Noncommutative Power Series and Formal Lie-algebraic Techniques in Nonlinear Control Theory

Matthias Kawski[1]
Arizona State University
Tempe, Arizona, U.S.A.

Héctor J. Sussmann[2]
Rutgers University
New Brunswick, NJ, U.S.A.

Abstract: In nonlinear control, it is helpful to choose a formalism well suited to computations involving solutions of controlled differential equations, exponentials of vector fields, and Lie brackets. We show by means of an example —the computation of control variations that give rise to the Legendre-Clebsch condition— how a good choice of formalism, based on expanding diffeomorphisms as products of exponentials, can simplify the calculations. We then describe the algebraic structure underlying the formal part of these calculations, showing that it is based on the theory of formal power series, Lie series, the Chen series —introduced in control theory by M. Fliess— and the formula for the dual basis of a Poincaré-Birkhoff-Witt basis arising from a generalized Hall basis of a free Lie algebra.

1 Introduction

In the theory of nonlinear control systems, A. Agrachev and R. Gamkrelidze introduced a formalism for computing the flow maps arising from various controls, based on "chronological exponentials." It turns out that, underlying this formalism, there is a rich and interesting algebraic structure, involving algebras of formal power series (whose use in control theory was advocated by M. Fliess, cf., e.g, [7]), exponential Lie series, and the Chen-Fliess series. Since controls give rise to diffeomorphisms, it is natural to seek asymptotic expansions of flow maps as products of exponentials.

[1] Supported in part by NSF-grant DMS 93-08289
[2] Supported in part by NSF Grant DMS95-00798 and AFOSR Grant 0923.

We will outline how such an expansion can be obtained, show how it can be applied to the derivation of high-order necessary conditions for optimality, and then describe the basic algebraic fact underlying such an expansion, namely, the formula for the dual basis of a Poincaré-Birkhoff-Witt basis of the universal enveloping algebra of a free Lie algebra.

2 Manifolds, Vector Fields, and Control Systems

Throughout this paper, "smooth" means "of class C^∞," and "manifold" means "finite-dimensional, second-countable smooth manifold without boundary." We use TM, T^*M, T_pM, T_p^*M, to denote, respectively, the tangent and cotangent bundles of a manifold M, and the tangent and cotangent spaces at a point $p \in M$.

If M is a manifold, we use $\mathcal{E}(M)$ to denote the commutative algebra —over $\mathbb{R}$— of smooth real-valued functions on M, topologized in the usual way. (A sequence $\{\varphi_j\}$ converges to a limit φ in $\mathcal{E}(M)$ iff for every $k \geq 0$ and every k-tuple $(X_1, \ldots, X_k)$ of smooth vector fields on M the functions $X_1X_2 \ldots X_k\varphi_j$ converge to $X_1X_2 \ldots X_k\varphi$ uniformly on compact sets.) We let $\mathcal{E}'(M)$ denote the dual space of $\mathcal{E}(M)$, i.e. the space of compactly supported Schwartz distributions on M.

It is well known that M is completely determined by $\mathcal{E}(M)$, in the following sense. A point p of M gives rise to a linear functional $\delta_p \in \mathcal{E}'(M)$ —the Dirac delta function at p— defined by $\delta_p(\varphi) = \varphi(p)$. This linear functional is *multiplicative*, i.e. is a homomorphism of $\mathbb{R}$-algebras. For any commutative algebra A over a field $\mathbf{k}$, let us use $\sigma(A)$ to denote the *spectrum* of A, i.e. the set of all maps $\mu : A \to \mathbf{k}$ that are algebra homomorphisms and do not vanish identically. Then *the map $p \to \delta_p$ is a bijection from M to $\sigma(\mathcal{E}(M))$*. So a manifold can be canonically identified with the spectrum of its algebra of smooth functions.

Via the map $p \to \delta_p$, we regard M as embedded in $\mathcal{E}'(M)$. Moreover, many other objects related to M can also be naturally regarded as members of $\mathcal{E}'(M)$. For example, TM is embedded in $\mathcal{E}'(M)$ as follows: a tangent vector v at a point is a linear functional on $\mathcal{E}(M)$, which maps $\varphi \in \mathcal{E}(M)$ to $v\varphi$. Since this functional is clearly continuous, v belongs to $\mathcal{E}'(M)$.

If X and Y are smooth vector fields, then the product XY, evaluated at p, is a well defined member of $\mathcal{E}'(M)$, namely, the map $\varphi \to (XY\varphi)(p)$. So the product XY —i.e. the map $p \to XY(p)$— is a well defined map from M to $\mathcal{E}'(M)$. The difference $[X,Y] = XY - YX$ —the *Lie bracket* of X and Y— is also a map from M to $\mathcal{E}'(M)$, that happens to take values in TM, and is in fact a vector field.

On the other hand, $\mathcal{E}'(M)$ is a topological linear space, so linear operations and limiting processes that in principle appear not to make intrinsic sense on M can be meaningful in $\mathcal{E}'(M)$. For example, if $\gamma : [0, \varepsilon] \to M$ is a curve, and $\gamma(0) = p$, then we would like to define the tangent vector $\dot\gamma(0)$ by just letting

$$\dot\gamma(0) \stackrel{\text{def}}{=} \lim_{h \to 0} h^{-1}(\gamma(h) - p)\,. \tag{1}$$

This may look unacceptable, since M is not a linear space, and $\gamma(h)-p$ does not make sense. However, Formula (1) is perfectly meaningful, and gives the right answer, if we regard it as an identity in $\mathcal{E}'(M)$, since $\lim_{h\to 0} h^{-1}(\gamma(h)-p)$ is the distribution that maps $\varphi\in\mathcal{E}(M)$ to the number $\lim_{h\to 0} h^{-1}(\varphi(\gamma(h))-\varphi(p))$, i.e. directional differentiation at p in the direction of $\dot\gamma(0)$.

To make the formalism work well, it is convenient to write $\pi\varphi$ —rather than $\pi(\varphi)$— for the value of the linear functional $\pi\in\mathcal{E}'(M)$ at the function $\varphi\in\mathcal{E}(M)$. With this notation, when $p\in M$ and $\varphi\in\mathcal{E}(M)$ the value $\varphi(p)$ is now written $p\varphi$. If f is a smooth vector field and $\varphi\in\mathcal{E}(M)$, then $f\varphi$ is in $\mathcal{E}(M)$, and the number usually written as $(f\varphi)(p)$ now becomes $p(f\varphi)$. On the other hand, if we also write pf rather than the more common $f(p)$, the number $p(f\varphi)$ is also equal to $(pf)\varphi$, so we can just write $pf\varphi$ without parentheses.

More generally, maps Φ from M to any manifold N should act on points of M on the right, so we will write $p\Phi$ rather than $\Phi(p)$. A smooth map $\Phi: M\to N$ gives rise to a map $\tilde\Phi:\mathcal{E}'(M)\to\mathcal{E}'(N)$, defined by letting $\tilde\Phi(D)$, for $D\in\mathcal{E}'(M)$, be the distribution on N given by $\tilde\Phi(D)(\psi)=D(\psi\circ\Phi)$, i.e., in our formalism, $(D\tilde\Phi)\psi\stackrel{\text{def}}{=}D(\Phi\psi)$, where, of course, $\Phi\psi$ is the function usually referred to as the "pullback" $\Phi^*(\psi)$ of ψ to M via Φ. (Naturally, since maps act on the right, $\Phi\psi$ is the composite map obtained by applying Φ first, and then ψ.) When D is actually a point $p\in M$, then $(D\tilde\Phi)\psi$ is what is normally written as $\psi(\Phi(p))$, so $p\tilde\Phi=p\Phi$. So $\tilde\Phi$ is an extension to $\mathcal{E}'(M)$ of the map Φ, originally defined on the subset M of $\mathcal{E}'(M)$. It is convenient to just use Φ for this extension. With this notation, if $p\in M$ and v is a tangent vector at p, then $v\Phi$ is the tangent vector at $p\Phi$ usually written as $\Phi_*(v)$, or $D\Phi(p).v$.

In particular, we will use e^{tf} to denote the time t flow map arising from a vector field f, and will write it as acting on the right. So $t\to pe^{tf}$ is the integral curve $\xi_{f,p}$ of f that goes through p at time 0. This curve satisfies the equation

$$(d/dt)(pe^{tf})=pe^{tf}f\,, \tag{2}$$

which is a much more elegant way of writing the formula $(d/dt)(\xi_{f,p}(t))=f(\xi_{f,p}(t))$, and amply justifies the use of the exponential notation.

With this formalism, several important formulas involving Lie brackets become completely trivial formally, and the formal calculations can be rigorously justified by working in $\mathcal{E}'(M)$. For example, let f,g be vector fields, and let $p\in M$. Then

$$\begin{aligned}&(d/dt)(pe^{tf}e^{tg}e^{-tf}e^{-tg})\\ =\;& pe^{tf}fe^{tg}e^{-tf}e^{-tg}+pe^{tf}e^{tg}ge^{-tf}e^{-tg}-pe^{tf}e^{tg}e^{-tf}fe^{-tg}-pe^{tf}e^{tg}e^{-tf}e^{-tg}g\,,\end{aligned}$$

whose value for $t=0$ is $pf+pg-pf-pg$, i.e. 0.

The second derivative $\frac{d^2}{dt^2}(pe^{tf}e^{tg}e^{-tf}e^{-tg})$ is then equal to

$$\begin{aligned}& pe^{tf}f^2e^{tg}e^{-tf}e^{-tg}+pe^{tf}fe^{tg}ge^{-tf}e^{-tg}-pe^{tf}fe^{tg}e^{-tf}fe^{-tg}-pe^{tf}fe^{tg}e^{-tf}e^{-tg}g\\ &+pe^{tf}fe^{tg}ge^{-tf}e^{-tg}+pe^{tf}e^{tg}g^2e^{-tf}e^{-tg}-pe^{tf}e^{tg}ge^{-tf}fe^{-tg}-pe^{tf}e^{tg}ge^{-tf}e^{-tg}g\\ &-pe^{tf}fe^{tg}e^{-tf}fe^{-tg}-pe^{tf}e^{tg}ge^{-tf}fe^{-tg}+pe^{tf}e^{tg}e^{-tf}f^2e^{-tg}+pe^{tf}e^{tg}e^{-tf}fe^{-tg}g\\ &-pe^{tf}fe^{tg}e^{-tf}e^{-tg}g-pe^{tf}e^{tg}ge^{-tf}e^{-tg}g+pe^{tf}e^{tg}e^{-tf}fe^{-tg}g+pe^{tf}e^{tg}e^{-tf}e^{-tg}g^2\,,\end{aligned}$$

whose value for $t=0$ is $2pfg-2pgf$, i.e. $2p[f,g]$. So we have shown that

$$pe^{tf}e^{tg}e^{-tf}e^{-tg} = p + t^2 p[f,g] + O(t^2) \text{ as } t \to 0\,, \tag{3}$$

which is the familiar formula describing how the Lie bracket measures the failure to close of the "square" described by the left-hand side of (3).

As a second example, we write another familiar formula, namely,

$$(d/dt)(pe^{tf}ge^{-tf}) = pe^{tf}fge^{-tf} - pe^{tf}gfe^{-tf} = pe^{tf}[f,g]e^{-tf}\,, \tag{4}$$

which says that, if we define a vector field v_g along an integral curve of f by letting $v_g(t) = pe^{tf}g$, and then move $v_g(t)$ back to T_pM via the differential of the diffemorphism e^{-tf} —which sends pe^{tf} to p— then the result is the derivative of the pullback of $v_{[f,g]}(t)$ via the differential of e^{-tf}.

Now consider a control system of the form

$$\dot{x} = x(f+ug)\,,\ \ |u|\leq 1\,, \tag{5}$$

where f and g are smooth vector fields on M, and we are using the previously described formalism in which vector fields act on the right.

For a control $\eta : [a,b] \to [-1,1]$, let us use the expression $\boxed{xe^{\overrightarrow{\int}_a^t (f+\eta(s)g)ds}}$ to denote the point $\xi(t)$, if ξ is the trajectory of (5) corresponding to η and the initial condition $\xi(a)=x$. (This "chronological exponential" formalism was introduced by Agrachev and Gamkrelidze. Notice that $xe^{\overrightarrow{\int}_a^t (f+\eta(s)g)ds} = xe^{\int_a^t (f+\eta(s)g)ds}$ if η is constant, and that the derivative of $xe^{\overrightarrow{\int}_a^t (f+\eta(s)g)ds}$ is $xe^{\overrightarrow{\int}_a^t (f+\eta(s)g)ds}(f+\eta(t)g)$ for a. e. t, justifying the use of an exponential notation.)

Given a measurable control $\eta : [a,b] \to [-1,1]$, a *point variation* of η at a time $t_0 \in [a,b]$ is a family $\boldsymbol{\eta} = \{\eta^\varepsilon\}_{0\leq\varepsilon\leq\bar{\varepsilon}}$ of controls $\eta^\varepsilon : [a,b] \to [-1,1]$ such that $\eta^0 = \eta$, having the property that for every $\delta > 0$ there exists $\tilde{\varepsilon}(\delta) \in]0,\bar{\varepsilon}]$ for which the set $\{t : \eta^\varepsilon(t) \neq \eta(t)\} \cap [t_0-\delta, t_0+\delta]$ is empty whenever $\varepsilon \in]0,\tilde{\varepsilon}(\delta)]$.

For a point variation $\boldsymbol{\eta}$ and an $x \in M$ we define the *endpoint curve* $\gamma_{\boldsymbol{\eta},x}$ by

$$\gamma_{\boldsymbol{\eta},x}(\varepsilon) = xe^{\overrightarrow{\int}_a^b (f+\eta^\varepsilon(s)g)ds}\,. \tag{6}$$

If $\gamma_{\boldsymbol{\eta},x}(\varepsilon)$ is continuous near $\varepsilon=0$ and differentiable at $\varepsilon=0$, then the vector

$$\tilde{v}_{\boldsymbol{\eta},x} \stackrel{\text{def}}{=} (d/d\varepsilon)\Big|_{\varepsilon=0}\Big(xe^{\overrightarrow{\int}_a^b (f+\eta^\varepsilon(s)g)ds}\Big) \tag{7}$$

is the *terminal variational vector* corresponding to the variation $\boldsymbol{\eta}$. The pullback of this vector to the point $xe^{\overrightarrow{\int}_a^t (f+\eta(s)g)ds}$ via the diffeomorphism $\Big(e^{\overrightarrow{\int}_t^b (f+\eta(s)g)ds}\Big)^{-1}$ is the *variational vector* $v_{\boldsymbol{\eta},x,t}$ generated at time t by $\boldsymbol{\eta}$ and the initial condition x. So, if $\delta(\varepsilon)\geq 0$ is chosen for every $\varepsilon \in]0,\bar{\varepsilon}]$ in arbitrary fashion, subject only to the condition that $\{t : \eta^\varepsilon(t) \neq \eta(t)\} \subseteq [a, t_0+\delta(\varepsilon)]$, then

$$v_{\boldsymbol{\eta},x,t_0} = (d/d\varepsilon)\Big|_{\varepsilon=0}\Big(xe^{\overrightarrow{\int}_a^{t_0+\delta(\varepsilon)}(f+\eta^\varepsilon(s)g)ds}\Big(e^{\overrightarrow{\int}_{t_0}^{t_0+\delta(\varepsilon)}(f+\eta(s)g)ds}\Big)^{-1}\Big)\,. \tag{8}$$

Point variations and their corresponding variational vectors occur in the Pontryagin Maximum Principle and its "high-order" generalizations. For example, a necessary condition for a trajectory ξ of (5) and corresponding control η to have the property that the reachable set from $\xi(a)$ is not a neighborhood of $\xi(b)$ is that there exist a nontrivial "adjoint vector" —i.e. an absolutely continuous field of covectors $[a,b] \ni t \to p(t) \in T^*_{\xi(t)}M$ along ξ that satisfies the *adjoint equation* $\boxed{\dot{p}(t) = -p(t).(\partial/\partial x)(f+\eta(t)g)(\xi(t))}$ for a.e. t— for which

$$\langle p(t), (f+\eta(t)g)(\xi(t))\rangle = \max\{\langle p(t), (f+ug)(\xi(t))\rangle : -1 \le u \le 1\} \tag{9}$$

for almost every t. Condition (9) is equivalent to the statement that, for almost all t, $\langle p(t), v\rangle \le 0$ whenever v is a tangent vector at $\xi(t)$ such that v is of the form $\xi(t)((u-\eta(t))g)$ for some $u \in [-1,1]$. These vectors v are precisely the variational vectors arising from "needle variations" $\boldsymbol{\eta}_{u,t}$ at time t.

In various "high-order" versions of the Maximum Principle, Condition (9) is replaced by the much stronger requirement that $\langle p(t), v\rangle \le 0$ whenever v is a tangent vector at $\xi(t)$ which is of the form $v_{\boldsymbol{\eta},\xi(a),t}$ for some point variation of η at t. (The precise statement of the high-order maximum principle requires that the variational vectors satisfy an extra "compatibility" condition. We will ignore this complication, and point out that the compatibility condition is always satisfied when all the pairs (t,v) under considerations satisfy *Knobloch's condition* (cf. Knobloch [10]): $v = \lim_{j\to\infty} v_j$, where the v_j are variational vectors arising from point variations $\boldsymbol{\eta}_j$ at time t_j and $t_j \to t$, $t_j \ne t$.)

Theorem 2.1 *Let $\xi : [a,b] \to M$ be a trajectory of (5) corresponding to a measurable control $\eta : [a,b] \to [-1,1]$. Let $\bar{t} \in [a,b[$ be such that $|\eta(\bar{t})| < 1$ and $\bar{t}$ is a Lebesgue point of η. Then there exists a point variation $\boldsymbol{\eta}$ of η at $\bar{t}$ that gives rise to a variational vector $v = v_{\boldsymbol{\eta},\xi(a),\bar{t}} = -[g,[f,g]](\xi(\bar{t}))$.*

The vector v given by the above theorem clearly satisfies Knobloch's condition, since the set of Lebesgue points of η has full measure, and every interval $[\bar{t}, \bar{t}+\varepsilon]$ must contain, for small enough ε, a subset of positive measure where $|\eta| < 1$, since otherwise the integral $\int_{\bar{t}}^{\bar{t}+\varepsilon} |\eta(t) - \eta(\bar{t})| dt$ would be bounded below by a positive constant times ε, contradicting the hypothesis that $\bar{t}$ is a Lebesgue point of η.

Theorem 2.1 says that at almost all times t such that $|\eta(t)| < 1$ the adjoint vector p may be required to satisfy the inequality $\langle p(t), [g,[f,g]](\xi(t))\rangle \ge 0$, in addition to the condition that $\langle p(t), g(\xi(t))\rangle = 0$, which follows from (9). Since the Hamiltonian H is given by $H(x,p,u) = \langle p, (f+ug)(x)\rangle$, we see that the new inequality says that $(\partial/\partial u)(d^2/dt^2)(\partial H/\partial u) \ge 0$, along our trajectory, which is the usual Legendre-Clebsch condition.

Proof of Theorem 1. Without loss of generality, we assume that $\bar{t} = 0$. We are going to construct our variation $\boldsymbol{\eta}$ by letting η^ε be, for sufficiently small ε, a control of the form $\eta^\varepsilon = \eta + \theta^\rho$, with $\rho = k\varepsilon^{1/3}$, k a constant to be chosen later, and θ^ρ a function that vanishes outside the interval $[0,\rho]$. Once we have explained how to choose θ^ρ, we will want to study the effect of the resulting control $\eta + \theta^\rho$ on the interval $[0,\rho]$, and for this purpose we will first look at an arbitrary control

$u:[0,\rho]\to[-1,1]$ and study the map $e^{\overrightarrow{\int_0^\rho}(f+u(t)g)dt}$. In particular, we will want to compute a product asymptotic expression for his map that will be exact modulo errors of order $o(\rho^3)$. Once this is done, we will try to choose θ^ρ so as to match all the factors of this expansion except one, and the one factor that is not matched will become the leading term of the variation we want.

So, let $u:[0,\rho]\to[-1,1]$ be measurable, and write $x_1(t)=x\,e^{\overrightarrow{\int_0^t}(f+u(s)g)ds}$, for a fixed $x\in M$. Let $x_1(t)=x_2(t)e^{tf}$. Then $\dot x_1(t)=\dot x_2(t)e^{tf}+x_2(t)e^{tf}f=\dot x_2(t)e^{tf}+x_1(t)f$. Comparing this with $\dot x_1(t)=x_1(t)(f+u(t)g)$, we find $\dot x_2(t)e^{tf}=x_1(t)u(t)g$, so $\dot x_2(t)=x_2(t)e^{tf}u(t)ge^{-tf}$.

Now, if X, Y are vector fields, then $(d/dt)e^{tY}Xe^{-tY}=e^{tY}[Y,X]e^{-tY}$ (cf. (4)), so $e^{tY}Xe^{-tY}=X+R_1(X,Y,t)$, where $R_1(X,Y,t)=\int_0^t e^{sY}[Y,X]e^{-sY}ds$.

Iterating this we get

$$e^{tY}Xe^{-tY}=X+t[Y,X]+R_2(X,Y,t)\,, \tag{10}$$

where
$$R_2(X,Y,t)=\int_0^t\int_0^{s_1}e^{s_2Y}[Y,[Y,X]]e^{-s_2Y}ds_2ds_1\,. \tag{11}$$

Then
$$e^{tY}Xe^{-tY}=X+t[Y,X]+\frac{t^2}{2}[Y,[Y,X]]+R_3(X,Y,t)\,, \tag{12}$$

where
$$R_3(X,Y,t)=\int_0^t\int_0^{s_1}\int_0^{s_2}e^{s_3Y}[Y,[Y,[Y,X]]]e^{-s_3Y}ds_3ds_2ds_1\,. \tag{13}$$

Then
$$e^{tf}ge^{-tf}=g+t[f,g]+\frac{t^2}{2}[f,[f,g]]+R_3(g,f,t)\,, \tag{14}$$

so $\dot x_2(t)=x_2(t)(u(t)g+X_t)$, with $X_t=tu(t)[f,g]+\frac{t^2u(t)}{2}[f,[f,g]]+u(t)R_3(g,f,t)$.

Now let $\boxed{U_1(t)=\int_0^t u(s)ds}$, and write $x_2(t)=x_3(t)e^{U_1(t)g}$. Then

$$\dot x_2(t)=\dot x_3(t)e^{U_1(t)g}+x_3(t)e^{U_1(t)g}u(t)g=\dot x_3(t)e^{U_1(t)g}+x_2(t)u(t)g\,. \tag{15}$$

So
$$\dot x_3(t)=x_3(t)e^{U_1(t)g}X_te^{-U_1(t)g}\,. \tag{16}$$

We now use (10) and write $e^{U_1(t)g}X_te^{-U_1(t)g}=X_t+U_1(t)[g,X_t]+R_2(X_t,g,U_1(t))$, so

$$\begin{aligned}\dot x_3(t)&=x_3(t)\Big(X_t+U_1(t)[g,X_t]+R_2(X_t,g,U_1(t))\Big)\\&=x_3(t)\Big(tu(t)[f,g]+\tfrac{t^2u(t)}{2}[f,[f,g]]+tu(t)U_1(t)[g,[f,g]]+S_1(t)\Big),\end{aligned}$$

where $S_1(t)$ is a complicated expression which is $O(t^3)$. Now define U_2 by letting $\boxed{U_2(t)=\int_0^t su(s)ds}$. Then write $x_3(t)=x_4(t)e^{U_2(t)[f,g]}$. We then get

$$\dot x_4(t)=x_4(t)\Big(\frac{t^2u(t)}{2}[f,[f,g]]+tu(t)U_1(t)[g,[f,g]]+S_2(t)\Big)\,, \tag{17}$$

where $S_2(t)$ is also $O(t^3)$. Next, define $\boxed{U_3(t)=\frac{1}{2}\int_0^t s^2u(s)ds}$, and then write $x_4(t)=x_5(t)e^{U_3(t)[f,[f,g]]}$. Then

$$\dot x_5(t)=x_5(t)\Big(tu(t)U_1(t)[g,[f,g]]+S_3(t)\Big)\,, \tag{18}$$

where $S_3(t)$ is also $O(t^3)$. Finally, if we let $\boxed{U_4(t)=\int_0^t su(s)U_1(s)ds}$, we can write $x_5(t)=x_6(t)e^{U_4(t)[g,[f,g]]}$, and conclude that $\dot x_6(t)=x_6(t)S_4(t)$, where $S_4(t)$ is $O(t^3)$. Since $x_6(0)=x$, we have $x_6(t)=x+O(t^4)$. Then

$$\boxed{x_1(t)=xe^{U_4(t)[g,[f,g]]}e^{U_3(t)[f,[f,g]]}e^{U_2(t)[f,g]}e^{U_1(t)g}e^{tf}+O(t^4)}\,, \tag{19}$$

i.e. $$\boxed{e^{\int_0^t (f+u(s)g)ds} = e^{U_4(t)[g,[f,g]]}e^{U_3(t)[f,[f,g]]}e^{U_2(t)[f,g]}e^{U_1(t)g}e^{tf} + O(t^4)}\,. \qquad (20)$$

In Formulas (19), (20), the product of five exponentials that appears in the right-hand side is the initial part —up to terms of degree 3— of the general *product expansion of the Chen series*, a purely formal Lie-algebraic result that gives, on a formal level, an expansion of a the diffeomorphism $e^{\int_0^t (f+u(s)g)ds}$ as a product of diffeomorphisms. The formulas themselves are special cases of the general proposition that *the Chen series associated to an expression such as* $e^{\int_0^t (f+u(s)g)ds}$ *gives an asymptotic expansion for this expression.*

We now use these asymptotic formulas to construct our variation. In our computations, we will use U_i^ζ to denote, for $i = 1, 2, 3, 4$, the functions U_i arising from a particular control ζ.

Fix once and for all a $c > 0$ such that $-1 \le \eta(0) - 2c$ and $\eta(0) + 2c \le 1$. Define the "good set" G_ρ and the "bad set" B_ρ by

$$G_\rho \stackrel{\text{def}}{=} \Big\{t \in [0,\rho] : |\eta(t) - \eta(0)| \le c\Big\}, \quad B_\rho = [0,\rho]\backslash G_\rho\,. \qquad (21)$$

Use $|E|$ to denote the Lebesgue measure of a measurable subset E of $\mathbb{R}$. Then $c|B_\rho| \le \int_0^\rho |\eta(t) - \eta(0)|dt = o(\rho)$, so $|B_\rho| = o(\rho)$, and then $|G_\rho| = \rho - o(\rho)$.

Let $P^\rho(t) = p_0^\rho + p_1^\rho t + p_2^\rho t^2 + p_3^\rho t^3$ be the unique cubic polynomial such that

$$\int_{G_\rho} t^i P^\rho(t)\, dt = \rho^4 \delta_{i3} \quad \text{for} \quad i = 0, 1, 2, 3\,, \qquad (22)$$

where "δ" is Kronecker's delta. (Clearly, P^ρ exists and is unique for small ρ.)

If $\hat{G}_\rho = \{s\in[0,1] : \rho s\in G_\rho\}$, then $|\hat{G}_\rho|\to 1$ as $\rho\to 0$. Define $\hat{G}^0 = [0,1]$. For $\rho \ge 0$, let $Q^\rho(s) = \sum_{i=0}^3 q_i^\rho t^i$ be such that $\int_{\hat{G}_\rho} Q^\rho(s) s^i ds = \delta_{i3}$ for $i = 0, 1, 2, 3$. Then

$$P^\rho(t) = Q^\rho(\rho^{-1}t) = \sum_{i=0}^3 q_i^\rho \rho^{-i} t^i \quad \text{if} \quad \rho > 0\,. \qquad (23)$$

As $\rho\to 0$, the q_i^ρ converge to q_i^0. (Explicitly, $Q^0(t) = 2800t^3 - 4200t^2 + 1680t - 140$.) Let $\kappa = 1 + \max\{|Q^0(s)| : 0 \le s \le 1\}$. Then $|Q^\rho(t)| \le \kappa$ for all $t \in [0,1]$, if ρ is small.

Take $$\theta^\rho(t) = \frac{c}{\kappa} P^\rho(t) \;\text{ for }\; t \in G_\delta\,, \quad \theta^\rho(t) = 0 \;\text{ for }\; t \notin G_\delta\,. \qquad (24)$$

Let $\eta^\varepsilon = \eta + \theta^\rho$. Then η^ε is admissible if ε is small, and $U_i^{\eta^\varepsilon}(\rho) = U_i^\eta(\rho)$ for $i = 1, 2, 3$.

We now compute $U_4^{\eta^\varepsilon}(\rho) - U_4^\eta(\rho)$. For a general control ζ, we have

$$U_4^\zeta(t) = \int_0^t s\zeta(s)U_1^\zeta(s)ds = \frac{1}{2}\int_0^t s\frac{d}{ds}(U_1^\zeta(s))^2 ds\,. \qquad (25)$$

Integration by parts then yields $U_4^\zeta(\rho) = \frac{\rho}{2}(U_1^\zeta(\rho))^2 - \frac{1}{2}\int_0^\rho (U_1^\zeta(s))^2 ds$. Since $U_1^{\eta^\varepsilon}(\rho)$ is equal to $U_1^\eta(\rho)$, we have $U_4^{\eta^\varepsilon}(\rho) - U_4^\eta(\rho) = \frac{1}{2}\int_0^t \Big(U_1^\eta(s)^2 - U_1^{\eta^\varepsilon}(s)^2\Big)ds$.

If we let $\Theta^\rho(t) = \int_0^t \theta^\rho(s)ds$, then $U_1^{\eta^\varepsilon}(t) = U_1^\eta(t) + \Theta^\rho(t)$, from which it follows that $U_1^{\eta^\varepsilon}(t)^2 = U_1^\eta(t)^2 + \Theta^\rho(t)^2 + 2U_1^\eta(t)\Theta^\rho(t)$, and then

$$U_4^{\eta^\varepsilon}(\rho) - U_4^\eta(\rho) = -\frac{1}{2}\int_0^\rho \Theta^\rho(s)^2 ds - \int_0^\rho U_1^\eta(s)\Theta^\rho(s)ds\,. \qquad (26)$$

Now, if $0 \leq s \leq 1$, we have

$$\Theta^\rho(\rho s) = \int_0^{\rho s} \theta^\rho(w)dw = \frac{c}{\kappa}\int_0^{\rho s} P^\rho(w)dw = \frac{c\rho}{\kappa}\int_0^s P^\rho(\rho t)dt = \frac{c\rho}{\kappa}\int_0^s Q^\rho(t)dt\,. \quad (27)$$

Then

$$\int_0^\rho \Theta^\rho(t)^2 dt = \rho\int_0^1 \Theta^\rho(\rho s)^2 ds = \frac{c^2\rho^3}{\kappa^2}\int_0^1\left(\int_0^s Q^\rho(t)dt\right)^2 ds = \frac{\nu c^2\rho^3}{\kappa^2} + o(\rho^3)\,, \quad (28)$$

where $\nu = \int_0^1 \left(\int_0^s Q^0(t)dt\right)^2 ds$.

On the other hand, the integral $I = \int_0^\rho U_1^\eta(s)\Theta^\rho(s)ds$ can also be computed by parts, and is equal to $V^\eta(\rho)\Theta^\rho(\rho) - \int_0^\rho V^\eta(s)\theta^\rho(s)ds$, where $V^\eta(t) = \int_0^t U_1^\eta(s)ds$. Since $\Theta^\rho(\rho) = \int_0^\rho \theta^\rho(t)dt = 0$, I is in fact equal to $-\int_0^\rho V^\eta(s)\theta^\rho(s)ds$. Since $U_1^\eta(t) = \int_0^t \eta(s)ds$, we have $U_1^\eta(t) = t\eta(0) + o(t)$. Then $V^\eta(t) = \frac{t^2}{2}\eta(0) + o(t^2)$. Since $\int_0^\rho t^2\theta^\rho(t)dt = 0$, we have $I = o(\rho^3)$.

Therefore

$$U_4^{\eta^\varepsilon}(\rho) - U_4^\eta(\rho) = -\tfrac{\nu c^2\rho^3}{2\kappa^2} + o(\rho^3) = -\tfrac{\nu c^2 k^3\varepsilon}{2\kappa^2} + o(\varepsilon).$$

Choosing k such that $\nu c^2 k^3 = 2\kappa^2$, we have $U_4^{\eta^\varepsilon}(\rho) - U_4^\eta(\rho) = -\varepsilon + o(\rho^3)$. Then, if $x \in M$, we have

$$x e^{\int_0^\rho (f+\eta^\varepsilon(s)g)ds}\left(e^{\int_0^\rho (f+\eta(s)g)ds}\right)^{-1} = -\varepsilon\, x\,[g,[f,g]] + o(\varepsilon)\,,$$

and our proof is complete.

3 Basic algebraic structures

In this and the following sections, we work with a fixed field $\mathbf{k}$ *of scalars, assumed to be of characteristic zero. We use LS, AA, LA, CA, CP as abbreviations for "linear space," "associative algebra," "Lie algebra," "chronological algebra," and "chronological product," respectively. All LS's, AA's, LA's, and CA's assumed to be over* $\mathbf{k}$*. Every AA* $\mathcal{A}$ *is automatically regarded as a LA, with the Lie bracket* $[x,y]$ *of* $x, y \in \mathcal{A}$ *defined by* $[x,y] = xy - yx$*.*

The calculations of the preceding sections were carried out for vector fields, but it is clear that a large part of what was done was purely formal. The purpose of this section is to exhibit the basic algebraic structures underlying the formal part of our arguments, to review some of their properties, and to fix terminology and notation. For a detailed description the reader is referred to [16].

We recall that, if L is a LA, then a *universal enveloping algebra* (abbr. UAE) of L is a pair (μ, U) such that U is an AA and $\mu : L \to U$ is a LA-homomorphism, having the property that, if $\mu' : L \to U'$ is any other LA-homomorphism from L to an AA U', then there exists a unique AA-homomorphism $\nu : U \to U'$ such that $\nu\mu = \mu'$. The existence of (μ, U) is trivial, since one can let U be the quotient of the tensor algebra $T(L)$ over L by the two-sided ideal generated by the elements $x\otimes y - y\otimes x - [x,y]$, for $x, y \in L$. The *Poincaré-Birkhoff-Witt Theorem* (abbr. PBWT) says that, if (μ, U) is a UEA of L, $\mathcal{B}$ is any basis of L, endowed with a total ordering $\preceq$, and we use $PWB(\mathcal{B},\preceq)$ to denote the set of all products $\mu(B_1)\cdot\mu(B_2)\cdot\ldots\cdot\mu(B_m)$, for all finite sequences $(B_1, B_2, \ldots, B_m)$ of members of $\mathcal{B}$ such that $B_1 \preceq B_2 \preceq \ldots \preceq B_m$, then $PWB(\mathcal{B},\preceq)$ is a basis of U. It follows from the PBWT that μ *is necessarily injective.* In view of this, we will only consider from now on UEA's (μ, U) such that $L \subseteq U$ and $\mu : L \to U$ is the inclusion map.

Let $\mathbf{X}$ be a set (also called an *alphabet*) of noncommuting indeterminates (or *letters*). Use $W(\mathbf{X})$ to denote the set $\cup_{n=0}^{\infty}\mathbf{X}^n$, i.e. the union of the Cartesian products $\mathbf{X}^n$ of n copies of $\mathbf{X}$. We refer to $W(\mathbf{X})$ as the *free monoid* generated by $\mathbf{X}$. The members of $W(\mathbf{X})$ —called *monomials*, or *words*— are the finite sequences $w = (x_1, x_2, \ldots, x_n)$ of letters, for which we will use the notation $w = x_1x_2\ldots x_n$. If $w \in \mathbf{X}^n$ then w is a *word of length* n, or a *monomial of degree* n, and we write $n = |w|$. The *product* ww' of $w = x_1x_2\ldots x_n$, $w' = x'_1x'_2\ldots x'_\nu$ is just the concatenation $x_1x_2\ldots x_nx'_1x'_2\ldots x'_\nu$. Clearly, $|ww'| = |w| + |w'|$. Also, we write $lett(w)$ to denote the set of all letters $x \in \mathbf{X}$ that occur in w. We write 1 —or $\emptyset$— to denote the only word of length zero, i.e. the empty word. Clearly, $1w = w1 = w$ for all $w \in W(\mathbf{X})$. We use $W^+(\mathbf{X})$ to denote the set $\{w \in W(\mathbf{X}) : w \neq 1\}$.

We let $A_{\mathbf{k}}(\mathbf{X})$ denote the free AA generated by $\mathbf{X}$ with coefficients in $\mathbf{k}$. Then $A_{\mathbf{k}}(\mathbf{X})$ is the set of all formal linear combinations $p = \sum_{w\in W(\mathbf{X})} p_w w$ of monomials with coefficients $p_w \in \mathbf{k}$ such that the set $\{w : p_w \neq 0\}$ is finite. The members of $A_{\mathbf{k}}(\mathbf{X})$ are the *noncommuting polynomials* in the letters of $\mathbf{X}$. We let $A^+_{\mathbf{k}}(\mathbf{X}) \stackrel{\text{def}}{=} \{P = \sum_w p_w w \in A_{\mathbf{k}}(\mathbf{X}) : p_1 = 0\}$, so $A^+_{\mathbf{k}}(\mathbf{X})$ is a two-sided ideal of $A_{\mathbf{k}}(\mathbf{X})$.

We use $\hat{A}_{\mathbf{k}}(\mathbf{X})$ to denote the completion of $A_{\mathbf{k}}(\mathbf{X})$ with respect to the uniform structure in which a basis of "entourages" of the diagonal is given by the sets $\mathcal{U}_F = \{(P,Q) \in A_{\mathbf{k}}(\mathbf{X}) \times A_{\mathbf{k}}(\mathbf{X}) : P - Q \in U_F\}$, for all finite subsets F of $W(\mathbf{X})$, where $U_F = \{P = \sum_w p_w w \in A_{\mathbf{k}}(\mathbf{X}) : p_w = 0 \text{ whenever } w \in F\}$. Then $A_{\mathbf{k}}(\mathbf{X})$ is the set of all sums $S = \sum_w s_w w$, with $\{s_w\}_{w\in W(\mathbf{X})}$ an arbitrary family of members of $\mathbf{k}$ (so in particular $\{w : s_w \neq 0\}$ is not required to be finite). The members of $\hat{A}_{\mathbf{k}}(\mathbf{X})$ are then the *formal power series in the letters of* $\mathbf{X}$ *with coefficients in* $\mathbf{k}$. The multiplication map $(P,Q) \to PQ$ from $A_{\mathbf{k}}(\mathbf{X}) \times A_{\mathbf{k}}(\mathbf{X})$ to $A_{\mathbf{k}}(\mathbf{X})$ is uniformly continuous, so it extends to a continuous map —also denoted by $(S,T) \to ST$— from $\hat{A}_{\mathbf{k}}(\mathbf{X}) \times \hat{A}_{\mathbf{k}}(\mathbf{X})$ to $\hat{A}_{\mathbf{k}}(\mathbf{X})$, with respect to which $\hat{A}_{\mathbf{k}}(\mathbf{X})$ is an AA.

Associated to the basis $W(\mathbf{X})$ of $A_{\mathbf{k}}(\mathbf{X})$ is a pairing $<\cdot,\cdot>: \hat{A}_{\mathbf{k}}(\mathbf{X}) \times A_{\mathbf{k}}(\mathbf{X}) \to \mathbf{k}$, that sends $(S,T) \in \hat{A}_{\mathbf{k}}(\mathbf{X}) \times A_{\mathbf{k}}(\mathbf{X})$ to $< S,T > = \sum_w S_w T_w$, if $S = \sum_w S_w w$ and $T = \sum_w T_w w$. Then $S = \sum_w < S,w > w$ if $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, so $< S,T > = \sum_w < S,w >< T,w >$. With the pairing $<\cdot,\cdot>$, $\hat{A}_{\mathbf{k}}(\mathbf{X})$ is the algebraic dual of $A_{\mathbf{k}}(\mathbf{X})$. Moreover, the continuous linear functionals on $\hat{A}_{\mathbf{k}}(\mathbf{X})$ are exactly the maps $S \to < S,T >$, for $T \in A_{\mathbf{k}}(\mathbf{X})$, so $A_{\mathbf{k}}(\mathbf{X})$ is the topological dual of $\hat{A}_{\mathbf{k}}(\mathbf{X})$.

We use $L_{\mathbf{k}}(\mathbf{X})$ to denote the Lie subalgebra generated by $\mathbf{X}$ of the LA $A_{\mathbf{k}}(\mathbf{X})$. It is well known —and follows easily from the PBWT— that $L_{\mathbf{k}}(\mathbf{X})$ is a realization of the *free LA over* $\mathbf{X}$. (That is, if Λ is any LA over $\mathbf{k}$ and $\mu : \mathbf{X} \to \Lambda$ is a mapping, then μ can be extended in a unique way to a LA-homomorphism $\tilde{\mu} : L_{\mathbf{k}}(\mathbf{X}) \to \Lambda$.)

It is also easy to show that $A_{\mathbf{k}}(\mathbf{X})$ is a UAE of $L_{\mathbf{k}}(\mathbf{X})$.

A formal series $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$ is called a *Lie series* if for every n the homogeneous component $S^{hom,n} \stackrel{\text{def}}{=} \sum_{|w|=n} <S, w> w$ is in $L_{\mathbf{k}}(\mathbf{X})$. We use $\hat{L}_{\mathbf{k}}(\mathbf{X})$ to denote the set of all Lie series. It is easy to see that $\hat{L}_{\mathbf{k}}(\mathbf{X})$ is a Lie subalgebra of $\hat{A}_{\mathbf{k}}(\mathbf{X})$.

The family $\{v \otimes w\}_{v,w \in W(\mathbf{X})}$ is a basis of the tensor product $A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$, so every $P \in A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$ is expressible in a unique way as a sum $P = \sum_{v,w} p_{v,w} v \otimes w$ such that $\{(v, w) : p_{v,w} \neq 0\}$ is finite. We use $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$ to denote the completion of $A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$ with respect to the uniform structure in which a basis of "entourages" of the diagonal is given by the sets

$$\mathcal{U}_F^2 = \{(P, Q) \in (A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})) \times (A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})) : P - Q \in U_F^2\},$$

for all finite subsets F of $W(\mathbf{X}) \times W(\mathbf{X})$, where

$$U_F^2 = \{P = \textstyle\sum_{v,w} p_{v,w} v \otimes w \in A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X}) : p_{v,w} = 0 \text{ whenever } (v, w) \in F\}.$$

Then $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$ can be thought of as the set of all sums $S = \sum_{v,w} s_{v,w} v \otimes w$, with $\{s_{v,w}\}_{v,w \in W(\mathbf{X})}$ a completely arbitrary family of members of $\mathbf{k}$ (so in particular $\{(v, w) : s_{v,w} \neq 0\}$ is not required to be finite). The members of $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$ are exactly the formal sums $S = \sum_{v \in W(\mathbf{X})} v \otimes S_v$, with all the S_v in $\hat{A}_{\mathbf{k}}(\mathbf{X})$, so $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$ is the space $\hat{A}_{\hat{A}_{\mathbf{k}}(\mathbf{X})}(\mathbf{X})$ of formal power series in $\mathbf{X}$ with coefficients in $\hat{A}_{\mathbf{k}}(\mathbf{X})$. There is a natural pairing $(S, P) \to < S, P >$ that sends $S \in A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$, $P \in A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$ to $< S, P > = \sum_{v,w} s_{v,w} p_{v,w} \in \mathbf{k}$, if $S = \sum_{v,w} s_{v,w} v \otimes w \in A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$, $P = \sum_{v,w} p_{v,w} v \otimes w \in A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$. With respect to this pairing, $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$ is the algebraic dual of $A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$, and $A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$ is the topological dual of $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$.

An important subspace of $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$ is $A_{\mathbf{k}}(\mathbf{X}) \tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$, the set of all sums $S = \sum_{v,w} s_{v,w} v \otimes w$ such that for each v the set $\{w : s_{v,w} \neq 0\}$ is finite. The members of $A_{\mathbf{k}}(\mathbf{X}) \tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$ are the sums $S = \sum_{v \in W(\mathbf{X})} v \otimes S_v$ with all the S_v in $A_{\mathbf{k}}(\mathbf{X})$, so $A_{\mathbf{k}}(\mathbf{X}) \tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$ is the space $\hat{A}_{A_{\mathbf{k}}(\mathbf{X})}(\mathbf{X})$ of formal power series in $\mathbf{X}$ with coefficients in $A_{\mathbf{k}}(\mathbf{X})$. Finally, the tensor product $\hat{A}_{\mathbf{k}}(\mathbf{X}) \otimes \hat{A}_{\mathbf{k}}(\mathbf{X})$ is naturally identified with the set of all sums $S = \sum_{v,w} s_{v,w} v \otimes w$ such that the matrix $(s_{v,w})_{v,w \in W(\mathbf{X})}$ has finite rank.

Clearly, $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$ is an AA, in which the product PQ is given by $PQ = \sum_{v,w} (PQ)_{v,w} v \otimes w$, where

$$(PQ)_{v,w} = \textstyle\sum_{v',v'',w',w'' \in W(\mathbf{X}),\, v'v''=v,\, w'w''=w} < v' \otimes w', P > < v'' \otimes w'', W > .$$

(The summations defining the $(PQ)_{v,w}$ are clearly finite.) Then $A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$, $A_{\mathbf{k}}(\mathbf{X}) \tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$, and $\hat{A}_{\mathbf{k}}(\mathbf{X}) \otimes \hat{A}_{\mathbf{k}}(\mathbf{X})$ are subalgebras of $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$.

The *diagonal map* $\Delta : A_{\mathbf{k}}(\mathbf{X}) \to A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X})$ is the $\mathbf{k}$-algebra homomorphism defined on generators $x \in \mathbf{X}$ by $\Delta(x) = x \otimes 1 + 1 \otimes x$. It is easy to see that Δ is uniformly continuous. So Δ extends uniquely to a continuous algebra homomorphism —also called Δ— from $\hat{A}_{\mathbf{k}}(\mathbf{X})$ to $A_{\mathbf{k}}(\mathbf{X}) \hat{\otimes} A_{\mathbf{k}}(\mathbf{X})$. Clearly, $\Delta(S) = \sum_n \Delta(S^{hom,n})$ for $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$.

The well known *Friedrichs' criterion* —which is a fairly easy consequence of the PBWT— says that, if $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, then $S \in \hat{L}_{\mathbf{k}}(\mathbf{X})$ iff $\boxed{\Delta(S) = S \otimes 1 + 1 \otimes S}$.

The *shuffle product* is the bilinear map $ш : A_{\mathbf{k}}(\mathbf{X}) \times A_{\mathbf{k}}(\mathbf{X}) \to A_{\mathbf{k}}(\mathbf{X})$ such that

$$< S, v ш w > = < \Delta(S), v \otimes w > \qquad \text{for } S \in \hat{A}_{\mathbf{k}}(\mathbf{X})\,, \;\; v, w \in A_{\mathbf{k}}(\mathbf{X})\,. \tag{29}$$

So $ш$, regarded as a linear map $A_{\mathbf{k}}(\mathbf{X}) \otimes A_{\mathbf{k}}(\mathbf{X}) \to A_{\mathbf{k}}(\mathbf{X})$, is the transpose of Δ.

It is also possible to characterize the shuffle product recursively, by first letting $w ш 1 = 1 ш w = w$, and then defining $(xv) ш (yw) = x(v ш (yw)) + y((xv) ш w)$ for $x, y \in \mathbf{X}$, $v, w \in W(\mathbf{X})$. It is easy to show that $A_{\mathbf{k}}(\mathbf{X})$, endowed with $ш$, is an associative and commutative algebra.

Friedrichs' criterion is equivalent to the statement that *an element* $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$ *is a Lie series if and only if (i)* $< S, 1 > = 0$, *and (ii)* S *is orthogonal to all nontrivial shuffles, i.e.* $< S, v ш w > = 0$ *for all* $v, w \in W^+(\mathbf{X})$.

Next, we let $\hat{G}_{\mathbf{k}}(\mathbf{X}) \subseteq \hat{A}_{\mathbf{k}}(\mathbf{X})$ denote the set of all *exponential Lie series*, that is the set of all formal power series $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$ such that there is a $Z \in \hat{L}_{\mathbf{k}}(\mathbf{X})$ for which $S = \exp(Z) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{1}{k!} Z^k$. Friedrichs' criterion easily implies that a series $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$ is in $\hat{G}_{\mathbf{k}}(\mathbf{X})$ if and only if $S \neq 0$ and $\Delta(S) = S \otimes S$. From this it follows in particular that if $S \in \hat{G}_{\mathbf{k}}(\mathbf{X})$ then $< S, 1 > = 1$, so $S^{-1} = \sum_{k=0}^{\infty} (1 - S)^k$ exists. It also follows that $\hat{G}_{\mathbf{k}}(\mathbf{X})$ is a group under multiplication. (The fact that $S_1, S_2 \in \hat{G}_{\mathbf{k}}(\mathbf{X})$ implies $S_1 S_2 \in \hat{G}_{\mathbf{k}}(\mathbf{X})$ is the well known *Campbell-Hausdorff formula.* If $S_i = \exp(Z_i)$ for $i = 1, 2$, then $S_1 S_2 = \exp(\mathbf{P}(Z_1, Z_2))$, where $\boxed{\mathbf{P}(x, y) = x + y + \frac{1}{2}[x, y] + \frac{1}{12}[x, [x, y]] + \frac{1}{12}[y, [x, y]] + \ldots}$.)

Since $< \Delta(S), v \otimes w > = < S, v ш w >$, the condition that $\Delta(S) = S \otimes S$ is equivalent to the property that $< S, v ш w > = < S, v >< S, w >$ for all words v, w. i.e. that "the coefficients $< S, w >$ of S satisfy the shuffle relations." This observation, known as *Ree's theorem*, says that $S \in \hat{G}_{\mathbf{k}}(\mathbf{X})$ if and only if the linear map $T \to < S, T >$, from $A_{\mathbf{k}}(\mathbf{X})$ to $\mathbf{k}$, is nonzero and multiplicative (with respect to $ш$), i.e. is an algebra homomorphism from $(A_{\mathbf{k}}(\mathbf{X}), ш)$ to $\mathbf{k}$ that sends 1 to 1.

There is a clear analogy between the facts of the previous paragraphs and our discussion in §2. The commutative algebra $A_{\mathbf{k}}(\mathbf{X})$ can be realized as an algebra of functions on $\hat{G}_{\mathbf{k}}(\mathbf{X})$, by mapping each $P \in A_{\mathbf{k}}(\mathbf{X})$ to the function $\tilde{P}$ given by $\tilde{P}(S) = < S, P >$ for $S \in \hat{G}_{\mathbf{k}}(\mathbf{X})$. Since $< S, P_1 >< S, P_2 > = < S, P_1 ш P_2 >$ for $S \in \hat{G}_{\mathbf{k}}(\mathbf{X})$, we see that under the map $P \to \tilde{P}$ the shuffle product in $A_{\mathbf{k}}(\mathbf{X})$ corresponds to ordinary pointwise multiplication of functions on $\hat{G}_{\mathbf{k}}(\mathbf{X})$. Moreover, $\hat{G}_{\mathbf{k}}(\mathbf{X})$ is embedded in $\hat{A}_{\mathbf{k}}(\mathbf{X})$, the dual of $A_{\mathbf{k}}(\mathbf{X})$, and Ree's theorem tells us that $\hat{G}_{\mathbf{k}}(\mathbf{X})$ is exactly the spectrum of $A_{\mathbf{k}}(\mathbf{X})$, i.e. that the nonzero linear functionals $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$ that are multiplicative are exactly those that belong to $\hat{G}_{\mathbf{k}}(\mathbf{X})$.

So $\hat{G}_{\mathbf{k}}(\mathbf{X})$ may be regarded as a formal analogue of the manifold M of §2, with $A_{\mathbf{k}}(\mathbf{X})$ playing the role of $\mathcal{E}(M)$ and $\hat{A}_{\mathbf{k}}(\mathbf{X})$ that of $\mathcal{E}'(M)$. So it is natural to call the elements of $\hat{G}_{\mathbf{k}}(\mathbf{X})$ *formal points.* Clearly, $\hat{G}_{\mathbf{k}}(\mathbf{X})$ is a "formal Lie group," and $\hat{L}_{\mathbf{k}}(\mathbf{X})$ is its "Lie algebra." Pursuing our analogy, we define a *formal tangent vector* to $\hat{G}_{\mathbf{k}}(\mathbf{X})$ at a point $S \in \hat{G}_{\mathbf{k}}(\mathbf{X})$ to be a linear functional $V : A_{\mathbf{k}}(\mathbf{X}) \to \mathbf{k}$ such that $V(P \sqcup\!\sqcup Q) = V(P)Q(S) + V(Q)P(S)$ for all $P, Q \in A_{\mathbf{k}}(\mathbf{X})$. Using the identification of $\hat{A}_{\mathbf{k}}(\mathbf{X})$ with the dual of $A_{\mathbf{k}}(\mathbf{X})$, the linear functional V is of the form $P \to \,<W, P>$ for some $W \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, and we can write $W = SZ$, for $Z \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, since S is invertible. Then the functional $P \to \,<SZ, P>$ is a formal tangent vector at S if and only if $<SZ, P \sqcup\!\sqcup Q> = <SZ, P> <S, Q> + <S, P> <SZ, Q>$ for $P, Q \in A_{\mathbf{k}}(\mathbf{X})$, i.e. iff $<\Delta(SZ), P \otimes Q> = <SZ \otimes S, P \otimes Q> + <S \otimes SZ, P \otimes Q>$ for all $P, Q \in A_{\mathbf{k}}(\mathbf{X})$. This happens iff $\Delta(SZ) = SZ \otimes S + S \otimes SZ$, i.e. —since $\Delta(SZ) = \Delta(S)\Delta(Z) = (S \otimes S)\Delta(Z)$, and $SZ \otimes S + S \otimes SZ = (S \otimes S)(Z \otimes 1 + 1 \otimes Z)$— iff $\Delta(Z) = Z \otimes 1 + 1 \otimes Z$. So *the formal tangent vectors to $\hat{G}_{\mathbf{k}}(\mathbf{X})$ at S are exactly the functionals $A_{\mathbf{k}}(\mathbf{X}) \ni P \to \,<SZ, P> \in \mathbf{k}$, for $Z \in \hat{L}_{\mathbf{k}}(\mathbf{X})$*. In particular, the members Z of $\hat{L}_{\mathbf{k}}(\mathbf{X})$ must be thought of as tangent vectors to $\hat{G}_{\mathbf{k}}(\mathbf{X})$ at 1. If $Z \in \hat{L}_{\mathbf{k}}(\mathbf{X})$ then the map $S \to SZ$ is a *formal left-invariant vector field* on $\hat{G}_{\mathbf{k}}(\mathbf{X})$.

In agreement with the notation $pV\varphi$ introduced in §2, for a point p, a vector field V and a function φ, the expression $<SZ, P>$ can be thought of as the result of applying SZ —regarded as a tangent vector at S— to the function $P \in A_{\mathbf{k}}(\mathbf{X})$. Naturally, then, we define L_Z —the operator of "formal Lie differentiation in the direction of Z"— to be the map that assigns to every $P \in A_{\mathbf{k}}(\mathbf{X})$ the function $Q = L_Z P \in A_{\mathbf{k}}(\mathbf{X})$ such that $<SZ, P> = <S, Q>$ for all $S \in \hat{G}_{\mathbf{k}}(\mathbf{X})$. Then $<SZ, P> = <S, Q>$ for all $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, so $L_Z : A_{\mathbf{k}}(\mathbf{X}) \to A_{\mathbf{k}}(\mathbf{X})$ is just the transpose of the map $\hat{A}_{\mathbf{k}}(\mathbf{X}) \ni S \to SZ \in \hat{A}_{\mathbf{k}}(\mathbf{X})$.

For $S \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, $P, Q \in A_{\mathbf{k}}(\mathbf{X})$, we have

$$
\begin{aligned}
&<S, L_Z(P \sqcup\!\sqcup Q)> = <SZ, P \sqcup\!\sqcup Q> = <SZ, P><S, Q> + <S, P><SZ, Q> \\
&= <S, L_Z P><S, Q> + <S, P><S, L_Z Q> = <S \otimes S, L_Z P \otimes Q + P \otimes L_Z Q> \\
&= <\Delta(S), L_Z P \otimes Q + P \otimes L_Z Q> = <S, (L_Z P) \sqcup\!\sqcup Q + P \sqcup\!\sqcup (L_Z Q)> .
\end{aligned}
$$

So $L_Z(P \sqcup\!\sqcup Q) = (L_Z P) \sqcup\!\sqcup Q + P \sqcup\!\sqcup (L_Z Q)$ for $P, Q \in A_{\mathbf{k}}(\mathbf{X})$, showing that L_Z is a derivation on the algebra $A_{\mathbf{k}}(\mathbf{X})$ equipped with the shuffle product.

If $Z = x \in \mathbf{X}$, then L_x is easily seen to be the map characterized by $L_x(wy) = 0$ if $y \neq x$, $w \in W(\mathbf{X})$, $L_x(wx) = w$ if $w \in W(\mathbf{X})$, and $L_x(1) = 0$.

This characterization implies, in particular, that *for every family $\{P^x\}_{x \in \mathbf{X}}$ of members of $A_{\mathbf{k}}(\mathbf{X})$ indexed by $\mathbf{X}$ there exists a unique $Q \in A_{\mathbf{k}}(\mathbf{X})$ such that $L_x Q = P^x$ for all $x \in \mathbf{X}$ and $<1, Q> = 0$.* (Indeed, letting $P^x = \sum_{w \in W(\mathbf{X})} p^x_w w$, $Q = \sum_{w \in W(\mathbf{X})} q_w w$, we have $L_x Q = \sum_{w \in W(\mathbf{X})} q_{wx} w$, so Q satisfies our conditions iff $q_1 = 0$ and $q_{wx} = p^x_w$ for all $w \in W(\mathbf{X})$, $x \in \mathbf{X}$, from which the existence and uniqueness of Q follows trivially.)

4 Chronological algebras, iterated integrals, and Chen series

Chronological algebras play a fundamental role in control (cf. [2, 3, 18, 9]), simplify formulas in combinatorics (cf. [8]), and are closely related to the *Leibniz algebras* that have recently been investigated in the algebraic literature (cf. [11, 12]).

A *(right) chronological algebra* is a linear space A over a field $\mathbf{k}$ endowed with a bilinear operation $* : A \times A \to A$ that satisfies the *right chronological identity*:

$$x * (y * z) = (x * y) * z + (y * x) * z \qquad \text{for all } x, y, z \in A\,. \tag{30}$$

(One can also define *left CA's*, in which the identity $(x * y) * z = x * (y * z) + x * (z * y)$ holds. In this note only right CA's will be used, so we will omit the word "right." The notion of CA introduced here is similar to that of Agrachev and Gamkrelidze [3], which also has been studied under the name of *Leibniz algebra* by Loday [11, 12] and others. The key identity for that other notion is the formula $(x\sharp y)\sharp z - (y\sharp x)\sharp z = x\sharp(y\sharp z) - y\sharp(x\sharp z)$, which says that $L_{x\sharp y - y\sharp x} = [L_x, L_y]$, where L_u is the map $z \to u\sharp z$. A typical example of a CA in this other sense is obtained by tensoring a Lie algebra with a CA in our sense.)

If $(A, *)$ is a CA, and we define $P \sqcup\!\sqcup Q \stackrel{\text{def}}{=} P * Q + Q * P$ for $P, Q \in A$, then it is easy to see that $\sqcup\!\sqcup$ is commutative and associative.

As a first example of a CP, we define $P * Q$, for $P, Q \in A_{\mathbf{k}}(\mathbf{X})$, by letting $P * Q$ be the unique $R \in A_{\mathbf{k}}(\mathbf{X})$ such that $L_x(R) = P \sqcup\!\sqcup L_x Q$ for all $x \in \mathbf{X}$ and $<1, R> = 0$. (The existence and uniqueness of R follows from the last remark of the previous section.) The CP then satisfies $L_x(P * Q) = P \sqcup\!\sqcup L_x Q$ for all $x \in \mathbf{X}$. It then follows easily that $P * Q + Q * P = P \sqcup\!\sqcup Q$ whenever $<1, P> <1, Q> = 0$, since $L_x(P*Q+Q*P) = L_x(P*Q)+L_x(Q*P) = P \sqcup\!\sqcup L_x Q + Q \sqcup\!\sqcup L_x P = L_x(P \sqcup\!\sqcup Q)$, $<1, P * Q + Q * P> = 0$, and $<1, P \sqcup\!\sqcup Q> = <1, P> <1, Q>$. This implies that, on the algebra $A^+_{\mathbf{k}}(\mathbf{X})$, the map $*$ is a CP, since

$$\begin{aligned} &L_x(P * (Q * R)) = P \sqcup\!\sqcup L_x(Q * R) = P \sqcup\!\sqcup (Q \sqcup\!\sqcup L_x R) = (P \sqcup\!\sqcup Q) \sqcup\!\sqcup L_x R \\ &= L_x((P \sqcup\!\sqcup Q) * R) = L_x((P * Q + Q * P) * R) = L_x((P * Q) * R + (Q * P) * R)\,, \end{aligned}$$

so $P * (Q * R) = (P * Q) * R + (Q * P) * R$.

We refer to $(A^+_{\mathbf{k}}(\mathbf{X}), *)$ as the *free CA over* $\mathbf{k}$ *in the indeterminates* $\mathbf{X}$, because $(A^+_{\mathbf{k}}(\mathbf{X}), *)$ is a "free CA" in the usual sense: if B is any CA over $\mathbf{k}$ and $\mu : \mathbf{X} \to B$ is a map, then μ can be extended to a unique CA-homomorphism $\tilde{\mu}$ from $A^+_{\mathbf{k}}(\mathbf{X})$ to B. (We construct $\tilde{\mu}$ recursively by letting $\tilde{\mu}(x) = \mu(x)$ for $x \in \mathbf{X}$, and $\tilde{\mu}(wx) = \tilde{\mu}(w) * \mu(x)$ for $x \in \mathbf{X}$, $w \in W(X)$. It is then not hard to verify that $\tilde{\mu}$ has the desired property, using the fact that $wx = w * x$. Uniqueness is trivial.)

There are numerous other examples of CA's. For example, if $\mathbf{k} = \mathbb{R}$, and $\mathcal{AC}$ is the space of locally absolutely continuous functions $f : [0, +\infty[\mapsto \mathbb{R}$ such that $f(0) = 0$, then we can define $f * g$, for f, g in $\mathcal{AC}$, by $(f * g)(t) = \int_0^t f(s) g'(s)\, ds$, where g' is the derivative of g. Then $f * g + g * f = fg$, and the CP identity says that $\int f (\int g h')' = \int (fg) h'$, which is trivially true. So $(\mathcal{AC}, *)$ is a CA over $\mathbb{R}$, and the commutative product associated to $*$ is just ordinary multiplication.

There are several natural chronological subalgebras of $(\mathcal{AC}, *)$. For example, the set of $f \in \mathcal{AC}$ such that $f \in C^\infty$, or the set $f \in \mathcal{AC}$ that are polynomial functions. In the latter example, a basis of the algebra is given by the monomials x^m, $m = 1, 2, \ldots$, and $x^n * x^m = \frac{1}{n} x^{n+m}$.

Finally, if $\mathbf{k}$ is any field of characteristic zero, then in the algebra $\mathbf{k}^+[X]$ of polynomials of a single variable over $\mathbf{k}$ with zero constant terms, we can define $x^n * x^m = \frac{m}{m+n} x^{n+m}$. Then $(\mathbf{k}^+[X], *)$ is a chronological algebra. Naturally, if

$P, Q \in \mathbf{k}^+[X]$, then $P * Q = \smallint(PQ')$ where, for $S \in \mathbf{k}^+[X]$, $\smallint S$ is the unique polynomial $T \in \mathbf{k}^+[X]$ such that $T' = S$.

The $\mathbf{k}$-algebra $\mathbf{k}[X_1, \dots X_m]$ of polynomials in m variables with coefficients in $\mathbf{k}$ can be represented as an algebra of functions on $\mathbf{k}^m$, namely, the algebra $\mathbf{k}[x_1, \dots x_n]$ of polynomial functions in m $\mathbf{k}$-valued *variables*, i.e. the subalgebra of $\mathrm{Map}(\mathbf{k}^m, \mathbf{k})$ generated by the projection maps $\mathbf{k}^m \ni (p_1, \dots, p_m) \to x_i(p) \stackrel{\mathrm{def}}{=} p_i \in \mathbf{k}$, where $\mathrm{Map}(\mathbf{k}^m, \mathbf{k})$ is the set of all maps from $\mathbf{k}^m$ to $\mathbf{k}$, regarded as an algebra with pointwise multiplication.

Similarly, we will represent the free CA $(A_{\mathbf{k}}(\mathbf{X}), *)$ as an algebra of "dynamic functionals" $\mathcal{U}_{\mathbf{k}}^{\mathbf{X}} \to \mathcal{U}_{\mathbf{k}}$, where $\mathcal{U}_{\mathbf{k}}$ is a CA of "time-varying scalars," i.e. of $\mathbf{k}$-valued functions $t \to f(t)$ of "time." One has substantial freedom in choosing the basic CA $\mathcal{U}_{\mathbf{k}}$. Here, for the sake of clarity, we specialize to a familiar setting. We work with $\mathbf{k} = \mathbb{R}$, and choose $\mathcal{U}_{\mathbf{k}} = \mathcal{AC}$. Then $\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$ is the set of all families $\{U_x : x \in \mathbf{X}\}$ of locally absolutely continuous real-valued functions on $[0, \infty[$ that vanish at 0. We use π_y to denote, for each $y \in \mathbf{X}$, the canonical projection $\mathcal{U}_{\mathbf{k}}^{\mathbf{X}} \ni \{U_x : x \in \mathbf{X}\} \to U_y \in \mathcal{U}_{\mathbf{k}}$.

We use $\mathrm{Map}(\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}, \mathcal{U}_{\mathbf{k}})$ to denote the set of all maps from $\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$ to $\mathcal{U}_{\mathbf{k}}$. Then $\mathrm{Map}(\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}, \mathcal{U}_{\mathbf{k}})$ is a CA under pointwise chronological multiplication: for Φ, Ψ in $\mathrm{Map}(\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}, \mathcal{U}_{\mathbf{k}})$ and $U \in \mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$, we define $(\Phi * \Psi)(U) = \Phi(U) * \Psi(U)$.

The CA $\mathcal{IIF}_{\mathbf{k}}(\mathbf{X})$ of *iterated integral functionals on* $\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$ is the chronological subalgebra of $\mathrm{Map}(\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}, \mathcal{U}_{\mathbf{k}})$ generated by the set $\{\pi_x : x \in \mathbf{X}\}$. So, if we use $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}$ to denote the unique CA-homomorphism from $A_{\mathbf{k}}(\mathbf{X}) \to \mathrm{Map}(\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}, \mathcal{U}_{\mathbf{k}})$ that sends x to π_x for each $x \in \mathbf{X}$, then $\mathcal{IIF}_{\mathbf{k}}(\mathbf{X}) = \mathcal{I}_{\mathbf{k}}^{\mathbf{X}}\Big(A_{\mathbf{k}}(\mathbf{X})\Big)$. Clearly, $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}} : A_{\mathbf{k}}(\mathbf{X}) \to \mathcal{IIF}_{\mathbf{k}}(\mathbf{X})$ is a surjective CA-homomorphism. We will see later that $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}$ is also injective.

We now define the *Chen-Fliess series* S_U of an input $U \in \mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$. For this purpose, we first extend the CP notation and define $(F * G)(t) = \int_0^t F(s)G'(s)ds$ for functions F, G on $[0, \infty[$ with values in any, not necessarily commutative, $\mathbb{R}$-algebra. The *universal control system with inputs in* $\mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$ is the system

$$\Sigma(\mathbf{X}) \; : \qquad (d/dt)S(t) = S(t)\Bigg(\sum_{x \in \mathbf{X}} x\dot{U}_x(t)\Bigg), \quad S(0) = 1, \tag{31}$$

evolving in $\hat{A}_{\mathbf{k}}(\mathbf{X})$. For any family $U = \{U_x\}_{x \in \mathbf{X}}$ of locally integrable functions on $[0, \infty[$, (31) has a unique solution $[0, \infty[\ni t \to S_U(t) \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, known as *the Chen-Fliess series for the input* U. Moreover, the Friedrichs criterion easily implies that *(31) actually evolves in* $\hat{G}_{\mathbf{k}}(\mathbf{X})$, i.e. that $S_U(t) \in \hat{G}_{\mathbf{k}}(\mathbf{X})$ for all U, t.

If we let $Z_U(t) = \sum_{x \in \mathbf{X}} xU_x(t)$, we see that (31) says that $\dot{S}_U = S_U \dot{Z}_U$ and $S_U(0) = 1$, i.e. that $S_U(t) = 1 + \int_0^t S_U(s)\dot{Z}_U(s)ds$ or, equivalently, $S_U = 1 + S_U * Z_U$. Using $1 * V = V$, this implies that $S_U = 1 + Z_U + (S * Z_U) * Z_U$. It is then easy to show, by successive iterations, that

$$S_U = 1 + Z_U + Z_U * Z_U + ((S_U * Z_U) * Z_U) * Z_U \,,$$
$$S_U = 1 + Z_U + Z_U * Z_U + (Z_U * Z_U) * Z_U + (((S_U * Z_U) * Z_U) * Z_U) * Z_U \,,$$

and so on, so that, finally, $S_U = \sum_{k=0}^{\infty} Z_U^{(*k)}$, where $Z_U^{(*k)}$ is defined recursively by $Z_U^{(*0)} = 1$, $Z_U^{(*(k+1))} = Z_U^{(*k)} * Z_U$.

Clearly, $Z_U^{(*k)} = \sum_{w\in W(\mathbf{X}):|w|=k} wU_w$, where U_w is defined recursively by $U_\emptyset = 1$, $U_{wx} = U_w * U_x$ for $w \in W(\mathbf{X})$, $x \in \mathbf{X}$. So $\boxed{S_U(t) = \sum_{w\in W(\mathbf{X})} w\,U_w(t)}$. If $w \in W(\mathbf{X})$, then $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(w)$ is the functional that assigns to $U \in \mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$ the function U_w. Therefore $\boxed{S_U(t) = \sum_{w\in W(\mathbf{X})} w\,\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(w)(U)(t)}$. We now define $\mathbf{CH}^{\mathbf{X}}$ to be the series

$$\mathbf{CH}^{\mathbf{X}} \stackrel{\text{def}}{=} \sum_{w\in W(\mathbf{X})} w \otimes \mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(w)\,, \tag{32}$$

so $\mathbf{CH}^{\mathbf{X}} \in \hat{A}_{\mathcal{IIF}_{\mathbf{k}}(\mathbf{X})}(\mathbf{X})$. The natural evaluation pairing from the Cartesian product $\mathcal{IIF}_{\mathbf{k}}(\mathbf{X}) \times \mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$ to $\mathcal{U}_{\mathbf{k}}$ that sends (Φ, U) to $\Phi(U)$ induces an "evaluation map" $\mathbf{E}^{\mathbf{X}} : \hat{A}_{\mathcal{IIF}_{\mathbf{k}}(\mathbf{X})}(\mathbf{X}) \times \mathcal{U}_{\mathbf{k}}^{\mathbf{X}} \to \hat{A}_{\mathcal{U}_{\mathbf{k}}}(\mathbf{X})$. It is then clear that, if $U \in \mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$, then $\mathbf{E}^{\mathbf{X}}(\mathbf{CH}^{\mathbf{X}}, U) = S_U$. If we let

$$\overline{\mathbf{CH}}^{\mathbf{X}} \stackrel{\text{def}}{=} \sum_{w\in W(\mathbf{X})} w \otimes w\,, \tag{33}$$

then $\overline{\mathbf{CH}}^{\mathbf{X}} \in \hat{A}_{\mathbf{k}}(\mathbf{X})\tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$, and $\mathbf{CH}^{\mathbf{X}} = \left(\mathrm{id} \otimes \mathcal{I}_{\mathbf{k}}^{\mathbf{X}}\right)(\overline{\mathbf{CH}}^{\mathbf{X}})$.

If $P \in \hat{A}_{\mathbf{k}}(\mathbf{X})$, then $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(P)$ is an iterated integral functional, which can be evaluated at any input $U \in \mathcal{U}_{\mathbf{k}}^{\mathbf{X}}$, yielding a function $\varphi_{P,U} : [0, \infty[\to \mathbb{R}$ given by $\varphi_{P,U}(t) = \mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(P)(U)(t)$. If P is a word $v \in W(\mathbf{X})$, then

$$< S_U(t), P > = < \textstyle\sum_w w\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(w)(U)(t), P > = \mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(P)(U)(t) = \varphi_{P,U}(t)\,.$$

It follows by linearity that $\varphi_{P,U}(t) = < S_U(t), P >$ for all P, U. If $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}(P) = 0$ as a member of $\mathcal{IIF}_{\mathbf{k}}(\mathbf{X})$, then $\varphi_{P,U}(t) = 0$ for all U, t, so $< S_U(t), P > = 0$ for all U, t. It follows in particular that $< S, P > = 0$ for every member $S \in \hat{G}_{\mathbf{k}}(\mathbf{X})$ which is of the form $Q = e^{t_1 x_1} e^{t_2 x_2} \dots e^{t_k x_k}$ for some $x_1, \dots, x_k$ in $\mathbf{X}$, $t_1, \dots, t_k$ in $\mathbb{R}$. Successive differentiations of these identities with respect to $t_1, \dots, t_k$ yield $< x_1 x_2 \dots x_k, P > = 0$. So $< w, P > = 0$ for every $w \in W(\mathbf{X})$, and then $P = 0$. This proves that *the map $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}$ is an isomorphism from $\hat{A}_{\mathbf{k}}(\mathbf{X})$ onto $\mathcal{IIF}_{\mathbf{k}}(\mathbf{X})$.*

Now that we know that the map $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}$ is an isomorphism, we can conclude that $\mathrm{id} \otimes \mathcal{I}_{\mathbf{k}}^{\mathbf{X}}$ is an isomorphism as well, so we can identify the spaces $\hat{A}_{\mathbf{k}}(\mathbf{X})\tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$ and $\hat{A}_{\mathbf{k}}(\mathbf{X})\tilde{\otimes}\mathcal{IIF}_{\mathbf{k}}(\mathbf{X})$. In particular, any expansion we obtain for $\mathbf{CH}^{\mathbf{X}}$ will yield a similar expansion for $\overline{\mathbf{CH}}^{\mathbf{X}}$.

We remark that the space $\hat{A}_{\mathbf{k}}(\mathbf{X})\tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$ is naturally identified with the space $\mathrm{Hom}_{\mathbf{k}}(A_{\mathbf{k}}(\mathbf{X}), A_{\mathbf{k}}(\mathbf{X}))$ of linear endomorphisms of $A_{\mathbf{k}}(\mathbf{X})$, by assigning to each map $\Lambda \in \mathrm{Hom}_{\mathbf{k}}(A_{\mathbf{k}}(\mathbf{X}), A_{\mathbf{k}}(\mathbf{X}))$ the series $\sum_{w\in W(\mathbf{X})} w \otimes \Lambda(w)$. Under this identification, $\overline{\mathbf{CH}}^{\mathbf{X}}$ corresponds to the identity map of $A_{\mathbf{k}}(\mathbf{X})$. So the Chen series is none other than the identity map of $A_{\mathbf{k}}(\mathbf{X})$, modulo several natural identifications, showing that $\mathbf{CH}^{\mathbf{X}}$ is a natural object.

5 Exponential product expansions and dual PBW-bases

In §2 we showed how to compute the first few factors of an expansion as a product of exponentials of flow maps $e^{\int_0^t(\sum_{i=1}^m u_i(s)f_i)ds}$ determined by m smooth vector fields f and g, by means of successive applications of the method of variations of constants. In the situation discussed in §2 we had $m = 2$, and $u_1(t) \equiv 1$, and we just computed the first five factors. It turns out that the formal calculation can be pursued for any number of factors, for a general m, and for general inputs $u = \dot{U} \in \mathcal{AC}^m$.

Remarkably, the algebra works out in such a way that one obtains a formula expressing the Chen series $\mathbf{CH}^{\mathbf{X}}$ as an infinite product of exponentials

$$\mathbf{CH}^{\mathbf{X}} = \overleftarrow{\prod}_{B\in\mathcal{B}} e^{B\otimes\Phi_B}\,, \tag{34}$$

where $\mathcal{B}$ is any "generalized Hall basis" (abbr. GHB) of $\hat{A}_{\mathbf{k}}(\mathbf{X})$, the coefficients $\Phi_B \in \mathcal{IIF}_{\mathbf{k}}(\mathbf{X})$ are iterated integral functionals given by simple formulas, as explained below, and the symbol $\overleftarrow{\prod}$ indicates that the factors are ordered from right to left, following the ordering of $\mathcal{B}$.

GHB's arise when one seeks to spell out explicit combinatorial rules to write bases of $L_{\mathbf{k}}(\mathbf{X})$. Several such schemes have been proposed, but all were shown by Viennot [19] (see also [14] for a modern discussion) to arise from the same underlying principle, resulting in what is now known as GHB's, which is a special type of basis $\mathcal{B}$ of $L_{\mathbf{k}}(\mathbf{X})$, endowed with a total ordering $\preceq$. (We refer the reader to [19, 14] for the precise definition of a GHB.)

Applying the method of variation of constants, Formula (34) was derived in Sussmann [18] in 1986, together with an explicit recursive formula for the functionals Φ_B. If $B = x \in \mathbf{X}$, then $\Phi_x = \pi_x$. If $B \in \mathcal{B}$ but $B \notin \mathbf{X}$, then write $B = \mathrm{ad}_{B_1}^{m_1}\mathrm{ad}_{B_2}^{m_2}\ldots\mathrm{ad}_{B_k}^{m_k}(B_{k+1})$, with the B_i in $\mathcal{B}$, $B_1 \succ B_2 \succ \ldots \succ B_k$, $B_k \prec B_{k+1}$, and $m_1,\ldots,m_k$ positive integers —it is a fact that every $B \in \mathcal{B}$ can be so expressed— and then Φ_B is given by

$$\Phi_B = \Big(\prod_{i=1}^{k}\frac{1}{m_i!}\Phi_{B_i}^{m_i}\Big) * \Phi_{B_{k+1}}\,. \tag{35}$$

(The derivation given in [18] was for classical P. Hall bases, but the proof applies without change to any generalized Hall basis.)

Expanding the exponentials of (34), one gets the formula

$$\mathbf{CH}^{\mathbf{X}} = \sum_{k=0}^{\infty}\ \sum_{B_1\succ B_2\succ\ldots\succ B_k}\ \sum_{\mu_1,\mu_2,\ldots,\mu_k} B_1^{\mu_1}B_2^{\mu_2}\ldots B_k^{\mu_k}\otimes\frac{\Phi_{B_1}^{\mu_1}\Phi_{B_2}^{\mu_2}\ldots\Phi_{B_k}^{\mu_k}}{\mu_1!\mu_2!\ldots\mu_k!}\,. \tag{36}$$

Via the inverse of the isomorphism $\mathrm{id}\otimes\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}$, and recalling that $\mathcal{I}_{\mathbf{k}}^{\mathbf{X}}$ is an isomorphism from $A_{\mathbf{k}}(\mathbf{X})$ with the shuffle product to $\mathcal{IIF}_{\mathbf{k}}(\mathbf{X})$ with ordinary multiplication, we can transform (36) into an expansion

$$\overline{\mathbf{CH}}^{\mathbf{X}} = \sum_{k=0}^{\infty}\ \sum_{B_1\succ B_2\succ\ldots\succ B_k}\ \sum_{\mu_1,\mu_2,\ldots,\mu_k} B_1^{\mu_1}B_2^{\mu_2}\ldots B_k^{\mu_k}\otimes\frac{\Psi_{B_1}^{⧢,\mu_1} ⧢ \Psi_{B_2}^{⧢,\mu_2} ⧢ \ldots ⧢ \Psi_{B_k}^{⧢,\mu_k}}{\mu_1!\mu_2!\ldots\mu_k!}\,, \tag{37}$$

where $\Psi_x = x$ for $x \in \mathbf{X}$ and, if $B \in \mathcal{B} \backslash \mathbf{X}$, then

$$\Psi_B = \left(\frac{\Psi_{B_1}^{ш, m_1}}{m_1!} ш \frac{\Psi_{B_2}^{ш, m_2}}{m_2!} ш \ldots ш \frac{\Psi_{B_k}^{ш, m_k}}{m_k!} \right) * \Psi_{B_{k+1}}, \tag{38}$$

if $B = \mathrm{ad}_{B_1}^{m_1} \mathrm{ad}_{B_2}^{m_2} \ldots \mathrm{ad}_{B_k}^{m_k} (B_{k+1})$, with the B_i in $\mathcal{B}$, $B_1 \succ B_2 \succ \ldots \succ B_k$, $B_k \prec B_{k+1}$, and $m_1, \ldots, m_k$ positive integers. (The ш symbols accompanying the exponents are there as a reminder that all the powers are taken in the sense of the shuffle product.)

Formulas (37) and (38) give the expansion of $\overline{\mathbf{CH}}^{\mathbf{X}}$ —i.e. the identity element of $\mathrm{Hom}_{\mathbf{k}}(A_{\mathbf{k}}(\mathbf{X}), A_{\mathbf{k}}(\mathbf{X}))$, regarded as a member of $\hat{A}_{\mathbf{k}}(\mathbf{X}) \tilde{\otimes} A_{\mathbf{k}}(\mathbf{X})$— in terms of the Poincaré-Birkhoff-Witt basis of $A_{\mathbf{k}}(\mathbf{X})$ associated to $\mathcal{B}$. (Recall that $A_{\mathbf{k}}(\mathbf{X})$ is a UEA of $L_{\mathbf{k}}(\mathbf{X})$.) So the coefficients $\frac{\Psi_{B_1}^{ш, \mu_1} ш \Psi_{B_2}^{ш, \mu_2} ш \ldots ш \Psi_{B_k}^{ш, \mu_k}}{\mu_1! \mu_2! \ldots \mu_k!}$ appearing in (37) are the members of the *dual basis of the PBW basis arising from $\mathcal{B}$.*

The derivation outlined here was given for $\mathbf{k} = \mathbb{R}$, but one can easily see in various ways that the formula is valid for any field $\mathbf{k}$ of characteristic zero. (For example, (37) and (38) are identities between formal power series with rational coefficients, so they are valid over any field of characteristic zero.)

A similar formula was derived by Melançon and Reutenauer in 1989 in [13] by combinatorial means using rewriting systems. Using different notations, one can also find this formula in Schützenberger's 1958 notes [17].

References

[1] Agrachev, A. and A. Sarychev, Abnormal sub-Riemannian geodesiscs: Morse Index and Rigidity, Ann. Inst. H. Poincaré **13** (1996), pp. 635-690.

[2] Agrachev, A. and R. Gamkrelidze, Exponential representation of flows and chronological calculus, Math. USSR Sbornik (Russian) **107**, N4 (1978), pp. 487-532.

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

[4] Bourbaki, N., Lie Groups and Lie algebras, Hermann, Paris, 1975.

[5] Chen, K. T., Integration of paths, geometric invariants, and a generalized Baker-Hausdorff formula, Annals of Math. **67** (1957), pp. 164-178.

[6] Crouch, P. and R. Grossman, Numerical Integration of Ordinary Differential Equations on Manifolds, J. Nonlinear Science **3** (1993), pp. 1-33.

[7] Fliess, M., Fonctionnelles causales non linéaires et indeterminées non commutatives, Bull. Soc. Math. France **109** (1981), pp.3-40.

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

[9] Kawski, M., Nonlinear control and combinatorics of words, in: Nonlinear Feedback and Optimal control, B. Jakubczyk and W. Respondek, eds., Dekker, 1997 (to appear).

[10] Knobloch, H. W., High Order Necessary Conditions in Optimal Control, Springer-Verlag, 1975.

[11] Loday, J.-L, Une version non commutative des algèbres de Lie: les algèbres de Leibniz, L'Enseignement Mathématique **39** (1993), pp. 268-293.

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

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

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

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

[16] Reutenauer, C., Free Lie Algebras, London Math. Soc. monographs, new series **7**, Oxford, 1993.

[17] Schützenberger, M., Sèminaire Dubreil, Facultè de Sciences de Paris, 1958.

[18] Sussmann, H. 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), pp. 323–335.

[19] Viennot, G. Algèbres de Lie Libres et Monoïdes Libres, Lect. Notes, Math., 692, Springer, Berlin, 1978.

Matthias Kawski
Department of Mathematics
Arizona State University
Tempe, Arizona 85287, USA
Phone: (602) 965 3376
Fax: (602) 965 0461
kawski@asu.edu
http://math.la.asu.edu/~kawski

Héctor J. Sussmann
Department of Mathematics
Rutgers University
Piscataway, NJ 08855, USA
Phone: (732)445-5407
Fax: (732)445-5530
sussmann@hamilton.rutgers.edu
http://math.rutgers.edu/~sussmann