The Combinatorics of Nonlinear Controllability and Noncommuting Flows

Matthias Kawski*

Department of Mathematics, Arizona State University, Tempe, Arizona, USA

Lectures given at the
Summer School on Mathematical Control Theory
Trieste, 3-28 September 2001

LNS028005

---

*kawski@asu.edu

Abstract

These notes accompany four lectures, giving an introduction to new developments in, and tools for problems in nonlinear control. Roughly speaking, after the successful development, starting in the 1960s, of methods from linear algebra, complex analysis and functional analysis for solving linear control problems, the 1970s and 1980s saw the emergence of differential geometric tools that were to mimic that success for nonlinear systems. In the past 30 years this theory has matured, and now connects with many other branches of mathematics.

The focus of these notes is the role of algebraic combinatorics for both illuminating structures and providing computational tools for nonlinear systems. On the control side, we focus on problems connected with controllability, although the combinatorial tools obviously have just as much use for other control problems, including e.g. path-planning, realization theory, and observability.

The lectures are meant to be an introduction, sketching the road from the comparatively naive, bare-handed constructions used in the early years, to the elegant and powerful insights from recent years. One of the main targets is to develop an explicit, continuous analogue of the classical Campbell-Baker-Hausdorff formula, and of a related exponential product expansion. The purpose of such formulae is to separate the time-dependent and control-dependent parts of solution curves from the invariant underlying geometrical structure inherent in each control system.

The key theme is that effective tools (including effective notation) from algebraic combinatorics are essential, for both theoretical analysis and for practical computation (beyond some miniscule academic examples). On a practical level we want the reader to take home the message to never write out complicated iterated integrals, as it is both a waste of paper and time, as it obscures the underlying structure. On the theoretical level, the key object is the chronological algebra isomorphism from the free chronological algebra to an algebra of iterated integral functionals, denoted by Υ in our exposition.

Reiterating, these notes are meant to be an introduction. As such, they provide many examples and exercises, and they emphasize as much getting a hands-on experience and intuitive understanding of various structural terms, as they are meant to establish the need for, and appreciation of tools from algebraic combinatorics. We leave a formal treatment of the abstract structures and isomorphism to future lectures, and until then refer the reader to pertinent recent literature.

Contents

0 Organization and objectives

These notes contain the background information and the contents (in roughly the same order) of four 75 minute lectures given during the 2001 summer school on mathematical control. They shall provide an introduction to nonlinear controllability and the algebraic-combinatorial tools used to study it. An effort is made to keep the level elementary, assuming familiarity primarily with the theory of differential equations and knowledge from selected preceding lectures in this summer school that addressed geometric methods in control, and an introduction to nonlinear control systems. Consequently, in several places a comparatively *"pedestrian approach"* is taken which may not be the cleanest or most elegant formulation, as the latter may typically presume more advanced ways of thinking in differential geometry or algebraic combinatorics. However, in most such places comments point to places in the literature where more advanced approaches may be found.
Similarly, proofs are given or sketched where they are illuminating and of reasonable length when using tools at the level of this course. In other cases comments refer to the literature where detailed, or more efficient proofs may be found.

Several examples are provided, and revisited frequently, both in order to provide motivation, and to provide the hands-on experience that is so important for making sense of otherwise abstract recipes, and to provide the ground for further developments. In this sense, the exercises imbedded in the notes are an essential component and the reader is urged to get her/his hands *dirty* by working out the details.

Aside from providing an introductory survey of some aspects of modern differential geometric control theory, the overarching objective is to develop a sense of necessity, and an appreciation of the algebraic and combinatorial tools, which provide as much an elegant algebraization of the theory as they provide the essential means that allow one to carry out real calculations that without these tools would be practically almost impossible.

1 Nonlinear controllability

1.1 Introductory examples

The problem of *parallel parking a car* provides one of the most intuitive introductions to many aspects of nonlinear control, especially controllability, and it may be analyzed at many different levels. Here we introduce a simplified version of the problem, and use it to motivate questions which naturally beg for generalization. The example will be revisited in later sections as a model case on which to try out newly developed tools and algorithms.

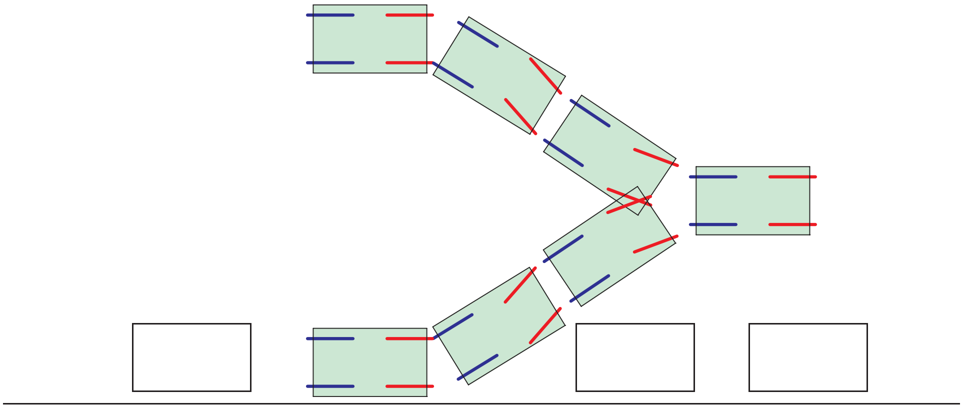


Figure 1. Parallel parking a car (exaggerated parallel displacement)

Example 1.1

Think about driving a real car, and the experience of parallel parking a car in an empty spot along the edge of the road. If the open gap is large, this is very easy – but it becomes more challenging when the length of the gap is just barely larger than the length of the car. For the sake of definiteness, suppose the initial position and orientation of the car as indicated in the diagram (with much exaggerated parallel displacement, and an exaggerated length of the gap), with steering wheels in the direction of the road.
Everyday experience says that, while it is impossible to directly move the car sideways, it is possible to do so indirectly via a series of careful maneuvers

that involve going back and fourth with suitably matching motions of the steering wheels.

One may consider different choices as possible controls. In this case let us use the forward acceleration of the rear wheels as one control, and the steering angle as a second control.

Exercise 1.1 *Develop different possible series of maneuvers that result in a car that is in the same location, with zero speed, but rotated by* $\frac{\pi}{2}$ *of by* π*. Describe the maneuvers verbally, and sketch the states as functions of time*

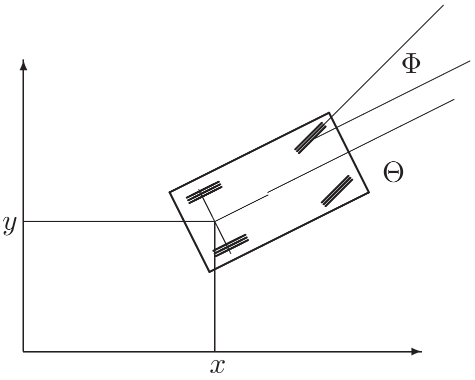


Figure 2. Defining the *states* of the system

To obtain a mathematical model consider the simpler (less controversial case as it does not require *differentials* to adjust for the different speeds of *inside* and *outside* wheels) of a bicycle! In particular, let $(x, y) \in \mathbb{R}^2$ denote the point of contact of the rear wheel with the plane (center of rear axle in the case of a car). Let $\theta \in S^1$ be the angle of the bicycle with the x_1-axis, and by $\phi \in S^1$ the angle of the front wheel(s) with the direction of the bicycle. An algebraic constraint captures that the distance between front and rear wheel is constant, equal to the length L. Thus the position of the front wheel (point of contact with plane) is $(x+L\cos\theta, y+L\sin\theta)$. The conditions that the wheels can slip neither forward nor sideways, each can only roll in the direction of the wheel is captured in

$$\left\{\begin{array}{lcl} 0 & = & \cos\theta\, dy - \sin\theta\, dx \\ 0 & = & \sin(\theta+\phi)\, d(x+L\cos\theta) - \cos(\theta+\phi)\, d(y+L\sin\theta) \end{array}\right. \tag{1}$$

Introducing the speed $v = \|\dot{x}^2 + \dot{x}^2\|$ of the rear wheel (or of the center of the rear axle), write $\dot{x} = v\cos\theta$ and $\dot{y} = v\sin\theta$.

Exercise 1.2 *Discuss what happens in this model when the forward speed of the rear wheel is zero and the angle of the steering wheel is* $\phi = \pi/2$. *Can the bicycle move?*
Develop an alternative front-wheel drive *model, i.e. with controlled speed* v *of front wheel. Continue working that model in parallel to the one discussed here in the notes.*

Using the first constraint, solve the second constraint for

$$d\theta = \frac{v\,dt}{L} \cdot \frac{\cos\theta \cdot \tan(\theta+\phi) - \sin\theta}{\cos\theta + \tan(\theta+\phi)\sin\theta} = \frac{v}{L} \cdot \tan\phi\,dt \qquad (2)$$

(The last step is immediate from basic trigonometric identities after multiplying numerator and denominator by $\cos(\theta+\phi)$.) Write the model as a system of controlled ordinary differential equations (for simplicity we choose units such that $L = 1$)

$$\left\{ \begin{array}{rcl} \dot{\phi} & = & u_1 \\ \dot{v} & = & u_2 \\ \dot{x} & = & v\cos\theta \\ \dot{\theta} & = & v\tan\phi \\ \dot{y} & = & v\sin\theta \end{array} \right. \qquad (3)$$

Exercise 1.3 *Using your practical driving experience, suggest specific control functions* u_1, u_2 *(e.g. piecewise constant or sinusoidal, with switching times as parameters to be determined) such that the corresponding solution* steers *the system from* $(\phi, v, x, \theta, y)(0) = (0,0,0,0,0)$ *to* $(\phi, v, x, \theta, y)(T) = (0,0,0,0,H)$ *for some* $T > 0$ *and* $H \neq 0$.
Sketch the graphs of the states as functions of time (compare figure 3).

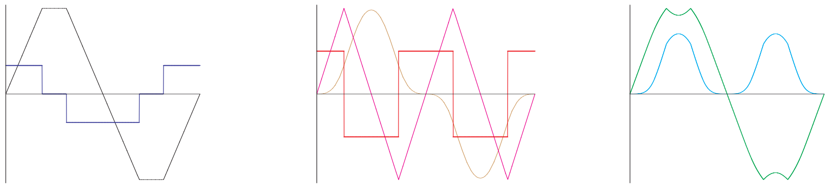


Figure 3. One possible, very symmetric, parallel parking maneuver

Exercise 1.4 *In figure 3, identify which curve represents which state or control.*

Another well-studied [3] introductory example is that of a *rolling penny* in he plane.

Example 1.2

Consider a disk of radius a and negligible thickness *standing on its edge* that may roll without slipping in the plane, and which may rotate about its vertical axis. Denoting by $(x_1, x_2) \in \mathbb{R}^2$ its point of contact with the plane, by $\theta \in S^1$ its angle with the x_1-axis, and by $\phi \in S^1$ its *rolling* angle from a fixed reference angle, the non-slip constraints may be written as:

$$\left\{ \begin{array}{rcl} \cos\theta \, dx_1 + \sin\theta \, dx_2 & = & a \, d\phi \\ \sin\theta \, dx_1 - \cos\theta \, dx_2 & = & 0 \end{array} \right. \tag{4}$$

Equivalently, considering the angular velocities as controls the system is written as

$$\left\{ \begin{array}{rcl} \dot{\phi} & = & u_1 \\ \dot{\theta} & = & u_2 \\ \dot{x}_1 & = & au_1 \cos\theta \\ \dot{x}_2 & = & au_1 \sin\Theta \end{array} \right. \tag{5}$$

Alternatively, considering the accelerations, or rather the torques as controls (suitably scaled), the system is described by

$$\left\{ \begin{array}{rcl} \dot{\omega_1} & = & u_1 \\ \dot{\omega_2} & = & u_2 \\ \dot{\phi} & = & \omega_1 \\ \dot{\theta} & = & \omega_2 \\ \dot{x}_1 & = & au_1 \cos\theta \\ \dot{x}_2 & = & au_1 \sin\theta \end{array} \right. \tag{6}$$

One of the more intriguing question is whether it is possible to roll, and turn the penny in such a way that at the end it is back at its original location with original orientation but rotated about a desired angle about its horizontal axis.

Moreover, one may ask if it always possible to achieve such a reorientation without moving far from the starting state. Alternatively, one may ask whether one can in any arbitrarily small time interval achieve at least a small reorientation.

Exercise 1.5 *Develop an intuitive strategy that results in such a reorientation. I.e. describe the maneuver in words, and sketch the general shapes of the states as functions of time.*

Exercise 1.6 *Develop an intuitive strategy that results in such a reorientation. I.e. describe the maneuver in words, and sketch the general shapes of the states as functions of time.*

Exercise 1.7 *Find an analytic solution using piecewise constant controls defined on an arbitrary short time-interval* $[0, T]$ *that rotates the penny by a given angle* $\varepsilon \in \mathbb{R}$.

Exercise 1.8 *Repeat the previous exercise using controls that are piecewise trigonometric functions of time, or that are trigonometric polynomials.*

With such mechanical examples as there is no question about the model and we concentrate on the analysis and geometry. But the methodology developed in sequel is just applicable to controlled dynamical systems that arise in electric and communication networks, in biological and bio-medical systems, in macro-economic and financial systems etc.

1.2 Controllability

For a given control $u(t)$, a control system $\dot{x} = f(x, u)$ with initial value $x(0)$ is simply a dynamical system, which is straightforward to analyze and *solve* using basic techniques from differential equations. What makes control so much more intellectually challenging is the inverse nature of most questions – e.g. given a target $x(T)$, *find* a control u that steers from $x(0)$ to $x(T)$. The first step, before one may start any construction or optimization, is to ask whether there exists any solution in the first place. This is the question about controllability.

Exercise 1.9 *Review the examples and exercises in the previous section, and relate the notion of controllability to the questions raised in that section.*

One may well say that the study of controllability is analogous, and just as fundamental as the questions of existence and uniqueness of solutions of differential equations. In further analogy, the study of controllability actually leads one to algorithmic constructions of more advanced problems such as path planning, much in the same way as proofs for existence and uniqueness of solutions of differential equations yield e.g. recipes for obtaining infinite series and numerical solutions.

Recall the case of linear systems $\dot{x} = Ax + Bu$ (with state and control vectors x and u and matrices A and B of appropriate sizes). Using variation of parameters one quickly obtains a formula for the solution curve

$$x(t) = x(0)e^{tA} + \int_0^t e^{(t-s)A} Bu(s)\, ds \tag{7}$$

It is readily apparent that the set of points that can be reached from $x(0) = 0$ (via piecewise constant, measurable controls or any similar sufficiently rich class) is always a subspace of the state space. Moreover, scaling of the control $u \mapsto cu$ immediately carries over to the solution curve $x(t, cu) = cx(t, u)$ (assuming $x(0) = 0$). Consequently, the size of the control is no major factor in the discussion of linear controllability, and neither is the time T that is available. The scaling invariance implies that most local properties and global properties are the same. Finally, the solution formula (7) also quickly yields (e.g. via Taylor expansions and the Cayley-Hamilton theorem) a simple algebraic criterion for linear controllability:

Theorem 1.1 (Kalman rank condition) *The linear system* $\dot{x} = Ax + Bu$ *with* $x \in \mathbb{R}^n$ *is controllable (for* any *reasonable technical definition of controllable) iff the block-matrix* $(B, AB, A^2B, \ldots A^(n-1)B)$ *has full rank.*

In the case of nonlinear systems almost *everything* is different: There are many, many equally reasonable notions of controllability which are not equivalent to each other. Local and global notions are generally very different. The class of admissible controls has to be very carefully stated – e.g. bounds on the control size can make all the difference. Assumptions about regularity properties of both the vector fields (e.g. smooth versus analytic) and the controls (e.g. measurable versus piecewise constant) are critically important. Similarly, a system may be controllable (in a reasonable) sense given sufficiently much time, but may be uncontrollable for small positive times.

In these lectures we shall concentrate on one of the best studied notions, and which is of significant importance for a variety of further theories (e.g. a sufficient condition for some notions of feedback stabilizability, and duality to optimality). Thus from now on, unless otherwise stated the following blanket assumptions shall generally apply: We consider affine, analytic systems that are of the form

$$\dot{x}(t) = f_0(x(t)) + \sum_{i=1}^{m} u_i(t)\, f_i(x(t)) \text{ usually initialized at } x(0) = 0 \tag{8}$$

where f_i are (real) analytic vector fields, and the controls u are assumed to be measurable with respect to time, and assumed to take values in a compact subset $U \subset \mathbb{R}^m$, often taken as $[-1,1]^m$. The vector field f_0 is called the drift vector field, while f_i for $i \geq 1$ are called the control (or controlled) vector fields. In the case that $f_0 \equiv 0$ (i.e. is absent) the system (8) is called *"without drift"*.

Much different techniques are needed when allowing more general dependence of the dynamics on the control $\dot{x} = f(x,u)$, compare the lectures by Jacubczyk in this series. One may also demand less regularity, e.g. only Lipschitz-continuity of the vector fields associated to fixed values of the controls. A mature theoretical framework for that case provided by *differential inclusions*, compare the lectures by Frankowska in this series.

Revisit the parking example of the first section and introduce standard, uniform notation by defining $x = (x_1, x_2, x_3, x_4, x_5) \overset{\text{def}}{=} (\phi, v, x, \theta, y)$, and write the system (3) in the form (8)

$$f_0(x) = \begin{pmatrix} 0 \\ 0 \\ x_2 \cos x_4 \\ x_2 \tan x_1 \\ x_2 \sin x_4 \end{pmatrix}, \quad f_1(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } f_2(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{9}$$

This is a system with drift – which corresponds to the *unforced* dynamics of the car, and with two controlled vector fields which correspond to forward acceleration/deceleration, and to changing the steering angle.

Exercise 1.10 *Revisit the second example, the rolling penny, from the first section. Again write the states as* $x = (x_1, x_2, \ldots)$ *and write the systems* (5) *and* (6) *in the form* (8) *(i.e. identify the controlled vector field(s), and the drift vector field.*

Explain in practical terms how the choice of controls as acceleration or as velocities effects the presence of a drift. (Note, these are models for the kinetic versus dynamic behaviours).

Definition 1.1

The reachable set $\mathcal{R}_p(T)$ *of system* (8) *at time* $T \geq 0$*, subject to the initial condition* $x(0) = p$ *is the set*

$$\mathcal{R}_p(T) = \{x(T,u) \colon x(0) = p \ \textit{ and } \ u \colon [0,T] \mapsto U \ \textit{measurable} \ \} \tag{10}$$

Definition 1.2 *The system* (8) *is accessible from* $x(0) = p$*, if the reachable sets* $\mathcal{R}_p(t)$ *have non-empty interior for all* $t > 0$.
The system (8) *is small-time locally controllable* (STLC) *about* $x(0) = p$*, if* $x(0) = p$ *is contained in the interior or the reachable sets* $\mathcal{R}_p(t)$ *for all* $t > 0$.
Consider

Accessibility and controllability (STLC) are generally different from each other.

$$\left\{ \begin{array}{lcl} \dot{x}_1 & = & u \\ \dot{x}_2 & = & x_1^k \end{array} \right. \qquad \begin{array}{rcl} x(0) & = & 0 \\ \|u(\cdot)| & \leq & 1 \end{array} \tag{11}$$

with measurable controls $u(\cdot)$ bounded by $\|u(\cdot)| \leq 1$ and $k \in \mathbf{Z}^+$ fixed. Using e.g. piecewise constant controls with a single switching one easily that the reachable sets $\mathcal{R}_0(t)$ have two dimensional interior for all $t > 0$ while $x(0) \notin \mathcal{R}_0(t)$ for all $t \geq 0$ if k is even.

Exercise 1.11 *Using methods from optimal control, one may show that the boundaries of the reachable sets at time* $T \geq 0$ *of the system* (11) *are contained in the set of endpoints of trajectories resulting from* bang-bang controls *with at most one switching, i.e. controls of the form* $u_{+-,t_1}(t) = 1$ *if* $0 \leq t \leq t_1$ *and* $u_{+-,t_1}(t) = -1$ *if* $t < t_1 \leq T$*, or* $u_{-+,t_1}(t) = 1$ *if* $0 \leq t \leq t_1$ *and* $u_{-+,t_1}(t) = -1$ *if* $t < t_1 \leq T$*. Calculate these curves of endpoints (as curves parameterized by* t_1*). Rewrite these as (unions of) graphs of functions* $x_2 = f(x_1)$*, and sketch the reachable sets.*

Exercise 1.12 *Continuing the previous exercise in the case of* k *an even integer, identify all pairs of switching times* t_1*,* t_2 *such that* $x(1; u_{+-,t_1}(t) = x(1; u_{-+,t_2})(t)$*.*

A few further remarks about controllability: Clearly, controllability is a geometric notion, independent of any choice of local coordinates. While for calculations it often is convenient to choose and fix a set of specific coordinates, it is desirable to obtain conditions for controllability that are geometric, too (compare the Kalman rank condition which involves the *geometric* property of the rank). In these notes we are concerned only with local properties. Consequently, we generally may assume that the underlying manifold is $\mathbb{R}^n$. In particular, when working with approximating vectors we shall conveniently identify the tangent spaces $T_p\mathbb{R}^n$ to $\mathbb{R}^n$with $\mathbb{R}^n$. Nonetheless, occasionally we may phrase our observations and results so as to emphasize that they really also apply to general manifolds.

In the linear setting there is a very distinctive duality between controllability and observability. In the nonlinear case this moves more to the background. However, STLC is *dual* to optimality: Controllability means that one can reach a neighborhood, whereas optimality means that a trajectory lies on the boundary of the *funnel* of all reachable sets. Consequently, necessary conditions for STLC translate immediately into necessary conditions for optimality, and vice versa.

Also the implication that controllable systems are feedback stabilizable carries over to the nonlinear framework when using the notion of STLC and *time-periodic static feedback* as shown by Coron [7]. Note, that a simple reversal of time, i.e. replacing each f_i by $(-f_i)$ makes the obvious conceptual transition from *controllable from* an equilibrium p, to *stabilizable to* an equilibrium p.

1.3 Piecewise constant controls and the CBH formula

A natural first approach to studying controllability is to start with the analysis of trajectories corresponding to piecewise constant controls. As illustrated in the explorations in the parallel parking example in the first section, it is the lack of commutativity that is the key to obtaining *new directions* by conjugation of flows corresponding to different (constant) control values. This section further explores piecewise constant controls and their connection to Lie brackets, which measure the lack of commutativity.

Consider a collection of switching times $0 = t_0 \leq t_1 < t_2 < \ldots < t_{s-1} < t_s = T$ and fixed control values $c_1, c_2, \ldots c_s \in U$, and define the control $u = u_{t_1,t_2,\ldots,t_s;c_1,c_2,\ldots c_s}\colon [0,T] \mapsto U$ by $u(t) = c_i$ if $t_{i-1} < t \leq t_i$ (and e.g. $u(0) = 0$). As a piecewise constant control, u is measurable and thus admissible. The endpoint $x(T,u)$ of the trajectory starting at $x(0)$ is obtained by concatenating the solutions of s differential equations $\dot{x} = f_0(x) + \sum_{j=1}^{m} c_{ij} f_j(x)$. In other words, $x(T,u)$ is obtained from $x(0)$ by composing the flows for times $(t_i - t_{i-1})$ of the vector fields $F_i = f_0 + \sum_{j=1}^{m} c_{ij} f_j$, $i = 1 \ldots s$ and evaluating the composition at $x(0)$. It is customary to write this as a *product of exponentials*

$$x(T,u) = e^{(t_s - t_s - 1)F_s} \cdots e^{(t_3 - t_2)F_1} e^{(t_2 - t_1)F_1} e^{t_1 F_1} x(0) \tag{12}$$

Here the exponential is just a convenient shorthand notation for the flow $(t,p) \mapsto e^{tX}p$ for the flow of the vector field X, i.e. defined by $e^{0X}p = p$ and $\frac{d}{dt}e^{tX}p$ equals the value of the vector field X at the point pe^{tX} for every t (in the domain of the flow).

A word of caution:

While practices vary around the world, and change with time, it is customary in geometric control theory to adopt the convention of writing xf for the value of a function f at a point x (replacing the traditional $f(x)$. In particular, one writes $\frac{d}{dt}pe^{tX} = pe^{tX}X$ for the value of the vector field X at the point pe^{tX}.

In more generality, in an expression $pe^{X}Ye^{Z}\phi$ it is understood that p is a point (on a manifold M), e^{X} the flow of the vector field X at time 1, Y is a vector field on M, e^{Z} is the tangent map of the flow of the vector field Z at time 1, and ϕ is a function on M. Particularly nice is that there is no need for parentheses, or a need to write additional *"stars"* for the tangent maps (see below). E.g. $pe^{X}Y$ is a tangent vector at the point pe^{X}, while e.g. $Y\phi$ is a function on M, $p\phi$ and $pY\phi$ are numbers.

It is important to remember at all times that these exponentials denote flows, and thus they are manipulated exactly as flows are manipulated. In particular, in general $e^{tX}e^{sY} \neq e^{sY}e^{tX}$. However, with careful attention to the legal rules of operation, this proves to be very effective notation for many calculations. For some impressive examples of substantial calculations see [35]. For an extension of this symbolism to *time varying vector fields* see Agrachev [1, 2]. We note on the side, that in the differentiation rules $\frac{d}{ds}pe^{tX}e^{sY} = pe^{tX}e^{sY}Y$ and $\frac{d}{dt}pe^{tX}e^{sY} = pe^{tX}Xe^{sY}$ exponentials to the right of a tangent vector (like pXe^{tY}) stand for the tangent maps of the flows, which in classical differential geometry is often denoted by a lower star: If $\Phi\colon M \mapsto N$ is a map between differentiable manifolds, and $p \in M$ then $\Phi_{*p}\colon T_pM \mapsto T_{\Phi(p)}N$ and $\Phi_{*}\colon TM \mapsto TN$ denote the tangent map in classical notation. In our case, the position of the exponentials will always make it clear which map it stands for, i.e. there is no need to write stars.

In these introductory notes we shall **not** follow this convention. There are just too many examples and calculations from areas other than geometric control where a consistent application of these rules would look very awkward (just think of $x\sqrt{}$ sin). However, we note that the reversal of the order in which certain expressions are to be interpreted will cause the (dis)appearance of sign correction factors $(-)^{k}$ in many places, i.e. one has to be very careful when combining formulas from different sources.

One major advantage of the exponential notation is that it not only matches the symbols used in the study of Lie groups and the symbolism used in formal power series, but that the properties and rules for manipulating them are often identical, making it very easy to mentally move back and fourth.

Rather than directly constructing a control that steers to any given point in a neighborhood of $x(0)$, the first simplification results from using the implicit or inverse function theorem. The basic idea is to construct a comparatively simple control, parameterized e.g. by a finite number of switching times (and/or control values) that *returns* the state to the starting point. If these data are interior (i.e. not extreme values of U), one can to conclude STLC if the Jacobian matrix of this endpoint map has full rank. This basic construction is applicable to much more general settings, compare e.g. the discussion of controllability of partial differential equations in the lectures by Coron. (However, for daily computations in finite dimensional systems we now know simpler tests that will be discussed in the sequel).

Example 1.3 (Stefani [59], 1985)

$$\left\{\begin{array}{lcl} \dot{x}_1 & = & u \\ \dot{x}_2 & = & x_1 \\ \dot{x}_3 & = & x_1^3 x_2 \end{array}\right. \qquad \begin{array}{lcl} x(0) & = & 0 \\ \|u(\cdot)| & \leq & 1 \\ & & \end{array} \tag{13}$$

Consider the piecewise constant controls that take values $+1, -1, +1, -1, 0$ on the intervals $[0, a]$, $(a, a+b]$, $(a+b, a+b+c]$, $(a+b+c, a+b+c+d]$, and $(a+b+c+d, T]$, and calculate the endpoint (using a computer algebra system)

$$\begin{array}{rcl} x_1(T, u_{+-+-;a,b,c,d}) & = & a + c - b - d, \\ x_2(T, u_{+-+-;a,b,c,d}) & = & \frac{1}{2}a^2 + ab - \frac{1}{2}b^2 - bc + ac + \frac{1}{2}c^2 + ad + cd - bd - \frac{1}{2}d^2, \\ x_3(T, u_{+-+-;a,b,c,d}) & = & -2bcd^4 + \frac{1}{12}a^6 + \frac{1}{2}a^5 b + 2ab^3cd - 4a^3bcd - 9a^2bc^2d \end{array}$$

$$\begin{array}{l} -3a^2bcd^2 - 8abc^3d + 6ab^2c^2d + 6ab^2cd^2 + 6abcd^3 - \frac{1}{4}a^4b^2 - \frac{2}{3}a^3b^3 \\ +a^2b^4 - \frac{1}{2}ab^5 + b^2c^4 + \frac{1}{2}b^5c + \frac{1}{12}c^6 + \frac{1}{12}d^6 + 4a^2b^3c - 2a^3b^2c - \frac{5}{2}ab^4c \\ +ab^3c^2 + 2ab^2c^3 - 2bac^4 - \frac{1}{2}ba^4c - 2ba^3c^2 - 3ba^2c^3 + \frac{1}{12}b^6 + 2acd^4 \\ -2ac^2d^3 + \frac{5}{2}ac^4d - ac^3d^2 - \frac{5}{2}abd^4 - 5ab^2d^3 - \frac{5}{2}ab^4d - 5ab^3d^2 \\ +2bc^2d^3 - \frac{5}{2}bc^4d + bc^3d^2 - \frac{1}{4}b^4c^2 - \frac{2}{3}b^3c^3 - \frac{1}{2}bc^5 + \frac{5}{4}a^2c^4 + \frac{1}{2}a^5c \\ +\frac{5}{4}a^4c^2 + \frac{5}{3}a^3c^3 + \frac{1}{2}ac^5 - \frac{1}{2}ad^5 + a^2d^4 + \frac{1}{2}a^5d - \frac{1}{4}a^4d^2 - \frac{2}{3}a^3d^3 \\ +\frac{1}{2}bd^5 + \frac{5}{4}b^2d^4 + \frac{5}{3}b^3d^3 + \frac{1}{2}b^5d + \frac{5}{4}b^4d^2 - \frac{1}{2}a^4bd + \frac{5}{2}a^4cd - 2a^3b^2d \\ -2a^3bd^2 + 5a^3c^2d - a^3cd^2 + 4a^2bd^3 + 4a^2b^3d + 6a^2b^2d^2 - 2a^2cd^3 \\ +5a^2c^3d - 3/2a^2c^2d^2 + 4b^2c^3d - \frac{1}{2}b^4cd - 2b^3c^2d - 2b^3cd^2 - 3b^2cd^3 \\ -\frac{1}{2}cd^5 + c^2d^4 - \frac{2}{3}c^3d^3 + \frac{1}{2}c^5d - \frac{1}{4}c^4d^2. \end{array} \tag{14}$$

It is easy to check that $x(10, u_{+-+-;1,1+\sqrt{2},1+\sqrt{2},1}) = (0,0,0)$, and that

$$\text{rank } \left. \frac{\partial x(10, u_{+-+-;a,b,c,d})}{\partial(a,b,c,d)} \right|_{(1,1+\sqrt{2},1+\sqrt{2},1)} = 3 \tag{15}$$

Thus, by the implicit function theorem, there exists some open neighborhood W of $x(0) = x(10, u_{+-+-;(1,1+\sqrt{2},1+\sqrt{2},1)} = 0$ such that for every $p \in W$, there exists some values (a,b,c,d) near $(1, 1+\sqrt{2}, 1+\sqrt{2}, 1)$ such that $x(10, u_{+-+-;a,b,c,d)} = p$. Thus the system is locally controllable about 0, and via some simple arguments using homogeneity (see the next sections), also STLC about 0.

Challenge exercise 1.13 *(use CAS!). Consider the slightly modified system*

$$\left\{ \begin{array}{lcl} \dot{x}_1 & = & u \\ \dot{x}_2 & = & x_1^3 \\ \dot{x}_3 & = & x_1 x_2 \end{array} \right. \qquad \begin{array}{rcl} x(0) & = & 0 \\ \|u(\cdot)| & \leq & 1 \end{array} \tag{16}$$

Find a piecewise constant, bang-bang control $u\colon [0,T] \mapsto \{-1,+1\}$ *(for some* $T > 0$*) such that corresponding trajectory of* (16) *returns to* 0*, and such that the Jacobian matrix of partial derivatives of the endpoint* $x(T,u)$ *with respect to the switching times has rank* 3 *at your choice of switching times. (Use a computer algebra system.)*

Challenge exercise 1.14 *(use CAS!). Repeat the previous exercise, but now with the values of the piecewise constant control considered as variables, while the switching times are considered fixed. I.e. find a piecewise constant control* $u\colon [0,T] \mapsto (-1,1)$ *(for some* $T > 0$*)* $u_i(t) = c_i$ *if* $t_{i-1} \leq t \leq t_i$ *such that corresponding trajectory of* (16) *returns to* 0*, and such that the Jacobian matrix of partial derivatives of the endpoint* $x(T,u)$ *with respect to the values* c_i *of the control has rank* 3 *at your choice of control values. Use a computer algebra system.*

In terms of compositions of flows or products of exponentials the previous example employed

$$x(10,u) = e^{(10-a-b-c-d)f_0} e^{d(f_0-f_1)} e^{c(f_0+f_1)} e^{b(f_0-f_1)} e^{a(f_0+f_1)}(0) \tag{17}$$

and found that in particular

$$x(10,u_*) = e^{(8-2\sqrt{2})f_0} e^{(f_0-f_1)} e^{(1+\sqrt{2})(f_0+f_1)} e^{(1+\sqrt{2})(f_0-f_1)} e^{(f_0+f_1)}(0) = 0 \tag{18}$$

Differentiation of (17) with respect to the times a, b, c, d then was used to establish controllability. (Compare exercise 4.12 for manual symbolic manipulations.) For many systems this approach is impractical as e.g. exact switching times which return the system to the starting point may be difficult or practically impossibleto find. Thus it is natural to look for alternative methods. In particular, the key is to study the lack of commutativity.

Recall that the Lie bracket $[F_1, F_2]$ of two smooth vector fields F_1 and F_2 on a manifold M is algebraically defined as the vector field $[F_1, F_2]\colon C^\infty(M) \mapsto C^\infty(M)$ via $[F_1, F_2]\phi = F_1(F_2\phi) - F_1(F_2\phi)$.
In coordinates, with vector fields written as column vectors, and denoting the Jacobian matrix by D, one calculates the Lie bracket as $[F_1, F_2] = (DF_2)F_1 - (DF_1)F_2$.

Example 1.4

Consider $f_0(x) = x_1 \frac{\partial}{\partial x_2}$ and $f_1(x) = \frac{\partial}{\partial x_1}$. Then

$$\begin{aligned}[f_0, f_1](x) &= \left(x_1 \frac{\partial}{\partial x_2}\right) \circ \left(\frac{\partial}{\partial x_1}\right) - \left(\frac{\partial}{\partial x_1}\right) \circ \left(x_1 \frac{\partial}{\partial x_2}\right) \\ &= x_1 \left(\frac{\partial^2}{\partial x_1 \partial x_2} - \frac{\partial^2}{\partial x_2 \partial x_1}\right) - \frac{\partial x_1}{\partial x_1}\frac{\partial}{\partial x_2} = -\frac{\partial}{\partial x_2} \end{aligned} \tag{19}$$

In matrix / column vector notation the same calculation reads

$$[\begin{pmatrix} 0 \\ x_1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}] = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ x_1 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \tag{20}$$

Exercise 1.15 *For the vector fields* $f_0(x) = x_1^4 \frac{\partial}{\partial x_2}$ *and* $f_1(x) = \frac{\partial}{\partial x_1}$ *calculate the iterated Lie brackets* $[f_0, f_1]$, $[f_0, [f_0, f_1]]$, *and* $[[f_0, f_1], f_1]$.
Find an iterated Lie bracket f_π *(of higher order) such that* $f_\pi(0) = \frac{\partial}{\partial x_2}$.

Geometrically,the Lie bracket is defined by the limit (for $\phi \in C^\infty(M)$)

$$[f_1, f_2]\phi(p) = \lim_{t \longrightarrow 0} \frac{1}{t^2} \left(\phi(e^{-tf_2} e^{-tf_1} e^{tf_2} e^{tf_1} p) - \phi(p)\right) \tag{21}$$

i.e. as the infinitesimal measure of the lack of commutativity of the flows of f_1 and f_2 at the point p. It is very instructive to calculate these flows in a simple explicit example, and see how the limit gives rise to a new direction.

Back to example 1.4. Staring at $p = (p_1, p_2)$, calculate

$$\begin{array}{rcl} p' = e^{tf_0}(p) & = & (p_1, p_2 + tp_1) \\ p'' = e^{tf_1}(p') & = & (p_1 + t, p_2 + tp_1) \\ p''' = e^{-tf_0}(p'') & = & (p_1 + t, p_2 - t^2) \\ p'''' = e^{-tf_1}(p''') & = & (p_1, p_2 - t^2) \end{array} \tag{22}$$

and thus

$$[f_0, f_1]\phi(p) = \lim_{t\to 0} \frac{1}{t^2}(\phi(p'''') - \phi(p)) \lim_{t\to 0} \frac{1}{t^2}(\phi(p_1, p_2 - t^2) - \phi(p_1, p_2)) = -\frac{\partial \phi}{\partial x_2}(p) \tag{23}$$

which is in agreement with the earlier algebraic calculation that yielded $[f_0, f_1] = -\frac{\partial}{\partial x_2}$.

Exercise 1.16 *Check the calculations in* (22)*.* (Note, a common source of confusion is the very questionable use of the same symbol t for both the length of each interval and also as integration variable along each piece.)*. Sketch the curves, and repeat all with the order of* f_1 *and* f_2 *reversed. Work out the special case of* $p = 0$ *– this is a useful case to remember as it helps to recover the sign-convention used in any locality.*

One of the major goals of these lectures is to develop methods and tools that allow one to more easily work with the compositions of noncommuting flows. One of the oldest such tools is the classical Campbell Baker Hausdorff formula which asserts that

$$e^X \cdot e^Y = e^{\log(e^X \cdot e^Y)} = e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]+\ldots} \tag{24}$$

One of the nice features of this formula is that it is just as correct in the sense of formal power series in *noncommuting indeterminates* X and Y, as it is correct for analytic vector fields X and Y (as long as all flows are defined). This is no accident! It is easy to informally verify this identity by simply using the standard Taylor expansion for exponentials, formally expanding both sides and recursively using the definition $[V, W] = VW - WV$. A rigorous justification that one can indeed go easily back and fourth between geometric/analytic and algebraic/combinatorial interpretations can be made in many ways – indeed, manipulations of analytic objects are by nature often purely algebraic. Arguably one of the more elegant ones starts with the classical identification of points on a manifold with multiplicative functionals on the algebra of smooth functions on a manifold, and then proceeds with identifying flows with formal partial differential operators of infinite order, compare e.g. [19, 35].

Exercise 1.17 *Repeatedly use the CBH-formula to write the end point of 4 flows corresponding to* bang-bang *controls as a single* exponential*:*

$$\begin{aligned} x(T, u_{t_1,t_2,t_3}) &= 0 \cdot e^{t_1(f_0+f_1)} \cdot e^{(t_2-t_1)(f_0-f_1)} \cdot e^{(t_3-t_2)(f_0+f_1)} \cdot e^{(T-t_3)(f_0-f_1)} \\ &\overset{!}{=} \; e^{p_0(t)f_0+p_1(t)f_1+p_{01}(t)[f_0,f_1]+p_{011}(t)[f_0,[f_0,f_1]]+p_{101}(t)[f_1,[f_0,f_1]]+\cdots}(0) \end{aligned} \tag{25}$$

Find explicit formulas for polynomial expressions $p_I(t) = p_I(t_1, t_2, t_3, t_4)$ (in the switching times) for $I = 0, 1, 01, 011, 110$.

The following lectures aim at obtaining similar formulas that are easier to use, and that also allow for controls that are not necessarily piecewise constant! The starting point will be the Chen Fliess series expansion.

1.4 Approximating cones and conditions for STLC

Instead of constructing controls that steer exactly to a specific point, which generally is very hard, *analysis* is about building arguments that use approximate directional information obtained from derivatives. This discussion also establishes a close link between STLC and optimal control.

The key idea is to develop a tangent, or derivative object for the reachable sets that is easy to constuct/compute, that has reasonably nice convexity properties, and that nicely approximates the reachable sets. While prior efforts in control largely focused on constructing specific kinds of control variations and then created arguments why these control variations can be combined especially, Frankowska [16, 17] pioneered a different approach that provides a very general open mapping principle under very general hypotheses, which then may be applied to many special cases after simply checking that the specific tangent vectors satisfy some mild conditions. Its general theory applies in Banach space settings, and requires only minimal smoothness (thus the name *nonsmooth analysis*. In these notes we generally follow the original definitions and constructions of [16, 17]. The open mapping theorem stated below is a very special case of the general results in [16, 17], and we use its simple language as it is meets exactly the needs of our systems.

The following is one of the most simple possible notions of tangent vectors, yet in the special case of affine smooth systems these vectors are automatically in the *contingent cone* and thus Frankowska's general open mapping principle [17] applies.

Definition 1.3 ([17, 29]) *Consider systems of form* (8) *on* $\mathbb{R}^n$ *with* $f_0(0) = 0$ *and* $0 \in \operatorname{int} U$*. A vector* $\xi \in R^n$ *is called a k-th order tangent vector to* *the family* $\{\mathcal{R}_t(0)\}_{t\geq 0}$ *at* 0 *if there exists a parameterized family of control variations* $u_s\colon [0,s] \mapsto U$, $s \geq 0$, *such that*

$$x(s, u_s) = 0 + s^k \xi + o(s^k). \tag{26}$$

The set of all k-th order tangent vectors (to $\{\mathcal{R}_t(0)\}_{t\geq 0}$ at zero) *is denoted by* C^k, *while* $\overline{C^k} = \bigcup_{\lambda>0} \lambda C^k$ *is the set of* tangent rays *to* $\{\mathcal{R}_t(0)\}_{t\geq 0}$ *at zero.*

The parameterization $s \mapsto u_s$ is not required to be smooth. Indeed, it suffices to require sequences $s_k \searrow 0$.

Exercise 1.18 *Find 6 families of control variations* $u_s^{\pm i}\colon [0,s] \mapsto [-1,1]$ *that generate the tangent vectors* $\pm \frac{\partial}{\partial x_i}\Big|_0$, $i = 1,2,3$, *for the system* (13).

Exercise 1.19
Repeat the previous exercise using the control sizes, as in exercise 1.14, *as parameter* s.

The following properties are easy to establish:

Proposition 1.2 ([16, 17, 29])
- (a) *If* $\lambda^k \in [0,1]$, *then* $\lambda^k C^k \subseteq C^k$.
- (b) *If* $k \leq \ell$ *then* $C^k \subseteq C^\ell$.
- (c) *If* $v_1,\, v_2 \in C^k$ *and* $\lambda^k \in [0,1]$ *then* $\lambda^k v_1 + (1-\lambda)^k v_2 \in C^k$.

Thus the sets C^k form an increasing sequence of truncated convex cones. The *approximation* property of these cones is established by:

Theorem 1.3 ([16, 17, 18, 29])
If $\overline{C'}$ *is a closed convex cone (with vertex* $0 \in \mathbf{R}^n$*) such that* $\overline{C'} \setminus \{0\} \subseteq \mathrm{int}\overline{C^k}$ *for some* $k < \infty$, *then there exist* $c > 0$, $T > 0$ *such that* $\overline{C'} \cap B(0, ct^k) \subseteq \mathcal{A}(t)$ *for all* $0 \leq t \leq T$.

In [29] an explicit constructive proof is given for the theorem in the special case for systems of form (8), which has subsequently used as a starting point for feedback stabilization. But analogous results hold in much more generality, see. Frankowska [17, 18] for infinite dimensional versions only requiring minimal regularity.

Corollary 1.4 *If* $\overline{C^k} = \mathbf{R}^n$ *then there are constants* $c > 0$, $T > 0$ *such that* $B(0, ct^k) \subseteq \mathcal{A}(t)$ *for all* $0 \leq t \leq T$.

These approximating cones are just as useful for obtaining high-order versions of the maximum principle of optimal control as basically the dichotomy is the same: Does the reference trajectory (here $x \equiv 0$ lie in the interior or on the boundary of the reachable sets?

The preceding discussions, examples, and exercises suggest that there should be universal families of control variations that generate specific Lie brackets as tangent vectors to the reachable sets. This is indeed the case – and much research in the 1980 focused on developing the following conditions, which really emanate from arguments why certain families of control variations generate some Lie brackets. First introduce the following notation:

For smooth vector fields f and g define recursively define $(ad^0 f, g) = g$ and $(\mathrm{ad}^{k+1} f, g) = [f, (\mathrm{ad}^k (f, g)]$. For smooth vector fields $f_0, f_1, \ldots f_m$ let $L(f_0, f_1, \ldots f_m)$ denote the Lie algebra spanned by all iterated brackets of the vector fields f_i.
For any multi-index $r = (r_0, r_1, \ldots r_m) \in \mathbf{Z}^{+(m+1)}$ let $L^r(f_0, f_1, \ldots f_m)$ be the subspace spanned by all iterated brackets with r_i factors f_i, $i = 0, 1, \ldots m$. Also write $\mathcal{S}^k(f_0, f_1, \ldots f_m)$ for the subspace spanned by all iterated brackets exactly k factors from $f_1, , \ldots f_m$ and any numbers of factors f_0. For a set S of vector fields and a point p, we write $S(p)$ for the set $\{v(p) \colon v \in S\}$.
Note that it is possible to have e.g. $0 \neq [f_1, [f_1, f_0]] \in L^{(0,1)}(f_0, f_1)$ as $[f_1, [f_1, f_0]] = f_1$ is possible, e.g. if $f_1 = \frac{\partial}{\partial x_1}$ and $f_2 = \frac{1}{2} x_1^2 \frac{\partial}{\partial x_1}$. See the next chapter for more elaborate language that carefully distinguishes *binary labeled trees* (or *formal brackets* from Lie brackets. Since all our considerations are local, we identify the tangent space $T_0 \mathbb{R}^n$ with $\mathbb{R}^n$.

Recall that for nonlinear systems in general accessibility and controllability are not equivalent. For analytic systems, accessibility is comparatively easy to decide.

Theorem 1.5
The system (8) *initialized at* $x(0) = 0$ *is accessible if and only if* $\dim L(f_0, f_1, \ldots f_m)(0) = n$.

If the system is reversible, e.g. if U is symmetric about 0 and the system has no drift ($f_0 \equiv 0$), then accessibility implies STLC. (Consequently, the controlled kinematics are comparatively easy to deal with as opposed to the full dynamic model. Compare (5) versus (6).)
The closest analogue of the Kalman rank condition (theorem 1.1) for linear systems is the following condition, which basically says that if the Taylor

linearization is linearly controllable, then the original system is controllable (STLC) in the sense of nonlinear systems.

Theorem 1.6 (Linear Test) *If* $\mathcal{S}^1(0) = \mathbf{R}^n$ *then the system* (8) *is STLC.*

A complementary necessary condition for single-input system is a special case of the Clebsch-Legendre condition of optimal control:

Theorem 1.7
If $m = 1$ *and the system* (8) *is STLC then* $[f_1, [f_0, f_1]](0) \in \mathcal{S}^1(f_0, f_1)(0)$.

The exercises in the preceding section and above, aimed at generating tangent vectors via families of control variations should have suggested that for certain brackets their negatives are generated by the negatives of the controls. On the other hand system (11) shows that at least some *even* powers may be *obstructions* to STLC. The correctness of this intuition is formally established in:

Theorem 1.8 (Hermes [22], Sussmann [62]) *If* $m = 1$ *and* (8) *is accessible, and* $\mathcal{S}^{2k}(f_0, f_1)(0) \subseteq \mathcal{S}^{2k-1}(f_0, f_1)(0)$ *for all* $k \in \mathbf{Z}^+$ *then* (8) *is STLC.*

A complementary necessary condition is:

Theorem 1.9 (Stefani [60]) *If* $m = 1$ *and the system* (8) *is STLC then* $(ad^{2k} f_0, f_1)(0) \in \mathcal{S}^{2k-1}(f_0, f_1)(0)$ *for all* $k \in \mathbf{Z}^+$.

Finally, the most general sufficient condition known today allows one to weight the drift and the controlled fields differently when counting the order of a bracket.

Theorem 1.10 (Sussmann [65]) *If the system the system* (8) *is accessible and there exists a weight* $\theta \in (0, 1]$ *such that for all odd* k *and even* $\ell_1, \ldots \ell_m$

$$L^{(k,\ell_1,\ldots\ell_m)}(f_0, \ldots f_m)(0) \subseteq \sum_{(k^i,\ell^i)} L^{(k^i,\ell^i_1,\ldots\ell^i_m)}(f_0, \ldots f_m)(0) \qquad (27)$$

where the sum extends over all (k^i, ℓ^i) *such that*

$$\theta k^i + \ell^i_1 + \ldots + \ell^i_m < \theta k + \ell_1 + \ldots + \ell_m \qquad (28)$$

then the system (8) *is STLC.*

Loosely phrased, this theorem singles out brackets of type (odd,even,. . . even) as *potential obstructions* to STLC, and it describes a way how these may be *neutralized* so that the system is STLC. The commonly used term *bad brackets*[??] for the potential obstructions is unfortunate since the obstructions should be identified with *supporting hyperplanes* of the the approximating cones, and thus are elements of the *dual space* of a free Lie algebra, compare [39]. (It is also possible to also use different weights for the controlled fields. The best notation uses weight 1 for f_0 and weights $\sigma_i = \frac{1}{\theta_i} \in [1, \infty)$ for the fields f_i.) Several small extensions, and specific examples that explore the region between the necessary and sufficient conditions can be found in the literature, see e.g. [32] for an overview.

Example (1.3) revisited. Extract the vector fields f_0 andf_1 from (13) and calculate iterated Lie brackets.

$$f_0(x) = \begin{pmatrix} 0 \\ x_1 \\ x_1^3 x_2 \end{pmatrix},\ f_1(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ [f_0, f_1](x) = \begin{pmatrix} 0 \\ -1 \\ 3x_1^2 x_2 \end{pmatrix},\ [f_0, [f_0, f_1]](x) = \begin{pmatrix} 0 \\ 0 \\ 6x_1^3 \end{pmatrix}$$

$$[f_1, [f_1, f_0]](x) = \begin{pmatrix} 0 \\ 0 \\ -6x_1 x_2 \end{pmatrix},\ (\mathrm{ad}^3 f_1, f_0)(x) = \begin{pmatrix} 0 \\ 0 \\ -6x_2 \end{pmatrix},\ (\mathrm{ad}^4 f_1, f_0)(x) = 0 \quad (29)$$

$$[f_0, [f_1, [f_1, f_0]](x) = [f_1, [f_0, [f_1, f_0]](x) = \begin{pmatrix} 0 \\ 0 \\ 12x_1^3 \end{pmatrix},\ [f_1, f_0], (\mathrm{ad}^3 f_1, f_0)](x) = \begin{pmatrix} 0 \\ 0 \\ -6 \end{pmatrix}$$

Note that $(\mathrm{ad}^k f_0, f_1)(x) = 0$ when $k > 2$. Since $\dim L(f_0, f_1)(0) = 3$ the system is accessible, but is *not linearly controllable* due to $\dim \mathcal{S}^1(f_0, f_1)(0) = 2 < 3$. Since $[f_1, [f_1, f_0]](0) = 0 \in \mathcal{S}^1(f_0, f_1)(0)$ and similarly $(\mathrm{ad}^4 f_1, f_0)(0) = 0 \in \mathcal{S}^3(f_0, f_1)(0)$, Stefani's necessary conditions are satisfied. The only brackets which give the $\frac{\partial}{\partial x_3}$ direction have 4 factors of f_1, and thus the Hermes' condition does not apply. However, since all brackets with an even number of factors f_1 and an odd number of factors f_0 vanish at 0, Sussmann's condition affirms STLC – something which we proved earlier by a brute-force construction.

Historical note: This example by Stefani was first shown to be STLC using the method we exhibited in the previous section. It clearly shows that the Hermes condition is far from necessary, and it served as a substantial motivation for the eventual sharpened version of Sussmann's general theorem that was proven in [65].

With a computer algebra system such calculations are very easy and quickly executed – but the big question is: Which brackets does one have to calculate? For very short brackets it is quite obvious that e.g. only one of $\{[f_0,[f_0,f_1]],\ [f_0,[f_1,f_0]],\ [f_1,[f_0,f_0]],\ [[f_0,f_0],f_1],\ [[f_0,f_1],f_0],\ [[f_1,f_0],f_0]\}$ needs to be computed (due to anticommutativity $[X,Y]=-[Y,X]$ and the Jacobi identity $[X,[Y,Z]]+[Z,[X,Y]]+[Y,[Z,X]]=0$ for all X,Y,Z in a Lie algebra). But as the length increases, the number of a-priori possible brackets sky-rockets, yet it is apparent that there will be lots of duplication. The subsequent lectures on combinatorics and algebra will provide nice answers by providing bases that are very easily constructed. The question "when can one stop?" is also answered in the next lecture (for nilpotent systems).

Exercise 1.20 *Determine whether the car model* (9) *from example* (1.1) *is STLC.*

Exercise 1.21 *Determine whether the models* (5) *and* (6) *for the kinematics and the dynamics of the rolling penny example* (1.2) *are STLC.*

2 Series expansion, nilpotent approximating systems

2.1 Introduction to the Chen Fliess series

Much classical work investigated the whether the sets of points reachable by piecewise constant controls agree with those reachable by means of arbitrary measurable controls, see e.g. [15] Grasse (late 1980s). But one may expect that in general one may need very large numbers of *pieces* in order to well approximate measurable controls. The subsequent very large number of repeated applications of the CBH-formula is even less attractive. Thus one is led to look for expansions that do not rely on piecewise constancy, and which allow one to combine a large number of *pieces* in one step.

One of the most basic formulas is obtained by simple Picard iteration. For an analytic *output function* $\phi\colon \mathbb{R}^n \mapsto \mathbb{R}$ (especially, for $\phi = x_i$ a coordinate function) first rewrite the system of differential equations with initial condition

$$\frac{d}{dt}\phi(x(t)) = \sum_{i=0}^{n} u_i(t)(f_i\phi)(x(t)), \quad \phi(x(0)) = \phi(p) \tag{30}$$

as an equivalent integral equation, and then iterate the analogous rewriting for the subsequently appearing Lie derivatives $(f_{i_s} \ldots f_{i_2} f_{i_1} \phi)$ of ϕ

$$\begin{aligned}
\phi(x(t)) = \phi(p) + & \int_0^t \sum_{i_1=0}^m u_{i_1}(t_1)(f_{i_1}\phi)(x(t_1))\, dt_1 \\
= \phi(p) + & \int_0^t \sum_{i_1=0}^m u_{i_1}(t_1)\Big((f_{i_1}\phi)(p) + \int_0^{t_1} \sum_{i_2=0}^m u_{i_2}(t_2)(f_{i_2} f_{i_1}\phi)(x(t_2))\, dt_2 \Big)\, dt_1 \\
= \phi(p) + & \int_0^t \sum_{i_1=0}^m u_{i_1}(t_1)\Big((f_{i_1}\phi)(p) + \int_0^{t_1} \sum_{i_2=0}^m u_{i_2}(t_2)\Big((f_{i_2} f_{i_1}\phi)(p) \\
& \qquad + \int_0^{t_2} \sum_{i_3=0}^m u_{i_3}(t_3)(f_{i_3} f_{i_2} f_{i_1}\phi)(x(t_3))\, dt_3 \Big)\, dt_2 \Big)\, dt_1
\end{aligned} \tag{31}$$

and so on. Note that each of these is an *exact* equation, where the last term to be considered an *error term* in a finite series approximation. The usefulness of this expansion is that it separates the time- and control dependence of the solution from the nonvarying geometry of the system which is captured by the vector fields f_i (or in the iterated Lie derivatives $(f_{i_s} \ldots f_{i_2} f_{i_1} \phi)$ which may be computed off-line, and only once). For compatibility with later formulas, we reverse the names of the integration variables and indices used, e.g. rename i_1 to become i_3 and vice versa, and expand the sums

$$\begin{aligned}
\phi(x(t)) = \phi(p) + & \sum_{i_1=0}^m \left(\int_0^t u_{i_1}(t_1) dt_1 \right) \cdot (f_{i_1}\phi)(p) \\
+ & \sum_{i_2=0}^m \sum_{i_1=0}^m \left(\int_0^t \int_0^{t_1} u_{i_2}(t_2) u_{i_1}(t_1) dt_1 dt_2 \right) (f_{i_1} f_{i_2} \phi)(p) \\
+ & \sum_{i_3=0}^m \sum_{i_2=0}^m \sum_{i_1=0}^m \left(\int_0^t \int_0^{t_3} \int_0^{t_2} u_{i_3}(t_3) u_{i_2}(t_2) u_{i_1}(t_1) dt_1 dt_2 dt_3 \right) (f_{i_1} f_{i_2} f_{i_3} \phi)(x(t_3))
\end{aligned} \tag{32}$$

Note that the indices in the partial derivatives and in the integrals are in *opposite order*. The pattern emerges clearly, and this procedure may be iterated ad infinitum, yielding a formal infinite series. (In a later lecture we shall repeat this derivation using the very compact notation of chronological products, without writing any integrals.)

Definition 2.1 (Chen-Fliess series) *For any measurable control* $u\colon [0,T] \mapsto \mathbb{R}^{m+1}$ *and a set of* $(n+1)$ *indeterminates* X_0, X_1,$\ldots X_m$ *define the formal* *series*

$$S_{CF}(T,u)=\sum_I \underbrace{\int_0^T\int_0^{t_{p-1}}\cdots\int_0^{t_3}\int_0^{t_2} u^{i_p}(t_p)\ldots u^{i_1}(t_1)\,dt_1\ldots dt_p}_{\Upsilon^I(u)(T)} \underbrace{X_{i_1}\ldots X_{i_p}}_{X_I} \tag{33}$$

where the sum ranges over all multi-indices $I=(i_1,\ldots i_s)$*,* $s\geq 0$ *with each* $i_j\in\{0,1,\ldots m\}$*.*

This series originates in K. T. Chen's study [6] in the 1950s of geometric invariants of curves in $\mathbb{R}^n$. In the early 1970s Fliess recognized its utility for the analysis of control systems. Using careful analytic estimates one may prove (compare [62]) that this so far only formal series actually converges

Theorem 2.1 *Suppose* f_i *are analytic vector fields on* $\mathbb{R}^n$*,* $\phi\colon\mathbb{R}^n\mapsto\mathbb{R}$ *is analytic and* $U\subset\mathbb{R}^{m+1}$ *is compact. Then for every compact set* $K\subseteq\mathbb{R}^n$*, there exists* $T>0$ *such that the series (with* Υ^I *and the range of the sum as above)*

$$S_{CF,f}(T,u)(\phi)=\sum_I \Upsilon^I(u)(T)\cdot(f_I\phi)(p) \tag{34}$$

converges uniformly to the solution $x(t,u)$ *of* (30) *with* $x(0)=p$ *for* $p\in K$ *and* $u\colon[0,T]\mapsto U$*.*

This series solution is not just good for piecewise constant controls, but for all measurable controls. To get a better feeling for the terms, revisit example 13 with

$$f_0=x_1\frac{\partial}{\partial x_2}+x_1^3x_2\frac{\partial}{\partial x_3} \text{ and } \quad f_1=\frac{\partial}{\partial x_1} \tag{35}$$

and consider the Chen-Fliess series for the coordinate functions $\phi=x_i$ about $p=0$. As usual we use $u_0\equiv 1$ and write $u_1=u$. Obviously, for $\phi=x_1$, the series collapses to a single term, yielding $x_1(T,u)(u)(T)=\Upsilon^1(u)(T)=\int_0^T u(t)dt$. For $\phi=x_2$, the series collapses to the single term corresponding to the multi-index $(1,0)$(or *"word"* 10)

$$x_2(T,u)=\Upsilon^{10}(u)(T)=\textstyle\int_0^T\int_0^{t_2}u(t_2)u_0(t_1)\,dt_1\,dt_2=\int_0^T\int_0^{t_2}u(t_1)\,dt_1\,dt_2 \tag{36}$$

As expected, the series just returns the integral form for the linear, double integrators part of the system 13.

For $\phi=x_3$ note that $f_1x_3\equiv 0$ and $f_0x_3=x_1^3x_2$. (Meticulous attention to the *two slots* of differential operators and careful notation are advised: A differential operator X acts on a function Φ and is evaluated at a point p – the usual identification of points with their coordinates causes the appearance of the same symbol x in both slots!) Next, e.g. $f_1 f_0 x_3 = 3x_1^2 x_2$, while $f_0 f_0 x_3 = x_1^4$. We leave further calculations to

Exercise 2.1 (Important!) *Continuing this example, find all partial derivatives $f_I x_3$ which are not identically zero. What is the highest order non-zero derivative (length of the word I)? How could you have found that length* by inspection*, without calculating any partial derivatives? Find all words I for which $(f_I x_3)(0) \neq 0$ and calculate the values of these derivatives at $p = 0$. Write out the corresponding iterated integrals and explicitly write out the Chen-Fliess series expansion for $x_3(T, u)$.*

The previous exercise, and the following challenge are excellent motivation for all later work. It really helps to first get one's hands dirty with comparatively naive and messy hand-calculations. This way the later elegant combinatorial and algebraic simplifications will be much more appreciated!

Exercise 2.2 (Important!) *Compare the resulting expression of the previous exercise with the* obvious *integral formula*

$$x_3(T, u) = \int_0^T \left(\left(\int_0^{t_3} u(t_2)\, dt_2 \right)^3 \cdot \left(\int_0^{t_3} \int_0^{t_2} u(t_1)\, dt_1\, dt_2 \right) \right) dt_3. \qquad (37)$$

Reconcile these expressions via repeated integration by parts and suitably combining terms.

The example considered above is apparently very special, yielding finite, *polynomial* series expansions in terms of *iterated integrals*. This property is easily traced to the *triangular nature* of the (Jacobian matrices of the) vector fields f_i *together* with their polynomial entries. Such very desirable structure is indeed the objective of nilpotent approximations, to be discussed in after we introduce some technical tools in the next section.

2.2 Families of dilations

For (nonconstant) polynomials of one variable, each derivative lowers the degree by one – something similar clearly is happening in the iterated Lie derivatives in the examples considered in previous sections. This is complemented by the degree with respect to the switching times in the *responses*

$x(T, u)$ in the explicit constructions of the previous chapter. The formal definitions of *families of dilations* are useful to capture the apparent patterns, and for e.g. to identify *leading terms* allowing one to construct approximating systems. Working with fixed coordinates $(x_1, x_2, \dots x_n)$ it is convenient to make the following definition which is a special case of the general geometric, coordinate free, notion of homogeneity of [33]:

Definition 2.2 *Consider* $\mathbb{R}^n$ *with fixed coordinates* $(x_1, x_2, \dots x_n)$ *and* $r_1, r_2, \dots r_n \geq 1$. *A* one-parameter family of dilations *is a map* $\Delta\colon \mathbb{R}^+ \times \mathbb{R}^n$ *defined by*

$$\Delta_s(x) = (s^{r_1} x_1, s^{r_2} x_2, \dots, s^{r_n} x_n). \tag{38}$$

A smooth function $\phi\colon \mathbb{R}^n \mapsto \mathbb{R}$ *and a smooth vector field* F *on* $\mathbb{R}^n$ *are homogeneous of degrees* m *and* k *(with respect to* Δ*), written* $\phi \in H_m$ *and* $F \in \underline{n}_k$*, respectively, if*

$$\phi \circ \Delta_s = s^m \phi \quad \textit{and } F x_k \in \mathcal{H}_{m+r_k} \quad \textit{for } k = 1, 2, \dots, n. \tag{39}$$

The Euler vector field *of this dilation is the vector field*

$$\nu(x) = r_1 x_1 \frac{\partial}{\partial x_1} + r_2 x_2 \frac{\partial}{\partial x_2} + \dots + r_n x_n \frac{\partial}{\partial x_n}. \tag{40}$$

For example consider $n = 3$, $r = (1, 2, 6)$. The practical meaning of the *exponents* r_i are as *weights* of the coordinate functions, i.e. $x_1 \in \mathcal{H}_1$, $x_2 \in \mathcal{H}_2$ and $x_3 \in \mathcal{H}_6$. With these weights e.g. $\phi(x) = x_1 x_3 - x_1^7 + x_1 x_2^3 \in \mathcal{H}_7$ is homogeneous of degree 7.
Similarly, the coordinate vector fields are homogeneous of degrees $\frac{\partial}{\partial x_1} \in \underline{n}_{-1}$, $\frac{\partial}{\partial x_2} \in \underline{n}_{-2}$, and $\frac{\partial}{\partial x_3} \in \underline{n}_{-6}$. The Lie derivatives of the homogeneous polynomial ϕ in the directions of the coordinate fields are again homogeneous $\frac{\partial}{\partial x_1}\phi \in \mathcal{H}_6$, $\frac{\partial}{\partial x_1}\phi \in \mathcal{H}_5$, and $\frac{\partial}{\partial x_1}\phi \in \mathcal{H}_6$.

The following properties hold also for more general, geometric dilations as defined in [33].

Proposition 2.2
Let Δ *be a one-parameter family of dilations on* $\mathbb{R}^n$ *with coordinates* $(x_1, \dots x_n)$.
If $\phi \in \mathcal{H}_m$*, and* $\psi \in \mathcal{H}_k$*, then* $\phi\psi \in \mathcal{H}_{m+k}$.
If $F \in \underline{n}_m$*, and* $G \in \underline{n}_k$ *then* $[F, G] \in \underline{n}_{m+k}$.
If $\phi \in \mathcal{H}_m$*, and* $F \in \underline{n}_k$*, then* $F\phi \in \mathcal{H}_{m+k}$.
If $m < -r_n$ *then* $\underline{n}_m = \{0\}$.

Together with the obvious properties for sums, these properties provide the algebras of polynomials and of polynomial vector fields with graded structures: E.g. every polynomial can be uniquely written as a sum of homogeneous polynomials, and every polynomial vector field can be uniquely decomposed into a sum of homogeneous vector fields. This is used e.g. in nilpotent approximating systems and for high-order analogues of linear stability (if the *leading term* of a dynamical system is asymptotically stable, then the system is locally asymptotically stable, compare e.g. [23, 56]

Exercise 2.3 *Prove the assertions made in proposition* (2.2)*.*

The Euler vector field ν is (up to rescaling by a logarithm) the *infinitesimal generator* of the dilation group Δ, and it allows for particularly elegant characterizations of homogeneity.

Proposition 2.3 [33]
Let Δ be a one-parameter family of dilations on $\mathbb{R}^n$ with coordinates $(x_1, \ldots x_n)$.
A smooth function ϕ on $\mathbb{R}^n$is homogeneous $\phi \in \mathcal{H}_m$ iff $\nu\phi = m\phi$.
A smooth vector field on $\mathbb{R}^n$is homogeneous $F \in \underline{n}_m$, and $G \in\in \underline{n}_k$ iff $[\nu, F] = mF$.

Actually, one may start with a vector field ν such that $\dot{x} = -\nu(x)$ is globally asymptotically stable, and then define a dilation Δ associated to ν via the properties in proposition 2.3 [33].

Exercise 2.4 *Prove the assertions made in proposition* (2.3)*.*

A typical use of the last property in proposition (2.2) is to justify stopping to compute Lie brackets of vector fields after reaching a certain maximal length. E.g. suppose that the vector fields f_0 and f_1 have polynomial components and for a some choice of exponents $(r_1, \ldots r_n)$ they are sums of homogeneous vector fields all of which have negative degrees. Then every bracket of length larger than $-r_n$ is identically zero.
Reconsider the vector fields $f_0 = (0, x_1, x_1^3 x_2)^T$ and $f_1 = (1, 0, 0)^T$ from example (1.3). These are homogeneous of degrees $f_0 \in \underline{n}_0$ and $f_1 \in \underline{n}_{-1}$ with respect to the dilation defined by $r = (1, 1, 4)$, while they are homogeneous of degrees $f_0, f_1 \in \underline{n}_{-1}$ with respect to the dilation defined by $r = (1, 2, 7)$. Using the second dilation we conclude that any Lie bracket involving more than 7 factors f_0 or f_1, in any order, with any bracketing is identically zero.
Recall:

Definition 2.3 *A Lie algebra L is called* nilpotent *if there exists a number s such that every iterated Lie bracket of elements of L of length greater than s is zero.*

Thus in the example, we conclude that $L(f_0, f_1)$ is nilpotent. It can be shown [31] that if the Lie algebra $L(f_0, f_1, \ldots f_n)$ is nilpotent, then the control system (8) can be brought into a strictly lower triangular form with polynomial vector fields (with well-defined maximal degrees) via a change of local coordinates: In the new coordinates each component $f_i x_j$ is a polynomial in $x_1, x_2, \ldots x_{j-1}$ only! Consequently, solution curves corresponding to any control $u(t)$ can be found by simple integrations of functions of a single variable, no nontrivial differential equations need to be integrated! This makes nilpotent systems a very attractive class to work with, and predestined to serve as a class of approximating systems – to be discussed in the next section.

The examples, and especially exercises 1.18 and 1.19, using piecewise constant controls also illustrated that, at least in the case of *homogeneous* systems, the length of each Lie bracket corresponds to the degree of the the polynomial expression in the data (switching times, control values). This is made precise using the notion of homogeneity.
Fix a control $u\colon [0,T] \mapsto U$. For $\varepsilon, \delta \in [0,1]$ define the families of rescaled controls

$$u_{\varepsilon,\delta}\colon [0, \delta T] \mapsto \varepsilon U \subseteq U \quad \text{by} \quad u_{\varepsilon,\delta}(\delta t) = \varepsilon u(t) \tag{41}$$

For the scaling by *amplitude*, using ε, to make sense, assume that the set U is star-shaped with respect to zero, i.e. $[0,1]U \subseteq U$ (meaning $\lambda c \in U$ for all $c \in U$ and all $0 \leq \lambda \leq 1$).

Proposition 2.4 *Suppose $\Delta\colon (s,x) \mapsto \Delta_s(x) = (s^{r_1} x_1, \ldots s^{r_n} x_n)$ is a family of dilations on $\mathbb{R}^n$. If the system is homogeneous such that $f_1 \in \underline{n}_{-1}$ and $f_0 \in \underline{n}_{-\theta}$ for some $\theta \in [0,1]$ then*

$$x(s^\theta T, u_{s^{1-\theta}, s^\theta}) = \Delta_s(x(T,u)) \;\; \textit{for all} \;\; s \in [0,1]. \tag{42}$$

Of particular importance are the special cases $\theta = 0$ and $\theta = 1$ which yield, respectively:

$$\text{if } f_0 \in \underline{n}_0,\ f_1 \in \underline{n}_{-1} \;\text{ then }\; x(T, u_{\varepsilon,1}) = \Delta_\varepsilon(x(T,u)) \;\text{ for all }\; \varepsilon \in [0,1] \tag{43}$$

$$\text{if } f_0, f_1 \in \underline{n}_{-1} \qquad \text{then }\; x(\delta T, u_{1,\delta}) = \Delta_\delta(x(T,u)) \;\text{ for all }\; \delta \in [0,1] \tag{44}$$

Many control systems, especially *free nilpotent* systems, admit several different dilations so that with respect to one dilation one may e.g. have $f_0 \in \underline{n}_0$ while w.r.t. another dilation one has $f_0 \in \underline{n}_{-1}$. In such case one may directly use separate scalings for time and size:

$$x(\delta T, u_{\varepsilon,\delta}) = \Delta^{(1)}_{\varepsilon}(\Delta^{(2)}_{\delta}(x(T,u))) \text{ for all } \varepsilon, \delta \in [0,1]. \tag{45}$$

A simple proof uses uniqueness of solutions of initial value problems, showing that both the right and left hand side of (45) are solutions of the same dynamical system, see e.g. [32].
This proposition is at the heart of many classical sufficient conditions for STLC as it basically allows one to construct control variations that will generate a specific tangent vector to the reachable sets, and which in some sense singles out the lowest order term or bracket according to some weighting scheme. The classical needle variations are built around arguments involving basically the dilations $\Delta_{1,\delta}$ (i.e. $\theta = 1$), while a Taylor expansion in the control sizes, and Hermes' sufficient condition is built around the dilation $\Delta_{\varepsilon,1}$ (i.e. $\theta = 0$). Sussmann's general sufficient condition allows a trade-off between the time-scale and amplitude.

Exercise 2.5 *If possible find a one-parameter family of dilations so that the following system, considered by Jakubczyk in the 1970s, is homogeneous. Find all values of* $\frac{r_2 - r_1}{r_1}$, *or of "*θ*" for which the term* x_1^3 *is of lower order than the definite term* x_2^2 *(which appears a potential obstruction to STLC) (compare theorem* 1.10*). Also, compute all nonzero Lie brackets of the vector fields* f_0 *and* f_1 *defining this system.*

$$\left\{ \begin{array}{lcll} \dot{x}_1 & = & u & |u(\cdot)| \leq c \\ \dot{x}_2 & = & x_1 & x(0) = 0 \\ \dot{x}_3 & = & x_2^2 + x_1^3 & \end{array} \right. \tag{46}$$

2.3 Nilpotent approximating systems

When a nonlinear control system of form (8) is controllable by virtue of the linear condition (theorem (1.6)), then it makes sense for many applications (that involve only/primarily the local behaviour near the equilibrium) to approximate the system (8) by the linear system $\dot{x} = Ax + Bu$ where $A = (Df_0)(0)$ equals the Jacobian matrix of partial derivatives if the drift f_0, and where the i-th column of B equals the value of $f_i(0)$, $i = 1, \ldots, m$. (Of course, this can be (but rarely is) formulated in a coordinate-free geometric way that does not mix up the state space and its tangent spaces.)

Exercise 2.6 *Calculate the standard linearized systems for the models* (9) *of* a car/bicycle (example (1.1)) and for the models (5) and (6) for the dynamics of a rolling penny (example (1.2)). Discuss the (linear) controllability properties of the linearized systems, and contrast these with the earlier findings from the first sections.

The exercises make it clear that for some nonlinear systems that are reasonably *"realistic"* the standard linearization causes a dramatic loss of information. Thus one asks for alternatives: Reasonable demands are that the approximating systems are elements of a reasonably rich class of systems that allows for the preservation of controllability or stabilizability properties, that systems in this class are amenable to reasonable analysis and computation, and that the approximation is algorithmic and allows for explicit computation. At this time no such ideal approximating scheme is known – the main culprit being the lack of conditions for STLC that are both necessary and sufficient. However, a very good solution is known that preserves STLC for virtually all systems that are known to be STLC by virtue of Sussmann's general sufficiency condition, theorem 1.10. However, there exist STLC systems for which the standard algorithms yield a nilpotent approximating system that is not STLC. But such systems are considered to be quite exotic – the most simple case is the system (54) considered at the end of this section.

We give a crude outline of an algorithm attributed to Hermes (compare the review [23]) (very similar constructions were employed at almost the same time by Stefani and others), omitting some technical steps that are not central and not essential here. See the review [23], or the original references for more details, especially Stefani [59] for details about adapted charts.) Assuming that the original systems of form (8) is STLC by virtue of theorem 1.10, the objective of this procedure is to construct a nilpotent approximating systems, on the *same* state space $\mathbb{R}^n$, of the same form

$$\dot{y}(t) = g_0(y) + \sum_{i=1}^{m} u_i(t)\; g_i(y) \tag{47}$$

(together with coordinates $y_1, \ldots, y_n$) such that not only $L(g_0, g_1, \ldots, g_n)$ is nilpotent, but so that in addition the vector fields g_j are polynomial and (their Jacobian matrices of partial derivatives w.r.t. y_j are) strictly lower triangular. Recall, that for any such system the solution curves for any given function $u(t)$ are obtained explicitly via simple quadratures (no solution of

nonlinear differential equations is needed). Thus, one considers *nilpotent approximations* as the natural nonlinear analogue of linearizations for systems that exhibit truly nonlinear behaviour, i.e. are more than just nonlinear perturbations of linearly controllable systems.
Start with calculating iterated Lie brackets of the vector fields of increasing length until their values at $x_0 = 0$ span the tangent space $T_0\mathbb{R}^n$. If necessary, continue further until brackets are found that *neutralize possible obstructions to STLC* as defined in theorem 1.10 for a suitable weight $\theta \in (0, 1]$. (It may happen that one can choose among different weights, and thus construct many different nilpotent approximating systems.) It is always possible to choose all weights to be rational. Determine the Lie brackets f_{π_i} such that

$$\operatorname{span}\{f_{\pi_1}(0), f_{\pi_2}(0), \ldots, f_{\pi_n}(0)\} = T_0\mathbb{R}^n \tag{48}$$

and they are of lowest possible weight, defined as the weighted sum of θ times the number of factors f_0 plus the number of factors of the controlled fields f_i, $i \geq 1$ in f_{π_i}. (This is very sloppy, see the discussion of *formal brackets* in the next section.) Define the exponents r_i to equal these weighted sums. If necessary, perform a linear coordinate change such that $f_{\pi_1}(0) = \frac{\partial}{\partial x_i}$ for $i = 1, 2, \ldots n$. Commonly one thrives to have $1 \leq r_1 \leq r_2 \leq \ldots, r_n$. (If homogeneity of the new vector fields is needed, e.g. as for feedback stabilization techniques, a strictly triangular polynomial coordinate change may have to be performed, see [59] for *"adapted charts".*) Using the new coordinates, again called $(x_1, \ldots, x_n)$, define a group of dilations by $\Delta_s(x) = (s^{r_1}x_1, \ldots, s^{r_n}x_n)$.
Expand each component $f_i x_j$ in a Taylor series in the new coordinates, and truncate each expansion keeping only polynomials $p_{ij}(x)$ of order less or equal to $r_j - 1$ for $i \geq 1$, and $r_j - \theta$ for $i = 0$. Define the vector fields $g_j = \sum_{j=1}^n p_{ij}(x)\frac{\partial}{\partial x_j}$. These are easily checked to be (sums of) homogeneous vector fields of negative degree of homogeneity and thus, they generate a nilpotent Lie algebra. The preservation of STLC properties follows from the observation that if g_σ is an iterated Lie bracket of the g_i, and f_σ is the corresponding bracket of the f_i, then their components $g_i x_j$ and $f_i x_j$ agree up to a well-defined degree, and in particular,

$$f_{\pi_i}(0) = g_{\pi_i}(0) \ \text{ for all } \ i = 1, \ldots n. \tag{49}$$

Note that this is only a rough outline of the procedure as a precise description requires a few more technical details and symbols. See the original references of the survey [23] for details.

For illustration consider the model (9) of a car/bicycle (example 1.1). Recall:

$$f_0(x) = \begin{pmatrix} 0 \\ 0 \\ x_2 \cos x_4 \\ x_2 \tan x_1 \\ x_2 \sin x_4 \end{pmatrix}, \qquad f_1(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } \qquad f_2(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{50}$$

One readily computes $[f_1, f_2] \equiv 0$. Selected other brackets are:

$$[f_1, f_0] = \begin{pmatrix} 0 \\ 0 \\ 0 \\ x_2 \sec^2 x_1 \\ 0 \end{pmatrix}, \ [f_2, f_0] = \begin{pmatrix} 0 \\ 0 \\ \cos x_4 \\ \tan x_1 \\ \sin x_4 \end{pmatrix}, \ [[f_0, f_1], f_2] = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sec^2 x_1 \\ 0 \end{pmatrix},$$

$$\text{and finally} \qquad [[f_0, f_2], [[f_0, f_1], f_2]] = \begin{pmatrix} 0 \\ 0 \\ -\sec^2 x_1 \sin x_4 \\ 0 \\ \sec^2 x_1 \cos x_4 \end{pmatrix}. \tag{51}$$

In principle there is a large number of other brackets that should be calculated, too. However, advanced knowledge from the next lectures (Hall bases) allow one to calculate only a minimal number of brackets. And once Sussmann's theorem 1.10 applies one always can stop. Note that at the origin these vector fields have the values:

$$\begin{aligned} f_1(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ f_2(0) &= \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ [f_2, f_0](0) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \\ f_{\pi_4}(0) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \ f_{\pi_5}(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \end{aligned} \tag{52}$$

where $f_{\pi_4} = [[f_0, f_1], f_2]$ and $f_{\pi_5} = [[f_0, f_2], [[f_0, f_1], f_2]]$. These iterated brackets span the tangent space at the origin, thereby guaranteeing accessibility. Clearly the system is not linearly controllable (it does not satisfy the conditions in theorem 1.6).

Exercise 2.7 *Explain why no matter how many brackets one uses that contain any number of factors* f_0*, but only a single factor* f_1 *or* f_2*, their values at* 0 *will never span* $T_0\mathbf{R}^5$*.*

While technically one needs to verify that indeed no *lower order possible obstructions* are nonzero at 0, it is quite apparent that no surprises can happen. (For a rigorous argument, use Hall bases from the next section, and check *ALL* brackets of length at most 5 that appear in such a basis.) Define $f_{\pi_1} = f_1$, $f_{\pi_2} = f_2$, and $f_{\pi_3} = [f_0, f_2]$.
As no *potential obstructions to STLC had to be neutralized*, we are free to choose any weight $\theta \in (0, 1]$, e.g. $\theta = 1$. Thus the weight of each of the five selected brackets agrees with its length (see next chapter for more precise language), and we obtain $r = (1, 1, 2, 3, 5)$. There is no need to perform any linear coordinate change as already $f_{\pi_i}(0) = \frac{\partial}{\partial x_i}\Big|_0$ for $i = 1, 2, 3, 4, 5$.
Expanding the components of $f_i x_j$ into Taylor series and keeping in each component $f_j x_i$ only the terms Δ-lowest term of degree no larger than $r_i - 1$ (and for f_0, in general, no larger than $r_i - \theta$) one obtains the approximating fields

$$g_0(x) = \begin{pmatrix} 0 \\ 0 \\ x_2 \\ x_1 x_2 \\ x_2 x_4 \end{pmatrix}, \; g_1(x) = f_1(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } g_2(x) = f_2(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{53}$$

Exercise 2.8 *Verify directly, i.e. using the theorem* 1.10 *that this nilpotent approximating system* (53) is indeed STLC about 0. Moreover verify that the corresponding brackets f_{π_j} and g_{π_j} have the same values at 0.

Exercise 2.9 *Give a (counter)example of a system that illustrates that the choice of the weight* $\theta = 0$ *may yield an approximating system that is not necessarily nilpotent. (Remark: However, the Lie algebra will be solvable, and thus still allow for a choice of coordinates in which the approximating vector fields are polynomial and triangular, thus allowing for still comparatively simple calculations of trajectories, compare Crouch* [8]*).*

Exercise 2.10 *Calculate an STLC nilpotent approximating systems for the models* (5) and (6) for the dynamics of a rolling penny (example (1.2)).

Finally consider the following system

$$\left\{\begin{array}{lcllcl} \dot{x_1} & = & u & x(0) & = & 0 \\ \dot{x_2} & = & x_1 & |u(\cdot)| & \leq & \varepsilon_0 \\ \dot{x_3} & = & x_1^3 & & & \\ \dot{x_4} & = & x_3^2 + x_2^7 & & & \end{array}\right. \tag{54}$$

which has been shown to be STLC in [30]. However, for every weighting $\theta \in [0,1]$ of $\varepsilon = s^{1-\theta}$ and $\delta = s^\theta$, the definite term $\int_0^{\delta T} x_3^2(t, u_{\varepsilon,\delta})dt = \varepsilon^6\delta^9 \int_0^T x_3^2(t, u_{1,1})dt$ is of lower order in s than the term $\int_0^{\delta T} x_2^7(t, u_{\varepsilon,\delta})dt = \varepsilon^7\delta^{15} \int_0^T x_2^7(t, u_{1,1})dt$ which provides controllability! As a consequence, none of the traditional control variations can be used to generate $-\left.\frac{\partial}{\partial x_4}\right|_0$ as a tangent vector to the reachable sets in order to conclude STLC, and different kinds of families of control variations were invented [30].

Exercise 2.11 *Calculate* all *iterated Lie brackets for the fields in system* (54) *that are nonzero at* 0 *and* recover *the scaling exponents* $(6,9)$ *and* $(7,15)$. *Verify that for no choice of* $\theta \in (0,1]$ *the system* (54) *satisfies Sussmann's sufficient conditions in theorem* 1.10.

3 Combinatorics of words and free Lie algebras

3.1 Intro: Trying to partially factor the Chen Fliess series

This section shall serve as the final motivation to get rid of all excessive symbols, such as iterated integrals, when facing either large computations or deeper theoretical analysis. While the sample calculations may appear rather simple and naive, past experience shows that for many a reader of the subsequent abstract material, they are an essential guide that connects the combinatorial structures with control.

Consider a single input system of form (8), i.e. with $m = 1$ and $u_0 \equiv 1$, $u = u$. Write out the first few terms in the Chen Fliess series (34)

$$\begin{aligned} S_{CF,f}(T,u)(\phi) = 1\cdot\phi(0) + \int_0^T 1dt\cdot(f_0\phi)(0) + \int_0^T u(t)dt\cdot(f_1\phi)(0) \\ + \int_0^T\int_0^{t_2} 1{\cdot}1dt_1dt_2{\cdot}(f_0f_0\phi)(0) + \int_0^T\int_0^{t_2} u(t_2)u(t_1)dt_1dt_2{\cdot}(f_1f_1\phi)(0) \\ + \int_0^T\int_0^{t_2} u(t_2){\cdot}1\,dt_1dt_2{\cdot}(f_0f_1\phi)(0) + \int_0^T\int_0^{t_2} 1{\cdot}u(t_1)dt_1dt_2{\cdot}(f_1f_0\phi)(0) \end{aligned} \tag{55}$$

$$
\begin{aligned}
&+\int_0^T\int_0^{t_3}\int_0^{t_2} 1\cdot 1\cdot 1 dt_1 dt_2 dt_3\cdot(f_0 f_0 f_0\phi)(0)\\
&+\int_0^T\int_0^{t_3}\int_0^{t_2} u(t_3)\cdot 1\cdot 1 dt_1 dt_2 dt_3\cdot(f_0 f_0 f_1\phi)(0)\\
&+\int_0^T\int_0^{t_3}\int_0^{t_2} 1\cdot u(t_2)\cdot 1 dt_1 dt_2 dt_3\cdot(f_0 f_1 f_0\phi)(0)\\
&+\int_0^T\int_0^{t_3}\int_0^{t_2} 1\cdot 1\cdot u(t_1) dt_1 dt_2 dt_3\cdot(f_1 f_0 f_0\phi)(0)\\
&+\int_0^T\int_0^{t_3}\int_0^{t_2} u(t_3)\cdot u(t_2)\cdot 1 dt_1 dt_2 dt_3\cdot(f_0 f_1 f_1\phi)(0)\\
&+\int_0^T\int_0^{t_3}\int_0^{t_2} u(t_3)\cdot 1\cdot u(t_1) dt_1 dt_2 dt_3\cdot(f_1 f_0 f_1\phi)(0)\\
&+\int_0^T\int_0^{t_3}\int_0^{t_2} 1\cdot u(t_2) u(t_1) dt_1 dt_2 dt_3\cdot(f_1 f_1 f_0\phi)(0)\\
&+\int_0^T\int_0^{t_3}\int_0^{t_2} u(t_3) u(t_2) u(t_1) dt_1 dt_2 dt_3\cdot(f_1 f_1 f_1\phi)(0)\\
&+\ \text{higher order terms}
\end{aligned}
$$

This is just the beginning, and one never should manually manipulate such a huge expression. Indeed, each of the summands is identified by a simple *word* such as 101 or 10 (to be read as finite sequence, like $(1,0,1)$ or $(1,0)$. The *identification* is captured in form of the two maps

$$
\mathcal{F}\colon w = a_1 a_2 \dots a_s \mapsto \Big(\phi \mapsto (f_w\phi)(0) = (f_{a_1}\circ f_{a_2}\circ\cdots f_{a_s}\phi)(0)\Big), \quad \text{and} \tag{56}
$$

$$
\Upsilon\colon w = a_1 a_2 \dots a_s \mapsto \left(u \mapsto \int_0^T u_{a_s}(t_s)\int_0^{t_s}\!\!\dots\int_0^{t_2} u_{a_1}(t_1)dt_1\dots dt_{s-1}dt_s\right) \tag{57}
$$

These two maps take the advanced point of view that each image is itself an operator: In the first case the image is a partial differential operator on (output) functions on the state space. In the second case, the image is an *iterated integral functional* on the space of admissible controls on an interval

$[0, T]$. For later convenience we already define a companion map Υ in terms of the primitives $U(t) = \int_0^t u(s)ds$ of the usual controls.

$$\Upsilon\colon w = a_1 a_2 \ldots a_s \mapsto \left(U \mapsto \int_0^T U'_{a_s}(t_s) \int_0^{t_s} \ldots \int_0^{t_2} U'_{a_1}(t_1) dt_1 \ldots dt_{s-1} dt_s \right) \tag{58}$$

It is well known that there are many ways to rewrite the huge expression of the Chen Fliess series, ways which are better in the sense of both providing much more insight for theoretical analysis and for being much more amenable for calculation and design (such as path planning). Such alternative forms may be obtained through direct simultaneous manipulation of the analytical objects on right hand sides of (56) and (57), or alternatively through purely algebraic and combinatorial manipulation of the combinatorial objects on the left hand side of (56) and (57).

For illustration, we shall perform some of the analytic operations for a typical objective on some of the low order terms written out above. Then we will repeat the same working only with the indices w. This hopefully will lead even the last skeptics to look positively on combinatorics, and it will motivate the *chronological algebra* structure which makes Υ a *chronological algebra homomorphism.*

One reasonable question to ask in view of this series, and in view of the ubiquitous presence of iterated Lie brackets (and their important geometric roles) in nonlinear control, as exhibited in the previous section, is: *"Where are the Lie brackets in the Chen Fliess series"* (or in above big expression (3.1)). The previous chapters analyzed systems using almost exclusively vector fields which are first order derivatives (all Lie brackets are vector fields!), whereas above formula contains primarily partial differential operators of arbitrarily high order!

Let us consider the terms containing one f_0 and one f_1, followed by looking at the terms containing one f_0 and two f_1. In particular, noting that $[f_1, f_0]\phi = f_1 f_0 \phi - f_0 f_1 \phi$, we add and subtract the following term (which does not appear in the series!) (alternative choices are possible)

$$\int_0^T \int_0^{t_2} 1 \cdot u(t_1) dt_1 dt_2 \cdot (f_0 f_1 \phi)(0)$$

then combine the results appropriately (alternatively integrate by parts)

$$\int_0^T \int_0^{t_2} u(t_2) \cdot 1 \, dt_1 dt_2 \cdot (f_0 f_1 \phi)(0) + \int_0^T \int_0^{t_2} 1 \cdot u(t_1) \, dt_1 dt_2 \cdot (f_1 f_0 \phi)(0) =$$

$$\begin{aligned}
&= \left(\int_0^T u(t_2) \int_0^{t_2} 1 \, dt_1 dt_2 + \int_0^T 1 \cdot \int_0^{t_2} u(t_1) \, dt_1 dt_2 \right) \cdot (f_0 f_1 \phi)(0) \\
&\qquad\qquad\qquad + \int_0^T \int_0^{t_2} 1 \cdot u(t_1) dt_1 dt_2 \cdot ((f_1 f_0 - f_0 f_1)\phi)(0) \\
&= \left(\int_0^T u(t) dt \right) \cdot \left(\int_0^T 1 \, dt \right) \cdot (f_0 f_1 \phi)(0) \\
&\qquad\qquad\qquad + \left(\int_0^T 1 \cdot \left(\int_0^{t_2} u(t_1) \, dt_1 \right) dt_2 \right) \cdot ([f_1, f_0]\phi)(0)
\end{aligned} \tag{59}$$

An important observation is that above sum of two second order partial derivatives with iterated integral coefficients is now expressed as a sum of one first order derivative with an iterated integral coefficient and a second order partial derivative with a product of integrals as coefficient.

For comparison let us write down the bare essentials to code all the terms in above calculation.

$$01 \otimes 01 + 10 \otimes 10 = (01 + 10) \otimes 01 + 10 \otimes (10 - 01) = (0 \sqcup\!\sqcup 1) \otimes 01 + 10 \otimes [10]$$

Barely one line, and already providing a preview of a product on *words* that will encode the pointwise multiplication of functions of a single variable, or of iterated integral functionals. This *shuffle product* shall be studied formally in subsequent sections.

Now consider the third order terms that contain exactly two factors of f_1 and one f_0. This time strategically integrate by parts repeatedly, instead of judiciously adding and subtracting terms. This has the same effect, and illustrates the duality. (The first approach, which focused on the vector fields as opposed to the integrals, though, appears to be closer to the technique of *rewriting systems* of algebraic combinatorics, compare [48] and [54]. (Caveat: The following might be done a little faster, but in the end one should always use the algebra, instead of trying to improve the lengthy integrations by parts.) Start with integrating by parts the inside integral in the first term

$$\begin{aligned}
&\int_0^T u(t_3) \int_0^{t_3} u(t_2) \int_0^{t_2} 1 \, dt_1 dt_2 dt_3 \cdot (f_0 f_1 f_1 \phi)(0) \\
&\qquad + \int_0^T u(t_3) \int_0^{t_3} 1 \int_0^{t_2} u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_0 f_1 \phi)(0)
\end{aligned}$$

$$+\int_0^T 1\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)dt_1dt_2dt_3\cdot(f_1f_1f_0\phi)(0)$$

$$=\left(\int_0^T u(t_3)\left(\cdot\left(\int_0^{t_3} u(t_2)dt_2\right)\cdot\left(\int_0^{t_3} 1dt_2\right)-\int_0^{t_3} 1\int_0^{t_2} u(t_1)dt_1dt_2\right)dt_3\right)\cdot(f_0f_1f_1\phi)(0)$$

$$+\int_0^T u(t_3)\int_0^{t_3} 1\int_0^{t_2} u(t_1)\,dt_1dt_2dt_3\cdot(f_1f_0f_1\phi)(0)$$

$$+\int_0^T 1\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)\,dt_1dt_2dt_3\cdot(f_1f_1f_0\phi)(0)$$

After suitably regrouping the first term again integrate by parts, and combine the second and third term, recognizing that $f_0f_1f_1 - f_0f_1f_1 = [f_1, f_0]f_1$

$$=\left(\left(\int_0^T 1\,dt\right)\cdot\left(\int_0^T u(t_3)\int_0^{t_3} u(t_2)dt_2dt_3\right)-\int_0^T 1\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)dt_1dt_2dt_3\right)\cdot(f_0f_1f_1\phi)(0)$$

$$+\left(\int_0^T\left(u(t_3)\int_0^{t_3} 1\int_0^{t_2} u(t_1)\,dt_1dt_2\right)dt_3\right)\cdot([f_1,f_0]f_1\phi)(0)$$

$$+\int_0^T 1\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)dt_1dt_2dt_3\cdot(f_1f_1f_0\phi)(0)$$

Combine the second and fourth terms, and integrate the third term by parts (outer integral). Also write the first term as a product of three integrals.

$$=\tfrac{1}{2}\cdot\left(\int_0^T 1\,dt\right)\cdot\left(\int_0^T u(t)\,dt\right)^2\cdot(f_0f_1f_1\phi)(0)$$

$$+\int_0^T 1\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)\,dt_1dt_2dt_3\cdot(f_1f_1f_0-f_0f_1f_1\phi)(0)$$

$$+\left(\left(\int_0^T u(t)dt\right)\left(\int_0^T 1\int_0^{t_3} u(t_2)dt_2dt_3\right)-\int_0^T 1\cdot\left(\int_0^{t_3} u(t_2)dt_2\right)^2 dt_3\right)\cdot([f_1,f_0]f_1\phi)(0)$$

Finally combine the second and fourth terms, recognizing that the integral in the fourth term is twice the integral in the second term.

$$=\tfrac{1}{2}\cdot\left(\int_0^T 1dt\right)\cdot\left(\int_0^T u(t)dt\right)^2\cdot(f_0f_1f_1\phi)(0)$$

$$+\left(\int_0^T u(t)dt\right)\cdot\left(\int_0^T 1\int_0^{t_3} u(t_2)dt_2dt_3\right)\cdot([f_1,f_0]f_1\phi)(0)$$

$$+\left(\int_0^T 1\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)dt_1dt_2dt_3\right)\cdot[f_1,[f_1,f_0]]\phi)(0)\ .$$

The last step used that

$$f_1 f_1 f_0 - f_0 f_1 f_1 - 2[f_1, f_0] f_1 = f_1 f_1 f_0 - 2 f_1 f_0 f_1 + f_0 f_1 f_1 = [f_1, [f_1, f_0]]. \quad (60)$$

What matters, aside from experiencing the painful book-keeping, is that again the three third-order partial derivatives with iterated integral coefficients of the original series can be written as a sum of a first, a second order and third order partial derivative, with corresponding products of iterated integrals as coefficients. The emerging pattern is very suggestive. However, this naive approach of repeatedly integrating by parts is no way to deal with the infinite series.

To illustrate the usefulness of this expression, suppose that $f(0) = 0$ and ϕ is a function such that $f_1\phi \equiv 0$ and $[f_1, f_0]\phi(0) = 0$ (this is very similar to the examples discussed in the first chapter). In this case the *leading term* in the rewritten Chen Fliess series (assuming that similar calculations to above have been carried out with analogous results for the other *homogeneous components*) is the last term in the result of our previous calculation. If the iterated integrals corresponding to the *words* 1 and 10 both vanish (by, say, a judicious choice of a piecewise constant control), then again the lowest order nonvanishing term is the last term in our result. Note, in the first argument we used the *product structure* of the partial differential operators (e.g. if $f_1\phi \equiv 0$ then $f_\pi f_1 \phi \equiv 0$ for *every* partial differential operator f_π). In the second argument we used the product structure of the rewritten iterated integrals that appear as coefficients of the non-first order operators. Clearly, there are lots of opportunities to combine these arguments, and indeed this is a route towards obtaining conditions for STLC and for optimality!

It turns out that the *expected* result is true, and even more: The entire series can be written as a product of nice flows (of constant vector fields!), or as the exponential of a single field. A partial factorization was used for obtaining a new necessary condition for STLC in [28], but it was clear that this is not the way to go. In [64] Sussmann managed to factor the entire series using differential equations techniques, compare section 4.3. An elegant alternative is to do away with all integrals and such, and proceed purely combinatorially, which allows one to focus on the underlying algebraic structure.

We conclude this last motivation for *combinatorics of words* with the combinatorial analogue of above calculation:

$$011 \otimes 011 + 101 \otimes 101 + 110 \otimes 110 =$$

$$= ((011 + \overbrace{101) - 101}) \otimes 011 + 101 \otimes 101 + 110 \otimes 110$$

$$
\begin{aligned}
&= ((011{+}101{+}\overbrace{110) \; - \; 110}) \otimes 011 + 101 \otimes (101{-}011) + 110 \otimes 110 \\
&= (011{+}101{+}110) \otimes 011 + ((101{+}\overbrace{2{*}110) - 2{*}110}) \otimes (101{-}011) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \; 110 \; \otimes \; (110{-}011) \\
&= (011{+}101{+}110){\otimes}011 + (101{+}2{*}110){\otimes}(101{-}011) \\
&\qquad\qquad\qquad\qquad\qquad\qquad + \; 110 \otimes ((110{-}011) - 2 * (101 - 011)) \\
&= \tfrac{1}{2}(0 \sqcup\!\sqcup 1 \sqcup\!\sqcup 1) \otimes 011 + (10 \sqcup\!\sqcup 1) \otimes [10]\,1 + 110 \otimes [1[10]]
\end{aligned}
$$

with the last line containing the abbreviated form involving *shuffle products*, see below. At this time the combinatorial *rewriting rules* used here may still look unfamiliar, but they simply code *integration by parts*. The following lectures shall give an introduction into this world of a different algebra. We shall aim first for a formal definition of the product $\sqcup\!\sqcup$ on words that encodes products of iterated integral functionals is needed. Together with a *systematic* choice of bases, it should reduce the above calculations to simply inverting the matrix corresponding to a change of basis in some vector space. Being able to use simple linear algebra, it will turn out to be rather easy to compute a powerful continuous analogue of the Campbell Baker Hausdorff formula [39]

3.2 Combinatorics and various algebras of words

This section provides a very basic introduction to the terminology commonly used in an area of combinatorial algebra commonly known as *combinatorics of words*. For a comprehensive introduction accessible to the non-specialist we refer the reader to consult the book *Combinatorics on words* by *"Lothaire"* [46] with the same title. For a more advanced treatment of many of the objects with applications to nonlinear control, we refer to the book *Free Lie algebras* by Reutenauer [54].

The basic idea from the control-perspective is to directly manipulate the *multi-indices* that appeared in the preceding calculations, rather than carry around the bulky overhead of iterated integrals, control functions and vector fields, when carrying out what effectively are purely algebraic or combinatorial manipulations. Moreover, as indicated previously, there is a need to work with formal brackets as opposed to brackets of vector fields (which are just vector fields, and thus have no numbers of factors etc.). Finally, there are many algebraic theorems and constructions available, starting with constructions of bases and formulas for their dual bases, that are very useful on control.

Start with a set Z whose elements are in one-to-one correspondence with the vector fields $f_0, f_1, \dots f_m$ and with the controls $u_0 \equiv 1, u_1, \dots u_m$. Occasionally it is convenient to simply use the indices $Z = \{0, 1, 2, \dots, m\} \subseteq \mathbf{Z}_0^+$ considered as formal symbols (not as integers). In general $Z = \{X_0, X_1, \dots X_m\}$ is just a set of *indeterminates* X_i. In the sequel we shall refer to this set as an *alphabet*, and to its elements as *letters*. In principle, this set can be infinite, but for most of our purposes finite sets suffice (Lazard elimination in chapter 4 is an exception). A *word* is a finite sequence $(a_1, \dots a_s)$ with $a_i \in Z$ and $s \in \mathbf{Z}_0^+$. It is customary to write $a_1a_2a_3 \dots a_s$ for the sequence $(a_1, \dots a_s)$, to use $a, b, c, \dots$ for letters in Z while u, v, w, z for words, to write e or 1 for the *empty word* defined by $we = ew = w$ for all words w. Write Z^+ for the set of all nonempty words and $Z^* = Z^+ \cup \{e\}$ for the set of all words. The set Z^* of all words forms a free monoid (semigroup) (associative, but noncommutative) under the concatenation product

$$(a_1a_2 \dots a_s \ , \ b_1b_2 \dots b_r) \mapsto a_1a_2 \dots a_sb_1b_2 \dots b_r \tag{61}$$

From the control perspective, on the side of the vector fields, this concatenation product clearly just corresponds (via the map $\mathcal{F}$ in (56)) to compositions of partial differential operators. But on the control and iterated integrals side, via the map Υ from (58) (or $\Upsilon\!\!\Upsilon$ from (57)) it is much more interesting as a product $\Upsilon(w)\Upsilon(z)$ of iterated integrals in the form special form as they arose in the derivation (31) of the Chen Fliess series is *not* an iterated integral of the same form – although, conceivably, through laborious repeated integration by parts, it can be written as a linear combination of iterated integrals in that special form. One of the purposes of this section is to take care of that kind of manipulation once for all!

As linear combinations are clearly needed, we consider the *free associative algebra* $A(Z) = A_{\mathbf{R}}(Z)$ of all finite linear combinations (with real coefficients) of words in Z^*, and linearly extending the concatenation product in the obvious way. (This algebra is also known as the *algebra of polynomials in noncommuting variables.*) Write $\hat{A}(Z) = \hat{A}_{\mathbf{R}}(Z)$ for the algebra of formal power series over Z (with the same concatenation product).

Define the Lie bracket as the bilinear map $[\cdot,\cdot]\colon A(Z) \times A(Z) \mapsto A(Z)$ that satisfies $[w, z] = wz - zw$ for words $w, z \in Z^*$.

Exercise 3.1 *Verify that is* $(A, \circ)$ *is any associative algebra, then the commutator* $[\cdot,\cdot]\colon A \times A \mapsto A$ *defined by* $[x, y] = x \circ y - y \circ x$ *satisfies the Jacobi identity.*

An element of $A(Z)$ is called a Lie polynomial if it lies in the smallest subspace of $A(Z)$ that contains Z and that is closed under the Lie bracket. It is nontrivial, requiirng one consequence of the Poincaré-Birkhoff-Witttheorem (4.6) (compare [54]) to show that this subspace of Lie polynomials (with the Lie bracket as above) is *the free Lie algebra* over Z, denoted $L(Z)$.

The next section will address the quest for bases of the free Lie algebra $L(Z)$. Under the natural extension of the map $\mathcal{F}$ in (56) to $A(Z)$ and thus to $L(Z)$, any such basis maps to a spanning set of $L(f_0, f_1, \ldots f_m)$, i.e. provides a *minimal* set of Lie brackets to calculated / considered in control applications.

Next we try to distill the essence of the algebraic structure of the iterated integrals in (31). After little reflection it is clear that the construction of iterating the integral form of the differential equation invariably leads one to the noncommutative product

$$* : (U(t),\ V(t)) \mapsto (U * V)(t) = \int_0^t U(s)V'(s)ds \qquad \left(= \int_0^t V'(s)U(s)ds \right). \tag{62}$$

(or to its mirror image). In (31) consider for example

$$U(t) = \int_0^t u_{i_3}(t_3) \int_0^{t_3} u_{i_2}(t_2) \int_0^{t_2} u_{i_1}(t_1)dt_1dt_2dt_3 \text{ and } V(t) = \int_0^t u_{i_4}(s)ds \tag{63}$$

and their *chronological product*

$$(U * V)(t) = \int_0^t u_{i_4}(t_4) \int_0^{t_4} u_{i_3}(t_3) \int_0^{t_3} u_{i_2}(t_2) \int_0^{t_2} u_{i_1}(t_1)dt_1dt_2dt_3dt_4 \tag{64}$$

We shall quickly identify the *defining identity* satisfied by this product, and then equip the free associative algebra $A(Z)$ with an analogous product, so that the map Υ, linearly extended to $A(Z)$ will be an homomorphism for the resulting algebra structure.

Looking for a three term identity (analogous to associativity or the Jacobi identity) that might possibly characterize this algebra structure, consider the products (of say, absolutely continuous functions) $f, g, h\colon \mathbb{R} \mapsto \mathbb{R}$ taken in different orders:

$$(f * (g * h))(t) = \int_0^t f(s) \cdot g(s) \cdot h'(s)\, ds \quad \text{and} \tag{65}$$

$$(f * g) * h)(t) = \int_0^t \left(\int_0^s f(\sigma) \cdot g'(\sigma) d\sigma \right) h'(s)\, ds \tag{66}$$

This reminds (with good reason) of the laborious integrations by parts in the previous chapter. Indeed, it is almost immediately apparent that this product satisfies, for all (absolutely continuous) functions $f, g, h \colon \mathbb{R} \mapsto \mathbb{R}$ (that vanish at 0) the three term *right chronological identity*

$$f * (g * h) = (f * g) * h + (g * f) * h \tag{67}$$

Definition 3.1 *A (right) chronological algebra is a vector space C with a bilinear product $* \colon C \times C \mapsto C$ that satisfies the* right chronological identity (67) *for all $fg, h \in C$.*

The naturalness and usefulness of this algebra structure for nonlinear control, as well as its natural appearance as the *(Koszul-)dual structure* to that of *Leibniz algebras* which recently have received much attention by algebraists [12, 13, 20, 42, 43, 44, 45, 55], suggest that one study this algebra structure in its own right, just like associative, commutative, and Lie algebras. Refer to [35] for some more abstract investigations. In these notes we shall basically just *use* this product.

Exercise 3.2 *Let $V = L^1_{\text{loc}}(\mathbb{R})$ be the space of locally integrable functions on $\mathbb{R}$ and define Verify that $(V, \star)$ is a chronological algebra with the product $\star; \colon V \times V \mapsto V$ defined by*

$$(f \star g)(t) = \left(\int_0^t f(s) ds \right) \cdot g(t) \tag{68}$$

There are many interesting chronological subalgebras of the algebras $AC_0(\mathbb{R})$ of absolutely continuous functions that vanish at 0, and of the algebra L^1_{loc} locally integrable functions that deserve attention in their own right. The following exercise gives further examples that open new doors for constructing further chronological algebras.

Exercise 3.3 *Verify directly that each of the products on polynomials and exponentials defined below is a right chronological product.*

$$\begin{array}{rclcrcl} X^k \star X^\ell & = & \frac{1}{k+1} X^{k+\ell+1} & \qquad & e^{ikt} \star e^{i\ell t} & = & \frac{(-i)}{k} e^{i(k+\ell)t} \\ X^k * X^\ell & = & \frac{\ell}{k+\ell} X^{k+\ell} & & e^{ikt} * e^{i\ell t} & = & \frac{\ell}{k+\ell} e^{i(k+\ell)t} \end{array} \tag{69}$$

It is not surprising that one can make sense of a *free chronological algebra* $C(Z)$ and even construct it from the free associative algebra $A(Z)$ by defining a *chronological product of words* in terms of the concatenation product. For any letter $a \in Z$, and words *nonempty* $w, z \in Z^+$ define inductively (on the length of the second word)

$$w * a \overset{\text{def}}{=} wa \qquad \text{and} \qquad w * (za) = (w * z + z * w)a \tag{70}$$

and extend bi-linearly to the subspace $A^+(Z)$ of $A(Z)$ that is spanned by all nonempty words. Note, it is impossible to extend the definition to all of $A(Z) \times A(Z)$ without loosing some key properties. (However, for some purposes it will be convenient to allow *one* factor to be the empty word e and set $w * e = 0$ and $e * w = w$ if $w \in Z^* \setminus \{e\}$.) With these definitions it is apparent that the following holds:

Theorem 3.1 *The map Υ from $C(Z)$ to a chronological algebra of iterated integral functionals $\mathcal{IIF}(\mathcal{U})$ is a chronological algebra homomorphism, i.e. for any $w, z \in C(Z)$*

$$\Upsilon(w * z) = \Upsilon(w) * \Upsilon(z) \tag{71}$$

For details on suitable domains $\mathcal{U}$ of the iterated integral functionals see [35]. Here we only note that any such set of functionals immediately inherits the chronological algebra structure from its domain, e.g. the set of absolutely continuous functions. One can show that the map Υ is actually is a chronological algebra *isomorphism* provided the class of admissible inputs is *sufficiently rich*, compare [35]. We sketch the key *inductive step* for nonempty words $w, z \in Z^*$, a letter $a \in Z$, $U \in AC_0([0, T]$ and $T \geq 0$, written out in mini-steps:

$$
\begin{aligned}
\Upsilon_{w*(za)}(U)(T) &= \Upsilon_{(w*z+z*w)a}(U)(T) \\
&= \int_0^T U_a'(t)\cdot\Upsilon_{w*z+z*w}(U)(t)dt \\
&= \int_0^T U_a(t)\cdot\left(\Upsilon_{w*z}(U)(t)+\Upsilon_{z*w}(U)(t)\right)dt \qquad (72)\\
&= \int_0^T U_a'(t)\cdot\Big(\int_0^t(\Upsilon_z(U))'\cdot\Upsilon_w(U)(s)ds+\int_0^t(\Upsilon_w(U))'\cdot\Upsilon_z(U)(s)ds\Big)dt \\
&= \int_0^T U_a'(t)\Upsilon_z(U)(t)\cdot\Upsilon_w(U)(t)dt \\
&= \int_0^T \frac{d}{dt}\Big(\int_0^t U_a'(s)\Upsilon_z(U)(s)ds\Big)\cdot\Upsilon_w(U)(t)dt \\
&= (\Upsilon_w(U)*\Upsilon_{za}(U))(T)
\end{aligned}
$$

The first and last step use the definition (58). In between, aside from using the linearity of Υ and the induction hypothesis, the key steps are integration by parts followed by suitable regrouping – exactly the steps from the section 2.1 that we wanted to combinatorially encode.

The symmetrization of the chronological product of functions (or iterated integral functionals) yields the pointwise multiplication, which is both commutative and associative. Since this product is also routinely used in control, it makes sense to formally define and name the corresponding product on the free associative algebra $C(Z)$, or its extension to the free associative algebra $A(Z)$.

Definition 3.2 *The shuffle product is the bilinear map* $\sqcup\!\sqcup : A(Z)\times A(Z)\mapsto A(Z)$ *that satisfies* $w\sqcup\!\sqcup e = e\sqcup\!\sqcup w = w$ *for all* $w\in A(Z)$ *and*

$$
w \sqcup\!\sqcup z = w*z+z*w \qquad \text{for all } w,z\in C(Z) \tag{73}
$$

Corollary 3.2 *The map* Υ *from* $A(Z,\sqcup\!\sqcup)$ *to a associative algebra of iterated integral functionals* $\mathcal{IIF}(\mathcal{U})$ *(with pointwise mutliplication) is a associative algebra homomorphism.*

$$
\Upsilon(w\sqcup\!\sqcup z)=\Upsilon(w)\cdot\Upsilon(z) \tag{74}
$$

Using the recursive definition (70) of the chronological product one immediately obtains a recursive formula for the shuffle product. For letters $a,b\in Z$

and words $w, z \in Z^*$

$$(wa) ш (zb) = (w ш (zb))a + ((wa) ш z)b \tag{75}$$

Exercise 3.4 *Verify by direct computation, using* (67)*, that the shuffle product is associative.*

Exercise 3.5 *Calculate at least a handful of shuffle products to get a feeling for it. E.g. calculate the following (but feel free to make your own choices)* $a ш b$, $a ш a$, $a ш a ш a$, $a ш a ш b$, $a ш b ш c$, $(ab) ш c$, $(ab) ш b$, $(ab) ш a$, $(ab) ш (cd)$, $(ab) ш (ab)$, *...*

Our definition of the shuffle product as the symmetrized chronological product makes sense from the point of view of nonlinear control – but this does not do it justice as its algebraic characterization shows that it is a fundamental map. First introduce yet another product, actually a *"coproduct"*.

Definition 3.3
Define the coproduct $\Delta\colon A(Z) \mapsto A(Z) \otimes A(Z)$ *as the linear algebra homomorphism defined on generators* $a \in Z$ *by (using* 1 *for the* empty word *(previously denoted by* e*).*

$$\Delta\colon a \mapsto a \otimes 1 + 1 \otimes a \tag{76}$$

Use [54] as a gentle introduction, and source for more advanced references for the realm of coproducts, co-gebras, bi-grebas, and ultimately Hopf-algebras. (They appear to play a quite useful, though still largely unrecognized role in nonlinear control.)
It is instructive to work a few examples. Suppose $a, b \in Z$. Then

$$\begin{array}{rcl} \Delta(ab) & = & \Delta(a)\Delta(b) \\ & = & (a \otimes 1 + 1 \otimes a)(b \otimes 1 + 1 \otimes b) \\ & = & ab \otimes 1 + a \otimes b + b \otimes a + 1 \otimes ab \\ & \neq & ab \otimes 1 + 1 \otimes ab \end{array} \tag{77}$$

By symmetry, it is clear that

$$\Delta([a, b]) = [a, b] \otimes 1 + 1 \otimes 1, b] \tag{78}$$

The previous calculation not only holds for $a, b \in Z$, but in much more generality, it is true for any $p, q \in A(Z)$ provided $\Delta(p) = p \otimes 1 + 1 \otimes p$ and $\Delta(q) = q \otimes 1 + 1 \otimes q$. Thus the set of $p, q \in A(Z)$ for which this holds is, maybe not surprisingly:

Theorem 3.3 *A polynomial $p \in A(Z)$ is a Lie polynomial if and only if*

$$\Delta(p) = p \otimes 1 + 1 \otimes p \tag{79}$$

Exercise 3.6 *Prove theorem* 3.3 *using the characterization that the set of Lie polynomials is the smallest subspace of $A(Z)$ that contains Z and is closed under the Lie bracket.*

Before returning to the shuffle product, we note that one may define an inner product on $A(Z)$ by demanding that the *basis* Z^* of $A(Z)$ is an orthonormal basis, i.e. define $<\cdot,\cdot>: A(Z) \times A(Z) \mapsto \mathbb{R}$ for $P_w, Q_w \in \mathbb{R}$ by

$$<P,Q>=\sum_{w\in Z^*} P_w Q_w \quad \text{if} \quad P=\sum_{w\in Z^*} P_w w, Q=\sum_{w\in Z^*} Q_w w \in A(Z), \tag{80}$$

This map extends immediately to a map $<\cdot,\cdot>: \hat{A}(Z) \times A(Z) \mapsto \mathbb{R}$ (or $<\cdot,\cdot>: A(Z) \times \hat{A}(Z) \mapsto \mathbb{R}$), which then is considered as a *bilinear pairing* upon noting that the algebra $\hat{A}(Z)$ of noncommuting formal power series is the *algebraic dual* (space of all linear functionals) of the algebra $\times A(Z)$of noncommuting polynomials. In turn, with the usual topology (compare [35, 54]), $A(Z)$ is the *topological dual* (space of *continuous* linear functionals of $\hat{A}(Z)$. (Note, that if $P=\sum_{w\in Z^*} P_w w$ then $P_w = 0$ for all except a finite number of $w \in Z^*$, and thus the sum in (80) is finite.

Remark 3.4 In the case of a finite alphabet Z, the topology can be characterized by the metric $\|P, Q\| = 2^{-k}$ where k is the *length* of the shortest word $w \in Z^*$ such that $<w, Q-P> \neq 0$. Alternatively, a sequence $P^{(n)} \in A(Z)$ converges to $Q \in A(Z)$ if for every $M > 0$ there exists $N < \infty$ such that $<w, Q - P^{(n)}> = 0$ for all words of length at most N. In the case of an infinite one uses a similar topology, albeit its characterization in terms of neighborhood bases is slightly more technical [54].

Algebraically, the shuffle product is defined in elementary terms using the coproduct:

Definition 3.4 *(Alternate definition) The shuffle product* $ш: \hat{A}(Z)\times\hat{A}(Z) \mapsto \hat{A}(Z)$ *is the transpose of the coproduct* $\Delta: A(Z) \mapsto A(Z) \otimes A(Z)$*:*

$$<u \sqcup\!\sqcup v, w>=<u \otimes v, \Delta(w)> \quad \text{for all } u, v, w \in A(Z) \tag{81}$$

Exercise 3.7 *Proof by induction on the lengths of the words that this algebraic definition is equivalent to the recursive combinatorial definition in equation* (75) *or alternatively equations* (70) *and* (73).

While there are many formulas that mix shuffle and concatenation product, they naturally *live* on spaces that are dual to each other, the algebra $A(Z)$ of noncommutative polynomials and the algebra $\hat{A}(Z)$ of noncommutative formal power series. But the latter naturally contains the former, and one may equip each with both products, giving rise to two *Hopf algebra* structures on $A(Z)$ see [54] for details. In these notes we shall not go deeper into this, for details see e.g. [54, 35].
This algebraic characterization makes it easy to establish that the set of Lie polynomials is *orthogonal* to nontrivial shuffles:

Proposition 3.5 *If $p \in L(Z) \subseteq A(Z)$ is a Lie polynomial and and $u, v \in Z^* \setminus \{1\}$ are nonempty words, then*

$$<u ш v, p> = 0. \tag{82}$$

The proof is a short calculation, using the natural pairing of $\hat{A}(Z) \otimes \hat{A}(Z)$ with $A(Z) \otimes A(Z)$. For a Lie polynomial p and nonempty words u, v calculate

$$\begin{array}{rcl} <u \amalg v, p> & = & <u \otimes v, \Delta(p)> \\ & = & <u \otimes v, p \otimes 1 + p \otimes 1> \\ & = & <u, p> \cdot <v, 1> + <u, 1> \cdot <v, p> \\ & = & 0. \end{array} \tag{83}$$

Finally consider the action and *anti-action* [?] of $A(Z)$ on $A(Z)$ by right and translations, i.e. for $a \in Z$ and $w \in Z^*$ define

$$\varrho_a, \lambda_a \colon A(Z) \mapsto A(Z), \qquad \varrho_a(w) = wa \quad \text{and} \quad \lambda_a(w) = aw \tag{84}$$

Note that $\varrho_a \varrho_b(w) = wba$ reverses the order, while $\lambda_a \lambda_b(w) = abw$ preserves order. It is easy to extend ϱ_w and λ_w to $w \in A(Z)$ but we shall have no need for this. Instead, we are interested in the transposes $\varrho_a^\dagger, \lambda_a^\dagger \colon \hat{A}(Z) \mapsto \hat{A}(Z)$ of these translations which are defined on words $w, z \in Z^*$ by

$$<\varrho_a^\dagger w, z> = <w, za> \qquad \text{and} \qquad <\lambda_a^\dagger w, z> = <w, az> \tag{85}$$

Clearly, if $a, b \in Z$ then $\rho_a^\dagger b = \lambda_a^\dagger b = 1$ if $a = b$ and 0 else. If $b \in Z$ and $w \in Z^* \setminus Z$ are words that are not letters, then

$$\varrho_a^\dagger(wb) = \left\{ \begin{array}{ll} w & \text{if } a = b \\ 0 & \text{else} \end{array} \right\} = \lambda_a^\dagger(bw) \tag{86}$$

Recall that a linear map $D \colon A \mapsto A$ on an algebra A with product $\cdot$ is called a derivation if $D(f \cdot g) = (D(f)) \cdot g + f \cdot (D(g))$ for all $f, g \in A$.

Exercise 3.8 *Show that the composition* $D_2D_1\colon f \mapsto D_2(D_1(f))$ *of two derivations* D_1 *and* D_2 *need not be a derivation, but the commutator* $D_2D_1 - D_1D_2$ *is always a derivation. Show that the set of derivations on an algebra always has a Lie algebra structure with the commutator as product.*

The following observation appears to have very profound consequences, compare the next section and the next chapter.

Theorem 3.6 *Both transposes* $\varrho_a^\dagger$ *and* $\lambda_a^\dagger$ *are derivations on the shuffle algebra* $(\hat{A}(Z), \text{ш})$*, but* only *the transpose* $\lambda_a^\dagger$ *of the left translation* λ_a *by a letter* a *is a derivation on the chronological algebra* $(C(Z), *)$*.*

Exercise 3.9 *Prove theorem* (3.6) *using the recursive combinatorial definitions of the shuffle and chronological products in equations* (75) *and* (70)*, and/or the algebraic definition in* (3.4)*.*

3.3 Hall Viennot bases for free Lie algebras

The goal of this section is to develop bases for free Lie algebras, and get some insight into their background, constructions and properties. We start with some remarks introducing binary labeled trees which will be necessary to later construct Hall bases as these critically depend on the notion of left and right factors. Returning to Lie algebras, we then consider the process of Lazard elimination which is the main procedure for generating bases of a free Lie algebra. (However, [54] provides an independent argument.) Finally we survey Hall-Viennot bases and their most important properties.

First recall some of the problems we encountered earlier. In every Lie algebra $[x,[y,[x,y]]] = [y,[x,[x,y]]]$ because $[[x,y],[x,y]] = 0$. This re-emphasizes that a Lie-bracket does not have well-defined left and right factors, and a need for a language of *formal brackets.* Such language also allows one to more precisely phrase the conditions for STLC of the previous chapter: Recall if $f_1 = \frac{\partial}{\partial x_1}$ and $f_0 = x_1^2 \frac{\partial}{\partial x_1}$ then strictly speaking the number of factors f_1 and f_0 in $[f_1,[f_1,f_0]$ is not well defined as e.g. $[f_1,[f_1,f_0] = 2\frac{\partial}{\partial x_1} = f_1$.
Formally, introduce the *"free magma"* $\mathcal{M}(Z)$, the set of rooted binary trees labeled in Z (also called *parenthesized words*). This set is recursively constructed by $M^1(Z) = \{Z\}$,

$$M^{n+1}(Z) = \cup_{k=1}^{n} M^k(Z) \times M^{n+1-k}(Z) \quad \text{and} \quad \mathcal{M}(Z) = \cup_{n=1}^{\infty} M^n(Z) \quad (87)$$

There are canonical maps $\varphi\colon \mathcal{M}(Z) \mapsto L(Z)$ and $\psi\colon \mathcal{M}(Z) \mapsto A(Z)$ defined for $a \in Z$,

$$\varphi(a) = \psi(a) = a \quad \text{and} \quad \varphi(t', t'') = [\varphi(t'), \varphi(t'')]a \quad \text{and} \quad \psi(t', t'') = \psi(t')\psi(t'') \tag{88}$$

Note that every tree $t \in \mathcal{M}(Z)$ is either a letter $t \in Z$ or it has well-defined left and right subtrees $t', t'' \in \mathcal{M}(Z)$, i.e. $t = (t', t'')$. Also, for each tree the numbers of times that each letter appears as a leaf are well-defined. Formally, for $a \in Z$ define $\| \cdot \|_a\colon \mathcal{M}(Z) \mapsto \mathbf{Z}_0^+$ by $\|a\|_a = 1$, $\|b\|_a = 0$ if $a \neq b \in Z$, and recursively for general trees $\|(t', t'')\|_a = \|t'\|_a = \|t''\|_a$. For example if $t = (a(ab))$ then $\|t\|_a = 2$ and $\|t\|_b = 1$. This map naturally carries over to Z^* to Lie monomials (images of single trees under φ. However, the notions of left and right subtree do *not* carry over to $L(Z)$, e.g. consider $Z = \{a, b\}$. Then $(a, b) \in \mathcal{M}(Z)$ maps to $\varphi((a, b)) = [a, b] = [-b, a] \in L(Z)$. Similarly, $(a, (b, (a, b))) \neq (a, (b, (a, b))) \in \mathcal{M}(Z)$, yet

$$\varphi((a, (b, (a, b))) = [a, [b, [a, b]] = [b, [a, [a, b]] = \varphi((a, (b, (a, b))) \tag{89}$$

due to anti-commutativity and the Jacobi-identity in $L(Z)$).

The language of formal brackets or trees is the one which allows for very precise statements of the theorems for controllability, such as Sussmann's theorem (1.10). However, one finds an abundance of rather sloppy statements of this and similar conditions in the recent literature, and our presentation in section 1.4 is just barely *border-line*. See the original article (1.10) for utmost precision.

It may take some time to get used to the precision needed to describe formal operations such as *expressing each iterated Lie bracket as a linear combination of a set of specified brackets*. E.g. working in $L(Z)$ it is correct to write $[[x, y], z] = [x, [y, z]] - [y, [z, x]]$, but for trees $((x, y), z) \neq (x, (y, z)) - (y, (z, x))$ (it is straightforward to introduce linear combinations of trees), visualized as

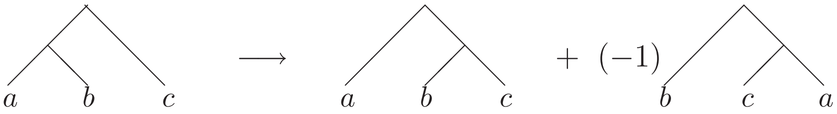


Figure 4. Jacobi identity, Leibniz rule and rewriting

This distinction is precisely what is needed in order to specify an algorithm how to *rewrite* formal brackets (and trees), and then map to identities of Lie brackets in $L(Z)$.

A tree $t \in \mathcal{M}(Z)$ is called a *Dynkin tree*, written $Dyn(Z)$, if either $t \in Z$ or $t = (a, t'')$ with $a \in Z$ and $t'' \in Dyn(Z)$. Note that these trees correspond exactly to the iterated Lie derivatives that appear in control when repeatedly differentiating e.g. output functions along solution curves of the system. The next exercise shows that the image $\varphi(Dyn(Z)) \subseteq L(Z)$ spans $L(Z)$. But since e.g. $(a,(b,(a,b))),(a,(b,(a,b))) \in Dyn(Z)$, yet $[a,[b,[a,b]]] = [b,[a,[a,b]] \in L(Z)$ it is clear that (even after removing obviously trivial trees like (a,a)) that $\varphi(Dyn(Z))$ is not a basis for $L(Z)$. (Nonetheless, this set is often called a *Dynkin basis*, clearly a bad misnomer.) Note that a Lie bracket may be a the image of a Dynkin tree even if it looks quite different, e.g. $[[y,[[x,y],x]],x] =\in \varphi(Dyn(Z))$ because $[[y,[[x,y],x]],x] = [x,[y,[x,[x,y]]]] = \varphi(\,(x,(y,(x,(x,y))))\,) \in \varphi(Dyn(Z))$.

Exercise 3.10 *Show that* $\varphi(Dyn(Z)) \subseteq L(Z)$ *spans* $L(Z)$. Work with trees, and adapt the *"rewriting rule"* $((x,y),z) \longrightarrow \{(x,(y,z)),\quad (y,(x,z))\}$ to recursively reduce any tree to a subset of $Dyn(Z)$. For precise, technical notions of *rewriting systems* see [54].)

The basic construction of bases for $L(Z)$, as well as e.g. the construction of Sussmann's infinite exponential product expansion of theorem (4.10) are fundamentally based on the concept of Lazard elimination. This rests on the following basic theorem, simple proofs (but using technical language beyond the scope of these notes) of which may be found in [4, 54]. We shall be satisfied with a simple illustration of the elimination process.

Theorem 3.7 (Lazard elimination). *Suppose* Z *is a (not necessarily finite) set and* $a \in Z$*. Then the free Lie algebra* $L(Z)$ *is the direct sum of the one-dimensional subspace* $\{a\}_{\mathbb{R}} = \{\lambda a \colon \lambda \in \mathbb{R}\}$*, spanned by* a*, and a Lie algebra that is freely generated (as a Lie algebra) by the set* $\{(\mathrm{ad}^{\mathrm{k}}\mathrm{a},\mathrm{b})\colon \mathrm{k} \geq 0,\ \mathrm{b} \in \mathrm{Z} \setminus \{\mathrm{a}\}\ \}$.

Exercise 3.11 *Adapt the* rewriting *process from exercise* (3.10) *to show that* $L(\{a,b\}) \subseteq \{a\}_{\mathbb{R}} \;\oplus\; L(\{(\mathrm{ad}^{\mathrm{k}}\mathrm{a},\mathrm{b})\colon \mathrm{k} \geq 0,\ \mathrm{b} \in \mathrm{Z} \setminus \{\mathrm{a}\}\ \})$. In view of the exercise (3.10), it suffices to show (by induction) that every Dynkin bracket

$[a_{i_r}, [a_{i_{r-1}}, [, \ldots [a_{i_2}, a_{i_1}]] \ldots]$ with $a_{i_j} \in \{a, b\}$ can be written as a linear combination of brackets of the form $[(\mathrm{ad}^{i_s} a, b), \ldots [(\mathrm{ad}^{i_2} a, b), (\mathrm{ad}^{i_1} a, b)]] \ldots]$ with $i_j \geq 0$. Work either on the level of trees or in $L(Z)$, but carefully reflect on your choices.

For illustration consider a two letter alphabet $Z = \{a, b\}$. Then

$$\begin{aligned}
L(Z) &= \{a\}_{\mathbb{R}} \bigoplus \{b, [a,b], [a,[a,b]], \ldots\} \\
&= \{a\}_{\mathbb{R}} \oplus \{b\}_{\mathbb{R}} \oplus \{[a,b], [a,[a,b]], \ldots, [b,[a,b]], [b,[b,[a,b]], \ldots\} \\
&= \{a\}_{\mathbb{R}} \oplus \{b\}_{\mathbb{R}} \oplus \{[a,b]\}_{\mathbb{R}} \oplus \{(\mathrm{ad}^k[a,b], (\mathrm{ad}^j b, (\mathrm{ad}^i a, b))) \colon i,j,k \geq 0\} \\
&= \{a\}_{\mathbb{R}} \oplus \{b\}_{\mathbb{R}} \oplus \{[a,b]\}_{\mathbb{R}} \oplus \{[a,[a,b]]\}_{\mathbb{R}} \\
&\qquad \oplus \{(\mathrm{ad}^\ell[a,[a,b]], (\mathrm{ad}^k[a,b], (\mathrm{ad}^j b, (\mathrm{ad}^i a, b)))) \colon i,j,k,\ell \geq 0\} \\
&= \{a\}_{\mathbb{R}} \oplus \{b\}_{\mathbb{R}} \oplus \{[a,b]\}_{\mathbb{R}} \oplus \{[a,[a,b]]\}_{\mathbb{R}} \oplus \{[b,[a,b]]\}_{\mathbb{R}} \oplus \cdots
\end{aligned} \tag{90}$$

Note that at every stage the infinite dimensional part is replaced by a new infinite dimensional part that is *generated by infinitely many times "more" generators* than the previous. Nonetheless, on can anticipate the *convergence* as all these generators become *"longer"* and longer (provided the eliminated brackets are chosen properly), compare remark 3.4. What is important to remember for applications and in the sequel, are the successive elimination of one bracket at a time, thereby conceivably constructing a basis, and the type of bracket that is common to all generators, and thus also to all eliminated elements. Note, the above elimination process should again be done on trees, and then mapped to $L(Z)$ – however, we presented it in the more traditional Lie algebra setting.

One defines *Lazard sets* as subsets of $\mathcal{M}(Z)$, that, roughly speaking, arise from infinite repetition of the illustrated Lazard elimination process – the technical statement is quite lengthy [54], and we omit it here. What matters is the following (again, for a proof see [54])

Theorem 3.8 *The image of a Lazard set $\mathcal{L} \subseteq \mathcal{M}(Z)$ under the map φ is a basis for $L(Z)$.*

Starting with Marshall Hall in the 1940s, whose work builds on Phillip Hall's studies of commutator groups in the 1930s, several bases for free Lie algebras have been proposed. Aside from *Hall bases*, the best known names are *Lyndon bases* and *Sirsov bases*. The latter two were eventually found to basically be same. In the 1970s Viennot [68] slightly relaxed the conditions in the construction of Hall bases, and showed that with that relaxed notion Lyndon (and thus also Sirsov) bases are just special cases of the generalized Hall bases. When we want to emphasize the distinction, we will refer to the latter as *Hall Viennot bases*.

Definition 3.5 *A Hall set over a set* Z *is any strictly ordered subset* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(\mathbf{Z})$ *that satisfies*

(i) $Z \subseteq \tilde{\mathcal{H}}$

(ii) Suppose $a \in Z$*. Then* $(t,a) \in \tilde{\mathcal{H}}$ *iff* $t' \in \tilde{\mathcal{H}}$*,* $t' < a$ *and* $a < (t',a)$*.*

(iii) Suppose $u, v, w, (u,v) \in \tilde{\mathcal{H}}$*.*
Then $(t',(t''',t''''))\in \tilde{\mathcal{H}}$ *iff* $t''' \leq t' \leq (t''',t'''')$ *and* $t' < (t',(t''',t''''))$*.*

The definition given here twice reverses the convention of the one given in [54]: All trees have left and right factors swapped, and larger and less have been reversed. On the other hand, the definition given here is compatible with the conventions of [4].

The original Hall bases, as presented in Bourbaki [4], require that ordering be compatible with the length, i.e. if the length of the word $\psi(t)$ is less then the length of the word $\psi(t')$, then $t < t'$. Viennot replaced this condition (and minor other parts) by condition (ii) in definition (3.5).

For the sake of completeness, we state (even without having given a definition of Lazard sets)

Theorem 3.9 *Every Hall set is a Lazard set.*

Theorem 3.10 *The image of a Hall set under the map* $\varphi\colon \mathcal{M}(Z) \mapsto L(Z)$ *is an ordered basis for* $L(Z)$*.*

While this is an immediate corollary of theorems 3.9 and 3.8, its importance earns it the title of a theorem. Also Reutenauer [54] also gives a direct proof that is not based on Lazard sets.

It is very straightforward to inductively construct Hall sets, especially when choosing an order that is compatible with the length of the associated word. See the example given in figure 5. However, for many applications Lyndon bases seem to be even more efficient – and effective algorithms have been coded that factor e.g. every word into a product of Lyndon words etc.

Definition 3.6 *Order the alphabet Z. A word $w \in Z^+$ is a* Lyndon word *if it is strictly smaller than its cyclic rearrangements with respect to the lexicographical ordering on Z^*, i.e.*

$$\textit{If } w = uv \textit{ with } u, v \in Z^+ \textit{ then } uv < vu. \tag{91}$$

This ordering is compatible with the choices for Hall bases in [54]. To match our choices, we need to read the words backwards *and* replace *strictly smaller* by *strictly larger* (i.e. reverse the ordering of Z). Note that it is very easy to tell whether a *given* word is a Lyndon word, and it is also easy to factor:

b
$(((((ab)b)b)b)b)$
$((((ab)b)b)b)$
$(a((((ab)b)b)b))$
$(((ab)b)b)$
$((ab)(((ab)b)b))$
$(a(((ab)b)b))$
$(a(a(((ab)b)b)))$
$((ab)b)$
$((ab)((ab)b))$
$(a((ab)((ab)b)))$
$(a((ab)b))$
$\boxed{((ab)(a((ab)b)))}$
$(a(a((ab)b)))$
$(a(a(a((ab)b))))$
(ab)
$((a(ab))(ab))$
$(a((a(ab))(ab)))$
$(a(ab))$
$(a(a(ab)))$
$(a(a(a(ab))))$
$(a(a(a(a(ab)))))$
a

$((a(ab))(b(ab)))$
$((ab)(b(b(ab))))$
$((ab)(b(a(ab))))$
$((ab)(a(a(ab))))$
$(b(b(b(b(ab)))))$
$(b(b(b(a(ab)))))$
$(b(b(a(a(ab)))))$
$(b(a(a(a(ab)))))$
$((ab)(b(ab)))$
$((ab)(a(ab)))$
$(b(b(b(ab))))$
$(b(b(a(ab))))$
$(b(a(a(ab))))$
$(a(a(a(ab))))$
$(b(b(ab)))$
$(b(a(ab)))$
$(a(a(ab)))$
$(b(ab))$
$(a(ab))$
(ab)
b
a

Figure 5. Doubly reversed Lyndon trees and Ph. Hall trees

Note that for short words the reversed Lyndon words with $b < a$ are almost the same as the usual Lyndon words with $a < b$. The framed tree in figure 5 shows that this is just a coincidence for small trees.

Suppose $w \in Z^+$ is a Lyndon word. If $w \in Z$ is a letter there is nothing to be factored. If not, then there exists a pair $(u, v) \in Z^+ \times Z^+$ such that u is the longest *left factor* of w that is a Lyndon word. It can be shown that then v is also a Lyndon word. Repeating this factorization recursively (i.e. for the left and right factors) one obtains a map from the set of Lyndon words into the set $\mathcal{M}(Z)$ of trees which is the right inverse (on the set of Lyndon words) of the map $\psi: \mathcal{M}(Z) \mapsto Z^*$ which *forgets* the tree structure and maps each tree to its (ordered sequence of) leaves.

Exercise 3.12
Construct all Lyndon trees with at most 5 *leaves for a three letter alphabet* $Z = \{0, 1, 2\}$.

Exercise 3.13 *Consider the three letter alphabet* $Z = \{0, 1, 2\}$ *and construct an ordered subset of a set of Hall trees containing all trees with at most* 5 *leaves. Be aware of the freedom to choose the ordering of newly constructed trees (and the consequences of the choice upon later constructed trees).*

Exercise 3.14 *Verify that the restriction of the map* $\psi: \mathcal{M}(\{0, 1, 2\}) \mapsto Z^+$ *to the set of Hall trees given in figure 5 is one-to-one. Write down the image of this set in* Z^+*, and develop an algorithm that recovers the trees* $t \in \mathcal{M}(\{a, b\})$ *from the images* $\psi(t) \in Z^+$.

Exercise 3.15 *Verify that the restriction of the map* $\psi: \mathcal{M}(\{0, 1, 2\}) \mapsto Z^+$ *to the set constructed in exercise* 3.13 *is one-to-one. Write down the image of this set in* Z^+*, and develop an algorithm that recovers the trees* $t \in \mathcal{M}(\{0, 1, 2\})$ *from the images* $\psi(t) \in Z^+$.

The one-to-one-ness of the restrictions of the map $\psi: \mathcal{M}(\{0, 1, 2\}) \mapsto Z^+$ to Hall sets $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ is one of the most important properties of Hall sets. As a practical consequence, it allows one to carry out most calculations using the words $\psi(t)$ (e.g. as indices) rather than the trees t themselves (which take much more effort to write without mistakes).

Definition 3.7 *Consider a fixed Hall set* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$. *A word* $h \in Z^+$ *is called a* Hall word *if it is the image* $\psi(t)$ *of a Hall-tree* $t \in \tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$.

In many practical cases one may be quite sloppy, identifying a tree t with the Hall-word $h = \psi(t)$. However, it is very important to understand that one must specify which Hall set one is working with, as over every alphabet Z with at least two letters there exists many Hall sets, compare the next exercise.

Exercise 3.16 *Construct an example of different trees* $t_1 \neq t_2$ *that belong to different Hall sets* $t_i \in \tilde{\mathcal{H}}_i \subseteq \mathcal{M}(Z)$ *(over the same alphabet* Z*), but which have the same* foliage $\psi_1(t_1) = \psi_2(t_2)$ *under the* "forget" *maps* $\psi_i \colon t_i \in \mathcal{H}_i \mapsto Z^+$.

Directly related to this one-to-one-ness of the restriction of the maps ψ to Hall-Viennot sets, is the fact that every word $w \in Z^+$ has a unique factorization into a structured product, as made precise in the next theorem. This property has been characterized by Viennot as one of the fundamental building blocks of Hall-Viennot sets, and as *the* property that makes Hall-Viennot bases *optimal* [48, 49, 54, 68]. Compare also the Poincaré-Birkhoff-Witt theorem 4.6.

Theorem 3.11 *Suppose* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ *is a Hall Viennot set and* $\mathcal{H} = \psi(\tilde{\mathcal{H}}) \subseteq Z^+$ *is the corresponding set of Hall words with the induced ordering. Then every word* $w \in Z^+$ factors *uniquely into a nonincreasing product of Hall words, i.e. there exist unique* $s \geq 0$, $h_j \in \mathcal{H}$, *such that*

$$w = h_1 h_2 \ldots h_s \quad \textit{and} \quad h_1 \geq h_2 \geq \ldots \geq h_s \tag{92}$$

Exercise 3.17 *To get a feeling for this factorization, consider the sets given in figure 5, write down a list of increasingly longer* random *words (e.g. up to length* 20*) and factor according to theorem* (3.11).

Exercise 3.18

Repeat the preceding exercise for the sets constructed in exercises (3.12) *and* (3.13).

An important special case is the case that $w = h \in \mathcal{H}$ is itself a Hall word. Clearly h is its own unique factorization into a nonincreasing product of Hall words. However, if we truncate the last letter of h, e.g. if $h = za$ with $a \in Z$, then we may consider the unique factorization of z into a nonincreasing product $z = h_1 H - 2 \ldots h_s$ of Hall words. A little reflection shows that necessarily also $h_s < a$, as illustrated in the figure 6. We will return to this diagram when discussing the structure formula for the dual Poincaré-Birkhoff-Witt bases in section 4.3, compare figure 7.

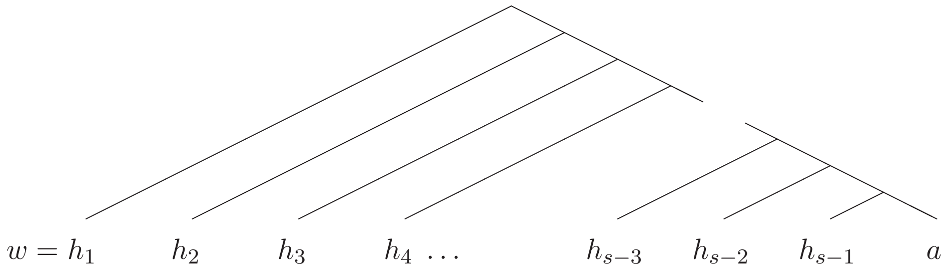


Figure 6. Structure and factorization of Hall trees

This tree very suggestively shows how Hall sets *grow out* of successive Lazard elimination. Note also the close correspondence to theorem 3.6, that while $\varrho_a^\dagger$ and $\lambda_a^\dagger$ are derivations on the shuffle algebra $(\hat{A}(Z), ш)$, *only* the transpose $\lambda_a^\dagger$ of the left translation λ_a by a letter a is a derivation on the chronological algebra $(C(Z), *)$.

While it is convenient to write a word $w \in Z^+$ where one really means the iterated Lie bracket $\varphi(\psi^{-1}(w)) \in L(Z)$, care needs to be taken not to confuse these two. This becomes even more important as the coordinate $< \varphi(\psi^{-1}(w)), w >$ of w in the Lie polynomial $\varphi(\psi^{-1}(w)) \in L(Z)$ (with respect to the basis Z^* of $A(Z)$) generally is not zero. Indeed, for some Hall bases (especially, Lyndon bases) the word w is always the smallest (largest) word that appears with nonzero coefficient in the Lie polynomial $\varphi(\psi^{-1}(w))$. For further details see [48, 49, 54].

4 A primer on exponential product expansions

4.1 Ree's theorem and exponential Lie series

Many of the constructions and properties of noncommuting polynomials and Lie polynomials from section 3.2 carry directly over to infinite series and infinite Lie series, although some extra caution needs to be taken when working with infinite alphabets. In the following we assume that Z is finite unless otherwise noted. We only summarize a few key notions, and concentrate on what is new and relevant for control. For a detailed description see [54].

Definition 4.1 *A formal series $f \in \hat{A}(Z)$ is a Lie series, written $f \in \hat{L}(Z)$ if for every $n \in \mathbf{Z}^+$ the* homogeneous component *f_n is a Lie polynomial, where*

$$f_n \overset{\text{def}}{=} \sum_{|w|=n} <f, w> w \quad \in L(Z) \tag{93}$$

For example, the characterization of Lie polynomials in theorem 3.3 carries directly over to Lie series.

Theorem 4.1 *A formal power series $f \in \hat{A}(Z)$ is a Lie series if and only if*

$$\Delta(f) = f \otimes 1 + 1 \otimes f \tag{94}$$

Exercise 4.1 *Prove theorem* 4.1 *using definition* 4.1 *and theorem* 3.3.

Just as in the case of Lie polynomials, see (82), one immediately obtains that a series $f \in \hat{A}(Z)$ is a Lie series if and only if both $<f, 1> = 0$ and f is orthogonal to all nontrivial shuffles $<f, u \sqcup\!\sqcup v> = 0$ for all $u, v \in A(Z)$.

Two common ways in which series arise from polynomials (and from other series) are the exponential map and its inverse, both defined on suitable domains.

Definition 4.2
For any power series $s \in \hat{A}(Z)$ with zero constant term define the exponential by

$$e^s = \sum_{k=0}^{\infty} \tfrac{s^k}{k!} = 1 + s + \tfrac{s^2}{2} + \tfrac{s^3}{6} + \tfrac{s^4}{24} + \ldots \tag{95}$$

Definition 4.3
For any power series $s \in \hat{A}(Z)$ with constant term $<s, 1> = 1$, define the logarithm by

$$\log s = \sum_{k=1}^{\infty} \tfrac{(-)^{k+1}}{k}(s-1)^k = (s-1) - \tfrac{(s-1)^2}{2} + \tfrac{(s-1)^3}{3} + \tfrac{(s-1)^4}{4} - \ldots \tag{96}$$

Exercise 4.2
Verify that exp: $s \mapsto e^s$ *and* log, *as defined above, are right, respectively left, inverses of each other. Carefully state the domains on which they are inverses. Check which of the* usual *identities for exponentials and logarithms hold for these maps on $\hat{A}(Z)$ defined as series.*

Theorem 4.2 (Friederich's criterion) *A power series* $p \in \hat{A}(Z)$ *with constant term* $< p, 1 >= 1$*, is an exponential Lie series (i.e.* $\log p \in \hat{L}(Z)$*), written* $p \in \hat{G}(Z)$*, if and only if*

$$\Delta(p) = p \otimes p. \tag{97}$$

Formally this is the result of a short calculation, shown below in one direction (assuming that $f \in \hat{L}(Z)$), using the continuity of the exponential, of the coproduct, and of the maps $f \mapsto 1 \otimes f$ and $f \mapsto f \otimes 1$. (For careful justifications of all steps see theorem 3.2 in [54].)

$$\Delta(e^f) = e^{\Delta(f)} = e^{f\otimes 1+1\otimes f} = e^{f\otimes 1}e^{1\otimes f} = \Big(e^f \otimes 1\Big)\cdot\Big(1 \otimes e^f\Big) = e^f \otimes e^f \tag{98}$$

Exercise 4.3 *Suppose* $s \in \hat{A}(Z)$ *is a formal power series with constant term* $<s, 1>= 1$*. Verify that* $s^{-1} \in \hat{A}(Z)$ *and* $s^{-1}s = ss^{-1} = 1$ *where*

$$s^{-1} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty}(-)^k(s-1)^k = 1-(s-1)+(s-1)^2-(s-1)^3+(s-1)^4+\dots \tag{99}$$

Use your result to find all series $s \in \hat{A}(Z)$ *which are invertible in this sense.*

Exercise 4.4 *Show that* $\hat{G}(Z)$ *is a group under multiplication, i.e. verify that* Δ *extends* continuously *to an associative algebra homomorphism on* $\hat{A}(Z)$*, and then use theorem* 4.2 *to verify that if* $p, q \in \hat{G}(Z)$*, then also* $p \cdot q \in \hat{G}(Z)$ *and* $p^{-1} \in \hat{G}(Z)$*.*

This justifies the name *"group-like elements"* for power series in $\hat{G}(Z) \subseteq \hat{A}(Z)$, i.e. $\hat{G}(Z)$ is a formal Lie group whose algebra is $\hat{L}(Z)$. Translating back to control, the Lie series $\log p \in \hat{L}(Z)$ correspond to (infinite) linear combination of) vector fields, while the exponential $p \in \hat{G}(Z)$ corresponds to the flow of $\log p$ evaluated at time $t = 1$ (or the flow of $\frac{1}{c}\log p$ evaluated at time c). Thus $\hat{G}(Z)$ corresponds to a formal group of diffeomorphisms, and we think of $p \in \hat{G}(Z)$ as a *point*. Alternatively, these points may be characterized as *multiplicative linear functionals* on the associative algebra of *smooth functions on the state space of the system*, see [35] for more details. In the algebraic setting of this section, this corresponds to the following theorem (recall that Υ maps shuffle multiplication to pointwise multiplication of functions, see corollary 3.2):

Proposition 4.3 *The* points $p \in \hat{G}(Z)$ *are* multiplicative linear maps *on the algebra* $A(Z, ⧢)$.

The proof is a straightforward calculation using the algebraic characterization (81) of the shuffle product.

$$<p, \phi \,\text{Ш}\, \psi> \,=\, <\Delta(p), \phi \otimes \psi> \,=\, <p \otimes p, \phi \otimes \psi> \,=\, <p, \phi> \,\cdot\, <p, \psi> \tag{100}$$

This may be thought of as a multi-dimensional, noncommutative analogue of the identification of the *point* $p \in (\mathbb{R}, +)$ with the translation $\tau_p \colon x \mapsto (x + p)$ (considered as a diffeomorphism of $\mathbb{R}$), and with the Taylor formula (for more advanced analysis see e.g. [1, 2, 19, 35])

$$\begin{array}{rcl} p & : & f \mapsto f(p). \\ \updownarrow & & \\ \tau_p & : & f(\cdot) \mapsto f(\cdot + p). \\ \updownarrow & & \\ e^{p\frac{d}{dx}}\Big|_0 = \displaystyle\sum_{k=0}^{\infty} \frac{p^k}{k!} \left(\frac{d}{dx}\right)^n \Big|_0 & : & f \mapsto \displaystyle\sum_{k=0}^{\infty} \frac{(p-0)^k}{k!} f^{(n)}(0) \end{array} \tag{101}$$

Continuing with making connections to control, we have the following theorem which follows almost immediately from re-reading (100). (For a short technical proof of the *"if"* direction see theorem 3.2 in [54].)

Theorem 4.4 (Ree's theorem*) A noncommutative power series*

$$p = 1 + \sum_{w \in Z^+} p_w w \in A(Z)$$

(*with constant term* $<p, 1> = 0$) *is an exponential Lie series,* (*i.e.* $p \in \hat{G}(Z)$ *or equivalently* $\log p \in \hat{L}(Z)$) *if and only if its* coefficients satisfy the shuffle relations*, i.e. if the map*

$$w \mapsto p_w \overset{\text{def}}{=} \; <w, p> \tag{102}$$

is an (associative algebra) homomorphism from $A(Z, \,\text{ш}\,)$ *to* $\mathbb{R}$*.*

Note that in (102) any $w \in A(Z)$ is allowed, whereas previously p_w was defined only for $w \in Z^*$. In view of theorem 3.1 and corollary 3.2, that the map Υ is both a chronological and an associative algebra homomorphism, this yields right away the following fundamental fact. (We will return to this in the subsequent sections, e.g. in (115)).

Proposition 4.5 *The Chen Fliess series is the image of an exponential Lie series.*

4.2 From infinite series to infinite products

The main item of this section is the Poincaré-Birkhoff-Witt theorem which relates Lie algebras to certain associative algebras, specifically relating their bases. This naturally leads one to consider infinite exponential products, especially for the Chen Fliess series for which the coefficients will be determined in the next section. But we start with a brief review - using the more compact notation and terminology developed in recent chapters. Recall from section 2.1 how the Chen Fliess series as an *infinite series* arose from solving a universal control system by *iteration*. (In contrast, the infinite exponential product in the next section will arise from *variation of parameters*).

Definition 4.4 *For any finite alphabet* Z *the* universal control system *is the formal bilinear system on* $\hat{A}(Z)$

$$\dot{s} = s \cdot \sum_{a \in Z} u_a a \quad \textit{with initial condition} \;\; s(0) = 1. \tag{103}$$

Here $u_a \colon t \mapsto u_a(t)$ *are locally integrable scalar controls and* $s \colon t \mapsto \hat{A}(Z)$ *is the solution curve.*

Exercise 4.5 *For the case of the three-letter alphabet* $Z = \{0, 1, 2\}$ *using the* basis Z^**, write out the first few components of the system* (103)*, i.e.* $\dot{s}_e = \ldots$, $\dot{s}_a = \ldots$, $\dot{s}_{ab} = \ldots$, *etc. (Here* e *is the empty word, and* $a, b \in Z$*). Write out the components* $s_w(t)$ *of the solution curve* $s(t)$ *using iterated integrals of the controls.*

Using the chronological product $(U * V)(t) = \int_0^t U(\tau)V'(\tau)d\tau$, and writing $U_a(t) = \int_0^t u_a(\tau)\, d\tau$ for the integrals of the controls, the integrated form of the *universal control system* (103)

$$s(t) = 1 + \int_0^t s(\tau)F'(\tau)\, d\tau \quad \text{with} \;\; F = \sum_{a \in Z} U_a a, \tag{104}$$

is most compactly written as

$$s = 1 + s * F \tag{105}$$

Iteration yields the explicit series expansion

$$\begin{array}{rcl} s & = & 1+(1+s*F)*F \\ & = & 1+F+((1+s*F)*F)*F \\ & = & 1+F+(F*F)+(((1+s*F)*F)*F)*F \\ & = & 1+F+(F*F)+((F*F)*F)+((((1+s*F)*F)*F)*F)*F \\ & \vdots & \\ & = & 1+F+(F*F)+((F*F)*F)+(((F*F)*F)*F)\ldots \end{array}$$

Using intuitive notation for chronological powers (compare definition 4.8) this solution formula in the form of an infinite series is compactly written as

$$s=\sum_{n=0}^{\infty} F^{*n} \quad = \; 1+F+F^{*2}+F^{*3}+F^{*4}+F^{*5}+F^{*6}+\ldots \tag{106}$$

After this review of how *solving differential equations by iteration* yields infinite series expressions for the solution curves, we develop some abstract background that will lead to effective product expansions of the solution curves.

Every Lie algebra can be *imbedded* into an associative algebra: The universal enveloping algebra $\mathcal{U}$ of a Lie algebra $\mathcal{L}$ (with natural Lie algebra homomorphism $\imath$) is, by definition, the associative algebra (which exists, and is unique up to homomorphism) such that whenever A is an associative algebra and $\Phi{:}\,\mathcal{L}\mapsto A$ is a Lie algebra homomorphism, then there exists a map $\Psi{:}\,\mathcal{U}\mapsto A$ such that $\Phi=\Psi\circ\imath$. The fundamental theorem (following [54]) is

Theorem 4.6 (Poincaré-Birkhoff-Witt theorem) *Suppose* $\mathcal{B}=\{b_\alpha{:}\,\alpha\in I\}$ *is an ordered basis for a Lie algebra* $\mathcal{L}$ *. Further suppose* $\mathcal{U}$ *is the universal enveloping algebra of* $\mathcal{L}$ *with inclusion map* $\imath{:}\,\mathcal{L}\mapsto\mathcal{U}$*. Then a basis for* $\mathcal{U}$ *is given by the set of decreasing products*

$$\{\imath(b_{\alpha_n})\imath(b_{\alpha_{n-1}})\ldots\imath(b_{\alpha_2})\imath(b_{\alpha_1}) : \alpha_n\geq\alpha_{n-1}\geq\ldots\alpha_2\geq\alpha_1,\;\; \alpha_i\in I\} \tag{107}$$

For a proof of the Poincaré-Birkhoff-Witt theorem see e.g. [46] or any textbook on Lie algebras. Note, we earlier used a consequence of this theorem to conclude that the Lie algebra $L(Z)$ of all Lie polynomials over Z is indeed the free Lie algebra over Z.

Of interest to us in the next section is the structure of the Poincaré-Birkhoff-Witt *basis* of the universal enveloping algebra, which in the case of $\mathcal{L}=L(Z)$

being the free Lie algebra agrees with the free associative algebra $A(Z) = \mathcal{U}$ over Z. More specifically, the set Z^* of all words over Z forms one basis of $A(Z)$, while every basis $\mathcal{B}$ of $L(Z)$ (in particular, each Hall-Viennot basis of the previous chapter) gives rise to a different Poincaré-Birkhoff-Witt *basis* of $A(Z)$.

Exercise 4.6 *Fix a Hall set* $\tilde{\mathcal{H}}$ *for the two element alphabet* $Z = \{a, b\}$ (*compare figure* 5 *in section* 3.3) *For each homogeneous component* $A^{(k,\ell)}(Z)$ *(i.e. the subspace spanned by all words* $w \in Z^*$ *with* $\|w\|_a = k$ *and* $\|w\|_b = \ell$*) with* $k + \ell \leq 4$ *write out the induced bases that arise from the Poincaré-Birkhoff-Wittbasis* $\mathcal{P}$ *built from* $\tilde{\mathcal{H}}$*, and find the transition matrix for the basis change from the standard basis* Z^**.* (This requires the expansion of Lie polynomials. See also the next example.)

For illustration consider the subspace $A^{(1,2)}(\{0,1\})$ whose standard basis (coming from Z^*) is $\mathcal{B}^{(1,2)} = \{110, 101, 011\}$. Considering a Hall set starting with $\tilde{\mathcal{H}} = \{0, 1, (1,0), (1,(1,0)),\ (0,(1,0)), \ldots\}$, the subset of the induced PBW-basis $\mathcal{P}$ for this homogeneous component is (compare with the calculations in section 2.1).

$$\mathcal{P}^{(1,2)} = \{[1,[10]], [1,0]1, 011\} = \{110 - 2\cdot 010 + 011, 101 - 011, 011\} \tag{108}$$

and the transition matrix between the two bases is

$$\begin{pmatrix} [1,[10]] \\ [1,0]1 \\ 011 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 110 \\ 010 \\ 011 \end{pmatrix} \tag{109}$$

The inverse of this matrix transforms the *dual bases* (see below) according to

$$\begin{pmatrix} \tilde{\xi}_{[1,[10]]} \\ \tilde{\xi}_{[1,0]} \sqcup\!\sqcup \tilde{\xi}_1 \\ \tilde{\xi}_0 \sqcup\!\sqcup \tilde{\xi}_1 \sqcup\!\sqcup \tilde{\xi}_1 \end{pmatrix} = \begin{pmatrix} \tilde{\xi}_{[1,[10]]} \\ \tilde{\xi}_{[1,0]1} \\ \tilde{\xi}_{011} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 110 \\ 010 \\ 011 \end{pmatrix} \tag{110}$$

To see the importance of the dual bases for control, recall the illustrative manipulations of the Chen Fliess series in section 2.1. Working with the words, or multi-indices, we started with the expression

$$110 \otimes 110 + 101 \otimes 101 + 011 \otimes 011 \tag{111}$$

and using integration by parts and collecting Lie polynomials transformed it into an expression of the form

$$\tilde{\xi}_{[1,[10]]} \otimes [1,[1,0]] + \tilde{\xi}_{[1,0]1} \otimes [1,0]1 + \tilde{\xi}_{011} \otimes 011 \qquad (112)$$

One way to think of this is as *resolving* the *identity map* (on a finite dimensional vector space) with respect to two different bases. E.g. suppose $\{e_1, \ldots e_n\}$ and $\{v_1, \ldots v_n\}$ are bases for V, and $\{e^1, \ldots e^n\}$ and $\{v^1, \ldots v^n\}$ are the corresponding dual bases for the dual space V^*, then upon the isomorphism $\mathrm{Hom}(V,V) \sim V^* \otimes V$ the identity map $\mathrm{id}_{\mathrm{v}}: V \mapsto V$ may be identified with

$$\mathrm{id}_{\mathrm{v}} \sim e^1 \otimes e_1 + e^2 \otimes e_2 + e^3 \otimes e_3 = v^1 \otimes v_1 + v^2 \otimes v_2 + v^3 \otimes v_3 \qquad (113)$$

In case of the Chen Fliess series the vector space $V = \hat{A}(Z)$ is infinite dimensional. But due to its graded structure one may consider one homogeneous component at a time. In case of a finite alphabet Z each such component is finite dimensional. The case of $A^{(1,2)}(\{0,1\})$ is typical.
Stepping back, we have ordered bases (Hall-Viennot bases) for $L(Z)$, and they induce PBW-bases on its universal enveloping algebra which is $A(Z)$. Thus it is straightforward to write down the products on the *right hand side*, e.g. $\{[1,[1,0]], [1,0]1, 011\}$ (corresponding to the partial differential operators $\{[f_1,[f_1,f_0]], [f_1,f_0]f_1, f_0f_1f_1\}$ in control). The interesting question is about the structure of the terms on the *left hand side*, algebraically the dual Poincaré-Birkhoff-Witt bases (in control the corresponding iterated integral functionals). This section is to give an elegant formula for these dual bases $\tilde{\xi}_p$.

Note the similarity of the *resolution of the identity* (113) with the Chen Fliess series: Indeed, the series is the image of the resolution of the identity map $\mathrm{id}: \hat{A}(Z) \mapsto \hat{A}(Z)$ with respect to the basis Z^* under the map

$$\Upsilon \otimes \mathcal{F}: A(Z) \otimes \hat{A}(Z) \mapsto \mathcal{IIF}(\mathcal{U}) \otimes \hat{A}(\{f_a: a \in Z\})\Big|_0 \qquad (114)$$

(to a series of partial differential operators – evaluated at "zero" – with iterated integral functionals as coefficients), i.e.

$$\Upsilon \otimes \mathcal{F}: \sum_{w \in Z^*} w \otimes w \mapsto \sum_{w \in Z^*} \Upsilon(w) \otimes \mathcal{F}(w) \qquad (115)$$

But as observed above, instead of using the standard basis Z^* for $A(Z)$, one may resolve the identity using any other basis. Of course, most useful will be the Poincaré-Birkhoff-Witt bases built on Hall bases for the free Lie algebra $L(Z)$. More specifically, suppose $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ is a Hall set. We introduce the following convenient notation:

Notation: If $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$, then write $[\,\cdot\,] = \varphi \circ \psi^{-1} \colon \psi(\tilde{\mathcal{H}}) \subseteq Z^* \mapsto L(Z)$ for the map that sends each Hall word to the corresponding Lie bracket.

The Poincaré-Birkhoff-Witt bases corresponding to the Hall set $\tilde{\mathcal{H}}$ is the set

$$\mathcal{P} = \{[h_n][h_{n-1}]\cdots[h_3][h_2][h_1] \colon h_n \geq h_{n-1} \geq \ldots h_2 \geq h_1,\ h_k \in \mathcal{H},\ n \geq 0\} \tag{116}$$

where the products of $[h_k]$ are taken in $A(Z)$ identified with the universal enveloping algebra $\mathcal{U}$ of $L(Z)$. In agreement with the prior usage in examples and in (110) formally define

Definition 4.5 *For a Poincaré-Birkhoff-Witt basis* $\mathcal{P} \subseteq A(Z)$ *denote by* $\tilde{\xi}_v \in A(Z)$ *the* dual basis elements *that are uniquely determined by*

$$<\tilde{\xi}_v, p> = \delta_{v,p} \ \textit{ for all } \ p \in \mathcal{P} \quad \textit{(Kronecker delta)} \tag{117}$$

In analogy to (113), the preimage of the Chen-Fliess series (from (115)) may thus equally be resolved as

$$\mathrm{id}_{\mathrm{A(Z)}} \sim \sum_{w \in Z^*} w \otimes w = \sum_{v \in \mathcal{P}} \tilde{\xi}_v \otimes v \tag{118}$$

(using that Z^* is self-dual) where $\mathcal{P}$ is any Poincaré-Birkhoff-Witt basis for $A(Z)$.
Earlier manipulations, e.g. (110) and section 2.1 demonstrated that explicit formulas for the dual bases elements $\tilde{\xi}_v$ (for Poincaré-Birkhoff-Witt bases over Hall sets) critically encode the iterated integral functionals in effective solution formulas. Later in this chapter we will see that indeed it suffices to obtain formulas for $\tilde{\xi}_{[h]}$ for Hall words $h \in \mathcal{H}$.

From the previous sections it is known that the Chen Fliess series is an exponential Lie series, i.e. for any basis, especially Hall basis of $L(Z)$ there exist $\zeta_h \in A(Z)$ such that

$$\sum_{w \in Z^*} w \otimes w = e^{\sum_{h \in H} \zeta_h \otimes [h]} \tag{119}$$

Such expression may be considered the formal analogue (preimage under the map $\Upsilon \otimes \mathcal{F}$) of a *continuous* Campbell-Baker-Hausdorff formula. One can obtain simple formulas for the entire exponent $\log\left(\sum_{w\in Z^*} w\otimes w\right)$ [54], compare also [64] for a cleaned up version – but such infinite linear combinations that do not use a basis obviously do not have uniquely determined coefficients. Using a Hall basis, explicit formulae for the case of a two-letter alphabet (alas a single input system with drift, or two input system without drift, in control) have been calculated for all terms up to fifth order [39]. Most recently [55] has used purely algebraic means to develop a reasonably simple, general formula for the exponent that, while not using a basis, uses a *comparatively small* spanning set for $L(Z)$. From that formula, one can obtain formulae for the ζ_h using a Hall basis, but these are again less attractive.

Indeed, the search for such simple expressions for ζ_h as h ranges over a basis of $L(Z)$ is still subject of ongoing research, with much evidence pointing to the need for a completely different construction of bases for $L(Z)$ (as Hall Viennot bases together with the Lazard elimination process are *inextricably* linked to the exponential product expansions discussed in the sequel.

Exercise 4.7 *Use the definitions* (95) *and* (96) *and the identity* (118) *to calculate explicit formulae for* ζ_h *for short Hall words from the initial segment* $\{0, 1, 10, 110, 001, 1110, 0110,\ 0010, \ldots\}$ *of a Hall set.* (This is basically a linear algebra exercise.)

An alternative to writing the series (119) as the exponential of an infinite Lie series, is to write it as an *infinite directed product of exponentials* (where both the directed product and the exponential still need to be defined, see below)

$$\sum_{w\in Z^*} w\otimes w = \overrightarrow{\prod}_{h\in H} e^{\xi_h\otimes[h]} \tag{120}$$

Exercise 4.8 *Referring to the formal definitions of directed products, use the definition* (95) *and structure of the the Poincaré-Birkhoff-Witt basis* (107) *to infer that the coefficients* ξ_h *in* (120) *indeed agree for* $[h]$ *in a Hall basis with the definition of the dual Poincaré-Birkhoff-Witt basis* $\tilde{\xi}_{[h]}$ *in* (4.5).

Notation: Since the key information is contained in the formulas $\tilde{\xi}_{[h]} = \xi_h$ for Hall elements h, and the map from Hall-trees to Hall words is injective, it is convenient to use the (deparenthesized) Hall words $h \in Z^*$ as indices rather than the coresponding Lie polynomials $[h] = \varphi(\psi^{-1}(h)) \in A(Z)$.

It remains to formally define directed products, and to consider the convergence properties of infinite products, compare remark 3.4.

Definition 4.6 *For a sequence* $\{s_k\colon k \in \mathbf{Z}_0^+\} \subseteq \hat{A}(Z)$ *of formal series inductively define the directed products via*

$$\overrightarrow{\prod}_{\emptyset} s_k = \overleftarrow{\prod}_{\emptyset} s_k = 1 \textit{ and}$$

$$\prod_{k=1}^{\longrightarrow n+1} s_k = \left(\prod_{k=1}^{\longrightarrow n} s_k\right)\cdot s_{n+1} \quad \textit{and} \quad \prod_{k=1}^{\longleftarrow n+1} s_k = s_{n+1}\cdot\left(\prod_{k=1}^{\longleftarrow n} s_k\right) \tag{121}$$

Proposition 4.7 *Suppose* $\{s_k\colon k \in \mathbf{Z}^+\} \subseteq\in \hat{A}(Z)$ *is a sequence of formal series such that for every* $N < \infty$ *there exists* $k_N < \infty$ *such that* $<s_k, w> = 0$ *for all* $k > k_N$ *and for all words* $w \in Z^*$ *with length* $\|w\| < N$*. Then the infinite directed products* $\prod_{k=1}^{\longrightarrow\infty} s_k$ *and* $\prod_{k=1}^{\longleftarrow\infty} s_k$ *are well defined.*

Exercise 4.9 *Prove proposition* (4.7) *and using remark* 3.4

Corollary 4.8 *Suppose* $\{f_k\colon k \in \mathbf{Z}^+\} \subseteq\in \hat{L}(Z)$ *is a sequence of Lie series such that for every* $N < \infty$ *there exists* $k_N < \infty$ *such that* $<f_k, w> = 0$ *for all* $k > k_N$ *and for all words* $w \in Z^*$ *with length* $\|w\| < N$*. Then the infinite directed products* $\prod_{k=1}^{\longrightarrow\infty} e^{f_k}$ *and* $\prod_{k=1}^{\longleftarrow\infty} e^{f_k}$ *are well defined.*

Exercise 4.10 *Prove corollary* (4.8) *assuming proposition* (4.7) *and using remark 3.4*

4.3 Sussmann's exponential product expansion

This section demonstrates how to write the Chen Fliess series as an infinite exponential product. The approach follows the construction originally given by Sussmann [64] (but utilizing terminology form prior lectures in this series). Alternative constructions have been given using repeated differentiation and analysis of the derivatives [21], and by using entirely combinatorial and algebraic methods [48, 49, 54, 58]. The approach relies on repeatedly employing the method of *variation of parameters* from differential equations to develop a formula for the solution of the *universal control system* (103) The key strategy is to methodically match the recursive design with the Lazard elimination process, compare theorem 3.7. To improve the readability, we concentrate on the differential equations formulation (e.g. work directly with iterated integrals) and only state the analogous combinatorial formulas (in terms of formulas in $\hat{A}(Z) \otimes A(Z)$).

We begin with a review of some technical manipulations that are essential for the variation of parameters approach. The following formulas is one of the most useful and most often used elementary formulas in control. It is instructive to compare the algebraic and geometric/analytic proofs and definitions, and the quite different appearance. The following again just reiterates that in the *analytic setting* many apparently analytic arguments are really purely algebraic.

Proposition 4.9 *If* $x \in \hat{L}(Z)$ *is a Lie series and* $y \in \hat{A}(Z)$ *then*

$$e^x y e^{-x} = e^{\mathrm{ad}x} y = \sum_{k=0}^{\infty} \tfrac{1}{k!}(\mathrm{ad}^k x, y) \ = y + [x,y] + \tfrac{1}{2}[x,[x,y]] + \tfrac{1}{6}[x,[x,[x,y]] + \ldots \tag{122}$$

We will give a *formal differential equations* argument below. However, it is instructive to write out the first few terms by hand, to get a feeling how the words *combine into Lie polynomials.*

Exercise 4.11 *By formally expanding each exponential into its series (using the definition of the exponential), formally derive at least the first few terms of the formula (which is the form in which the previous formula is used most often in control)*

$$e^f g = (e^f g e^{-f}) e^f = (g + [f,g] + \tfrac{1}{2}[f,[f,g]] + \tfrac{1}{6}[f,[f,[f,g]]] + \ldots) e^f \tag{123}$$

This simple formula is very useful in control as differentiation of products of flows typically yields expressions like $e^{t_3 f_3} e^{t_2 f_2} f_2 e^{t_1 f_1} p$ (corresponding to a *variation* at $e^{t_1 f_1} p$ *transported* along the flows of first f_2, and then f_3 to the same terminal point $e^{t_3 f_3} e^{t_2 f_2} e^{t_1 f_1} p$ where different such tangent vectors are combined in the usual arguments of approximating cones yielding conditions for optimality and controllability). This means finding a formula for $\tilde{f}$ such that

$$e^{t_3 f_3} e^{t_2 f_2} f_2 e^{t_1 f_1} p = \tilde{f} e^{t_3 f_3} e^{t_2 f_2} e^{t_1 f_1} p \tag{124}$$

Clearly, f_2 commutes with $e^{t_2 f_2}$, but the tangent map of the third flow has an effect which is quantified by the above formula.

Exercise 4.12 (Control application) *Differentiate* (17) *with respect to each of the switching times* a, b, c *and* d, *and then use formula* (123) *to move each of the vector fields to the left of all exponentials, i.e. (geometrically) transport each vector back to the same point* $x(10, u)$.

Suppose $z \in \hat{L}(Z)$ is a Lie series with constant term $<z,1>=0$. Consider the *curve* $\gamma\colon \mathbb{R} \mapsto \hat{A}(Z)$ defined by $\gamma\colon t \mapsto e^{tZ}$. Define the *derivative* $\gamma'\colon \mathbb{R} \mapsto \hat{A}(Z)$ of γ at t as

$$\gamma'(t) = \tfrac{d}{dt} e^{tz} \overset{\text{def}}{=} e^{tz} z. \tag{125}$$

This allows an elegant formal derivation of (122), connecting algebra and geometry in a bootstrapping argument. Suppose $x \in \hat{L}(Z)$ is a Lie series and $y \in \hat{A}(Z)$. Consider the curve $\gamma\colon \mathbb{R} \mapsto \hat{A}(Z)$ defined by $\gamma\colon t \mapsto e^{tx} y e^{-tx}$ and differentiate.

$$\gamma'(t) = e^{tx} x y e^{-tx} + e^{tx} y e^{-tx}(-x) = e^{tx}(xy - yx)e^{-tx} = e^{tx}(\mathrm{ad} x, y)e^{-tx}. \tag{126}$$

using that $e^{tz} z = z e^{tz}$ for all $z \in \hat{L}(Z)$. Recursively, replacing y in above calculation by $(\mathrm{ad}^k x, y)$ one obtains

$$\left(\tfrac{d}{dt}\right)^k \bigg|_{t=0} e^{tx} y e^{-tx} = e^{tx}(\mathrm{ad}^k x, y)e^{-tx}\bigg|_{t=0} = (\mathrm{ad}^k x, y) \tag{127}$$

It is helpful to recall that in traditional notation in differential geometry, e.g. Spivak [57], the expression $e^{tf} g e^{-tf}$ is written as $\Phi_{t*} g$ where $(t, q) \mapsto \Phi_t(q)$ denotes the flow of the vector field f. The second exponential corresponds to the ever-present *inverse* in the *push-forward* (of a vector field), or in the *tangent map* (of the diffeomorphism) Φ_t, e.g. written as

$$(\Phi_{t*} g)(p) = \Phi_{t*\Phi_{-t}(p)} \left(g(\Phi_{-t}(p))\right) \tag{128}$$

says that the value of the vector field g pushed forward by the tangent map Φ_{t*} (bundle to bundle) is the same as the value of the vector field g at the *preimage* $\Phi_{-t}(p)$, i.e. the tangent vector $g(\Phi_{-t}(p))$, mapped forward by the tangent map $\Phi_{t*\Phi_{-t}(p)}$ from the fibre $T_{\Phi_{-t}(p)}M$ to the fibre $T_p M$. To complete the side-trip, recall the definition of the Lie derivative of a vector field g in the direction of a vector field f at a point p in terms of the flow Φ of f:

$$(L_f g)(p) = \lim_{t \to 0} \tfrac{1}{t} \Big(g(p) - (\Phi_{t*} g)(p)\Big) \tag{129}$$

which resounds well with the differential equations argument given above (127) (read backwards in the case of $k = 1$), written as $[f, g] = (\mathrm{ad} f, g) = \lim_{t \to 0} \frac{1}{t} \left(e^{tf} g e^{-tf}\right)$.

We are now ready to apply this formula in the variation of parameters approach that leads to Sussmann's exponential product expansion of the

Chen Fliess series. For illustration consider a two input system, i.e. a two-letter alphabet $Z = \{a, b\}$ and the solution curve $y(\cdot)$ taking values in $\hat{G}(\{a, b\}) \subseteq \hat{A}(\{a, b\})$. The *controls* $U'_a, U'_b \colon [0, t] \mapsto \mathbb{R}$ are assumed to be integrable.

$$y'(t) = y(t) \cdot (U'_a(t) \cdot a + U'_b(t) \cdot b) \tag{130}$$

Make the *Ansatz*

$$y(t) = y_1(t) \cdot e^{U_a(t)a} \;\text{ for some }\; y_1(\cdot) \in \hat{A}(Z) \tag{131}$$

Here $e^{U_a(t)a} \in \hat{A}(Z)$ may be thought of as the solution of the initial value problem $y'_1 = za$ with $y_1(0) = 1$ evaluated at *time* $U_a(t) = \int_0^t U'_a(s) ds$. Differentiate (131) and use (130) to obtain a differential equation for $y_1(\cdot)$

$$\left(y_1(t) \cdot e^{U_a(t)a}\right) \cdot (U'_a(t) \cdot a + U'_b(t) \cdot b) = y'_1(t) \cdot e^{U_a(t)a} + y_1(t) \cdot e^{U_a(t)a} \cdot (U'_a(t) \cdot a) \tag{132}$$

i.e. after collecting like terms

$$\begin{aligned} y'_1(t) &= y_1(t) \cdot \left(e^{U_a(t)a} \cdot (U'_a(t) \cdot a + U'_b(t) \cdot b) - e^{U_a(t)a} \cdot (U'_a(t) \cdot a)\right) \cdot e^{-U_a(t)a} \\ &= y_1(t) \cdot e^{U_a(t)a} \cdot (U'_b(t) \cdot b)) \cdot e^{-U_a(t)a} \\ &= y_1(t) \cdot \left(\sum_{k=0}^{\infty} \tfrac{1}{k!} U_a^k(t) U'_b(t) \cdot (\mathrm{ad}^k a, b)\right) \end{aligned} \tag{133}$$

Note that the resulting differential equation for $y_1(t)$ is of the same form as the original one (130) for $y(t)$, albeit now with an infinite linear combination of *control vector fields* $(\mathrm{ad}^k a, b)$. Important is that these are all elements of a Hall-basis for $L(\{a, b\})$, and

$$U'_{a^k b}(t) \overset{\text{def}}{=} \tfrac{1}{k!} \left(U_a(t)\right)^k \cdot U'_b(t) \tag{134}$$

plays a role as a *virtual* control associated to the *vector field* $(\mathrm{ad}^k a, b) \in L(Z)$. Iterating this process, we make the *Ansatz*

$$y_1(t) = y_2(t) \cdot e^{U_b(t)b} \;\text{ for some }\; y_2(\cdot) \in \hat{A}(Z) \tag{135}$$

Differentiate (135) and use (133) to obtain a differential equation for $y_2(\cdot)$

$$\left(y_2(t) \cdot e^{U_b(t)b}\right) \cdot \sum_{k=0}^{\infty} U'_{a^k b}(t) \cdot (\mathrm{ad}^k a, b) = y'_2(t) \cdot e^{U_b(t)b} + y_2(t) \cdot e^{U_b(t)b} \cdot (U'_b(t) \cdot b) \tag{136}$$

which after collecting like terms becomes

$$\begin{aligned} y_2'(t) &= y_2(t)\cdot\Big(e^{U_b(t)b}\cdot(\sum_{k=0}^{\infty} U'_{a^kb}(t)\cdot(\mathrm{ad}^k a, b)) - e^{U_b(t)b}\cdot(U_b'(t)\cdot b)\Big)\cdot e^{-U_b(t)b} \\ &= y_2(t)\cdot e^{U_b(t)b}\cdot\Big(\sum_{k=1}^{\infty} U'_{a^kb}(t)\cdot(\mathrm{ad}^k a, b))\Big)\cdot e^{-U_b(t)b} \\ &= y_2(t)\cdot\Big(\sum_{\ell=0}^{\infty}\sum_{k=1}^{\infty}\tfrac{1}{\ell!}\tfrac{1}{k!}U_b^{\ell}(t)U'_{a^kb}(t)\cdot(\mathrm{ad}^{\ell} b,(\mathrm{ad}^k a, b)\Big) \end{aligned} \tag{137}$$

The resulting differential equation for $y_2(t)$ is again of the same form, now with a doubly-infinite linear combination of *control vector fields* $(\mathrm{ad}^{\ell} b,(\mathrm{ad}^k a, b))$. Again, these are all elements of a Hall-basis for $L(\{a,b\})$, and

$$U'_{b^{\ell}a^kb}(t) = \tfrac{1}{\ell!}U_b^{\ell}(t)\cdot U_{a^kb}(t) \tag{138}$$

plays the role of a *virtual* control associated to $(\mathrm{ad}^{\ell} b(\mathrm{ad}^k a, b)) \in L(Z)$.

This process may be iterated infinitely many times. The critical choice at each step is the selection of the *smallest* Hall element among the *control fields*

$$(\mathrm{ad}^{k_n} h_n, (\mathrm{ad}^{k_{n-1}} h_{n-1}, \ldots (\mathrm{ad}^{k_3} h_3, (\mathrm{ad}^{k_2} h_2, h_1))\ldots)) \tag{139}$$

in the previous stage. This will assure that in the next step again all *control vector fields* will be of the same form, again with $h_{j+1} > h_j$ for all j. By virtue of the Lazard elimination process 3.7 they will again be Hall elements. Combining the *Ansätze* of all steps yields

$$y(t) = y_1(t)e^{U_a(t)a} = y_2(t)e^{U_b(t)b}e^{U_a(t)a} = y_3(t)e^{U_{ab}(t)ab}e^{U_b(t)b}e^{U_a(t)a} = \ldots \tag{140}$$

For Hall sets (over finite alphabets) whose ordering is compatible with the lengths of the words (i.e. $\|h\| < \|h'\|$ implies $h \prec h'$, such as in the definition of Philipp Hall bases given in Bourbaki [4]) it is easily seen that this iterative process converges in the topology considered here (compare remark 3.4). However, a little reflection shows that such formula actually also holds for infinite alphabets, and also for Hall sets whose ordering is not necessarily compatible with the length (e.g. recall that Lyndon bases have *infinite segments*).

We refer the reader to the original references [64], also [35], for detailed proofs, and only state the main result as it applies to analytic control systems on finite dimensional manifolds such as those encountered in the examples of the first chapter:

Theorem 4.10 (Sussmann [64]) *Suppose* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ *is a Hall set,* f_a, $a \in Z$, *are analytic vector fields on a manifold* M, *and* u_a, $a \in Z$, *are measurable bounded controls defined on* $[0, \infty]$. *Then for every compact set* $K \subseteq M$ *there exists* $T > 0$ *such that for all initial conditions* $x_0 \in K$ *the infinite directed exponential product*

$$s(t) = \overrightarrow{\prod}_{h \in \mathcal{H}} e^{U_h(t)\, f_{[h]}} \tag{141}$$

converges for $0 \leq t \leq T$ *uniformly on* K *to the solution of the control system*

$$\dot{x} = \sum_{a \in Z} u_a f_a(x), \quad x(0) = x_0 \tag{142}$$

Here $f_{[h]} = \mathcal{F}([h])$ *is the vector field on* M *that is obtained from the commutator* $[h] \in L(Z)$ *by* substituting *the control vector fields* f_a *for the letters* $a \in Z$ *in* $h \in \mathcal{H}$. *The iterated integrals* U_h *for Hall words* $h \in \mathcal{H} \setminus Z$ *satisfy*

$$U_h(t) = \tfrac{1}{k!} \cdot \int_0^t U_{h_1}^k(\tau) U'_{h_2}(\tau)\, d\tau \tag{143}$$

if $h = h_1^k h_2 \in \mathcal{H}$, $h_1 > h_2$ *and either* $h_2 \in Z$ *is a letter, or the* left factor *of* h_2 *is strictly smaller than* h_1.

Exercise 4.13 *Consider the nilpotent control system* $\dot{x_1} = u_1$, $\dot{x_2} = u_2$, *and* $\dot{x_3} = x_2 u_1$ *on* $\mathbb{R}^3$. *Identify the system vector fields* f_1 *and* f_2, *and explicitly write out the images* $f_{[h]}$ *of a suitable Hall-basis for* $L(\{1, 2\})$, *the associated iterated integrals* $U_h(t)$ *and the* flows *in the product* (141). *Verify that the solutions* $x(t, u)$ *agree with the products of the flows as in the theorem.*

Exercise 4.14 *Repeat the previous exercise for the nilpotent control system* $\dot{x_1} = u$, $\dot{x_2} = x_1^p$ *on* $\mathbb{R}^2$, *for* $p = 2, 3$ *or any other integer.*

Exercise 4.15 *Returning to the previous exercise in the case of* $p = 2$ *investigate how the choice of* $0 < 1$ *or* $1 < 0$ *in the Hall set over* $Z = \{0, 1\}$ *affects the structure of the iterated integrals* U_h.

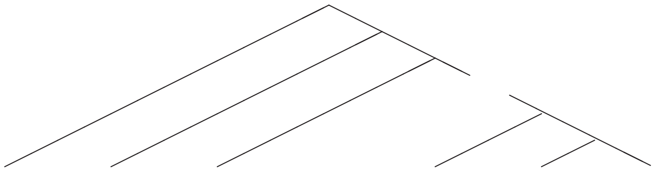
The essence in the combinatorial analogue of the theorem 4.10 is captured in the formula for the elements ξ_v of the dual basis for $A(Z)$ indexed by elements $v \in \mathcal{P}$ of a Poincaré-Birkhoff-Witt basis $\mathcal{P}$ for $A(Z)$ that is built on a Hall set, or Hall basis for $L(Z)$. We first give a technical definition.

Definition 4.7 *Suppose* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ *is a Hall set. Define the function* $\mu \colon Z^* \mapsto \{\frac{1}{n} \colon n \in \mathbf{Z}^+\}$ *by*

$$\mu(wa) = \mu(w) = \frac{1}{m_1! m_2! \cdot m_{s-1}! m_s!} \tag{144}$$

if $a \in Z$ *and* $h = wa \in \mathcal{H}$ *factors uniquely into* $h = h_1^{m_1} h_2^{m_2} \cdots h_{s-1}^{m_{n-1}} h_n^{m_n} a$ *with* $a \in Z$, $m_i \in \mathbf{Z}^+$, $h_i \in \mathcal{H}$ *and* $h_1 > h_2 > \ldots > h_{s-1} > h_s < a$ (compare theorem (3.11)).

Theorem 4.11 *Suppose* $\hat{\mathcal{H}} \subseteq \mathcal{M}(Z)$ *is a Hall set and the Hall word* $h \in \mathcal{H} \subseteq Z^*$ *factors uniquely into Hall words* $h = h_1 h_2 \cdots h_{n-1} h_n a$ *with* $a \in Z$ *and* $h_i \in \mathcal{H}$ *and* $h_1 \geq h_2 \geq \ldots \geq h_n < a$. *Then*



$$\xi_h = \frac{1}{\mu(h)} (\xi_{h_1} * \quad (\xi_{h_2} * \quad (\xi_{h_3} * (\ldots \qquad (\xi_{h_{s-2}} * \quad (\xi_{h_{s-1}} * \quad \xi_a))...)))) \tag{145}$$

Figure 7. Structure of the dual Poincaré-Birkhoff-Wittbases for Hall trees

Compare the unique factorization theorem 3.11 for Hall Viennot words and the similar figure 6. Note also the close correspondence to theorem 3.6, that while $\varrho_a^\dagger$ and $\lambda_a^\dagger$ are derivations on the shuffle algebra $(\hat{A}(Z), ш)$, *only* the transpose $\lambda_a^\dagger$ of the left translation λ_a by a letter a is a derivation on the chronological algebra $(C(Z), *)$.

One way to establish theorem 4.11 as a consequence of theorem 4.10 is to use that the map Υ is a chronological algebra isomorphism from the free chronological algebra $C(Z)$ onto the space of $\mathcal{IIF}(\mathcal{U})$ of iterated integral functionals – provided the space $\mathcal{U}$ of admissible controls is sufficiently large (compare [35]). An alternative proof of a purely combinatorial nature was given by Melancon and Reutenauer [48, 49]. Essentially the same formula may also be found in Schützenberger [58] and Grayson and Grossman [21].

Introducing left and right chronological powers not only facilitates the writing, but in some cases it may *make factorials disappear*. More specifically, rewriting formulas with another product may result in the disappearance of factorials, e.g. via the Taylor expansions of $\frac{1}{1-x}$ with one product, becomes the Taylor expansion e^x with respect to another product.

Definition 4.8 *For* $w \in A^+(Z) = A(Z^+)$ *define* $w^{*1} = \lambda^1(w) = w^{\sqcup\!\sqcup\, 1} = w$, *and inductively for* $n \geq 1$ (sometimes it is convenient also allow $w^{*0} = 1 = w^{\sqcup\!\sqcup\, 0}$ and $\lambda^0(w) = 0$)

$$\begin{array}{rcl} \lambda^{n+1}(w) & = & w * \lambda^n(w) \\ w^{*(n+1)} & = & w^{*n} * w \\ w^{\sqcup\!\sqcup\,(n+1)} & = & w \sqcup\!\sqcup\, w^{\sqcup\!\sqcup\, n} = w^{\sqcup\!\sqcup\, n} \sqcup\!\sqcup\, w \end{array}$$

Proposition 4.12 *For* $w \in A^+(Z) = A(Z^+)$ *and* $n \in \mathbf{Z}^+$ *the following identities hold:*

$$\begin{array}{rcll} w * w^{*(n-1)} & = & (n-1) \cdot w^{*n} & \\ \lambda^n(w) & = & (n-1)! \cdot w^{*n} & \\ w^{\sqcup\!\sqcup\, n} & = & n! \cdot w^{*n} & (= n\lambda^n(w)) \end{array} \tag{146}$$

Exercise 4.16 *Prove the identities in proposition* 4.12*.* (Take advantage of bilinearity and first prove the identities, by induction on n, for words $w \in Z^+$.)

For the sake of completeness we also note the structure of the complete dual Poincaré-Birkhoff-Witt bases built on Hall sets. A complete combinatorial proof is given in [48], also see [54]. Alternatively, use the chronological algebra isomorphism Υ to obtain this result from theorem 4.10.

Proposition 4.13 *Suppose* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ *is a Hall set and* $\mathcal{P}$ *the associated Poincaré-Birkhoff-Witt basis for* $A(Z)$*. If* $w = [h_1]^{m_1}[h_2]^{m_2}\ldots[h_n]^{m_n} \in \mathcal{P}$ *for Hall words* $h_i \in \mathcal{H}$ *with* $h_i > h_{i+1}$*, then the dual basis elements are*

$$\tilde{\xi}_v = \frac{1}{m_1!m_2!\cdots m_n!} \cdot \xi_{h_n}^{\sqcup\!\sqcup\, m_n} \sqcup\!\sqcup\, \xi_{h_{n-1}}^{\sqcup\!\sqcup\, m_{n-1}} \sqcup\!\sqcup\, \ldots \sqcup\!\sqcup\, \xi_{h_2}^{\sqcup\!\sqcup\, m_2} \sqcup\!\sqcup\, \xi_{h_1}^{\sqcup\!\sqcup\, m_1} \tag{147}$$

We conclude the section with a series of challenge exercises that aim at bridging the combinatorial and analytical / differential equations arguments that culminate in theorems 4.10 and 4.11.

Exercise 4.17
Verify that the product $*\colon (A(Z) \otimes \hat{A}(Z)) \times (A(Z) \otimes \hat{A}(Z)) \mapsto (A(Z) \otimes \hat{A}(Z))$ *defined by*

$$(w \otimes f) * (z \otimes g) = (w * z) \otimes (fg) \tag{148}$$

is a chronological product, i.e. it satisfies the right chronological identity (67)*.*

Exercise 4.18 *Rewrite the universal control system* (103) *as an equation on* $A(Z) \otimes \hat{A}(Z)$ *using the chronological product from the previous exercise.*

Exercise 4.19 *Capture the combinatorial essence of the* variation of parameters *technique using chronological products – first rewrite the differential equation, e.g.* (130) *as an equivalent integral equation, and then use chronological products.*

4.4 Free nilpotent systems

A simple way to state and remember, and a very useful application for the formula (145) in theorem 4.11 is as a normal form for free nilpotent systems. Recall that nilpotent control systems are systems of the form (8) for which the Lie algebra $L(f_0, f_1, \ldots f_m)$ generated by the system vector fields is nilpotent. Via a local coordinate change they can always be brought into a form in which the vector fields are polynomial and have a *cascade* structure. Such systems are sufficiently rich that they have good approximation properties (controllability, stabilizability etc.), and they are very manageable: E.g. solution curves can be computed by simple quadratures, requiring no intractable solution of nonlinear differential equations.
A natural objective is to write down a canonical form for the most general such system (up to a certain order). However, any naive try such as the one starting with

$$\left\{\begin{array}{lllllllll} \dot{x}_1 & = & u & \dot{x}_5 & = & x_4 & \dot{x}_9 & = & x_1^3 \\ \dot{x}_2 & = & x_1 & \dot{x}_6 & = & x_5 & \dot{x}_{10} & = & x_9 \\ \dot{x}_3 & = & x_2 & \dot{x}_7 & = & x_2^2 & \dot{x}_{11} & = & x_1^2 x_2 \\ \dot{x}_4 & = & x_1^2 & \dot{x}_8 & = & x_1 x_3 & \dot{x}_{12} & = & x_1^4 \end{array}\right. \tag{149}$$

does not do the job as the system is not accessible due to redundant terms: Along every solution curve the function $\Phi(x) = x_7 + x_8 - x_2 x_3$ is constant.

Exercise 4.20 *Verify by direct calculation of the iterated Lie brackets of the system vector fields that the system* (149) *does not satisfy the Lie algebra rank condition for accessibility.*

To make precise what we mean by a (maximally) free nilpotent system, define:

Definition 4.9 *For any integer* $r > 0$ *define* $L^{(r)}(Z)$ *to be quotient of* $L(Z)$ *by the ideal* $L(Z) \cap \cup_{k=r+1}^{\infty} A^{(k)}(Z)$.

Definition 4.10 *Suppose* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ *is a Hall set and* $H = \psi(\mathcal{H}) \subseteq Z^*$ *is the set of corresponding Hall words. Define* $\mathcal{H}^{(r)} \overset{\mathrm{def}}{=} \{h \in \mathcal{H} \colon |h| \leq r\}$ *to be the subset of Hall words of length at most* r.

Definition 4.11 *Suppose* $\tilde{\mathcal{H}} \subseteq \mathcal{M}(Z)$ *is a Hall set and* $r > 0$. *The* normal form of the free nilpotent system determined by $\mathcal{H}^{(r)}$ *is the control system*

$$
\begin{array}{rcll}
\dot{x}_a & = & u_a & \text{if } a \in Z \\
x_{wz} & = & x_w * x_z & \text{if } w,\, z,\, wz \in \mathcal{H}^{(r)} \subseteq Z^*
\end{array}
$$

Theorem 4.14 *The Lie algebra* $L(\{f_a \colon a \in Z\})$ *generated by the vector fields of the system* (4.11) *(written in the form* $\dot{x} = \sum_{a \in Z} u_a f_a(x)$*) is free nilpotent of step* r.

Example: A normal from for a free nilpotent system (of rank $r = 5$) using a typical Hall set on the alphabet $Z = \{0, 1\}$ is

$$
\begin{array}{rclcllrcl}
\dot{x}_0 & = & u_0 & & & & & & \\
\dot{x}_1 & = & u_1 & & & & & & \\
\dot{x}_{01} & = & x_0 \cdot \dot{x}_1 & = & x_0\, u_1 & & & & \\
\dot{x}_{001} & = & x_0 \cdot \dot{x}_{01} & = & x_0^2\, u_1 & \text{from} & \psi^{-1}(001) & = & (0(01)) \\
\dot{x}_{101} & = & x_1 \cdot \dot{x}_{01} & = & x_1 x_0\, u_1 & \text{from} & \psi^{-1}(101) & = & (1(01)) \\
\dot{x}_{0001} & = & x_0 \cdot \dot{x}_{001} & = & x_0^3\, u_1 & \text{from} & \psi^{-1}(0001) & = & (0(0(01))) \\
\dot{x}_{1001} & = & x_1 \cdot \dot{x}_{001} & = & x_1 x_0^2\, u_1 & \text{from} & \psi^{-1}(1001) & = & (1(0(01))) \\
\dot{x}_{1101} & = & x_1 \cdot \dot{x}_{101} & = & x_1^2 x_0\, u_1 & \text{from} & \psi^{-1}(1101) & = & (1(1(01))) \\
\dot{x}_{00001} & = & x_0 \cdot \dot{x}_{001} & = & x_0^4\, u_1 & \text{from} & \psi^{-1}(00001) & = & (0(0(0(01)))) \\
\dot{x}_{10001} & = & x_1 \cdot \dot{x}_{0001} & = & x_1 x_0^3\, u_1 & \text{from} & \psi^{-1}(10001) & = & (1(0(0(01)))) \\
\dot{x}_{11001} & = & x_1 \cdot \dot{x}_{1001} & = & x_1^2 x_0^2\, u_1 & \text{from} & \psi^{-1}(11001) & = & (1(1(0(01)))) \\
\dot{x}_{01001} & = & x_{01} \cdot \dot{x}_{001} & = & x_{01} x_0^3\, u_1 & \text{from} & \psi^{-1}(01001) & = & ((01)(0(01))) \\
\dot{x}_{01101} & = & x_{01} \cdot \dot{x}_{101} & = & x_{01} x_1^2 x_0\, u_1 & \text{from} & \psi^{-1}(01101) & = & ((01)(1(01)))
\end{array}
\tag{150}
$$

Remark 4.15 *It is noteworthy, and almost essential for effective calculations that the coordinates* x_h *are indexed by Hall words, rather than by consecutive natural numbers!*
On the other hand, by virtue of the unique factorization theorem 3.11, *one may use the Hall words as indices, and* does not *to use trees or* parenthesized *words (which would make for very cumbersome notation).*

Exercise 4.21 *Write the system* (150) *in the form* $\dot{x} = u_0 f_0(x) + u_1 f_1(x)$*, and calculate iterated Lie brackets of* f_0 *and* f_1 *of length at most* 5 *(using the same Hall set). Verify that for each such iterated Lie bracket* $f_w = \mathcal{F}(w)$*, its value* $f_w(0)$ *at* $x = 0$ *is a multiple of the corresponding coordinate direction* $\frac{\partial}{\partial x_w}\Big|_0$

Exercise 4.22 *Use a different Hall-Viennot basis for* $Z = \{0, 1\}$ *(e.g. the Lyndon basis from figure* 5*), to construct a different representation of the free nilpotent system* (150)*.*
Demonstrate that these systems are equivalent under a global, polynomial coordinate change.

Exercise 4.23 *Use a Hall-Viennot basis for* $Z = \{0, 1, 2\}$ *to construct an explicit coordinate representation similar to* (150) *for a free nilpotent (of order* 4*) two-input system with drift* $\dot{x} = f_0(x) + u_1 f_1(x) + u_2 f_2(x)$*.*

Exercise 4.24 *Prove that the Lie algebra* $L(\{f_a : a \in Z\})$ *of the vector fields of the system* (4.11) *is nilpotent.* (Introduce a suitable family of dilations so that all vector fields are homogeneous of strictly negative order.)

Exercise 4.25 *Prove that the Lie algebra* $L(\{f_a : a \in Z\})$ *of the vector fields of the system* (4.11) *is isomorphic to* $L^{(r)}(Z)$*.* (Demonstrate that the Lie algebra has maximal dimension by showing that $\mathcal{F}([h])(0) = c_h \frac{\partial}{\partial x_h}(0)$ for some $c_h \neq 0$.)

On the side we mention another useful number, the dimensions of the *homogeneous* components $L^{(\alpha)}(Z) = L(Z) \cap A^{(\alpha)}(Z)$ of the free Lie algebra $L(Z)$. Here $A^{(\alpha)}(Z)$ is the linear span of all words containing α_a times the letter a (for each $a \in Z$). First recall the Moebius function from enumerative combinatorics [4]:

Definition 4.12 *The Moebius function* $\mathrm{Moe}\colon \mathbf{Z}^+ \mapsto \{-1, 0, 1\}$ *is defined by* $\mathrm{Moe}(n) = 0$ *if* n *is divisible by the square of a prime, and else* $\mathrm{Moe}(n) = (-)^s$ *if* s *is the number of distinct prime factors of* n*.*

The Moebius function can also be characterized as the unique function from $\mathbf{Z}^+$ to $\{-1, 0, 1\}$ such that $\mathrm{Moe}(1) = 1$ and

$$\sum_{d|n} \mathrm{Moe}(d) = 0 \quad \text{for all } n \in \mathbf{Z}^+ \tag{151}$$

Proposition 4.16
Suppose $\alpha_a \geq 0$ *for* $a \in Z$ *are nonnegative integers. Then the dimension of* $L^{(\alpha)}(Z)$ *is*

$$\dim L^{(\alpha)}(Z) = \frac{1}{|\alpha|} \sum_{d|\alpha} \mathrm{Moe}(d) \frac{(|\alpha|/d)!}{(\alpha/d)!} = \frac{1}{\sum_{a \in Z} \alpha_a} \sum_{d|\alpha} \mathrm{Moe}(d) \frac{(\sum_{a \in Z} \alpha_a)/d)!}{\prod_{a \in Z} (\alpha_a/d)!} \tag{152}$$

Exercise 4.26 *Calculate, and tabulate, the dimensions of the homogeneous components* $L^{(\alpha)}(\{a,b\})$ *for* $|\alpha| \leq 6$.

Exercise 4.27 *Calculate, and tabulate, the dimensions of the homogeneous components* $L^{(\alpha)}(\{a,b,c\})$ *for* $|\alpha| \leq 4$.

We conclude with a few comments about the path planning problem, which given a system of form (8) and two points $p, q \in \mathbb{R}^n$ in the state space, asks for a control u, defined on some time interval $[0,T]$ which steers the system from $x(0) = p$ to $x(T,u) = q$. For a detailed discussion and a variety of results see e.g. [24, 25, 41, 51, 52, 66].

A reasonably tractable class consists of nilpotent systems (possibly used as approximating systems, yielding approximate path planning results). One of the most useful features of free nilpotent systems is that they can provide *universal* solutions to the path planning problems, as any specific nilpotent system *lifts* to a free system. In other words, the trajectories of the free system map, or project to the trajectories of the specific system. Thus the general solution of the problem for the free system yields also a (many) solutions(s) for the specific problem.

More specifically, suppose $\Sigma\colon \dot{x} = \sum_{i=1}^{m} u_i f_i(x)$ is a specific system such that $L(f_1, \ldots f_m)$ is nilpotent of order r. Then let $\Sigma^{(r)}\colon \dot{x} = \sum_{i=1}^{m} u_i F_i(x)$ be a free nilpotent system (of order r) on $\mathbb{R}^N$. Due to the *freeness* there exists a smooth map $\Phi\colon \mathbb{R}^N \mapsto \mathbb{R}^n$ that maps trajectories of $\Sigma^{(r)}$ to trajectories of Σ corresponding to the same controls. Thus in order to steer the system Σ from $p \in \mathbb{R}^n$ to $q \in \mathbb{R}^n$ one may use any control of the presumedly solved path planning problem steering $\Sigma^{(r)}$ from any $P \in \Phi^{-1}(p) \subseteq \mathbb{R}^N$ to any $Q \in \Phi^{-1}(q) \subseteq \mathbb{R}^N$.

This short discussion justifies that one take a closer look at the general path planning problem for free nilpotent systems. For systems with nonzero drift the possible lack of controllability remains a formidable obstacle to a general solution. Thus here we only take a brief look at systems without drift (for which accessibility is the same as controllability). For practical solutions, a key step is to reduce the problem from the very large space $\mathcal{U}$ of all possible controls to smaller sets, typically finite dimensional subspaces. Typical examples include those spanned by trigonometric polynomials (with fixed base frequency and maximal order), polynomial controls, and piecewise constant or piecewise polynomial controls.

For illustration, in the following exercises, calculate the iterated integrals $U_h(T)$ for $h \in \mathcal{H}^{(r)}(\{0,1\})$ with $r = 3, 4$ or 5 (depending on available computer algebra system resources) for the specified parameterized families of controls u_α. Note that this is tantamount to calculating the trajectories of the system (150) for $p = 0$.
Then analyze the nature of the inverse problem of finding the control parameterized by α that yields a given target point $Q = (U_h(T))_{h \in \mathcal{H}^{(r)}(\{0,1\})} = (U_0(T), U_1(T), U_{10}(T), U_{110}(T), \ldots)$.

Exercise 4.28 *Consider polynomial controls (e.g.* $(u_0, u_1)(t) = (\alpha_1 + \alpha_2 t + \alpha_3 t^2, \alpha_4 + \alpha_5 t + \alpha_6 t^2)$*). – Use as many parameters as the dimension of the subsystem you are working with. Is the map from* α *to the endpoint (or sequence of iterated integrals) injective? invertible? What problems do you see as the dimension increases?*

Exercise 4.29 *Consider controls that are trigonometric polynomials (e.g.* $(u_0, u_1)(t) = (\alpha_1 \cos t + \alpha_2 \cos 2t + \alpha_3 \cos 3t, \alpha_4 \sin t + \alpha_5 \sin 2t + \alpha_6 \sin 3t)$*. Use as many parameters as the dimension of the subsystem you are working with. Is the map from* α *to the endpoint (or sequence of iterated integrals) injective? invertible? Contrast this map with the analogous map for linear systems! What problems do you see as the dimension increases?*

These exercises open many new question in an area that still leaves a lot to be explored. One suggestion for exploration is, instead of considering the full free nilpotent system, to restrict one's attention to some special class of system – e.g. one class of systems that has been popular in the 1990s is that of systems in *chain-form*, compare e.g. [67].

Acknowledgments

Supported in part by NSF-grant DMS 00-72369. The author greatly appreciates and the support and hospitality of the Abdus Salam Institute of

Theoretical Physics and of SISSA, who provided this fantastic opportunity to interact with this new generation of bright scientists, and who provided the locality for advancing my own research through continuous daily interactions with the best of my peers from around the world.

My special thanks go to the organizers of this summer school, A. Agrachev, B. Jakubczyk, and C. Lobry. I am most thankful for the continuous encouragement and support by all participants, especially the unusually diverse group of student-participants without whom these lectures would not have happened, without whom these notes would not exist, and who greatly contributed through excellent observations, conjectures and questioning.

With such a superb new generation of geometric control scientists becoming ready to lead the world, I am excited that this wonderful subject will not only stay alive, but prosper in the future, and contribute to making this an even better world.

References

[1] A. Agrachev and R. Gamkrelidze, *Exponential representation of flows and chronological calculus*, Math. USSR Sbornik (Russian) **107**, N4 (1978), 487–532. Math. USSR Sbornik, **35** (1978) 727–

[2] A. Agrachev and R. Gamkrelidze, *Chronological algebras and nonstationary vector fields*, Journal Soviet Math. **17**, no.1 (1979), 1650–1675.

[3] A. Bloch and H. McClamroch, "Controllability and stabilizability properties of nonholonomic control systems", Proc. of 29th IEEE CDC (1990) 1312–1317.

[4] N. Bourbaki, Lie Groups and Lie algebras, (1989) (Springer).

[5] R. Brockett, *Differential geometric methods in system theory*, in Proc. IEEE Conf.Dec.Cntrl. (1971) 176-180.

[6] K. T. Chen, *Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula*, Annals of Mathematics **65** (1957) 163–178.

[7] J.-M. Coron, *On the stabilization in finite time of locally controllable systems by means of continuous time-varying feedback law*, SIAM J. Cntrl. & Opt., **33** (1995) 804-833.

[8] P. Crouch, *Solvable Approximations to Control Systems*, SIAM J. Cntrl. & Opt., **22** (1984) 40–45.

[9] P. Crouch and F. Lamnabhi-Lagarrigue, *Algebraic and multiple integral identities*, Acta Appl. Math. **15** (1989) 235–274.

[10] P. Crouch and R. Grossman, *The explicit computation of integration algorithms and first integrals for ordinary differential equations with polynomial coefficients using trees*, Proc. Int. Symposium on Symbolic and Algebraic Computation, ACM Press (1992) 89-94.

[11] P. Crouch and F. Leite, *the dynamical interpolation problem: on Riemannian manifolds, Lie groups and symmetric spaces*, J. Dynam.Control Systems **1** (1995) 177 - 202.

[12] A. Dzhumadil'daev, *Non-associative algebras without unit*, Comm. Algebra (2002) (to appear).

[13] A. Dzhumadil'daev, *Trees, free right-symmetric algebras, free Novikov algebras and identities*, Homotopy, Homology and Applications (2001) (to appear).

[14] M. Fliess, *Fonctionelles causales nonlinéaires et indeterminées noncommutatives*, Bull. Soc. Math. France **109** (1981) 3–40.

[15] K. Grasse, *On the relation between small-time local controllability and normal self-reachability* Math. Control Signals and Systems, **5** (1992) 41 – 66.

[16] H. Frankowska, *An open mapping principle for set-valued maps*, J.Math.Anal.Appl. **127** (1987) 172–180.

[17] H. Frankowska, *Local controllability of control systems with feedback*, J.Opt.Theory Applics. **60** (1989) 277–296.

[18] H. Frankowska, *A conical open mapping principle for set-valued maps*, Bull.Australian Math.Soc **45** (1992) 53–60.

[19] I. Gelfand, D. Raikov, and G. Shilov, *Commutative normed rings*, **1991** (1964 Chelsea).

[20] V. Ginzburg and M. Kapranov, *Koszul duality for operads*, Duke Math.J. **76** (1994) 203–272.

[21] M. Grayson and R. Grossman, *Models for free nilpotent algebras*, J. of Algebra, **135** (1990) 177–191.

[22] H. Hermes, *Controlled Stability*, Annali di Matematica pura ed applicata IV, **CXIV** (1977) 103-119.

[23] H. Hermes, *Nilpotent and high-order approximations of vector field systems*, SIAM Review **33** (1991) 238–264.

[24] G. Jacob, *Lyndon discretization and exact motion planning*, Proc. Europ. Control Conf., Grenoble (1991), pp. 1507–1512.

[25] G. Jacob, *Motion planning by piecewise constant or polynomial inputs*, Proc. NOLCOS, Bordeaux (1992).

[26] V. Jurdjevic and H. Sussmann, *Controllability of nonlinear systems*, J. Diff. Eqns. **12** (1972) 95–116.

[27] V. Jurdjevic and H. Sussmann, *Control systems on Lie groups*, J. Diff. Eqns. **12** (1972) 313–329.

[28] M. Kawski, *A new necessary condition for local controllability*, AMS Contemporary Mathematics **68** (1987) 143–156.

[29] *An angular open mapping theorem*, in: Analysis and Optimization of Systems, A. Bensoussan and J. L. Lions, eds., Lect. Notes in Control and Information Sciences **111** (1988) 361-371.

[30] M. Kawski, *Control variations with an increasing number of switchings*, Bull.Amer.Math.Soc. **18** (1988) 149-152.

[31] M. Kawski, *Nilpotent Lie algebras of vector fields*, J. Reine & Angewandte Mathematik, **188** (1988) pp. 1-17.

[32] M. Kawski, M. Kawski, *High-order small-time local controllability*, in: Nonlinear Controllability and Optimal Control, H. Sussmann, ed. (1990) 441-477 (Dekker)

[33] M. Kawski, *Geometric homogeneity and applications to stabilization* in: Nonlinear Control Systems Design, A.Krener and D.Mayne, eds., (1995), pp.147-152.

[34] M. Kawski, *Chronological algebras and nonlinear control*, Proc. Asian Conf. Control, Tokyo, 1994.

[35] M. Kawski and H. J. Sussmann *Noncommutative power series and formal Lie-algebraic techniques in nonlinear control theory*, in: Operators, Systems, and Linear Algebra, U. Helmke, D. Prätzel-Wolters and E. Zerz, eds., Teubner (1997), 111–128.

[36] M. Kawski, *Controllability via chronological calculus*, Proc. 38th IEEE Conf.Dec.Contr. (1999) 2920-2925.

[37] M. Kawski, *Calculating the logarithm of the Chen Fliess series*, submitted to Int. J.Cntrol (special issue Proc. MTNS 2000 Perpignan).

[38] M. Kawski, *Chronological algebras: combinatorics and control*, Itogi Nauki i Techniki, vol.68 (2000) pp.144-178. (translation in J. Math. Sciences).

[39] M. Kawski, *Controllability and coordinates of the first kind*, in: Contemporary Trends in Non-linear Geometric Control Theory and its Applications, eds. A. Anzaldo-Meneses, B. Bonnard, J.P. Gauthier, F. Monroy-Perez. (World Sci. Publ.) (to appear)

[40] H. Knobloch, *Higher Order Necessary Conditions in Optimal Control Theory*, Lect. Notes in Control and Information Sciences **34** (1981) (Springer)

[41] G. Laffarière and H. J. Sussmann, *Motion planning for controllable systems without drift*, IEEE Conf. Robotics and Automation, (1991), pp. 1148–1153.

[42] J.-L. Loday, *Une version non commutative des algèbres de Lie: les algèbres de Leibniz*, L'Enseignement Mathématique **39** (1993), pp. 269–293.

[43] J.-L. Loday and T. Pirashvili, *Universal enveloping algebras of Leibniz algebras and (co)homology*, Math. Annalen **196** (1993), pp. 139–158.

[44] J.-L. Loday and T. Pirashvili, *Leibniz representations of Lie algebras*, J. Algebra **181** (1996), pp. 414–425.

[45] J.-L. Loday, *Cup-product for Leibniz cohomology and dual Leibniz algebras*, Math. Scand. 77 (1995), no. 2, pp. 189–196.

[46] M. Lothaire, *Combinatorics on Words*, Addison-Wesley, Reading, Mass., (1983).

[47] W. Magnus, *On the exponential solution of differential equations for a linear operator*, Copmm. Pure Appl. Math. **VII** (1954), 649–673.

[48] G. Melançon and C. Reutenauer, *Lyndon words, free algebras and shuffles*, Canadian J. Math. **XLI** (1989), 577–591.

[49] G. Melançon and C. Reutenauer C., *Combinatorics of Hall trees and Hall words*, J. Comb. Th. Ser. A **59** (1992), 285–299.

[50] H. Munthe-Kaas and B. Owren, *Computations in a Free Lie Algebra*, report no. 148, Dept. Informatics, Univ. Bergen, 1998.

[51] R. Murray and S. Sastry, *Nonholonomic path planning: steering with sinusoids*, IEEE T. Automatic Control, **38** (1993) 700–716.

[52] R. Murray, *Nilpotent bases for a class of nonintegrable distribution swith applications to trajectory generation for nonholonomic systems*, Math. Controls, Signals, & Systems, **7** (1994) 58–75.

[53] R. Ree, *Lie elements and an algebra associated with shuffles*, Annals of Mathematics **68** (1958) 210–220.

[54] C. Reutenauer, *Free Lie algebras*, (1993), Oxford (Clarendon Press).

[55] E. Rocha, *On computation of the logarithm of the Chen-Fliess series for nonlinear systems*, Proc. NCN Sheffield (2001).

[56] L. Rosier, *Homogeneous Lyapunov functions for homogeneous continuous vector fields*, Systems and Control Letters, **19** (1992) 467 – 473.

[57] M. Spivak, A comprehensive introduction to differential geometry (5 volumes). Publish or Perish.

[58] M. Schützenberger, *Sur une propriété combinatoire des algèbres de Lie libres pouvant être utilisée dans un probléme de mathématiques appliquées*, Séminaire P. Dubreil, Algèbres et Théorie des Nombres, Faculté des Sciences de Paris (1958/59).

[59] G. Stefani, *On the local controllability of a scalar-input system*, in "Theory and Applications of Nonlinear Control Systems," C. I. Byrnes and A. Lindquist eds., Elsevier, North-Holland (1986) 167–179.

[60] G. Stefani, *Polynomial approximations to control systems and local controllability* Proc. 25th IEEE CDC (1985) 33–38

[61] H. J. Sussmann, *An extension of a theorem of Nagano on transitive Lie algebras*, Proc. Amer. Math. Soc., **45** (1974) 349–356.

[62] H. J. Sussmann, *Lie brackets and local controllability: a sufficient condition for scalar-input systems*, SIAM J. Cntrl. & Opt., **21** (1983) 686–713.

[63] H. J. Sussmann, *Lie brackets, Real Analyticity, and Geometric Control*, in Differential Geometric Control, (R. W. Brockett, R. S. Millman, H. J. Sussmann, eds.) (1983) 1–116.

[64] H. J. Sussmann, *A product expansion of the Chen series*, in "Theory and Applications of Nonlinear Control Systems," C. I. Byrnes and A. Lindquist eds., Elsevier, North-Holland (1986), 323–335.

[65] H. J. Sussmann, *A general theorem on local controllability*, SIAM J. Control & Opt., **25** (1987) 158-194

[66] H. J. Sussmann, *New differential geometric methods in nonholonomic path finding*, Progress Systems and Control Theory, **12**(1992), pp. 365.

[67] D. Tilbury, R. Murray and S. Sastry, *Trajectory generation for the N-trailer problem using Goursat normal form*, IEEE Trans.Aut.Cntrl. **40** (1995) 802–819.

[68] G. Viennot, *Algèbres de Lie Libres et Monoïdes Libres*, Lecture Notes, Math., 692, Springer, Berlin, 1978.