Hopf algebras in dynamical systems theory

José F. Cariñena*, Kurusch Ebrahimi-Fard†,

Héctor Figueroa‡ and José M. Gracia-Bondía§‖

*Departamento de Física Teórica, Universidad de Zaragoza, Zaragoza 50006, Spain

†Max-Planck-Institut für Mathematik, Vivatsgasse 7 D-53111 Bonn, Germany

‡Departamento de Matemáticas, Universidad de Costa Rica, San Pedro 2060, Costa Rica

§Departamento de Física Teórica I, Universidad Complutense, Madrid 28040, Spain

‖Departamento de Física, Universidad de Costa Rica, San Pedro 2060, Costa Rica

December 28, 2006

Abstract

The theory of exact and of approximate solutions for non-autonomous linear differential equations forms a wide field with strong ties to physics and applied problems. This paper is meant as a stepping stone for an exploration of this long-established theme, through the tinted glasses of a (Hopf and Rota–Baxter) algebraic point of view. By reviewing, reformulating and strengthening known results, we give evidence for the claim that the use of Hopf algebra allows for a refined analysis of differential equations. We revisit the renowned Campbell–Baker–Hausdorff–Dynkin formula by the modern approach involving Lie idempotents. Approximate solutions to differential equations involve, on the one hand, series of iterated integrals solving the corresponding integral equations; on the other hand, exponential solutions. Equating those solutions yields identities among products of iterated Riemann integrals. Now, the Riemann integral satisfies the integration-by-parts rule with the Leibniz rule for derivations as its partner; and skewderivations generalize derivations. Thus we seek an algebraic theory of integration, with the Rota–Baxter relation replacing the classical rule. The methods to deal with noncommutativity are especially highlighted. We find new identities, allowing for an extensive embedding of Dyson–Chen series of time- or path-ordered products (of generalized integration operators); of the corresponding Magnus expansion; and of their relations, into the unified algebraic setting of Rota–Baxter maps and their inverse skewderivations. This picture clarifies the approximate solutions to generalized integral equations corresponding to non-autonomous linear (skew)differential equations.

Mathematics Subject Classification 2000: 16W25, 16W30, 37B55, 37C10

2006 PACS: 02.20.Qs, 02.30.Hq, 02.40.Gh, 45.30.+s

Keywords: Differential equations, Lie–Scheffers systems, Rota–Baxter operators, Hopf algebra, Spitzer's identity, Magnus expansion

To Ennackal Chandy George Sudarshan,
with admiration, on his 75th birthday

Contents

1 Aim, plan of the article and preliminaries

This paper studies non-autonomous differential equations of the general type

$$\dot g(t)g^{-1}(t) = \xi(t), \qquad g(t_0) = 1_G \quad \text{or} \tag{1.1}$$
$$g^{-1}(t)\dot g(t) = \eta(t), \qquad g(t_0) = 1_G, \tag{1.2}$$

where the unknown $g : \mathbb{R}_t \to G$ is a curve on a (maybe infinite-dimensional) local Lie group G, with 1_G the neutral element; and $\xi(t), \eta(t)$ are given curves on the tangent Lie algebra $\mathfrak{g}$ of G. Before proceeding note that, if $g(t)$ solves (1.1), then $g^{-1}(t)$ solves (1.2) for $\eta(t) = -\xi(t)$.

Such equations are pervasive in mathematics, physics and engineering. To begin with, G can have a faithful finite-dimensional representation. For instance, consider (affine) linear differential equations on $\mathbb{R}^n$,

$$\dot x = A(t)x + b(t) \quad \text{with} \quad x(t_0) = x_0. \tag{1.3}$$

They are exactly solved by

$$x(t) = \mathcal{G}(t,t_0)x_0 + \int_{t_0}^{t} \mathcal{G}(t,t')b(t')\,dt';$$

where the *Green's function* $\mathcal{G}(t,t_0)$ is the matrix satisfying

$$\frac{d\mathcal{G}(t,t_0)}{dt}\,\mathcal{G}^{-1}(t,t_0) = A(t), \quad \mathcal{G}(t_0,t_0) = 1_n$$

of the kind (1.1). More generally, dynamical systems admitting a superposition principle can be reduced to the form (1.1), with G a finite dimensional Lie group. This assertion is part of the classical *Lie–Scheffers* theory [1], reviewed in Section 3 as part of and motivation for the whole enterprise. Even more generally, any non-autonomous dynamical system, given in local coordinates by

$$\frac{dx^i}{dt} = Y^i(t;x(t)), \qquad x = (x^1,\dots,x^n) \tag{1.4}$$

corresponds to a 'time-dependent vector field' Y with $Y(t)$ belonging to $\mathfrak{X}(M)$, the Lie algebra of all vector fields on a manifold M. Then the solution of (1.4) is given by the solution of an equation like (1.2); this remark will be formalized in Section 4. The crucial difference is dimensionality of the (pseudo-)group. In practice, almost always we must content ourselves with approximate solutions —Lie–Scheffers systems are not solvable by quadratures in general— and actually those are our main concern.

We reformulate (1.1) within the framework of Hopf algebra and Rota–Baxter operator theory —the latter has become popular recently in relation with the Connes–Kreimer paradigm for renormalization theory in perturbative quantum field theory. The convenience of such algebraic approach stems already from that, unless G is a matrix group, $\dot g$ and g^{-1} cannot be multiplied, and then equations (1.1) and (1.2) have no meaning, strictu sensu. Hopf algebras generalize both Lie groups and Lie algebras, so the problem does not present itself in a Hopf algebra formalism. Another advantage is that Hopf algebra and Rota–Baxter theory allow for efficient and meaningful comparisons among the different techniques for solving (1.1), proposed over the years. The main aim of this paper is to show these and other benefits of our algebraic viewpoint. They have been patent for a while to people working on control theory, but mostly ignored by the wider community of mathematicians and mathematical physicists.

We presume the readers acquainted with standard tools of differential analysis, like for instance in [2]: primarily the notion of tangent map and the exponential map $\exp : \mathfrak{g} \to G$. For the benefit of the readers, Lie group and Lie algebra actions are reviewed in Appendix A. The basics of Hopf algebra are a prerequisite. Unless otherwise specified, we consider Hopf algebras over the complex numbers. We briefly introduce our notations for them, which are like in [3]; the pedagogical paper [4] is recommended as well. Given an associative algebra with unit $H \ni 1_H =: u(1_\mathbb{C})$, then $H \otimes H$ is associative with bilinear multiplication given by $(a \otimes b)(c \otimes d) = ac \otimes bd$ on decomposable tensors, and unit $1_H \otimes 1_H$. Write just 1 for the unit element of H henceforth. One says H is a bialgebra if algebra morphisms $\eta : H \to \mathbb{C}$ (augmentation) and $\Delta : H \to H \otimes H$ (coproduct) are defined, such that the maps $(\eta \otimes \mathrm{id})\Delta$ and $(\mathrm{id} \otimes \eta)\Delta$ from H to H coincide with the identity map id and $(\Delta \otimes \mathrm{id})\Delta$ and $(\mathrm{id} \otimes \Delta)\Delta$ from H to $H \otimes H \otimes H$ also coincide (we omit the sign $\circ$ for composition of linear maps). One says H is a Hopf algebra if it furthermore possesses an antiautomorphism S, the *antipode*, such that $m(S \otimes \mathrm{id})\Delta = m(S \otimes \mathrm{id})\Delta = u\eta$; where $m : H \otimes H \to H$ denotes the algebra map. Familiarity with enveloping algebras and the Poincaré–Birkhoff–Witt and Cartier–Milnor–Moore theorems in particular will be helpful. Both results can be summarized in the statement that a connected cocommutative Hopf algebra is the enveloping algebra of a Lie algebra, as an algebra, and cofree, as a coalgebra. At any rate, we discuss a strong version of the Poincaré–Birkhoff–Witt theorem in Section 4, and the Cartier–Milnor–Moore theorem at the end of Section 5. The necessary notions of Rota–Baxter operator theory will be introduced and explained in due course.

Whereas the Hopf algebraic description springs up naturally from the intrinsic geometrical approach, we have found it expedient to smooth this transition with the help of Lie–Rinehart algebras: these constitute the "noncommutative geometry" version of Lie algebroid technology.

The plan of the work is as follows. In this section we explain our main aims and fix some notations of frequent use. Next we recommend a look at Appendix A, indispensable for everything that follows; most readers will just need to scan it for the notations. In Section 2 we address for the first time Lie–Scheffers systems; they are intimately linked to equations (1.1) and (1.2).

After this, two paths are possible: either reading Appendices C and D for motivation, or not. Most of the stuff in them could be regarded as preceding Sections 3 and following; but it gets in the way of our algebraic business, and this is why it has been confined to the end. Section 3 plunges the reader at once into an application of Hopf algebra to differential geometry. This is due to Rinehart and Huebschmann, and deserves to be better known, as it clarifies several questions; one should compare the treatment of differential operators given here with that of [5, Chapter 3]. Readers less familiar with Hopf algebra might wish to read this in parallel with Section 4.

Sections 4 to 6 are largely expository. In Section 4 we leisurely *construct* the Hopf algebra structure governing our approach to equations (1.1) and (1.2) from the geometric notions. Section 5 recalls some structure results for Hopf algebras.

After that, our master plan is to transplant the usual paradigmatic strategies for dealing with (1.1) and (1.2) to the Hopf algebraic soil, which on the one hand will prove to be their native one, and on the other naturally leads to far-reaching generalizations. At the outset, in Section 6 we consider the Campbell–Baker–Hausdorff–Dynkin (CBHD) development, which we proceed to derive in Hopf algebraic terms. In turn, that development is the natural father of the Magnus expansion method [6]. We eventually derive the Magnus series with the help of Rota–Baxter theory. For the purpose, the Riemann integral is treated in this paper as a particular Rota–Baxter operator of weight zero.

We first show in Section 7 that skewderivations and Rota–Baxter operators (of the same weight) are natural inverses. The ordinary Spitzer formula is revisited in Section 8, together with a nonlinear CBHD recursion due to one of us. The latter is instrumental in obtaining the *noncommutative Spitzer formula*. Also, inspired by the work of Lam, we give a noncommutative generalization of the Bohnenblust–Spitzer formula. In the next two sections, the Magnus expansion is arrived at as a limiting case of that formula. All along, we try to distinguish carefully which statements are valid for general Rota–Baxter operators, which for Rota–Baxter operators of vanishing weight, and which just for the Riemann integral. The main alternative integration method, the Dyson–Chen 'expansional' [7–9], flows from the Magnus series, and vice versa, by our Hopf algebraic means in Section 11. In turn, it reveals itself useful to understand the quirks of the Magnus expansion, and to solve the weight-zero CBHD recursion. Section 12 explores by means of pre-Lie algebras with Rota–Baxter maps the solution of that nonlinear recursion in the general case. Section 13 is the conclusion, whereupon perspectives for research are discussed.

As said, Appendix A reviews the basics of Lie group and Lie algebra actions on manifolds. As also hinted at, Appendices B and C run a parallel, complementary strand to the main body of the paper. They contain more advanced material on the topic of dynamical systems with symmetry; their treatment here naturally calls for the Darboux derivative of Lie algebroid theory. The main point is to show how one is led to the arena of Lie algebra and *geometrical integration*, for the solution of differential equations we are concerned with. This provides a rationale for our choice of the Magnus series, and its generalizations, as the primary approximation method in the body of the paper —see the discussion at the beginning of Section 10.

Appendix D gives the Hopf algebraic vision of a theorem of Lie and Engel.

Several notational conventions are fixed next. Let M be a (second countable, smooth, without boundary) manifold of finite dimension n. The space $\mathcal{F}(M)$ of (real or) complex smooth functions on M is endowed with the standard commutative and associative algebra structure. Let $\tau_M : TM \to M$ denote the tangent bundle to M. Vector fields on M can be defined either as sections for τ_M, that is, maps $X : M \to TM$ such that $\tau_M \circ X = \mathrm{id}_M$, or as derivations of $\mathcal{F}(M)$. When we wish to distinguish between those roles, we denote by $\mathcal{L}_X$ the differential operator corresponding to the vector field X. Because the commutator of two derivations is again a derivation, the space $\mathfrak{X}(M)$ of vector fields has a Lie algebra structure; we choose to define the bracket there as the opposite of the usual one: in local coordinates,

$$[X, Y]^i = Y^j\, \partial_j X^i - X^j\, \partial_j Y^i;$$

so $\mathfrak{X}(M) \equiv \mathfrak{diff}(M)$, the Lie algebra of the infinite-dimensional Lie group $\mathrm{Diff}(M)$ [10]. Also $\mathfrak{X}(M)$ becomes a faithful $\mathcal{F}(M)$-module when one defines $hX(x) = h(x)X(x)$ for $h \in \mathcal{F}(M)$. Together, $\mathcal{F}(M)$ and $\mathfrak{X}(M)$ constitute a Lie–Rinehart algebra in the sense of [11].

Given a smooth map $f : N \to M$, the pull-back $f^* : \mathcal{F}(M) \to \mathcal{F}(N)$ is defined as $f^*h = h \circ f$. A vector field $X \in \mathfrak{X}(N)$ is said to be f-related with the vector field $Y \in \mathfrak{X}(M)$ if $Tf \circ X = Y \circ f$; we then say that the vector field X is f-projectable onto the vector field Y, and write $X \sim_f Y$. We have $X \sim_f Y$ iff the maps $\mathcal{L}_X \circ f^*$ and $f^* \circ \mathcal{L}_Y$ from $\mathcal{F}(M)$ to $\mathcal{F}(N)$ coincide. If $X_1, X_2 \in \mathfrak{X}(N)$ are f-related with Y_1, Y_2 respectively, then $X_1 + X_2$ and $[X_1, X_2]$ are also f-related, respectively with $Y_1 + Y_2$ and $[Y_1, Y_2]$. A given $X \in \mathfrak{X}(N)$ will not be f-projectable in general. However, if f is a diffeomorphism, then every vector field $X \in \mathfrak{X}(N)$ is projectable onto a unique vector field on M, to wit, $Y = Tf \circ X \circ f^{-1}$, and we say X is the pull-back of Y. A vector field $X \in \mathfrak{X}(M)$ is *invariant* under a diffeomorphism f of M iff $X \sim_f X$. On $\mathbb{R}_t$ (or on an open interval $I \subset \mathbb{R}_t$) there is a canonical vector field d/dt. A curve $\gamma : \mathbb{R}_t \to M$ is said to be an integral curve for a vector field $X \in \mathfrak{X}(M)$ if d/dt and X are γ-related: $\dot\gamma := T\gamma \circ d/dt = X \circ \gamma$. Well-known theorems assert that the integral curves of a vector field define a local $\mathbb{R}_t$-action or flow [12].

By a *vector field along* f we understand a map $Y : N \to TM$ such that $\tau_M \circ Y = f$:

$$\begin{tikzcd} & TM \arrow[d, "\tau_M"] \\ N \arrow[ur, "Y"] \arrow[r, "f"'] & M \end{tikzcd}$$

It is clear that the concept is just a particular case of a more general one: section along the map f over a general bundle $\pi : E \to M$. Vector fields along f can also be regarded as f-derivations, in an obvious sense. The right hand side of the non-autonomous dynamical system (1.4) is just the vector field along the map $\pi_2 : \mathbb{R}_t \times M \to M$ expressed in local coordinates by

$$Y = Y^i(t; x(t))\, \partial_i \circ \pi_2.$$

Also, clearly any curve $\gamma : \mathbb{R}_t \to M$ defines a vector field $\dot\gamma$ along γ. We envisage here the concept of integral curves of vector fields Y along maps f. These are curves $\gamma : \mathbb{R}_t \to N$ such that the image under $Tf \circ T\gamma$ of the vector field d/dt coincides with the vector field along $f \circ \gamma$ given by $Y \circ \gamma$ —depending on f, there might be vector fields along it without integral curves. Under this definition $t \mapsto (t, \gamma(t))$ is always the integral curve of $(t, \gamma(t)) \mapsto \dot\gamma(t)$ along $\pi_2 : \mathbb{R}_t \times M \to M$.

2 The Lie–Scheffers theorem

Definition 1. The system (1.4) of differential equations admits a *superposition principle* —or possesses a set of fundamental solutions— if a superposition function $\Psi : \mathbb{R}^{n(m+1)} \to \mathbb{R}^n$ exists, written

$$x = \Psi(x_{(1)}, \dots, x_{(m)}; k_1, \dots, k_n), \tag{2.1}$$

such that the general solution of (1.4) can be expressed (at least for small t) as the functional

$$x(t) = \Psi(x_{(1)}(t), \dots, x_{(m)}(t); k_1, \dots, k_n), \tag{2.2}$$

where $\{ x_{(a)} : a = 1, \dots, m \}$ is a set of particular solutions and $k_1, \dots, k_n$ denote n arbitrary parameters. The latter must be *essential* in the sense that they can be solved from the solution functional:

$$k = \Xi\big(x_{(1)}(t), \dots, x_{(m)}(t); x(t)\big), \quad \text{with} \quad k := (k_1, \dots, k_n). \tag{2.3}$$

The Lie–Scheffers theorem [1] asserts:

Theorem 2.1. *For* (1.4) *to admit a superposition principle it is necessary and sufficient that the time-dependent vector field Y be of the form*

$$Y(t;x) = Z_1(t)X_1(x) + \cdots + Z_r(t)X_r(x), \tag{2.4}$$

where, as indicated in the notation, the r scalar functions Z_a depend only of t and the r vector fields X_a depend only on the variables x; and these fields close to a real Lie algebra $\mathfrak{g}$. That is, the X_a are linearly independent and there exist suitable structure constants f_{ab}^c such that

$$[X_a, X_b] = \sum_{c=1}^{r} f_{ab}^c\, X_c.$$

Moreover, the dimension r of $\mathfrak{g}$ is not greater than nm. Systems fulfilling the conditions of the theorem are called here Lie–Scheffers systems associated to $\mathfrak{g}$.

Modern reviews of this subject include [13–17]. We ponder a few pertinent examples of Lie–Scheffers systems next.

Linear systems (1.3) are Lie–Scheffers systems. If $n+1$ particular solutions $x_{(1)}, \dots, x_{(n+1)}$ of (1.3) are known, such that $x_{(2)}(t) - x_{(1)}(t), \dots, x_{(n+1)}(t) - x_{(1)}(t)$ are independent, and $H(t)$ is the regular matrix with these vectors as columns, then the vector of parameters (2.3) is given by

$$k = H^{-1}(t)\big(x(t) - x_{(1)}(t)\big).$$

This follows from the fact that the transformation

$$x'(t) = H^{-1}\big(x(t) - x_{(1)}(t)\big)$$

reduces the system to $dx'/dt = 0$.

A famous example for $n = 1$ is provided by the Riccati equation:

$$\dot{x} = a_0(t) + a_1(t)x + a_2(t)x^2. \tag{2.5}$$

One can understand by Hopf algebraic methods why, up to diffeomorphisms, Riccati's is the only nonlinear Lie–Scheffers differential equation on the real line; this was indicated in [18] and it is spelled in Appendix D. Also (2.5) is the simplest Lie–Scheffers system not solvable by quadratures; and other Lie–Scheffers equations on the line are reductions of it, in an appropriate sense —see Appendix B. The superposition principle for the Riccati equation is given by

$$k = \frac{(x - x_{(2)})(x_{(1)} - x_{(3)})}{(x - x_{(1)})(x_{(2)} - x_{(3)})}.$$

The one-dimensional example

$$\dot{x} = b(t)\chi(x) \tag{2.6}$$

is instructive. We assume χ does not change sign in the interval of interest. Let

$$\phi(x) := \int^x \frac{dx'}{\chi(x')}.$$

Then

$$x(t) = \phi^{-1}\Big(k' + \int b(t)\,dt\Big)$$

is the general solution. We have therefore a superposition rule of the form

$$x(t) = \phi^{-1}\big(\phi(x_{(1)}(t)) + k\big),$$

with $m = n = r = 1$: only one particular solution is required. Notice that the local diffeomorphism ϕ projects the vector field corresponding to the right hand side in (2.6) to the vector field $b(t)\partial_x$: in the language of Lie, the flow associated to this problem is locally similar to a translation.

Now we can tackle at last the question of giving intrinsic geometrical meaning to (1.1). It turns out to correspond to a Lie–Scheffers system on a Lie group. For an arbitrary curve $t \mapsto g(t)$ on the r-dimensional Lie group G, we have the vector field along the curve given by $\dot{g}(\cdot)$ as discussed at the end of Section 1. Then we *define* the left hand side of (1.1) as

$$\dot{g}(t)\, g^{-1}(t) := T_{g(t)}R_{g^{-1}(t)}\dot{g}(t). \tag{2.7}$$

By construction, for each value of the parameter t, this vector lies in $T_1G \equiv \mathfrak{g}$, the tangent Lie algebra of G. We obtain in this way a curve on $\mathfrak{g}$. Note that, if $g(t) = \exp(t\eta)$, then simply $\dot{g}(t)\, g^{-1}(t) = \eta$. Now, if γ^a is a basis for $\mathfrak{g}$, then

$$\dot{g}(t)\, g^{-1}(t) = \sum_{a=1}^{r} Z_a(t)\gamma^a =: \xi(t) \in \mathfrak{g}, \tag{2.8}$$

for some functions $Z_a(t)$. We realize that $g(t)$ is an integral curve of the right invariant vector field along $\mathbb{R}_t \times G \to G$:

$$\xi_G(t, g) := \sum_{a=1}^{r} Z_a(t)\gamma^a_G(g) = \sum_{a=1}^{r} Z_a(t)X^R_{\gamma^a}(g), \quad \text{with} \quad \gamma^a_G(1_G) = \gamma^a.$$

Here γ^a_G is the fundamental vector field or infinitesimal generator of the left group translations generated by γ^a, which is a *right* invariant vector field —see Appendix A for the notations. In the language of (1.4), the differential system is

$$\dot{g}(t) = \sum_{a=1}^{r} Z_a(t)\,\gamma^a_G(g(t)). \tag{2.9}$$

The theorem says that for every choice of the functions $Z_a(t)$ we have a (right invariant) Lie–Scheffers system on the Lie group G, and any such system is of this form. The reader should be aware, nevertheless, that for a system of the type (2.9) there might be more than one superposition rule [17]. The reason one needs only one particular solution is precisely the right invariance of the last equation: if $g(t)$ is the solution such that $g(t_0) = 1$, then consider $\bar{g}(t) := g(t)\, g_0$ for each $g_0 \in G$. We have

$$\dot{\bar{g}}(t) = T_{g(t)}R_{g_0}(\dot{g}(t)) = T_{g(t)}R_{g_0}\left[T_1R_{g(t)}\left(\sum_{a=1}^{r} Z_a(t)\gamma^a\right)\right]$$

$$= T_1 R_{g(t)g_0} \left(\sum_{a=1}^{r} Z_a(t)\gamma^a \right) = \sum_{a=1}^{r} Z_a(t)\gamma^a_G(\bar{g}(t)).$$

Equation $X^R(g) = T_1 R_g X^R(1_G)$ has been used. Therefore $\bar{g}(t)$ is the solution of the same equation (2.9) with $\bar{g}(t_0) = g_0$: the solution curves of the system are obtained from just one of them by right-translations. In other words, the superposition functional (2.2) can be symbolically expressed by $\Psi(g_{(1)}, k) = g_{(1)}k$, with $k \in G$; for which always $m = 1$.

Lie–Scheffers systems live on manifolds which are not groups in general; however, they are always associated with the action of a finite-dimensional Lie group on the manifold on which $Y(t; x)$ is defined; and this symmetry of the differential equation can be powerfully exploited through the action of the group of curves on the group manifold on a set of systems of the same type. This variant of Lie's reduction method is of wide applicability; it is explained in Appendix B.

Let us finally note than in control theory, say on $M \equiv \mathbb{R}^n$, business is often with equations of a form not unrelated to (2.9):

$$\dot{x}(t) = X_1(x(t)) + \sum_{a=2}^{r} Z_a(t)\, X_a(x(t));$$

the functions $Z_2, \dots, Z_r$ being the controls. In the most important cases the $X_1, \dots, X_r$ vector fields close to a finite-dimensional Lie algebra, or $X_2, \dots, X_r$ close to a finite-dimensional Lie algebra.

3 Differential operators on Lie–Rinehart algebras

Let R be a commutative, unital ring and $\mathcal{A}$ a commutative algebra over R be given. A derivation δ of $\mathcal{A}$ is a R-linear map from $\mathcal{A}$ to itself, such that $\delta(ab) = \delta a\, b + a\, \delta b$. Since $\mathcal{A}$ is commutative, the linear space $\operatorname{Der}(\mathcal{A})$ of such maps becomes an $\mathcal{A}$-module when we define $(a\delta)b = a\, \delta b$. Moreover, with the usual bracket given by the commutator $\operatorname{Der}(\mathcal{A})$ is a Lie algebra. In this paper $R = \mathbb{C}$ nearly always.

A left (right) *action* of a Lie algebra $\mathfrak{g}$ on $\mathcal{A}$ is a Lie algebra homomorphism (antihomomorphism) $\alpha : \mathfrak{g} \to \operatorname{Der}(\mathcal{A})$.

Definition 2. Assume that we are given a commutative algebra $\mathcal{A}$ and a Lie algebra $\mathfrak{g}$ which is also a faithful $\mathcal{A}$-module. The pair $(\mathcal{A}, \mathfrak{g})$ is a *Lie–Rinehart algebra* if there exists an $\mathcal{A}$-module morphism $\alpha : \mathfrak{g} \to \operatorname{Der}(\mathcal{A})$, called the *anchor*, satisfying the compatibility condition

$$a[X, Y] = [X, a\, Y] - \alpha(X)a\, Y, \tag{3.1}$$

for $a \in \mathcal{A}, X, Y \in \mathfrak{g}$. If we write m_a for the multiplication operator $m_a(X) = a\, X$ and ad_X, as usual, for the adjoint operator $\operatorname{ad}_X(Y) = [X, Y]$, the compatibility condition is rewritten as

$$[\operatorname{ad}_X, m_a] = m_{\alpha(X)a}.$$

Often, in the definition of Lie–Rinehart algebra, the apparently stronger condition that the anchor be also a left action of $\mathfrak{g}$ on $\mathcal{A}$ is required. But, as it turns out, these two definitions are equivalent.

The concept essentially coincides with Kastler and Stora's Lie–Cartan pairs [19]. As already indicated $\big(\mathcal{F}(M), \mathfrak{X}(M), \mathrm{id}\big)$ is a Lie–Rinehart algebra. A more general example of Lie–Rinehart algebra may be $\big(\mathcal{F}(M), \Gamma(M,E), \alpha\big)$, where $\Gamma(M,E)$ is the $\mathcal{F}(M)$-module of sections of a vector bundle E over M, on which a Lie bracket and an anchor α (hence a vector bundle map $E \to TM$, denoted in the same way) are supposed given. If the fibres have dimension bigger than one, then a linear map satisfying (3.1) is not only automatically a Lie algebra morphism, but also a $\mathcal{F}(M)$-module morphism. These geometrical examples are called Lie algebroids. A Lie algebroid is called *transitive* when it is onto fibrewise; totally intransitive when $\alpha = 0$. For examples of this, consider a principal bundle $P(M, G, \pi)$ over M; if VP is the vertical bundle over P, we have the exact sequences of vector bundles

$$0 \to VP \hookrightarrow TP \to TM \to 0 \quad \text{and} \quad 0 \to VP/G \hookrightarrow TP/G \to TM \to 0;$$

the second being essentially the Atiyah sequence; and then $(C^\infty(M), \Gamma(M, TP/G), T\pi/G)$ is a transitive Lie algebroid; while $\big(C^\infty(M), \Gamma(M, VP/G)\big)$ is totally intransitive.

Whenever we have a Lie–Rinehart algebra, we can algebraically define a differential calculus. For instance a n-form is a skewsymmetric n-linear map from $\mathfrak{g}$ to $\mathcal{A}$. If we define d on 1-forms by

$$d\beta(X,Y) = \alpha(X)\beta(Y) - \alpha(Y)\beta(X) - \beta\big([X,Y]\big),$$

then certainly d can be extended so $d^2 = 0$. Also, let V be an $\mathcal{A}$-module. A V-connection in the sense of [19, 20] is a linear assignment to each element $X \in \mathfrak{g}$ of a linear map $\rho(X) : V \to V$ such that, for $v \in V$,

$$\big(a\rho(X)\big)v = a\big(\rho(X)\big)v; \qquad \rho(X)(av) = a\,\rho(X)v + \alpha(X)a\,v.$$

If V is moreover a $\mathfrak{g}$-module, the connection is flat (as the curvature defined in the obvious way vanishes).

A morphism $(\mathcal{A}, \mathfrak{g}) \to (\mathcal{A}', \mathfrak{g}')$ of Lie–Rinehart algebras is a pair (ϕ, ψ) of an algebra morphism $\phi : \mathcal{A} \to \mathcal{A}'$ and an $\mathcal{A}$-module morphism $\psi : \mathfrak{g} \to \mathfrak{g}'$, where the action of $\mathcal{A}$ on $\mathfrak{g}'$ is given by $aX' := \phi(a)X'$, intertwining the anchors:

$$\phi\big(\alpha(X)\,a\big) = \alpha'\big(\psi(X)\big)\,\phi(a).$$

An important example by Grabowski is as follows [21]. A linear operator $D : \mathfrak{g} \to \mathfrak{g}$ is called a *quasi-derivation* for $\mathcal{A}$, and we write $D \in \mathrm{Qder}_{\mathcal{A}}(\mathfrak{g})$, if for each $a \in \mathcal{A}$ there exists $\widehat{D}(a) \in \mathcal{A}$ —necessarily unique— such that $[D, m_a] = m_{\widehat{D}(a)}$, where the bracket is the usual commutator. It is easily seen that $\widehat{D} \in \mathrm{Der}(\mathcal{A})$. Then $(\mathrm{id}_{\mathcal{A}}, \mathrm{ad}) : (\mathcal{A}, \mathfrak{g}) \to \big(\mathcal{A}, \mathrm{Qder}_{\mathcal{A}}(\mathfrak{g})\big)$, where $\mathrm{ad} : X \mapsto \mathrm{ad}_X$, is a morphism of Lie–Rinehart algebras.

Our first example of Hopf algebra comes now across: the (universal) enveloping algebra $\mathcal{U}(\mathfrak{g})$ of the Lie algebra $\mathfrak{g}$. The enveloping algebra is Hopf because there is the diagonal algebra homomorphism

$$\Delta : \mathcal{U}(\mathfrak{g}) \to \mathcal{U}(\mathfrak{g} \oplus \mathfrak{g}) \simeq \mathcal{U}(\mathfrak{g}) \otimes \mathcal{U}(\mathfrak{g}) \quad \text{by} \quad X \mapsto X \oplus X \mapsto X \otimes 1 + 1 \otimes X,\ \Delta 1 = 1 \otimes 1,$$

for every $X \in \mathfrak{g}$.

When $\mathfrak{g} = \mathfrak{X}(M)$, the enveloping algebra should not be confused with the algebra (with the usual composition product) of differential operators $\mathbb{D}(M)$. It is true that the first-order elements of both are the vector fields, thus coincide. However, we are going to show that if first order differential operators are to be considered primitive elements, then $\mathbb{D}(M)$ cannot be given a natural Hopf algebra structure; we thank P. Aschieri for making us aware of the following argument, since published [22]. Consider the linear map $\mathcal{L} : \mathcal{U}(\mathfrak{X}(M)) \to \mathbb{D}(M)$ obtained by extending the Lie derivative. This map is not onto because zeroth order differential operators are functions, whereas the zeroth order elements of $\mathcal{U}(\mathfrak{X}(M))$ are just scalars. Actually, $\mathbb{D}(M)$ is a $\mathcal{F}(M)$-module, while $\mathcal{U}(\mathfrak{X}(M))$ is not. For this very reason $\mathbb{D}(M)$ cannot be a Hopf algebra. Consider two commuting linearly independent vector fields X, Y nonvanishing on a common domain (e.g. locally let X be the partial derivative ∂_i and Y a different one ∂_j), and the vector fields aX, aY, where a is an arbitrary function, nonvanishing on the same domain. The composition $a\mathcal{L}_X\mathcal{L}_Y = a\mathcal{L}_Y\mathcal{L}_X$ is an element in $\mathbb{D}(M)$. Suppose *arguendo* that there exist on $\mathbb{D}(M)$ a coproduct δ compatible with composition of operators, and such that vector fields are primitives. We would have

$$
\begin{aligned}
&\delta(a\mathcal{L}_X\mathcal{L}_Y) = a\mathcal{L}_X\mathcal{L}_Y \otimes 1 + a\mathcal{L}_X \otimes \mathcal{L}_Y + \mathcal{L}_Y \otimes a\mathcal{L}_X + 1 \otimes a\mathcal{L}_X\mathcal{L}_Y \quad \text{and} \\
&\delta(a\mathcal{L}_Y\mathcal{L}_X) = a\mathcal{L}_Y\mathcal{L}_X \otimes 1 + a\mathcal{L}_Y \otimes \mathcal{L}_X + \mathcal{L}_X \otimes a\mathcal{L}_Y + 1 \otimes a\mathcal{L}_Y\mathcal{L}_X.
\end{aligned}
$$

Now, the right hand sides are not equal. As a corollary we have that the map $\mathcal{L}$ is not injective: the fact that in $\mathcal{U}(\mathfrak{X}(M))$ one has a good coproduct implies that $aX \cdot Y$ is there different ¿from $aY \cdot X$, with $\cdot$ the product in the enveloping algebra. Thus the kernel of $\mathcal{L}$ contains $aX \cdot Y - aY \cdot X$. Notice that the argument fails if M is one-dimensional. Notice as well that δ makes $\mathbb{D}(M)$ into a good coalgebra over $\mathcal{F}(M)$. However, the product is then not $\mathcal{F}(M)$-linear.

In spite of the above, Hopf algebra renders us a first great service in helping to manufacture $\mathbb{D}(M)$ out of $\mathcal{U}(\mathfrak{X}(M))$. This involves a purely algebraic construction [11, 23] suggested by the previous discussion and better presented in the context of Lie–Rinehart algebras.

Assume for simplicity that $\mathcal{A}$ is unital. The *universal object* of $(\mathcal{A}, \mathfrak{g})$ is by definition a triple $(\mathcal{U}(\mathcal{A}, \mathfrak{g}), \imath_\mathcal{A}, \imath_\mathfrak{g})$, where $\mathcal{U}(\mathcal{A}, \mathfrak{g})$ is an associative algebra, therefore a Lie algebra with the usual commutator, together with morphisms $\imath_\mathcal{A} : \mathcal{A} \to \mathcal{U}(\mathcal{A}, \mathfrak{g})$ and $\imath_\mathfrak{g} : \mathfrak{g} \to \mathcal{U}(\mathcal{A}, \mathfrak{g})$, respectively of algebras and Lie algebras, such that

$$
\imath_A(a)\imath_\mathfrak{g}(X) = \imath_\mathfrak{g}(aX); \qquad [\imath_\mathfrak{g}(X), \imath_A(a)] = \imath_A(\alpha(X)a);
$$

and $(\mathcal{U}(\mathcal{A}, \mathfrak{g}), \imath_\mathcal{A}, \imath_\mathfrak{g})$ is universal among these triples; that is, for a similar triple $(B, \phi_\mathcal{A}, \phi_\mathfrak{g})$, there is a unique algebra morphism $\Phi_B : \mathcal{U}(\mathcal{A}, \mathfrak{g}) \to B$ such that $\Phi_B\imath_\mathcal{A} = \phi_\mathcal{A}$ and $\Phi_B\imath_\mathfrak{g} = \phi_\mathfrak{g}$.

To construct $\mathcal{U}(\mathcal{A}, \mathfrak{g})$ we employ $\mathcal{U}(\mathfrak{g})$ in the following way. The condition that the anchor maps into derivations means precisely that $\mathcal{A}$ is a Hopf $\mathcal{U}(\mathfrak{g})$-module [3]. One may keep using the same notation α for the new action, as for the generators; $\alpha(1) = 1$. Consider now the smash product or crossed product algebra $\mathcal{A} \rtimes \mathcal{U}(\mathfrak{g})$. This is the vector space $\mathcal{A} \otimes \mathcal{U}(\mathfrak{g})$ with the product defined on simple tensors by

$$
(a \otimes u)(b \otimes v) := a\alpha(u_{(1)})b \otimes u_{(2)}v,
$$

where we use the standard Sweedler notation $\Delta u = u_{(1)} \otimes u_{(2)}$. There are obvious morphisms $\imath'_\mathcal{A} : \mathcal{A} \to \mathcal{A} \rtimes \mathcal{U}(\mathfrak{g})$ and $\imath'_\mathfrak{g} : \mathfrak{g} \to \mathcal{A} \rtimes \mathcal{U}(\mathfrak{g})$. Now, let J be the right ideal generated in $\mathcal{A} \rtimes U(\mathfrak{g})$

by the elements $ab \otimes X - a \otimes bX$. One has

$$
\begin{aligned}
(c \otimes X)(ab \otimes Y - a \otimes bY) &= cab \otimes XY + c\alpha(X)(ab) \otimes Y - ca \otimes XbY - c\alpha(X)a \otimes bY \\
&= cab \otimes XY + c\alpha(X)a\, b \otimes Y + ca\alpha(X)b \otimes Y \\
&\quad - ca \otimes bXY - ca \otimes \alpha(X)bY - c\alpha(X)a \otimes bY \\
&= cab \otimes XY - ca \otimes bXY + c\alpha(X)a\, b \otimes Y - c\alpha(X)a \otimes bY \\
&\quad + ca\alpha(X)b \otimes Y - ca \otimes \alpha(X)bY,
\end{aligned}
$$

where in the last equality we just reordered terms; hence J is two-sided. By construction it is clear that the quotient $\mathcal{U}(\mathcal{A}, \mathfrak{g}) := \mathcal{A} \rtimes U(\mathfrak{g})/J$ together with the obvious quotient morphisms $\imath_{\mathcal{A}}$ and $\imath_{\mathfrak{g}}$ possesses the universal property. Note that $\imath_{\mathcal{A}}$ is injective. A morphism $(\phi, \psi) : (\mathcal{A}, \mathfrak{g}) \to (\mathcal{A}', \mathfrak{g}')$ induces a morphism of algebras $\mathcal{U}(\phi, \psi) : \mathcal{U}(\mathcal{A}, \mathfrak{g}) \to \mathcal{U}(\mathcal{A}', \mathfrak{g}')$, and vice versa; this is an equivalence of categories.

One obtains by this construction the ordinary algebra of differential operators $\mathbb{D}(M) = \mathcal{U}(\mathcal{F}(M), \mathfrak{X}(M))$. It is also clear that $\mathcal{U}(\mathbb{C}, \mathfrak{g}) = \mathcal{U}(\mathfrak{g})$ with trivial action of $\mathfrak{g}$. Just like the enveloping algebra, the universal algebra $\mathcal{U}(\mathcal{A}, \mathfrak{g})$ is filtered, with an associated graded object $\operatorname{gr} \mathcal{U}(\mathcal{A}, \mathfrak{g})$, which is a commutative graded $\mathcal{A}$-algebra. There is also a Poincaré–Birkhoff–Witt theorem for $\mathcal{U}(\mathcal{A}, \mathfrak{g})$ when $\mathfrak{g}$ is projective over $\mathcal{A}$ —which is the case in the geometrical examples. It claims that if $S_{\mathcal{A}}[\mathfrak{g}]$ is the symmetric $\mathcal{A}$-algebra on $\mathfrak{g}$, then the natural surjection

$$S_{\mathcal{A}}[\mathfrak{g}] \to \operatorname{gr} \mathcal{U}(\mathcal{A}, \mathfrak{g}), \tag{3.2}$$

is an isomorphism of $\mathcal{A}$-algebras, rather like the $S[\mathfrak{g}] \simeq \operatorname{gr} \mathcal{U}(\mathfrak{g})$ effected, say, through the ‘symmetrization’ map $\sigma : S[\mathfrak{g}] \to \mathcal{U}(\mathfrak{g})$. In that case $\imath_{\mathfrak{g}}$ is of course injective.

4 Coming by Hopf algebra

We begin here a journey from the geometrical to the Hopf world. Let us start by some well-known observations [24]. A smooth manifold M is determined by the linear space $\mathcal{F}(M)$, in the sense that points of M are in one-to-one correspondence with a particular class of linear functionals on $\mathcal{F}(M)$, to wit, multiplicative ones. One writes $\langle x, h\rangle := h(x)$ to express this correspondence. As a consequence $M \hookrightarrow \mathcal{F}'(M)$, where $\mathcal{F}'(M)$ is the space of compactly supported distributions on M. We denote by $\mathbb{C}M$ the subspace of $\mathcal{F}'(M)$ generated by the points of M. It will sometimes be convenient to write $Th = \langle T, h\rangle$ for the value of the distribution $T \in \mathcal{F}'(M)$ at $h \in \mathcal{F}(M)$; accordingly we abbreviate to $xh = h(x)$. Many other geometrical objects can be expressed as functionals in this way; for instance, if $v_x \in T_xM$, then $\langle v_x, h\rangle \equiv v_xh$ is the derivative of h in the direction of the tangent vector v_x at x. Therefore $TM \hookrightarrow \mathcal{F}'(M)$, too. If X is a smooth vector field and $Xh := X(h)$, then xX is defined naturally by

$$(xX)h = x(Xh).$$

That is, $xX = X(x)$; and we may omit the parentheses in xXh. An advantage of thinking in this way is that operations in principle not meaningful on M make sense in $\mathcal{F}'(M)$. For instance, given a curve $\gamma : \mathbb{R}_t \to M$ with $\gamma(0) = x_0$, the definition

$$\dot\gamma(0) = \lim_{\varepsilon\downarrow 0} \frac{\gamma(\varepsilon) - x_0}{\varepsilon},$$

which looks unacceptable on M, in $\mathcal{F}'(M)$ just means that for all $h \in \mathcal{F}(M)$:

$$\dot\gamma(0)h := \Big\langle \lim_{\varepsilon\downarrow 0} \frac{\gamma(\varepsilon) - x_0}{\varepsilon}, h \Big\rangle := \lim_{\varepsilon\downarrow 0} \frac{h(\gamma(\varepsilon)) - h(x_0)}{\varepsilon}.$$

Let now $f : N \to M$ be smooth. If $S \in \mathcal{F}'(N)$, a corresponding element S_f is defined in $\mathcal{F}'(M)$ by

$$S_f h := S(f^* h).$$

Clearly, $x_f = xf = f(x)$, and so $S \to S_f$ extends f to a map from $\mathcal{F}'(N)$ to $\mathcal{F}'(M)$. Somewhat rashly, one denotes the extension by the same letter; with this notation, if v_x is a tangent vector at $x \in N$, the tangent vector $T_x f(v_x)$ at $f(x) \in M$ would become $v_x f$.

We may freely use the notation e^{tX} for the flow generated by a vector field X: if $\gamma_X(t, x_0)$ is the integral curve of X going through x_0 at $t = 0$ and $xe^{tX} := \gamma_X(t, x)$, then the identity $\frac{d}{dt}\big(\gamma_X(t,x)\big) = X\big(\gamma_X(t,x)\big)$ acquires the linear look $\frac{d}{dt}\big(xe^{tX}\big) = xe^{tX}X$. We have indeed linearized the dynamical system associated to X. In more detail: if $xe^{tX}h := h\big(\gamma_X(t,x)\big)$, then

$$\begin{aligned}\frac{d}{dt}\big(xe^{tX}\big)h &= \sum_{i=1}^n \frac{dx^i}{dt}\frac{\partial h}{\partial x^i}\big(\gamma_X(t,x)\big) = \sum_{i=1}^n X^i\big(\gamma_X(t,x)\big)\frac{\partial h}{\partial x^i}\big(\gamma_X(t,x)\big) \\ &= Xh\big(\gamma_X(t,x)\big) =: xe^{tX}Xh.\end{aligned}$$

Linearization works as well for non-autonomous dynamical systems. Recall (1.4) under the form:

$$\dot x = Y(t; x(t)); \qquad x(t_0) = x_0. \tag{4.1}$$

For t given, the vector $Y(t; x(t))$ lives in the fibre over $x(t)$. Denote

$$L(t,t_0)h(x) = h(x(t)), \quad \text{for } h \in \mathcal{F}(M);\text{ then } \quad \frac{dL(t,t_0)}{dt} = L(t,t_0)Y(t);$$

with Y interpreted as the corresponding time-dependent vector field. This, as announced in Section 1, is in the guise of (1.2). The Cauchy problem

$$\dot x = xL(t,t_0); \qquad x(t_0) = x_0$$

has that of (4.1) as unique solution [25]. The difference with the finite-dimensional case is of course substantial; at the analytical level this is discussed at the end of Section 11.

Linearization is precisely what Hopf algebra is about. Things become really interesting when there is a symmetry group G of the manifold M. As pointed out in [26], linearization is then a particularly good idea; for instance, often the action Φ is indecomposable (think of the case $M = G$ and lateral action) and so contains little information; whereas the linear actions of G on $\mathcal{F}(M)$ and $\mathcal{F}'(M)$ are generally decomposable (for instance when $G = M = \mathbb{S}^1$). This is the point of harmonic analysis. For a fully algebraic description of these phenomena, we try to regard $\mathcal{F}'(G)$ and $\mathcal{F}(G)$, eventually restricting appropriately the functors $\mathcal{F}', \mathcal{F}$, as Hopf algebras.

There is no trouble in recognizing an algebra structure for the *whole* of $\mathcal{F}'(G)$: this is given just by *convolution*, which is a map $\mathcal{F}'(G) \otimes \mathcal{F}'(G) \hookrightarrow \mathcal{F}'(G \times G) \to \mathcal{F}'(G)$. If $S_1, S_2 \in \mathcal{F}'(G)$, then $S_1 * S_2$ is defined as the image $\mu(S_1, S_2)$, of the extended group multiplication $\mu : G \times G \to G$. This is an associative operation [27]. We have in particular $g_1 * g_2 = g_1 g_2$, for $g_1, g_2 \in G$. The unit element in $\mathcal{F}'(G)$ is 1_G. An augmentation on $\mathcal{F}'(G)$ is given by evaluation on the function $1 \in \mathcal{F}(G)$:

$$\eta(S) = S1.$$

In particular $\eta(g) = 1$ for all $g \in G$. Clearly $\eta(S_1 * S_2) = \eta(S_1)\eta(S_2)$. A candidate antipode is the extension of the inversion diffeomorphism $\imath : g \mapsto g^{-1}$; certainly it is an algebra antiautomorphism:

$$\imath(S_1 * S_2) = \imath(S_2) * \imath(S_1).$$

In the sequel, the integral notation for convolution

$$(S_1 * S_2)h' = \int dS_1(g')\, dS_2(g)\, h'(g'g),$$

will be handy. A locally summable function h defines a distribution by $h' \mapsto \int h(g)h'(g)\, dg$, with dg a (left) invariant measure on G. For instance $1 \in \mathcal{F}'(G)$ if G is compact. Now $S * h$ is defined by:

$$\int dS(g')\, h(g)h'(g'g)\, dg = \int dS(g')\, h(g'^{-1}g)h'(g)\, dg;$$

so we identify it with the function

$$g \mapsto \int dS(g')\, h(g'^{-1}g) = \int dS(gg')\, h(g). \tag{4.2}$$

Similarly, for $h * S$:

$$\int dS(g)\, h(g')h'(g'g)\, dg' = \int dS(g)\, h(g'g^{-1})h'(g')\, \delta(g^{-1})\, dg';$$

where δ is the modular function, so we identify $h * S$ with $g' \mapsto \int dS(g)\, h(g'g^{-1})\, \delta(g^{-1})$. In particular,

$$h_1 * h_2(g) = \int_G h_1(g')h_2(g'^{-1}g)\, dg' = \int_G h_1(gg')h_2(g'^{-1})\, dg' = \int h_1(gg'^{-1})h_2(g')\, \delta^{-1}(g')\, dg'.$$

To give $\mathcal{F}'(G)$ a coalgebra structure, one might try the following strategy. The diagonal homomorphism $d : G \to G \times G$, given by $g \mapsto (g, g)$, extends to $d : \mathcal{F}'(G) \to \mathcal{F}'(G \times G)$, by the repeatedly used procedure. However, $\mathcal{F}'(G \times G)$ is vastly bigger than $\mathcal{F}'(G) \otimes \mathcal{F}'(G)$. So we look for convolution subalgebras $O(G)$ of $\mathcal{F}'(G)$ for which $O(G \times G) \simeq O(G) \otimes O(G)$. For a start, $\mathbb{C}G$ will do; and naturally the elements of G, when regarded as elements of $\mathcal{F}'(G)$, are *grouplike* in the sense of Hopf algebra theory: for $h_1, h_2 \in \mathcal{F}(G)$:

$$\langle \Delta g, h_1 \otimes h_2 \rangle = \langle g, \mu(h_1 \circ d \otimes h_2 \circ d) \rangle = \langle g \otimes g, h_1 \otimes h_2 \rangle.$$

The Hopf algebra $\mathbb{C}G$ is too small for our purposes. Nonetheless, recall that the tangent algebra $\mathfrak{g}$ of G also sits inside $\mathcal{F}'(G)$. The discussion around (A.6) in Appendix A allows us to regard the elements of $\mathfrak{g}$ as right invariant differential operators on G. So let us focus on the subalgebra $\mathbb{D}^R(G) \subsetneq \mathbb{D}(G)$ of right invariant differential operators on $\mathcal{F}(G)$. There is great advantage in regarding any $D \in \mathbb{D}(G)$ as extended to distributions by $(DS)h = D(Sh)$. Now, $\mathbb{D}^R(G)$ can be realized as the algebra of distributions on G with support at 1_G. In effect, look first at the fundamental vector fields $\xi_G \equiv X^R_\xi$. From (4.2), we see that in general

$$\xi_G(S * h) = \xi_G(S) * h; \quad \text{as } 1_G * h = h, \text{ we conclude} \quad \xi_G(h) = \xi_G(1_G) * h.$$

For any element D of $\mathbb{D}^R(G)$ analogously $D(h) = D(1_G){*}h$. Note that $D(1_G)$ is a distribution concentrated at 1_G. Moreover,

$$DD'h = D(1_G) * D'(1_G) * h,$$

so the map $D \mapsto D(1_G)$ is a homomorphism. It is in fact an isomorphism, as any distribution vanishing outside a point is a finite sum of derivatives of a Dirac function; thus conversely D can be written as a polynomial in the right invariant vector fields.

Therefore we have a new subalgebra of $\mathcal{F}'(G)$. Let us just write ξ for $\xi_G(1_G)$. Furthermore, by the Leibniz rule we are able to define the shuffle coproduct:

$$\Delta\xi = \xi \otimes 1 + 1 \otimes \xi. \tag{4.3}$$

This extends to $\mathbb{D}^R(G)(1_G)$ as an algebra homomorphism. Naturally (4.3) says that the elements of $\mathfrak{g}$, when regarded as elements of $\mathcal{F}'(G)$, are primitive in the sense of Hopf algebra theory. A little more work shows that in fact $\mathbb{D}^R(G) \simeq \mathcal{U}(\mathfrak{X}^R(G)) \equiv \mathcal{U}(\mathfrak{g})$, the algebra of right invariant differential operators *coincides* with the enveloping algebra of the Lie algebra of fundamental vector fields for the left action of G on itself. Also we remark here that the equivalence of the Lie algebra structures considered on T_1G in Appendix A can be seen from

$$\xi * \eta - \eta * \xi = [\xi, \eta];$$

see [28].

Our $O(G)$ will be the convolution algebra generated by $\mathcal{U}(\mathfrak{g}) \equiv \mathbb{D}^R(G)(1_G)$ and $\mathbb{C}G$; this is a Hopf crossed product [3], as $g * \xi * g^{-1} = \mathrm{Ad}_g\, \xi$, and similarly for more general elements of $\mathcal{U}(\mathfrak{g})$; it can be also regarded as a completion of the latter. Note $\Delta \circ i = (i \otimes i)\Delta$ as well; we invite the reader to check the rest of the expected Hopf algebra properties. By the way, extending Ad to $S(\mathfrak{g})$ as well, it is found that the symmetrization map mentioned after (3.2) intertwines both actions of G. The centre $Z(\mathfrak{g})$ of left *and* right (Casimir) invariant differential operators is clearly a commutative algebra, isomorphic to the algebra of the G-invariant elements in $S(\mathfrak{g})$ —this states a strong form of the Gelfand–Harish–Chandra theorem.

One begins to feel the power of the Hopf algebra approach: equations (1.1) and (1.2) make sense in $O(G)$ without further ado; we are allowed to write for them

$$\dot g(t) * g^{-1}(t) = \xi(t), \quad g(t_0) = 1_G \quad \text{or} \quad \dot g(t) = \xi(t) * g(t), \quad g(t_0) = 1_G; \tag{4.4}$$

similarly for (1.2):

$$g^{-1}(t) * \dot g(t) = \eta(t), \quad g(t_0) = 1_G \quad \text{or} \quad \dot g(t) = g(t) * \eta(t), \quad g(t_0) = 1_G;$$

The rigorous but roundabout arguments at the end of Section 3 are simplified thereby. Moreover the possibility of considering interpolated equations, of the form

$$g^{-a}(t) * \dot g(t) * g^{-b}(t) = \kappa(t), \quad \text{with } a + b = 1,$$

opens distinctly [29]. This is uninvestigated.

Before leaving $\mathcal{F}'(G)$, let us note that we refrained from pondering distributions with point supports other than the elements of G. This would have allowed us in particular to consider $TG \hookrightarrow \mathcal{F}'(G)$; and then $g * v_{g'} = gv_{g'}$, $v_{g'} * g = v_{g'}g$ —see (A.4) in Appendix A. However, one should not believe $v_g * v_{g'} = v_g v_{g'}$; that is, $O(TG)$ does not embed into $\mathcal{F}'(G)$.

It is well known that the subalgebra $\mathcal{R}(G)$ of 'representative functions' in $\mathcal{F}(G)$, with its ordinary commutative multiplication, is also a Hopf algebra. The space $\mathcal{R}(G)$ is made of those functions whose translates $x \mapsto h(xt)$, for all $t \in G$, generate a finite-dimensional subalgebra of $\mathcal{F}(G)$. Then also $\mathcal{R}(G)$ is endowed with a coproduct in which

$$\Delta h \in \mathcal{R}(G) \otimes \mathcal{R}(G) \quad \text{is given by} \quad \Delta h(x,y) := \big(h_{(1)} \otimes h_{(2)}\big)(x,y) := h(xy); \tag{4.5}$$

which is not cocommutative, unless G is abelian. One has:

$$\eta(h) = h(1); \qquad Sh(g) = h(g^{-1}).$$

Both previous constructions of $O(G)$ and $\mathcal{R}(G)$ are mutually dual. Questions of duality are delicate in Hopf algebra theory; fortunately we need not deal with them in particular detail. The main point there is the following. Given any Hopf algebra H and an algebra A, one can define [3] the *algebraic convolution* of two $\mathbb{C}$-linear maps $f, h \in \operatorname{Hom}(H, A)$ as the map $f * h \in \operatorname{Hom}(H, A)$ given by the composition

$$H \xrightarrow{\Delta} H \otimes H \xrightarrow{f \otimes h} A \otimes A \xrightarrow{m_A} A.$$

Here m_A denotes the product map from $A \otimes A$ to A. Because of coassociativity of Δ, the triple $\big(\operatorname{Hom}(H, A), *, u_A\eta_H =: \eta_A\big)$ is an associative algebra with unit. Now, for $A = \mathbb{C}$, the Hopf algebraic definition of convolution on $O(G)$ as a dual of $\mathcal{R}(G)$ *coincides* with the analytical one.

Algebra morphisms respect convolution, in the following way

$$\ell(f * h) = \ell f * \ell h; \quad \text{similarly} \quad (f * h)\ell = f\ell * h\ell,$$

if ℓ is a coalgebra morphism. Clearly the antipode S is the inverse of the identity map id for the convolution product of endomorphisms of H [3]. If $f \in \operatorname{Hom}(H, A)$ is an algebra morphism, using the convolution product of $\operatorname{End}(H)$ one finds that its composition fS with the antipode is a convolution inverse for f:

$$f * fS = f(\mathrm{id} * S) = f u_H \eta_H = \eta_A = f(S * \mathrm{id}) = fS * f.$$

Denote by $\operatorname{Hom}_{\mathrm{alg}}(H, A)$ the convolution monoid —with unit element the map η_A— of multiplicative morphisms of H on the algebra A. In general fS does not belong to $\operatorname{Hom}_{\mathrm{alg}}(H, A)$; but it does when the algebra A is commutative. Moreover, if A is commutative, the convolution product of two multiplicative maps is again multiplicative, so $\operatorname{Hom}_{\mathrm{alg}}(H, A)$ becomes a group, that we may call $G_H(A)$. In particular, this happens for the set $G_H(\mathbb{C})$ of scalar characters, and for $G_H(H)$ if H is commutative. Thus we have a (representable by definition) functor G_H going from commutative Hopf algebras to groups. We may call G_H an 'affine group scheme'. If we suppose H to be graded, connected (meaning that the scalars are the only elements in degree zero) and of finite type, then $G_H(\mathbb{C})$ is a projective limit of triangular matrix groups. An important example is studied in Appendix D.

In general there will be only an embedding —that can be made continuous— of G into the group $G_{\mathcal{R}(G)}(\mathbb{R})$ of characters of $\mathcal{R}(G)$; under favourable circumstances (for instance, for G compact, thanks to the Peter–Weyl theorem) both groups coincide. Also if A is commutative then $\mathrm{Hom}(H, A)$ is an A-algebra. Then a Lie algebra $\mathfrak{g}_H(A)$ can be obtained as well by considering the elements L ('infinitesimal characters') of $\mathrm{Hom}(H, A)$ satisfying the Leibniz rule

$$L(cd) = \eta_A(c)Ld + \eta_A(d)Lc,$$

for all $c, d \in H$. The bracket $[L_1, L_2] := L_1 * L_2 - L_2 * L_1$ of two infinitesimal characters is an infinitesimal character, and so we have a functor $\mathfrak{g}_H$ from commutative algebras to Lie algebras. Needless to say, under favourable circumstances $\mathfrak{g}_{\mathcal{R}(G)}(\mathbb{R})$ is just $\mathfrak{g}$.

To summarize, the situation is here quite different of that examined in Section 2, whereby we showed $\mathbb{D}(G) \nsim \mathcal{U}(\mathfrak{X}(G))$, whereas $\mathbb{D}^R(G) \sim \mathcal{U}(\mathfrak{X}^R(G))$. Notice that $\mathbb{D}^R(G)$ can be expressed directly in Hopf theoretic terms, as follows: a derivation of the commutative algebra $\mathcal{R}(G)$ belongs to $\mathbb{D}^R(G)$ iff it is of the form $L * \mathrm{id}$, with L an infinitesimal character of $\mathcal{R}(G)$. Here the convolution of an endomorphism of $\mathcal{R}(G)$ and an element of $O(G)$ is clearly well defined; and indeed

$$L * \mathrm{id}(h_1h_2) = D(h_1)h_2 + h_1D(h_2),$$

after a short calculation. Right invariance of $L*\mathrm{id}$ is clear. Reciprocally ηD is an infinitesimal character. All this is in [30].

The books [28, 31, 32] and the review article [33] are good references for most of this section.

5 Some structure results for Hopf algebras

Familiarity with the tensor $\mathcal{T}(V)$ and cotensor (or shuffle) $\mathcal{T}^*(V)$ Hopf algebras is very convenient; we survey them here. Consider a countable basis $B = \{ v_1, \dots, v_p, \dots \}$ of the vector space V, and think of it as an alphabet, a word of this alphabet being a finite sequence of v's. We let $\mathcal{T}^*(V)$ be the vector space generated by the set of words and 1 (corresponding to the empty word). The length of a word $w = v_{i_1} \cdots v_{i_n}$ is denoted by $|w| = n$; naturally $|1| = 0$. Introduce a noncocommutative (deconcatenation) coproduct on $\mathcal{T}^*(V)$ by the formulae $\Delta 1 = 1 \otimes 1$ and

$$\Delta w = \sum_{p=0}^{n} v_{i_1} \cdots v_{i_p} \otimes v_{i_{p+1}} \cdots v_{i_n},$$

with the agreement that when all the terms are on the one side of the tensor sign there is a 1 on the other side. Notice that

$$(\Delta \otimes \mathrm{id})\Delta(v_{i_1} \cdots v_{i_n}) = \sum_{0 \le p \le q \le n} v_{i_1} \cdots v_{i_p} \otimes v_{i_{p+1}} \cdots v_{i_q} \otimes v_{i_{q+1}} \cdots v_{i_n} = (\mathrm{id} \otimes \Delta)\Delta(v_{i_1} \cdots v_{i_n}),$$

understanding that when $p = q$ the middle term of the summand is 1. Hence $\big(\mathcal{T}^*(V), \Delta, \eta\big)$, where $\eta : \mathcal{T}^*(V) \to \mathbb{R}$ is defined by $\eta(1) = 1$ and $\eta(w) = 0$ if $|w| > 0$, is indeed a coalgebra; by the way, any commutative $\mathbb{Q}$-algebra at the place of the real numbers would do here. The dual vector space consists of all infinite series of the form $\sum \lambda_I v'_I$, where $I = \{ i_1, \dots, i_n \}$ and v'_I denotes the dual of $v_I := v_{i_1} \cdots v_{i_n}$. It becomes an algebra with product

$$\langle v'_J v'_K, v_I \rangle := \langle v'_J \otimes v'_K, \Delta v_I \rangle.$$

Since the right hand side vanishes unless $J \cup K = I$ as ordered sets, and in that case equals 1, this product is simply concatenation:

$$v'_J v'_K = v'_{j_1} \cdots v'_{j_m} v'_{k_1} \cdots v'_{k_l}.$$

In other words, this dual is the algebra $\mathbb{R}[[B']]$ of noncommutative formal power series in the variables v_i, which is the (Krull topology) completion of the algebra $\mathbb{R}[B']$ of noncommutative polynomials in the same variables —that is the tensor algebra $\mathcal{T}(V)$, as tensor product is given by concatenation.

It is clear that $\mathcal{T}(V)$ is a free associative algebra on B' [34]. Moreover, if $\mathcal{L}(B')$ is the free Lie algebra on B', from the universal properties of $\mathcal{L}(B')$ and of $\mathcal{U}\big(\mathcal{L}(B')\big)$ it follows that also $\mathcal{U}\big(\mathcal{L}(B')\big)$ is a free associative algebra on B'; therefore $\mathcal{T}(V) = \mathcal{U}\big(\mathcal{L}(B')\big)$. In particular, we have a Hopf algebra structure on $\mathcal{T}(V)$, which is inherited by its completion. Their cocommutative coproduct is given on monomials by the formula

$$\Delta(v'_{i_1} \cdots v'_{i_n}) = \sum_{p=0}^{n} \sum_{\sigma \in S_{n,p}} v'_{\sigma(i_1)} \cdots v'_{\sigma(i_p)} \otimes v'_{\sigma(i_{p+1})} \cdots v'_{\sigma(i_n)}.$$

Here we deal with $(p, n-p)$-shuffles, that is, permutations σ of $[n] = \{ 1, \dots, n \}$ such that $\sigma(1) < \cdots < \sigma(p)$ and $\sigma(p+1) < \cdots < \sigma(n)$; we write $\sigma \in S_{n,p}$. From this coproduct we obtain a product $\pitchfork$ on $\mathcal{T}^*(V)$ by dualization:

$$\langle v'_I, v_J \pitchfork v_K \rangle := \langle \Delta v'_I, v_J \otimes v_K \rangle. \tag{5.1}$$

Explicitly, the commutative *shuffle* product $\pitchfork$ is given by

$$v_{i_1} \cdots v_{i_p} \pitchfork v_{i_{p+1}} \cdots v_{i_n} = \sum_{\sigma \in S_{n,p}} v_{i_{\sigma(1)}} \cdots v_{i_{\sigma(n)}}.$$

For instance $v_i v_j \pitchfork v_k = v_i v_j v_k + v_i v_k v_j + v_k v_i v_j$ and

$$v_i v_j \pitchfork v_k v_l = v_i v_j v_k v_l + v_i v_k v_j v_l + v_i v_k v_l v_j + v_k v_i v_j v_l + v_k v_i v_l v_j + v_k v_l v_i v_j.$$

It is also easy to check the following formula, which can be employed as a recursive definition of the shuffle product:

$$v_{i_1} \cdots v_{i_p} \pitchfork v_{i_{p+1}} \cdots v_{i_n} = (v_{i_1} \cdots v_{i_p} \pitchfork v_{i_{p+1}} \cdots v_{i_{n-1}}) v_{i_n} + (v_{i_{p+1}} \cdots v_{i_n} \pitchfork v_{i_1} \cdots v_{i_{p-1}}) v_{i_p}.$$

The coproduct on $\mathcal{T}^*(V)$ and $\pitchfork$ are compatible since they are respectively obtained by dualization of the product and coproduct of the Hopf algebra $\mathbb{R}[[B']]$. The resulting commutative, connected, graded Hopf algebra $\mathrm{Sh}(V) \equiv \mathcal{T}^*(V)$ is called the shuffle Hopf algebra over V. The construction does not depend on the choice of the basis B, since all the algebras involved only depend on the cardinality of B. The antipode on $\mathcal{T}(V)$ is given by

$$S(v'_1 \cdots v'_n) = (-1)^n v'_n \cdots v'_1;$$

by duality the same formula holds on $\mathrm{Sh}(V)$.

Every polynomial $P \in \mathbb{R}[B']$ can be written in the form $P = \sum_n P_n$ where P_n is the sum of all monomials of P of degree n. It is called a *Lie element* if $P_0 = 0$ and each P_n belongs to the free Lie algebra generated by the v'_i. The following is a classical theorem by Friedrichs.

Theorem 5.1. *A polynomial P is a Lie element if, and only if, it is primitive.*

Proof. The 'only if' part follows easily by induction from the obviously true assertion for $n = 1$. To prove the converse we invoke in context the Dynkin operator D, for whose study we recommend [29,35,36]. An abstract definition of D is $D = S*Y$, where Y is the derivation given by the grading; equivalently $\mathrm{id} * D = Y$. If P_n is primitive, then so are

$$nP_n = YP_n = \pi(\mathrm{id} \otimes D)(1 \otimes P_n + P_n \otimes 1) = D(P_n);$$

and vice versa. But $D(P_n)$ is a Lie element, as it corresponds to the left-to-right bracketing:

$$D(x_{i_1} \dots x_{i_n}) = [\dots[[x_{i_1}, x_{i_2}], x_{i_3}], \dots, x_{i_n}] \quad \text{(the Dynkin–Specht–Wever theorem).}$$

To prove the last equality, note that it is true for $|w| = 1$. Assume that it holds for all words of degree less than n, and let $w = xx_{i_n} = x_{i_1} \cdots x_{i_n}$. Then

$$\Delta w = \Delta x \Delta x_{i_n} = (x_{(1)} \otimes x_{(2)})(x_{i_n} \otimes 1 + 1 \otimes x_{i_n}) = x_{(1)}x_{i_n} \otimes x_{(2)} + x_{(1)} \otimes x_{(2)}x_{i_n}.$$

Since $Sx_{i_n} = -x_{i_n}$ and $\eta(x) = 0$,

$$\begin{aligned}(S * Y)w &= S(x_{(1)}x_{i_n})Yx_{(2)} + Sx_{(1)}Y(x_{(2)}x_{i_n}) \\ &= S(x_{i_n})Sx_{(1)}Yx_{(2)} + Sx_{(1)}Yx_{(2)}x_{i_n} + Sx_{(1)}x_{(2)}Yx_{i_n} \\ &= -x_{i_n}Sx_{(1)}Yx_{(2)} + Sx_{(1)}Yx_{(2)}x_{i_n} \\ &= [Dx, x_{i_n}] = [\dots[[x_{i_1}, x_{i_2}], x_{i_3}], \dots, x_{i_n}],\end{aligned}$$

upon using the induction hypothesis in the last equality. (The definition $D = Y * S$ would work the same, yielding right-to-left bracketing.) $\square$

When V is finite dimensional, the previous argument of Friedrichs' theorem goes through for formal power series, because the homogeneous components are polynomials, and there is only a finite number of words of a given length. In the infinite-dimensional case Lie series are defined as those such that their projections to any finite-dimensional subspace $\tilde{V}$ are Lie series over $\tilde{V}$, so the theorem also holds for series in the infinite-dimensional context [34, Section 3.1].

Given a power series Z, let us denote by (Z, w) the coefficient of the word w in Z. The topology in $\mathbb{R}[[B']]$ alluded above is the weakest topology such that for each w the mapping $Z \mapsto (Z, w)$ is continuous, when $\mathbb{R}$ is equipped with the discrete topology. In particular, the neighbourhoods of 0 are indexed by finite sets of words, and correspond to those series whose coefficients vanish on all the words of the given finite set. Thus, given a sequence of series (Z_n) such that for each neighbourhood of 0, all but a finite number of Z_n's are in this neighbourhood, their sum $\sum_n Z_n$ is defined as the power series Z satisfying

$$(Z, w) = \sum_n (Z_n, w).$$

This sum makes sense since only finitely many terms are different ¿from zero for each w. Notice that Z can be written as $Z = \sum_w (Z, w)w$, where the sum runs over the set of words. (Henceforth we shall no longer be fussy on 'topological' matters.)

When Z is a series such that $(Z,1)=0$, then the expression $\sum_n \lambda_n Z^n$ has a meaning for any choice of the numbers λ_n. In particular, we may define exponentials and logarithms as usual:

$$\exp(Z)=\sum_{n=0}^{\infty}\frac{Z^n}{n!} \quad \text{and} \quad \log(1+Z)=\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}Z^n,$$

As expected

$$\log\bigl(\exp(Z)\bigr)=Z, \quad \text{and} \quad \exp\bigl(\log(1+Z)\bigr)=1+Z.$$

and routine calculations establish that exp is a bijection from the set of primitive elements in the completion of $\mathcal{T}(V)$ into the set of grouplike elements, and vice versa for log.

Equation (5.1) entails

$$\Delta Z=\sum_{w,x}(Z,w\pitchfork x)w\otimes x.$$

Since for grouplike elements $\Delta Z=\sum_{w,x}(Z,w)(Z,x)w\otimes x$, it follows that

$$(Z,w\pitchfork x)=(Z,w)(Z,x), \tag{5.2}$$

for them. This of course means that the grouplike elements of $\mathbb{R}[[B']]$ are precisely those Z for which the map $w\mapsto(Z,w)$ is an algebra homomorphism for the shuffle product. This characterization is originally due to Ree [37].

We collect next some elements of structure theory of commutative or cocommutative Hopf algebras —mostly due to Patras [38, 39]— beginning by a 'double series' argument similar to the one in [34] for the shuffle-deconcatenation Hopf algebra.

Consider, for $H=\bigoplus_m^\infty H^{(m)}$ a graded connected *commutative* Hopf algebra with augmentation ideal H_+ and graded dual H', a suitable completion $H\mathbin{\overline{\otimes}}H'$ of the tensor product $H\otimes H'$. This is a unital algebra, with product $m\otimes\Delta^t$ and unit $1\otimes 1$. Now by Leray's theorem —an easy dual version of the Cartier–Milnor–Moore theorem— our H is a symmetric algebra over a supplement V of H_+^2 in H_+ [3, 39]. Let A index a basis for V, let $\tilde{A}$ (the monoid freely generated by A) index the words X_u, and let Z_u denote an element of the dual basis in H'; then the product on $H\mathbin{\overline{\otimes}}H'$ is given by the double series product:

$$\biggl(\sum_{u,v\in\tilde{A}}\alpha_{uv}X_u\otimes Z_v\biggr)\biggl(\sum_{w,t\in\tilde{A}}\beta_{wt}X_w\otimes Z_t\biggr):=\sum_{u,v,w,t\in\tilde{A}}\alpha_{uv}\beta_{wt}\,X_uX_w\otimes Z_vZ_t.$$

The linear embedding $\operatorname{End}H\to H\mathbin{\overline{\otimes}}H'$ given by

$$f\mapsto\sum_{u\in\tilde{A}}f(X_u)\otimes Z_u,$$

is really a convolution algebra embedding

$$(\operatorname{End}H,*)\to(H\mathbin{\overline{\otimes}}H',m\otimes\Delta^t).$$

Indeed,

$$
\begin{aligned}
\Big(\sum_{u\in\tilde A} f(X_u)\otimes Z_u\Big)\Big(\sum_{v\in\tilde A} g(X_v)\otimes Z_v\Big) &= \sum_{u,v\in\tilde A} f(X_u)g(X_v)\otimes Z_uZ_v \\
&= \sum_{t\in\tilde A}\Big(\sum_{u,v\in\tilde A} f(X_u)g(X_v)\,\langle Z_uZ_v, X_t\rangle\Big)\otimes Z_t \\
&= \sum_{t\in\tilde A}\Big(\sum_{u,v\in\tilde A} f(X_u)g(X_v)\,\langle Z_u\otimes Z_v, \Delta X_t\rangle\Big)\otimes Z_t \\
&= \sum_{t\in\tilde A} f*g(X_t)\otimes Z_t. \qquad (5.3)
\end{aligned}
$$

Notice that the identities $u\eta$ for convolution and id for composition in $\operatorname{End} H$ correspond respectively to

$$u\eta \mapsto 1\otimes 1 \quad \text{and} \quad \mathrm{id} \mapsto \sum_{u\in\tilde A} X_u\otimes Z_u.$$

Denote

$$\pi_1(X_w) := \sum_{k\geq 1}\frac{(-1)^{k-1}}{k}\sum_{u_1,\dots,u_k\neq 1}\langle Z_{u_1}\cdots Z_{u_k}, X_w\rangle\, X_{u_1}\cdots X_{u_k} =: \log^*\mathrm{id}\, X_w.$$

Using the same idea as in (5.3), we get

$$
\begin{aligned}
\log\Big(\sum_{u\in\tilde A} X_u\otimes Z_u\Big) &:= \sum_{k\geq 1}\frac{(-1)^{k-1}}{k}\Big(\sum_{u\neq 1} X_u\otimes Z_u\Big)^k \\
&= \sum_{k\geq 1}\frac{(-1)^{k-1}}{k}\sum_{u_1,\dots,u_k\neq 1} X_{u_1}\cdots X_{u_k}\otimes Z_{u_1}\cdots Z_{u_k} \\
&= \sum_{w\in\tilde A}\sum_{k\geq 1}\frac{(-1)^{k-1}}{k}\sum_{u_1,\dots,u_k\neq 1}\langle Z_{u_1}\cdots Z_{u_k}, X_w\rangle\, X_{u_1}\cdots X_{u_k}\otimes Z_w \\
&= \sum_{w\in\tilde A}\pi_1(X_w)\otimes Z_w. \qquad (5.4)
\end{aligned}
$$

We moreover consider the endomorphisms $\pi_n := \pi_1^{*n}/n!$ so that, by (5.3):

$$\sum_{w\in\tilde A}\pi_n(X_w)\otimes Z_w = \frac{1}{n!}\Big(\sum_{v\in\tilde A}\pi_1(X_v)\otimes Z_v\Big)^n.$$

We may put $\pi_0 := u\eta$. Thus, if $a\in H$ is of order n, $\pi_m(a) = 0$ for $m > n$. Furthermore, for $n > 0$,

$$\mathrm{id}^{*l}a = \exp^*(\log^*(\mathrm{id}^{*l}))\,a = \sum_{m=1}^{n}\frac{(\log^*(\mathrm{id}^{*l}))^m}{m!}a = \sum_{m=1}^{n} l^m\frac{(\log^*\mathrm{id})^m}{m!}a = \sum_{m=1}^{n} l^m\pi_m(a). \qquad (5.5)$$

In particular $\mathrm{id} = \sum_{m\geq 0}\pi_m$. The graded maps id^{*n} are called the *Adams operations* or characteristic endomorphisms of H; they play an important role in the (Hochschild, cyclic) cohomology of commutative algebras [40–42]. The π_n are often called *Eulerian idempotents*. We have for them:

Proposition 5.2. *For any integers n and k,*

$$\mathrm{id}^{*n}\,\mathrm{id}^{*k} = \mathrm{id}^{*nk} = \mathrm{id}^{*k}\,\mathrm{id}^{*n}. \tag{5.6}$$

and

$$\pi_m \pi_k = \delta_{mk}\,\pi_k. \tag{5.7}$$

Proof. The first assertion is certainly true for $k = 1$ and all integers n, and if it is true for some k and all integers n, then taking into account that id is an algebra homomorphism, the induction hypothesis gives

$$\mathrm{id}^{*n}\,\mathrm{id}^{*k+1} = \mathrm{id}^{*n}(\mathrm{id}^{*k} * \mathrm{id}) = \mathrm{id}^{*nk} * \mathrm{id}^{*n} = \mathrm{id}^{*n(k+1)}.$$

Substituting the final expression of (5.5) in (5.6), with very little work one obtains (5.7). So indeed the π_k form a family of orthogonal projectors. □

Thus the space $H = \bigoplus_m^\infty H^{(m)}$ always has the direct sum decomposition

$$H = \bigoplus_{n\geq 0} H_n := \bigoplus_{n\geq 0} \pi_n(H). \tag{5.8}$$

Moreover, from (5.5),

$$\mathrm{id}^{*l} H_n = l^n H_n,$$

so the H_n are the common eigenspaces of the operators id^{*l} with eigenvalues l^n. Thus, the decomposition (5.8) turns H into a graded algebra. Indeed, if $a \in H_r$ and $b \in H_s$, then

$$\mathrm{id}^{*l}(ab) = \mathrm{id}^{*l}a\,\mathrm{id}^{*l}b = l^{r+s}(ab),$$

and therefore m sends $H_r \otimes H_s$ into H_{r+s}. We shall denote by $\pi_n^{(m)}$ the restriction of π_n to $H^{(m)}$, the set of elements of degree m, with respect to the original grading.

If H is cocommutative instead of commutative, the previous arguments go through. One then has

$$\log\Bigl(\sum_{u\in\tilde{A}} X_u \otimes Z_u\Bigr) = \sum_{w\in\tilde{A}} X_w \otimes \pi_1(Z_w).$$

Furthermore, in this case the Eulerian idempotents of H are the transpose of the Eulerian idempotents of the graded commutative Hopf algebra H'. In particular, for H cocommutative, $\pi_1(H) = P(H)$, the Lie algebra of primitive elements in H. This is easily sharpened into the following version [39] of the Cartier–Milnor–Moore theorem: the inclusion $\pi_1(H) \hookrightarrow H$ extends to an isomorphism of $\mathcal{U}\bigl(\pi_1(H)\bigr)$ with H.

6 The CBHD *development* and Hopf algebra

There are three paradigmatic methods (and sundry hybrid forms) to deal with first order non-autonomous differential equations: the iteration formula or Dyson–Chen expansional, the Magnus expansion and the product integral. For reasons expounded later, at the beginning of Section 10, in this paper we look first for the Magnus expansion [6]. In the influential paper [43] dealing with the latter method (although Magnus' seminal contribution is not mentioned) the famous Campbell–Baker–Hausdorff–Dynkin (CBHD) formula in Lie algebra theory is shown to be a special case of general formulas for the solution of (1.2). This is scarcely surprising, as that solution involves some kind of exponential with non-commuting exponents; also the quest for 'continuous analogues' of the CBHD formula was a motivation for Chen's work. Conversely, a heuristic argument for obtaining Magnus' expansion from the CBHD formula has been known for some time [13, 44]; and a routine, if rigorous and Hopf flavoured as well, derivation of Magnus' method from CBHD is available in [45]. Hence the interest, as a prelude to our own derivation of the Magnus expansion from the CBHD development (that will employ the concept of nonlinear CBHD *recursion* and Rota–Baxter theory techniques) of rendering the proof of the CBHD expansion in Hopf algebraic terms. This was recognized as the deeper and more natural approach to the subject some fifteen years ago, but remains to date woefully ignored. Standard treatments of the CBHD development can be found in good Lie group theory books like [46].

In the sequel we follow [47] and [48]. It will be soon clear to the reader, according to the previous discussion, that the CBHD formulae are universal; thus we can as well return to the case where H is the Hopf tensor algebra $\mathcal{T}(V)$ and where V possesses a basis $B = \{ X_1, \dots, X_n \}$. The CBHD series $\sum_{m\geq 1} \Phi_m(X_1, \dots, X_n)$ is defined by

$$\sum_{m\geq 1} \Phi_m(X_1, \dots, X_n) = \log\bigl(e^{X_1} \cdots e^{X_n}\bigr),$$

where $\Phi_m(X_1, \dots, X_n)$ are homogeneous polynomials of degree m.

Now, if a is a grouplike element in a Hopf algebra H, and $f, h \in \mathrm{Hom}(H, A)$, where A is a unital algebra, then

$$f * h(a) = f(a)h(a).$$

In particular

$$\log\bigl(e^{X_1}e^{X_2} \cdots e^{X_n}\bigr) = \log^* \mathrm{id}\bigl(e^{X_1}e^{X_2} \cdots e^{X_n}\bigr) =: \pi_1\bigl(e^{X_1}e^{X_2} \cdots e^{X_n}\bigr). \tag{6.1}$$

Take first $n = 2$. Then $\Phi_m(X, Y) = \pi_1^{(m)}\bigl(e^X e^Y\bigr)$. The Cauchy product gives

$$e^X e^Y = \sum_{m\geq 0} \left(\sum_{i=0}^{m} \frac{X^i}{i!} \, \frac{Y^{n-i}}{(n-i)!} \right),$$

hence

$$\Phi_m(X, Y) = \sum_{i+j=m} \frac{1}{i!j!} \, \pi_1^{(m)}(X^i Y^j).$$

A similar argument entails the following proposition.

Proposition 6.1.

$$\Phi_m(X_1, \dots, X_n) = \sum \frac{1}{i_1! \cdots i_n!} \, \pi_1^{(m)}(X_1^{i_1} X_2^{i_2} \cdots X_n^{i_n}), \tag{6.2}$$

where the sum runs over all vectors $(i_1, \dots, i_n)$, with nonnegative coordinates, such that $i_1 + \cdots + i_n = m$.

Denote by $\varphi_n(X_1,\dots,X_n)$ the 'multilinear' part of $\Phi_n(X_1,\dots,X_n)$ —that is, the homogeneous polynomial of degree n that consist of those monomials of $\Phi_n(X_1,\dots,X_n)$ that include all the X_i's. This amounts to take $X_i^2=0$ in $\Phi_n(X_1,\dots,X_n)$, for all i. So by (6.1)

$$\varphi_n(X_1,\dots,X_n)=\pi_1^{(n)}(X_1\cdots X_n),$$

since in that case

$$\begin{aligned} e^{X_1}\cdots e^{X_n} &= (1+X_1)\cdots(1+X_n)\\ &= \sum_i X_i+\sum_{i<j}X_iX_j+\sum_{i<j<k}X_iX_jX_k+\cdots+X_1\cdots X_n. \end{aligned}\tag{6.3}$$

Now, if $X_\sigma:=(X_1,\dots,X_n)\cdot\sigma:=(X_{\sigma(1)},\dots,X_{\sigma(n)})$ denotes the standard right action of the symmetric group S_n on $V^{\otimes n}$, then the monomials that include all the X_i's are of the form X_σ, therefore

$$\pi_1^{(n)}(X_1\cdots X_n)=\sum_{\sigma\in S_n}c_\sigma X_\sigma,$$

for some coefficients c_σ, that we shall determine in a moment.

Proposition 6.2.

$$\pi_1^{(n)}(X_1\cdots X_n)=\sum_{\sigma\in S_n}\frac{(-1)^{d(\sigma)}}{n}\binom{n-1}{d(\sigma)}^{-1}X_\sigma.\tag{6.4}$$

where $d(\sigma)$ is the number of descents *of σ, that is, the number of 'errors' in ordering consecutive terms in $\sigma(1),\dots,\sigma(n)$.*

Proof. Assume that σ has d descents, say in $n_0, n_0+n_1, n_0+n_1+\cdots+n_{j-1}$, set $n_j=n-n_0-\cdots-n_{j-1}$ and let $Z=\sum_i X_i+\sum_{i<j}X_iX_j+\cdots+X_1\cdots X_n$. By (6.3), $e^{X_1}\cdots e^{X_n}=Z+Y$, where Y is a collection of terms that contains at least one factor of the form X_i^2, therefore they will not contribute to the coefficient of X_σ, and we neglect them. Now, since $\log(1+Z)=\sum\frac{(-1)^j}{j}Z^j$ we have to compute the contribution $c(j)$ from each power Z^j.

Suppose that the monomial $X_{\sigma(1)}\cdots X_{\sigma(n_0)}$ is built ¿from j_1 monomials of Z, and in general that each monomial $X_{\sigma(n_0+\cdots+n_{i-1}+1)}\cdots X_{\sigma(n_0+\cdots+n_i)}$ is the product of j_i monomials of Z. Notice that there are $\binom{n_i-1}{j_i-1}$ manners to construct each monomial, in such a way, because $X_{\sigma(n_0+\cdots+n_{i-1}+1)}$ is always in the first monomial, and once the first j_i-1 monomials are chosen, the last monomial is fixed since σ is increasing in each segment. Thus

$$c(j)=\sum_{(j_0,\dots,j_d)}\binom{n_0-1}{j_0-1}\binom{n_1-1}{j_1-1}\cdots\binom{n_d-1}{j_d-1},$$

where the sum extends over all vectors $(j_0,\dots,j_d)$ satisfying $j_0+\cdots+j_d=j$. Since $\binom{n_k-1}{j_k-1}$ is the coefficient of x^{j_k-1} in the binomial expansion of $(1+x)^{n_k-1}$, and $\sum_{i=0}^d(j_i-1)=j-d-1$, $c(j)$ is the coefficient of x^{j-d-1} in

$$\prod_{i=0}^d(1+x)^{n_k-1}=(1+x)^{\sum_{i=0}^d(n_i-1)}=(1+x)^{n-d-1},$$

we therefore conclude that

$$c(j) = \binom{n-d-1}{j-d-1}.$$

Now, we have $j \le n$ since X_σ has n letters. Also $j \ge d+1$ as X_σ is broken in $d+1$ parts. Therefore

$$c_\sigma = \sum_{j=d+1}^{n} \frac{(-1)^{j-1}}{j} \binom{n-d-1}{j-d-1} = (-1)^d \sum_{i=0}^{m} \frac{(-1)^i}{i+d+1} \binom{m}{i},$$

where $m = n-d-1$. Now, from the binomial identity

$$\int_0^1 (1-x)^m x^d\,dx = \sum_{i=0}^{m} (-1)^i \binom{m}{i} \int_0^1 x^{i+d}\,dx = \sum_{i=0}^{m} \frac{(-1)^i}{i+d+1} \binom{m}{i}.$$

Finally, a simple induction, using integration by parts, gives

$$\int_0^1 (1-x)^m x^d\,dx = \frac{d!\,m!}{(m+d+1)!} = \frac{1}{m+d+1} \binom{d+m}{d}^{-1} = \frac{1}{n} \binom{n-1}{d}^{-1}.$$

Our task is over. But the number of descents will reappear soon enough. $\square$

This construction performed here is arguably more elegant and simpler than the standard treatments of the CBHD development by purely Lie algebraic methods. We came in by the backdoor, using the bigger free associative algebra, knowing that $\log\left(e^{X_1} \cdots e^{X_n}\right)$ —and each of its homogeneous parts— is primitive, i.e., a Lie element; and that we have the Dynkin operator to rewrite it in terms of commutators.

Let us exemplify with the case $n = 2$. Obviously we have

$$\Phi_1(X,Y) = X + Y; \qquad \Phi_2(X,Y) = \frac{1}{2}[X,Y].$$

Now,

$$\pi_1^{(3)}(X_1X_2X_3) = \frac{1}{3}X_{(123)} - \frac{1}{6}\left(X_{(132)} + X_{(213)} + X_{(231)} + X_{(312)}\right) + \frac{1}{3}X_{(321)}.$$

Therefore

$$\begin{aligned}
\Phi_3(X,Y) &= \frac{1}{2}(\pi_1^{(3)}(X^2Y) + \pi_1^{(3)}(XY^2)) \\
&= \frac{1}{2}\left(\frac{1}{6}X^2Y - \frac{1}{3}XYX + \frac{1}{6}YX^2 + \frac{1}{6}XY^2 - \frac{1}{3}YXY + \frac{1}{6}Y^2X\right) \\
&= \frac{1}{12}\left([[X,Y],Y] - [[X,Y],X]\right).
\end{aligned}$$

Both cubic Lie elements appear in Φ_3. Similarly

$$\begin{aligned}
\pi_1^{(4)}(X_1X_2X_3X_4) = \frac{1}{4}X_{(1234)} - \frac{1}{12}\Big(&X_{(1243)} + X_{(1324)} + X_{(1342)} + X_{(1423)} + X_{(2134)} \\
&+ X_{(2314)} + X_{(2341)} + X_{(2413)} + X_{(3124)} + X_{(3412)} + X_{(4123)}\Big) \\
&+ \frac{1}{12}\Big(X_{(1432)} + X_{(2143)} + X_{(2431)} + X_{(3142)} + X_{(3214)} + X_{(3241)} \\
&+ X_{(3421)} + X_{(4132)} + X_{(4213)} + X_{(4231)} + X_{(4312)}\Big) - \frac{1}{4}X_{(4321)}
\end{aligned}$$

We concentrate on X^2Y^2, as it is clear that most terms coming from X^3Y or XY^3 will vanish; and in fact the corresponding contributions *in toto* come to naught. We obtain

$$\begin{aligned}\Phi_4(X,Y) &= -\frac{1}{192}\big(4XYXY + 2XY^2X + 2YX^2Y - 3YXYX - 2XY^2X - 3YX^2Y\big)\\ &= -\frac{1}{24}[[[X,Y],X],Y].\end{aligned}$$

The identity of Jacobi has been used, under the form

$$[[[X,Y],X],Y] = [[[X,Y],Y],X].$$

It is remarkable that the other quartic Lie elements, $[[[X,Y],X],X]$ and $[[[X,Y],Y],Y]$, do not appear in the fourth degree term.

7 Rota–Baxter maps and the algebraization of integration

This paper draws inspiration partly from [30], where Connes and Marcolli have introduced logarithmic derivatives in the context of Hopf algebras. Our intent and methods are different; but it is expedient to dwell here a bit on their considerations. Given H and A commutative as in the last part of Section 4, and a derivation δ on A, for a multiplicative map $\phi \in G_H(A)$ Connes and Marcolli define two maps in $\mathrm{Hom}(H,A)$ by $\delta(\phi) := \delta \circ \phi$, and then

$$D_\delta(\phi) := \phi^{-1} * \delta(\phi)$$

This yields an A-valued infinitesimal character. Indeed, using Sweedler's notation and multiplicativity of $\phi \in G_H(A)$, one has

$$\begin{aligned} D_\delta(\phi)[cd] &= \phi^{-1} * \delta(\phi)[cd] = m_A(\phi^{-1} \otimes \delta(\phi))\Delta(cd) = \phi^{-1}(c_{(1)}d_{(1)})\delta(\phi(c_{(2)}d_{(2)}))\\ &= \phi^{-1}(c_{(1)})\phi^{-1}(d_{(1)})\big(\delta(\phi(c_{(2)}))\phi(d_{(2)}) + \phi(c_{(2)})\delta(\phi(d_{(2)}))\big)\\ &= \phi^{-1}(c_{(1)})\delta(\phi(c_{(2)}))\phi^{-1}(d_{(1)})\phi(d_{(2)}) + \phi^{-1}(c_{(1)})\phi(c_{(2)})\phi^{-1}(d_{(1)})\delta(\phi(d_{(2)}))\\ &= D_\delta(\phi)[c]\eta_A(d) + \eta_A(c)D_\delta(\phi)[d]. \end{aligned}$$

Therefore $D_\delta(\phi)$ belongs to $\mathfrak{g}_H(A)$.

The Dynkin operator appearing in Section 5 —one of the fundamental Lie idempotents in the theory of free Lie algebras [34,36]— is a close cousin of the logarithmic derivative $D_\delta(g)$. Consider $G_H(H)$, for H connected and graded. The grading operator Y is a derivation of H

$$Y(hh') = Y(h)h' + hY(h') =: |h|hh' + hh'|h'|.$$

The map Y extends naturally to a derivation on $\mathrm{End}(H)$. With $f, g \in \mathrm{End}(H)$ and $h \in H$ we find

$$\begin{aligned} Y(f*g)(h) &:= f*g\,(Y(h)) = |h|(f*g)(h) = |h|f(h^{(1)})g(h^{(2)})\\ &= |h^{(1)}|f(h^{(1)})g(h^{(2)}) + |h^{(2)}|f(h^{(1)})g(h^{(2)}) = Yf*g\,(h) + f*Yg\,(h), \end{aligned}$$

where we used that $\Delta(Y(h)) = |h|\Delta(h) = \big(|h^{(1)}| + |h^{(2)}|\big)\, h^{(1)} \otimes h^{(2)}$. Now, as before, convolution of the antipode S with the derivation Y of H defines a Dynkin operator, to be interpreted as an H-valued infinitesimal character [29].

Suppose we have a smooth map $t \mapsto L(t)$ from $\mathbb{R}_t$ to $\mathfrak{g}_H(A)$. We could say that one of the main aims of this paper is to solve for $g(t)$ the initial value scheme

$$D_{d/dt}\big(g(t)\big) = L(t); \qquad g(0) = \eta_A, \tag{7.1}$$

at least for (real and) complex points. Now, both the classical notions of derivation and integration have interesting generalizations. It would then be a pity to limit ourselves to the classical framework; and so we now jump onto a somewhat more adventurous path.

For integration, one lacks a good algebraic theory similar to the one developed in [49], say. Next we elaborate on a somewhat unconventional presentation of the integration-by-parts rule using the algebraic notion of the weight-θ Rota–Baxter relation corresponding to the generalization of the Leibniz rule in terms of weight-θ *skewderivations*. One should strive for nothing less ambitious than developing Rota's program, beautifully outlined in [50] in the context of Chen's work [8], of establishing an algebraic theory of integration in terms of generalizations of the integration-by-parts rule.

Let us recall first the integration-by-parts rule for the Riemann integral map. Let $A := C(\mathbb{R})$ be the ring of real continuous functions. The indefinite Riemann integral can be seen as a linear map on A

$$I : A \to A, \qquad I(f)(x) := \int_0^x f(t)\,dt. \tag{7.2}$$

Then, integration-by-parts for the Riemann integral can be written as follows. Let

$$F(x) := I(f)(x) = \int_0^x f(t)\,dt, \qquad G(x) := I(g)(x) = \int_0^x g(t)\,dt;$$

then

$$\int_0^x F(t)\frac{d}{dt}\big(G(t)\big)\,dt = F(x)G(x) - \int_0^x \frac{d}{dt}\big(F(t)\big)G(t)\,dt.$$

More compactly, this well-known identity is written

$$I(f)(x)I(g)(x) = I\big(I(f)g\big)(x) + I\big(fI(g)\big)(x), \tag{7.3}$$

dually to the Leibniz rule.

Now, we introduce so-called skewderivations of weight $\theta \in \mathbb{R}$ on an algebra A [51]. A skewderivation is a linear map $\delta : A \to A$ fulfilling the condition

$$\delta(ab) = a\delta(b) + \delta(a)b - \theta\delta(a)\delta(b). \tag{7.4}$$

We call skewdifferential algebra a double $(A, \delta; \theta)$ consisting of an algebra A and a skewderivation δ of weight θ. A skewderivation of weight $\theta = 0$ is just an ordinary derivation. An induction argument shows that if A is commutative we have

$$\delta(a^n) = \sum_{i=1}^{n} \binom{n}{i} (-\theta)^{i-1} a^{n-i} \delta(a)^i.$$

Also

$$\delta^n(ab) = \sum_{i=0}^{n} \binom{n}{i} \sum_{j=0}^{n-i} \binom{n-i}{j} (-\theta)^i \delta^{n-j}(a)\delta^{i+j}(b).$$

Both formulae generalize well-known identities for an ordinary derivation. We mention examples. First, on a suitable function algebra A the simple finite difference operation $\delta : A \to A$ of step λ,

$$\delta(f)(x) := \frac{f(x-\lambda) - f(x)}{\lambda}, \tag{7.5}$$

satisfies identity (7.4) with $\theta = -\lambda$. See [52] for an interesting application of the $\lambda = 1$ case in the context of multiple zeta values. A closely related, though at first sight different, example is provided by the q-difference operator

$$\delta_q f(x) := \frac{f(qx) - f(x)}{(q-1)x} \tag{7.6}$$

which satisfies the q-analog of the Leibniz rule,

$$\delta_q(fg)(x) = \delta_q f(x) g(x) + f(qx)\delta_q g(x) = \delta_q f(x) g(qx) + f(x)\delta_q g(x).$$

This corresponds to relation (7.4) for $\theta = (1-q)$, modulo the identity

$$\delta_q(fg)(x) = \delta_q f(x) g(x) + f(x)\delta_q g(x) + x(q-1)\delta_q f(x)\delta_q g(x);$$

defining now $\bar{\delta}_q = x\delta_q$, it is a simple matter to check that $\bar{\delta}_q$ is a skewderivation of weight $1-q$.

We may ask for an *integration operator* corresponding to the skewderivation in (7.5). On a suitable class of functions, we define the summation operator

$$Z(f)(x) := \sum_{n\geq 1} \theta f(x+\theta n). \tag{7.7}$$

For δ being the finite difference map of step θ,

$$\begin{aligned} Z\delta(f)(x) &= \sum_{n\geq 1} \theta\delta(f)(x+\theta n) = \sum_{n\geq 1} \theta \, \frac{f(x+\theta n-\theta) - f(x+\theta n)}{\theta} \\ &= \sum_{n\geq 1} f\big(x+\theta(n-1)\big) - f(x+\theta n) = \sum_{n\geq 0} f(x+\theta n) - \sum_{n\geq 1} f(x+\theta n) = f(x). \end{aligned}$$

As δ is linear we find as well $\delta Z(f) = f$. Observe, moreover, that

$$\begin{aligned} &\Big(\sum_{n\geq 1} \theta f(x+\theta n)\Big)\Big(\sum_{m\geq 1} \theta g(x+\theta m)\Big) = \sum_{n\geq 1, m\geq 1} \theta^2 f(x+\theta n) g(x+\theta m) \\ &= \Big(\sum_{n>m\geq 1} + \sum_{m>n\geq 1} + \sum_{m=n\geq 1}\Big)\theta^2 f(x+\theta n) g(x+\theta m) \\ &= \sum_{m\geq 1}\Big(\sum_{k\geq 1} \theta^2 f\big(x+\theta(k+m)\big)\Big) g(x+\theta m) + \sum_{n\geq 1}\Big(\sum_{k\geq 1} \theta^2 g\big(x+\theta(k+n)\big)\Big) f(x+\theta n) \\ &+ \sum_{n\geq 1} \theta^2 f(x+\theta n) g(x+\theta n) = Z\big(Z(f)g\big)(x) + Z\big(fZ(g)\big)(x) + \theta Z(fg)(x). \end{aligned} \tag{7.8}$$

Related to the q-difference operator (7.6) there is the *Jackson integral*

$$J[f](x) := \int_0^x f(y)\, d_q y = (1-q)\sum_{n\geq 0} f(xq^n) x q^n \qquad (0 < q < 1).$$

This can be written in a more algebraic way, using the operator $P_q[f] := \sum_{n>0} E_q^n[f]$, with the algebra endomorphism (q-dilatation) $E_q[f](x) := f(qx)$, for $f \in A$. The map P_q is a Rota–Baxter operator of weight -1 and hence, $\mathrm{id} + P_q =: \hat{P}_q$ is of weight $+1$, see [60]. Jackson's integral is given in terms of the above operators P_q and the multiplication operator $M[f](x) := xf(x)$, $f \in A$, by $J[f](x) = (1-q)\hat{P}_q M[f](x)$. The modified Jackson integral $\bar{J}$, defined by $\bar{J}[f](x) = (1-q)\hat{P}_q[f](x)$, satisfies the relation

$$\bar{J}[f]\,\bar{J}[g] + (1-q)\bar{J}[f\,g] = \bar{J}\big[f\,\bar{J}[g]\big] + \bar{J}\big[\bar{J}[f]\,g\big].$$

For motivational reasons we remark that the map $\hat{P}_q$ is of importance in the construction of q-analogs of multiple-zeta-values. The examples motivate the generalization of the dual relation between the integration-by-parts rule and the Leibniz rule for the classical calculus.

Definition 3. A *Rota–Baxter map* R of weight $\theta \in \mathbb{R}$ on a not necessarily associative algebra A, commutative or not, is a linear map $R : A \to A$ fulfilling the condition

$$R(a)R(b) = R(R(a)b) + R(aR(b)) - \theta R(ab), \qquad a, b \in A. \tag{7.9}$$

The reader will easily verify that $\tilde{R} := \theta\,\mathrm{id} - R$ is a Rota–Baxter map of the same weight, as well. We call a pair (A, R), where A is an algebra and R a Rota–Baxter map of weight θ, a *Rota–Baxter algebra* of weight θ. The indication 'not necessarily associative' is indispensable in this paper, as we soon meet Rota–Baxter algebras that are neither Lie nor associative.

We state a few simple observations, which will be of use later. The so-called *double Rota–Baxter* product

$$x *_R y := xR(y) + R(x)y - \theta xy, \qquad x, y \in A, \tag{7.10}$$

endows the vector space underlying A with another Rota–Baxter algebra structure, denoted by (A_R, R). In fact, R satisfies the Rota–Baxter relation for the new product. One readily shows, moreover:

$$R(x *_R y) = R(x)R(y) \quad \text{and} \quad \tilde{R}(x *_R y) = -\tilde{R}(x)\tilde{R}(y), \qquad x, y \in A. \tag{7.11}$$

This construction may be continued, giving a hierarchy of Rota–Baxter algebras.

Proposition 7.1. *Let (A, R) be an associative Rota–Baxter algebra of weight $\theta \in \mathbb{R}$. The Rota–Baxter relation extends to the Lie algebra A with the commutator $[x, y] := xy - yx$,*

$$[R(x), R(y)] + \theta R\big([x, y]\big) = R\big([R(x), y] + [x, R(y)]\big)$$

making $(A, [.,.], R)$ into a Rota–Baxter Lie algebra.

This is a mere algebra exercise. A more exotic result coming next will prove to be important in the context of Magnus' expansion and beyond.

Proposition 7.2. *Let (A, R) be an associative Rota–Baxter algebra of weight $\theta \in \mathbb{R}$. The binary composition*

$$a \cdot_R b := [a, R(b)] + \theta ba \tag{7.12}$$

defines a right pre-Lie (or Vinberg) product such that A becomes a Rota–Baxter right pre-Lie algebra.

Proof. Recall that for a pre-Lie algebra $(A,\cdot)$ the (right) pre-Lie property is weaker than associativity

$$a\cdot(b\cdot c)-(a\cdot b)\cdot c=a\cdot(c\cdot b)-(a\cdot c)\cdot b,\quad \forall a,b,c\in A,$$

As the Jacobiator is the total skewsymmetrization of the associator, the pre-Lie relation is enough to guarantee that the commutator $[a,b]:=a\cdot b-b\cdot a$ satisfies the Jacobi identity. For the sake of brevity we verify only the weight-zero case and leave the rest to the reader.

$$\begin{aligned}
a\cdot_R(b\cdot_R c)-(a\cdot_R b)\cdot_R c&=\big[a,R([b,R(c)])\big]-\big[[a,R(b)],R(c)\big]=\\
&=\big[a,R([b,R(c)])\big]+\big[[R(c),a],R(b)\big]+\big[a,[R(c),R(b)]\big]\\
&=\big[a,R([c,R(b)])\big]-\big[[a,R(c)],R(b)\big]=:a\cdot_R(c\cdot_R b)-(a\cdot_R c)\cdot_R b;\\
\text{and}\quad R(a)\cdot_R R(b)&=[R(a),R(R(b))]=R\big([R(a),R(b)]\big)+R\big([a,R(R(b))]\big)\\
&=R\big(R(a)\cdot_R b\big)+R\big(a\cdot_R R(b)\big).
\end{aligned}$$

Here we used Proposition 7.1 as well as the Jacobi identity. □

The Lie algebra bracket corresponding to the double Rota–Baxter product (7.10) is the double Rota–Baxter Lie bracket $[a,b]_R:=a{*_R}b-b{*_R}a=a{\cdot_R}b-b{\cdot_R}a$, known since the work of Semenov-Tian-Shansky [53]. We should mention that these little calculations become more transparent using the link between associative Rota–Baxter algebras and Loday's dendriform algebras [54, 55].

As a corollary to the last propositions we add the following identity which will also be useful later

$$\begin{aligned}
R(a\cdot_R b)=R([a,R(b)])+\theta ba)&=R([b,R(a)])+[R(a),R(b)]+\theta R(ab)\\
&=R(b\cdot_R a)+[R(a),R(b)],
\end{aligned}\tag{7.13}$$

which is another way of saying that

$$R([a,b]_R)=R(a*_R b-b*_R a)=[R(a),R(b)]=R(a\cdot_R b-b\cdot_R a).$$

The triple $(A,\delta,R;\theta)$ will denote an algebra A endowed with a skewderivation δ and a corresponding Rota–Baxter map R, both of weight θ, such that $R\delta a=a$ for any $a\in A$ such that $\delta a\neq 0$, as well as $\delta Ra=a$ for any $a\in A, Ra\in 0$. We check consistency of the conditions (7.9) and (7.4) imposed on R,δ. Respectively

$$\begin{aligned}
\theta\delta R(ab)&=R(a)b+aR(b)-\delta(R(a)R(b))=R(a)b+aR(b)-R(a)b-aR(b)+\theta ab=\theta ab;\\
R\delta(ab)&=R(a\delta(b))+R(\delta(a)b)-\theta R(\delta(a)\delta(b))=R(a\delta(b))+R(\delta(a)b)\\
&\quad-R(a\delta(b))-R(\delta(a)b)+ab=ab.
\end{aligned}$$

The moral of the story is that Rota–Baxter maps are *generalized integrals*, skewderivations and Rota–Baxter operators being natural (partial) inverses. As an example we certainly have $(C(\mathbb{R}),d/dt,\int;0)$, with $\delta=$ the derivative (with only the scalars in its kernel). Another example is given by the aforementioned triple $(A,\delta,Z;-\theta)$ of the finite difference map δ of step θ and the summation Z in (7.7).

Rota–Baxter algebras have attracted attention in different contexts, such as perturbative renormalization in quantum field theory (see references further below) as well as generalized shuffle relations in combinatorics [56]. A few words on the history of the Rota–Baxter relation are probably in order here. In the 1950's and early 1960's, several interesting results were obtained in the fluctuation theory of probability. One of the better known is Spitzer's classical identity [57] for sums of independent random variables. In an important 1960 paper [58], the American mathematician G. Baxter developed a combinatorial point of view on Spitzer's result, and deduced it from the above operator identity (7.9), in the context where the algebra A is associative, unital and commutative. Then G.-C. Rota started a careful in depth elaboration of Baxter's article in his 1969 papers [59, 60], where he solved the crucial "word problem", and in [61], where he established several important results. During the 1960's and 1970's, further algebraic, combinatorial and analytic aspects of Baxter's identity were studied by several people, see [62–64] for more references. Recently, the Rota–Baxter relation became popular again as a key element of the Connes–Kreimer [65–67] algebraic approach to renormalization.

At an early stage the mathematician F. V. Atkinson made an important contribution, characterizing such algebras by a simple decomposition theorem.

Theorem 7.3. (Atkinson [68]) *Let A be an algebra. A linear operator $R : A \to A$ satisfies the Rota–Baxter relation* (7.9) *if and only if the following two statements are true. First, $A_+ := R(A)$ and $A_- := (\theta\,\mathrm{id} - R)(A)$ are subalgebras in A. Second, for $X, Y, Z \in A$, $R(X)R(Y) = R(Z)$ implies $(\theta\,\mathrm{id} - R)(X)(\theta\,\mathrm{id} - R)(Y) = -(\theta\,\mathrm{id} - R)(Z)$.*

This result degenerates in the case $\theta = 0$, whereby $R = -\tilde{R}$. A trivial observation is that every algebra is a Rota–Baxter algebra (of weight 1); in fact, the identity map and the zero map are a natural Rota–Baxter pair. The case of an idempotent Rota–Baxter map implies $\theta = 1$ and, more importantly, $A_- \cap A_+ = \{0\}$, corresponding to a direct decomposition of A into the image of R and $\tilde{R}$.

Atkinson made another observation, formulating the following theorem, which describes a multiplicative decomposition for associative unital Rota–Baxter algebras.

Theorem 7.4. *Let A be an associative complete filtered unital Rota–Baxter algebra with Rota–Baxter map R. Assume X and Y in A to solve the equations*

$$X = 1_A + R(a\,X) \quad \text{and} \quad Y = 1_A + \tilde{R}(Y\,a), \tag{7.14}$$

for $a \in A^1$. Then we have the following factorization

$$Y(1_A - \theta a)X = 1_A, \quad \text{so that} \quad 1_A - \theta a = Y^{-1}X^{-1}. \tag{7.15}$$

For an idempotent Rota–Baxter map this factorization is unique.

Proof. First recall that a complete filtered algebra A has a decreasing filtration $\{\,A^n\,\}$ of sub-algebras

$$A = A^0 \supset A^1 \supset \cdots \supset A^n \supset \dots$$

such that $A^m A^n \subseteq A^{m+n}$ and $A \cong \varprojlim A/A^n$, that is, A is complete with respect to the topology determined by the $\{\,A^n\,\}$. Also, note that

$$R(a)\tilde{R}(b) = \tilde{R}(R(a)b) + R(a\tilde{R}(b)),$$

and similarly exchanging R and $\tilde{R}$. Then the product YX is given by

$$\begin{aligned} YX &= \big(1_A + \tilde{R}(Y\,a)\big)\big(1_A + R(a\,X)\big) = 1_A + R(a\,X) + \tilde{R}(Y\,a) + \tilde{R}(Y\,a)R(a\,X) \\ &= 1_A + \tilde{R}\Big(Y\,a\,\big(1_A + R(a\,X)\big)\Big) + R\Big(\big(1_A + \tilde{R}(Y\,a)\big)\,a\,X\Big) \\ &= 1_A + R(Y\,a\,X) + \tilde{R}(Y\,a\,X) = 1_A + \theta X\,a\,Y. \end{aligned}$$

Hence we obtain the factorization (7.15). Uniqueness for idempotent Rota–Baxter maps is easy to show [69]. $\square$

In summary, finite difference as well as q-difference equations play a role in important applications; thus it is useful to consider generalizations of the classical apparatus for solving differential equations. In the next section, by exploiting and complementing the CBHD development of the previous one, we make preparations to extend the work by Magnus on exponential solutions for non-autonomous differential equations to general Rota–Baxter maps, beyond the Riemann integral.

8 The Spitzer identities and the CBHD *recursion*

In the last section we mentioned Spitzer's classical identity as a motivation for Baxter's work. Now we spell out what that is. Spitzer's identity can be seen as a natural generalization of the solution of the simple initial value problem (1.1) on the commutative algebra A of continuous functions over $\mathbb{R}$,

$$\frac{df(t)}{dt} = a(t)f(t), \quad f(0) = 1, \quad a \in A. \tag{8.1}$$

This has, of course, a unique solution $f(t) = \exp\Big(\int_0^t a(u)\,du\Big)$. Transforming the differential equation into an integral equation by application of the Riemann integral $I : A \to A$ to (8.1),

$$f(t) = 1 + I(af)(t), \tag{8.2}$$

we arrive naturally at the not-quite-trivial identity

$$\exp\left(\int_0^t a(u)\,du\right) = \exp\big(I(a)(t)\big) = 1 + \sum_{n=1}^{\infty} \underbrace{I\Big(aI\big(a\cdots I(a)\cdots\big)\Big)}_{n\text{-times}}(t). \tag{8.3}$$

Taking into account the weight-zero Rota–Baxter rule (7.3) for I, the last identity follows simply from

$$\big(I(a)(t)\big)^n = n!\,\underbrace{I\Big(aI\big(a\cdots I(a)\cdots\big)\Big)}_{n\text{-times}}(t). \tag{8.4}$$

Let now (A, R) to be a *commutative* Rota–Baxter algebra of weight $\theta \neq 0$. We formulate Spitzer's finding in the ring of power series $A[[t]]$, which is a complete filtered algebra with the decreasing filtration given by the powers of t, $A^n := t^n A[[t]]$, $n \geq 0$. Notice that the power series algebra $A[[t]]$ with the operator $\mathcal{R} : A[[t]] \to A[[t]]$ acting on a series via R through the coefficients, $\mathcal{R}\left(\sum_{n\geq 0} a_n t^n\right) := \sum_{n\geq 0} R(a_n)t^n$, is Rota–Baxter as well. Then we have

Theorem 8.1. (Spitzer's identity) *Let (A, R) be a unital commutative Rota–Baxter algebra of weight $\theta \neq 0$. Then for $a \in A$,*

$$\exp\left(-R\left(\frac{\log(1-a\theta t)}{\theta}\right)\right) = \sum_{n=0}^{\infty}(t)^n \underbrace{R\Big(aR\big(a\cdots R(a)\cdots\big)\Big)}_{n\text{-times}} \tag{8.5}$$

in the ring of power series $A[[t]]$.

Analytic as well as algebraic proofs of this identity can be found in the literature, see for instance [61, 67]; and anyway it is a corollary of our work further below. Observe that $-\theta^{-1}\log(1-a\theta t) \xrightarrow{\theta\downarrow 0} at$. Thus indeed (8.5) generalizes (8.3).

Moreover, identity (8.4) generalizes to the *Bohnenblust–Spitzer formula* [61] of weight θ. This is as follows. Let (A, R) be a commutative Rota–Baxter algebra of weight θ and fix $s_1, \ldots, s_n \in A$, $n > 0$. Let S_n be the set of permutations of $\{1, \ldots, n\}$. Then

$$\sum_{\sigma\in S_n} R\Big(s_{\sigma(1)}R\big(s_{\sigma(2)}\cdots R(s_{\sigma(n)})\cdots\big)\Big) = \sum_{\mathcal{T}\in\Pi_n} \theta^{n-|\mathcal{T}|}\prod_{T\in\mathcal{T}}(|T|-1)!\,R\Big(\prod_{j\in T}s_j\Big), \tag{8.6}$$

Here $\mathcal{T}$ runs through all unordered set partitions of $\{1, \ldots, n\}$; by $|\mathcal{T}|$ we denote the number of blocks in $\mathcal{T}$; by $|T|$ the size of the particular block T. The Rota–Baxter relation itself appears as a particular case for $n = 2$. The weight $\theta = 0$ case reduces the sum over $\mathcal{T}$ to $|\mathcal{T}| = n$:

$$\sum_{\sigma\in S_n} R\Big(s_{\sigma(1)}R\big(s_{\sigma(2)}\cdots R(s_{\sigma(n)})\cdots\big)\Big) = \prod_{j=1}^{n} R\big(s_j\big). \tag{8.7}$$

Also, for $n > 0$ and $s_1 = \cdots = s_n = x$ we find in (8.6):

$$R\Big(xR\big(x\cdots R(x)\cdots\big)\Big) = \frac{1}{n!}\sum_{\mathcal{T}\in\Pi_n}\theta^{n-|\mathcal{T}|}\prod_{T\in\mathcal{T}}(|T|-1)!\,R\big(x^{|T|}\big). \tag{8.8}$$

Relation (8.6) follows from Spitzer's identity (8.5) by expanding the logarithm and the exponential on the left hand side, and comparing order by order the infinite set of identities in $A[[t]]$.

Spitzer's classical identity constitutes therefore an interesting generalization of the initial value problem (8.1), respectively the integral equation (8.2), to more general integration-like operators R, satisfying the identity (7.9). Again we refer the reader to [61, 64] for examples of such applications in the context of renormalization in perturbative quantum field theory, q-analogs of classical identities, classical integrable systems and multiple zeta values. Also, Atkinson's factorization Theorem 7.4 is obvious from Spitzer's identity. The right hand side of identity (8.5) is a solution to $X = 1_A + tR(a\,X)$ in $A[[t]]$ corresponding to the factorization of the element $1_A - \theta at$. (One ought to be careful here, since Spitzer's identity as well as (8.4) are only valid for commutative Rota–Baxter algebras of weight θ, whereas Atkinson's factorization result applies to general associative unital Rota–Baxter algebras.)

Let us adopt an even more general point of view. For functions with image in a noncommutative algebra, say $n \times n$ matrices with entries in $\mathbb{R}$, relation (8.3) is not valid anymore as a solution to (8.1); nor is identity (8.4) valid. From our present perspective, however, it does seem quite natural to approach the problem of finding a solution to the initial value problem, as well as relations (8.3) and (8.4), on a noncommutative function algebra A by looking for a generalization of Spitzer's identity to noncommutative unital associative Rota–Baxter algebras of weight θ. This latter problem was finally solved in [70, 71] —see also [69], where the reader may find more detail and earlier references. We will review briefly those results, prior to extend our findings by indicating a noncommutative generalization of the Bohnenblust–Spitzer formula.

We then take the first steps towards the *noncommutative Spitzer* identity. Let A be a complete filtered associative algebra. Bring in from Section 6 the Campbell–Baker–Hausdorff–Dynkin (CBHD) formula for the product of exponentials of two non-commuting objects x, y

$$\exp(x)\exp(y) = \exp\bigl(x + y + \mathrm{CBHD}(x, y)\bigr), \quad \text{where} \quad \sum_{m\geq 2} \Phi_m(x, y) =: \mathrm{CBHD}(x, y).$$

Now let $P : A \to A$ be *any* linear map preserving the filtration and $\tilde{P} = \theta\,\mathrm{id} - P$, with θ an arbitrary nonzero complex parameter. For $a \in A^1$, define the nonlinear map

$$\chi^{\theta,\tilde{P}}(a) = \lim_{n\to\infty} \chi^{\theta,\tilde{P}}_{(n)}(a)$$

where $\chi^{\theta,\tilde{P}}_{(n)}(a)$ is given by the so-called CBHD *recursion*,

$$\begin{aligned} \chi^{\theta,\tilde{P}}_{(0)}(a) &:= a, \\ \chi^{\theta,\tilde{P}}_{(n+1)}(a) &= a - \frac{1}{\theta}\,\mathrm{CBHD}\bigl(\tilde{P}(\chi^{\theta,\tilde{P}}_{(n)}(a)), P(\chi^{\theta,\tilde{P}}_{(n)}(a))\bigr), \end{aligned} \tag{8.9}$$

and where the limit is taken with respect to the topology given by the filtration. Then the map $\chi^{\theta,\tilde{P}} : A^1 \to A^1$ satisfies

$$\chi^{\theta,\tilde{P}}(a) = a - \frac{1}{\theta}\,\mathrm{CBHD}\bigl(\tilde{P}(\chi^{\theta,\tilde{P}}(a)), P(\chi^{\theta,\tilde{P}}(a))\bigr). \tag{8.10}$$

We call $\chi^{\theta,\tilde{P}}$ the CBHD recursion of weight θ, or just the θ-CBHD recursion. In the following we do not index the map $\chi^{\theta}(a) := \chi^{\theta,\tilde{P}}$ by the operator $\tilde{P}$ involved in its definition, when it is obvious from context. One readily observes that χ^{θ} reduces to the identity for commutative algebras.

The following theorem states a general decomposition on the algebra A implied by the CBHD recursion. It applies to associative as well as Lie algebras.

Theorem 8.2. *Let A be a complete filtered associative (or Lie) algebra with a linear, filtration preserving map $P : A \to A$ and $\tilde{P} := \theta\,\mathrm{id} - P$. For any $a \in A^1$, we have*

$$\exp(\theta a) = \exp\bigl(\tilde{P}(\chi^{\theta}(a))\bigr)\exp\bigl(P(\chi^{\theta}(a))\bigr). \tag{8.11}$$

Under the further hypothesis that the map P is idempotent (and $\theta = 1$), we find that for any $x \in 1_A + A^1$ there are unique $x_- \in \exp\bigl(\tilde{P}(A^1)\bigr)$ and $x_+ \in \exp\bigl(P(A^1)\bigr)$ such that $x = x_- x_+$.

For proofs we refer the reader to [69]. Using this factorization one simplifies (8.10) considerably.

Lemma 8.3. *Let A be a complete filtered algebra and $P : A \to A$ a linear map preserving the filtration, with $\tilde{P}$ as above. The map χ^θ in* (8.10) *solves the following recursion*

$$\chi^\theta(u) := u + \frac{1}{\theta}\,\mathrm{CBHD}\big(\theta u, -P(\chi^\theta(u))\big), \quad u \in A^1. \tag{8.12}$$

The convolution algebra $(\mathrm{Hom}(H, A), *)$, for H a connected graded commutative Hopf algebra, will also be complete filtered. We may immediately apply the above factorization theorem, giving rise to a factorization of the group $G_H(A)$ of A-valued characters, upon choosing any filtration-preserving linear map on $\mathrm{Hom}(H, A)$. In fact, we find for $\theta = 1$ in the definition of χ the following result.

Proposition 8.4. *Let A be a commutative algebra and H a connected graded commutative Hopf algebra. Let P be any filtration preserving linear map on* $\mathrm{Hom}(H, A)$*. Then we have for all $\phi \in G_H(A)$ and $Z := \log(\phi) \in g_H(A)$ the characters $\phi_-^{-1} := \exp\big(\tilde{P}(\chi(Z))\big)$ and $\phi_+ := \exp\big(P(\chi(Z))\big)$ such that*

$$\phi = \phi_-^{-1} * \phi_+. \tag{8.13}$$

If P is idempotent this decomposition is unique.

A natural question is whether one can find closed expressions for the map χ^θ. The answer is certainly affirmative in some non-trivial particular cases [69].

Corollary 8.5. *In the setting of the last proposition we find for the particular choice of $P = \pi_- : H \to H$ being the projection to the odd degree elements in H (hence $\theta = 1$)*

$$\chi(Z) = Z + \mathrm{CBHD}\Big(Z, -\pi_-(Z) - \tfrac{1}{2}\,\mathrm{CBHD}\big(Z, Z - \pi_-(Z)\big)\Big), \quad Z \in g_H(A).$$

Before proceeding, we must underline that the factorization is due solely to the map χ^θ; in fact, the map P —respectively $\tilde{P}$— involved in its definition has only to be linear and filtration preserving. The role played by this map is drastically altered when we assume it moreover to be *Rota–Baxter* of weight θ, on a complete filtered Rota–Baxter algebra. This we do next, to rederive and generalize the Magnus expansion. Also, it will soon become clear what χ^0 is. One of the aims of this paper is to attack the solution of the CBHD recursion when P is Rota–Baxter.

We noted earlier Atkinson's multiplicative decomposition of associative complete filtered Rota–Baxter algebras. Let from now on (A, R) denote one such, of weight $\theta \neq 0$. Observe the useful identity

$$\theta \prod_{i=1}^{n} R(x_i) = R\Big(\prod_{i=1}^{n} R(x_i) - \prod_{i=1}^{n} \tilde{R}(-x_i)\Big), \quad \text{for} \quad x_i \in A,\ i = 1, \ldots, n. \tag{8.14}$$

This comes from the Rota–Baxter relation (7.9). The case $n = 2$ simply returns it. The reader should check it with the help of the double Rota–Baxter product (7.10). In the following we consider $\chi^\theta := \chi^{\theta,\tilde{R}}$ on A^1. Using (8.14), for $\theta^{-1}\log(1_A - \theta a) =: u \in A^1$ one readily computes

$$\exp\big(-R(\chi^\theta(u))\big) = 1_A + \sum_{n>0} \frac{R(-\chi^\theta(u))^n}{n!}$$

$$= 1_A + \sum_{n>0} \frac{(-1)^n}{n!\theta} R\big(R(\chi^\theta(u))^n - (-\tilde{R}(\chi^\theta(u))^n)\big)$$
$$= 1_A + \frac{1}{\theta} R\Big(\exp\big(-R(\chi^\theta(u))\big) - \exp\big(\tilde{R}(\chi^\theta(u))\big)\Big) = 1_A + R\Big(a \exp\big(-R(\chi^\theta(u))\big)\Big).$$

In the last step we employed the factorization Theorem 8.2 corresponding to χ^θ. Therefore, on the one hand we have found that $X := \exp\big(-R(\chi^\theta(u))\big) \in 1_A + A^1$ solves $X = 1_A + R(a\,X)$, one of Atkinson's recursions in Theorem 7.4. On the other hand, a solution to this recursion follows from the iteration

$$X = 1_A + \sum_{n>0} \underbrace{R\big(aR(aR(a \cdots R(a))}_{n \text{ times}} \dots)\big). \tag{8.15}$$

Hence:

Theorem 8.6. *The natural generalization of Spitzer's identity* (8.5) *to noncommutative complete filtered Rota–Baxter algebras of weight $\theta \neq 0$ is given by*

$$\exp\bigg(-R\bigg(\chi^\theta\bigg(\frac{\log(1_A - \theta a)}{\theta}\bigg)\bigg)\bigg) = \sum_{n=0}^{\infty} \underbrace{R\big(aR(aR(a \cdots R(a))}_{n \text{ times}} \dots)\big), \tag{8.16}$$

for $a \in A^1$. Recall that χ^θ reduces to the identity for commutative algebras, yielding Spitzer's classical identity.

So far we have achieved the following. First we derived the general factorization Theorem 8.2 for complete filtered algebras, upon the choice of an arbitrary linear filtration preserving map. Specifying the latter to be Rota–Baxter of weight θ, that is, identity (8.14), we have been able to show that Atkinson's recursion equations in Theorem 7.4 have exponential solutions. It is now natural to ask whether the Bohnenblust–Spitzer formula (8.6) valid for weight-θ commutative Rota–Baxter algebras can be generalized to noncommutative ones. The answer is yes! We outline next this generalization, postponing detailed proof to the forthcoming [72], to keep this long work within bounds. First, by using the pre-Lie (7.12) and the double (7.10) Rota–Baxter products, we find

$$R\big(x_1 R(x_2)\big) + R\big(x_2 R(x_1)\big) = R(x_1)R(x_2) + R(x_2 \cdot_R x_1) = R(x_1 *_R x_2) + R(x_2 \cdot_R x_1).$$

Recall the relations (7.11). One may now check by a tedious calculation that

$$\begin{aligned}
&\sum_{\sigma \in S_3} R\Big(x_{\sigma_1} R\big(x_{\sigma_2} R(x_{\sigma_3})\big)\Big) = R(x_1 *_R x_2 *_R x_3) + R((x_2 \cdot_R x_1) *_R x_3) \\
&+R((x_3 \cdot_R x_1) *_R x_2) + R(x_3 \cdot_R (x_2 \cdot_R x_1)) + R(x_1 *_R (x_3 \cdot_R x_2)) + R(x_2 \cdot_R (x_3 \cdot_R x_1)) \\
&= R(x_1)R(x_2)R(x_3) + R(x_2 \cdot_R x_1)R(x_3) + R(x_3 \cdot_R x_1)R(x_2) \\
&+R(x_3 \cdot_R (x_2 \cdot_R x_1)) + R(x_1)R(x_3 \cdot_R x_2) + R(x_2 \cdot_R (x_3 \cdot_R x_1)).
\end{aligned} \tag{8.17}$$

We obtain special cases of the above when $x_1 = x_2 = x_3 = x$

$$\begin{aligned}
3!R\Big(xR\big(xR(x)\big)\Big) &= R(x *_R x *_R x) + 2R((x \cdot_R x) *_R x) + 2R(x \cdot_R (x \cdot_R x)) \\
+ R(x *_R (x \cdot_R x)) &= R(x)^3 + 2R(x \cdot_R x)R(x) + 2R(x \cdot_R (x \cdot_R x)) + R(x)R(x \cdot_R x).
\end{aligned}$$

Equation (8.17) is an instance of the following result, that seems to be new:

Theorem 8.7. *Let (A, R) be an associative Rota–Baxter algebra of weight θ. For $x_i \in A$, $i = 1, \ldots, n$, we have*

$$\sum_{\sigma \in S_n} R\Big(x_{\sigma_1} R\big(x_{\sigma_2} \ldots R(x_{\sigma_n}) \ldots\big)\Big) = \sum_{\sigma \in S_n} R\Big(x_{\sigma_1} \diamond_1 x_{\sigma_2} \diamond_2 \cdots \diamond_n x_{\sigma_n}\Big)$$

where

$$x_{\sigma_i} \diamond_i x_{\sigma_{i+1}} = \begin{cases} x_{\sigma_i} *_R x_{\sigma_{i+1}}, & \sigma_i < \min(\sigma_j | i < j) \\ x_{\sigma_i} \cdot_R x_{\sigma_{i+1}}, & \textit{otherwise}; \end{cases}$$

furthermore consecutive $\cdot_R$ products should be performed ¿from right to left, and always before the $*_R$ product.

This is the *noncommutative Bohnenblust–Spitzer* formula. Obviously, in the commutative case, i.e., when $a \cdot_R b = \theta ab$, we just recover the classical Bohnenblust–Spitzer identities (8.6) and (8.8). On the other hand, anticipating on coming sections, the case $\theta = 0$ reduces to Lam's factorization theorem [73], stated in the context of a weight-zero Rota–Baxter algebra of operator valued functions $(B, \int; 0)$. Let us adopt Lam's notation for the n-fold right bracketed pre-Lie product by

$$C_n^R := C_n^R(x) := R(x \cdot_R (x \cdot_R \ldots (x \cdot_R x) \ldots)). \tag{8.18}$$

Also, we introduce a notation for the so-called Rota–Baxter words:

$$(Rx)^{[n+1]} = R\big(x(Rx)^{[n]}\big), \tag{8.19}$$

with the convention that $(Rx)^{[0]} = 1$. Then we obtain the general expression

$$(Rx)^{[n]} = \sum_{l=1}^{n} \sum_{\substack{k_1, \ldots, k_l \in \mathbb{N}^* \\ k_1 + \cdots + k_l = n}} \frac{C_{k_1}^R \cdots C_{k_l}^R}{k_l(k_{l-1} + k_l) \cdots (k_1 + \cdots + k_l)}. \tag{8.20}$$

That is to say, we sum over the compositions of n. The simplest cases already examined now are written

$$\begin{aligned} 2!(Rx)^{[2]} := 2!R\big(xR(x)\big) &= (C_1^R)^2 + C_2^R, \\ 3!(Rx)^{[3]} := 3!R\Big(xR\big(xR(x)\big)\Big) &= (C_1^R)^3 + 2C_2^R C_1^R + C_1^R C_2^R + 2C_3^R. \end{aligned} \tag{8.21}$$

Later we make use of those expansions in relation with the Magnus expansion and the Dyson–Chen series. In fact, the left hand side of the above expressions are the second and third order terms in the path- or time-ordered expansion, in the context when the map R is the Riemann integral. We just generalized this to general-weight Rota–Baxter operators.

9 The zero-weight recursion

Let us come back to the CBHD recursion χ^θ. The question of the limit $\theta \downarrow 0$ becomes subtler than in the commutative case, due to the particular properties of relation (8.10). Now, in general we may write $\Phi(a, b) = \Phi_1(a, b) + \mathrm{CBHD}(a, b)$ as a sum

$$\Phi(a, b) = \sum_{n \geq 1} H_n(a, b),$$

where $H_n(a,b)$ is the part of $\Phi(a,b)$ which is homogenous of degree n with respect to a. For $n=1$ we have [34]:

$$H_1(a,b) = \frac{\operatorname{ad} b}{e^{\operatorname{ad} b} - 1_A}(a) = \frac{\operatorname{ad} b}{2}\Big(\coth \frac{\operatorname{ad} b}{2} - 1\Big)(a). \tag{9.1}$$

In the limit $\theta \downarrow 0$ all higher order terms $H_{n>1}$ vanish and from (8.12) we get a nonlinear map χ^0 inductively defined on A^1 by the formula

$$\chi^0(a) = \frac{\operatorname{ad} P\big(\chi^0(a)\big)}{\mathrm{e}^{\operatorname{ad} P(\chi^0(a))} - 1_A}(a) = \Big(1_A + \sum_{n>0} b_n \Big[\operatorname{ad} P\big(\chi^0(a)\big)\Big]^n\Big)(a). \tag{9.2}$$

We may call this the *weight-zero* CBHD *recursion.* The coefficients are $b_n := B_n/n!$ with B_n the Bernoulli numbers. For $n = 1, 2, 4$ we find $b_1 = -1/2, b_2 = 1/12$ and $b_4 = -1/720$. We have $b_3 = b_5 = \cdots = 0$. The first terms in (9.2) are then easily written down:

$$\begin{aligned} \chi^0(a) &= a - \frac{1}{2}[P(a),\, a] + \frac{1}{4}\big[P\big([P(a), a]\big),\, a\big] + \frac{1}{12}\big[P(a), [P(a), a]\big] \\ &- \frac{1}{24}P\Big(\big[P([P(a), [P(a), a]]), a\big] + \big[P(a), [P([P(a), a]), a]\big] + \big[[P([P(a), a]), [P(a), a]\big]\Big) \\ &- \frac{1}{8}P\Big(\big[P([P([P(a), a]), a]), a\big]\Big) + \cdots . \end{aligned} \tag{9.3}$$

We pause here to note that (9.1) is but an avatar of the formula

$$D(e^a) = e^a\,\frac{1 - e^{-\operatorname{ad} a}}{\operatorname{ad} a} Da, \quad \text{for } D \text{ a derivation;}$$

a noncommutative chain rule familiar from linear group theory. See [74, Chapter 1] for instance. Apparently this is due to F. Schur (1891), and was taken up later by Poincaré and Hausdorff. One may also consult the charming account of the determination of a local Lie group from its Lie algebra, using canonical coordinates of the first kind, in [75, Chapter 13]. The appearance of the Bernoulli numbers is always fascinating. The deep reason for it is that we are trying to express elements of the enveloping algebra in terms of the symmetric algebra.

Now suppose P is a weight-zero Rota–Baxter operator, denoted R henceforth. The noncommutative generalization of Spitzer's identity in the case of vanishing weight is captured in the following corollary.

Corollary 9.1. *Let (A, R) be a complete filtered Rota–Baxter algebra of weight zero. For a in A^1 the weight-zero* CBHD *recursion $\chi^0 : A^1 \to A^1$ is given by equation* (9.2)*:*

$$\chi^0(a) = \frac{\operatorname{ad} R\big(\chi^0(a)\big)}{\mathrm{e}^{\operatorname{ad} R(\chi^0(a))} - 1_A}(a).$$

Moreover:

1. *The equation $x = 1 + R(a\,x)$ has a unique solution $x = \exp\big(R(\chi^0(a))\big)$.*

2. *The equation $y = 1 - R(y\,a)$ has a unique solution $y = \exp\big(R(-\chi^0(a))\big)$.*

We will see pretty soon that the 0-CBHD recursion gives Magnus' expansion. The diagram (9.6) further below summarizes the foregoing relations, generalizing the simple initial value problem (8.1) in a twofold manner. First we go to the integral equation (8.2). Then we replace the Riemann integral by a general Rota–Baxter map and assume a noncommutative setting. That is, we start with a complete filtered noncommutative associative Rota–Baxter algebra (A, R) of nonzero weight θ in the appropriate field. The top of the diagram (9.6) contains the solution to the equation

$$X = 1_A + R(a\,X), \quad \text{for} \quad a \in A^1, \tag{9.4}$$

generalized to associative, otherwise arbitrary Rota–Baxter algebras (8.16),

$$X = \exp\biggl(-R\biggl(\chi^\theta\biggl(\frac{\log(1_A - \theta a)}{\theta}\biggr)\biggr)\biggr). \tag{9.5}$$

The θ-CBHD recursion χ^θ is given in (8.10). The left wing of (9.6) describes the case when first the weight θ goes to zero, hence reducing $\chi^\theta \to \chi^0$; see (9.2). This is the algebraic structure underlying Magnus's Ω series of the next section. Then, we let the algebra A become commutative, which reduces χ^0 to id. The right wing of diagram (9.6) just describes the alternative reduction, i.e., we first make the algebra commutative, which gives the classical Spitzer identity for nonzero weight commutative Rota–Baxter algebras, see (8.5). Then we take the limit $\theta \downarrow 0$.

$$\begin{tikzcd}
 & \underset{\theta\neq 0,\ \text{noncom}}{\exp\bigl(-R\bigl(\chi^\theta(\theta^{-1}\log(1-\theta a))\bigr)\bigr)} \arrow[dl, "\substack{\theta\downarrow 0\\ \text{noncom}}"'] \arrow[dr, "\substack{\theta\neq 0\\ \text{com}}"] \arrow[dd, "\substack{\text{com}\\ \theta\downarrow 0}"] & \\
\underset{\text{Magnus}}{\exp\bigl(R\bigl(\chi^0(a)\bigr)\bigr)} \arrow[dr, "\text{com}"] & & \underset{\text{Spitzer}}{\exp\bigl(-R\bigl(\theta^{-1}\log(1-\theta a)\bigr)\bigr)} \arrow[dl, "\theta\downarrow 0"'] \\
 & \underset{\theta=0,\ \text{com}}{\exp(R(a))} &
\end{tikzcd} \tag{9.6}$$

Both paths eventually arrive at the elementary fact that equation (9.4) is solved by a simple exponential in a commutative, weight-zero Rota–Baxter setting. We have succeeded in finding the general algebraic structure underlying the initial value problem for generalized integrals, that is, Rota–Baxter operators.

10 On Magnus' commutator series

It is high time for us to declare why we choose to deal with first order non-autonomous differential equations primarily via the Magnus expansion method. The latter has a somewhat chequered history. To attack the initial value problem of the type (1.1):

$$\frac{d}{dt}F(t) = a(t)F(t), \qquad F(0) = 1, \tag{10.1}$$

with F a matrix-valued function, say, Magnus proposed the exponential Ansatz

$$F(t) = \exp\bigl(\Omega[a](t)\bigr),$$

with $\Omega[a](0) = 0$. He found a series for $\Omega[a]$:

$$\Omega[a](t) = \sum_{n>0} \Omega_n[a](t), \tag{10.2}$$

in terms of multiple integrals of nested commutators, and provided a differential equation which in turn can be easily solved recursively for the terms $\Omega_n[a](t)$

$$\frac{d}{dt}\Omega[a](t) = \frac{\operatorname{ad}\Omega[a]}{e^{\operatorname{ad}\Omega[a]} - 1}(a)(t). \tag{10.3}$$

It is worth indicating that originally Magnus was motivated by Friedrich's theorem of our Section 5. We already mentioned, however, that one of the papers most influential on the subject [43] was written without knowledge of Magnus' paper. In the 1990's, several mathematicians interested in approximate integrators for differential equations developed the discipline of geometrical integration. Originally also unaware of Magnus' work, they derived anew Magnus' expansion. The point was to make sure that the approximate solutions evolve in the Lie group if $\xi(t)$ in (4.4) remains in the Lie algebra. This is *not* true of the iterative Dyson–Chen method —no finite truncation of the latter is the exact solution of any approximating system. By construction χ^0 respects the Lie algebra structure in (4.4), and thus truncations of the series are sure to remain in the Lie group. This is one reason why —in view also of the considerations in Appendix C— we give priority to the Magnus method. For geometrical integration, consult [76, 77].

Comparison with (9.2) settles the matter of the link between Magnus series and the CBHD recursion in the context of vanishing Rota–Baxter weight. Namely,

Corollary 10.1. *Let A be a function algebra over $\mathbb{R}_t$ with values in an operator algebra. Let R denote the indefinite Riemann integral operator. Magnus' Ω expansion is given by the formula*

$$\Omega[a](t) = R\big(\chi^0(a)\big)(t).$$

In conclusion, we could say that the θ-CBHD recursion (8.10) generalizes Magnus' expansion to general weight $\theta \neq 0$ Rota–Baxter operators R by replacing the weight-zero Riemann integral in $F = 1 + R\{aF\}$. Corollary 9.1 represents a more modest generalization, to zero-weight Rota–Baxter operators different ¿from the ordinary integral.

Let us write explicitly the first few terms of the Magnus expansion using (9.3), when R is the Riemann integral operator. The function $a = a(t)$ is defined over $\mathbb{R}$ and takes values in a noncommutative algebra, say of matrices of size $n \times n$. We obtain

$$\begin{aligned}
&R(a)(t) = \int_0^t a(t_1)\, dt_1, && (10.4)\\
&-\frac{1}{2}R\big([R(a),\, a]\big)(t) = \frac{1}{2}\int_0^t \int_0^{t_1} [a(t_1), a(t_2)]\, dt_2\, dt_1, &&\\
&\frac{1}{4}R\Big(\big[R\big([R(a),\, a]\big),\, a\big]\Big)(t) = \frac{1}{4}\int_0^t \int_0^{t_1} \int_0^{t_2} \big[[a(t_3),\, a(t_2)],\, a(t_1)\big]\, dt_3\, dt_2\, dt_1, &&\\
&\frac{1}{12}R\Big(\big[R(a), [R(a), a]\big]\Big)(t) = \frac{1}{12}\int_0^t \int_0^{t_1} \int_0^{t_1} \big[a(t_3), [a(t_2), a(t_1)]\big]\, dt_3\, dt_2\, dt_1. &&
\end{aligned}$$

This gives indeed the first terms of the expansion *precisely* in the form that Magnus derived it. However, in later works [43–45,78,79] the terms in the Magnus' expansion are presented as iterated commutator brackets of strictly time-ordered Riemann integrals. Especially in [43] Strichartz succeeded in giving a closed solution to Magnus' expansion —and hence to our recursion $\chi^{0,R}$ when R is the Riemann integral. With the notation of Proposition 6.2, he found

$$\Omega[a](t) = \sum_{n>0} \Omega_n[a](t), \quad \text{with} \tag{10.5}$$
$$\Omega_n[a](t) = \sum_{\sigma \in S_n} \frac{(-1)^{d(\sigma)}}{n^2\binom{n-1}{d(\sigma)}} \int_0^t \int_0^{t_1} \dots \int_0^{t_{n-1}} \big[[\dots[a(t_1), a(t_2)]\dots], a(t_n)\big]\, dt_n \dots dt_2\, dt_1.$$

This formula clearly points to the close relation between the CBHD expansion and Magnus' series, although the appearance of the number of descents is still 'unexplained'. However, it is not to everyone's taste. It is immediate from the formula that

$$\Omega_2[a](t) = \frac{1}{2} \int_0^t \int_0^{t_1} [a(t_1), a(t_2)]\, dt_2\, dt_1,$$

coincident with the second term in (10.4); and clear enough that

$$\Omega_3[a](t) = \frac{1}{6} \int_0^t \int_0^{t_1} \int_0^{t_2} \Big(\big[[a(t_1), a(t_2)], a(t_3)\big] - \big[[a(t_2), a(t_3)], a(t_1)\big] \Big)\, dt_3\, dt_2\, dt_1; \tag{10.6}$$

however the number of terms grows menacingly with $n!$, it is never evident when we will find cancellations, and one quickly concludes that the beauty of (10.5) hides its computational complexity. Nor is it entirely obvious, although of course it is true, that (10.6) coincides with the sum of the third and fourth terms in (10.4).

The best policy, in our opinion, is to invoke the alternative Dyson–Chen solution at this point. This attacks three problems: systematic writing of the Magnus series simplifies; the zero-weight recursion is solved; and the comparison between different expressions for the same terms is made easier.

11 Enter the Dyson–Chen series

The first order initial value problem (10.1), respectively the corresponding recursion

$$F(t) = 1 + R(aF)(t),$$

where R is the Riemann integral operator, possess a natural solution in terms of iteration, see (8.15). The resulting infinite series is called here Dyson–Chen integral. In the physics literature those series are often referred to as time-ordered exponentials or path-ordered integrals; their importance can hardly be overstated. To reflect such nomenclature in the notation, write

$$\mathrm{T}\, e^{\int_0^t a(t_1)\, dt_1} = \mathrm{T}\, e^{R(a)(t)} := 1 + \sum_{n>0} \underbrace{R\big(aR(aR(a \cdots R(a))}_{n \text{ times}} \dots)\big)(t).$$

The operator T implies the strict iteration of the integral corresponding to the 'time ordering'. A short presentation of Chen's work on this kind of integrals can be found in [80]; the findings of Magnus and Chen played a decisive role, especially for Rota and his followers. Directly from the group property of the flow, we have for the Dyson–Chen integral the factorization

$$\mathrm{T}\, e^{\int_0^t a(t_1)dt_1} = \mathrm{T}\, e^{\int_0^{t'} a(t_1)\, dt_1}\ \mathrm{T}\, e^{\int_{t'}^t a(t_1)\, dt_1},$$

giving rise to many identities of integrals and concatenation products of series, which we need not go into. This factorization might be compared with the quite different decomposition induced by the CBHD recursion (8.11). The major result of the theory is the following theorem.

Theorem 11.1. *The logarithm of a Chen series is a Lie series.*

The direct proof of this statement uses Hopf algebra, to wit, the shuffle product algebra of our Section 5. It is just a matter of verifying Ree's condition (5.2) inductively. In our present context, the Dyson–Chen expansional is the solution to Atkinson's recursion (9.4), and the theorem scarcely needs justification.

Simply by taking the logarithm in

$$\exp(\Omega[a](t)) = \mathrm{T}\, e^{R(a)(t)},$$

we obtain

$$\Omega_n[a] = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} \sum_{\substack{l_1,\ldots,l_k\in\mathbb{N}^* \\ l_1+\cdots+l_k=n}} (Ra)^{[l_1]}\ldots(Ra)^{[l_k]}. \tag{11.1}$$

This was of course known to the practitioners —see [79, 81] and references there. It is derived in [3] by use of the Faà di Bruno Hopf algebra. Inverting these relations, one finds the $(Ra)^{[n+1]}$'s in terms of the $\Omega_m[a]$'s

$$(Ra)^{[n]} = \sum_{k=1}^{n} \frac{1}{k!} \sum_{\substack{l_1,\ldots,l_k\in\mathbb{N}^* \\ l_1+\cdots+l_k=n}} \Omega_{l_1}[a]\ldots\Omega_{l_k}[a]. \tag{11.2}$$

The first examples are:

$$\begin{aligned} 2!(Ra)^{[2]} &= \Omega_1^2[a] + 2\Omega_2[a], \\ 3!(Ra)^{[3]} &= \Omega_1^3[a] + 3\big(\Omega_1^2[a]\Omega_2[a] + \Omega_2[a]\Omega_1^2[a]\big) + 6\Omega_3[a]. \end{aligned}$$

that might be compared with (8.21); of course $Ra = \Omega_1[a] = C_1^R$ in the occasion. Now, both sets of equations (11.1) and (11.2) simply describe how to link Magnus' expansion to the Dyson–Chen expansional. They purely follow from the Rota–Baxter relation as well as the CBHD formula. Therefore they are valid for *any* weight-zero Rota–Baxter operator R.

By inverting the Rota–Baxter map, we solve moreover the zero-weight CBHD recursion:

$$\chi_n^0(a) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} \sum_{\substack{l_1,l_2,\ldots,l_k\in\mathbb{N}^* \\ l_1+l_2+\cdots+l_k=n}} \big(a(Ra)^{[l_1-1]}(Ra)^{[l_2]}\cdots(Ra)^{[l_k]}+$$

$$+ (Ra)^{[l_1]}a(Ra)^{[l_2-1]}\cdots(Ra)^{[l_k]} + \cdots + (Ra)^{[l_1]}(Ra)^{[l_2]}\cdots a(Ra)^{[l_k-1]}\big).$$

Next we are set to give an alternative formula to (10.5), keeping left-to-right bracketing. This is better explained by way of example. Bring in Heaviside's step function,

$$\Theta_{1,2}(t_1,t_2) := \Theta(t_1-t_2) := \begin{cases} 1, & \text{if } t_1-t_2>0,\\ 0, & \text{otherwise;}\end{cases}$$

with its help, iterated Riemann integrals can be rewritten

$$R\big(aR(b)\big)(t) = \int_0^t a(t_1)\int_0^{t_1} b(t_2)\,dt_2\,dt_1 = \int_0^t\int_0^t \Theta(t_1-t_2)a(t_1)b(t_2)\,dt_2\,dt_1. \tag{11.3}$$

More generally,

$$\Theta_{i,j}(t_1,t_2,\dots,t_n) := \Theta(t_i-t_j), \quad \text{for } 1\le i,j\le n,$$

and we can write

$$\mathrm{T}\,e^{\int_0^t a(t_1)dt_1} = 1 + \int_0^t a(t_1)\,dt_1 + \sum_{n=2}^{\infty}\int_0^t\cdots\int_0^t \Theta_{1,2}\cdots\Theta_{n-1,n}\,a(t_1)\cdots a(t_n)\,dt_n\dots dt_1.$$

For instance, for the third term of the Magnus series, applying (11.1),

$$\Omega_3[a] = \int_0^t\int_0^t\int_0^t \big(\Theta_{1,2}\Theta_{2,3} - \tfrac12\Theta_{1,2} - \tfrac12\Theta_{2,3} + \tfrac13\big)a(t_1)a(t_2)a(t_3)\,dt_3\,dt_2\,dt_1.$$

Now, we know —if only from theorem (11.1)— this is a Lie element, so we can apply at once the Dynkin operator to rewrite it with nested commutators:

$$\Omega_3[a] = \int_0^t\int_0^t\int_0^t \big(\Theta_{1,2}\Theta_{2,3} - \tfrac12\Theta_{1,2} - \tfrac12\Theta_{2,3} + \tfrac13\big)[[a(t_1),a(t_2)],a(t_3)]\,dt_3\,dt_2\,dt_1.$$

We see now that the last term actually does not contribute to the integral. With very little work, just using $\Theta_{1,2}+\Theta_{2,1} = \Theta_{2,3}+\Theta_{3,2} = 1$, one recovers (10.6). An explicit formula for all terms along these lines, fully equivalent to, but simpler to work with, than Strichartz's, is easily obtained [45]; we do not bother to write it. We must avow, however, that we do not see a way to write terms like the third one in the integral above as a combination of iterations and products of the R operators; thus we must conclude that formulae like (10.5) and (10.6) are only valid for the Riemann integral.

For general zero-weight Rota–Baxter operators we may fall back on (11.1). Magnus himself did not use any property of the map R beyond integration-by-parts, and only presented the expansion in a form equivalent to (9.3). Of course, even using purely the weight-zero Rota–Baxter relation, there are many equivalent ways of writing the same. For instance, simply by Proposition 7.1, one finds that the term at third order in Magnus' expansion (10.4) is rewritten

$$\frac13 R\Big(\big[R\big([R(a),a]\big),a\big]\Big)(t) - \frac1{12}\big[R\big([R(a),a]\big),R(a)\big](t).$$

It is worthwhile to mention that Iserles and Norsett use binary rooted trees to achieve a better understanding of Magnus' expansion [76, 77].

We have long taken the algebraic tack. But what about convergence of the Magnus series? Note that Dyson–Chen series converge absolutely for all t if a is bounded, and this is why they are preferred in quantum field theory; however this good property is not transmitted in general to Magnus series via (11.1), as there is an infinite resummation involved. Excellent bounds at small t have been found recently [82] for matrix systems. Strichartz linearizes arbitrary initial-value problems, for which we cannot expect convergence in general in the smooth category; but he does not fail to observe that Magnus' expansion has especially good properties for Lie–Scheffers systems [43, Section 3]. This is because the closing of the involved vector fields to a finite-dimensional Lie algebra sharply improves the estimates. Furthermore, for those systems Magnus' exponential can be interpreted as the exponential map of Lie theory.

12 Towards solving the θ-weight recursions

Let us now come back to Proposition 8.7 and take the first steps in going from the Dyson–Chen series to the θ-weight CBHD recursion. This looks somewhat hard; but recall that Lam found, in the context of the Riemann integral, another way to relate the terms in the Dyson–Chen series to those in the Magnus expansion —consult [73, 83]. In fact, Lam's findings are true in a much more general sense, i.e., for general weight Rota–Baxter algebras, as we will indicate here. The attentive reader will remember the weight-zero pre-Lie product (7.12), that allows for the following way of writing the weight zero CBHD recursion $\chi^0(a)$, see (9.3):

$$\begin{aligned}\chi^0(a) &= a + \frac{1}{2} a \cdot_R a + \Big(\frac{1}{4}\big(a \cdot_R (a \cdot_R a)\big) + \frac{1}{12}\big((a \cdot_R a) \cdot_R a\big)\Big) \\ &\quad + \frac{1}{24} R\Big(a \cdot_R \big((a \cdot_R a) \cdot_R a\big) + \big(a \cdot_R (a \cdot_R a)\big) \cdot_R a + (a \cdot_R a) \cdot_R (a \cdot_R a)\Big) \\ &\quad + \frac{1}{8} R\Big(a \cdot_R \big(a \cdot_R (a \cdot_R a)\big)\Big) + \cdots . \end{aligned} \tag{12.1}$$

This contains in germ the main idea. Remember (8.18) in terms of the (double and) pre-Lie Rota–Baxter product. Lam made an exponential Ansatz

$$\sum_{n\geq 0} R(a)^{[n]} = \exp\Big(\sum_{m>0} K_m(a)\Big)$$

and derived the following formulae for the K_i's in terms of $C^R_1(a), \ldots, C^R_i(a)$:

$$\begin{aligned} K_1(a) &= C^R_1(a), \quad K_2(a) = \frac{1}{2} C^R_2(a), \quad K_3(a) = \frac{1}{3} C^R_3(a) + \frac{1}{12}[C^R_2(a), C^R_1(a)], \\ K_4(a) &= \frac{1}{4} C^R_4(a) + \frac{1}{12}[C^R_3(a), C^R_1(a)], \ldots \end{aligned}$$

The weight-θ Rota–Baxter relation enters at the level of identity (8.20), hence implying the particular form of the K_i's. This naturally demands a comparison with the CBHD recursion, respectively the generalized Spitzer identity.

Theorem 12.1. *Let (A, R) be an associative Rota–Baxter algebra of weight θ. Then for $K_i = K_i(C^R_1(a), \ldots, C^R_i(a); \theta)$ we have*

$$\sum_{i>0} K_i t^i = -R\big(\chi^\theta\big(\theta^{-1}\log(1_A - \theta a t)\big)\big).$$

Hence, finding a formula for the K_i's gives a solution, in the sense of a closed expression, to the CBHD recursion χ^θ, which follows from the Rota–Baxter relation. A full proof of this statement lies beyond the scope of this work and will be provided elsewhere [72]. In the context of Hopf and Rota–Baxter theory, it is the generalization of the shuffle relation to the quasi-shuffle (or mixed-shuffle) identity [84] underlying the algebraic structure encoded in the Rota–Baxter relation of nonzero weight, which generalizes the integration-by-parts rule (7.3) corresponding to the shuffle relation.

Of course, when $\theta = 0$ Lam's K_i's are just the Magnus Ω_i's. These are expressed as sums of commutators of *right-to-left* bracketed integrals, when R is the Riemann operator. This turns out to be the most efficient method for the expansion, as well. For instance, the expression of K_5 contains just six terms, whereas Ω_5 is written usually with 22 terms [83].

We close this section with a simple but striking observation flowing ¿from the last theorem. Defining $u(at) := \theta^{-1}\log(1_A - \theta at)$, we recover $-\chi^\theta(u(at))$ from $\chi^0(at)$, that is, from (12.1), simply by using the weight-θ pre-Lie product 7.12). A full proof of this statement will be given elsewhere. But we show this here up to third order. Using $\theta^{-1}\log(1_A - \theta at) = -\sum_{n>0}\frac{\theta^{n-1}}{n(at)^n}$, we find for

$$-\chi^\theta(u(at)) = at - \sum_{n>0}\chi^\theta_n(u(a))t^{n+1}$$

the following

$$\begin{aligned}
\chi^\theta_{(1)}(u(a)) &= \frac{1}{2}\theta a^2 - \frac{1}{2}[R(a), a],\\
\chi^\theta_{(2)}(u(a)) &= \frac{1}{3}\theta^2 a^3 - \frac{1}{4}\theta\big([R(a^2), a] + [R(a), a^2]\big) + \frac{1}{4}\big[a, R\big([a, R(a)]\big)\big]\\
&\quad + \frac{1}{12}\Big(\big[[a, R(a)], R(a)\big] - \theta\big[a, [a, R(a)]\big]\Big).
\end{aligned}$$

Let us go back to (12.1) and use the pre-Lie product $a \cdot_R b := [a, R(b)] + \theta ba$ of (7.12). We obtain at second order

$$\frac{1}{2}a \cdot_R a = \frac{1}{2}[a, R(a)] + \frac{1}{2}\theta a^2.$$

At third order we calculate:

$$\begin{aligned}
&\frac{1}{4}\big(a \cdot_R (a \cdot_R a)\big) + \frac{1}{12}\big((a \cdot_R a) \cdot_R a\big)\\
&= +\frac{1}{4}[a, R(a \cdot_R a)] + \frac{1}{4}\theta(a \cdot_R a)a + \frac{1}{12}[(a \cdot_R a), R(a)] + \frac{1}{12}\theta a(a \cdot_R a)\\
&= \frac{1}{4}\big[a, R\big([a, R(a)]\big)\big] + \frac{1}{4}\theta[a, R(a^2)] + \frac{1}{4}\theta[a, R(a)]a + \frac{1}{4}\theta^2 a^3\\
&\quad + \frac{1}{12}\big[[a, R(a)], R(a)\big] + \frac{1}{12}\theta[a^2, R(a)] + \frac{1}{12}a[a, R(a)] + \frac{1}{12}\theta a^3\\
&= \frac{1}{3}\theta^2 a^3 + \frac{1}{12}\big[[a, R(a)], R(a)\big] + \frac{1}{4}\big[a, R\big([a, R(a)]\big)\big] + \frac{1}{4}\theta\big([a, R(a^2)] + [a^2, R(a)]\big)\\
&\quad + \frac{1}{4}\theta[a, R(a)]a - \frac{1}{6}\theta[a^2, R(a)] + \frac{1}{12}\theta a[a, R(a)]\\
&= \frac{1}{3}\theta^2 a^3 + \frac{1}{4}\theta\big([a, R(a^2)] + [a^2, R(a)]\big) + \frac{1}{4}\big[a, R\big([a, R(a)]\big)\big]
\end{aligned}$$

$$+\frac{1}{12}\big[[a,R(a)],R(a)\big]-\frac{1}{12}\theta\big[a,[a,R(a)]\big].$$

Earlier in Section 10 we have seen how the Magnus expansion naturally follows from the CBHD recursion in the limit $\theta \downarrow 0$. In turn we see here the advantage of reformulating Magnus' expansion in terms of the Rota–Baxter pre-Lie product of weight θ yielding the CBHD recursion.

13 Conclusion and outlook

Our purpose in this paper was twofold. Starting from the innocent-looking dynamical system (1.1) —of classical Lie–Scheffers type when G is an ordinary Lie group— we sought to reformulate it in Hopf algebraic terms, thus being led to generalized derivation and integration (Rota–Baxter) operators. Whereby we show that two of the three main ordinary strategies to attack non-autonomous linear differential equations (linked respectively to the names of Magnus and Dyson–Chen) still make sense in the broader context. In particular, the noncommutative version of the Bohnenblust–Spitzer identity has been found, and we blaze a trail to solve the nonlinear recursion introduced earlier by one of us in relation with the noncommutative Spitzer formula.

There is no doubt that the product integral method to attack (1.1), often linked with the name of Fer [85], is also susceptible to our kind of algebraic reinterpretation and generalization. However, with a heavy heart, we leave this for a later occasion: the present paper is already long enough.

Needless to say, the programmatic purpose of this work was to propagandize the Hopf algebra approach to differential equations. The lure of presenting classical subjects under a new light explains why we spent much space on a smooth transition from the standard to a Hopf-flavoured view of dynamical systems; and indeed this article became a powerful spur to revisit the traditional proof of the Lie–Scheffers theorem, and plug its gaps [17]. On the other hand, many of our findings and procedures will surely not raise an eyebrow of people working in sophisticated methods for control theory —on which we confess no expertise. There is, at any rate, plenty left to do. Avenues open for possible research include:

- The Cariñena–Ramos' approach to Lie–Scheffers systems, based on connections, should be recast in the noncommutative mould, in the light of [20] and [30].
- To relate and compare the action algebroid approach to group & Lie algebra actions with the Hopf algebra approach.
- Investigation of the product integral method.
- Further exploration of the theory of Rota–Baxter operators as natural inverses to skewderivations; that is, developing Rota's proposal of an algebraic theory of integration.
- Definitive clarification of the noncommutative Spitzer formula and the noncommutative Bohnenblust–Spitzer identity in the light of Lam's expansion.
- The bridge to control theory and chronological products, via Loday's dendriform algebras in particular, should be enlarged and strengthened. In this respect, Lie–Butcher theory [86, 87] shows great promise.

A Précis on group actions

Definition 4. A (left) action of a Lie group G on a manifold M is a homomorphism Φ of G into Diff M. For $x \in M$, and $g \in G$ we denote

$$\Phi_g := \Phi(g) \quad \text{and} \quad \Phi(g, x) := \Phi_g x.$$

A right action is just an antihomomorphism of G into Diff M. The orbits of Φ are the subsets of M of the form $\Phi(G, x)$ for a fixed $x \in M$; they are homogeneous manifolds, on which the action is transitive. We will call Φ_x the map from G to M defined by $g \mapsto \Phi(g, x)$. Recall that a flow is an action of $\mathbb{R}$ on M. When Φ with the indicated properties is given, we say M is a G-manifold. A Lie group action is proper if given any pair K, L of compacts subsets of M, the set $\{ g \in G : gK \cap L \neq \emptyset \}$ is compact. The stabilizer or isotropy subgroups are then compact. Proper actions, in particular compact group actions of general Lie groups, have good properties: for instance the orbits of a proper action are closed submanifolds of M [88]. An action is faithful (or effective, or essential) when the map $g \to \Phi_g$ is injective; if the kernel of this map is discrete, we say the action is almost faithful.

A good reference for Lie group actions is [2, Chapter 4]. As for the examples, any Lie group G acts on itself by left and right translations $L_g, R_g : G \to G$ respectively given for each $g \in G$ by

$$g' \mapsto gg', \qquad g' \mapsto g'g.$$

The inverse diffeomorphisms are $L_g^{-1} = L_{g^{-1}}$ and $R_g^{-1} = R_{g^{-1}}$. This action is free and transitive. Also G acts on itself by conjugation:

$$g' \mapsto gg'g^{-1} =: \mathrm{Ad}(g)g'.$$

This action is neither free nor transitive; it is almost faithful iff the centre of G is discrete.

Definition 5. Suppose G acts both on N by Φ^N and on M by Φ^M. A smooth map $f : N \to M$ between these manifolds is *equivariant* (with respect to the actions) if $f \circ \Phi_g^N = \Phi_g^M \circ f$ for each $g \in G$. The maps $\Phi_x : G \to M$, where Φ is a left (right) action are equivariant for all x, with respect to the left (right) action of G on itself and Φ:

$$\Phi_x \circ L_g = \Phi_g \circ \Phi_x \quad \text{or} \quad \Phi_x \circ R_g = \Phi_g \circ \Phi_x, \tag{A.1}$$

as the case may be.

If G acts on M, then G also acts on TM by

$$(g, v_x) \mapsto (\Phi_g x, T_x\Phi_g v_x) =: \Phi^T(g, v_x), \quad \text{for } v_x \in T_xM.$$

When Φ is described in local coordinates, say by

$$\Phi_i(g, x) = h_i(g, x^1, \dots, x^n), \quad \text{then} \quad T_x\Phi_g v_x = \sum_{j=1}^n \frac{\partial h_i}{\partial x^j}(g, x^1, \dots, x^n) v_x^j.$$

Clearly the map $v_x \mapsto \Phi^T(g, v_x)$ from T_xM into $T_{\Phi(g,x)}M$ is linear and the canonical projection $\tau_M : TM \to M$ is equivariant with respect to these actions: $\tau_M\big(\Phi^T(g, v_x)\big) = \Phi\big(g, \tau_M(v_x)\big)$. We then say that Φ is *equilinear* [28]. For vector fields, then, there is the action:

$$(g, X) \mapsto T\Phi_g \circ X \circ \Phi_{g^{-1}}. \tag{A.2}$$

Corresponding to group translations we have then equilinear left and right actions of G on TG; as well as actions on $\mathfrak{X}(G)$. In view of (A.2), a vector field X^L on G is left invariant if for all $g \in G$, $X^L \circ L_g = TL_g \circ X^L$; this means that X^L is L_g-related to itself for all $g \in G$. Therefore the left invariant vector fields constitute a Lie subalgebra $\mathfrak{X}^L(G) =: \mathfrak{g}_L$ of $\mathfrak{X}(G)$. Replacing L_g by R_g we obtain right invariant vector fields $X^R \in \mathfrak{X}^R(G)$ and a Lie subalgebra $\mathfrak{X}^R(G) =: \mathfrak{g}_R$. In particular, X^L, X^R are determined by their values in the neutral element:

$$X^L(g) = T_1 L_g X^L(L_g^{-1} g) = T_1 L_g X^L(1_G); \quad \text{similarly} \quad X^R(g) = T_1 R_g X^R(1_G); \tag{A.3}$$

for typographical simplicity we write T_1 instead of T_{1_G}. The dimension of $\mathfrak{X}^L(G)$ or of $\mathfrak{X}^R(G)$ is thus that of the group. We denote by X^L_ξ, X^R_ξ the left invariant, respectively right invariant, vector field associated to $\xi \in T_1 G$. The (complete) flow of X^L_ξ is $(t, g) \mapsto g \exp(tX^L_\xi)$ and the flow of X^R_ξ is $(t, g) \mapsto \exp(tX^R_\xi) g$.

We remark that $\mathfrak{g}_L$ is the commutant of $\mathfrak{g}_R$ in $\mathfrak{X}(G)$, and vice versa. For instance, thinking of the affine group of orientation-preserving transformations of the line as a neighbourhood of $(1, 0)$ with the multiplication rule:

$$(x^1, x^2) \cdot (y^1, y^2) = (x^1 y^1, x^1 y^2 + x^2),$$

then a basis for left (respectively right) invariant vector fields is

$$(X^L_1, X^L_2) := (x^1 \partial_1, x^1 \partial_2); \quad \text{respectively} \quad (X^R_1, X^R_2) := (x^1 \partial_1 + x^2 \partial_2, \partial_2).$$

With our Lie bracket, by the way: $[X^R_1, X^R_2] = X^R_2$. It is an easy exercise to check that if $a_1(x^1, x^2)\partial_1 + a_2(x^1, x^2)\partial_2$ commutes with X^L_1, X^L_2, then it is a linear combination of X^R_1, X^R_2 with scalar coefficients.

Consider the tangent map $T\imath : TG \to TG$ lifting the inversion diffeomorphism $\imath : g \mapsto g^{-1}$ on the base; it carries left invariant vector fields into right invariant ones. The vector fields $T\imath \circ X^L_\xi$ and $-X^R_\xi \circ \imath$ along $\imath$ coincide, that is, X^L_ξ is $\imath$-projectable on $-X^R_\xi$. This simply because $\big(g^{-1} \exp(tX^L_\xi)\big)^{-1} = \exp(-tX^R_\xi) g$. Therefore $[X^L_\xi, X^L_\eta]$ projects into $[X^R_\eta, X^R_\xi]$.

Now, TG is itself a group, with product $T\mu$ lifted from the product $\mu : G \times G \to G$. The short exact sequence (where $T_1 G$ is the additive group of this tangent linear space)

$$0 \to T_1 G \to TG \to G \to 1$$

splits, which means $TG \sim T_1 G \rtimes G$, with $T_1 G$ embedded in TG as a normal subgroup. In particular TG is a trivial vector bundle. We have in TG:

$$g v_{g'} = TL_g v_{g'}; \qquad v_{g'} g = TR_g v_{g'}. \tag{A.4}$$

Clearly, the action of G on $T_1 G$ is just $\mathrm{Ad}^T_{1_G}$. Henceforth we write Ad for this adjoint action of G on $T_1 G$. A Lie bracket can now be defined directly on $T_1 G$ by $[\xi, \eta] := \mathrm{ad}(\xi)\eta := T_1 \mathrm{Ad}(\xi)\eta$. One could also transfer to $T_1 G$ the Lie algebra structure from $\mathfrak{X}^L(G)$ or $\mathfrak{X}^R(G)$, say $[\xi, \eta] := [X^R_\xi, X^R_\eta](1)$. That these and other natural definitions amount to the same is standard fare [89, Appendix III]. The space $T_1 G$ with any of these equivalent structures is what people call the tangent (Lie) algebra $\mathfrak{g}$ of G.

Definition 6. A left (right) Lie algebra (infinitesimal) action λ on M is a Lie algebra homomorphism (antihomomorphism) $\mathfrak{g} \ni \xi \mapsto \lambda_\xi \in \mathfrak{X}(M)$; we say M is a $\mathfrak{g}$-manifold. The action is said *transitive* at x when the $\lambda_\xi(x)$ span T_xM. It is furthermore *primitive* when the stabilizer $\mathfrak{g}_x$ is a maximal subalgebra; these concepts are analogous to the case of Lie group actions. When the action is transitive at all points of M, we say infinitesimally transitive. Given $\lambda : \mathfrak{g} \to \mathfrak{X}(M)$, if $\lambda_\mathfrak{g}$ is made up of complete vector fields (in particular when M is compact, guaranteeing completeness of all vector fields) and G is the simply connected Lie group with Lie algebra $\mathfrak{g}$, then there is a unique $\Phi : G \to \operatorname{Diff} M$ such that $T_1\Phi = \lambda$. This lifting to a group action always exists locally. We remark as well that our choice of sign for the bracket of vector fields insures that the derivative $T_1\Phi$ of a left action is a left action.

For the infinitesimal description of actions, the following notion is essential.

Definition 7. Let Φ denotes an action of G on M. For $\xi \in \mathfrak{g}$, the map $(t, x) \mapsto \Phi(\exp t\xi, x)$ is a flow on M. The *fundamental vector field* or *infinitesimal generator* ξ_M^Φ of Φ corresponding to ξ is the vector field

$$\xi_M^\Phi(x) := \frac{d}{dt}\bigg|_{t=0} \Phi(\exp t\xi, x) = T_1\Phi_x(\xi)\,. \tag{A.5}$$

The superscript Φ is omitted in the notation when the action is clear in the context. The image of $\mathfrak{g}$ under $T_1\Phi_x$ is the tangent bundle $T(G \cdot x)$ of the Φ-orbit. The corresponding differential operator is given by

$$\xi_M^\Phi f(x) := \frac{d}{dt}\bigg|_{t=0} f\big(\Phi(\exp t\xi, x)\big).$$

The anchor map $\xi \mapsto \xi_M^\Phi$ from the tangent algebra $\mathfrak{g}$ to $\mathfrak{X}(M)$ constitutes a Lie–Rinehart algebra; the corresponding Lie algebroid will be transitive when the action of $\mathfrak{g}$ on M is infinitesimally transitive.

For example, when Φ is $L_g : G \to G$, we know that the corresponding flow is $(t, g') \mapsto R_{g'} \exp t\xi$. Therefore

$$\xi_G(g') = T_1R_{g'}\xi = X_\xi^R(g'), \tag{A.6}$$

the *right* invariant vector field associated to ξ. By the same token $\xi_G^R(g) = X_\xi^L(g)$.

If M is a G-manifold, the flow of ξ_M is given by $\Phi_{\exp t\xi}$. Indeed,

$$\begin{aligned}\frac{d}{dt}\Phi(\exp t\xi, x) &= \frac{d}{ds}\bigg|_{s=0} \Phi(\exp(s+t)\xi, x) \\ &= \frac{d}{ds}\bigg|_{s=0} \Phi(\exp s\xi, x) \circ \Phi(\exp t\xi, x) = \xi_M \circ \Phi_{\exp t\xi}(x).\end{aligned}$$

As a consequence ξ_M is complete. The reader will have little difficulty in verifying the following

Proposition A.1. *Let N, M be G-manifolds with respective actions Φ^N, Φ^M, and $f : N \to M$ a smooth map equivariant with respect to these actions; then $\xi_N \sim_f \xi_M$, that is $Tf \circ \xi_N = \xi_M \circ f$. More precisely, $\xi_N \sim_f \xi_M$ iff the flows verify*

$$f \circ \Phi^N_{\exp(t\xi)} = \Phi^M_{\exp(t\xi)} \circ f.$$

Proposition A.2. *For every $\xi, \eta \in \mathfrak{g}$ we have*

$$[\xi_M, \eta_M] = [\xi, \eta]_M.$$

In other words: $\xi \mapsto \xi_M$ is a left Lie algebra action.

Proof. A simple calculation gives

$$(\mathrm{Ad}_g\, \xi)_M = T\Phi_{g^{-1}} \xi_M.$$

We obtain the result immediately by differentiation. Our unconventional choice of sign for the Lie bracket of vector fields avoids the obnoxious minus signs of the usual treatments. □

The action Φ of G on M lifts naturally to representations of G on the various linear spaces associated with M —for instance to representations on spaces of sections of vector bundles [90] or on morphisms of vector bundles. We will limit ourselves to some simple cases, needed in the main text. For $f \in \mathcal{F}(M)$, we consider $(g \cdot f)(x) := f\big(\Phi(g^{-1}, x)\big)$; then for $T \in \mathcal{F}'(M)$ and for $D \in \mathbb{D}(M)$:

$$\langle g \cdot T, f \rangle := \langle T, g^{-1} \cdot f \rangle, \quad \text{respectively} \quad (g \cdot D)f := g \cdot D(g^{-1} \cdot f).$$

Invariant functions, distributions and differential operators are defined in the obvious way.

B Differential equations on homogeneous spaces

The problem of solving non-autonomous differential equations on *homogeneous spaces* of Lie groups is intimately linked to Lie–Scheffers theory: given an arbitrary Lie group G and an action of it on a manifold M, for most purposes one can restrict oneself to the orbits of the action, that is, the points of M/G; these are (immersed) submanifolds of M of the form G/G_x, with $G_x := \{\, g \in G : \Phi(g, x) = x \,\}$ the stabilizer of a point x of the orbit, a closed Lie subgroup of G. From this perspective, Lie–Scheffers systems are precisely those that can be rewritten in the form

$$\dot{x}(t) = \lambda_{\xi(t)}\big(x(t)\big), \tag{B.1}$$

where $A : \mathbb{R} \to \mathfrak{g}$ is a curve on the Lie algebra $\mathfrak{g}$ and λ denotes an infinitesimal action. If $\lambda = T\Phi$ for some action Φ of G on M and $g(t)$ solves the initial value problem (2.9):

$$\dot{g}(t) = \xi_G(t, g(t)); \qquad g(0) = 1_G, \tag{B.2}$$

then the solution of (B.1) with initial condition $x(0) = x_0$ is given by the integrated action: $x(t) = \Phi(g(t), x_0)$. At this point we again advise the reader to consult [17].

In practice we consider transitive actions on $M \equiv G/G_x$. Suppose that $x_{(1)}$ is a particular solution of (B.1) satisfying $x(0) = x_0$. Let $g_1 \in \mathrm{Map}(\mathbb{R}_t, G)$ such that $x_{(1)}(t) = \Phi(g_1(t), x_0)$. Such curve is not unique in general; but, if g_2 is another one, then $g_2(t) = g_1(t)h(t)$ with h in $\mathrm{Map}(\mathbb{R}_t, G_{x_0})$. It is convenient to choose h so that g_2 is the fundamental solution of (B.1):

$$\dot{g}_2(t) = T_1 R_{g_2(t)} \xi(t),$$

upon using (A.6) in the last equality. Then h is the fundamental solution of the Lie–Scheffers system associated to the curve $B : \mathbb{R} \to \mathfrak{g}_{x_0}$, given by [91]:

$$B(t) = T_1 L_{g_1(t)^{-1}} \big(T_1 R_{g_1(t)} \xi(t) - \dot{g}_1(t) \big).$$

Therefore the knowledge of a particular solution of (B.1) that satisfies $x_{(1)}(0) = x_0$ reduces the problem of finding the fundamental solution for G to finding the fundamental solution for the subgroup G_{x_0}. Naturally if more particular solutions are known, whose values at 0 are $x_1, \dots, x_r$, then we can reduce the problem to solving a Lie–Scheffers system in the subgroup $G_{x_0} \cap \dots \cap G_{x_r}$. When this group is discrete, one can explicitly compute the fundamental solution for G, from which the general solution of the original Lie–Scheffers system can be derived. This is known as the Lie reduction method.

A variant of the Lie reduction method was studied in the language of gauge theory in [16]. Without actually invoking connections, we illustrate the approach in this last reference with the Riccati equation (2.5). The latter seeks the integral curves of the vector field along $\pi_2 : \mathbb{R}_t \times M \to M$:

$$\bar{Y} = \big(a_0(t) + a_1(t)x + a_2(t)x^2 \big) \frac{\partial}{\partial x}.$$

For vector fields $E_+ = \partial/\partial x, H = x\,\partial/\partial x$ and $E_- = x^2\,\partial/\partial x$ we observe the commutation relations

$$[H, E_+] = E_+; \qquad [E_+, E_-] = -2H; \qquad [H, E_-] = -E_-, \tag{B.3}$$

exactly those of the matrices $E'_+ := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$; $H' := \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}$; $E'_- := \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}$. Therefore $E_\pm, H$ realize the (perfect) Lie algebra $\mathfrak{sl}(2;\mathbb{R})$ of the group $SL(2;\mathbb{R})$. The corresponding flows of $\mathbb{R}$ are respectively

$$x_0 \longmapsto x_0 + t; \qquad x_0 \longmapsto x_0 e^t; \qquad x_0 \longmapsto \frac{x_0}{1 - x_0 t};$$

the last one blows up for $x_0 > 0$ in finite time, indicating that E_- is not complete. This can be corrected by adding to $\mathbb{R}$ the point at infinity. More precisely, we have the well-known action of the projective group $SL(2;\mathbb{R})/Z_2$ on the projective line $\mathbb{R} \cup \infty$ —to wit, the projectivization of the fundamental action of $SL(2;\mathbb{R})$ on $\mathbb{R}^2$. Just as well, in the spirit of this article, we can decide to regard the action as a local one, defined on the open set of $SL(2;\mathbb{R}) \times \mathbb{R}$ given by the pairs such that $cx + d \neq 0$.

Now, consider the group $\mathrm{Map}\big(\mathbb{R}_t, SL(2;\mathbb{R})\big)$ of curves acting on the set of Riccati equations (that is, the group of automorphisms of the trivial principal bundle $SL(2;\mathbb{R}) \times \mathbb{R}_t \to \mathbb{R}_t$) corresponding to the indicated action, expressed by:

$$\Phi\big(A(t), x(t)\big) = \Phi\left(\begin{pmatrix} \alpha(t) & \beta(t) \\ \gamma(t) & \delta(t) \end{pmatrix}, x(t) \right) = \frac{\alpha(t)x(t) + \beta(t)}{\gamma(t)x(t) + \delta(t)},$$

together with the other obvious cases. When $x(t)$ is a solution of the Riccati equation (2.5), then $x'(t) := \Phi\big(A(t), x(t)\big)$ is also a solution of a Riccati equation with coefficients

$$\begin{pmatrix} a'_2(t) \\ a'_1(t) \\ a'_0(t) \end{pmatrix} = \begin{pmatrix} \delta^2 & -\delta\gamma & \gamma^2 \\ -2\beta\delta & \alpha\delta + \beta\gamma & -2\alpha\gamma \\ \beta^2 & -\alpha\beta & \alpha^2 \end{pmatrix} \begin{pmatrix} a_2(t) \\ a_1(t) \\ a_0(t) \end{pmatrix} + \begin{pmatrix} \gamma\dot{\delta} - \delta\dot{\gamma} \\ \delta\dot{\alpha} - \alpha\dot{\delta} + \beta\dot{\gamma} - \gamma\dot{\beta} \\ \alpha\dot{\beta} - \beta\dot{\alpha} \end{pmatrix}.$$

The second term on the right hand side is a 1-cocycle for the linear action on the coefficients of the Riccati equation given by the first term. If a particular solution $x_{(1)}(t)$ of (2.5) is known, the element $A_1(t) = \begin{pmatrix} 1 & 0 \\ -x_{(1)}^{-1}(t) & 1 \end{pmatrix} \in \mathrm{Map}\big(\mathbb{R}_t, SL(2;\mathbb{R})\big)$, transforms the original Riccati equation into the linear equation $dx'/dt = \big(2x_{(1)}^{-1}(t)a_0(t) + a_1(t)\big)x' + a_0$, thereby reducing the group $SL(2;\mathbb{R})$ to the subgroup $A(1;\mathbb{R})$. When a second particular solution $x_{(2)}(t)$ of (2.5) is given, then $x' = x_{(1)}x_{(2)}/(x_{(1)} - x_{(2)})$ satisfies the linear equation, therefore we obtain the corresponding homogeneous linear equation using the matrix $A_2 = \begin{pmatrix} 1 & -x_{(1)}x_{(2)}(x_{(1)} - x_{(2)})^{-1} \\ 0 & 1 \end{pmatrix}$. Concretely, the change of variables

$$x'' = \Phi(A_2, x') = \Phi(A_2A_1, x) = \frac{x_{(1)}^2(x - x_{(2)})}{(x_{(2)} - x_{(1)})(x - x_{(1)})}$$

leads to the homogeneous linear equation $dx''/dt = \big(2x_{(1)}^{-1}(t)a_0(t) + a_1(t)\big)x''$. Finally, if $x_{(3)}$ is a third particular solution of (2.5), then $z = x_{(1)}^2(x_{(2)} - x_{(3)})/(x_{(2)} - x_{(1)})(x_{(1)} - x_{(3)})$ solves this linear equation, thus if $A_3 = \begin{pmatrix} z^{-1/2} & 0 \\ 0 & z^{1/2} \end{pmatrix}$, the transformation

$$x''' = \Phi(A_3A_2A_1, x) = \frac{(x - x_{(2)})(x_{(1)} - x_{(3)})}{(x - x_{(1)})(x_{(2)} - x_{(3)})}$$

gives the reduced equation $dx'''/dt = 0$, which is the superposition principle (2.3) for the Riccati equation.

We are not likely to find an exact solution for (B.2) in most cases. This is one reason why we concentrate on approximate solutions in this paper. To attack (2.8), it is generally a good strategy to move on to an equivalent system on the tangent algebra of G —a coordinate space for G which enjoys the advantage of being a linear space. To effect properly the method of working on the tangent algebra, one needs to ponder equivariant maps between homogeneous spaces. We go to this in the next appendix.

C More on the same

Consider again the canonical action of G on its tangent algebra $\mathfrak{g}$, and let $f : \mathfrak{g} \to G$ be a local coordinate map. The exponential map is an example, but of course there are slight variants of it (see below); or we could employ, if available, the Cayley map [92]. A local action B^f of G on $\mathfrak{g}$ is constructed by $B^f_g = f^{-1} \circ L_g \circ f$. This is a (somewhat skew) generalized version of the CBHD map, since, if f is the exponential map, then for $\eta \in \mathfrak{g}$ we obtain:

$$B^{\exp}(g, \eta) = \log(g \exp \eta) = \log g + \eta + \mathrm{CBHD}(\log g, \eta),$$

with the notation of Section 8. Similarly for right actions.

By definition the map f is equivariant with respect to B^f and left translations. Since the maps Φ_x are also equivariant, their composition $\Phi_x \circ f : \mathfrak{g} \to G \to M$ is equivariant, and we have the following commutative diagram relating the flows on M, G and $\mathfrak{g}$:

$$\begin{tikzcd}
\mathfrak{g} \arrow[r, "f"] & G \arrow[r, "\Phi_x"] & M \\
\mathfrak{g} \arrow[u, "B^f_{e^{t\xi}}"] \arrow[r, "f"'] & G \arrow[u, "L_{e^{t\xi}}"'] \arrow[r, "\Phi_x"'] & M, \arrow[u, "\Phi_{e^{t\xi}}"']
\end{tikzcd}$$

with the notation of Section 4 for $\exp(t\xi)$. By Proposition A.1, this commutative diagram can be extended to:

$$\begin{tikzcd}
T\mathfrak{g} \approx \mathfrak{g} \times \mathfrak{g} \arrow[r, "Tf"] & TG \arrow[r, "T\Phi_x"] & TM \\
\mathfrak{g} \arrow[u, "\xi_{\mathfrak{g}}"] \arrow[r, "f"] & G \arrow[u, "\xi_G"'] \arrow[r, "\Phi_x"] & M \arrow[u, "\xi_M(x)"'] \\
\mathfrak{g} \arrow[u, "B^f_{e^{t\xi}}"] \arrow[r, "f"'] & G \arrow[u, "L_{e^{t\xi}}"'] \arrow[r, "\Phi_x"'] & M; \arrow[u, "\Phi_{e^{t\xi}}"']
\end{tikzcd}$$

in particular

$$\xi_M(x) \circ \Phi_x \circ f = T\Phi_x \circ Tf \circ \xi_{\mathfrak{g}}. \tag{C.1}$$

The overarching question is now: what is the concrete description of $\xi_{\mathfrak{g}}$? This we answer next, and we obtain a congenial reply. Write $g = f(u)$ with $u \in \mathfrak{g}$, to distinguish the role of the points of $\mathfrak{g}$ as coordinates for G. The map $T_u f : T\mathfrak{g} \to TG$ can be factorized into a map from $T_u\mathfrak{g} \approx \mathfrak{g}$ to $\mathfrak{g}$, say A^f_u, and the translation $T_1 R_{f(u)}$. Now, in view of (A.6), $Tf \circ \xi_{\mathfrak{g}} = \xi_G \circ f$ gives

$$T_1 R_{f(u)} \circ A^f_u \circ \xi_{\mathfrak{g}}(u) = T_1 R_{f(u)}\, \xi;$$

therefore

$$\xi_{\mathfrak{g}}(u) = (A^f_u)^{-1}\xi, \tag{C.2}$$

where $A^f_u = T_{f(u)} R_{f^{-1}(u)} \circ T_u f$ is the *Darboux derivative* of f, a map $T\mathfrak{g} \to \mathfrak{g} \times \mathfrak{g}$ yielding the pullback via f of the right Maurer–Cartan form on G (a $\mathfrak{g}$-valued 1-form on $\mathfrak{g}$). Then one recovers the 'static' version of (2.7) from a slightly different viewpoint. Note the double role of $\mathfrak{g}$ in the construction: on the one hand, its elements are parameters of the infinitesimal generators on M; on the other hand they serve as coordinates of the *linear* space on which we want to solve a differential equation equivalent to the one originally given on M. The general Darboux derivative for group-valued maps on manifolds is a key ingredient in the study of connections via transitive Lie algebroids [5].

In summary, a differential equation on a homogenous G-manifold M —described by infinitesimal generators of the Lie group action along the projection $\mathbb{R}_t \times M \to M$— has been transformed to a 'pulled-back' equation on the tangent algebra of the group:

$$\dot{u} = \xi_{\mathfrak{g}}(u; t), \tag{C.3}$$

by means of the commutative diagram

$$\begin{tikzcd}
T\mathfrak{g} \arrow[r, "T\Phi_x \circ Tf"] & TM \\
\mathbb{R}_t \times \mathfrak{g} \arrow[u, "\xi_{\mathfrak{g}}"] \arrow[r, "\Phi_x \circ f"] & \mathbb{R}_t \times M, \arrow[u, "\xi_M"']
\end{tikzcd}$$

where $\xi_{\mathfrak{g}}$ is the vector field along $\mathbb{R}_t \times \mathfrak{g} \to \mathfrak{g}$ associated to the curve $t \mapsto \xi(t)$, explicitly given by formula (C.2). The equation evolving on the Lie algebra is susceptible of attack by geometrical integration techniques, a point made in [93].

To exemplify, let us look at Riccati's equation (2.5) again. Consider

$$L := a_0(.)E'_+ + a_1(.)H' + a_2(.)E'_- \in \mathrm{Map}\big(\mathbb{R}_t, \mathfrak{sl}(2;\mathbb{R})\big).$$

We know that if we are able somehow to solve the equation

$$\frac{dg}{dt} = L(t)g(t), \quad \text{with} \quad g(t_0) = 1_{SL(2;\mathbb{R})}, \tag{C.4}$$

then (2.5) is entirely solved by the 'Green operator'

$$x(t) = \Phi(g(t), x_0).$$

where Φ is the integrated action considered in the previous section. To search for that solution, let us bring in a variant of the exponential map [94]. Using canonical coordinates of the second kind for the element $g(t) \in G$, write:

$$\begin{aligned} g(t) = f(u(t)) &:= \exp(u^0(t)E'_+)\exp(u^1(t)H')\exp(u^2(t)E'_-) \\ &=: \exp(u^0(t)L'_0)\exp(u^1(t)L'_1)\exp(u^2(t)L'_2). \end{aligned} \tag{C.5}$$

Therefore f denotes the defined locally bijective map from $\mathfrak{sl}(2;\mathbb{R})$ onto $SL(2;\mathbb{R})$, with $u \equiv (u^0, u^1, u^2)$. (Incidentally, this means that we seek the general solution of the Riccati equation under the form

$$x(t) = \frac{e^{u^1(t)}x_0}{1 - u^2(t)x_0} + u^0(t);$$

then, taking $x_0 = \infty, 0, 1$, three particular solutions are obtained, and the reader will see at once that the superposition formula (2.3) follows from here.)

Replacing $g(t)$ in (C.4) by (C.5), upon using the commutation relations we obtain

$$\begin{aligned} \frac{dg(t)}{dt}g^{-1}(t) &= \dot{u}^0E'_+ + \dot{u}^1e^{u^0E'_+}H'e^{-u^0E'_+} + \dot{u}^2e^{u^0E'_+}e^{u^1H'}E'_-e^{-u^1H'}e^{-u^0E'_+} \\ &= \dot{u}^0E'_+ + \dot{u}^1\exp(u^0\,\mathrm{ad}\,E'_+)H' + \dot{u}^2\exp(u^0\,\mathrm{ad}\,E'_+)\exp(u^1\,\mathrm{ad}\,H')E'_- \\ &= \dot{u}^0E'_+ + \dot{u}^1(H' - u^0E'_+) + \dot{u}^2e^{-u^1}\exp(u^0\,\mathrm{ad}\,E'_+)E'_- \\ &= \dot{u}^0E'_+ + \dot{u}^1(H' - u^0E'_+) + \dot{u}^2e^{-u^1}(E'_- - 2u^0H' + (u^0)^2E'_+) \\ &= (\dot{u}^0 - u^0\dot{u}^1 + (u^0)^2e^{-u^1}\dot{u}^2)E'_+ + (\dot{u}^1 - 2u^0e^{-u^1}\dot{u}^2)H' + e^{-u^1}\dot{u}^2E'_- \\ &= a_0(t)E'_+ + a_1(t)H' + a_2(t)E'_-. \end{aligned}$$

This leads to the following differential equations for the u-variables:

$$\begin{aligned} \dot{u}^0 &= a_0(t) + a_1(t)u^0 + a_2(t)(u^0)^2 \\ \dot{u}^1 &= a_1(t) + 2a_2(t)u^0 \\ \dot{u}^2 &= a_2(t)e^{u^1}, \end{aligned} \tag{C.6}$$

to be solved under the initial conditions $u^0(t_0) = u^1(t_0) = u^2(t_0) = 0$. With the chosen map f, the first equation of this system is the same Riccati equation we started with. This we contrived to make the point again that *one* particular solution needs to be known, for the general solution to be obtainable by quadratures. The explicit form of the Darboux derivative A^f_u in our example is

$$(u^0, u^1, u^2; v^0, v^1, v^2) \mapsto \sum_{k=0}^{2} \mathrm{Ad}_{\prod_{i=0}^{k-1} \exp(u^i L'_i)}\, v^k L'_k.$$

The inversion of this map was just performed, with result the field corresponding to the system (C.6), to wit,

$$\big(a_0(t) + a_1(t)u^0 + a_2(t)(u^0)^2\big)\partial_0 + \big(a_1(t) + 2a_2(t)u^0\big)\partial_1 + a_2(t)e^{u^1}\partial_2,$$

which is our (C.3). Once the latter equation is solved, the rest is obvious: as repeatedly said, one just uses the map $\Phi \circ f$, to go back to M.

The perceptive reader would ask at this point: what about transferring the convolution algebra in Section 4 to the tangent algebra, too? We know nowadays that for conjugation invariant distributions this can be done [95].

Further work on connections *à la* Lie–Rinehart in the respect of Lie–Scheffers systems is in progress [96].

D Faà di Bruno Hopf algebra and the Lie–Engel theorem

Due to its fundamental nature, the Hopf algebra we conjure next is ubiquitous. Let $\mathrm{Diff}^+_0(\mathbb{R})$ be the group of orientation-preserving formal diffeomorphisms of $\mathbb{R}$ (similarly for $\mathbb{C}$) leaving 0 fixed. We think of them as exponential power series:

$$f(t) = \sum_{n=1}^{\infty} \frac{f_n}{n!}\, t^n \quad \text{with} \quad f_1 > 0. \tag{D.1}$$

On $\mathrm{Diff}^+_0(\mathbb{R})$ we consider the coordinate functions

$$a_n(f) := f_n = f^{(n)}(0), \quad n \geq 1.$$

Now,

$$h(t) = \sum_{k=1}^{\infty} \frac{f_k}{k!} \bigg(\sum_{l=1}^{\infty} \frac{g_l}{l!}\, t^l \bigg)^k,$$

where h is the composition $f \circ g$ of two such diffeomorphisms. Therefore, from Cauchy's product formula, the nth coefficient $h_n = a_n(h)$ is

$$h_n = \sum_{k=1}^{n} \frac{f_k}{k!} \sum_{l_i \geq 1,\, l_1 + \cdots + l_k = n} \frac{n!\, g_{l_1} \cdots g_{l_k}}{l_1! \cdots l_k!}.$$

To rewrite h_n in a compact form, it is convenient to introduce the notation

$$\binom{n}{\lambda; k} := \frac{n!}{\lambda_1! \lambda_2! \dots \lambda_n! (1!)^{\lambda_1} (2!)^{\lambda_2} \dots (n!)^{\lambda_n}}.$$

Then, taking in consideration that the sum $l_1 + \cdots + l_k = n$ can be rewritten as

$$\lambda_1 + 2\lambda_2 + \cdots + n\lambda_n = n, \quad \text{where} \quad \lambda_1 + \cdots + \lambda_n = k$$

if there are λ_1 copies of 1, λ_2 copies of 2, and so on, among the l_i; and that the number of contributions from g of this type is precisely the multinomial coefficient

$$\binom{k}{\lambda_1 \cdots \lambda_n} = \frac{k!}{\lambda_1! \cdots \lambda_n!},$$

it follows:

$$h_n = \sum_{k=1}^{n} f_k \sum_{\lambda \vdash n, |\lambda|=k} \binom{n}{\lambda; k} g_1^{\lambda_1} \ldots g_n^{\lambda_n} =: \sum_{k=1}^{n} f_k \, B_{n,k}(g_1, \ldots, g_{n+1-k}). \tag{D.2}$$

We have used notations of the theory of partitions of integers. The $B_{n,k}$ are called the (partial, exponential) *Bell polynomials*, often defined via the expansion

$$\exp\biggl(u \sum_{m \geq 1} g_m \frac{t^m}{m!}\biggr) = 1 + \sum_{n \geq 1} \frac{t^n}{n!} \biggl[\sum_{k=1}^{n} u^k B_{n,k}(g_1, \ldots, g_{n+1-k})\biggr],$$

which is a particular case of (D.2). Each $B_{n,k}$ is a homogeneous polynomial of degree k.

According to (4.5), a coproduct on $\mathcal{R}(\mathrm{Diff}_0^+(\mathbb{R}))$, which we realize as the polynomial algebra $\mathbb{R}[a_1, a_2, \ldots]$, is given by $\Delta a_n(g, f) = a_n(f \circ g)$. This entails

$$\Delta a_n = \sum_{k=1}^{n} \sum_{\lambda \vdash n, |\lambda|=k} \binom{n}{\lambda; k} a_1^{\lambda_1} a_2^{\lambda_2} \ldots a_n^{\lambda_n} \otimes a_k. \tag{D.3}$$

The flip of f and g is done to keep the tradition of writing the linear part on the right of the tensor product; this amounts to taking the opposite coalgebra structure. With (D.3) we have a bialgebra structure. In a Hopf algebra grouplike elements are invertible: $g^{-1} = Sg$. Since a_1 is grouplike, to have an antipode one must either adjoin an inverse a_1^{-1}, or put $a_1 = 1$. The latter is equivalent to work with the subgroup $\mathrm{Diff}_{0,1}^+(\mathbb{R})$ of $\mathrm{Diff}_0^+(\mathbb{R})$, of diffeomorphisms tangent to the identity at 0, that is, to consider power series such that $f_0 = 0$ and $f_1 = 1$. The coproduct formula is accordingly simplified to:

$$\Delta a_n = \sum_{k=1}^{n} \sum_{\lambda \vdash n, |\lambda|=k} \binom{n}{\lambda; k} a_2^{\lambda_2} a_3^{\lambda_3} \cdots \otimes a_k = \sum_{k=1}^{n} B_{n,k}(1, \ldots, a_{n+1-k}) \otimes a_k.$$

The resulting graded connected Hopf algebra $\mathcal{F} = \mathcal{R}^{\mathrm{cop}}(\mathrm{Diff}_{0,1}^+(\mathbb{R}))$, where the superindex stands for the opposite coalgebra structure, was baptized Faà di Bruno Hopf algebra by Joni and Rota [97]. The degree is then given by $|a_n| = n - 1$.

Several comments are in order. Formula (D.2) can be directly expressed in terms of partitions of finite sets; consult [3] or [98, Chapter 5]. The happy fact that the algebra of representative functions $\mathcal{R}^{\mathrm{cop}}(\mathrm{Diff}_{0,1}^+(\mathbb{R}))$ is graded is related to the linearity of the product $f \circ g$ in one of the coordinates. This also means that $\mathrm{Diff}_{0,1}^+(\mathbb{R})$ is the inverse limit of finite-dimensional matrix groups, and that it possesses a (necessarily unique) right invariant connection with vanishing torsion and curvature. Also, although 'formal' may sound a bit dismissive, one should remember that, in view of E. Borel's theorem, expression (D.1) does represent a smooth function; and that (D.2) can be used to show without having recourse to complex variables that the composition of analytic functions on appropriate domains is analytic [99].

Let us turn our attention to the dual of the Faà di Bruno Hopf algebra $\mathcal{F}$. Since we are dealing with a graded connected Hopf algebra it is natural to consider the *graded dual*, that we denote simply by $\mathcal{F}'$; for which $\mathcal{F}'' = \mathcal{F}$. (From the discussion in Sections 4 and 5 we know there exist bigger duals, for instance $\mathcal{F}'$ does not have grouplike elements apart from its unit η.) Let a'_n be the linear functionals defined by $\langle a'_n, P\rangle = \partial P/\partial a_n(0)$, where P is a polynomial in $\mathbb{R}[a_2, a_3, \dots]$. In particular the a'_n kill non-trivial products of the a_q generators. Also, taking in consideration that the counit $\eta(P) = P(0)$ of $\mathcal{F}$ is the unit in $\mathcal{F}'$

$$\langle \Delta a'_n, P\otimes Q\rangle = \langle a'_n, m(P\otimes Q)\rangle = \langle a'_n, PQ\rangle = \frac{\partial(PQ)}{\partial a_n}(0) = \langle a'_n\otimes 1 + 1\otimes a'_n, P\otimes Q\rangle,$$

Thus the a'_n are primitive. Using the definition of the Bell polynomials

$$\langle a'_n a'_m, a_q\rangle = \langle a'_n\otimes a'_m, \Delta a_q\rangle = \begin{cases} \binom{m+n-1}{n} & \text{if } q = m+n-1 \\ 0 & \text{otherwise.}\end{cases}$$

On the other hand, note that

$$\Delta(a_q a_r) = a_q a_r\otimes 1 + 1\otimes a_q a_r + a_q\otimes a_r + a_r\otimes a_q + R,$$

where R is either vanishing or a sum of terms of the form $b\otimes c$ with b or c a monomial in $a_2, a_3, \dots$ of degree greater than 1. Therefore

$$\langle a'_n a'_m, a_q a_r\rangle = \langle a'_n\otimes a'_m, \Delta(a_q a_r)\rangle = \begin{cases} 1 & \text{if } n = q \neq m = r \text{ or } n = r \neq m = q, \\ 2 & \text{if } m = n = q = r, \\ 0 & \text{otherwise.}\end{cases}$$

Similarly, since all the terms of the coproduct of three or more a_q's are the tensor product of two monomials where at least one of them is of order greater than 1, it follows that

$$\langle a'_n a'_m, a_{q_1} a_{q_2} a_{q_3}\cdots\rangle = 0.$$

Collecting all this together,

$$a'_n a'_m = \binom{m-1+n}{n} a'_{n+m-1} + \big(1+\delta_{nm}\big)(a_n a_m)'.$$

In particular,

$$[a'_n, a'_m] := a'_n a'_m - a'_m a'_n = (m-n)\frac{(n+m-1)!}{n!m!}\, a'_{n+m-1}.$$

Therefore, taking $b'_n := (n+1)! a'_{n+1}$, we get the simpler looking

$$[b'_n, b'_m] = (m-n)\, b'_{n+m}. \tag{D.4}$$

The Cartier–Milnor–Moore theorem implies that $\mathcal{F}'$ is the enveloping algebra of the Lie algebra spanned by the b'_n with commutators (D.4). Obviously $\mathcal{F}'$ can be realized by the vector fields $Z_n := x^{n+1}\,\partial/\partial x$, for $n \geq 1$, on the real line [100].

Consider the 'regular representation' of $\mathcal{F}$ given by $\langle a \triangleright a', b\rangle := \langle a', ba\rangle$ on $\mathcal{F}'$. Since

$$\langle b \triangleright (a \triangleright a'), c\rangle = \langle a \triangleright a', cb\rangle = \langle a', cba\rangle = \langle ba \triangleright a', c\rangle \quad \text{and} \quad \langle 1 \triangleright a', b\rangle := \langle a', b\rangle,$$

we do obtain a left module algebra over $\mathcal{F}$. Let now a be a primitive element of $\mathcal{F}$; using the Sweedler notation:

$$\begin{aligned}
\langle a \triangleright b'a', c\rangle &= \langle b'a', ca\rangle = \langle b' \otimes a', \Delta(ca)\rangle = \langle b' \otimes a', \Delta c\Delta a\rangle \\
&= \langle b' \otimes a', c_{(1)} \otimes c_{(2)}(a \otimes 1 + 1 \otimes a)\rangle \\
&= \langle a \triangleright b' \otimes a' + b' \otimes a \triangleright a', c_{(1)} \otimes c_{(2)}\rangle \\
&= \langle (a \triangleright b')a' + b'(a \triangleright a'), c\rangle,
\end{aligned}$$

so a acts as a derivation. In particular if $a \triangleright a' = a \triangleright b' = 0$, then $a \triangleright (a'b') = 0$, hence the kernel of the map $a \triangleright \cdot$ is a Lie subalgebra of vector fields, and we conclude that primitive elements of $\mathcal{F}$ identify finite-dimensional Lie subalgebras of vector fields. Now, the space $P(\mathcal{F})$ of primitive elements of $\mathcal{F}$ has *dimension two*. Indeed, $P(\mathcal{F}) = \big(\mathbb{R}1 \oplus \mathcal{F}_+'^{\,2}\big)^{\perp}$, where $\mathcal{F}'_+ := \ker \eta$ is the augmentation ideal of $\mathcal{F}'$. By (D.4) there is a dual basis of $\mathcal{F}'$ made of products, except for its first two elements. Hence $\dim P(\mathcal{F}) = 2$. A basis of $P(\mathcal{F})$ is given by $\{a_2, a_3 - \frac{3}{2}a_2^2\}$. This yields the equations $y'' = 0$ and $y'y''' - 3(y'')^2/2 = 0$, respectively solved by dilations and by the action of $SL(2;\mathbb{R})$ we know; translations do not show up because we made $a_1 = 1$.

The previous argument, together with the part of classical one [101] —more recently rehearsed in [28, Section XIX] or in [91]— to the effect that infinitesimally transitive actions on the line must correspond to Lie algebras of vector fields of dimension at most three, shows that Riccati's is the only nonlinear Lie–Scheffers differential equation on the real (or complex) line. Whether or not it is simpler to think in Hopf algebraic terms seems largely a matter of taste. We do contend that the Beatus Faà di Bruno algebra is too fundamental an object to ignore.

Acknowledgments

We are very thankful to J. C. Várilly for helpful comments. The first named author acknowledges partial financial support through grants BFM-2003-02532 and DGA-Grupos Consolidados E24/1. KEF greatly acknowledges support from the European Post-Doctoral Institute. HF acknowledges support from the Vicerrectoría de Investigación of the Universidad de Costa Rica. The first and fourth named authors are grateful for discussions with J. Grabowski and G. Marmo, during a visit to Università di Napoli Federico II in November 2005. JMGB moreover acknowledges partial support from CICyT, Spain, through grant FIS2005-02309. KEF, HF and JMGB thank the Departamento de Física Teórica of the Universidad de Zaragoza for its warm hospitality.

References

[1] S. Lie and G. Scheffers, *Vorlesungen über continuierliche Gruppen mit geometrischen und anderen Anwendungen*, Teubner, Leipzig, 1893.

[2] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, Benjamin, Reading, MA, 1978.

[3] H. Figueroa and J. M. Gracia-Bondía, Rev. Math. Phys. **17** (2005) 881.

[4] G. M. Bergman, Contemp. Math. **43** (1985) 25.

[5] K. C. H. Mackenzie, *General theory of Lie groupoids and Lie algebroids*, Cambridge University Press, Cambridge, 2005.

[6] W. Magnus, Commun. Pure Appl. Math. **7** (1954) 649.

[7] F. J. Dyson, Phys. Rev. **84** (1948) 486.

[8] K. T. Chen, Ann. Math. **65** (1957) 163.

[9] H. Araki, Ann. Sci. École Norm. Sup. **6** (1973) 67.

[10] H. Omori's, "Infinite-dimensional Lie groups", American Mathematical Society, Providence, RI, 1997.

[11] J. Huebschmann, J. reine angew. Math. **408** (1990) 57.

[12] J. C. Butcher, *T*he numerical analysis of ordinary differential equations, Wiley & Sons, Chichester, England, 2003.

[13] J. Nasarre, "Aplicación de métodos algebraicos y geométricos al estudio de la evolución de sistemas dinámicos", Ph. D. thesis, Zaragoza, 1999.

[14] J. F. Cariñena, J. Grabowski and G. Marmo, *Lie–Scheffers system: a geometric approach*, Bibliopolis, Naples, 2000.

[15] O. Stormark, *Lie's structural approach to PDE systems*, Cambridge University Press, Cambridge, 2000.

[16] J. F. Cariñena and A. Ramos, in *Differential Geometry and its applications*, Charles University, Prague, 2005; pp. 437–452.

[17] J. F. Cariñena, J. Grabowski and G. Marmo, "Superposition rules, Lie theorem and partial differential equations", arXiv: math-ph/0610013.

[18] H. Figueroa, J. M. Gracia-Bondía and J. C. Várilly, "Faà di Bruno Hopf algebras", in the *Encyclopaedia of Mathematics*, M. Hazewinkel (ed.), Springer-Verlag, Berlin, 2007. arXiv: math.CO/0508337.

[19] D. Kastler and R. Stora, J. Geom. Phys. **2**(3) (1985) 1.

[20] J. Huebschmann, Contemporary Mathematics **227** (1999) 145.

[21] J. Grabowski, Rep. Math. Phys. **52** (2003) 445.

[22] P. Aschieri, "Noncommutative symmetries and gravity", arXiv: hep-th/0608172.

[23] G. Rinehart, Trans. Amer. Math. Soc. **108** (1963) 195.

[24] M. Kawski and H. J. Sussmann, in *Operators, systems, and linear algebra*, Teubner, Stuttgart, 1997; pp. 111–128.

[25] A. A. Agrachev and Y. L. Sachkov, *Control theory from the geometric viewpoint*, Springer-Verlag, Berlin, 2004.

[26] M. Hazewinkel, CWI Quarterly **4** (1991) 3.

[27] J. Dieudonné, *Éléments d'Analyse III*, Gauthier-Villars, Paris, 1970.

[28] J. Dieudonné, *Éléments d'Analyse IV*, Gauthier-Villars, Paris, 1971.

[29] K. Ebrahimi-Fard, J. M. Gracia-Bondía and F. Patras, "A Lie theoretic approach to renormalization", arXiv: hep-th/0609035.

[30] A. Connes and M. Marcolli, "From physics to number theory via noncommutative geometry II", to appear in *Frontiers in Number Theory, Physics and Geometry.* arXiv: hep-th/0411114.

[31] G. Warner, *Harmonic Analysis on semi-simple Lie Groups I*, Springer-Verlag, Berlin, 1972.

[32] M. E. Taylor, *Noncommutative Harmonic Analysis*, American Mathematical Society, Providence, RI, 1986.

[33] P. Cartier, A primer on Hopf algebras, IHES preprint, August 2006.

[34] C. Reutenauer, *Free Lie Algebras*, Clarendon Press, Oxford, 1993.

[35] D. Wigner, Proc. Amer. Math. Soc. (1989) 639.

[36] F. Patras and C. Reutenauer, Adv. Appl. Math. **28** (2002) 560.

[37] R. Ree, Ann. Math. **68** (1958) 210.

[38] F. Patras, Ann. Inst. Fourier Grenoble **43** (1993) 1067.

[39] F. Patras, J. Algebra **170** (1994) 547.

[40] M. Gerstenhaber and S. D. Schack, J. Pure Appl. Algebra **48** (1987) 229.

[41] J. L. Loday, Invent. Math. **96** (1989) 205.

[42] M. Gerstenhaber and S. D. Schack, J. Pure Appl. Algebra **70** (1991) 263.

[43] R. S. Strichartz, J. Func. Anal. **72** (1987) 320.

[44] B. Mielnik and J. Plebański, Ann. Inst. Henri Poincaré A **XII** (1970) 215.

[45] R. Suárez and L. Saénz, J. Math. Phys. **42** (2001) 4582.

[46] V. S. Varadarajan, *Lie groups, Lie algebras and their representations*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[47] C. Reutenauer, Springer Lecture Notes in Mathematics **1234** (1986) 267.

[48] J. L. Loday, Expos. Math. **12** (1994) 165.

[49] E. R. Kolchin, *Differential Algebras and Algebraic Groups*, Academic Press, New York, 1973.

[50] G.-C. Rota, DMV Mittellungen, Heft **2** (1998) 45.

[51] A. Joseph, *Quantum groups and their primitive ideals*, Springer-Verlag, Berlin, 1995.

[52] W. Zudilin, "One parameter models of Hopf algebras associated with multiple zeta values", preprint, 2001.

[53] M. A. Semenov-Tian-Shansky, Funct. Anal. Appl. **17** (1983) 254.

[54] J.-L. Loday, Springer Lecture Notes in Mathematics **1763** (2001) 7.

[55] K. Ebrahimi-Fard, Lett. Math. Phys. **61** (2002) 139.

[56] K. Ebrahimi-Fard and Ph. Leroux, "Generalized shuffles related to Nijenhuis and TD-algebras", arXiv: math.RA/0606164.

[57] F. Spitzer, Trans. Amer. Math. Soc. **82** (1956) 323.

[58] G. Baxter, Pacific J. Math. **10** (1960) 731.

[59] G.-C. Rota, Bull. Amer. Math. Soc. **75** (1969) 325; *ibidem* 330.

[60] G.-C. Rota, in *Gian-Carlo Rota on Combinatorics, Introductory papers and commentaries*, J. P. S. Kung (ed.), Birkhäuser, Boston, 1995.

[61] G.-C. Rota and D. A. Smith, Istituto Nazionale di Alta Matematica, **IX** (1972) 179.

[62] J. F. C. Kingman, J. London Math. Soc. **37** (1962) 309.

[63] P. Cartier, Adv. Math. **9** (1972) 253.

[64] K. Ebrahimi-Fard and L. Guo, "Rota–Baxter Algebras in renormalization of perturbative quantum field theory", to appear in Fields Institute Communications. arXiv: hep-th/0604116.

[65] D. Kreimer, Adv. Theor. Math. Phys. **3** (2000) 627.

[66] A. Connes and D. Kreimer, Commun. Math. Phys. **210** (2000) 249.

[67] K. Ebrahimi-Fard and D. Kreimer, J. Phys. A: Math. Gen. **38** (2005) R385.

[68] F. V. Atkinson, J. Math. Anal. and Applications **7** (1963) 1.

[69] K. Ebrahimi-Fard, L. Guo and D. Manchon, Commun. Math. Phys. **267** (2006) 821.

[70] K. Ebrahimi-Fard, L. Guo and D. Kreimer, J. Phys. A **37** (2004) 11037.

[71] K. Ebrahimi-Fard, L. Guo and D. Kreimer, Ann. Henri Poincaré **6** (2005) 369.

[72] K. Ebrahimi-Fard, J. M. Gracia-Bondía and F. Patras, forthcoming.

[73] C. S. Lam, J. Math. Phys. **39** (1998) 5543.

[74] W. Rossmann, *Lie groups. An introduction through linear groups*, Oxford University Press, Oxford, 2002.

[75] E. C. G. Sudarshan and N. Mukunda, *Classical dynamics: a modern perspective*, Wiley & Sons, New York, 1974.

[76] A. Iserles and S. P. Norsett, Philos Trans. R. Soc. London A **357** (1999) 983.

[77] A. Iserles, Notices of the AMS **49** (2002) 430.

[78] R. M. Wilcox, J. Math. Phys. **8** (1967) 962.

[79] W. R. Salzman, J. Chem. Phys. **82** (1985) 822.

[80] S. Shnider and S. Sternberg, *Quantum groups. From coalgebras to Drinfel'd algebras. A guided tour*, Graduate Texts in Mathematical Physics, International Press, Cambridge, MA, 1993.

[81] S. Klarsfeld and J. A. Oteo, Phys. Rev. A **39** (1989) 3270.

[82] C. Moan and J. Niesen, "Convergence of the Magnus series", arXiv: math.CA/0609198.

[83] J. A. Oteo and J. Ros, J. Math. Phys. **41** (2000) 3268.

[84] K. Ebrahimi-Fard and L. Guo, J. Alg. Comb. **4** (2006) 83.

[85] F. Fer, Bull. Classe des Sci. Acad. Roy. Belg. **44** (1958) 818.

[86] H. Z. Munthe-Kaas and W. M. Wright, On the Hopf algebraic structure of Lie group integrators, arXiv: math.AC/0603023.

[87] D. Harrivel, Butcher series and control theory, arXiv: math.0C/0603133.

[88] P. W. Michor, Isometric actions of Lie groups and invariants, preprint, Wien, 1997.

[89] A. A. Kirillov, *Lectures on the orbit method*, American Mathematical Society, Providence, RI, 2004.

[90] M. Asorey, J.F. Cariñena and M.A. del Olmo, J. Phys. A **16** (1983) 1603.

[91] R. L. Bryant, in *Geometry and quantum field theory*, D. S. Freed and K. K. Uhlenbeck (eds.), American Mathematical Society, Providence, RI, 1995; pp. 1–181.

[92] B. Kostant and P. W. Michor, in *The orbit method in geometry and physics: in honor of A. A. Kirillov*, C. Duval, L. Guieu and V. Ovsienko (eds.), Birkhäuser, Basel, 2003; pp. 259–296.

[93] K. Engø, BIT **40** (2000) 41.

[94] J. Wei and E. Norman, Proc. Amer. Math. Soc. **15** (1964) 327.

[95] A. H. Dooley and N. J. Wildberger, Can. Math. Bull. **27** (1993) 25.

[96] J. F. Cariñena and H. Figueroa, forthcoming.

[97] S. A. Joni and G.-C. Rota, Contemp. Math. **6** (1982) 1.

[98] R. P. Stanley, *Enumerative combinatorics*, vol. 2, Cambridge University Press, Cambridge, 1999.

[99] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, Birkhäuser, Boston, 2002.

[100] A. Connes and H. Moscovici, Commun. Math. Phys. **198** (1998) 198.

[101] S. Lie and F. Engel, *Theorie der Transformationsgruppen III*, Teubner, Leipzig, 1893.