem 0.3 in detail.

**Theorem 0.4a.** A necessary condition that (1) is STLC is that

\[(\text{ad}^2 X^1; X^0)(0) \in \mathcal{A}^j(0).\]

This was first shown in [16] for the case \(n = 2\) and finally generalized in [26, 27] to give:

**Theorem 0.4.** A necessary condition that (1) is STLC is that for all \(k \in \mathbb{Z}^+\)

\[(\text{ad}^{2k} X^1; X^0)(0) \in \sum_{j=1}^{2k-1} \mathcal{A}^j(0).\]

Clearly theorem 0.3 contains theorem 0.2 as a special case (when \(\theta = 0\) (and theorem 0.2 in turn contains theorem 0.1 as a special case). Thus, just comparing theorems 0.3 and 0.4 shows how wide the gap between necessary and sufficient conditions for STLC of (1) is. This gap is precisely the place where this paper will fill in one more necessary condition similar to theorem 0.4a and discuss several particular control systems where all known necessary conditions hold but all known sufficient conditions fail. These examples illustrate what complications may occur with higher order Lie brackets and give some idea what a general theorem about STLC must include. Most important the methods developed to show that these systems are STLC — especially the new class of control variations introduced in §2.5.4 — might be useful tools for proving a general theorem in the future.
2.2 Discussion of theorems 0.1 through 0.4, motivation for and statement of the new results.

The philosophy behind this approach is that all the information about local properties of system (1) such as STLC is contained in the values of elements of \( L(X^1, X^0) \) evaluated at zero.

In many different approaches one observes that Lie brackets in \( L^{(k,l)}(X^1, X^0) \) usually are associated to coefficients involving \( e^k t^{k+l} \), e.g., in the Chen series [32], when expanding via the Campbell-Baker-Hausdorff formula, or very clearly in the proof of theorem 0.4 [33] and in the scaling lemma (§2.3.1).

The basic idea is that, when choosing \( \epsilon \) and \( t \) sufficiently small, the lower order brackets shall dominate the higher order brackets, where of course the notion of an ordering of the brackets has to be made more precise.

We will discuss certain families of control variations which correspond to letting \( \epsilon \) go to zero or letting both \( \epsilon \) and \( t \) go to zero, show what ordering of the brackets they induce, and how they relate to the theorems in §2.1.

However, as pointed out in [22], with different control variations one can achieve that if \( X^r \in L^{(k,l)} \), then \( (\text{ad}^\nu X^0, X^r) \in L^{(k,l+v)} \) also occurs with a coefficient involving \( e^k t^{k+l} \), for any \( \nu \in \mathbb{Z}_0^+ \), which of course leads to a different ordering of the Lie brackets in \( L(X^1, X^0) \) and gives results on STLC not covered by theorems 0.1 through 0.4.

Finally in §2.5.4 a very different form of control variations will be introduced which shall lead to completely new orderings of the Lie brackets in \( L(X^1, X^0) \) and hopefully will considerably narrow the gap between necessary and sufficient conditions.

Before continuing, we will precisely say what we mean by a tangent
vector in this paper. A vector $\xi \in \mathbb{R}^n$ is a **tangent vector to** $G_\epsilon(t_1)$ **at zero** (or just tangent vector) if there is a differentiable map $s \mapsto q(s) \in G_\epsilon(t_1)$, $s \geq 0$ with $q(0) = 0$ and $\lim_{s \to 0^+} q'(s) = \xi$. It suffices to consider $q(s)$ of the form $x(t_1, u_s)$ where $s \mapsto u_s$ is a one-parameter family of admissible controls with $u_0 \equiv 0$. For an attainable set $G_\epsilon(t_1)$ the set of all tangent vectors at zero forms a convex cone, denoted $K = K_\epsilon(t_1)$. It is well known that (1) is STLC if and only if $K_\epsilon(t_1) = \mathbb{R}^n$ for all $\epsilon, t_1 \geq 0$ [22].

One of the most common forms of control variations is generated as follows: Given $\epsilon_0, t_0 \geq 0$, choose numbers $|\epsilon_i| \leq \epsilon_0$, $t_i > 0$ with $\sum_{i=1}^r t_i \leq t_0$, $i = 1, \ldots, r$; $r \in \mathbb{Z}^+$, which constitute the perturbation data and will be denoted by $\Gamma$. ($\Gamma = \{r, \epsilon_1, \ldots, \epsilon_r, t_1, \ldots, t_r\}$.) Define a one-parameter family of admissible controls by

$$u_s^\Gamma(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq t_0 - \sum_{i=1}^r t_i \\ s \epsilon_j & \text{if } \left(t_0 - \sum_{i=1}^j t_i\right) < t \leq \left(t_0 - \sum_{i=1}^{j-1} t_i\right) \end{cases}$$

![Figure 1. A control variation.](image)
This leads to

\[ x(t_0, u^\Gamma s) = (\exp t_1 (X^0 + s e_1 X^1)) \cdots \circ (\exp t_r (X^0 + s e_r X^1)) \circ (\exp (t_{r+1} - \sum_{i=1}^r t_i) X^0)(0) \]

which after repeated use of the Campbell-Baker-Hausdorff formula takes the form

\[ x(t_0, u^\Gamma s) = \left\{ \exp \sum_{X^r \in L(X^1, X^0)} a^\Gamma_X (s) X^r \right\}(0) \]

with \( a^\Gamma_X \) polynomials in the perturbation data \( \Gamma \). One easily sees that if \( X^r \in L^{(k,l)}(X^1, X^0) \), then \( a^\Gamma_X (s) \) can be written as \( a^\Gamma_X (s) = s^k b^\Gamma_X \), where \( b^\Gamma_X \) is independent of \( s \) and is polynomial in the perturbation data, i.e.

\[ x(t_0, u^\Gamma s) = \exp \left\{ \sum_{k=1}^\infty s^k \sum_{X^r \in \mathcal{J}^k} b^\Gamma_X X^r \right\}(0) \]

By choosing the perturbation data \( \Gamma \) properly, one might be able to achieve that

\[ \sum_{X^r \in \mathcal{J}^j} b^\Gamma_X X^r (0) = 0 \text{ for all } 1 \leq j < k \text{ and } \sum_{X^r \in \mathcal{J}^k} b^\Gamma_X X^r (0) = \xi \in \mathbb{R}^a. \]

This then gives \( \lim_{s \to 0^+} \frac{d^j}{ds^j} x(t_0, u^\Gamma s) = 0 \) for \( 1 \leq j < k \) and

\( \lim_{s \to 0^+} \frac{d^k}{ds^k} x(t_0, u^\Gamma s) = \xi. \) After reparameterization, i.e. \( \sigma = s^k \) one obtains \( \xi \) as a tangent vector (to \( \iota_0 (t_0) \) at zero).

The induced ordering on the Lie monomials in \( L(X^1, X^0) \) clearly only involves the number of factors \( X^1 \), i.e. \( X^r \in L^{(k,l)} \) is of lower order than \( X^s \in L^{(k',l')} \) iff \( k < k' \).
Control variations of this form have been used extensively in proving theorems (0.1) and (0.2).

A more general class of control variations can be obtained when, after choosing $\Gamma$ as before, one defines for some fixed $\theta \in [0,1]$

$$u^\Gamma(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq \left(t_0 - s^\theta \sum_{i=1}^r t_i\right) \\ s^{1-\theta} \epsilon_j & \text{if } \left(t_0 - s^\theta \sum_{i=1}^j t_i\right) < t \leq \left(t_0 - s^\theta \sum_{i=1}^{j-1} t_i\right) \end{cases}$$

![Diagram of control variation](image)

Figure 2. A control variation associated to $\theta \in [0,1]$.

The typical term in the exponential formula is

$$\cdots \circ \left(\exp t \left(s^\theta X^1 + s \epsilon_1 X^1\right)\right) \cdots$$

When proceeding as before, i.e. expanding by the Campbell-Baker-Hausdorff formula and collecting same powers, one obtains

$$x(t_0, u^\Gamma) = \left\{ \exp \left(\sum_{k,l \geq 0} s^{k+i\theta} \sum_{X^s \in L(k,l)} b^\Gamma_s X^s\right)\right\}(0)$$

where again $b^\Gamma_s$ are polynomials in the perturbation data only.

These control variations induce an ordering on the Lie monomials in
L(X^1, X^0) by assigning to each bracket X^r \in L^{(k,l)} the weight (k + l\theta), and the relation to theorem 0.3 is immediate. Observe that with \( \theta = 0 \) these control variations coincide with those defined before and are closely related to theorem 0.2.

When listing the homogeneous components \( L^{(k,l)} \) as in Fig. 3, this ordering can be easily pictured as follows: For a given \( \theta \in [0,1] \) a bracket of type \( (k', l') \) is of lower weight than a bracket of type \( (k, l) \) if and only if \( (k', l') \) lies above the line through \( (k, l) \) with slope \( \theta \).

![Figure 3. Weight assignment to types of Lie products.](image)

A nice illustration of the usefulness of these control variations is the following system, discussed in [32]:

\[
\begin{cases}
\dot{x}_1 = u \\
\dot{x}_2 = x_1 \\
\dot{x}_3 = x_2^2 + x_1^3.
\end{cases}
\]

One easily computes \( X^1 \equiv \frac{\partial}{\partial x_1} \), \([X^0, X^1](0) = \frac{\partial}{\partial x_2} \) and

\[
\frac{1}{2} (\text{ad}^2[X^0, X^1]; X^0)(0) = -\frac{1}{6} (\text{ad}^3 X^1; X^0)(0) = \frac{\partial}{\partial x_3},
\]

all other brackets being Jacobi related or vanishing at zero. Here theorem 0.2
fails since \((\text{ad}^2 [X^0, X^1]; X^0)(0) \not\in \mathfrak{X}^1(0)\).

But when choosing \(\theta = 1\), \((\text{ad}^2 [X^0, X^1]; X^0)\) has weight 2+1·3 = 5, whereas \((\text{ad}^3 X^1; X^0)\) has weight 3+1·1 = 4 and the possibly "bad" bracket \((\text{ad}^2 [X^0, X^1]; X^0)\) (bad because it contains an even number of factors \(X^1\) and an odd number of factors \(X^0\)) is "neutralized" by the lower order bracket \((\text{ad}^3 X^1; X^0)(0)\) (with \(\theta = 1\)), thus giving STLC.

With \(\theta = 0\) \((\text{ad}^2 [X^0, X^1]; X^0)\) has weight 2, which is less than the weight 3 of \((\text{ad}^3 X^1; X^0)\), thus again explaining the failure of theorem 0.2.

Now suppose \(X^\sigma \in L^{(k_0, \ell_0)}\) is the only bracket (modulo the Jacobi identity) of weight \(k_0 + \ell_0\) not vanishing at zero and is such that for any choice of perturbation data \(\Gamma\), \(b^\Gamma_\sigma\) is always nonnegative, say a sum of squares, and that \(b^\Gamma_\sigma\) is always strictly positive if \(u \neq 0\). If one in addition can choose \(\Gamma\) such that \(\sum_{X^\sigma \in L^{(k_0)}} b^\Gamma_\sigma X^\sigma(0) = 0\) for all pairs \((k, l)\) with \(k + \ell l < k_0 + \ell l_0\), then one obtains \(\xi = X^\sigma(0)\) as a tangent vector. If furthermore \(\xi = X^\sigma(0)\) cannot be written as a linear combination of brackets of lower weight (w.r.t. \(\theta\)) evaluated at zero, then firstly \(-\xi\) and secondly all vectors \(\eta \in \mathbb{R}^n\) which are not linear combinations of \(\xi\) (with \(\xi \cdot \eta \geq 0\)) and brackets of lower weight than \(X^\sigma\) (w.r.t. \(\theta\)) evaluated at zero cannot be generated as tangent vectors using control variations of this particular form with this particular \(\theta\).

Very detailed analysis shows that if this situation occurs, then the lowest order term with definite coefficient \(b^\Gamma_\sigma\) is always associated to a bracket \(X^\sigma\) of type \((k_0, l_0)\) with \(k_0\) even and \(l_0\) odd, thus explaining why in theorem 0.4 brackets of type (even, odd) shall be neutralized by lower order brackets.
In the other cases, i.e. $k$ odd or $k$ and $l$ both even, the very fundamental ideas are that if $\xi = X^r(0)$ can be generated as a tangent vector to $G_\epsilon(t_0)$ at zero by using the control $u$, then replacing $u$ by $-u$ if $X^r \in L^{(\text{odd},*)}$ and by $u^{-1}$ if $X^r \in L^{(\text{even,even})}$ should essentially lead to generating $-\xi$ as a tangent vector. (Here $u^{-1}(t) = u(t_0 - t)$ is the time reversed control.)

This idea works for $k = 1$, i.e. in proving theorem 0.1, but although the proofs of theorem 0.2 and 0.3 are based on this idea, the details are much more intricate and consequently the proofs are very laborious [16, 19, 32].

Very loosely speaking we will refer to brackets $X^r \in L^{(k,l)}$ as "bad brackets" if they always exhibit this behavior described above, i.e., for every (possibly much more general) perturbation data $\Gamma$ $b^r_\Gamma$ is always definite and zero only if $u \equiv 0$.

Then theorems 0.3 and 0.4 can very loosely be rephrased as:

Only brackets of type (even, odd) are possibly bad. But system (1) is still STLC if all of these possibly bad brackets are neutralized by brackets of lower weight relative to some $\theta \in [0,1]$.

Brackets of type (even, 1) are always bad, and they are obstructions to STLC unless they are neutralized by brackets of lower weight relative to some $\theta \in [0,1]$.

The following questions arise and shall be dealt with here:

(1) Are all brackets of type (even, odd) necessarily bad?
(2) Can a bad bracket be balanced by a bracket of the same weight? In particular can two bad brackets of the same weight balance each other to give STLC.

(3) Are there other meaningful weight assignments for \( \{ X^\tau \in L(X^1, X^0) \} \) which do not come from a \( \theta \in [0,1] \) in the way described above and can be utilized in neutralizing bad brackets?

We have the following (partial) answers:

Similar to theorem 0.4a we will prove that brackets of type (2,3) are necessarily bad and show what precisely the set of brackets is that can neutralize these bad brackets.

Since it is obvious that \((\text{ad}^2 X^0; (\text{ad}^2 X^1; X^0))\) is bad, and must be neutralized by elements in \( \mathfrak{g}^1 \) for STLC by theorem 0.4a, we here only consider the "other" bracket in \( L^{(2,3)} \), namely \( X^{s_0} = (\text{ad}^2 [X^0; X^1]; X^0) \). We prove

**Theorem 1.** A necessary condition that (1) is STLC, is that \( X^{s_0}(0), X^{s_0} = (\text{ad}^2 [X^0; X^1]; X^0) \), can be written as a linear combination of

\[
\{ (\text{ad}^\nu X^0; X^1)(0), (\text{ad}^\nu X^0, (\text{ad}^3 X^1; X^0))(0) : \nu \in \mathbb{Z}_0^+ \}.
\]

As pointed out in [22], brackets of the form \((\text{ad}^\nu X^0; X^\tau)\) need not be bad, even if they are of type (even, odd). We will discuss the homogeneous component \( L^{(4,3)} \) in detail, show it is of dimension at most 5, show that 3 brackets of type (4,3) shall be considered bad and that one other bracket of the form \([X^0; X^\tau]\) with \( X^\tau \) of type (4,2) and the remaining one \([X^0; X^1]; ([\text{ad}^3 X^1; X^0]) \) shall not be considered bad. This discussion also shall motivate the conjecture that the only bad brackets in some suitable basis of \( L(X^1, X^0) \) are of the form
\[(\text{ad}^{\theta}X^0; (\text{ad}^2X^r_1; (\text{ad}^2X^r_2; \ldots; (\text{ad}^2X^r_i; X^0))) \ldots)\]

where \(X^{r+1}\) is of lower weight than \(X^r\) relative to some suitable weight assignment. (Or products with each factor of this form.)

In our answer to question (2), we show that for \(\theta = 1\) balancing may not be possible, but for \(\theta \in [0,1)\) balancing might give STLC. Furthermore bad brackets may balance each other, in particular even bad brackets of the same type.

In our answer to question (3) we show that apart from assigning the same weight to \((\text{ad}^{\theta}X^0; X^r)\) as to \(X^r\), there are other very interesting weight assignments different from those arising from \(\theta \in [0,1]\). In particular we shall introduce a new class of control variations which, for example, allows one to assign to \((\text{ad}^7[X^1; X^0]; X^0)\) of type \((7,8)\) a lower weight than to \((\text{ad}^2(\text{ad}^3X^1; X^0); X^0)\) of type \((6,3)\).

2.3 Technical Lemmata and Provision of the Tools for Proving Theorem and Discussing the Examples.

2.3.1 Scaling lemma. For a fixed control \(u = u_{1,1} : [0, t_0] \rightarrow [-1,1]\) and \(\epsilon \geq 0, \delta > 0\) define \(u_{\epsilon, \delta} : [0, \delta t_0] \rightarrow [-\epsilon, \epsilon]\) by

\[u_{\epsilon, \delta}(\delta t) = \epsilon u_{1,1}(t), \quad t \in [0, t_0].\]

(Sometimes we want the new control to rather be defined on \([0, t_0]\). In this case set

\[\tilde{u}_{\epsilon, \delta}(t) = \begin{cases} 0 & \text{if } 0 \leq t < (1-\delta)t_0 \\ u_{\epsilon, \delta}(t - (1-\delta)t_0) & \text{if } (1-\delta)t_0 \leq t \leq t_0. \end{cases}\]

For two finite sequences of positive integers \(\{r_i\}_{i=1}^{n}\) and \(\{s_i\}_{i=1}^{n}\) with
\( r_1 = s_1 = 1 \), define a two-parameter family of dilations on \( \mathbb{R}^n \) by

\[
\Delta_{\epsilon, \delta}(x) = (\epsilon^{r_1} \delta^{s_1} x_1, \ldots, \epsilon^{r_n} \delta^{s_n} x_n)
\]

for a given choice of local coordinates \( (x_1, \ldots, x_n) \) on \( \mathbb{R}^n \).

Of course, \( \Delta_{\epsilon, \delta} \) can be thought of as the composition of two one-parameter families of dilations \( \Delta^{1}_\epsilon \) and \( \Delta^{2}_\delta \) as they were introduced in chapter 1, only that we no longer can require that both \( \{r_i\}_{i=1}^n \) and \( \{s_i\}_{i=1}^n \) are nondecreasing. However for our purposes it is more convenient to think of \( \Delta_{\epsilon, \delta} \) as one two-parameter family of dilations.

Similarly to one-parameter families of dilations we introduce the following notions:

A polynomial \( p: \mathbb{R}^n \rightarrow \mathbb{R} \) (in \( (x_1, \cdots, x_n) \)) is homogeneous of degree \((m, m')\) w.r.t. \( \Delta_{\epsilon, \delta} \) if

\[
p(\Delta_{\epsilon, \delta}(x)) = \epsilon^m \delta^{m'} p(x) \quad \text{if} \quad m, m' \geq 0.
\]

\( H_{m,m'} \) is the set of all polynomials homogeneous of degree \((m, m')\) w.r.t. \( \Delta_{\epsilon, \delta} \), and \( H_{m,m'} = \{0\} \) if \( m < 0 \) or \( m' < 0 \).

A vectorfield \( X = \sum_{i=1}^n a_i(x) \frac{\partial}{\partial x_i} \in T\mathbb{R}^n \) with polynomial coefficients \( a_i(x) \) is homogeneous of weight \((-k, -k')\) w.r.t. \( \Delta_{\epsilon, \delta} \), \( k, k' \geq 0 \), if

\[
X H_{m,m'} \subseteq H_{m-k, m'-k'} \quad \text{for all} \quad m, m' \in \mathbb{Z}.
\]

This is equivalent to \( a_i \in H_{r_i-k, s_i-k'} \), \( i = 1, \cdots, n \).

The set of all vectorfields (with polynomial coefficients) homogeneous of weight \((-k, -k')\) is denoted \( \mathcal{U}_{(-k, -k')} \). (Let \( \mathcal{U}_{(k,k)} = \{0\} \) if \( k \) or \( k' > 0 \).)

Obviously \( \mathcal{U}_{(k,k)} \cap \mathcal{U}_{(t,t)} \subseteq \mathcal{U}_{(k+t, k'+t)} \). (Compare chapter 1.)
In the control system in question we will choose local coordinates such that \( X^1 \equiv \frac{\partial}{\partial x_1} \), and thus have \( X^1 \in U_{(1,-1)} \) (using \( r_1 = s_1 = 1 \)). By means of state feedback one always can achieve \( \langle dx_1, X^0 \rangle \equiv 0 \). In addition \( X^0(0) = 0 \) and very often we will work with dilations so that \( X^0 \) is homogeneous of weight \((0,-1)\). In this case \( L^{(k,l)}(X^1, X^0) \subseteq U_{(-k,-k-l)} \), which means a bracket \( X^\pi \) containing \( k \) factors \( X^1 \) and \( l \) factors \( X^0 \) is homogeneous of weight \((-k,-k-l)\).

We now state and prove the scaling lemma.

**Lemma 2.** If \( \Delta_{t,\delta} \) is a two-parameter family of dilations on \( \mathbb{R}^\pi \) such that \( X^0 \in U_{(0,-1)} \) and \( X^1 \equiv \frac{\partial}{\partial x_1} \in U_{(1,-1)} \), then \( x(\delta t, u_{t,\delta}) = \Delta_{t,\delta}(x(t, u_{1,1})) \).

**Proof:** We will show that both \( t \mapsto x(\delta t, u_{t,\delta}) \) and \( t \mapsto \Delta_{t,\delta}(x(t, u_{1,1})) \) are solutions of the same initial value problem

\[
\begin{align*}
\dot{x} &= \delta X^0(x) + \delta u_{t,\delta}(\delta t) X^1(x) \\
     &= \delta X^0(x) + \epsilon \delta u_{1,1}(t) X^1(x) \\
x(0) &= 0
\end{align*}
\]

which has a unique solution for any fixed control \( u = u_{1,1} \). Observe that by the hypotheses necessarily \( \langle dx_1, X^0 \rangle \equiv 0 \) must hold.

Clearly

\[
\begin{align*}
x(\delta 0, u_{t,\delta}) &= 0 = \Delta_{t,\delta}(x(0, u_{1,1})) \\
\frac{d}{dt} x_1(\delta t, u_{t,\delta}) &= \delta u_{t,\delta}(\delta t) = \epsilon \delta u_{1,1}(t)
\end{align*}
\]

and
\[
\frac{d}{dt} \left( \Delta_{x,t}(x(t, u_{1,1})) \right)_1 = \epsilon \delta \frac{d}{dt} x_1(t, u_{1,1}) = \epsilon \delta u_{1,1}(t).
\]

For \( i \geq 2 \),

\[
\frac{d}{dt} x_i(\delta t, u_{i,\epsilon}) = \delta a_i(x(\delta t, u_{i,\epsilon}))
\]

and

\[
\frac{d}{dt} \left( \Delta_{x,t}(x(t, u_{1,1})) \right)_i = \epsilon^{i-1} \delta^i \frac{d}{dt} x_1(t, u_{1,1})
\]

\[
= \delta^i \epsilon^{i-1} a_i(x(t, u_{1,1}))
\]

\[
= \delta^i a_i(\Delta_{x,t}(x(t, u_{1,1})))
\]

where \( X^0(x) = \sum_{i=2}^{n} a_i(x) \frac{\partial}{\partial x_i} \in \mathfrak{n}_{(0,-1)} \), i.e. \( a_i \in H_{i,-1} \).

One application of this lemma lies in showing that a particular system is STLC. Suppose the system is such that \( X^1 = \frac{\partial}{\partial x_1} \in \mathfrak{n}_{(-1,-1)} \) and \( X^0 \in \mathfrak{n}_{(0,-1)} \) w.r.t. some dilation \( \Delta_{x,\epsilon} \). If for some control bound \( \epsilon_0 \), say \( \epsilon_0 = 1 \), and some time \( t_0 > 0 \), the attainable set \( G_{\epsilon_0}(t_0) \) contains a neighborhood \( U \) of \( x = 0 \), then because of \( 0 \in \Delta_{x,\epsilon} U \subseteq \Delta_{x,\epsilon} G_{\epsilon_0, t_0} = G_{\epsilon, t_0} \), \( G_{\epsilon}(t) \) contains zero in its interior for all \( \epsilon, t > 0 \), and thus the system is STLC.

### 2.3.2 Independence of \( [X^0, X^1](0) \) from \( X^1(0) \)

We need to verify the following

**Claim.** If \( \{X^1(0), [X^0, X^1](0)\} \) are linearly dependent, then \( X^0(0) = (\text{ad}^2[\text{ad}^0 X^1; X^0])(0) \) is linearly dependent on \( \{X^1(0), (\text{ad}^2 X^1; X^0)(0), [X^0, (\text{ad}^2 X^1; X^0)](0)\} \).
Observe that $X^1(0) \neq 0$, and we thus may assume $[X^0, X^1] = \lambda X^1 + H$ with $H(0) = 0$ and $\lambda \in \mathbb{R}$. Then $(\text{ad}^2 X^0, X^1) = \lambda X^1 + \lambda H + [X^0, H]$ and using $X^0(0) = H(0) = [X^0, H](0) = 0$ the definition of $H$, and the Jacobi identity repeatedly, one obtains:

$$X^\nu(0) = [(\text{ad}^2 X^0, X^1); [X^0, X^1]](0) = \lambda [[X^0, H]; X^1]$$

$$= \lambda [X^1; X^0], H](0) + \lambda [[H; X^1]; X^0]$$

$$= \lambda [H - \lambda X^1, H](0) + \lambda [[[X^0, X^1]; X^1]; X^0]$$

$$= -\lambda^2 [X^1; [X^0, X^1]](0) + \lambda [[[X^0, X^1]; X^1]; X^0].$$

A nice generalization would answer: If $\nu \in \mathbb{Z}_0^+$ is maximal such that $\{(\text{ad}^j X^0, X^1)(0); j = 0, 1, \cdots, \nu\}$ are linearly independent, which brackets $X^\nu \in L(X^1, X^0)$, especially which brackets $X^\nu \in \mathcal{A}^2(X^1, X^0)$, can be linearly independent from lower order ones when evaluated at zero?

2.3.3 Chen and Lie series. The proof of theorem 1 will in some aspects be similar to the ones of theorems 0.4a and 0.4 by Sussmann and Stefani, as in the use of a certain series in the algebra of formal power series in the indeterminants $\hat{X}^1$ and $\hat{X}^0$ which is also known as "Chen Series". Here we will only sketch where this series comes from. For details, the interested reader is referred to [10, 12, 32].

At some points we will have to stress the distinction between the free Lie algebra $L(\hat{X}^1, \hat{X}^0)$ (and the free associative algebra $\mathcal{A}(\hat{X}^1, \hat{X}^0)$) generated by the two letters (indeterminants) $\hat{X}^1$ and $\hat{X}^0$ and the Lie algebra $L(X^1, X^0)$ of vector fields on $\mathbb{R}^n$ (and the associative algebra $\mathcal{A}(X^1, X^0)$ of partial differential operators on $\mathbb{R}^n$ generated by the two vector fields $X^1$ and $X^0$.}
The map $\hat{X}^1 \rightarrow X^1$ and $\hat{X}^0 \rightarrow X^0$ extends to a (Lie) algebra homomorphism in an obvious way. But to avoid further confusing notation we will not explicitly give this map a name, instead we simply will "drop the hat". For example if $\hat{X}^0$ denotes a specific Lie product, say $\hat{X}^0 = [\hat{X}^1; [\hat{X}^1, \hat{X}^0]] \in L(\hat{X}^1, \hat{X}^0)$, then $X^0 = [X^1; [X^1, X^0]] \in L(X^1, X^0)$.

Because of $X^0(0) = 0$, it will make sense to work mostly in

$$A(\hat{X}^1, \hat{X}^0) = A(\hat{X}^1, \hat{X}^0) / \hat{X}^0 A(\hat{X}^1, \hat{X}^0)$$

and $A(X^1, X^0)$ being defined similarly.

The homogeneous components spanned by all monomials (Lie monomials, respectively) with $k$ factors $\hat{X}^1$ (or $X^1$) and $l$ factors $\hat{X}^0$ (or $X^0$) are denoted by

$$A^{(k,l)}(\hat{X}^1, \hat{X}^0), \ A^{(k,l)}(X^1, X^0), \ L^{(k,l)}(\hat{X}^1, \hat{X}^0), \ etc.$$ and elements in these components sometimes will be referred to as being of type $(k, l)$.

The dimensions of the homogeneous components are

$$\dim A^{(k,l)}(\hat{X}^1, \hat{X}^0) = \binom{k+l}{k},$$

$$\dim A^{(k,l)}(\hat{X}^1, \hat{X}^0) = \binom{k+l-1}{k-1} \text{ and}$$

$$\dim L^{(k,l)}(\hat{X}^1, \hat{X}^0) = \frac{1}{k+l} \sum_{d \mid (k+l)} \mu(d) \frac{(k+l)!}{(k/d)!(l/d)!}$$

where $\mu(d) = \begin{cases} 0 & \text{if } p^2 \mid d \text{ for some prime } p \\ 1 & \text{if } d = p_1 p_2 \cdots p_k \text{ for } p_1, \cdots, p_k \text{ distinct primes} \end{cases}$

is the Moebius function [3].
(Without the hat, above equalities become inequalities, i.e., the \( = \) sign is to be replaced by \( \leq \).

Finally to discuss the Chen series, we also need the algebra of formal power series in \( \hat{X}^1 \) and \( \hat{X}^0 \), which will be denoted by \( \hat{A}(\hat{X}^1, \hat{X}^0) \). Now, given an element \( \hat{P} \in \hat{A}(\hat{X}^1, \hat{X}^0) \) and an admissible control function \( u: [0, t_0] \to [-\epsilon, \epsilon] \), one may consider the differential equation

\[
\frac{d\hat{S}}{dt} = \hat{S}(\hat{X}^0 + u\hat{X}^1) \quad \text{with initial condition}
\]

\[
\hat{S}(0) = \hat{P},
\]

which in the particular case \( P = 1 \) has the unique solution:

\[
\text{Ser}(u) = \hat{S}(t) = \sum_I a_I(t, u)\hat{X}^I
\]

(3a)

where \( I = (i_1, \ldots, i_r) \) ranges over all multi-indices with values \( i, j \in \{0, 1\} \), \( r \) ranges over \( \mathbb{Z}_0^+ \), \( \hat{X}^I = \hat{X}^{i_1} \cdots \hat{X}^{i_r} \) and

\[
a_I(u,t) = \int_0^t u_1 = \int_0^{s_1} u(s_1) \int_0^{s_{r-1}} u(s_{r-1}) \cdots \int_0^{s_2} u(s_2) ds_1 \cdots ds_r
\]

is an iterated integral with \( u_0 \equiv 1 \) and \( u_1 = u \) the actual control function from above. We let \( |I| = |(i_1 \cdots i_r)| = r \) and \( I_1 = \text{number of } i_j \text{ equal to one} \).

An interesting fact is that the map

\[
u \to \text{Ser}(u) = \sum_I a_I(t, u)\hat{X}^I \in \hat{A}(\hat{X}^1, \hat{X}^0),
\]

is injective. The importance of this series for our purposes lies in the following:

Given a control system of form (1), an admissible control \( u \) and an analytic function \( \phi: \mathbb{R}^n \to \mathbb{R} \), then
\[ \phi(x(t,u)) = \sum_i a_i(t,u)(X^i \phi)(0) \]

i.e., the series on the right side of (3) can be considered as an "asymptotic series for the propagation of \( \phi \) along trajectories of (1)," the convergence being uniform on some time interval \([0,T], T > 0\). A proof and details can be found in [32].

(Remark: In Lemma 4.3 of [32] Sussmann requires \( x \in K^\text{compact} \subseteq \mathbb{R}^n \) also to show uniform convergence. But since in our case we have no finite escape time, it suffices to require \( t \leq T \) and \( |u(t)| \leq \epsilon_0 \); this automatically gives \( x(t,u) \in K \).)

In proving theorems 0.4a, 0.4 and 1 one assumes that the hypothesis is violated, that means, the supposedly bad bracket cannot be neutralized by brackets of lower weight from the specified set. Then one constructs an analytic function \( \Phi: \mathbb{R}^n \rightarrow \mathbb{R} \) with \( \Phi(0) = 0 \) and \( (\text{grad } \phi)(0) \neq 0 \) and shows \( \Phi(x(t,u)) \geq 0 \) for all admissible controls \( u \), as long as \( t \) is sufficiently small. This then shows that the attainable set \( G_x(t) \) lies on one side of the (locally defined) manifold \( \Phi^{-1}(0) \) through \( x_0 = 0 \) and thus cannot contain \( x_0 = 0 \) in its interior.

To do this one works with the series (3), shows that a specific term associated to the bad bracket in question is always nonnegative and dominates all other terms in the series. This essentially amounts to estimating the various iterated integrals \( |a_i(t,u)| \) and the derivatives \( |X^i \phi(0)| \).

Since in this case it is a priori not at all clear which \( I \)'s correspond to the brackets \( (\text{ad}^2[X^0,X^1],X^0) \) and \( (\text{ad}^4X^0,(\text{ad}^3X^1,X^0)) \) occurring in theorem (1), we will start by doing some manipulations on the terms of the series (3a)
corresponding to $I_1 \leq 3$. These manipulations heavily rely on $X^0(0) = 0$, give
some more geometrical meaning to (3), and essentially amount to changing the
bases of $A^{(k,l)}(\hat{X}, \hat{X}^0)$ by introducing Lie polynomials on one hand, and doing
successive integrations by parts on the iterated integrals $a_1(t,u)$ on the other
hand.

Later, in the proof of theorem 1 it will become clear how "nice" the
resulting series will be. Unfortunately, for $I_1 \geq 4$, a similar rearrangement of
(3a) becomes considerably more complicated, mainly because there are too many
choices for a "nice" basis. However, from the geometric intuition, all of the
series (3a) should be expressible in similar geometrically meaningful terms.

Before doing the rearrangement for the general homogeneous
component $A^{(k,l)}$, we will do it for $A^{(2,3)}$ as a motivation, since in the general
case the idea tends to get obscured by too many notational compromises.

Let $\mathcal{M}(k,l)$ be the set of all multi-indices $I$ of length $k+l$ containing
$k$ ones and $l$ zeros and with $i_1 = 1$.

For easier reading, we will use for iterated integrals like

$$a_{(1,1,0,0,0)}(t,u) = \int_0^t \int_0^{t_2} \int_0^{t_4} \int_0^{t_5} u(t_2) u(t_4) dt_1 dt_2 dt_3 dt_4 dt_5$$

the short-hand notation

$$a_{11000} = \int \int \int u \int u, \quad \text{etc.,}$$

and write $\hat{X}^{11000}$ for $\hat{X}^{(1,1,0,0,0)}$ if it will not cause confusion. Then we have as
a partial sum of (3a)
\[
\sum_{I \in \mathcal{I}(2,3)} a_I(t,u) \hat{X}^I = a_{11000} \hat{X}^{11000} + \cdots + a_{10001} \hat{X}^{10001}
\]

\[
= \int \int \int \int \int u \int u \hat{X}^1 \hat{X}^0 \hat{X}^0 \hat{X}^0 + \cdots \int u \int \int \int \int u \hat{X}^1 \hat{X}^0 \hat{X}^0 \hat{X}^0 \hat{X}^1.
\]

Using Lie polynomials one easily computes (in $A$, not in $A$)

\[
\begin{pmatrix}
\hat{W}^{2,3}_1 \\
\hat{W}^{2,3}_2 \\
\hat{W}^{2,3}_3 \\
\hat{W}^{2,3}_4
\end{pmatrix} =
\begin{pmatrix}
(ad^2 \hat{X}^0, (ad^2 \hat{X}^1, \hat{X}^0)) \\
(ad^2 [\hat{X}^0, \hat{X}^1], \hat{X}^0) \\
(ad^2 \hat{X}^0, \hat{X}^1) [\hat{X}^0, \hat{X}^1] \\
(ad^3 \hat{X}^0, \hat{X}^1)
\end{pmatrix} =
\begin{pmatrix}
1 & -2 & 0 & 0 \\
0 & 1 & -3 & 2 \\
0 & 0 & 1 & -1 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\hat{X}^1 \hat{X}^0 \hat{X}^0 \hat{X}^0 \\
\hat{X}^1 \hat{X}^0 \hat{X}^0 \hat{X}^0 \\
\hat{X}^1 \hat{X}^0 \hat{X}^0 \hat{X}^0 \\
\hat{X}^1 \hat{X}^0 \hat{X}^0 \hat{X}^0 \hat{X}^1
\end{pmatrix}
\]

which we will write as $\hat{W} = \hat{W}^{2,3} = M \cdot \hat{V}^{2,3} = M \hat{V}$. ($M$ is the $4 \times 4$ matrix of integers, $\hat{V}$ the vector of $5$th-order products in $A(X^1, X^0)$ on the right.)

On the other hand, successive integration by parts gives

\[
a_{11000} = \int \int \int \int u \int u = \frac{1}{2} \int \int \int (\int u)^2,
\]

\[
a_{10100} = \int \int \int u \int \int u = \frac{1}{2} \int (\int \int u)^2 - \int \int \int (\int u)^2,
\]

\[
a_{10010} = \int \int u \int \int \int u = (\int \int u)(\int \int u) - \frac{3}{2} \int (\int \int u)^2,
\]

\[
a_{10001} = \int \int \int u \int \int \int u = (\int \int \int u)(\int \int u) - (\int \int \int u)(\int \int u) + \int (\int \int u)^2,
\]

and hence if $a^{2,3} = (a_{11000}, \cdots, a_{10001})$ and

\[
b^{2,3} = (\frac{1}{2} \int \int \int (\int u)^2, \frac{1}{2} \int (\int \int u)^2, (\int \int \int u)(\int \int u), (\int \int \int u)(\int \int u))
\]

one immediately has $a = bM$ with the same matrix $M$ as above. (I.e., the combinatorial manipulations by introducing Lie polynomials and successive integration by parts are essentially inverses of each other.)

Together these give

\[
\sum_{I \in \mathcal{I}(2,3)} a_I(t,u) = a \cdot \hat{V} = (bM) \cdot \hat{V} = b \hat{W} = \frac{1}{2} \int \int \int (\int u)^2(ad^2 \hat{X}^0, (ad^2 \hat{X}^1, \hat{X}^0))
\]
\[ + \frac{1}{2} \int (\int u)^2 (\text{ad}^2 [\dot{X}^0, \dot{X}^1]; \dot{X}^0) \]
\[ + (\int \int u)(\int u)(\text{ad}^2 \dot{X}^0, \dot{X}^1)[\dot{X}^0, \dot{X}^1] \]
\[ + (\int \int \int u)(\int u)(\text{ad}^3 \dot{X}^0, \dot{X}^1)[\dot{X}^0, \dot{X}^1], \]

which is "what one expects" when considering the following control systems and computing \( W^{2,3}_i(0), i = 1,2,3,4, \) for each of them:

(i) \[
\begin{align*}
\dot{x}_1 &= u \\
\dot{x}_2 &= \frac{1}{2}x_1^2 \\
\dot{x}_3 &= x_2 \\
\dot{x}_4 &= x_3
\end{align*}
\]

(ii) \[
\begin{align*}
\dot{x}_1 &= u \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= \frac{1}{2}x_2^2 \\
\dot{x}_4 &= x_2 + x_1 x_3
\end{align*}
\]

(iii) \[
\begin{align*}
\dot{x}_1 &= u \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= x_2 \\
\dot{x}_4 &= x_3 \\
\dot{x}_5 &= x_1 x_3 + x_4 u
\end{align*}
\]

(iv) \[
\dot{x}_1 = u \\
\dot{x}_2 = x_1 \\
\dot{x}_3 = x_2 \\
\dot{x}_4 = x_3 \\
\dot{x}_5 = \frac{dx}{dt} (x_2 x_3) \\
\dot{x}_6 = \frac{dx}{dt} (x_4 x_1)
\]

In some sense the Lie brackets / 2nd order partial differential operators \( W^{2,3}_i, \ldots W^{2,3}_4 \) are diagonal on these control systems when evaluated at zero, that is for example \( \text{ad}^2 [X^0, X^1]; X^0)(0) \) equals \( \frac{\partial}{\partial x_3} \big|_0 \) in the second system and vanishes in the three other systems.

Furthermore observe that the third and fourth system evolve on a lower dimensional manifold \( \dim(L(X^1, X^0)(0) < n) \), (e.g., \( x_4 - x_2 x_3 = \text{const.} \) along solutions in the third system.) In other words \( \frac{\partial}{\partial x_4} \big|_0 / \frac{\partial}{\partial x_5} \big|_0 \) in the third/fourth system cannot be obtained from Lie brackets, only from higher order partial differential operators in \( A(X^1, X^0) \). This nicely reflects the philosophy, that all the "important" information about the system lies in the Lie brackets evaluated at zero, the higher order operators just reflecting a not
perfect choice of local coordinates on the state space.

We would like to get similar nice sums \( bW = \sum b_i W_i \) at least for all homogeneous components \( A^{(k,l)} \) with \( k \leq 3 \) without having to multiply out Lie polynomials or keeping track of many successive integrations by parts, i.e. without computing matrices like \( M \) explicitly. These combinatorically intricate calculations can be avoided in the following way, again using \( A^{(2,3)} \) as an illustrating example.

First we know \( \dim A^{(k,l)}(\hat{X}_1, \hat{X}_0) = \binom{k+l-1}{k-1} = \lambda \), i.e. in the case \( (k,l) = (2,3) \), \( \lambda = 4 \). We next write down a set \( \{ \hat{W}_1^{k,l}, \ldots, \hat{W}_\lambda^{k,l} \} \) of elements of \( A^{(k,l)}(\hat{X}_1, \hat{X}_0) \) which we expect to be a new basis for \( A^{(k,l)} \) from our geometric intuition. In our example we choose \( \{ \hat{W}_1^{2,3}, \ldots, \hat{W}_4^{2,3} \} \) as above. We next write down a "prototype" control system of form (1) on some euclidean space \( \mathbb{R}^d \) and an analytic function \( \psi: \mathbb{R}^d \to \mathbb{R}^\lambda \). In the case \( (k,l) = (2,3) \), we combine the four systems given above in a suitable way:

\[
\begin{align*}
\dot{x}_1 &= u \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= x_2 \\
\dot{x}_4 &= x_3 \\
\dot{x}_5 &= \frac{1}{2} x_1^2 \\
\dot{x}_6 &= x_5 \\
\dot{x}_7 &= x_6 \\
\dot{x}_8 &= \frac{1}{2} x_2^2 .
\end{align*}
\]

Let \( \psi: \mathbb{R}^8 \to \mathbb{R}^4 \), \( \psi(x) = (x_7, x_8, x_2 x_3, x_1 x_4)^T \), and compute

\[
X^1 \equiv \frac{\partial}{\partial x_1}, [X^0, X^1] = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial x_5}, (\text{ad}^2 X^0, X^1) = \frac{\partial}{\partial x_3} + x_1 \frac{\partial}{\partial x_6} + x_2 \frac{\partial}{\partial x_8} \\
(\text{ad}^3 X^0, X^1) = \frac{\partial}{\partial x_4} + x_1 \frac{\partial}{\partial x_7} - x_1 \frac{\partial}{\partial x_8} \\
W_i(0) = (\text{ad}^2 X^0; (\text{ad}^2 X^1; X^0))(0) = \frac{\partial}{\partial x_7},
\]

\[ W_2(0) = \left( \text{ad}^2 [X^0, X^1]; X^0 \right)(0) = \frac{\partial}{\partial x_3} \]

\[ W_3(0) = \left( \text{ad}^2 X^0, X^1 \right)[X^0, X^1](0) = \frac{\partial^2}{\partial x_2 \partial x_3} \]

\[ W_4(0) = \left( \text{ad}^3 X^0, X^1 \right) X^1(0) = \frac{\partial^2}{\partial x_1 \partial x_4} \]

and therefore \( W\psi(0) = I_4 \), the identity \( 4 \times 4 \) matrix.

Associated with this control system is a two-parameter family of dilations, in our case:

\[ \Delta_{\epsilon, \delta}(x) = (\epsilon \delta x_1, \epsilon^2 \delta^2 x_2, \epsilon \delta^4 x_3, \epsilon \delta^4 x_4, \epsilon^2 \delta^4 x_5, \epsilon^2 \delta^4 x_6, \epsilon \delta^4 x_7, \epsilon \delta^4 x_8). \]

With respect to this dilation we have \( X^0 \in \mathfrak{n}_{(0,-1)}, X^1 = \mathfrak{n}_{(-1,-1)} \) and \( \psi \in H_{k,l} \)

and therefore \( X^1 \psi(0) = 0 \) for all \( I \not\in (k,l) \).

Since \( W\psi(0) = I_\lambda \), we have found an isomorphism from the linear span of \( \{ \hat{W}_1^{k,l}, \cdots, \hat{W}_\lambda^{k,l} \} \) considered as a subspace of \( A^{(k,l)}(\hat{X}^1, \hat{X}^0) \) onto \( \mathbb{R}^\lambda \),

defined by \( \sum_{i=1}^\lambda c_i \hat{W}_i \rightarrow \sum_{i=1}^\lambda c_i W_i \psi(0) \), \( c_i \in \mathbb{R} \), and hence \( \{ \hat{W}_1^{k,l}, \cdots, \hat{W}_\lambda^{k,l} \} \) is a basis

for \( A^{(k,l)}(\hat{X}^1, \hat{X}^0) \).

If we write \( \{ \hat{V}_1^{k,l}, \cdots, \hat{V}_\lambda^{k,l} \} \) for \( \{ \hat{X}^i : I \in (k,l) \} \), which of course is a basis for \( A^{(k,l)}(\hat{X}^1, \hat{X}^0) \), then there is an invertible matrix \( M = M^{(k,l)} \in \mathbb{R}^{\lambda \times \lambda} \)

such that \( \hat{W} = M \hat{V} \). Since \( M \) comes from multiplying Lie polynomials, it even has integer entries.

If we let \( \alpha = (\alpha_1(t,u), \cdots, \alpha_\lambda(t,u)) \) be an enumeration of \( \{ a_I(t,u), I \in \mathbb{I}(k,l) \} \) such that \( \sum_{i=1}^\lambda \alpha_i \hat{V}_i = a_X \hat{X}^1 \), we may define

\[ b = b(t,u) = \alpha(t,u)^\cdot M^{-1} \]
This way we have

\[ \sum_{I \in \mathcal{P}(k,l)} a_I \hat{X}^I = \alpha \hat{V} = (bM)(M^{-1}\hat{W}) = b\hat{W} = \sum_{i=1}^{\lambda} b_i(t,u)\hat{W}^i_{k,l} \]

We now use \( W\psi(0) = 1 \) to obtain:

\[ \psi(x(u,t)) = \sum_{I \in \mathcal{P}(k,l)} a_I(t,u)X^I\psi(0) = \sum_{i=1}^{\lambda} b_i(t,u)W^i_{k,l}\psi(0) \]

\[ = b(t,u)W^k_{k,l}\psi(0) = b(t,u) . \]

But \( \psi \) is chosen so that we can read off \( \psi(x(u,t)) \) and thus \( b(t,u) \) immediately, without ever computing \( M \). In our case:

\[ b(t,u) = \psi(x(t,u)) = (x_7(t,u), x_8(t,u), x_2(t,u)x_3(t,u), x_1(t,u)x_4(t,u)) . \]

\[ = b(\frac{1}{2}\int\int\int (\int u)^2, \frac{1}{2}\int (\int\int u)^2, (\int\int\int u)(\int u), (\int\int\int u)(\int u) . \]

Now we explicitly give \( W^k_{j,l} \) and \( b^k_{j,l} \) for \( k=1,2,3 \). In the case \( k=1 \) observe

\[ \hat{X}^1\hat{X}^0\dotsc\hat{X}^0 = (ad^\nu\hat{X}^0,\hat{X}^1) \mod \hat{X}^0 \mathcal{A}(\hat{X}^1,\hat{X}^0) \]

and thus

\[ \hat{X}^1\hat{X}^0\dotsc\hat{X}^0 = (ad^\nu\hat{X}^0,\hat{X}^1) \text{ in } \mathcal{A}(\hat{X}^1,\hat{X}^0) . \]

From now on we do all calculations in \( \mathcal{A}(\hat{X}^1,\hat{X}^0) \) and therefore may write

\[ \sum_{l=1}^{\infty} a_l(t,u)\hat{X}^l = \sum_{l=1}^{\infty} b^{l,l}_{0}(t,u)\hat{W}^l_{l,l} \text{ where } \hat{W}^{l,l} = (ad^l X^0, X^l), \ l \geq 0, \ and \]

\[ b^{l,0}(t,u) = a_1(t,u) = \int_0^t u(s)ds \text{ and inductively for } l \geq 1 . \]
\[ b^{l,l}(t,u) = \int_0^t b^{l,l-1}(s,u)ds = \int_0^t \frac{(t-s)^l}{l!} u(s)ds \]

\[ (= a_l(t,u) \text{ for } \|l\| = l + 1, l_1 = i_1 = 1) \]

for the last step compare lemma 7 (§2.3.5).

In the case \( k = 2 \) observe \( \dim A^{(2,l)}(\hat{X}, \hat{X}^0) = l + 1, \ l \geq 0 \), and define

\[ \hat{W}_{1,\nu} = (\text{ad}^{l-1-2\nu} \hat{X}^0, (\text{ad}^{2} \hat{X}^0, \hat{X}^0)) \text{ if } 0 \leq 2\nu \leq l - 1 \]

\[ \hat{W}_{2,\nu} = (\text{ad}^{\nu} \hat{X}^0, \hat{X}^0) (\text{ad}^{l-\nu} \hat{X}^0, \hat{X}^0) \text{ if } \frac{l}{2} \leq \nu \leq l . \]

The control system associated to the \((2, l)\) component is

\[
\begin{aligned}
\dot{x}_0 &= u \\
\dot{x}_{i+1} &= x_i, & 1 \leq i \leq l-1, \\
\dot{y}_{1,2i+1} &= \frac{\beta}{2} x_i^2, & 0 \leq 2i \leq l-1, \\
\dot{y}_{i,j+1} &= y_{ij}, & 0 \leq 2i \leq l-1, 2i+1 \leq j \leq l-1.
\end{aligned}
\]  

(Then \( \psi \) maps the \((x,y)\)-space analytically into \( \mathbb{R}^{l+1} \).

Here \( \alpha(i, j) = 1 \) if \( i \neq j \) and \( \alpha(i, j) = \frac{1}{2} \) if \( i = j \). In general we define

\[ \alpha(n_1, \ldots, n_r) = \frac{1}{k_1! \cdots k_r!} \]  

if for \( s \) pairwise distinct integers \( z_1, \ldots, z_s \) and some permutation \( \sigma \)

\[ z_1 = n_{\sigma(1)} = n_{\sigma(2)} = \cdots = n_{\sigma(k_1)}, \]

\[ z_2 = n_{\sigma(k_1+1)} = \cdots = n_{\sigma(k_1+k_2)}, \ldots, z_s = n_{\sigma(k_1+\cdots+k_{r-1}+1)} = \cdots = n_{\sigma(r)}, \]

where \( r = k_1 + \cdots k_s \). We later will only use \( \alpha(i,j,k) \) for various \( i,j,k \), which takes
values $1, \frac{1}{2}$ and $\frac{1}{6}$.

Let $\Delta_{x,y}(x,y) = (\epsilon^{\delta+1} x_{1}, \epsilon^{\delta+2} y_{1})$. W.r.t. this dilation $X^0 \in \mathbb{R}_{(0,1)}$

$X^1 \equiv \frac{\partial}{\partial x_0} \in \mathbb{R}_{(-1,-1)}$ and $\psi \in \mathbb{H}_{2,l}$. To show that $W\psi(0) = \mathbf{1}_{l+1}$, verify that for

$0 \leq \nu < l$, $(\text{ad}^\nu X^0; X^1) = \frac{\partial}{\partial x_\nu} + x_{\nu-1} \frac{\partial}{\partial y_{\nu,2\nu-1}} + R(\nu)$ where $R(\nu)$ is a sum of vectorfields of the form $r x_\mu \frac{\partial}{\partial y_{1,\lambda}}$ with $r \in \mathbb{R}$, $\mu < \nu - 1$, and $x_{\nu-1}$ shall be replaced by $0$ if $\nu = 0$ or $2\nu > l-1$. This is easily (inductively) derived from

$[X^0, \frac{\partial}{\partial x_\nu}] = \frac{\partial}{\partial x_{\nu+1}} + x_{\nu} \frac{\partial}{\partial y_{\nu,2\nu+1}}$ (the last term vanishing when $2\nu+1 > l$)

and $[X^0, x_\mu \frac{\partial}{\partial y_{\nu,2\nu+1}} + R(\nu)] = R(\nu+1)$ (using $X^0 x_\mu = x_{\mu-1}$ or $X^0 x_\mu = 0$).

This immediately leads to $(\text{ad}^2(\text{ad}^\nu X^0; X^1); X^0)(0) = \frac{\partial}{\partial y_{\nu,2\nu+1}}$ if

$0 \leq 2\nu \leq l-1$ and thus $W^{2,1}_{l,\nu}(0) = (\text{ad}^{l-2\nu-1} X^0, (\text{ad}^2(\text{ad}^\nu X^0; X^1); X^0))(0) = \frac{\partial}{\partial y_{\nu,l}}$

if $0 \leq 2\nu \leq l-1$ and $W^{2,1}_{2,\nu}(0) = (\text{ad}^\nu X^0; X^1)(\text{ad}^{l-\nu} X^0; X^1)(0) = \frac{\partial^2}{\partial x_\nu \partial y_{l-\nu}}$ if

$\frac{l}{2} \leq \nu \leq l$. We obtain $W\Psi(0) = \mathbf{1}_{l+1}$ and read off the coefficients $b^{2,l}_{\nu}(t,u)$

directly from the control system (4).

$$b^{2,l}_{1,\nu}(t,u) = \frac{1}{2} \int_{0}^{t} \int_{s}^{t} \int (\int_{0}^{s} \int_{0}^{u} \int_{0}^{v} \int_{0}^{w})^2$$

$$= \frac{1}{2} \int_{0}^{t} \int_{0}^{s} \frac{(s-\sigma)^{\nu}}{(l-2\nu-1)!} \int_{0}^{\sigma} u(\tau) d\tau ds$$ if $0 \leq 2\nu \leq l-1$ and

$$b^{2,l}_{2,\nu}(t,u) = \int_{0}^{t} \int_{0}^{s} \int_{0}^{t} \int_{0}^{u} \int_{0}^{v} (\int_{0}^{w} \int_{0}^{x})^2$$

$$= \frac{1}{2} \int_{0}^{t} \int_{0}^{s} \frac{(s-\sigma)^{\nu}}{(l-\nu+1)!} \int_{0}^{\sigma} u(\tau) d\tau ds$$ if $0 \leq \nu \leq l-1$.
\[ = \int_0^t \frac{(t-s)^\nu}{\nu!} u(s)ds \int_0^t \frac{(t-s)^{l-\nu}}{(l-\nu)!} u(s)ds \quad \text{if} \quad \frac{l}{2} \leq \nu \leq l. \]

In the case \( k = 3 \) observe \( \dim A^{(3,l)}(\hat{X}^1, \hat{X}^0) = \binom{l+2}{2} = \frac{(l+2)(l+1)}{2} \) and

\[
r_{3,l} = \dim L^{(3,l)}(\hat{X}^1, \hat{X}^0) = \begin{cases} 
\frac{(l+2)(l+1)}{6} & \text{if} \quad 3 \nmid l \\
\frac{l^2+2l}{6} & \text{if} \quad 3 \mid l
\end{cases}
\quad (6a)
\]

(Compare equation (2) at the beginning of §2.3.3).

For \( 0 \leq \nu \leq \mu \) with \( 2\nu + \mu \leq l-1 \) define

\[ \hat{W}_{3,\mu,\nu}^0 = -\langle \text{ad}^{l-2\nu-\mu-1} \hat{X}^0; (\text{ad}^\nu \hat{X}^0; \hat{X}^1); (\text{ad}^\mu \hat{X}^0; \hat{X}^1); \hat{X}^0 \rangle. \]

For \( l = 0 \) there is no admissible pair \((\mu, \nu)\). For \( l \geq 1 \), \( \hat{W}_{3,\mu,\nu}^0 \) is either of the form \([\hat{X}^0, \hat{W}_{3,\mu',\nu'}^{0,l-1}]\) with suitable \( \mu' \) and \( \nu' \) or \( 2\mu + \nu = l-1 \). In the latter case there are exactly \((\nu_i+1)\) choices for \( \nu = 0,1, \ldots, \nu_i \) for distinct admissible pairs \((\mu, \nu)\), where \( l-1 < 3\nu_i \leq l \); which fits well with \( r_{3,l} - r_{3,l-1} = \nu_i + 1 \), where \( r_{3,l} \) is as defined in \((6a)\). Again, later we will show that the brackets \( \hat{W}_{3,\mu,\nu}^0 \) indeed form a basis for \( L^{(3,l)}(\hat{X}^1, \hat{X}^0) \).

For all \( 0 \leq \mu, \nu \) with \( 2\mu + \nu \leq l-1 \) define

\[ \hat{W}_{2,\mu,\nu}^3 = \langle \text{ad}^{l-1-2\nu-\mu} \hat{X}^0; (\text{ad}^\mu \hat{X}^0; \hat{X}^1); \text{ad}^\nu \hat{X}^0; \hat{X}^1 \rangle. \]

For each \( 0 \leq \nu \leq l-1 \) one has \( \frac{1}{2}(l-\nu) \) choices for \( \mu \) if \( (l-\nu) \) is even and \( \frac{1}{2}(l-\nu+1) \) choices for \( \mu \) if \( (l-\nu) \) is odd.

Thus for \( l \) even one has
\[ q_{3,l} = \sum_{\nu=0}^{\frac{l-1}{2}} \left( \frac{l-2\nu}{2} + \frac{l-(2\nu+1)+1}{2} \right) = \frac{l^2+2l}{4} \]  \hfill (6b)

and for \( l \) odd one has

\[ q_{3,l} = \sum_{\nu=0}^{\frac{l-1}{2}} \frac{l-2\nu+1}{2} + \sum_{\nu=0}^{\frac{l-1}{2}-1} \frac{l-(2\nu+1)}{2} = \frac{(l+1)^2}{4} \]

distinct choices for admissible pairs \((\mu,\nu)\).

For all \( \mu,\nu \) with \( 0 \leq \nu \leq \mu \leq l-\mu-1 \) define

\[ \tilde{W}_{3,\mu,\nu} = \left( \text{ad}^{l-\mu-\nu} \tilde{X}_0, \tilde{X}_1 \right) \left( \text{ad}^{\mu} \tilde{X}_0, \tilde{X}_1 \right) \left( \text{ad}^{\nu} \tilde{X}_0, \tilde{X}_1 \right) . \]

The number of admissible choices \((\mu,\nu)\) is equal to the number of partitions of \( l \) into 3 nonincreasing nonnegative integers. As shown below this number is

\[ p_{3,l} = \frac{l^2+6l+\pi(l)}{12} \]  \hfill (6c)

where \( \pi(l) = \begin{cases} 12 & \text{if } 6 \mid l \\ 9 & \text{if } 3 \mid l \text{ but } 2 \nmid l \\ 8 & \text{if } 2 \mid l \text{ but } 3 \nmid l \\ 5 & \text{if } 2 \nmid l \text{ and } 3 \nmid l . \end{cases} \]

If \( l = l_1 + l_2 + l_3 \) is an admissible partition of \( l \) in the sense given above (i.e., \( l_1 \geq l_2 \geq l_3 \geq 0 \)), then either \( l_3 = 0 \) or \( l-3 = (l_1-1) + (l_2-1) + (l_3-1) \) is an admissible partition of \( l-3 \).

If \( l_3 = 0 \), then there are \( \frac{l+2}{2} \) distinct choices of admissible pairs \((n_1, n_2)\) if \( n \) is even, and \( \frac{l+1}{2} \) choices if \( n \) is odd. One easily checks that for \( l \leq 2 \) the only admissible partitions are \( 0 = 0 + 0 + 0 \), \( 1 = 1 + 0 + 0 \), \( 2 = 2 + 0 + 0 = 1 + 1 + 0 \), thus giving \( p_{3,0} = p_{3,1} = 1 \) and \( p_{3,2} = 2 \), which shows that validity of (7) for \( l \leq 2 \). Then inductively for \( l \) even
\[ p_{3, l} = p_{3, l-3} + \frac{l+2}{2} = \frac{(l-3)^2 + 6(l-3) + \pi(l-3) + 6(l+2)}{12} = \frac{l^2 + 6l + (\pi(l-3)+3)}{12} \]

and similarly for \( l \) odd

\[ p_{3, l} = p_{3, l-3} + \frac{l+1}{2} = \frac{l^2 + 6l + (\pi(l-3)-3)}{12} \]

which proves (6c) using \( \pi(l) = \pi(l-3) + 3(-1)^l \).

We combine (6a), (6b), and (6c) to check

\[ r_{3, l} + q_{3, l} + p_{3, l} = \dim A^3, (\hat{X}^1, \hat{X}^0) : \]

if \( 6 \mid l \)

\[ r_{3, l} + q_{3, l} + p_{3, l} = \frac{l^2 + 6l + 12}{12} + \frac{l^2 + 2l}{4} + \frac{l^2 + 3l}{6} = \left( \frac{l+2}{2} \right), \]

if \( 3 \mid l \) and \( 2 \nmid l \)

\[ r_{3, l} + q_{3, l} + p_{3, l} = \frac{l^2 + 6l + 9}{12} + \frac{(l+1)^2}{4} + \frac{l^2 + 3l}{6} = \left( \frac{l+2}{2} \right), \]

if \( 3 \nmid l \) and \( 2 \mid l \)

\[ r_{3, l} + q_{3, l} + p_{3, l} = \frac{l^2 + 6l + 8}{12} + \frac{l^2 + 2l}{4} + \frac{(l+2)(l+1)}{6} = \left( \frac{l+2}{2} \right), \]

if \( 3 \nmid l \) and \( 2 \nmid l \)

\[ r_{3, l} + q_{3, l} + p_{3, l} = \frac{l^2 + 6l + 5}{12} + \frac{(l+1)^2}{4} + \frac{(l+2)(l+1)}{6} = \left( \frac{l+2}{2} \right). \]

To show that \( \{ \hat{W}_3^j, J \text{ as defined above} \} \) is a basis for \( A^{3, l}(\hat{X}^1, \hat{X}^0) \), we thus only have to show linear independence. We proceed by giving the associated control system, dilation and \( \psi \).

\[
\begin{align*}
x_{0} & = u \\
x_{j+1} & = x_{j} & 1 \leq j+1 \leq l, \\
y_{2i+1} & = \frac{1}{2}x_{i}^{2} & 1 \leq 2i+1 \leq l, \\
y_{i,j+1} & = y_{ij} & 1 \leq 2i-1 \leq j \leq l-1, \\
z_{i,j,2i+j+1} & = \alpha(i,j) x_{i}^{2} x_{j} & 0 \leq i \leq j, 1 \leq 2i+j+1 \leq l, \\
z_{i,j,h+1} & = z_{i,j,h} & 0 \leq i \leq j, 1 \leq 2i+j+1 \leq h \leq l-1, \\
\end{align*}
\]
and define

\[ \psi(x, y, z) = (z_{i,j}, 0 \leq i \leq l-2i-1, \]
\[ y_{i,j} = 1 \leq 2i+1 \leq j \leq l \]
\[ \alpha(i,j,l-i-j)x_i x_{l-i-j}, 0 \leq i \leq l-i-j \]

in a suitable ordering of the components (where \( \alpha(\cdot) \) is defined as in (5)).

With the dilation \( \Delta_{\epsilon, \delta}(x, y, z) = (\epsilon \delta^{i+1} x_i, \epsilon^2 \delta^{i+2} y_{i,j}, \epsilon^3 \delta^{i+3} z_{i,j,h}) \) we again have \( \psi \in H_{3, l} \), \( X^0 \in \mathfrak{u}_{(0, -1)} \) and \( X^1 \equiv \frac{\partial}{\partial x_0} \in \mathfrak{u}_{(-1, -1)} \). We verify that

\[ W = \{ W_3^{3, l} \} \] as above is "diagonal" on \( \psi \) at zero (modulo permutation) and then read off the coefficients \( b_3^{3, l} \) of \( W_3^{3, l} \).

Again, first observe that for \( 1 \leq j \leq l \)

\[ (ad^1 X^0_j, X^1) = \frac{\partial}{\partial x_j} + x_{j-1} \frac{\partial}{\partial y_{j-1} z_{j-1}} + R(j), \]

the second term vanishing for \( j-1 \geq l \) and \( R(\nu) \) this time being a sum of vectorfields of the form \( rx_{\lambda} \frac{\partial}{\partial y_{i,j}} \) with \( r \in \mathbb{R} \) and \( \lambda < j-1 \) and of the form \( rx_{\lambda} \frac{\partial}{\partial z_{i,j,h}} \) with \( r \in \mathbb{R} \) and \( \lambda < j-1 \). This immediately gives

\[ W_{3, l}^{3, l}(0) = \frac{\partial^3}{\partial x_{\lambda} \partial y_{i,j} \partial x_{\mu}} \bigg|_0. \]

Next observe

\[ (ad^2 (ad^1 X^0_j, X^1); X^0)(0) = \frac{\partial}{\partial y_{i,j+1}} \bigg|_0 \]

and thus

\[ (ad^{i-2} X^0_j, (ad^2 (ad^1 X^0_j, X^1); X^0)(0) = \frac{\partial}{\partial y_{i,j}}, 2i+1 \leq j \leq l. \]

Since \((ad^1 X^0_j, X^1)\) does not contain any \( y \) in its coefficients

\[ W_{3, l}^{3, l}(0) = (ad^{i-2} \mu \nu X^0_j, (ad^2 (ad^1 X^0_j, X^1)); X^0)(0) = \frac{\partial^2}{\partial y_{i,j+1} \partial x_{\nu}} \bigg|_0. \]

Next check
\[(\text{ad}^2 (\text{ad}^\nu X^0, X^1), X^0) = \frac{\partial}{\partial y_{\nu, 2\nu + 1}} + \sum_{j=\nu}^{i-1-2\nu} x_j \frac{\partial}{\partial z_{\nu, j, 2\nu + j + 1}} + R(\nu)\]

where \(R(\nu)\) is a sum of vector fields of the form \(x_j \frac{\partial}{\partial z_{h, r}}\) with both \(j < \nu\) and \(h < \nu\). Thus using that \((\text{ad}^\lambda X^0, X^1)\) does not contain any \(y\) or \(z\) in its coefficients we obtain for \(\mu \geq \nu\)

\[W_{1,\mu, \nu}^{\lambda, i} = (\text{ad}^{i-1-2\nu-\mu} X^0, [(\text{ad}^\mu X^0, X^1); (\text{ad}^\nu X^0, X^1)](0)) = \frac{\partial}{\partial z_{\nu, \mu, i}}\]

Finally we can read off the coefficients \(b_{1, \mu, \nu}^{\lambda, i}(t,u)\)

\[b_{1, \mu, \nu}^{\lambda, i}(t,u) = z_{\nu, \mu}(t,u) = \alpha(\mu, \nu, \nu) \int_0^t \int_0^t \cdots \int_0^t \left( \int_0^t \cdots \int_0^t u \right)^2 (\int_0^t \cdots \int_0^t u)^{(i-2\nu-\mu)}(\int_0^t \cdots \int_0^t u)^{(i+1)}(\int_0^t \cdots \int_0^t u)^{(\nu+1)} \right) ds\]

\[= \alpha(\mu, \nu, \nu) \int_0^t \frac{(t-s)^{i-1-2\nu-\mu}}{(i-1-2\nu-\mu)!} \int_0^s \frac{(s-\sigma)^{\mu}}{\mu!} u(\sigma) d\sigma \int_0^s \frac{(s-\sigma)^{\nu}}{\nu!} u(\sigma) d\sigma ds\]

\[b_{2, \mu, \nu}^{\lambda, i}(t,u) = y_{\mu, \nu-\mu}(t,u) x_{\nu}(t,u) = \frac{1}{2} \left( \int_0^t \int_0^t \cdots \int_0^t \left( \int_0^t \cdots \int_0^t u \right)^2 (\int_0^t \cdots \int_0^t u)^{(i-2\nu-\mu)}(\int_0^t \cdots \int_0^t u)^{(i+1)}(\int_0^t \cdots \int_0^t u)^{(\nu+1)} \right) ds\]

\[= \frac{1}{2} \int_0^t \frac{(t-s)^{i-1-2\mu-\nu}}{(i-1-2\mu-\nu)!} \left( \int_0^s \frac{(s-\sigma)^{\mu}}{\mu!} d\sigma \right)^2 ds \int_0^t \frac{(t-s)^{\nu}}{\nu!} u(s) ds\]

\[b_{3, \mu, \nu}^{\lambda, i}(t,u) = \alpha(i-\mu-\nu, \mu, \nu) x_{i-\mu-\nu}(t,u) x_{\mu}(t,u) x_{\nu}(t,u)\]

\[= \alpha(i-\mu-\nu, \mu, \nu) \left( \int_0^t \int_0^t \cdots \int_0^t \left( \int_0^t \cdots \int_0^t u \right)^2 (\int_0^t \cdots \int_0^t u)^{(i-2\mu-\nu+1)}(\int_0^t \cdots \int_0^t u)^{(\mu+1)}(\int_0^t \cdots \int_0^t u)^{(\nu+1)} \right) ds\]

\[= \alpha(i-\mu-\nu, \mu, \nu) \int_0^t \frac{(t-s)^{i-1-2\mu-\nu}}{(i-\mu-\nu)!} u(s) ds \int_0^t \frac{(t-s)^{\mu}}{\mu!} u(s) ds \int_0^t \frac{(t-s)^{\nu}}{\nu!} u(s) ds\]

2.3.4 Estimating \((X^1 \phi)(0)\) for \(I_1 \leq k_0\):

In [32] the following lemma is proven.
Lemma 4: If \( X^0, \ldots, X^m \) are real analytic vectorfields on \( M \), \( \phi: M \to \mathbb{R} \) analytic and \( K \subseteq M \) compact, then there is a constant \( C < \infty \) such that 
\[
|(X^i \phi)(x)| \leq r!C^r
\]
for all \( x \in K \), all \( r \in \mathbb{Z}_0^+ \) and all choices of multi-indices \( I = (i_1, \ldots, i_m) \in \{0,1,\ldots,m\}^m \).

Here we need a similar estimate, which is slightly modified according to our new summation for \( I_1 \leq 3 \), and makes heavy use of \( X^0(0) = 0 \). Since we only have to evaluate derivatives at zero, we only consider this special case:

Lemma 5: If \( X^0 \) and \( X^1 \) are real analytic vectorfields on \( \mathbb{R}^m \) with \( X^0(0) = 0 \) and \( \phi: \mathbb{R}^n \to \mathbb{R} \) is analytic, then there is a constant \( C = C(s) < \infty \) such that for all multi-indices \( I \in (k,l) \) with \( k \leq s \)
\[
|(X^i \phi)(0)| \leq (sC_0)^{k+l} = C^{k+l}.
\]
Since a product of Lie polynomials in \( A_k^{(k,l)} \) can always be written as a sum of at most \( 2^{k+l} \) monomials \( \pm X^i \), an immediate consequence is

Corollary 6: In lemma 5 \( X^i \) may be replaced by a product of Lie polynomials in \( A_k^{(k,l)} \) resulting only in having to replace \( C \) by \( 2C \).

Lemma 5 and corollary 6 become trivial if \( X^1(0) = 0 \). We thus may assume \( X^1(0) \equiv \frac{\partial}{\partial x_1} \) relative to some suitable coordinates.

To prove lemma 5, we need to introduce some notation: For \( r \in \mathbb{Z}_0^+ \) and \( f: \mathbb{R}^n \to \mathbb{R}^m \) define
\[
\|f\|_r = \sum_{i=1}^m \sum_{|\alpha| \leq r} \frac{1}{\alpha!} \|D^{\alpha} f_i(0)\|,
\]
where \( |\alpha| \leq r \) has to be read as \( \alpha \) ranges over all multi-indices.
\[ \alpha = (\alpha_1, \cdots, \alpha_n), \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i \leq r, \text{ and } \alpha!, D^\alpha \text{ a.s.o. defined as usual.} \]

Similarly, for a \( C^\infty \)-vectorfield \( X = \sum_{i=1}^{n} a_i(x) \frac{\partial}{\partial x_i} \in \mathcal{T}R^n \) let

\[ \|X\|_r = \|a\|_r, \text{ where } a = (a_1, \cdots, a_n): \mathbb{R}^n \rightarrow \mathbb{R}^n. \]

With \( fg = \sum_{i=1}^{m} f_i g_i \) (for \( f, g: \mathbb{R}^n \rightarrow \mathbb{R}^m \)) the usual "dot-product",

\[ \|fg\|_r \leq \|f\|_r \|g\|_r \text{ and } \|fg\|_r \leq \|f\|_r \|g\|_{r-1} \text{ if } f(0) = 0, \text{ which come from the following short calculation, first done for } m = 1 \text{ only} \]

\[ \|fg\|_r = \sum_{|\beta| \leq r} \frac{1}{\alpha!} \|D^\alpha (fg)(0)\| \]

\[ \leq \sum_{|\beta| + \gamma \leq r} \frac{1}{(\beta + \gamma)!} \|D^\gamma f(0)\| \|D^\beta g(0)\| \]

\[ \leq \sum_{|\beta| \leq r} \|D^\beta f(0)\| \|D^\gamma g(0)\| = \|f\|_r \|g\|_r \]

If \( f(0) = 0 \), then in the second sum all terms corresponding to \( \gamma = 0 \) or equivalently \( |\beta| = r \) vanish, and thus in the last two sums \( |\beta| = r \) can be replaced by \( |\beta| \leq r-1 \), thus giving \( \|g\|_r \) instead of \( \|g\|_{r-1} \).

If \( m > 1 \),

\[ \|fg\|_r = \sum_{i=1}^{m} \|f_i g_i\|_r \leq \sum_{i=1}^{m} \|f_i\|_r \|g_i\|_r \]

\[ \leq (\sum_{i=1}^{m} \|f_i\|_r)(\sum_{i=1}^{m} \|g_i\|_r) = \|f\|_r \|g\|_r, \]

the case \( f(0) = 0 \) treated analogously.

Next observe that for \( f: \mathbb{R}^n \rightarrow \mathbb{R}, \|\nabla f\|_{r-1} \leq r \cdot \|f\|_r \), coming from
\[ \| \nabla f \|_{r-1} = \sum_{|\alpha| \leq r-1} \frac{1}{\alpha!} \sum_{i=1}^{n} |D^\alpha \frac{\partial f}{\partial x_i}(0)| \]

\[ = \sum_{1 \leq |\alpha| \leq r-1} \sum_{\alpha_i \neq 0} \frac{1}{(\alpha-e_i)!} \| (D^\alpha f)(0) \| \leq r \sum_{|\alpha| \leq r} \| (D^\alpha f)(0) \| \]

where \( e_i = (0, \cdots, 0, 1, 0, \cdots, 0) \) is the multi-index with a one in the \( i \)-th entry and zero elsewhere, and we use

\[ \sum_{i=1}^{n} \frac{1}{(\alpha-e_i)!} = \sum_{i=1}^{n} \frac{\alpha_i}{\alpha!} = \frac{|\alpha|}{\alpha!} \leq \frac{r}{\alpha!}. \]

With these tools, lemma 5 is easily proven:

\[ \| X^I f \|_{\lambda-1} \leq \| \nabla f \|_{\lambda-1} \leq \lambda \cdot \| f \|_{\lambda} \quad \text{and} \]

\[ \| X^0 f \|_{\lambda} = \| a \cdot \nabla f \|_{\lambda} \leq \| a \|_{\lambda} \cdot \| \nabla f \|_{\lambda-1} \leq \lambda \| a \|_{\lambda} \| f \|_{\lambda} \]

which hold for all \( f \in C^\infty(\mathbb{R}^n, \mathbb{R}) \), \( \lambda \in \mathbb{Z}^+ \).

If \( C \geq \max\{ r \| X^0 \|_r, \ r \| \phi \|_r \} \), one shows inductively \[ \| X^I \phi(0) \| = \| X^I \phi \|_0 \leq C^{k+l} \] for all \( I \in \mathcal{M}(k,l) \) and \( k \leq r \) (induction on the length \( k+l \) of \( I \)). For \( J \in \mathcal{M}(k,l) \) with \( k \leq r, \lambda \leq r \), \[ \| X^0 X^J \phi \| \leq \lambda \| X^0 \|_{\lambda} \| X^J \phi \|_{\lambda} \leq C \| X^J \phi \|_{\lambda} \] and for \( J \in \mathcal{M}(k,l) \) with \( k < r, \lambda \leq r \) \[ \| X^1 X^J \phi \|_{\lambda-1} \leq \lambda \| X^J \phi \|_{\lambda} \] and thus for all \( I \in \mathcal{M}(k,l) \) with \( k \leq r \)

\[ \| X^I \phi \|_0 \leq k! k^l \| X^0 \|_k^l \| \phi \|_k \leq k^{k+l} \| X^0 \|_k^l \| \phi \|_k \leq C^{k+l}. \]

(If \( k = 0 \), we immediately have \( \| X^I \phi \|_0 = (X^I \phi)(0) = 0 \).)

### 2.3.5 Some integral inequalities

In §2.3.3 we already used several times

**Lemma 7**: For \( k \geq 1 \), \( f: [0; t_0] \to [-\varepsilon, \varepsilon] \) measurable,
\[ \int \cdots \int f(t_1)dt_1 \cdots dt_k = \int \frac{(t-s)_\nu^{k-1}}{(k-1)!} f(s)ds, \]

which we briefly verify by induction on \( k \). For \( k = 1 \), the statement is trivially true, and the induction step is equivalent to showing

\[ \int_0^t \frac{(t-s)_\nu^{\nu+1}}{\nu!(\nu+1)!} g(s)ds = \int_0^t \int_0^s \frac{(s-\sigma)_\nu^\nu}{\nu!} g(\sigma)d\sigma ds, \]

which obviously is true for \( t = 0 \), and differentiating both sides w.r.t. \( t \) yields

\[ \frac{(t-t)_\nu^{\nu+1}}{(\nu+1)!} g(t) + \int_0^t \frac{(t-s)_\nu^\nu}{\nu!} g(s)ds, \]

on the left side and

\[ \int_0^t \frac{(t-\sigma)_\nu^\nu}{\nu!} g(\sigma)d\sigma \]

on the right side, both expressions being continuous in \( t \), and the lemma follows by uniqueness of solutions of initial value problems.

In the proof of theorem 1, two iterated integrals require special attention. We formulate the needed inequalities in

**Lemma 8.** For every \( t \geq 0 \), \( u: [0, t] \to [-\epsilon, \epsilon] \) measurable

\[ \frac{3}{2} \epsilon^2 \int_0^t (\int_0^s u)^2 \geq \int_0^t (\int u)^4 \quad \text{and} \quad (9) \]

\[ \frac{1}{2} \epsilon \int_0^t (\int_0^s u)^2 \geq \int_0^t (\int u)^2 |\int u| \quad \text{if} \int_0^t u = 0. \]

(10)
To prove lemma 8 we consider the following control system on $\mathbb{R}^4$
(with right side $C^0$, only).

\[
\begin{align*}
\dot{x}_1 &= u & x(0) &= 0 \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= \frac{1}{2} \epsilon x_2^2 - x_1^2 |x_2| \\
\dot{x}_4 &= \frac{3}{2} \epsilon^2 x_2^2 - x_1^4
\end{align*}
\] (11)

and we show that $x_3(t,u) \geq 0$, $x_4(t,u) \geq 0$ if $x_1(t,u) = 0$. Define two $C^1$ functions $\psi_1, \psi_2 : \mathbb{R}^4 \to \mathbb{R}$ by

\[
\begin{align*}
\psi_1(x) &= x_3 + \frac{1}{2} x_1 x_2 |x_2| \\
\psi_2(x) &= x_4 + x_1^3 x_2 + \frac{3}{2} \epsilon x_1 x_2 |x_2|
\end{align*}
\]

and compute

\[
(X^0 + uX^1) \psi_1(x) = \frac{1}{2} |x_2| (\epsilon |x_2| + ux_2) \geq 0 \quad \text{and}
\]

\[
(X^0 + uX^1) \psi_2(x) = \left( \frac{3}{2} \epsilon |x_2| + 3x_1^2 \right) (\epsilon |x_2| + ux_2) \geq 0.
\]

Thus for $i = 1, 2$, we have $\psi_i(0) = 0$, $t \to \psi_i(x(t,u))$ nondecreasing and $\psi_1(x) = x_3$, $\psi_2(x) = x_4$ when $x_1 = 0$.

Finally, the proof of theorem 1 heavily relies on the following estimate:

**Lemma 9:** For a control system of form (1), $\phi : \mathbb{R}^8 \to \mathbb{R}$ analytic, there is a constant $K < \infty$ depending on $\phi$ and $X^0$, but not on $u$, s.t.

\[
\left| \sum_{1 \leq i \leq 4} a_i(t,u)(X^i \phi)(0) \right| \leq Kt \int_0^t \left( \int_0^s u(\sigma) d\sigma \right)^4 ds
\] (12)
for \( t \in [0; t_0], \ t_0 > 0 \) sufficiently small and when \( \int_0^t u(s)ds = 0. \)

Lemma 9 is an essential part of the proof of theorem 0.4 in [27], and since the proof of this lemma 9 is very lengthy and technical we shall refer the interested reader to the original papers [26, 27].

(Remark: What we formulated as lemma 9 is part of the so called Property (P) in [27]. Property (P) is proven only for a special type of function \( \phi \), but the special form of \( \phi \) is only needed for \( 0 = \sum_{l \leq 3} |a_1(t,u)(X^l \phi)(0)| \), whereas the second part of the proof of property (P), the same as for lemma 9, is independent of this special form of the function \( \phi \).)

2.4 Proof of theorem 1.

We start with some remarks towards the length of the proof.

The proof of theorem 1 shall be expected to be considerably more complicated than the one of theorem 0.4 for reasons explained below. Indeed, it will by using lemma 9 (§2.3.5) almost use a special case of theorem 0.4. In proving theorem 0.4 one uses that the sum \( \sum_{l} a_l(X^l \Phi)(0) \) clearly splits into terms of lower order and terms of higher order relative to the term in question, say \( 1^0 \in m(2k,1) \). All lower order terms corresponding to multi-indices \( I \in m(j,l) \) with \( j < 2k \) vanish by a sophisticated choice of the function \( \Phi \) using the filtration \( \{ \phi^i; i \in \mathbb{Z}_0^+ \} \) on \( L(X^1, X^0) \). On the other hand all higher order terms corresponding to multi-indices \( I \in m(j,l) \) with \( j = 2k \) and \( l > 1 \) or \( j > 2k \) can be estimated to be in absolute values smaller than the only nontrivial term of type \( (2k,1) \). (Compare lemma 9 (§2.3.5) and the proof of
In our case the situation is considerably more intricate as even \( \dim L^{(2,3)}(\hat{X}^1, \hat{X}^0) > 1 \) and terms corresponding to multi-indices \( I \) with \( I_1 = 2 \) or \( I_1 = 3 \) may or may not dominate the term(s) in question of type \((2,3)\). To clear up this situation the machinery in §2.3.4 has been developed.

To see how sharp theorem 1 is the interested reader is referred to [22] where it is shown that (1) may be STLC whenever the hypothesis in theorem 1 holds, i.e., whenever \( X^\nu(0) = (\text{ad}^2[X^1, X^0]; X^0)(0) \) is linearly dependent on ("neutralized by") \{\( (\text{ad}^\nu X^0; X^1)(0), (\text{ad}^\nu X^0; (\text{ad}^3 X^1; X^0))(0): \nu \in \mathbb{Z}_0^+ \)\}.

A considerably less sharp condition, like: "If \( X^\nu(0) \) is linearly independent from \( \langle \omega^1 + \omega^2 + \omega^3 \rangle(0) \), then (1) is not STLC" should be considerably easier to prove and probably be relatively easily extendible to: "If \( \langle \text{ad}^2(\text{ad}^\nu X^0; X^1); X^0 \rangle(0) \) is linearly independent from all other brackets in
\[
\sum_{i=0}^{2\nu+1} \omega^i(0),
\]
then (conjecture!) (1) is not STLC."

To get a similarly sharp condition as theorem 1 for \( \nu > 1 \), one has to find out specifically which brackets in \( L(X^1, X^0) \) can dominate \( X^r = (\text{ad}^2(\text{ad}^\nu X^0; X^1); X^0) \). Clearly these include all brackets in \( \omega^\mu \) and all those of the form \( (\text{ad}^\mu X^0; X^\theta) \), \( \mu \in \mathbb{Z}_0^+ \), where \( X^\theta \) is of lower weight than \( X^r \) relative to \( \theta = 1 \). (Observe that the latter brackets \( (\text{ad}^\mu X^0; X^\theta) \) are in general not Hall elements.) But there are even more, for example, if \( \nu = 4 \), then \( (\text{ad}^4 X^0; (\text{ad}^2 X^0, (\text{ad}^3 X^1; X^0)); X^0) \) of type \((6, 7 + \mu)\) can dominate \( X^\nu \) of type \((2, 9)\) although it is not of the form \( (\text{ad}^\mu X^0; X^\theta) \) with \( X^\theta \) of lower weight than \( X^\nu \) for \( \theta = 1 \). (This can easily be shown when using control variations as in
§2.5.4.)

We will prove the contraposition of theorem 1, i.e., we assume

$$(\text{ad}^2[X^0;X^1];X^0)(0) \notin \text{span} \{(\text{ad}^\nu X^1;X^0)(0), (\text{ad}^\nu X^0;\text{ad}^3 X^1;X^0)(0) : \nu \in \mathbb{Z}_0^+\}$$

and then show that (1) is not STLC.

At one point we have to integrate by parts and therefore have to assume that $u$ is absolutely continuous, which can be done w.l.o.g. since it even suffices to consider $u$ piecewise constant. We later will choose local coordinates s.t. $X^1 \equiv \frac{\partial}{\partial x_1}$. By means of state feedback, which does not affect STLC, one always can achieve $<dx_1,X^0> \equiv 0$ and thus $<dx_1,X^*> \equiv 0$ for all Lie monomials $X^*$ of type $(k,l) \neq (1,0)$. We assume this holds throughout.

Choose $n$ vectorfields $Z^1, \cdots Z^n \in L(X^1,X^0)$ which are linearly independent at $x = 0$ and are such that

$$Z^i = (\text{ad}^{i-1}X^0;X^1), \quad i = 1,2, \cdots d_1,$$

$$Z^i = (\text{ad}^{i-d_1-1}X^0;\text{ad}^3 X^1;X^0), \quad i = d_1+1, \cdots d_2,$$

$$Z^d = (\text{ad}^2 [X^0;X^1];X^0), \quad d = d_2 + 1,$$

where $d_1$ and $d_2$ are maximal. By claim 2 (§2.3.2) we necessarily have $d_1 \geq 2$.

Let $\gamma(x) = (\exp x_1 Z^1) \cdots \circ (\exp x_n Z^n)(0)$. Then $\gamma^{-1}$ defines local coordinates $(x_1, \cdots x_n)$ in a neighborhood $U$ of $x_0 = 0$. From now on we will work entirely in this neighborhood $U$ and identify a point $\gamma(x)$ in $U$ with its coordinates $x$.

Let $\Phi: U \to \mathbb{R}$ be defined by $\Phi(x) = x_d = (\gamma^{-1})_d$. We will show that $\Phi(x(t,u)) \geq 0$ if $x_1(t,u) = 0$ and both $t \geq 0$ and the controlbound $\epsilon \geq |u(\cdot)|$
are sufficiently small.

The estimates on the next pages might be a notational nuisance, however they illustrate geometrically why certain brackets can while others cannot "dominate" \( X^\tau = (\text{ad}^2[X^0, X^1]; X^0) \) and what role the higher order partial differential operators play, namely essentially they come from a not yet perfect choice of local coordinates. That the estimates are not that technical can be seen in that we even can drop all factorials in the denominators.

For terms corresponding to \( I_1 = 1 \) we use that each \((\text{ad}^1 X^0; X^1)(0)\) is a linear combination of \( \{Z^1(0); \cdots Z^{d_1}(0)\} \) and obtain

\[
\sum_{l=1}^{\infty} a_l(t, u)(X^1 \Phi)(0) = \sum_{l=1}^{\infty} a_{l,0, \ldots, 0}(t, u)(\text{ad}^l X^0; X^1)\Phi(0) = 0.
\]

Terms corresponding to \( I_1 = 2 \).

We may assume \((\text{ad}^2 X^1; X^0)(0) \in \text{span} \nu^1(0)\) and thus \((\text{ad}^\nu X^0; (\text{ad}^2 X^1, X^0))(0) \in \text{span} \nu^1(0)\) for all \( \nu \in \mathbb{Z}_0^+ \) since otherwise (1) is not STLC by theorem 0.4a [32]. Similarly to the case \( I_1 = 1 \) this leads to

\[
(W_{l,0}^2 \Phi)(0) = (\text{ad}^{l-1} X^0; (\text{ad}^2 X^1, X^0))\Phi(0) = 0, \ l = 1, 2, \cdots;
\]

and because of \( X^1 \Phi \equiv 0 \) one obtains

\[
(W_{l,0}^2 \Phi)(0) = (\text{ad}^l X^0; X^1) X^1 \Phi(0) = 0, \ l = 1, 2, \cdots.
\]

Special consideration is required for the terms

\[
(W_{l,0}^2 \Phi)(0) = (\text{ad}^{l-1} X^0, X^1)[X^0, X^1]\Phi(0), \ l = 2, 3, \cdots.
\]

If \( l-1 \leq d_1 \), then \((W_{l,0}^2 \Phi)(0) = 0\) by construction of \( \Phi \). If \( l-1 > d_1 \), use that there are \( \lambda_{l, \nu} \in \mathbb{R} \) such that
\( (\text{ad}^{l-1}X^0, X^1)(0) = \sum_{\nu=2}^{d_1} \lambda_{l,\nu} (\text{ad}^{\nu}X^0, X^1)(0) \)

and hence by the same argument as above \((W^{2,1}_{2,l-1} \Phi)(0) = 0\) for \(l-1 \geq d_1\) and thus for all \(l = 1, 2, \ldots\).

The next interesting term is \((W^{2,3}_{1,1} \Phi)(0) = 1\), and if we let

\[
v(s) = \int_0^s \int_0^{\tau} u(\tau)d\tau d\sigma,
\]

then

\[
b^{2,3}_{1,1}(t,u)(W^{2,3}_{1,1} \Phi)(0) = \frac{1}{2} \int_0^t v^2(s)ds.
\]

These are the only \(J\) such that \((W^{2,J}_{2,l} \Phi)(0)\) necessarily vanishes or is known as for \(J = (1,1)\). For the others we use that by lemma 5 (§2.3.4) there is a constant \(C < \infty\) such that \(|(W^{2,J}_{1,l} \Phi)(0)| \leq C^{k+l}\) for \(k = 1, 2, 3; l = 1, 2, \ldots\)

and all \(J\). Recall that \(C\) only depends on \(\|X^0\|_3\) and \(\|\Phi\|_3\). Recall

\[
W_{1,\nu}^{2,l} = (\text{ad}^{l-1-2\nu}X^0, (\text{ad}^2(\text{ad}^{\nu}X^0, X^1); X^0)), \ 0 \leq 2\nu \leq l-1
\]

and

\[
W_{2,\nu}^{2,l} = (\text{ad}^{\nu}X^0, X^1)(\text{ad}^{l-\nu}X^0, X^1), \ \frac{l}{2} \leq \nu \leq l.
\]

For \(\nu = 1, l \geq 4\) we find

\[
|b^{2,1}_{1,1}(t,u)| = \frac{1}{2} \int_0^t \frac{(t-s)^{l-3}}{(l-3)!} v^2(s)ds \leq t^{l-3} \int_0^t v^2(s)ds,
\]

for \(4 \leq 2\nu \leq l-1\) we use Schwarz inequality to find

\[
|b^{2,1}_{1,\nu}(t,u)| = \frac{1}{2} \left| \int_0^t \frac{(t-s)^{l-2\nu-1}}{(l-2\nu-1)!} \left( \int_0^s \frac{(s-\sigma)^{\nu-2}}{\nu-2)!} v(\sigma)d\sigma \right)^2 ds \right|
\]
\[
\leq \frac{1}{2} t^{l-2\nu-1} \int_0^t \left[ \int_0^s \frac{(s-\sigma)^{2\nu-4}}{((\nu-3))!} d\sigma \right] \left( \int_0^s v^2(\sigma) d\sigma \right) ds
\]

\[
\leq t^{-2\nu-1} t^{2\nu-3} \int_0^t v^2(s) ds = t^{l-3} \int_0^t v^2(s) ds .
\]

Letting \( c_\nu(t,u) = \int_0^t \frac{(t-s)^\nu}{\nu!} v(s) ds, \ \nu = 0, 1, 2, \ldots \), we use Schwarz inequality to find

\[
|c_\nu(t,u)| \leq t^{\nu+1} \left( \int_0^t v^2(s) ds \right)^{\frac{1}{2}}.
\]

Noting that for \( \frac{l}{2} \leq \nu \leq l-2 \), \( b_{2,\nu}^l(t,u) = c_{\nu-2}(t,u) c_{l-\nu-2}(t,u) \) leads to

\[
|b_{2,\nu}^l(t,u)| \leq t^{\nu-3} t^{l-\nu-3} \int_0^t v^2(s) ds = t^{l-3} \int_0^t v^2(s) ds .
\]

Summing over \( I_1 = 2 \) with the exception of \( J = (1,1) \) gives

\[
\sum_{l_1=2} a_{l_1}(t,u)(X^1\Phi)(0) - b_{2,\nu}^l(t,u)(W_{2,\nu}^2,\Phi)(0) \leq \sum_{l=4}^{\infty} \sum_{\nu=2}^{l-1} C^{l+2} t^{l-3} \int_0^t v^2(s) ds
\]

\[
\leq C^5 \sum_{i=1}^{\infty} (i+2)(Ci)^i \int_0^t v^2(s) ds \leq \frac{1}{12} \int_0^t v^2(s) ds
\]

(14)

if \( t > 0 \) is sufficiently small.

Terms corresponding to \( I_1 = 3 \).

Of the vector fields \( W_{1,\mu,0}^3 \) defined in §2.3.3 there are two sets which require special attention. If \( \mu = \nu = 0 \), then \( W_{1,0,0}^3 = -(ad^{l-1}X^0, (ad^3 X^1, X^0)) \).

If \( l-1 \leq d_2 - d_1 - 1 \), clearly \( (W_{1,0,0}^3, \Phi)(0) = 0 \). If \( l-1 > d_2 - d_1 - 1 \), one can replace \( W_{1,0,0}^3(0) \) by a linear combination of \( \{Z^1(0), \cdots Z^{d_2}(0)\} \) and also
obtains \((W^{3,l}_{1,0,0}\Phi)(0) = 0\).

If \(\nu = 0\), \(\mu \geq 1\), we need inequality (10) from lemma 8 \((\S 2.3.5)\) and compute

\[
|b^{3,l}_{1,\mu,0}(t,u)| = \frac{1}{2} \left| \int_0^t \frac{(t-s)^{l-\mu-1}}{(l-\mu-1)!} \left( \int_0^s \frac{(s-\sigma)^\mu}{\mu!} u(\sigma)d\sigma \right) \left( \int_0^s u(\sigma)d\sigma \right)^2 ds \right|
\leq \frac{1}{2} t^{l-\mu-1} \int_0^t s^{\mu-1} |v(s)| \left( \int_0^s u(\sigma)d\sigma \right)^2 ds
\leq ct^{l-2} \int_0^t v^2(s)ds \text{ if } x_1(t,u) = \int_0^t u(s)ds = 0.
\]

For each \(l \geq 2\) there are \((l-1)\) admissible choices for \(1 \leq \mu \leq l-1\) and we obtain

\[
\left| \sum_{l=2}^{\infty} \sum_{\mu=1}^{l-1} b^{3,l}_{l,\mu,0}(W^{3,l}_{1,\mu,0}\Phi)(0) \right| \leq \sum_{l=2}^{\infty} C^l (ct)^{l-2} \int_0^t v^2(s)ds \leq \frac{1}{12} \int_0^t v^2(s)ds \quad (15)
\]

if both \(\epsilon > 0\) and \(t > 0\) are sufficiently small and \(x_1(t,u) = 0\).

If \(\nu = 1\), then \(\mu \geq 1\) and \(l \geq 4\) and straightforward estimation gives

\[
|b^{3,l}_{1,\mu,1}(t,u)| \leq \alpha(\mu,1,1) \left| \int_0^t \frac{1}{(l-3-\mu)!} \left( \int_0^s \frac{(s-\sigma)^\mu}{\mu!} \mu(\sigma)d\sigma \right) v^2(s)ds \right|
\leq t^{l-3-\mu} \epsilon t^{\mu+1} \int_0^t v^2(s)ds = ct^{l-2} \int_0^t v^2(s)ds.
\]

If \(\nu > 1\), then \(\mu > 1\) and \(l > 7\) and
\[ |b_{1,\nu,\nu}^{3,1}(t,u)| \leq \alpha(\mu,\nu,\nu) \left| \int_0^t \left( \begin{array}{c} (t-s)^{1-\mu-2\nu} \\ \mu! \\ (1-\mu-2\nu)! \end{array} \right) \left( \int_0^s \left( \begin{array}{c} (s-\sigma)^\mu \\ \mu! \\ \nu! \end{array} \right) u(\sigma) d\sigma \right) \left( \int_0^s \left( \begin{array}{c} (s-\sigma)^{\nu-2} \\ \nu! \end{array} \right) v(\sigma) d\sigma \right) ds \right| \]

\[ \leq t^{1-\mu-2\nu} \epsilon t^{\mu+1} \int_0^t \left( \int_0^s s^{2\nu-4} ds \right) \left( \int_0^s v^2(s) ds \right) ds \]

\[ \leq \epsilon t^{1-2\nu} t^{2\nu-2} \int_0^t v^2(s) ds = \epsilon t^{1-2} \int_0^t v^2(s) ds \]

Summing over these terms yields

\[ \left| \sum_{i=1}^\infty \sum_{2\nu+\mu \leq i-1} b_{1,\mu,\nu}^{3,1}(t,u)(W_{1,\mu,\nu}^{3,1}(0)) \right| \leq \epsilon C^5 \sum_{i=1}^\infty t^{2}(Ct)^{i-2} \int_0^t v^2(s) ds \]

\[ \leq \frac{1}{12} \int_0^t v^2(s) ds \quad (16) \]

if \( t > 0 \) is sufficiently small.

Of the second-order operators defined in §2.3.3

\[ W_{2,\mu,\nu}^{3,1} = (ad^{1-2\mu-\nu-1} X^0;(ad^2 X^0;X^1);X^0)) (ad^\nu X^0;X^1) \]

those with \( \mu = 0 \) require special attention.

If \( \nu = 0 \), then \( W_{2,0,0}^{3,1} \Phi \equiv 0 \) because of \( X^1 \Phi \equiv 0 \). If \( \nu = 1 \), replace \( (ad^{1-2} X^0;(ad^2 X^1;X^0))(0) \) by a linear combination of \( \{ Z^2(0), \cdots Z^4(0) \} \), i.e. there are \( \lambda_{i,j} \in \mathbb{R} \) such that

\[ W_{2,0,1}^{3,1}(0) = (ad^{1-2} X^0;(ad^2 X^1;X^0))[X^0;X^1] \Phi(0) \]

\[ = \sum_{i=2}^{d_1} \lambda_{i,j} (ad^{i-1} X^0;X^1)[X^0;X^1] \Phi(0) = 0 . \]

If \( \nu \geq 2, \) we no longer can expect \( W_{2,0,\nu}^{3,1}(0) \) to vanish in general, but now its
coefficient $b_{2,\nu}^{3,l}(t,u)$ is sufficiently small. Since we integrate by parts we have
to require that $u$ is absolutely continuous. By using Schwarz inequality we
obtain for $\mu = 0$, $\nu \geq 2$, $l \geq 3$

$$|b_{2,\nu}^{3,l}(t,u)| = \frac{1}{2} \int_0^t \frac{(t-s)^{l-3-\nu}}{(l-3-\nu)!} \left( \int_0^s u(\sigma)d\sigma \right)^2 ds \cdot \int_0^t \frac{(t-s)^{\nu}}{\nu!} u(s)ds$$

$$\leq t^{-l-\nu-1} \int_0^t \left( \int_0^s u(\sigma)d\sigma \right) \left( \int_0^\sigma u(\rho)d\rho \right) ds \cdot \int_0^t \frac{(t-s)^{\nu-2}}{(\nu-2)!} v(s)ds$$

$$= t^{l-\nu-1} \int_0^t u(s)(\int_0^s (t-\sigma) d\sigma)(\int_0^\sigma u(\tau)d\tau)d\sigma \cdot \int_0^t t^{2\nu-4} v(s)ds$$

$$\leq t^{l-\nu-1}(\int_0^t u^2(s)ds)^{\alpha}(\int_0^t v^2(s)ds)^{\alpha}(\int_0^t t^{2\nu-4} ds)^{\alpha}(\int_0^t v^2(s)ds)^{\alpha}$$

$$\leq \epsilon t^{-2} \int_0^t v^2(s)ds \text{ if } x_1(t,u) = \int_0^t u(s)ds = 0.$$

If $\mu = 1$, then for $\nu \geq 0$ and $l \geq 3+\nu$

$$|b_{2,1,\nu}^{3,l}(t,u)| = \frac{1}{2} \int_0^t \frac{(t-s)^{l-3-\nu}}{(l-3-\nu)!} v^2(s)ds \cdot \int_0^t \frac{(t-s)^{\nu}}{\nu!} u(s)ds$$

$$\leq t^{l-3-\nu} \epsilon t^{\nu+1} \int_0^t v^2(s)ds = \epsilon t^{-2} \int_0^t v^2(s)ds;$$

and if $\mu > 1$, then for $\nu > 0$ and $l \geq 5+\nu$

$$|b_{2,1,\mu,\nu}^{3,l}(t,u)| = \frac{1}{2} \int_0^t \frac{(t-s)^{l-1-2\mu-\nu}}{(l-1-2\mu-\nu)!} \left( \int_0^s \frac{(s-\sigma)^{\mu}}{\mu!} \mu(\sigma)d\sigma \right)^2 ds \cdot \int_0^t \frac{(t-s)^{\nu}}{\nu!} u(s)ds.$$
\[
\leq t^{l-1-2\mu-\nu} \int_0^t s^{2\mu-2} v^2(s) ds \cdot \epsilon t^{\nu+1} \leq \epsilon t^{l-2} \int_0^t v^2(s) ds.
\]

Summing over all these terms with \( \mu \geq 1 \) yields

\[
\left| \sum_{l=3}^{\infty} \sum_{\mu=1}^{\infty} \sum_{\nu=1}^{\infty} b_{3,\mu,\nu}(t,u) W_{3,\mu,\nu}^3 \Phi(0) \right| \leq C^5 \epsilon \sum_{l=3}^{\infty} \int_0^t v^2(s) ds
\]

\[
\leq \frac{1}{12} \int_0^t v^2(s) ds \text{ if } \epsilon > 0 \text{ and } t > 0 \text{ are sufficiently small.} \quad (17)
\]

Finally the third order operators defined in \( \S 2.3.3 \) are

\[
W_{3,\mu,\nu}^3 = (\text{ad}^{l-\mu-\nu} X^0; X^1)(\text{ad}^\mu X^0; X^1)(\text{ad}^\nu X^0; X^1)
\]

with \( 0 \leq \nu \leq \mu \leq l-\mu-\nu \). If \( \nu = 0 \), then \( W_{3,0,0}^3 \Phi(0) = 0 \) because of \( X^1 \Phi \equiv 0 \). If \( \nu = \mu = 1 \), then \( W_{3,1,1}^3 \Phi(0) = 0 \) either because \( 1 \leq l-2 \leq d_1 - 1 \) or because we can replace \( (\text{ad}^{l-\mu-\nu} X^0; X^1)(0) \) by a linear combination of \( \{Z^2(0), \ldots, Z^{d_1}(0)\} \).

Thus the only nonvanishing terms come from \( \mu \geq 2 \), \( \nu \geq 1 \) and thus

\( l-\mu-\nu \geq 2 \) and \( l \geq 5 \).

\[
|b_{3,\mu,\nu}(t,u)| = c_{l-\mu-\nu-2}(t,u) c_{\mu-2}(t,u) \int_0^t \frac{(t-s)^\nu}{\nu!} u(s) ds
\]

\[
\leq t^{l-\nu-3} \int_0^t v^2(s) ds \cdot \epsilon t^{\nu+1} = \epsilon t^{l-2} \int_0^t v^2(s) ds
\]

where again we used \( c_\lambda(t,u) = \int_0^t \frac{(t-s)^\lambda}{\lambda!} v(s) ds = \int_0^t \frac{(t-s)^\lambda}{\lambda!} \int_0^s u(\tau) d\tau d\sigma ds \), and

\[
|c_\lambda(t,u)| \leq t^{\lambda+\kappa} \left( \int_0^t v^2(s) ds \right)^\kappa. \text{ Summing over these terms yields}
\]
\[ \left| \sum_{i=2}^{\infty} \sum_{1 \leq j \leq t-u} b_{l,\mu,\nu}^j(t,u) W_{3,\mu,\nu}^j \Phi(0) \right| \leq C \sum_{i=2}^{\infty} t^2 (Ct)^{i-2} \int_0^t \int v^2(s) ds \]  

\[ \leq \frac{1}{12} \int_0^t \int v^2(s) ds \]  

if \( \epsilon > 0 \) and \( t \geq 0 \) are sufficiently small.

Finally by lemmata 8 and 9 (§2.3.5)

\[ \left| \sum_{i \geq 1} a_i(t,u)(X^i \Phi)(0) \right| \leq (1 + Kt) \int_0^t \left( \int_0^s u(\sigma) d\sigma \right) d\sigma \leq \frac{1}{12} \int_0^t \int v^2(s) ds \]

if \( t > 0 \) and \( \epsilon > 0 \) are sufficiently small and \( x_1(t,u) = 0 \). (K is some positive constant depending on \( X^0 \) and \( \Phi \) only, as in lemma 8.)

To finish the proof of theorem 1 we combine the inequalities (13) through (19) and obtain

\[ \left| \sum_{i \geq 1} a_i(t,u)(X^i \Phi)(0) \right| \geq b_{1,2}^2(t,u) W_{1,2}^2 \Phi(0) - \text{all others} \]

\[ \geq \frac{1}{2} \int_0^t v^2(s) ds - 6 \cdot \frac{1}{12} \int_0^t v^2(s) ds = 0 \]

when \( x_1(t,u) = \int_0^t u(s) ds = 0 \) and both \( t > 0 \) and \( \epsilon > 0 \) are sufficiently small.

(We required \( \epsilon > 0 \) and \( t > 0 \) sufficiently small only a finite number of times; thus there are \( t_0 > 0 \) and \( \epsilon_0 > 0 \) sufficiently small for all these inequalities to hold simultaneously.)

2.5 Examples.

2.5.1 A new good bracket of type (4,3). In this paragraph we will show that the system
\[
\begin{align*}
    \dot{x}_1 &= u & x(0) &= 0 \\
    \dot{x}_2 &= x_1 & |u(t)| &\leq \epsilon_0 \\
    \dot{x}_3 &= \frac{1}{5} x_1^3 \\
    \dot{x}_4 &= x_2 x_3
\end{align*}
\]

(20)

is STLC and afterwards discuss the Lie brackets of type (4,3) more generally.

For system (20) the only at \( x_0 = 0 \) nonvanishing brackets (modulo Jacobi identity) are: \( X^1(0) = \frac{\partial}{\partial x_1}, \quad [X^0; X^1](0) = \frac{\partial}{\partial x_2}, \quad -(\text{ad}^3 X^1; X^0) = \frac{\partial}{\partial x_3} \) and \( X^\theta(0) = (\text{ad}^2 X^0; X^1)(0) = \frac{\partial}{\partial x_4} \). The bracket \( X^\theta \) is of type (4,3), not of the form \([X^0; X^\sigma]\) for any \( X^\sigma \) of type (4,2), and for no \( \theta \in [0,1] \) can \( X^\theta(0) \) be written as a linear combination of brackets of strictly lower weight w.r.t. \( \theta \) evaluated at zero. Therefore by all our present knowledge \( X^\theta \) must be regarded as a possible obstruction to STLC of (20).

To show that (20) is STLC observe that if \( \Delta_{\epsilon,\delta}(x) = (\epsilon \delta x_1, \epsilon \delta^2 x_2, \epsilon^3 \delta^4 x_3, \epsilon^4 \delta^4 x_4) \) is a two-parameter family of dilations on \( \mathbb{R}^4 \), then \( X^0 \in \mathfrak{n}_{(0,-1)} \) and \( X^1 \equiv \frac{\partial}{\partial x_1} \in \mathfrak{n}_{(-1,-1)} \) and thus the scaling lemma (lemma 2) applies.

According to the remark at the end of \S 2.3.1 it suffices to show that for some time \( t_0 > 0 \) and some control bound \( \epsilon_0 > 0 \) (we take \( \epsilon_0 = 1 \)) the cone of tangent vectors to \( \mathcal{G}_{t_0}(t_0) \) at zero is all of \( \mathbb{R}^4 \). By using standard control variations one easily generates \( \pm \frac{\partial}{\partial x_i}, \quad i = 1,2,3 \), as tangent vectors. Thus we only have to show that both \( + \frac{\partial}{\partial x_4} \) and \( - \frac{\partial}{\partial x_4} \) are tangent vectors. For
\( \xi, \eta \geq 0 \) let \( T_1 = T_1(\xi) = 4 + 2 \xi \), \( T = T(\xi, \eta) = 2T_1 + \eta \), and define the control function \( u: [0, T] \rightarrow [-1, 1] \) by

\[
u(t) = u^{\xi, \eta} (t) = \begin{cases} 1 & \text{if } t \in [0, 1) \cup [2+\xi, 3+2\xi) \\ -1 & \text{if } t \in [1, 2+\xi) \cup [3+2\xi, T_1) \\ 0 & \text{if } t \in [T_1, T_1 + \eta] \\ -u(t-T_1-\eta) & \text{if } t \in (T_1+\eta, T] \end{cases}\]

![Diagram of control variation](image)

**Figure 4. A control variation**

Then elementary arithmetic gives for all \( s \in [0, \eta] \)

\[
x_1(T_1, u) = x_1(T_1+s, u) = 0,
\]

\[
x_2(T_1, u) = x_2(T_1+s, u) = 2 - \xi^2, \text{ and}
\]

\[
x_3(T_1, u) = x_3(T_1+s, u) = 2 - \xi^4.
\]

Similarly \( x_i(T, u) = 0 \) for \( i = 1, 2, 3 \) and all choices for \( \xi, \eta \geq 0 \). We let

\[
C(\xi) = x_4(T_1, u) = \int_0^{T_1} x_2(s, u)x_3(s, u)ds.
\]

One easily verifies \( C(0) > 0 \) and

\[
x_4(T, u) = 2C(\xi) + \eta x_2(T_1, u)x_3(T_1, u).
\]

Choosing \( \xi = \eta = 0 \), we obtain \( x_4(T, u) > 0 \) and thus \( \frac{\partial}{\partial x_4} \) as a tangent
vector. Choosing $\sqrt{2} < \xi < \frac{2}{\sqrt{2}}$ gives $x_2(T_1,u) > 0$ and $x_3(T_1,u) < 0$, and thus by choosing

$$\eta > \frac{2C(\xi)}{|x_2(T_1,u)x_3(T_1,u)|}$$

we obtain $x_4(T,u) < 0$, giving $-\frac{\partial}{\partial x_4}$ as a tangent vector.

In constructing the control $u^{\xi,\eta}$ explicitly we made use of that $\dot{x}_i = 0$ for $i = 1,2,3$ on the subspace $E_1 = \{x \in \mathbb{R}^4: x_4 = 0\}$ when $u = 0$, and that on the intersection of $E_1$ with any neighborhood of $x = 0$ $\dot{x}_4 = x_2x_3$ takes both positive and negative values.

However we conjecture that in much more general control systems if $+X^\xi(0)$ can be generated as a tangent vector, then so can $-X^\xi(0)$.

Recall, that the ideas in theorems 0.2 and 0.3 were: If the control $u$ generates the tangent vector $\xi$, then some variation of $-u$ or of the time-reversed control $u^{-1}$ may generate the tangent vector $-\xi$.

But here it is not at all clear, what the relation between the controls generating $+X^\xi(0)$ and $-X^\xi(0)$ as tangent vectors is in general.

**Note:** Example 1 generalizes in an obvious way when replacing $\dot{x}_3 = \frac{1}{6}x_1^3$ in (20) by $\dot{x}_3 = \frac{1}{(2k-1)!} x_1^{2k-1}$ for any $k = 2,3,4,\ldots$. The computations remain essentially unchanged, and if one chooses $\frac{2}{\sqrt{2}} < \xi < \frac{2}{\sqrt{2}}$ one obtains

$$-X^\xi(0) = -[(\text{ad}^{2k-1}X^1;X^0);(\text{ad}^2X^1;X^0)](0) = -\frac{\partial}{\partial x_4}$$

as a tangent vector, where $X^\xi$ is of type $(2k,3)$. $+\frac{\partial}{\partial x_4}$ is again obtained
from choosing $\xi = \eta = 0$.

From equation (2), §2.3.3, one computes $\dim L^{(4,3)}(\hat{X}^1, \hat{X}^0) = 5$. Although in general it does not seem clear how to choose a practical basis for $L^{(k,l)}(\hat{X}^1, \hat{X}^0)$ if $k \geq 4$, we here present a reasonably good choice in case $(k,l) = (4,3)$.

$$\hat{W}_1 = (\text{ad}^2 \hat{X}^0; (\text{ad}^4 \hat{X}^1, \hat{X}^0)),$$
$$\hat{W}_2 = (\text{ad}^2 (\text{ad}^2 \hat{X}^1, \hat{X}^0); \hat{X}^0),$$
$$\hat{W}_3 = (\text{ad}^2 [\hat{X}^0, \hat{X}^1]; (\text{ad}^2 \hat{X}^1, \hat{X}^0)),$$
$$\hat{W}_4 = -[\hat{X}^0, [\hat{X}^0, \hat{X}^1]; (\text{ad}^3 \hat{X}^1, \hat{X}^0)]$$
$$\hat{W}_5 = [[\text{ad}^2 \hat{X}^0, \hat{X}^1]; (\text{ad}^3 \hat{X}^1, \hat{X}^0)].$$

Consider the following control systems

<table>
<thead>
<tr>
<th>(I)</th>
<th>(II)</th>
<th>(III)</th>
<th>(IV)</th>
<th>(V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\dot{x}_1 = u$</td>
<td>$\dot{x}_1 = u$</td>
<td>$\dot{x}_1 = u$</td>
<td>$\dot{x}_1 = u$</td>
<td>$\dot{x}_1 = u$</td>
</tr>
<tr>
<td>$\dot{x}_2 = \frac{1}{4!} x_1^4$</td>
<td>$\dot{x}_2 = \frac{1}{2} x_1^2$</td>
<td>$\dot{x}_2 = x_1$</td>
<td>$\dot{x}_2 = x_1$</td>
<td>$\dot{x}_2 = x_1$</td>
</tr>
<tr>
<td>$\dot{x}_3 = x_2$</td>
<td>$\dot{x}_3 = \frac{1}{6} x_3 x_2$</td>
<td>$\dot{x}_3 = x_2$</td>
<td>$\dot{x}_3 = x_2$</td>
<td>$\dot{x}_3 = x_2$</td>
</tr>
<tr>
<td>$\dot{x}_4 = x_3$</td>
<td>$\dot{x}_4 = \frac{1}{2} x_2^2$</td>
<td>$\dot{x}_4 = \frac{1}{4} x_1^2 x_2^2$</td>
<td>$\dot{x}_4 = x_3$</td>
<td>$\dot{x}_4 = \frac{1}{6} x_1^3 x_3$</td>
</tr>
</tbody>
</table>

One easily computes that $W_i(0)$ equals $\frac{\partial}{\partial x_4}$ in the $i$th system and vanishes on the four other ones. From this, arguing similarly as in §2.3.3, or from computing the transformation matrix relative to a Hall basis as given in [3], one finds that $\{\hat{W}_1, \cdots, \hat{W}_5\}$ constitute a basis for $L^{(4,3)}(\hat{X}^1, \hat{X}^0)$.

In the first three systems $x_4(t,u) \geq 0$ for all $u$ admissible and all $t \geq 0$, and thus it makes sense to consider $\hat{W}_1, \hat{W}_2$ and $\hat{W}_3$ bad brackets.

Without the $x_4$-component system (IV) has been discussed in [25] and
shown to be STLC, a fact which later could be obtained directly from Theorem 0.3. In [22] it is shown that then (IV) including the last component is STLC, too.

Hence one shall consider the associated bracket \( \hat{\mathcal{W}}_4 \) of type (4,3) good.

Finally in our example 1 we essentially investigated the system (V) and showed it is STLC, hence \( \hat{\mathcal{W}}_3 \) shall be considered good, also.

However to assign to each \( \hat{\mathcal{W}}_i \) from above a coefficient \( b_i(t,u) \) as in §2.3.3 requires to consider all of \( A^{(4,3)}(\hat{\mathcal{X}}^1;\hat{\mathcal{X}}^0) \) which is of dimension 20 and find a suitable basis for it.

Here we will not go further, instead only ascertain that there are in general both bad and good brackets of type (even,odd) to be expected, the good ones not necessarily of the form \([X^0,X^\sigma]\).

2.5.2. Bad brackets of different types balancing each other.

Suppose in a control system of form (1) on \( \mathbb{R}^\sigma \), for some \( i \leq n \), there are two distinct brackets \( X^\sigma(0) = \lambda^\sigma \frac{\partial}{\partial x_i}, \quad X^\sigma(0) = \lambda^\sigma \frac{\partial}{\partial x_i} \) with \( \lambda^\sigma, \lambda^\sigma \in \mathbb{R} \setminus \{0\} \), which are linearly independent from all other brackets (modulo the Jacobi identity) and \( X^\sigma, X^\sigma \) are bad brackets insofar as they always occur with definite coefficients. The idea of balancing is that there may be different families of control variations \( s \rightarrow u^{(1)}_a \) and \( s \rightarrow u^{(2)}_a \) inducing different orderings (say w.r.t. different \( \theta \in [0,1] \) on the Lie monomials in \( L(X^1,X^0) \), such that w.r.t. one ordering \( X^\sigma \) is of lower weight than \( X^\sigma \) and gives \( + \frac{\partial}{\partial x_i} \) as a tangent vector, and w.r.t. the other ordering \( X^\sigma \) is of lower weight than \( X^\sigma \) and gives \( - \frac{\partial}{\partial x_i} \) as a tangent vector.
We discuss two examples of bad brackets of different types balancing each other. For these two systems it suffices to consider control variations corresponding to some $\theta \in [0,1]$ as described in §2.2.

Consider the following 2 systems on $\mathbb{R}^4$:

\begin{align*}
\begin{cases}
\dot{x}_1 = u & x(0) = 0 \\
\dot{x}_2 = x_1 & |u(\cdot)| \leq \epsilon_0 \\
\dot{x}_3 = x_2 & \lambda > 0 \\
\dot{x}_4 = x_2^2 - \lambda x_1^4 ,
\end{cases}
\end{align*} \tag{21}

and

\begin{align*}
\begin{cases}
\dot{x}_1 = u & x(0) = 0 \\
\dot{x}_2 = x_1 & |u(\cdot)| \leq \epsilon_0 \\
\dot{x}_3 = x_2 & \lambda > 0 \\
\dot{x}_4 = x_2^2 - \lambda x_1^4
\end{cases}
\end{align*} \tag{22}

(\text{where of course in (21) the } x_3 \text{ component is insignificant and is introduced only for notational ease when comparing (21) and (22).}) System (21) has in certain aspects been discussed in [28].

The at zero nonvanishing brackets are: $X^1(0) = \frac{\partial}{\partial x_1}$,

$[X^0;X^1](0) = \frac{\partial}{\partial x_2}$, \quad $(\text{ad}^2 X^0, X^1)(0) = \frac{\partial}{\partial x_3}$ \quad and \quad $(\text{ad}^4 X^1; X^0)(0) = -24\lambda \frac{\partial}{\partial x_4}$ \quad in both systems, $(\text{ad}^2 [X^0; X^1]; X^0)(0) = 2 \frac{\partial}{\partial x_4}$ \quad in (21) and

$(\text{ad}^2 (\text{ad}^2 X^0; X^1); X^0)(0) = 2 \frac{\partial}{\partial x_4}$ \quad in (22).

From the linear theory it is clear that $\pm \frac{\partial}{\partial x_i}$ for $i = 1,2,3$ can be generated as tangent vectors to $G_\epsilon(t)$ at zero for any $\epsilon,t > 0$. 
We will next show explicitly when and how one can generate \( \frac{\partial}{\partial x_4} \) as tangent vectors. It is more convenient to do the calculations for

\[
\begin{aligned}
\dot{x}_1 &= u \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= x_2 \\
\dot{x}_4 &= x_2^2 \\
\dot{x}_5 &= x_3^2 \\
\dot{x}_6 &= x_1^4
\end{aligned}
\]

which has the advantage of having a homogeneous right side w.r.t. the dilation

\[
\Delta_{\epsilon, \delta}(x) = (\epsilon \delta x_1, \epsilon \delta^2 x_2, \epsilon \delta^3 x_3, \epsilon^2 \delta^3 x_4, \epsilon^2 \delta^3 x_5, \epsilon^4 \delta^3 x_6)
\]

and thus the scaling lemma (lemma 2) applies and we can work with \( \epsilon_0 = 1, \) some \( T > 0, \) and later scale down.

First choose an arbitrary admissible control \( u = u_{1,1}: [0, T] \to [-1, 1], \) some \( 0 < T < \infty \) s.t. \( u \neq 0 \) and \( x_i(T, u_{1,1}) = 0 \) for \( i = 1, 2, 3. \) For example the controls pictured in figures 5 and 6 satisfy these requirements.

![Figure 5. A control variation.](image-url)
Let $C_i = x_i(T, u_{1,1})$ for $i = 4, 5, 6$; then

$$C_4 = \int_0^T x_2^2(x, u_{1,1}) \, ds > 0,$$

$$C_5 = \int_0^T x_3^2(x, u_{1,1}) \, ds > 0,$$

$$C_6 = \int_0^T x_4^2(s, u_{1,1}) \, ds > 0.$$

Define $u_{t, \delta} : [0, \delta T] \rightarrow [-\epsilon, \epsilon]$ as usual by $u_{t, \delta}(\delta t) = \epsilon u_{1,1}(t)$ ($u_{t,0} \equiv 0$). Using the scaling lemma (lemma 2) $x(\delta t, u_{t, \delta}) = \Delta_{t, \delta}(x(t, u_{1,1}))$ for system (23) and thus in system (21)

$$x(\delta t, u_{t, \delta}) = (0,0,0, \epsilon^2 \delta^2 (C_4 - \epsilon^2 \lambda C_6))$$

and in system (22)

$$x(\delta t, u_{t, \delta}) = (0,0,0, \epsilon^2 \delta^2 (C_5 - (\epsilon / \delta)^2 \lambda C_6)).$$

Now given $\epsilon_0, t_0 > 0$, we will show when we can get $\pm \frac{\partial}{\partial x_4}$ as tangent vectors in the systems (21) and (22) and how to choose $\epsilon$ and $\delta$ to do so.

Let $\eta_1 = \left( \frac{C_4}{\lambda C_6} \right)_{\epsilon_0}$ and $\eta_2 = \left( \frac{C_5}{\lambda C_6} \right)_{\epsilon_0}$. In system (22) $s \rightarrow u_{t_0, 2s}$ generates $\frac{\partial}{\partial x_4}$, $s \rightarrow u_{t_0, 2s}$ generates $\frac{\partial}{\partial x_4}$ as a tangent vector to $G_{t_0}(t_0)$.
at zero, where \( s \in [0, s_0] \) with \( s_0 = \min \left\{ \frac{\epsilon_0}{2\eta_2}, \frac{t_0}{2} \right\} \), and thus system (22) is STLC.

(To be precise one has to use \( \bar{u}_{\epsilon, \delta} : [0, t_0] \to [-\epsilon_0, \epsilon_0] \) defined by
\[
\bar{u}_{\epsilon, \delta}(t) = \begin{cases} 
0 & \text{if } 0 \leq t \leq t_0 - \delta T \\
\epsilon u_{1,1} \left( \frac{t - (t_0 - \delta T)}{\delta} \right) & \text{if } (t_0 - \delta T) < t \leq t_0 .
\end{cases}
\]

Since \( \epsilon \sim s \) and \( \delta \sim s \) \( \sim \) "proportional to") these families of control variations correspond to \( \theta = \frac{1}{2} \).

In system (21) the situation is more complicated. Fixing \( \delta = \frac{t_0}{T} \), the one-parameter family of control variations \( s \mapsto u_{s, \delta} \) with \( s \in [0; \epsilon_0] \) generates
\[ + \frac{\partial}{\partial x_4} \]
as a tangent vector to \( G_{\epsilon_0}(t_0) \) at zero. (Note \( x_4(t_0, u_{s, \delta}) > 0 \) for \( s \in (0, \eta_1) \).) But in general it is not possible to generate \( - \frac{\partial}{\partial x_4} \) as a tangent vector, since in lemma 8 we showed that (in the notation of (21)) \( x_4(t, u) \geq 0 \) when \( x_1(t, u) = 0 \) and \( \epsilon^2 \lambda \leq \frac{2}{3} \) and thus (21) is not STLC in our sense for any \( \lambda \in \mathbb{R} \). However for every \( \lambda > 0 \) \( 0 \in G_{t_0}(t_0) \) for all \( t_0 > 0 \) if \( \epsilon > 0 \) is sufficiently large(!). (Using similar control variations as in §2.5.3 and in §2.5.4 one can show \( 0 \in \text{int } G_{\epsilon_0}(t_0) \) if \( \epsilon^2 \lambda > \frac{2}{3} \) and \( t_0 > 0 \).

The difference of the systems (21) and (22) becomes very clear when comparing (24) to (25). In (25) one may fix the ratio \( \frac{\epsilon}{\delta} \) to make \( x_4 \) positive, zero, or negative, and then let \( \epsilon \) and \( \delta \) go to zero at the same rate without that \( x_4 \) will change the sign. In (24) it suffices to give a sufficiently small
control bound \( \epsilon_0 \), thus \( 0 \leq \epsilon \leq \epsilon_0 \) will result in \( x_4 \) being nonnegative. The only way to get \( x_4 \) negative for smaller and smaller \( \epsilon \) is to exchange the basic control \( u_{1,1} \), and thus make the ratio \( C_4/C_6 \) smaller, also. But as shown in lemma 8, this ratio is bounded below by \( 2/3 \).

Assuming that these examples exhibit typical behavior one may conclude that balancing bad brackets of different types to give STLC (using control variations corresponding to some \( \theta \in [0,1] \)) is possible when the brackets have the same weight for some \( \theta \in [0,1] \), and impossible if they have the same weight for \( \theta = 1 \), in the latter case \( 0 \in \text{int} \mathcal{G}_e(t) \) only if \( \epsilon \) is sufficiently large.

2.5.3. Bad brackets of the same type balancing each other. We consider a control system, where two bad brackets of the same type balance each other, and show explicitly how to control it. This example already gives a hint how a more general class of control variations might be used. We show how powerful these new control variations are by working another example in §2.5.4.

\[
\begin{aligned}
\dot{x}_1 &= u \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= x_1^2 \\
\dot{x}_4 &= x_3^2 - \lambda x_2 x_4^4.
\end{aligned}
\]  

(26)

In this case the significant brackets are \( X^1 \equiv \frac{\partial}{\partial x_1}, \) \( [X^0, X^1](0) = \frac{\partial}{\partial x_2}, \)

\( -\langle \text{ad}^3 X^1; X^0 \rangle(0) = 6 \frac{\partial}{\partial x_3}, \) which are of types \((1,0),(1,1)\) and \((3,1)\), respectively,

\( \langle \text{ad}^2(\text{ad}^3 X^1; X^0); X^0 \rangle(0) = 36 \frac{\partial}{\partial x_4} \) and \( \langle \text{ad}^2[X^0, X^1];(\text{ad}^4 X^1; X^0) \rangle = -48\lambda \frac{\partial}{\partial x_4}, \)

which are both of type \((6,3)\). W.r.t. the two-parameter family of dilations
\[ \Delta_{\epsilon, \delta}(x) = (\epsilon^6 x_1, \epsilon^6 x_2, \epsilon^3 \delta^4 x_3, \epsilon^6 \delta^9 x_4) \]

\( X^0 \in \mathbb{R}^{(0,-1)}, X^1 \in \mathbb{R}^{(-1,-1)} \) and thus the scaling lemma (lemma 2) applies. But unlike in systems (21) and (22) changing \( u_{1,1} \) into \( u_{\epsilon, \delta} \) does not distinguish between the two squares on the right side of \( \dot{x}_4 \).

We will show that (26) is STLC for all \( \lambda \in \mathbb{R}^+ \). It suffices to show that the cone of tangent vectors to \( G_{\epsilon_0}(t_0) \) is all of \( \mathbb{R}^4 \) for \( \epsilon_0 = 1 \) and some \( t_0 > 0 \), by virtue of lemma 2.

By standard theory \( \pm \frac{\partial}{\partial x_i} \) for \( i = 1, 2, 3 \) are tangent vectors.

To show that \( + \frac{\partial}{\partial x_4} \) is a tangent vector to \( G_1(4) \) at zero consider the following one-parameter family of control variations \( s \rightarrow u^\xi_s: [0,4] \rightarrow [-1,1], s \in [0,1] \) defined by

\[
 u^\xi_s(t) = \begin{cases} 
 s^{1/6} & \text{if } t \in [0,\xi) \cup [4-\xi, 4] \\
 0 & \text{if } t \in [\xi, 4-2\xi) \\
 -s^{1/6} & \text{if } t \in [\xi, 2\xi) \cup [4-2\xi, 4\xi) 
\end{cases}
\]

Figure 7. A control variation

Then for every choice of \( \xi \in (0; 1] \) and \( i = 1, 2, 3 \) \( x_i(4, u^\xi_4) = 0 \) and \( x_4(4, u^\xi_4) = sx_4(4, u^\xi_4) \). We will show that for every \( \lambda \in \mathbb{R} \) choosing \( \xi > 0 \) sufficiently small leads to \( x_4(4, u^\xi_4) > 0 \) and thus gives \( + \frac{\partial}{\partial x_4} \) as a tangent
vector. Observe that for $t \in (2\xi, 4-2\xi)$ we have $x_1(t, u^\xi_2) = 0$, $x_3(t, u^\xi_2) > 0$ and thus $x_4(t, u^\xi_2) > 0$.

Using that for $t \in [0, 2\xi] \cup [4-2\xi, 4]$ the parameter $\xi$ plays the same role as $\delta$ in lemma 2, one finds $x_4(2\xi, u^\xi_1) = \xi^3 x_4(2, u^\xi_1)$ and $x_3(2\xi, u^\xi_1) = \xi^4 x_3(2, u^\xi_1)$.

Letting $C_3 = x_3(2, u^\xi_1)$, $C_4 = x_4(2, u^\xi_1)$, we find $x_4(4, u^\xi_1) = 2\xi^3 C_4 + (4-4\xi) \xi^8 C_3^2 = 4C_3^2 \xi^8 + (2C_4 - 4C_3^2) \xi^9$, and since $C_3 \neq 0$, $x_4(4, u^\xi_1) > 0$ for $\xi > 0$ sufficiently small.

For $\lambda \gg 1$, $x_4(4, u^\xi_1) < 0$ and thus fixing $\xi = 1$ gives $-\frac{\partial}{\partial x_4}$ a tangent vector also.

The interesting case, however, is $0 < \lambda \ll 1$. (The case $\lambda$ of the order of one can easily be handled by combining controls given above with those given below.)

Thus we assume $0 < \lambda \ll 1$ and find a one-parameter family of controls $s \to u_s$: $[0, t_0] \to [-1, 1]$ for some $t_0 > 0$, such that $x_i(t_0, u_s) = 0$, $i = 1, 2, 3$, all $s$ and $x_4(t_0, u_s) < 0$ for all $s > 0$.

Let $T = 4 + 2\sqrt{2}$ and define $u^1 : [0, T] \to [-1, 1]$ by

$$u^1(t) = \begin{cases} 1 & \text{if } t \in (0, 1) \cup (2 + 2\sqrt{2}, 3 + 2\sqrt{2}) \\ -1 & \text{if } t \in (1, 2 + 2\sqrt{2}) \cup (3 + 2\sqrt{2}, T) \end{cases}$$
Figure 8. A control variation.

Figure 9. Projections of $x(t,u)$ on the $(x_1,x_2)$- and the $(x_1,x_3)$-plane.

For two control functions $v_i: [0,t_i] \to [-1,1], i = 1,2$, define their concatenation $v_1 \ast v_2: [0,t_1+t_2] \to [-1,1]$ by

$$v_1 \ast v_2(t) = \begin{cases} v_1(t) & \text{if } t \in [0,t_1) \\ v_2(t-t_1) & \text{if } t \in [t_1,t_1+t_2]. \end{cases}$$

Define $u^{-1}(t) = u(T-t)$, the time reversed control, and $u^{k+1} = u^1 \ast u^k, u^{-1-k} = u^{-1} \ast u^{-k}$, and finally $u_k = u^k \ast u^{-k}$, for all $k > 0$.

From $x_i(T,u^1) = x_i(T,u^{-1}) = 0$ for $i = 1,3$, it follows that $x_i(2kT,u_k) = 0$ for $i = 1,3$. Similarly from $x_2(T,u^1) = -x_2(T,u^{-1}) = 2 - \sqrt{2}$, $x_2((2k-j)T,u_k) = j(2 - \sqrt{2}), j = 0,1, \ldots k$. Let $C_i = \int_0^T x_i^2(s,u^1)ds$
and compute

$$\int_0^{2kT} x_1^4(s, u_k) x_2^2(s, u_k) ds = 2 \sum_{j=1}^{k-1} \int_0^T x_1(s, u)(j(2 - \sqrt{2}) + x_2(s, u^1))^2 ds
$$

$$= 2 \sum_{j=0}^{k-1} \left( A_2 j^2 + A_1 j + A_0 \right)$$

(with some constants $A_0, A_1, A_2$ and $A_2 > 0$),

$$= C_3 k^3 + o(k^3) \text{ with } C_3 > 0,$$

and $o(k^3)$ stands for terms such that $\lim_{k \to \infty} \frac{o(k^3)}{k^3} = 0$. Together these give

$$x_4(2kT, u_k) = 2kC_1 - \lambda(k^3 - C_3 + o(k^3)) < 0$$

for $k$ sufficiently large.

Detailed analysis gives $C_1 \approx 1 \approx C_3$ and therefore $k \approx 8\lambda^{-\frac{1}{2}}$ gives a rough estimate for the number of switches needed with this control variation to achieve $x_4(t_0, u) < 0$ while $x_i(t_0, u) = 0$ for $i = 1, 2, 3$.

It seems as if (26) becomes almost uncontrollable when $\lambda \to 0^+$, which reflects the fast growing number of switches needed to go in the $(-\frac{\partial}{\partial x_4})$-direction.

(Remark: We here are only concerned about controllability of (26) for all $\lambda > 0$ and make no claims about optimality. Although it does not seem very likely, there may be control variations requiring much fewer switches which allow one to go into the $(-\frac{\partial}{\partial x_4})$-direction for $\lambda > 0$ small.)

2.5.4. A new class of control variations. In this paragraph we introduce a new class of control variations and illustrate their usefulness by working an
example. In some sense these control variations resemble the ones used in §2.5.3, as again there is a parameter \( k \), related to the number of switches, which rapidly grows to infinity. But in 5.3 \( k \) only depended on the constant \( \lambda \) on the right hand side of the control system (26), and was the same for all \( \epsilon_0, t_0 > 0 \). Here instead \( k \) will depend on \( \epsilon_0 \) and \( t_0 \) and grow rapidly as they become small, i.e. the smaller \( \epsilon_0 \) and \( t_0 \) are, the more switches are needed to generate a certain tangent vector to \( G_{\epsilon_0}(t_0) \).

Before summarizing in what aspects these are new control variations, we illustrate their application in considering the following system on \( \mathbb{R}^4 \):

\[
\begin{align*}
\dot{x}_1 &= u \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= x_1^3 \\
\dot{x}_4 &= x_3^2 - x_2^2
\end{align*}
\quad (27)
\]

where the significant brackets are \( X^1 \equiv \frac{\partial}{\partial x_1}, \quad [X^0; X^1](0) = \frac{\partial}{\partial x_2}, \)

\[ -(\text{ad}^3 X^1; X^0)(0) = 6 \frac{\partial}{\partial x_3}, \quad W_1(0) = (\text{ad}^2(\text{ad}^3 X^1; X^0); X^0)(0) = 72 \frac{\partial}{\partial x_4} \]

and \( W_2(0) = (\text{ad}^7 [X^0; X^1]; X^0)(0) = 7! \frac{\partial}{\partial x_4} \), the last two brackets being of types (6,3) and (7,8), respectively.

Clearly, \( W_2 \neq (\text{ad}^7 X^0; (\text{ad}^7 X^1; X^0)) \) and thus by all usual assignments of weights, \( W_2 \) is of higher weight than the bad bracket \( W_1 \) (bad, since clearly coming from a square). Thus with the usual control variations \( W_1 \) will always dominate \( W_2 \), resulting in \( x_4 \geq 0 \). The tangent vectors \( \pm \frac{\partial}{\partial x_i} \) for \( i = 1,2,3 \) and \( + \frac{\partial}{\partial x_4} \) can easily be generated by using standard control variations, so...
that we only have to consider \(-\frac{\partial}{\partial x_4}\).

We will show that (27) is STLC, by showing that \(-\frac{\partial}{\partial x_4}\) is a tangent vector to \(G_\epsilon(t_0)\), any \(\epsilon,t_0 > 0\). This is done by explicitly constructing controls \(u: [0; t_0] \rightarrow [-\epsilon_0, \epsilon_0]\), such that \(x_i(t_0, u) = 0\) for \(i = 1, 2, 3\) and

\[
\int_0^{t_0} x_2^7(s, u) ds > \int_0^{t_0} x_3^2(s, u) ds, \tag{28}
\]
or equivalently \(x_4(t_0, u) < 0\).

To achieve that (28) holds, one would like to keep \(|x_3(\cdot, u)|\) small, while letting \(x_2(\cdot, u)\) be relatively large. However with the usual control variations the integral on the left side of (28) is of order \(\epsilon^7 t_0^{15}\), whereas the one on the right side is of order \(\epsilon^8 t_0^9\). The idea is to find controls \(u\) so that the projection of the path \(s \mapsto x(s, u)\) onto the \((x_2, x_3)\)-plane starts and ends at the origin and stays very close to the positive \(x_2\)-axis. Of course \(x_3(\cdot, u) \equiv 0\) would immediately require \(x_2(\cdot, u) \equiv 0\), but by choosing the number of switches sufficiently large, depending on \(\epsilon_0\) and \(t_0\), one can achieve that the path \(x(\cdot, u)\) is as described above and that finally (28) holds.

For computational ease we will work with

\[
\begin{align*}
\dot{x}_1 &= u \\
\dot{x}_2 &= x_1 \\
\dot{x}_3 &= x_1^3 \\
\dot{x}_4 &= x_3^2 \\
\dot{x}_4 &= x_2^7
\end{align*}
\tag{29}
\]

which is homogeneous w.r.t. the dilation.
\[ \Delta_{\epsilon, \delta}(x) = (\epsilon \delta x_1, \epsilon^2 \delta^2 x_2, \epsilon^3 \delta^3 x_3, \epsilon^8 \delta^{15} x_{41}, \epsilon^7 \delta^{15} x_{42}) \]

so that lemma 2 applies. Exactly as in §2.5.3 for system (2) we define

\[ T = 4 + 2\sqrt{2}, \text{ and } u^1 : [0,T] \rightarrow [-1,1] \text{ by} \]

\[ u^1(t) = \begin{cases} +1 & \text{if } t \in [0,1) \cup [2+\sqrt{2},3+\sqrt{2}) \cup [3+2\sqrt{2},T], \\ -1 & \text{if } t \in [1,2+\sqrt{2}) \cup [3+2\sqrt{2},T], \end{cases} \]

(30)

\[ u^{-1}(t) = u(T-t) = -u(t), \ u^{k+1} = u^1 \ast u^k, \ u^{-k-1} = u^{-1} \ast u^{-k}, \ k = 1,2,3, \ldots, \text{ and} \]

\[ u = u^k \ast u^{-k} \text{ (we called this } u_k \text{ in §2.5.3).} \]

One readily checks the following relations:

\[ x_i(x,u^{-1}) = -x_i(s,u^1), \ i = 1,2,3,4,2, \ s \in [0,T]; \]

\[ x_{41}(s,u^{-1}) = x_{41}(s,u^1), \ s \in [0,T]; \]

\[ x_i(T,u^1) = 0, \ i = 1,3; \]

\[ x_i(2kT,u) = 0, \ i = 1,2,3; \]

\[ x_{41}(2kT,u) = 2kC_{41}, \text{ where } C_{41} = x_{41}(T,u^1); \]

\[ x_2(T-s,u^{-1}) = x_2(s,u^1) \text{ for } s \in [0,T] \text{ and thus} \]

\[ x_2(2kT-s,u) = x_2(s,u) \text{ for } s \in [0,kT]. \]

Given \( \epsilon_0, t_0 > 0 \), choose \( 0 < \epsilon \leq \epsilon_0 \) and \( 0 < t_1 \leq t_0 \), let \( \delta = \delta(k) = \frac{t_1}{2kT} \)

and define \( u_{\epsilon,\delta} : [0,t_0] \rightarrow [-\epsilon,\epsilon_0] \) by

\[ u_{\epsilon,\delta}(\delta t) = \begin{cases} 0 & \text{if } t \in [0,t_0-t_1] \\ \epsilon u(t-t_0+t_1) & \text{if } t \in [t_0-t_1,t_0] \end{cases} \]

One easily computes

\[ x_{41}(t_0,u_{\epsilon,\delta}) = \epsilon^8 \delta^{15} \cdot 2kC_{41} = k^{-8} \epsilon^8 t_1^9 \frac{2C_{41}}{(2T)^9} \]

and
\[
x_{42}(t_0, u_{t, \delta}) = \int_0^{2kT} x_2^T(s, u_{t, \delta}) ds
\]
\[
= 2 \sum_{j=0}^{k-1} \int_0^T \epsilon^7 \delta^{14} (j(2 - \sqrt{2}) + x_2(s, u^1))^7 ds
\]
\[
= 2 \epsilon^7 \delta^{14} \sum_{j=0}^{k-1} (Tj^7 (2 - \sqrt{2})^7 + o(j^7))
\]
\[
= 2(8!)^{-1} \epsilon^7 \delta^{14} (Tk^8 (2 - \sqrt{2})^7(8!)^{-1} + o(k^8))
\]
\[
= \epsilon^7 t_1^{14} \left( \frac{(2 - \sqrt{2})^7}{(2T)^{13} 8!} + o(k^{-5}) \right)
\]

where \( o(k^\nu) \) stands for terms such that \( \lim_{k \to \infty} \frac{o(k^\nu)}{k^\nu} = 0 \).

Letting \( C_{42} = \frac{2(2 - \sqrt{2})^7}{8!(2T)^{13}} > 0 \),

\[
x_4(t_0, u_{t, \delta}) = \frac{\epsilon^7 t_1^{14}}{k^8} (C_{41} - \epsilon t_1^9 k^2 C_{42} + \epsilon t_1^9 o(k^2)), \quad \text{and by choosing } k \text{ sufficiently large, roughly of the size } k \approx \epsilon^{-4} t_1^{-\frac{5}{4}} \geq \epsilon_0^{-4} t_0^{-\frac{5}{4}} \text{ one obtains } x_4(t_0, u_{t, \delta}) < 0.
\]

Thus for every \( \epsilon_0, t_0 > 0 \) we obtain points in the intersection of \( G_0(t_0) \) and the negative \( x_4 \)-axis, and therefore \( -\frac{\partial}{\partial x_4} \) is a tangent vector to \( G_0(t_0) \) for every \( \epsilon_0, t_0 > 0 \), so that (27) is STLC.

Observe that even for fixed \( \epsilon_0, t_0 > 0 \), one cannot fix \( k \) and let \( \epsilon \) and \( t_1 \) go to zero, to obtain a sequence of points on the negative \( x_4 \)-axis approaching zero — rather \( k \) really depends on \( \epsilon \) and \( t_1 \), i.e. one may as well start with \( \epsilon = \epsilon_0 \) and \( t_1 = t_0 \).

Observe furthermore that for \( \epsilon_0, t_0 \) small, the number of switches is of order \( (\epsilon_0 t_0^5)^{-\frac{5}{4}} \) and that \( x_4(t_0, u_{t_0, \delta}) \) is of order \( \epsilon_0^{10} t_0^{19} \), and thus for practical
purposes controls like this do not seem to be very useful. However from the mathematical side controls like this may be very important to formulate a general theorem on necessary and sufficient conditions for STLC.

To compare this control variation to the usual ones, recall that one usually defines perturbation data $\Gamma$ as in (30) (compare to §2.2), fixes a functional dependence $\epsilon = \epsilon(s)$ and $\delta = \delta(s)$ (usually $\epsilon = s^{1-\theta}$, $\delta = s^\theta$, $\theta \in [0,1]$) and then only considers $s \rightarrow x(\cdot, u_{\epsilon,\delta})$.

Here instead we use the control $u$, as, for example, defined in (30), as a control element ("Baustein"), not only scale $u_{1,1}$ to $u_{\epsilon,\delta}$, but also concatenate an increasing number, depending on $\epsilon$ and $\delta$ (and thus on $\epsilon_0$ and $t_0$), of these control elements (which itself requires some scaling, so that this concatenation still is defined on $[0,t_0]$). This allows — within a certain range, still to be determined — to override the usual weight assignment $(k+l\theta)$ to terms corresponding to brackets $X^r$ or $(\text{ad}^r X^0; X^r)$ where $X^r$ is of type $(k,l)$.

There are many control systems exhibiting a similar behavior as (27) and which by using controls as above can be shown to be STLC. However system (27) seems to be the "lowest order" interesting system of this form. (Replacing $W_1$ of type $(6,3)$ by a bracket of type $(2m,1)$ will not yield similar behavior, since by theorems 0.2 and 0.4 one knows precisely which brackets may dominate it; namely only, but all, those of type $(k,l)$ with $k < 2m$. When considering theorem 1, brackets of type $(2,2m+1)$ do not seem to be nice candidates either. Of the three bad brackets of type $(4,3)$ two will lead to a loss of STLC already in some other component. The remaining one $(\text{ad}^2 [X^0; X^1]; (\text{ad}^2 X^1; X^0))$ is by rough estimates not an interesting candidate
either. Thus \( W_1 \) of type (6,3), as above, is the lowest order interesting bracket. \( W_2 \) shall have both more factors \( X^1 \) and \( X^0 \) than \( W_1 \), and not be of the form (\( \text{ad}^\nu X^0; X^r \)) with \( \nu > 0 \) and \( X^r \) of lower weight than \( W_1 \) relative to \( \theta = 1 \). Thus the choice \( W_2 = (\text{ad}^7 [X^0; X^1]; X^0) \) is the most simple one, too.

We leave it as an open question, which brackets may dominate which brackets, in particular when using control variations as discussed in this paragraph.

**2.6 Conclusion.**

With theorem 1 a necessary condition for STLC concerning brackets of type (2,3) has been proven, which points the way to similar necessary conditions involving brackets of type (2, 2j+1).

In the example in §2.5.1 it has been shown that not all brackets of type (even, odd) are bad, probably many of them are even good. But before one can make reliable statements about which brackets are good or bad, one has to find a basis for \( \text{L}(X^1, X^0) \) which suits practical purposes. We conjecture that the only bad brackets are all of the form

\[
(\text{ad}^\nu X^0; (\text{ad}^2 X^r; (\cdots (\text{ad}^2 X^r; X^0)) \cdots )
\]

or where \( X^r_1, \cdots X^r_t \) are Lie monomials with nondecreasing weight relative to some suitable weight assignment still to be found, and satisfy some condition like \( X^r_i(0) \neq 0 \) (or products with all factors of above form).

In §2.5.2, §2.5.3 and §2.5.4 we showed that there are more ways of neutralizing bad brackets (so that they are not obstructions to STLC) than the method given in theorem 0.3. Finally the control variations introduced in §2.5.4
with the number of switches rapidly increasing as the disposable (available) time $t_0$ goes to zero allows us to control many systems which formerly seemed uncontrollable and might be a key element in a general theorem about STLC.

We found a few very little pieces in the controllability puzzle, but it is still a very, very long way to a general theorem about when a map $L^1([0,t_0]; [-\epsilon, \epsilon]) \rightarrow \mathbb{R}^n$ given by (1) covers a full neighborhood of the origin.
References


