united nations educational, scientific and cultural organization

international atomic energy agency

the
abdus salam
international centre for theoretical physics

SMR1327/18

Summer School on Mathematical Control Theory

(3 - 28 September 2001)

The combinatorics of nonlinear controllability and noncommuting flows (I)

Matthias Kawski
Department of Mathematics
Arizona State University
Tempe, Arizona 85287-1804
U.S.A.

The combinatorics of nonlinear controllability and noncommuting flows

Matthias Kawski *

Department of Mathematics, Arizona State University, Tempe, Arizona 85287-1804, U.S.A.

Preliminary version: September 16, pages 00 – 27

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

LNS

*Supported in part by NSF-grant DMS 00-72369.

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 in 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 the most recent years. One of the main targets is the development of an explicit, continuous analogue of the classical Campbell-Baker-Hausdorff formula. The purpose of such formula 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, and notation, from algebraic combinatorics are essential, both for 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 the recent literature.

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.

Keywords: Nonlinear controllability, exponential Lie series, free Lie algebra, Hall bases, chronological algebra, combinatorics.

AMS classification: 93 X - xx

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 and clearest or most efficient formulation was 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, more advanced 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 revisited 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.

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 your car. For the sake of definiteness, suppose the initial position and orientation of the car as indicated in the diagram above (with much exaggerated parallel displacement, and an exaggerated length of the gap), with steering wheels in the direction of the road.

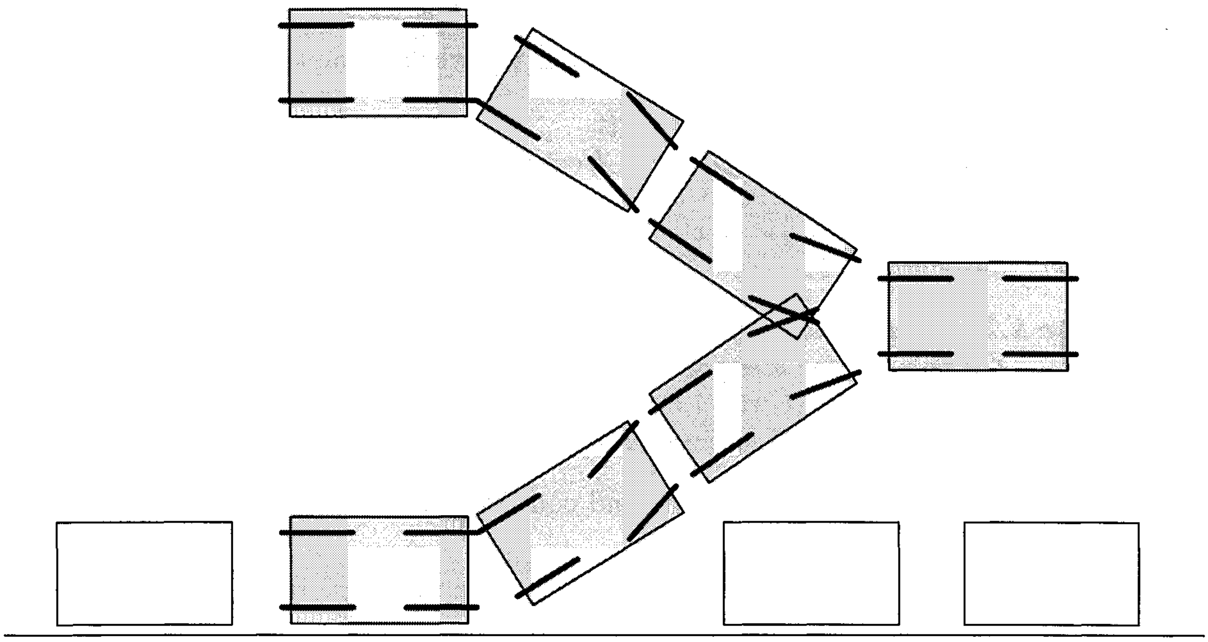

Figure 1. Parallel parking a car (exaggerated parallel displacement)

Everyday experience says that, while it is impossible to directly move the car sideways, it is possible to do so through 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 wheel 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*

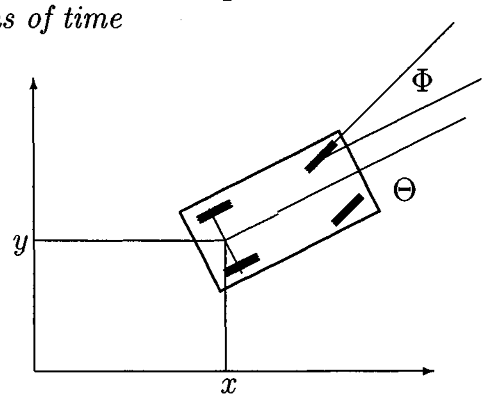

Figure 2. Defining the *states* of the system

To obtain a mathematical model consider the simpler (less controversial case as it does not require a differential) 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 neither wheel can slip 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 (center of the axle of the) rear wheel, we 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

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

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

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

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).

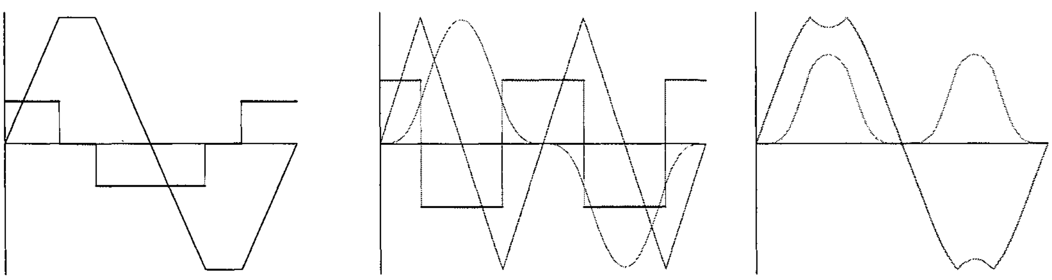

Figure 3. One possible, very symmetric, parallel parking maneuver.

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.*

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 an ordinary dynamical system, and it 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, or the 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, as is the time T allowed. Moreover, the scaling immediately connects local and global properties. Finally, the solution formula above formally also quickly yields (e.g. via Taylor expansions and the Cayley-Hamilton theorem) to a simple algebraic criterion for linear controllability:

Theorem 1.1 (Kalman rank condition) *The linear system $\dot{x} = Ax + Bu$ with $x \in \mathbf{R}^n$ 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 (e.g. measurable versus piecewise constant) are 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). Thus from now on, unless otherwise stated the following blanket assumptions shall generally apply: We consider affine, analytic systems that are of the form

$$\dot{x}(t) = f_0(x(t)) + \sum_{i=1}^{m} u_i(t)\ f_i(x(t)) \tag{8}$$

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 $U \subset \mathbf{R}^m$, 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 he control $\dot{x} = f(x,u)$, compare the lectures by Jacubczyk in this series. One may also demand less regularity, e.g. e.g. only Lipschitz-continuity of the vector fields associated to fixed values of the controls Good theoretical framework for that case provided by *differential inclusions*, compare the lectures by Frankowska in this series.

Revisiting the parking example of the first section we introduce standard, uniform notation by defining $x = (x_1, x_2, x_3, x_4, x_5) \stackrel{\text{def}}{=} (\phi, v, x, \theta, y)$. With this we write the system (3) in the form (8) with

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

Thus, this is a system with drift – which corresponds to the dynamics of the car, and with two controlled vector fields which correspond to forward accelartion/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 filed(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 sets $\mathcal{R}_p(t)$ of system (8) subject to the initial condition $x(0) = p$ is the set*

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

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 $t > 0$.*

The following most simple example clearly illustrates that accessibility and controllability (STLC) are generally different from each other.

$$\begin{cases} \dot{x}_1 &= u \qquad\qquad x(0) &= 0 \\ \dot{x}_2 &= x_1^k \qquad\quad \|u(\cdot)\| &\leq 1 \end{cases} \tag{11}$$

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 these 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 $\mathbf{R}^n$. Nonetheless, occasionally we may phrase our observations and results so as to emphasize that they really also apply to general manifolds. In particular, when working with approximating vectors we shall conveniently identify the tangent spaces to $\mathbf{R}^n$with $\mathbf{R}^n$.
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 are automatically necessary conditions for optimality, and vice versa.

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 $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_{s-1} \leq t_s = T$ 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 $\dot{x} = f_0(x) + \sum_{j=1}^m (c_i)_j f_j(x)$. 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 $F_i = f_0 + \sum_{j=1}^m (c_i)_j f_j$, $i = 1 \ldots s$ and evaluating the composition at $x(0)$. It is customary to write this as a *product of exponentials*

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

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 $pe^X Y e^Z \phi$ 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 are 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 MK and Sussmann [23]. 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 $pe^{tX}X$ 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 $\Phi_{*p}\colon T_pM \mapsto T_pN$ and $\Phi_*\colon TM \mapsto TN$ denote the tangent map in classical notation. In our case, the positioning 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. 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 may be able to conclude STLC if the Jacobian matrix f 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 [36], 1985)

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

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)

$$\begin{array}{rcl} x(T, u_{+-+-;a,b,c,d}) &=& (a+c-b-d, \\ && \frac{1}{2}a^2 + ab - \frac{1}{2}b^2 - bc + ac + \frac{1}{2}c^2 + ad + cd - bd - \frac{1}{2}d^2, \\ && -\frac{1}{4}b^4c^2 + b^2c^4 + ab^3c^2 - 2bac^4 - \frac{1}{2}ba^4c - 2ba^3c^2 - \ldots) \end{array} \tag{14}$$

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

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

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*

$$\left\{\begin{array}{lcl} \dot{x}_1 &=& u \\ \dot{x}_2 &=& x_3 \\ \dot{x}_3 &=& x_1 x_3 \end{array}\right. \qquad \begin{array}{rcl} x(0) &=& 0 \\ \|u(\cdot)| &\leq& 1 \\ && \end{array} \tag{16}$$

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 swithcing 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* $t_{i-1} \le t \le t_i$ *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. For many systems this approach is impractical as e.g. exact switching times which return the system to the starting point may be difficult to 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, one defines the Lie bracket via the limit (for $\phi \in C^\infty(M)$)

$$[F_1,F_2]\phi(p) = \lim_{t \to 0} \frac{1}{t^2}\left(\phi(e^{-tF_2}e^{-tF_1}e^{tF_2}e^{tF_1}p) - \phi(p)\right) \tag{21}$$

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

$$\begin{array}{rcl} p' = e^{tf_1}(p) & = & (p_1, p_2 + tp_1) \\ p'' = e^{tf_2}(p') & = & (p_1 + t, p_2 + tp_1) \\ p''' = e^{-tf_1}(p'') & = & (p_1 + t, p_2 - t^2) \\ p'''' = e^{-tf_2}(p''') & = & (p_1, p_2 - t^2) \end{array} \tag{22}$$

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 $[f_0, f_1] = -\frac{\partial}{\partial x_2}$.

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

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

On 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). 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, but they are, in general, beyond the scope of these notes. 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. [23].

Exercise 1.16 *Repeatedly use the CBH-formula to write the end point of 4 flows corresponding to* bang-bang *controls as a single* exponential*:*

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

Find explicit formulas for polynomial expressions $p_I(t)$ (in the switching times) for $I = 0, 1, 01, 011, 110$, but don't worry about higher order brackets.

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 go to a specific point, we continue to develop tools that build on arguments using approximate directional information obtained from derivatives. This discussion should also establish the close link between STLC and optimal control.
The key idea is to develop a tangent, or derivative object for the reachable sets that nicely approximates it, that is easy to constuct/compute, and which has reasonably nice convexity properties. For general systems the very well developed tools of nonsmooth analysis apply, especially the contingent cones (see the lectures by Frankowska in this series). However, for the affine, analytic systems (8) initialized at an equilibrium point the following much simpler notion does the job.

Definition 1.3 *Consider systems of the form* (8) *on* $\mathbf{R}^n$*with* $f_0(0) = 0$ *and* $0 \int U$. *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 1.17 *Find 6 families of control variations* $u_s^{\pm i}\colon [0,s] \mapsto [-1,1]$ *that generate the tangent vectors* $\frac{\pm\partial}{\partial x_i}$ *for the system* (13). *Hint: Start with the piecewise constant controls used in example 1.3, and analyze how the endpoint changes as the controls are scaled in time.*

Exercise 1.18
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 (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.

Theorem 1.3 (Kawski [17]) *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 are constants* $C > 0$, $T > 0$ *such that* $\overline{C'} \cap B(0, Ct^k) \subseteq \mathcal{A}(t)$ *for all* $0 \leq t \leq T$.

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* $0 \leq t \leq T$.

The preceding discussions, examples, and exercises suggest that there should be generic 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 $(ad^{k+1} f, g) = [f, (ad^k(f, g)]$. 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.
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 $f0$.
For a set S of vector fields and a point p, we write $S(p)$ for the set $\{v(p) : v \in S\}$. Since all our considerations are local, we identify the tangent space $T_0\mathbf{R}^n$ with $\mathbf{R}^n$.

First we need to distinguish between accessibility and controllability – for analytic vector fields, accessibility is comparatively easy to decide.

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

The closest analogue to the Kalman rank condition 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 1.1) for linear

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

A complementary necessary condition is the following, closely related to the Clebsch-Legendre condition of optimal control:

Theorem 1.7 (Hermes [12], Sussmann [39])
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 [12], Sussmann [39])
If $m = 1$ *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 system* (8) *is STLC.*

A complementary necessary condition is:

Theorem 1.9 (Stefani [37])
If $m = 1$ *and* (8) *is STLC then* $(ad^{2m}(f_0, f_1)(0) \in \mathcal{S}^{2m-1}(f_0, f_1)(0)$ *for all* $m \in \mathbf{Z}^+$.

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

Theorem 1.10 (Sussmann [42])
If there is a weight $\theta \in (0,1]$ such that for all brackets $f_\pi \in L^{(k,\ell_1,\ldots\ell_m)}(f_0,\ldots f_m)$ with k odd and all ℓ_j even, there are brackets $f_\pi \in L^{(k_j,\ell_{1,j},\ldots\ell_{m,j})}(f_0,\ldots f_m)$ with k f_{π_j} of type (k_j,l_j) with $\theta k_j + \ell_j < \theta k + \ell_j$ such that $f_\pi(0)$ is a linear combination of $f_{\pi_j}(0)$ then the system (8) is STLC.

(It is also possible to allow several different weights.) Several small extensions, and specific examples that test the ground between the necessary and sufficient conditions may be found in the literature, see e.g. [?] for an overview.

Example (1.3) revisited. Extract the vector fields f_0 andf_1 from (13) and calculate iterated Lie brackets – the big question: which ones?

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

$$[f_1,[f_1,f_0]](x) = \begin{pmatrix} 0 \\ 0 \\ -6x_1x_2 \end{pmatrix} \quad (ad^3 f_1, f_0)(x) = \begin{pmatrix} 0 \\ 0 \\ -6x_2 \end{pmatrix} \quad (ad^4 f_1, f_0)(x) = 0 \qquad (27)$$

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

$(ad^k f_0, f_1)(x) = 0$ when $k > 0$. Since $\dim L(f_0,f_1)(0) = 3$ the system is accessible, but since $\dim \mathcal{S}^1(f_0,f_1)(0) = 2 < 3$ is *not linearly controllable.* Since $[f_1,[f_1,f_0]](0) = 0 \in \mathcal{S}^1(f_0,f_1)(0)$ and similarly $(ad^4 f_1, f_0)(0) = 0 \in \mathcal{S}^3(f_0,f_1)(0)$, Stefani's necessary conditions are satisfied. The only brackets which gives 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.
Indeed, this example by Stefani was first shown to be STLC using the method we exhibited in the previous section, and it clearly shows that the Hermes' condition was far from necessary. It then served as a substantial motivation for the eventual sharpened version of Sussmann's general theorem.

With a computer algebra system such calculation is 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 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 very quickly 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 1.19 *Determine whether the car model (9) from example (1.1) is STLC.*

Exercise 1.20 *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 and bases

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. [?] 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 repeating the application of the CBH-formula is even less attractive. Thus one is lead to look for expansions that do not rely on piecewise constancy.
One of the most basic formulas is obtained by simple Picard iteration. First rewrite the system of differential equations with initial condition

$$\dot{x}(t) = \sum_{i=0}^{n} u_i(t) f_i(x(t)), \quad x(0) = p \tag{28}$$

as an equivalent integral equation, and then iterate this

$$\begin{aligned} x(t) &= p + \int_0^t \sum_{i=0}^{n} u_i(t_1) f_i(x(t_1))\, dt_1 \\ &= p + \int_0^t \sum_{i_1=0}^{n} u_{i_1}(t_1) f_{i_1}\Big(p + \int_0^{s_1} \sum_{i_2=0}^{n} u_{i_2}(t_2) f_{i_2}(x(t_2))\, dt_2 \Big)\, dt_1 \\ &= p + \int_0^t \sum_{i_1=0}^{n} u_{i_1}(t_1) f_{i_1}\Big(p + \int_0^{t_1} \sum_{i_2=0}^{n} u_{i_2}(t_2) f_{i_2}\Big(p + \int_0^{t_2} \sum_{i_3=0}^{n} u_{i_3}(t_3) f_{i_3}(x(t_3))\, dt_3 \Big)\, dt_2 \Big)\, dt_1 \end{aligned} \tag{29}$$

and so on. Assuming that the vector fields f_i are analytic, one may simplify the resulting expression (after infinite iteration) using Taylor expansions of f_i about the starting point p. (One may also start by expanding $h(x(T,u))$ for any analytic function $h \in C^\omega(0)$ into a Taylor series with respect to time about $T = 0$, compare [?]). A more attractive approach is to recall that vector field are first order partial differential operators, and then proceed on a more formal level. Using the chronological formalism we will give a simple, direct derivation in a later lecture. At this time let us just record the final result.

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*

$$S_{CF}(T,u) = \sum_I \underbrace{\int_0^T \int_0^{t_{p-1}} \cdots \int_0^{t_3} \int_0^{t_2} u^{i_p}(t_p) \ldots u^{i_1}(t_1)\, dt_1 \ldots dt_p}_{\Upsilon^I(T,u)} \; \underbrace{X_{i_1} \ldots X_{i_p}}_{X_I} \tag{30}$$

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

This series originates in X. T. Chen's study [6] of geometric invariants of curves in $\mathbf{R}^n$in the 1950s. In the early 1970s Fliess recognized its utility for the analysis of control systems. Using careful analytic estimates one may prove [39]

Theorem 2.1 *Suppose f_i are analytic vector fields on $\mathbf{R}^n$, $\varphi: \mathbf{R}^n \mapsto \mathrm{I\!R}$ is analytic and $\mathcal{U} \subset \mathrm{I\!R}^{m+1}$ is compact. Then for every compact set $K \subseteq \mathrm{I\!R}^n$, there exists $T > 0$ such that for every $p \in K$ and every $u: [0,T] \mapsto \mathcal{U}$ the series*

$$S_{CF,f}(T,u)(\varphi) = \sum_I \Upsilon^I(T,u) \cdot ((f_I \varphi)(p)) \tag{31}$$

converges to the solution $x(t,u)$ of (28) with initial condition $x(0) = p$.

In the last formula $(f_I\varphi)(p) = (f_{i_1} \circ f_{i_2} \cdots \circ f_{i_s}\varphi)(p)$ $\boxed{\textbf{Check the order}}$ is to be interpreted as an s-th order partial derivative of φ evaluated at p.
This series solution is not just good for piecewise constant controls, but for all measurable controls. To get a better feeling for the terms, consider again the example 13. Write out the vector fields

$$f_0(x) = x_1 \frac{\partial}{\partial x_2} + x_1^3 x_2 \frac{\partial}{\partial x_3} \text{ and } f_1(x) = \frac{\partial}{\partial x_1} \tag{32}$$

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

$$x_2(T,u) = \Upsilon^10(T,u) = \int_0^T \int_0^{t_2} u_1(t_2)u_0(t_1)\, dt_1\, dt2 = \int_0^T \int_0^{t_2} u_1(t_1)\, dt_1\, dt2 \tag{33}$$

As expected, the series just returns the integral form for the linear double integrators part of the system 13.
For $\varphi = x_3$ note that $f_1 x_3 \equiv 0$ and $f_0 x_3 = x_1^3 x_2$. (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 the

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 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*

$$x_3(T,u) = \int_0^T \left(\left(\int_0^{t_3} u(t_2)\,dt_2 \right)^3 \cdot \left(\int_0^{t_3} \int_0^{t_2} u(t_1)\,dt_1\,dt_2 \right) \right) dt_3. \tag{34}$$

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 niloptent approximations, to be discussed in the next chapter.

2.2 Families of dilations

For (nonconstant) polynomial functions 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 apparent structures of some sort of order or degree are quite useful, in particular for identifying *leading terms* and consequent constructions of approximations. Working with fixed coordinates $(x_1, x_2, \ldots x_n)$ it is convenient to make the following definition which is a special case of the general geometric (i.e. coordinate free) notion of homogeneity of [21]:

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

$$\Delta_s(x) = (s^{r_1}x_1, s^{r_2}x_2, \ldots, s^{r_n}x_n). \tag{35}$$

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

$$\phi \circ \Delta_s = s^m \phi \;\; \textit{and} \;\; F x_k \in H_{r_k} \;\; \textit{for} \; k = 1, 2, \ldots, n. \tag{36}$$

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

$$\nu(x) = r_1 x_1 \frac{\partial}{\partial x_1} + r_2 x_2 \frac{\partial}{\partial x_2} + \ldots + r_n x_n \frac{\partial}{\partial x_n}. \tag{37}$$

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. $x_1 \in H_1$, $x_2 \in H_2$ and $x_3 \in H_6$. With these weights e.g. $\phi(x) = x_1 x_3 - x_1^7 + x_1 x_2^3 \in H_7$ 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 $\frac{\partial}{\partial x_1}\phi \in H_6$, $\frac{\partial}{\partial x_1}\phi \in H_5$, and $\frac{\partial}{\partial x_1}\phi \in H_6$.

The following properties hold also for more general, geometric dilations as defined in [21].

Proposition 2.2
Let Δ be a one-parameter family of dilations on $\mathbb{R}^n$ with coordinates $(x_1, \ldots x_n)$.
If $\phi \in H_m$, and $\psi \in H_k$, then $\phi\psi \in H_{m+k}$.
If $F \in\in \underline{n}_m$, and $G \in\in \underline{n}_k$ are smooth then then $[F, G] \in \underline{n}_{m+k}$.
If $\phi \in H_m$, and $F \in\in \underline{n}_k$, then $F\phi \in H_{m+k}$.
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.

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

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

Proposition 2.3 ([21]) *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 $\mathbf{R}^n$is homogeneous $\phi \in H_m$ iff $\nu\phi = m\phi$.
A smooth vector field on $\mathbf{R}^n$is homogeneous $F \in\in \underline{n}_m$, and $G \in\in \underline{n}_k$ iff $[\nu, F] = mF$.

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

One of the typical uses of the last property stated in proposition (2.2) is to allow one to stop computing 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 specific 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 $f_0 \in underlinen_0$ and $f_1 \in underlinen_{-1}$ with respect to the dilation defined by $r = (1, 1, 4)$, while they are homogeneous of degrees $f_0, f_1 \in underlinen_{-1}$ 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 that essentially 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 (with well-defined maximal degrees) through a local coordinate change. I.e. 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 1.17 and 1.18, using piecewise constant controls also illustrated that, at least in the case of *homogeneous* systems, the order of each Lie bracket corresponded to the degree of the degree of the polynomial expression. 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

$$u_{\varepsilon,\delta}\colon [0\delta T] \mapsto \varepsilon U \subseteq U \;\text{ by } |; u_{\varepsilon,\delta}(\delta t) = \varepsilon(u(t) \tag{38}$$

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

Proposition 2.4 *Suppose that $\Delta^{(1)}$ and $\Delta^{(2)}$ are families of dilations as above.*
If the system is homogeneous such that $f_1 \in \underline{n}^{(1)}(-1)$ and $f_0 \in \underline{n}^{(1)}(0)$ with respect to $\Delta^{(1)}$ and such that the system is homogeneous such that $f_0, f_1 \in \underline{n}(2)(0)$ with respect to $\Delta^{(2)}$ then

$$x(\delta T, u_{\varepsilon,\delta} = \Delta^{(1)}_{\varepsilon}\Delta^{(2)}_{\delta}x(T, u_{1,1} \tag{39}$$

for all $\varepsilon, \delta \in [0,1]$.
If only one such dilation is known, then the statement holds true with a fixed parameter $\delta = 1$ (only $\Delta^{(1)}$-homogeneity) or with $\varepsilon = 1$ (only $\Delta^{(2)}$-homogeneity).

A simple proof uses uniqueness of solutions of initial value problems, showing that both the right and left hand side of (40) are solutions of the same control system, see e.g. [20].
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 $\Delta^{(2)}_{\delta}$, while a Taylor expansion in the control sizes, and Hermes sufficient condition s is built around the dilation $\Delta^{(2)}_{\varepsilon}$. Sussmann's general sufficient condition allows a trade-off between the time-scale and amplitude. Basically, relate the rates at which ε and δ go to zero by setting $\varepsilon = s$ and $\theta = s^{\theta}$ (for $0 \leq s \leq 1$ and $\theta \in (0,1]$.

Corollary 2.5 *Suppose that Δ is a family of dilations as above.*
If the system is homogeneous such that $f_1 \in \underline{n}^{(1)}(-1)$ and $f_0 \in \underline{n}^{(1)}(\Theta)$ with respect to Δ, then for all $0 \leq s \leq 1$

$$x(\delta T, u_{s,s^{\Theta}} = \Delta^{(1)}_{\varepsilon}\Delta^{(2)}_{\delta}x(T, u_{1,1} \tag{40}$$

Exercise 2.5 *Use the proposition to prove the corollary.*

Exercise 2.6 *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 $\frac{r_2}{r_1}$, 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 [?]). Also, compute all nonzero Lie brackets of the vector fields f_0 and f_1 defining this system.*

$$\left\{\begin{array}{rcl} \dot{x}_1 &=& u \\ \dot{x}_2 &=& x_1 \\ \dot{x}_3 &=& x_2^2 + x_1^3 \end{array}\right. \qquad \begin{array}{l} |u(\cdot)| \leq c \\ x(0) = 0 \\ \end{array} \tag{41}$$

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). I.e. one approximates he system
refeq1 by a linear system $\dot{x} = Ax + Bu$ where A equals the Jacobian matrix of partial derivatives if the drift $f_0(x)$, and where the i-th column of B equals the value of $g_i(0)$, $i = 1, \ldots, m$. (Of course, this can be formulated in a coordinate-free geometric way that does not mix up the state space and its tangent spaces.)

Exercise 2.7 *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 your earlier findings from the exercises in the first sections.

The exercises make it clear that for some nonlinear systems of reasonable *"reality"* the standard linearization causes a dramatic loss of information. Thus one may ask 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 properties, that systems in this class are amenable to reasonable analysis and computation, and that he approximation is algorithmic and allows for explicit computation. At this turn no such ideal approximating scheme has been been found – he 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, as a consequence of [18] this approach fails for the system

$$\left\{ \begin{array}{rcl} \dot{x_1} &=& u \\ \dot{x_2} &=& x_1 \\ \dot{x_3} &=& x_1^3 \\ \dot{x_4} &=& x_3^2 + x_2^7 \end{array} \right. \qquad \begin{array}{rcl} x(0) &=& 0 \\ |u(\cdot)| &\leq& \varepsilon_0 \end{array} \tag{42}$$

For systems of from (8) that are knwon to be STLC by Sussmann's condition theorem 1.10, the objective of this procedure is to obtain STLC, nilpotent approximating systems, on the *same state space* $\mathbf{R}^n$, of the same form

$$\dot{y}(t) = g_0(y) + \sum_{i=1}^{m} u_i(t)\; g_i(y) \tag{43}$$

(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 only (no solution of nonlinear differential equations is needed). Thus, one should consider *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 controllable linear systems.

The following procedure is due to Hermes (compare the review [13]), with very similar algorithms employed at almost the same time by Stefani, Bressan and others. We give a crude outline, omitting some technical steps that are not central and rarely needed. See the review [13], or the original references for more details, especially Stefani [36] for details about adapted charts.)

The basic assumption is that the system is of form (8) and is known to be STLC by virtue of theorem 1.10. Calculate iterated Lie brackets of the vector fields of increasing length until their values at 0 span the tangent space $T_0\mathbf{R}^n$. 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 is free to choose any such weights, and thus may 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

$$\operatorname{span}\{f_{\pi_1}(0), f_{\pi_2}(0), \ldots, f_{\pi_n}(0)\} = T_0\mathbf{R}^n \tag{44}$$

and they are of lowest possible weight, defined as the weighted sum of θ times the number of occurrences of each factor f_i 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 (actually, in general a strictly triangular polynomial coordinate change should be done, see [36] for *"adapted charts"* such that $f_{\pi_1}(0) = \frac{\partial}{\partial x_i}$ for $i = 1, 2, \ldots n$ and wlog. $1 \le r_1 \le r_2 \le \ldots, r_n$. 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, and truncate the expansion keeping only polynomials $p_{ij}(x)$ of order less or equal to $r_j - 1$ for $i \ge 1$, and $r_j - \theta$ for $i = 0$. Define the vector fields $g_j = \sum_{j=1}^n p_{ij}(x)\frac{\partial}{\partial x_j}$, which are easily checked to be of homogeneous degrees $g_i \in \underline{n}_{-\theta}$. Since all these degrees are strictly negative, they generate a nilpotent Lie algebra. The preservation of STLC properties follows from the observation that $g\sigma$ 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,

$$\operatorname{span}\{f_{\pi_1}(0), \ldots, f_{\pi_n}(0)\} = \operatorname{span}\{g_{\pi_1}(0), \ldots, g_{\pi_n}(0)\} \tag{45}$$

Note that this is only a rough outline of the procedure as a precise description requires substantially more technical symbols. See the original references of the survey [13] for details.

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

$$f_0(x) = \begin{pmatrix} 0 \\ 0 \\ x_2 \cos x_4 \\ x_2 \tan x_1 \\ x_2 \sin x_4 \end{pmatrix}, \qquad f_1(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } \qquad f_2(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{46}$$

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

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

and finally $\boxed{\text{check this}}$

$$[[f_0, f_2], [[f_0, f_1], f_2]] = \begin{pmatrix} 0 \\ 0 \\ -\sec^2 x_1 \sin x_4 \\ -2\sec^2 x_1 \tan^2 x_1 \\ \sec^2 x_1 \cos x_4 \end{pmatrix}. \tag{47}$$

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:

$$f_1(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; f_2(0) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; [f_0, f_2](0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \; f_{\pi_4}(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \; f_{\pi_5}(0) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{48}$$

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 2.8 *Check this. I.e. 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 section 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 $f_{\pi_i}(0) = \frac{\partial}{\partial x_i}$ for $i = 1, 1, 2, 3, 5$ and the r_i are nondecreasing.
Expanding the components of $f_i x_j$ into Taylor series and keeping only the Δ-lowest term, i.e. of degree $r_i - 1$ we obtain the approximating fields

$$g_0(x) = \begin{pmatrix} 0 \\ 0 \\ x_2 \\ x_1 x_2 \\ x_2 x_4 \end{pmatrix}, \quad g_1(x) = f_1(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and } \quad g_2(x) = f_2(x) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{49}$$

Exercise 2.9 *Verify directly, i.e. using the theorem* 1.10 *that this nilpotent approximating system* (49) 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 2.10 *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 [?]).*

Exercise 2.11 *Calculate an STLC nilpotent approximating systems for the models* (5) and (6) for the dynamics of a rolling penny (example (1.2)).

Exercise 2.12 *Verify that for no choice of* $\theta \in (0, 1]$ *the system* (42) satisfies Sussmann's sufficient conditions in theorem 1.10. (But it has been shown in [18] that the system (42) is nonetheless STLC.)
Show that for no choice of a dilation $\Delta_s(x) = (s^{r_1}x_1, s^{r_2}x_2, s^{r_3}x_3, s^{r_4}x_4)$ (with the same coordinates (x_1, x_2, x_3, x_4)) the definite term (*potential obstruction to STLC*) x_3^2 has a higher weight (order of homogeneity) than the indefinite term x_2^7.

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 computational challenges or for 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 (31)

$$
\begin{aligned}
S_{CF,f}(T,u)(\varphi) &= 1 \cdot \varphi(0) + \int_0^T 1 dt \cdot (f_0\varphi)(0) + \int_0^T u(t)dt \cdot (f_1\varphi)(0) \\
&+ \int_0^T \int_0^{t_2} 1 \cdot 1 dt_1 dt_2 \cdot (f_0 f_0 \varphi)(0) + \int_0^T \int_0^{t_2} u(t_2)u(t_1) dt_1 dt_2 \cdot (f_1 f_1 \varphi)(0) \\
&+ \int_0^T \int_0^{t_2} 1 \cdot u(t_1) dt_1 dt_2 \cdot (f_0 f_1 \varphi)(0) + \int_0^T \int_0^{t_2} u(t_2) \cdot 1 dt_1 dt_2 \cdot (f_1 f_0 \varphi)(0) \qquad (50) \\
&+ \int_0^T \int_0^{t_3} \int_0^{t_2} 1 \cdot 1 \cdot 1 dt_1 dt_2 dt_3 \cdot (f_0 f_0 f_0 \varphi)(0) + \int_0^T \int_0^{t_3} \int_0^{t_2} 1 \cdot 1 \cdot u(t_1) dt_1 dt_2 dt_3 \cdot (f_0 f_0 f_1 \varphi)(0) \\
&+ \int_0^T \int_0^{t_3} \int_0^{t_2} 1 \cdot u(t_2) \cdot 1 dt_1 dt_2 dt_3 \cdot (f_0 f_1 f_0 \varphi)(0) + \int_0^T \int_0^{t_3} \int_0^{t_2} u(t_3) \cdot 1 \cdot 1 dt_1 dt_2 dt_3 \cdot (f_1 f_0 f_0 \varphi)(0) \\
&+ \int_0^T \int_0^{t_3} \int_0^{t_2} u(t_3) \cdot u(t_2) \cdot 1 dt_1 dt_2 dt_3 \cdot (f_0 f_1 f_1 \varphi)(0) + \int_0^T \int_0^{t_3} \int_0^{t_2} u(t_3) \cdot 1 \cdot u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_0 f_1 \varphi)(0) \\
&+ \int_0^T \int_0^{t_3} \int_0^{t_2} 1 \cdot u(t_3) u(t_2) dt_1 dt_2 dt_3 \cdot (f_1 f_1 f_0 \varphi)(0) + \int_0^T \int_0^{t_3} \int_0^{t_2} u(t_3) u(t_2) u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_1 f_1 \varphi)(0) \\
&+ \text{ higher order terms}
\end{aligned}
$$

This is just the beginning, and one never should work with 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

$$
w = a_1 a_2 \dots a_s \mapsto \left(\phi \mapsto (f_w \phi)(0) = (f_{a_1} \circ f_{a_2} \circ \cdots f_{a_s} \phi)(0) \right), \quad \text{and} \qquad (51)
$$

$$
\Upsilon\colon w = a_1 a_2 \dots a_s \mapsto \left(u \mapsto \int_0^T \dots \int_0^{t_s} u_{a_s}(t_s) \dots u_{a_1}(t_1) dt_1 \dots dt_s \right) \qquad (52)
$$

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 $[0, T]$.

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 (51) and (52), or alternatively through purely algebraic and combinatorial manipulation of the combinatorial objects on the left hand side of (51) and (52).
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 one combinatorial algebraic operation which makes Υ an algebra homomorphism (for a suitable algebra structure).
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 (50)). 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 primaily 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_1s. In particular, noting that $[f_1, f_0]\phi = f_1 f_0 \phi - f_0 f_1 \phi$, we add and subtract the term (alternative choices are possible)

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

then combine the results appropriately (alternatively start by integrating by parts)

$$\begin{aligned}
&\int_0^T \int_0^{t_2} 1 \cdot u(t_1) dt_1 dt_2 \cdot (f_0 f_1 \varphi)(0) + \int_0^T \int_0^{t_2} u(t_2) \cdot 1 dt_1 dt_2 \cdot (f_1 f_0 \varphi)(0) = \\
&= \left(\int_0^T 1 \int_0^{t_2} u(t_1) dt_1 dt_2 + \int_0^T u(t_2) \int_0^{t_2} 1 dt_1 dt_2 \right) \cdot (f_0 f_1 \varphi)(0) + \int_0^T \int_0^{t_2} u(t_2) \cdot 1 dt_1 dt_2 \cdot ((f_1 f_0 - f_0 f_1)\varphi)(0) \\
&= \left(\int_0^T u(t) dt \right) \cdot \left(\int_0^T 1\, dt \right) \cdot (f_0 f_1 \varphi)(0) + \int_0^T u(t_2) \int_0^{t_2} 1 dt_1 dt_2 \cdot ([f_1, f_0]\varphi)(0)
\end{aligned} \tag{53}$$

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 repeatedly integrate by parts, instead of judiciously adding and subtracting terms, This has the same effect, but is closer to the popular technique of *rewriting systems* of algebraic combinatorics, compare [30] and [34]. (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. After all this is for illustration only.)

$$\int_0^T u(t_3)\int_0^{t_3} u(t_2)\int_0^{t_2} 1 dt_1 dt_2 dt_3 \cdot (f_0 f_1 f_1 \varphi)(0)$$

$$+\int_0^T u(t_3)\int_0^{t_3} 1\int_0^{t_2} u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_0 f_1 \varphi)(0) + \int_0^T\int_0^{t_3} 1 u(t_2)\int_0^{t_2} u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_1 f_0 \varphi)(0) =$$

integrating by parts the inside integral in the first term

$$= \left(\int_0^T u(t_3)\left(\cdot\left(\int_0^{t_3} u(t_2)dt_2\right)\cdot\left(\int_0^{t_3} 1 dt_2\right) - \int_0^{t_3} 1\int_0^{t_2} u(t_1)dt_1 dt_2\right) dt_3\right)\cdot(f_0 f_1 f_1 \varphi)(0)$$

$$+\int_0^T u(t_3)\int_0^{t_3} 1\int_0^{t_2} u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_0 f_1 \varphi)(0) + \int_0^T\int_0^{t_3} 1 u(t_2)\int_0^{t_2} u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_1 f_0 \varphi)(0)$$

Again integrate the first term by parts, after suitably regrouping the inside, and combine the second and third term, recognizing that $f_0 f_1 f_1 - f_0 f_1 f_1 = [f_1, f_0] f_1$

$$= \left(\left(\int_0^T \mathbb{1} dt\right)\cdot\left(\int_0^T u(t_3)\int_0^{t_3} u(t_2)dt_2 dt_3\right) - \int_0^T 1\cdot\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)dt_1 dt_2 dt_3\right)\cdot(f_0 f_1 f_1 \varphi)(0)$$

$$+\left(\int_0^T\left(u(t_3)\int_0^{t_3} 1\int_0^{t_2} u(t_1)dt_1 dt_2\right) dt_3\right)\cdot([f_1, f_0] f_1 \varphi)(0)$$

$$+\int_0^T\int_0^{t_3} 1 u(t_2)\int_0^{t_2} u(t_1) dt_1 dt_2 dt_3 \cdot (f_1 f_1 f_0 \varphi)(0)$$

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

$$= \frac{1}{2}\cdot\left(\int_0^T \mathbb{1} dt\right)\cdot\left(\int_0^T u(t)dt\right)^2 dt_3\cdot(f_0 f_1 f_1 \varphi)(0) + \int_0^T\int_0^{t_3} 1 u(t_2)\int_0^{t_2} u(t_1)dt_1 dt_2 dt_3\cdot(f_1 f_1 f_0 - f_0 f_1 f_1 \varphi)(0)$$

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

Finally write the inner integral in the fourth term as a double integral, as in the second term and combine them

$$= \frac{1}{2}\cdot\left(\int_0^T \mathbb{1} dt\right)\cdot\left(\int_0^T u(t)dt\right)^2 dt_3\cdot(f_0 f_1 f_1 \varphi)(0)$$

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

$$+\left(\int_0^T 1\int_0^{t_3} u(t_2)\int_0^{t_2} u(t_1)dt_1 dt_2 dt_3\right)\cdot[f_1, [f_1, f_0]]\varphi)(0) .$$

The last step used that

$$f_1 f_1 f_0 - f_0 f_1 f_1 - 2[f_1, f_0] f_1 = f_1 f_1 f_0 - 2 f_1 f_0 f_1 + f_0 f_1 f_1 = [f_1, [f_1, f_0]]. \tag{54}$$

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 products of iterated integrals. 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 the the usefulness of this expression, suppose ϕ is a function such that $f_1\phi \equiv\!= f_0\phi(0) = [f_1, f_0]\phi(0) = 0$ (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. Alternatively, if the integrals corresponding to the *words* $0, 1, 10$ all 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 opeartors (e.g. if $f_1\phi \equiv 0$ then $Xf_1\phi \equiv 0$ fir *every* partial differential operator X. 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 lost 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 [?], but it was clear that this is not the way to go. In [41] Sussmann managed to factor the entire series using differential equations techniques. An elegant alternative is o 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: using $(0 \sqcup\!\sqcup 1) = 10 + 01$ in the first step.

$$011\otimes011+101\otimes101+110\otimes110 = (((01+10)-10)1)\otimes011+101\otimes101+110\otimes110$$

$$\begin{aligned}
&= ((01+10)1\otimes011+101\otimes(101-011)+110\otimes110 \\
&= (((011+101+110)-110)\otimes011+101\otimes([1,0]1)+110\otimes110 \\
&= ((011+101+110)\otimes011+2*((110+011)-011)\otimes([1,0]1)+(110-110)\otimes110 \\
&= (0 \sqcup\!\sqcup 1 \sqcup\!\sqcup 1)\otimes011+(10 \sqcup\!\sqcup 1)\otimes[10]\,1+110\otimes[1[10]]
\end{aligned}$$

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 rather easy to compute a powerful continuous analogue of the Campbell Baker Hausdorff formula [27]