Clumped forests for the description of the Hopf algebra structures of exotic aromatic series and their applications in stochastic numerical analysis

Eugen Bronasco[1] and Adrien Laurent[2]

February 20, 2024

Abstract

We uncover the algebraic structures associated to the laws of composition and substitution of exotic aromatic S-series. For standard Butcher series, these laws are the convolution products of characters of the corresponding Hopf algebras of Connes-Kreimer and Calaque-Ebrahimi-Fard-Manchon, but this does not generalise to exotic aromatic S-series. To address this challenge, we introduce the notion of clumped forests and we use decorated forests to describe conveniently exotic aromatic forests. The formalism applies successfully in numerical analysis to the algebraic formulation of stochastic order theory, in the weak sense and for the invariant measure, as well as to the description of the composition of integrators, postprocessing, backward error analysis, and modified equations.

Keywords: geometric numerical integration, exotic aromatic series, Butcher-series, Hopf algebra, stochastic differential equations, invariant measure, backward error analysis.

AMS subject classification (2020): Primary: 16T05, 41A58, 60H35. Secondary: 37M25, 65L06, 70H45.

1 Introduction

The understanding of the algebraic and combinatorial structures underlying the Taylor expansions of the solution of ordinary differential equations and their approximations is a central tool in geometric numerical integration. In particular, the work [24] presents the two Hopf algebra structures governing the composition and the substitution of such Taylor expansions and applies these theoretical results to the formulation of a general theory for the creation and study of geometric integrators (see also [22]). The aim of the present work is to use the powerful tool that are exotic aromatic S-series to extend the work [24] for the approximation of stochastic differential equations. This is not straightforward as there are different types of approximations of stochastic dynamics: strong, weak, and for the invariant measure, there is no backward error analysis of SDEs in general [83], and the Hopf algebra structures used in the deterministic case do not extend naturally to the stochastic case. We shall ignore the strong approximations as it is beyond the scope of the present work. The ultimate goal of this paper is the rigorous formalisation of stochastic geometric numerical integration with exotic aromatic S-series.

Butcher-series were first introduced in [17, 39] (see also the textbooks [38, 18, 19] and the review [68]) for deriving order conditions for Runge-Kutta methods in numerical analysis.

---

[1]Université de Genève, Section de mathématiques, Genève, Switzerland. Eugen.Bronasco@unige.ch.

[2]Univ Rennes, INRIA (Research team MINGuS), IRMAR (CNRS UMR 6625) and ENS Rennes, France. Adrien.Laurent@INRIA.fr.

They were later applied successfully to a variety of different fields such as geometric numerical integration [38], rough paths theory [37, 40], or to stochastic numerical analysis. We mention in particular the aromatic extension of B-series introduced in [25, 43] for the study of volume-preserving integrators. They allow to compute the divergence of a B-series, and were later studied in [67, 71, 9, 33, 10, 51, 50] for their algebraic, geometric, and numerical properties. For stochastic numerical analysis, we mention in particular the early works [14, 45, 15] that first introduced stochastic trees and B-series for the strong convergence of SDEs, and the works by Rößler [76, 77, 78, 79] and Debrabant and Kværnø [28, 27, 29] for the design and analysis of high order strong and weak integrators on a finite time interval, as well as the works [8, 48]. The different stochastic extensions of B-series were recently formalised and unified into the exotic aromatic S-series formalism [53, 54, 49, 12] for the study of stochastic numerical analysis in the weak sense and for the invariant measure. It was then shown in [52] that the exotic aromatic formalism is precisely the natural extension of B-series and aromatic B-series as they all satisfy similar universal geometric properties. We present in this paper the algebraic and combinatorial structures related to the exotic aromatic S-series and study their concrete applications in stochastic numerical analysis.

We introduce the concept of decorated aromatic S-series, defined using decorated aromatic forests, to simplify the study of the algebraic structures related to the exotic aromatic S-series. Our investigation involves the study of the D-algebra structure [72] over the decorated aromatic forests, and we employ this structure to introduce a Hopf algebroid [62, 70], which we refer to as the Grossman-Larson Hopf algebroid. This Hopf algebroid is intricately linked to the composition law of decorated aromatic S-series. The substitution law does not generalise straightforwardly with decorated aromatic S-series and the main difficulty comes from the aromas, as first observed in [9] in the context of aromatic forests. We introduce the concept of clumped forests, which represent monomials of aromatic trees and form the universal enveloping algebra of aromatic trees. The difference between aromatic forests and clumped forests lies in the aromas that are attached to the rooted components in the latter case: $\vartheta_\bullet \cdot \mathbf{i} \neq \bullet \cdot \vartheta\mathbf{i}$. These clumped forests are essential in establishing the algebraic foundations required for the substitution law, in the spirit of [20, 63, 64, 65, 75]. Subsequently, we proceed to construct the space of exotic aromatic forests using the decorated aromatic forests we introduced earlier. We use our analysis of algebraic structures over decorated aromatic forests to present the composition [12] and substitution law for the exotic aromatic S-series.

The Hopf algebra structures related to exotic aromatic S-series allow us to elegantly formalise stochastic numerical analysis theory. The composition law gives the methodology to derive order conditions for arbitrarily high orders, and to study the accuracy of the composition of different numerical integrators and of postprocessors [85]. The substitution law is a powerful tool in geometric numerical integration, as it allows one to derive the modified vector field for backward error analysis. In the stochastic setting, the backward error analysis rewrites for the convergence for the invariant measure via the use of an additional operation, called the integration by parts [3, 53, 12]. The previous works give a methodology to construct a modified vector field for backward error analysis or modified equations, but there is no explicit formula and there is no proof that it fits into the exotic B-series formalism, making it tedious to compute even for low orders. The exotic aromatic S-series formalism allows us to derive a new explicit expression of a modified vector field that writes as an exotic B-series for backward error analysis and for creating modified equations in the Euclidean case. This simplifies greatly the computation of the modified vector field and the exotic B-series expression ensures that the modified vector field satisfies important natural geometric properties such as orthogonal equivariance [52].

The paper is organised so that the reader interested only in the numerical applications can skip the algebraic details (Section 2) and so that no knowledge of stochastic numerical analysis is needed to understand the algebraic results of the paper, that are interesting independently of their numerical applications. The detailed structure of the paper is the following. Section 2 presents the algebraic structure of decorated aromatic forests formalism, relying on the new idea of clumping. This structure is applied in Section 3 to the exotic aromatic forests and their Hopf algebra structures linked to composition and substitution with numerous examples. The variety of concrete numerical applications are gathered in Section 4 and include the algebraic formalisation of stochastic order theory in the weak sense and for the invariant measure, the application of the composition law to the composition of integrators and postprocessors, and the application of the substitution law to backward error analysis and modified equations.

2 Decorated clumped forests for the description of the Hopf algebraic structures of the composition and substitution laws

We consider graphs $\pi = (V, E, S)$ where V is a finite set of vertices and $E \subset V \times V$ is a set of directed edges. The empty graph is included and written $\mathbf{1}$. If $e = (v, w) \in E$, the edge e is going from the source v to the target w, v is a predecessor of w, and w is a successor of v. The stolons[1] in $S \subset V \times V/(x, y) = (y, x)$ link some of the vertices without successors by pairs, but these vertices are in at most one stolon. The vertices that don't have a successor and are not part of stolons are called roots and the vertices that are not in stolons and do not have predecessors are called leaves. The graphs have two kinds of connected components: the ones that have a root, called trees, and the other ones, called aromas. The aromas either have a stolon or have a cycle, that is a list of vertices $\{v_1, \ldots, v_n\}$ where v_{i+1} is the successor of v_i and v_1 is the successor of v_n. We call aromatic forests, gathered in the set AF, the equivalence classes of such graphs, where two graphs are equivalent if there exists a bijection between their sets of vertices and edges that preserve successors, predecessors, and stolons. When drawn, by convention, the orientation of the edges goes from top to bottom and in clockwise direction for cycles.

We consider the following subsets of AF. The set of trees T contains the connected aromatic forests with one root. The trees with up to four vertices are

$$\text{\small (trees diagram)} \cdot, \; \mathbf{!}, \; \mathsf{V}, \; \mathbf{!}, \; \mathsf{W}, \; \mathsf{V}, \; \mathsf{Y}, \; \mathbf{!}$$
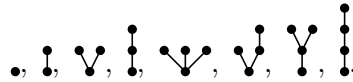
We obtain the set F of forests by taking all possible unordered monomials of trees, including the empty forest $\mathbf{1}$.

$$F = \{\mathbf{1}, \; \cdot, \; \cdot \cdot, \; \mathbf{!}, \; \cdot \cdot \cdot, \; \cdot \mathbf{!}, \; \mathsf{V}, \; \mathbf{!}, \ldots\}$$

The subset of aromatic forests with n roots is written AF_n, which gives us a first grading on AF. In particular, the set of aromas is

$$A = AF_0 = \{\mathbf{1}, \; \circlearrowleft, \; \multimap, \; \overbrace{}, \; \overset{\bullet}{\circlearrowleft}, \; \multimap\circlearrowleft, \; \multimap\; \multimap, \ldots\},$$

[1]In botany, stolons are horizontal connections that link the base of two plants, allowing a plant to clone itself. Strawberry plants are an example of plants with stolons.

with the double edge denoting a stolon, and the aromatic trees are in

$$AT = AF_1 = \{ \text{⬡⬡}, \text{⬡⬡}, \vee, \dots \}.$$

Note that the set $AF \simeq A \times F$ of aromatic forests is the product of the sets of aromas and forests. The vector space spanned by a given set is written in calligraphic font. For instance, we write $\mathcal{AF} = \mathrm{Span}_{\mathbb{R}}(AF)$.

Note that the definition of aromas taken here is more general than in [43, 25, 9], as we choose to include stolons among the aromas. The reason is that the combinatorial structure of stolonic and cyclic aromas is similar, and the exotic extension of the standard aromatic forests relies in particular on stolons.

Decorated aromatic forests are aromatic forests $\pi \in AF$ endowed with a decoration $\alpha : V(\pi) \to D$ for a given set D. A morphism $\varphi : (\pi_1, \alpha_1) \to (\pi_2, \alpha_2)$ between decorated aromatic forests is a graph morphism $\varphi : \pi_1 \to \pi_2$ such that $\alpha_1 = \alpha_2 \circ \varphi$. The group of automorphisms is denoted by $\mathrm{Aut}(\pi, \alpha)$ and $\sigma(\pi, \alpha) := |\mathrm{Aut}(\pi, \alpha)|$. The set of decorated aromatic forests is denoted by AF_D and the corresponding vector space by $\mathcal{AF}_D$.

We will often omit writing α for simplicity and will denote the elements of AF_D by π. The rest of this section is devoted to the study of the algebraic properties of decorated aromatic forests. We shall focus on a specific class of decorated aromatic forests, called exotic aromatic forests, in Section 3.

2.1 D-algebra of decorated aromatic forests

In this section, we introduce algebraic structures over decorated aromatic forests that are relevant in our description of the substitution law. Let $\curvearrowright : \mathcal{AT}_D \otimes \mathcal{AT}_D \to \mathcal{AT}_D$ denote the grafting product over decorated aromatic trees. The grafting product connects the root of the left operand to a vertex of the right operand in all possible ways, for example,

$$\text{⬡} \curvearrowright \text{⬡} = 2\,\text{⬡} + \text{⬡}.$$

Let us consider the divergence of an aromatic tree to be defined as a map $d : \mathcal{AT}_D \to \mathcal{A}_D$ that connects the root of an aromatic tree to one of its vertices in all possible ways, for example,

$$d(\text{⬡}\vee) = \text{⬡} + \text{⬡} + 2\,\text{⬡}.$$

Let $El_{\mathcal{A}_D}(\mathcal{AT}_D)$ be an algebra of endomorphisms $- \curvearrowright \tau : \mathcal{AT}_D \to \mathcal{AT}_D$. The pair $(\mathcal{AT}_D, \mathcal{A}_D)$ is a tracial pre-Lie-Rinehart algebra generated by the set D, that is, it satisfies the following properties:

1. $\mathcal{A}_D$ is a unital commutative algebra with concatenation product $\cdot : \mathcal{A}_D \otimes \mathcal{A}_D \to \mathcal{A}_D$,

2. $\mathcal{AT}_D$ is an $\mathcal{A}_D$-module with a pre-Lie product $\curvearrowright : \mathcal{AT}_D \otimes \mathcal{AT}_D \to \mathcal{AT}_D$, that is,

$$\tau \curvearrowright (\gamma \curvearrowright \nu) - (\tau \curvearrowright \gamma) \curvearrowright \nu = \gamma \curvearrowright (\tau \curvearrowright \nu) - (\gamma \curvearrowright \tau) \curvearrowright \nu,$$

for $\tau, \gamma \in \mathcal{AT}_D, \nu \in \mathcal{AT}_D \sqcup \mathcal{A}_D$,

4

3. for any $\tau \in \mathcal{AT}_D$, the map $\tau \curvearrowright - : \mathcal{A}_D \to \mathcal{A}_D$ is a derivation and the Leibniz rule holds,

$$\tau \curvearrowright (\omega \cdot \nu) = (\tau \curvearrowright \omega) \cdot \nu + \omega \cdot (\tau \curvearrowright \nu),$$

for $\omega \in \mathcal{A}_D$ and $\nu \in \mathcal{A}_D \sqcup \mathcal{AT}_D$.

4. there exists a map $t : El_{\mathcal{A}_D}(\mathcal{AT}_D) \to \mathcal{A}_D$ called a trace such that $t(\tilde{\tau} \circ \tilde{\gamma}) = t(\tilde{\gamma} \circ \tilde{\tau})$ and $t(\tau \curvearrowright \tilde{\gamma}) = \tau \curvearrowright t(\tilde{\gamma})$ with $\tau \in \mathcal{AT}_D, \tilde{\tau}, \tilde{\gamma} \in El_{\mathcal{A}_D}(\mathcal{AT}_D)$ and $\circ$ denotes the composition of endomorphisms. The divergence is then defined as $d(\tau) := t(- \curvearrowright \tau)$.

More details can be found in [33] where it is proven that $(\mathcal{AT}_D, \mathcal{A}_D)$ is a free tracial pre-Lie-Rinehart algebra. We extend the structure of the tracial pre-Lie-Rinehart algebra by considering a symmetric $\mathcal{A}_D$-bilinear form $\langle -, - \rangle : \mathcal{AT}_D \otimes \mathcal{AT}_D \to \mathcal{A}_D$ with the Leibniz rule:

$$\tau \curvearrowright \langle \gamma, \nu \rangle = \langle \tau \curvearrowright \gamma, \nu \rangle + \langle \gamma, \tau \curvearrowright \nu \rangle, \quad \text{for } \tau, \gamma, \nu \in \mathcal{AT}_D.$$

The aroma $\langle \gamma, \nu \rangle \in \mathcal{A}_D$ is called a stolon and is denoted by a horizontal double edge that connects the roots of the corresponding trees, for example, $\langle \overset{\curvearrowright}{\bullet}, \overset{\curvearrowright}{\vert} \rangle = \overset{\curvearrowright}{\bullet}\!\!\bullet$.

Proposition 2.1 *The tracial stolonic pre-Lie-Rinehart algebra $(\mathcal{AT}_D, \mathcal{A}_D)$ is free.*

The classical Guin-Oudom process [74] extends uniquely a pre-Lie product over a vector space V to a product over the symmetric algebra $S(V)$. We generalize and use this process to extend uniquely the pre-Lie-Rinehart product $\curvearrowright$ over a $\mathcal{A}_D$-module $\mathcal{AT}_D$ to the symmetric algebra $\mathcal{AF}_D := S_{\mathcal{A}_D}(\mathcal{AT}_D)$ of decorated aromatic forests.

Proposition 2.2 *There exists a unique extension of the $\curvearrowright$ product to $\mathcal{AF}_D$ such that*

(i) $\omega\mathbf{1} \curvearrowright \nu = \omega\nu, \quad$ for $\omega \in \mathcal{A}_D, \nu \in \mathcal{A}_D \sqcup \mathcal{AT}_D$,

(ii) $(\tau \cdot \pi) \curvearrowright \nu = \tau \curvearrowright (\pi \curvearrowright \nu) - (\tau \curvearrowright \pi) \curvearrowright \nu, \quad$ for $\tau \in \mathcal{AT}_D, \pi \in \mathcal{AF}_D$,

(iii) $\pi \curvearrowright (\mu_1 \cdot \mu_2) = \sum_{(\pi)} (\pi_{(1)} \curvearrowright \mu_1) \cdot (\pi_{(2)} \curvearrowright \mu_2), \quad$ for $\mu_1, \mu_2 \in \mathcal{AF}_D$,

with deshuffle coproduct $\Delta_{\mathcal{A}_D}(\pi) = \sum_{(\pi)} \pi_{(1)} \otimes_{\mathcal{A}_D} \pi_{(2)}$.

We follow the structure of the proof of Proposition 2.7 of [74] which proves an analogous statement for $S(\mathcal{AT}_D)$. We check that the relations $(i), (ii), (iii)$ are well-defined over $S_{\mathcal{A}_D}(\mathcal{AT}_D)$.

Proof It follows from (i), (iii), and the coassociativity of $\Delta_{\mathcal{A}_D}$ that

$$\tau \curvearrowright \mathbf{1} = 0, \quad \text{and} \quad \tau \curvearrowright (\pi_1 \cdots \pi_n) = \sum_{k=1}^{n} \pi_1 \cdots (\tau \curvearrowright \pi_k) \cdots \pi_n.$$

The relation (ii) is well-defined with respect to the choice of τ using the Lemma 2.5 of [74] which is based on induction on the length of the monomial and the pre-Lie relation. This means that $(\tau \cdot \pi) \curvearrowright \nu$ is well-defined for $\tau \cdot \pi \in S(\mathcal{AT}_D)$ using (i) and (ii). Let $\mathcal{J}$ be an ideal of $S(\mathcal{AT}_D)$

$$\mathcal{J} := \langle (\omega\tau \cdot \gamma - \tau \cdot \omega\gamma) \cdot \pi : \omega \in \mathcal{A}_D, \tau, \gamma \in \mathcal{AT}_D, \pi \in S(\mathcal{AT}_D) \rangle.$$

It remains to show that $\mathcal{J} \curvearrowright \nu = 0$ which follows from the property $(\omega\tau \cdot \pi) \curvearrowright \nu = \omega(\tau \cdot \pi \curvearrowright \nu)$, with $\omega \in \mathcal{A}_D$, proved by induction on the length of the monomial π. The initial step is shown below for $\gamma \in \mathcal{AT}_D$ using the $\mathcal{A}_D$-linearity in the left operand of $\curvearrowright$:

$$(\omega\tau \cdot \gamma) \curvearrowright \nu = \omega\tau \curvearrowright (\gamma \curvearrowright \nu) - (\omega\tau \curvearrowright \gamma) \curvearrowright \nu = \omega(\tau \cdot \gamma \curvearrowright \nu).$$

5

Assume the property to be true for monomials shorter than $\pi = \tau_1 \cdots \tau_n \in S(\mathcal{AT}_D)$ and recall that

$$\omega\tau \curvearrowright \pi = \sum_{k=1}^{n} \omega(\tau \curvearrowright \tau_k) \cdot \pi_{\hat{k}}, \quad \text{where } \pi_{\hat{k}} := \tau_1 \cdots \tau_{k-1}\tau_{k+1} \cdots \tau_n.$$

Then, by induction, we have,

$$(\omega\tau \curvearrowright \pi) \curvearrowright \nu = \Big(\sum_{k=1}^{n} \omega(\tau \curvearrowright \tau_k) \cdot \pi_{\hat{k}} \Big) \curvearrowright \nu = \omega\Big(\sum_{k=1}^{n} (\tau \curvearrowright \tau_k) \cdot \pi_{\hat{k}} \Big) \curvearrowright \nu = \omega((\tau \curvearrowright \pi) \curvearrowright \nu).$$

This allows us to prove the inductive step:

$$(\omega\tau \cdot \pi) \curvearrowright \nu = \omega\tau \curvearrowright (\pi \curvearrowright \nu) - (\omega\tau \curvearrowright \pi) \curvearrowright \nu = \omega(\tau \cdot \pi \curvearrowright \nu).$$

Therefore, relations (i) and (ii) extend $\curvearrowright$ to $S_{\mathcal{A}_D}(\mathcal{AT}_D) \otimes (\mathcal{A}_D \oplus \mathcal{AT}_D) \to (\mathcal{A}_D \oplus \mathcal{AT}_D)$. Due to the Leibniz rule, the property $(\omega\pi) \curvearrowright \nu = \omega(\pi \curvearrowright \nu)$, the cocommutativity and coassociativity of $\Delta_{\mathcal{A}_D}$, the relation (iii) is well-defined. Therefore, it defines $\curvearrowright$ on $\mathcal{AF}_D = S_{\mathcal{A}_D}(\mathcal{AT}_D)$. $\qquad \square$

Following the D-algebra structure from [72], see also [63, 64, 65], we define the tracial commutative D-algebra $(\mathcal{AF}_D, \curvearrowright)$ of decorated aromatic forests graded by the number of roots.

Definition 2.3 *Let $(A, \cdot)$ be a unital commutative graded algebra with unit $\mathbf{1}$. Let A be equipped with a non-associative product $\curvearrowright$ and let $x \in A_1$ and $a, b \in A$ satisfy the relation*

$$x \curvearrowright (a \cdot b) = (x \curvearrowright a) \cdot b + a \cdot (x \curvearrowright b).$$

The triple $(A, \cdot, \curvearrowright)$ is a commutative D-algebra if the following identities are satisfied

$$\mathbf{1} \curvearrowright a = a,$$
$$a \curvearrowright x \in A_1,$$
$$(\omega \cdot a) \curvearrowright b = \omega \cdot (a \curvearrowright b),$$
$$(x \cdot a) \curvearrowright b = x \curvearrowright (a \curvearrowright b) - (x \curvearrowright a) \curvearrowright b,$$

for $\omega \in A_0$, $x \in A_1$, $a, b \in A$. It is called tracial if there exists a trace $t : El_{A_0}(A_1) \to A_0$.

For a tracial commutative D-algebra A, the pair (A_1, A_0) is a tracial pre-Lie-Rinehart algebra. This implies that $(\mathcal{AF}_D, \curvearrowright)$ is the free tracial commutative D-algebra since $(\mathcal{AT}_D, \mathcal{A}_D)$ is the free tracial pre-Lie-Rinehart algebra and $(\mathcal{AF}_D, \curvearrowright)$ is obtained uniquely using the Guin-Oudom process. We extend the structure of the D-algebra $(\mathcal{AF}_D, \curvearrowright)$ with the $\mathcal{A}_D$-bilinear form $\langle -, - \rangle : \mathcal{AT}_D \otimes \mathcal{AT}_D \to \mathcal{A}_D$ and see that it remains free.

A map $\varphi : A \to A'$ between two tracial commutative D-algebras A and A' is a D-algebra morphism if $\varphi(A_1) \subset A'_1$ and

$$\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b), \quad \varphi(a \curvearrowright b) = \varphi(a) \curvearrowright \varphi(b),$$
$$\varphi(\langle x, y \rangle) = \langle \varphi(x), \varphi(y) \rangle, \quad \varphi(t(\tilde{x})) = t(\varphi(\tilde{x})),$$

for $a, b \in A$, $x, y \in A_1$, $\tilde{x} \in El_{A_0}(A_1)$.

Example 2.4 *Let $\mathcal{X}$ be the space of vector fields $\mathbb{R}^d \to \mathbb{R}^d$. The symmetric algebra $S_{\mathcal{C}^\infty(\mathbb{R}^d)}(\mathcal{X})$ over the ring of $\mathcal{C}^\infty(\mathbb{R}^d)$ maps is a commutative tracial D-algebra. The non-associative product is given by the covariant derivation, for example, let $f, g, h : \mathbb{R}^d \to \mathbb{R}^d$, then,*

$$(fg)[h] = \sum_{i,j=1}^d f^i g^j h_{ij}, \quad \text{where } h_{ij} := \frac{\partial^2 h}{\partial x_i \partial x_j}.$$

Divergence of a vector field is $d(f) = \sum_{i=1}^d f_i^i$ and the bilinear product is the inner product, i.e., $\langle f, g \rangle = \sum_{i=1}^d f^i g^i$.

We note that the algebra of endomorphisms generated by $\pi \curvearrowright - : \mathcal{AF}_D \to \mathcal{AF}_D$ for $\pi \in \mathcal{AF}_D$ is isomorphic to the algebra $(\mathcal{AF}_D, \diamond)$ with $\diamond$ denoting the Grossman-Larson product defined as

$$\pi_1 \curvearrowright (\pi_2 \curvearrowright -) = (\pi_1 \diamond \pi_2) \curvearrowright -, \quad \pi_1, \pi_2 \in \mathcal{AF}_D.$$

Let us obtain a Grossman-Larson $\mathcal{A}/_\mathbb{R}$-bialgebra [70] from the commutative D-algebra $\mathcal{AF}_D$.

Proposition 2.5 *Let $\epsilon_{\mathcal{A}_D} : \mathcal{AF}_D \to \mathcal{A}_D$ be a $\mathcal{A}_D$-linear map such that $\epsilon_{\mathcal{A}_D}(\mathbf{1}) = \mathbf{1}$ and $\epsilon_{\mathcal{A}_D}(\pi) = 0$ for $\pi \notin \mathcal{AF}_{D,0}$, then $B_{GL} := (\mathcal{AF}_D, \diamond, \mathbf{1}, \Delta_{\mathcal{A}_D}, \epsilon_{\mathcal{A}_D})$ is a $\mathcal{A}/_\mathbb{R}$-bialgebra [70] called Grossman-Larson $\mathcal{A}/_\mathbb{R}$-bialgebra. That is, it satisfies the following properties:*

(1) $\epsilon_{\mathcal{A}_D}(\mathbf{1}) = \mathbf{1}$,

(2) $\Delta_{\mathcal{A}_D}(\mathbf{1}) = \mathbf{1} \otimes_{\mathcal{A}_D} \mathbf{1}$,

(3) $\epsilon_{\mathcal{A}_D}(\pi \diamond \mu) = \epsilon_{\mathcal{A}_D}(\pi \diamond \epsilon_{\mathcal{A}_D}(\mu))$,

(4) $\Delta_{\mathcal{A}_D}(\pi \diamond \mu) = \Delta_{\mathcal{A}_D}(\pi) \diamond_\otimes \Delta_{\mathcal{A}_D}(\mu)$,

with $(\pi_{(1)} \otimes_{\mathcal{A}_D} \pi_{(2)}) \diamond_\otimes (\mu_{(1)} \otimes_{\mathcal{A}_D} \mu_{(2)}) = (\pi_{(1)} \diamond \mu_{(1)}) \otimes_{\mathcal{A}_D} (\pi_{(2)} \diamond \mu_{(2)})$.

Proof Since Grossman-Larson product is isomorphic to the composition of maps, it is associative. The properties (1) and (2) follow from the definition of the counit and coproduct. Property (4) can be proven using the definition of the Grossman-Larson product together with the relation *(iii)* from Proposition 2.2. To prove the property (3), we note that $\epsilon_{\mathcal{A}_D}(\pi_1 \diamond \pi_2)$ is non-zero if and only if $\pi_2 \in \mathcal{A}_D$ and is equal to the aromas obtained by grafting all trees of π_1 onto π_2 in all possible ways. $\square$

Proposition 2.5 can be applied to a general commutative D-algebra if the associativity of the Grossman-Larson product is proven analogously to [74, Lemma 2.10]. We note that if we exclude all aromas, i.e. $\mathcal{A}_D = \{\mathbf{1}\}$, the $\mathcal{A}/_\mathbb{R}$-bialgebra B_{GL} is a graded connected bialgebra which is also a Hopf algebra. We also note that the product $\diamond_\otimes$ is well-defined over $\Delta_{\mathcal{A}_D}(\mathcal{AF}_D) \subset \mathcal{AF}_D \otimes_{\mathcal{A}_D} \mathcal{AF}_D$, but not on $\mathcal{AF}_D \otimes_{\mathcal{A}_D} \mathcal{AF}_D$ since $\diamond$ is not $\mathcal{A}_D$-linear in its right operand.

Remark 2.6 *We say that B_{GL} is cocomplete since $B_{GL} = \bigcup_{n=0}^\infty B_{GL,n}$ and is a free $\mathcal{A}_D$-module. An analog of the Cartier-Milnor-Moore theorem is proven in [70] which states that a cocomplete and graded projective $\mathcal{A}/_\mathbb{R}$-bialgebra is the universal enveloping Lie-Rinehart algebra of the Lie-Rinehart algebra of its primitive elements. The freeness of the $\mathcal{A}_D$-module B_{GL} implies its graded projectiveness, therefore, the $\mathcal{A}/_\mathbb{R}$-bialgebra B_{GL} is the universal enveloping Lie-Rinehart algebra of the pre-Lie-Rinehart algebra $(\mathcal{AT}_D, \mathcal{A}_D)$.*

Remark 2.7 *If we replace the coalgebra structure of B_{GL} by the deshuffle coproduct $\Delta : \mathcal{AF}_D \to \mathcal{AF}_D \otimes \mathcal{AF}_D$ and $\epsilon : \mathcal{AF}_D \to \mathbb{R}$, then we get a Grossman-Larson Hopf algebra dual (up to the symmetry coefficients) to the Hopf algebra mentioned in [9, Thm. 4.4].*

$\mathcal{A}/_{\mathbb{R}}$-bialgebras are also called cocommutative bialgebroids. We show that B_{GL} defines a Hopf algebroid as introduced in [62].

Proposition 2.8 *Let $S_\diamond : (\mathcal{AF}_D, \diamond) \to (\mathcal{AF}_D, \diamond)$ be an algebra anti-isomorphism defined as*

(i) $S_\diamond(\omega) := \omega, \quad S_\diamond(\omega\pi) := S_\diamond(\pi) \diamond \omega, \quad for\ \omega \in \mathcal{A}_D, \pi \in \mathcal{AF}_D,$

(ii) $S_\diamond(\tau) := -\tau, \quad S_\diamond(\tau\pi) := -S_\diamond(\pi) \diamond \tau - S_\diamond(\tau \curvearrowright \pi), \quad for\ \tau \in \mathcal{T}_D.$

Then, $H_{GL} := (B_{GL}, S_\diamond)$ is the Grossman-Larson Hopf algebroid with $S_\diamond$ being called an antipode and satisfying the following conditions where $\pi \in \mathcal{AF}_D$ and $\Delta_{\mathcal{A}_D}(\pi) = \sum_{(\pi)} \pi_{(1)} \otimes_{\mathcal{A}_D} \pi_{(2)}$,

(1) $\sum_{(\pi)} S_\diamond(\pi_{(1)}) \diamond \pi_{(2)} = \mathbf{1}\epsilon_{\mathcal{A}_D}(S_\diamond(\pi)),$

(2) $\sum_{(\pi)} \hat{\pi}_{(1)} \diamond S_\diamond(\hat{\pi}_{(2)}) = \mathbf{1}\epsilon_{\mathcal{A}_D}(\pi),\ with\ \gamma(\pi_{(1)} \otimes_{\mathcal{A}_D} \pi_{(2)}) = \hat{\pi}_{(1)} \otimes \hat{\pi}_{(2)},$

where γ is the section of the projection $P : \mathcal{AF}_D \otimes \mathcal{AF}_D \to \mathcal{AF}_D \otimes_{\mathcal{A}_D} \mathcal{AF}_D$ that places all aromas on the left side of the tensor product, that is, $\hat{\pi}_{(2)}$ has no aromas.

Proof It can be seen that $S_\diamond$ defined this way is an anti-isomorphism due to the associativity of the Grossman-Larson product and the definition which can be rewritten as

$$S_\diamond(\omega \diamond \pi) = S_\diamond(\pi) \diamond S_\diamond(\omega), \quad S_\diamond(\tau \diamond \pi) = S_\diamond(\pi) \diamond S_\diamond(\tau).$$

Let us prove that $S_\diamond$ satisfies the condition (1) and (2). We start with (1) by noting that $S_\diamond(-) \diamond - : \mathcal{AF}_D \otimes_{\mathcal{A}_D} \mathcal{AF}_D \to \mathcal{AF}_D$ is well-defined since

$$S_\diamond(\omega\pi_{(1)}) \diamond \pi_{(2)} = S_\diamond(\pi_{(1)}) \diamond \omega \diamond \pi_{(2)} = S_\diamond(\pi_{(1)}) \diamond \omega\pi_{(2)}.$$

We see that (1) is satisfied for $\pi = \omega \in \mathcal{A}_D$. Let us check that (1) is satisfied for $\pi = \omega\tau \in \mathcal{AT}_D$:

$$S_\diamond(\tau) \diamond \omega + S_\diamond(\mathbf{1}) \diamond \omega\tau = -\tau\omega - \tau \curvearrowright \omega + \omega\tau = S_\diamond(\tau) \curvearrowright \omega = \mathbf{1}\epsilon_{\mathcal{A}_D}(S_\diamond(\omega\tau)),$$

where $\tau \in \mathcal{T}_D$. Let $\tau \in \mathcal{T}_D, \pi \in \mathcal{F}_D, \omega \in \mathcal{A}_D$, then, $\tau\pi\omega = \tau \diamond \pi\omega - \tau \curvearrowright \pi\omega$. We use induction on the length of π and assume that the condition (1) is satisfied for $\tau \curvearrowright \pi\omega$. We check that the left-hand side of (1) applied to $\tau \diamond \pi\omega$ is 0:

$$\sum_{(\tau \diamond \pi\omega)} S_\diamond((\tau \diamond \pi\omega)_{(1)}) \diamond (\tau \diamond \pi\omega)_{(2)} = \sum_{(\tau),(\pi\omega)} S_\diamond(\tau_{(1)} \diamond \pi_{(1)}) \diamond \tau_{(2)} \diamond \pi_{(2)}\omega$$

$$= \sum_{(\pi\omega)} S_\diamond(\pi_{(1)}) \diamond \Big(\sum_{(\tau)} S_\diamond(\tau_{(1)}) \diamond \tau_{(2)} \Big) \diamond \pi_{(2)}\omega = 0,$$

since $\sum_{(\tau)} S_\diamond(\tau_{(1)}) \diamond \tau_{(2)} = 0$. This implies that left-hand side of (1) applied to $\tau\pi\omega$ is

$$\sum_{(\tau\pi\omega)} S_\diamond((\tau\pi\omega)_{(1)}) \diamond (\tau\pi\omega)_{(2)} = -\mathbf{1}\epsilon_{\mathcal{A}_D}(S_\diamond(\tau \curvearrowright \pi\omega))$$

$$= \mathbf{1}\epsilon_{\mathcal{A}_D}(S_\diamond(\tau\pi\omega)) + \mathbf{1}\epsilon_{\mathcal{A}_D}(S_\diamond(\pi\omega) \diamond \tau) = \mathbf{1}\epsilon_{\mathcal{A}_D}(S_\diamond(\tau\pi\omega)).$$

This proves that $S_\diamond$ satisfies the condition (1). To prove condition (2), we recall that the Grossman-Larson product is $\mathcal{A}_D$-linear in its left operand and that γ places all aromas on the left side of the tensor product. This implies that both sides of the condition (2) are $\mathcal{A}_D$-linear and the condition is reduced to the analogous condition over the Grossman-Larson Hopf algebra over $\mathcal{F}_D$. This proves that $S_\diamond$ satisfies the condition (2). $\qquad\square$

The space of decorated aromatic forests with commutative product $\cdot$ and deshuffle coproduct forms a Hopf algebra $H := (\mathcal{AF}_D, \cdot, \mathbf{1}, \Delta_{\mathcal{A}_D}, \epsilon_{\mathcal{A}_D}, S)$. The Hopf algebra H together with the product $\curvearrowright: H \otimes H \to H$ forms a pre-Hopf algebroid which is a generalization of the pre-Hopf algebra [60] that satisfies the following conditions for $\pi, \mu, \eta \in H$, a map $\beta: H \to H$, and a section γ of the projection $P: H \otimes H \to H \otimes_{\mathcal{A}_D} H$ with $\gamma(\pi_{(1)} \otimes_{\mathcal{A}_D} \pi_{(2)}) = \hat{\pi}_{(1)} \otimes \hat{\pi}_{(2)}$:

(1) $\pi \curvearrowright (\mu \cdot \eta) = (\pi_{(1)} \curvearrowright \mu) \cdot (\pi_{(2)} \curvearrowright \eta),$ (3) $\hat{\pi}_{(1)} \curvearrowright (\beta(\hat{\pi}_{(2)}) \curvearrowright -) = \epsilon_{\mathcal{A}_D}(\pi) \, \mathrm{id},$

(2) $\pi \curvearrowright (\mu \curvearrowright \eta) = (\pi_{(1)} \cdot (\pi_{(2)} \curvearrowright \mu)) \curvearrowright \eta,$ (4) $\beta(\pi_{(1)}) \curvearrowright (\pi_{(2)} \curvearrowright -) = \epsilon_{\mathcal{A}_D}(\beta(\pi)) \, \mathrm{id}.$

Conditions (1) and (2) follow from the definition of the D-algebra and the conditions (3) and (4) are satisfied for β and γ being the anitipode $S_\diamond$ and γ from Proposition 2.8. We present an alternative proof for (8) of [60, Lemma 2.3].

Lemma 2.9 *For all $\pi, \mu \in H$, we have $\pi \curvearrowright S(\mu) = S(\pi \curvearrowright \mu)$.*

Proof Consider $\mu = \tau \in \mathcal{AT}_D$, then $\pi \curvearrowright \tau \in \mathcal{AT}_D$ and $\pi \curvearrowright S(\tau) = -\pi \curvearrowright \tau = S(\pi \curvearrowright \tau)$. Assume the statement is true for all monomials shorter than $\mu = \tau_1 \cdots \tau_n$ for $\tau_i \in \mathcal{AT}_D$, then,

$$S(\pi \curvearrowright \tau_1 \cdots \tau_n) = -\sum_{(\pi)} (\pi_{(1)} \curvearrowright \tau_1) S(\pi_{(2)} \curvearrowright \tau_2 \cdots \tau_n)$$
$$= -\sum_{(\pi)} (\pi_{(1)} \curvearrowright \tau_1)(\pi_{(2)} \curvearrowright S(\tau_2 \cdots \tau_n))$$
$$= -\pi \curvearrowright (\tau_1 S(\tau_2 \cdots \tau_n)) = \pi \curvearrowright S(\mu).$$

This finishes the proof. $\qquad\square$

We check that the subadjacent Hopf algebra with antipode $\hat{S}_\diamond$ defined in [60, Thm. 2.4] corresponds to the Grossman-Larson Hopf algebroid by showing that the antipodes coincide.

Lemma 2.10 *Let $\hat{S}_\diamond$ be the antipode defined in [60, Thm. 2.4], that is,*

$$\hat{S}_\diamond(\pi) := \sum_{(\pi)} \hat{S}_\diamond(\pi_{(1)}) \curvearrowright S(\pi_{(2)}), \quad for \ \pi \in H.$$

Then, $\hat{S}_\diamond = S_\diamond$ where $S_\diamond$ is defined in Proposition 2.8.

Proof We see that $\hat{S}_\diamond(\omega) = \omega$ and $\hat{S}_\diamond(\tau) = -\tau$ for $\omega \in \mathcal{A}_D$ and $\tau \in \mathcal{T}_D$. We check by induction and using the fact that $S_\diamond$ is a coalgebra homomorphism that $\hat{S}_\diamond(\omega\pi) = S_\diamond(\omega\pi)$ for $\pi \in \mathcal{AF}_D$:

$$\hat{S}_\diamond(\omega\pi) = \sum_{(\pi)} S_\diamond(\pi_{(1)}) \curvearrowright S(\omega\pi_{(2)}) = \sum_{(\pi)^2} (S_\diamond(\pi_{(1)}) \curvearrowright \omega)(S_\diamond(\pi_{(2)}) \curvearrowright S(\pi_{(3)}))$$
$$= \sum_{(\pi)} (S_\diamond(\pi_{(1)}) \curvearrowright \omega) S_\diamond(\pi_{(2)}) = S_\diamond(\pi) \diamond \omega.$$

We use the same properties to show that $\hat{S}_\diamond(\tau\pi) = S_\diamond(\tau\pi)$:

$$\hat{S}_\diamond(\tau\pi) = S_\diamond(\pi_{(1)}) \curvearrowright S(\tau\pi_{(2)}) + S_\diamond(\tau\pi_{(1)}) \curvearrowright S(\pi_{(2)})$$
$$= -S_\diamond(\pi_{(1)}) \curvearrowright \tau S(\pi_{(2)}) - (S_\diamond(\pi_{(1)}) \diamond \tau) \curvearrowright S(\pi_{(2)}) - S_\diamond(\tau \curvearrowright \pi_{(1)}) \curvearrowright S(\pi_{(2)})$$
$$= -S_\diamond(\pi) \diamond \tau - S_\diamond(\pi_{(1)}) \curvearrowright S(\tau \curvearrowright \pi_{(2)}) - S_\diamond(\tau \curvearrowright \pi_{(1)}) \curvearrowright S(\pi_{(2)})$$
$$= -S_\diamond(\pi) \diamond \tau - S_\diamond(\tau \curvearrowright \pi),$$

where we omit writting the sums to simplify the notation. Therefore, $\hat{S}_\diamond = S_\diamond$ following the definition from Proposition 2.8. $\qquad\square$

Following [7, Def. 1.1], the Hopf algebra H together with the Hopf algebroid H_{GL} forms a generalization of the Hopf brace, that is, the following compatibility condition is satisfied

$$\pi \diamond (\mu \cdot \eta) = \sum_{(\pi)^2} (\pi_{(1)} \diamond \mu) \cdot S(\pi_{(2)}) \cdot (\pi_{(3)} \diamond \eta),$$

with $\pi, \mu, \eta \in \mathcal{AF}_D$ and $\Delta^2(\pi) = \sum_{(\pi)^2} \pi_{(1)} \otimes \pi_{(2)} \otimes \pi_{(3)}$. A proof can be found in [60, Thm. 2.13].

2.2 Decorated aromatic S-series

We use the universal property of $(\mathcal{AF}_D, \curvearrowright)$ to define the unique morphism F_D from the tracial commutative D-algebra of decorated aromatic forests generated by D to the tracial commutative D-algebra $(S_{\mathcal{C}^\infty(\mathbb{R}^d)}(\mathcal{X}_n), -[-])$ of vector fields (Example 2.4) generated by f_d, $d \in D$ with $-[-]$ being the covariant derivation. The morphism F_D is defined as $F_D(d) = f_d$ for $d \in D$.

Let us introduce S-series over decorated aromatic forests. Let $\overline{\mathcal{AF}}_D$ be the space of formal sums of the following form

$$\sum_{\pi \in AF_D} a(\pi)\pi, \quad \text{with } a \in \mathcal{AF}_D^*.$$

It is the completion with respect to the graduation given by the number of vertices. We extend F_D by linearity to $\overline{\mathcal{AF}}_D$. Recall the symmetry coefficient $\sigma(\pi) := |\text{Aut}(\pi)|$. Let $\delta_\sigma : \mathcal{AF}_D^* \to \overline{\mathcal{AF}}_D$ be the isomorphism between the dual and the completion given by

$$\delta_\sigma(a) = \sum_{\pi \in AF_D} \frac{a(\pi)}{\sigma(\pi)}\pi,$$

We will abuse the notation and use δ_σ to denote analogous isomorphisms for other spaces. The space will be made clear from the context and will always be the domain of the functional to which δ_σ is applied.

Let $|.| : AF_D \to \mathbb{N}$ be a map defining a finite grading on AF_D, that we call the order on decorated aromatic forests.

Definition 2.11 *An S-series over decorated aromatic forests is defined for $a \in \mathcal{AF}_D^*$ and some smooth vector fields f_d, $d \in D$, as the following formal power series*

$$S^h_{(f_d)_{d \in D}}(a) := \sum_{\pi \in AF_D} h^{|\pi|} \frac{a(\pi)}{\sigma(\pi)} F_D(\pi), \quad \text{for } a \in \mathcal{AF}_D^*.$$

We often omit the dependency in the f_d and write for simplicity $S^h(a)$, as well as $S(a) = S^1(a)$. S-series are used to represent formal sums of differential operators. We use S-series to represent formal sums of vector fields by using the identification between first order operators and vector fields and by requiring the corresponding functional $a_0 : \mathcal{AF}_D \to \mathbb{R}$ to satisfy

$$\text{supp}(a_0) \subset \mathcal{AT}_D.$$

Such series are called B-series, written $B^h(a_0)$, and a_0 is called an infinitesimal character. B-series are used in deterministic numerical analysis [38] to represent numerical methods $y_1 := \Psi_h(y_0)$ applied to $y' = f(y)$ as $\Psi_h(y_0) = y_0 + B^h(a_0)(y_0)$ for an infinitesimal character a_0. This is typically done by Taylor expanding the numerical method with respect to its time stepsize h.

Given any smooth function $\phi : \mathbb{R}^d \to \mathbb{R}$, the composition of $\Psi_h : \mathbb{R}^d \to \mathbb{R}^d$ with ϕ is represented by the following exponential of B-series:

$$\phi \circ \Psi_h = F_D\Big(\exp^{\cdot}\big(\delta_\sigma(a_0)\big)\Big)[\phi], \quad \text{with } \exp^{\cdot}(x) := 1 + \sum_{k=1}^{\infty} \frac{x^{\cdot k}}{k!}.$$

We use the formalism of aromatic forests and the convolution product $\odot : \mathcal{AF}_D^* \otimes \mathcal{AF}_D^* \to \mathcal{AF}_D^*$ defined as $a \odot b := (a \otimes b) \circ \Delta$ with $\Delta : \mathcal{AF}_D \to \mathcal{AF}_D \otimes \mathcal{AF}_D$ being the deshuffle coproduct (not $\mathcal{A}_D$-linear) over decorated aromatic forests to write the exponential of B-series as an S-series as

$$\phi \circ \Psi_h = S^h(\exp^{\odot}(a_0))[\phi] \quad \text{with } \exp^{\odot}(x) := 1 + \sum_{k=1}^{\infty} \frac{x^{\odot k}}{k!}. \tag{1}$$

We use (1) and Theorem 2.13 to describe the composition of numerical methods that are expanded as B-series.

Definition 2.12 *[9] The Connes-Kreimer coproduct on $\mathcal{AF}_D$ is defined as*

$$\Delta_{CK}(\pi) := \sum_{\pi_0 \subset \pi} (\pi \setminus \pi_0) \otimes \pi_0,$$

where the sum runs over all rooted subforests $\pi_0 \in AF_D$ of π such that $\pi \setminus \pi_0 \in AF_D$ and there are no edges going from π_0 to $\pi \setminus \pi_0$ in π.

It was proved in [12] that the Connes-Kreimer coproduct is dual to the Grossman-Larson product up to the symmetry coefficient. This implies the following Theorem 2.13 which was first proven in [9] for classical aromatic forests and generalized in [12].

Theorem 2.13 (Composition law) *[9, 12] Let $S(a)$ and $S(b)$ be two S-series defined using the same morphism F_D and let ϕ be a test function. Then,*

$$S(a)[S(b)[\phi]] = S(a * b)[\phi], \quad \text{with } a * b = (a \otimes b) \circ \Delta_{CK}.$$

2.3 Decorated clumped forests

The commutative algebra $(\mathcal{AF}_D, \cdot)$ is defined as the symmetric algebra $S_{\mathcal{A}_D}(\mathcal{AT}_D)$ of decorated aromatic trees $\mathcal{AT}_D$ over the ring of decorated aromas $\mathcal{A}_D$. We consider the symmetric algebra $\mathcal{CF}_D := S(\mathcal{AT}_D)$ of decorated aromatic trees $\mathcal{AT}_D$ over the base field $\mathbb{R}$. The elements of $\mathcal{CF}_D$ are called clumped forests and we use the Guin-Oudom process [74] to define the product $\curvearrowright$ on $\mathcal{CF}_D$. The commutative D-algebra $(\mathcal{CF}_D, \curvearrowright)$ that we obtain in this way has $\mathcal{CF}_{D,0} = \{\mathbf{1}\}$ and is in many ways similar to the commutative D-algebra of classical forests $(\mathcal{F}_D, \curvearrowright)$. We note that $\overset{\circ}{\bullet} \cdot \overset{\bullet}{\bullet} \neq \overset{\bullet}{\bullet} \cdot \overset{\circ}{\bullet}$ in $\mathcal{CF}_D$. Some elements of $\mathcal{CF}_D$ are listed below:

$$\overset{\circ}{\bullet} \cdot \overset{\vee}{\bullet}, \quad \overset{\bullet}{\bullet} \cdot \overset{\circ}{\bullet}\overset{\vee}{}, \quad \overset{\circ}{\bullet}\overset{\vee}{} \cdot \overset{\bullet}{\bullet} \cdot \overset{\vee}{\bullet} \cdot \overset{\bullet}{\bullet}.$$

Decorated clumped forests have a convenient algebraic structure described in the following result.

Theorem 2.14 *The Grossman-Larson $\mathcal{A}/_\mathbb{R}$-bialgebra of $(\mathcal{CF}_D, \curvearrowright)$ is a Hopf algebra dual up to the symmetry to the Connes-Kreimer Hopf algebra over clumped forests.*

Proof Since $\mathcal{CF}_{D,0} = \{\mathbf{1}\}$, the $\mathcal{A}/\mathbb{R}$-bialgebra structure reduces to a graded connected bialgebra, that is, to a Hopf algebra. Its duality to the corresponding Connes-Kreimer Hopf algebra over clumped forests can be seen by following the proof for classical forests from [42]. $\square$

Let us consider the following inner product for $\pi, \mu \in \mathcal{CF}_D$ or $\mathcal{AF}_D$:

$$\langle \pi, \mu \rangle_\sigma := \begin{cases} \sigma(\pi), & \text{if } \pi = \mu, \\ 0, & \text{otherwise.} \end{cases}$$

We use it to obtain the following duality between concatenation product and deshuffle coproduct.

Lemma 2.15 *Let $\cdot$ be the concatenation product and let $a, b \in \mathcal{CF}_D^*$ or $\mathcal{AF}_D^*$, then,*

$$\delta_\sigma(a) \cdot \delta_\sigma(b) = \delta_\sigma(a \odot b), \quad \text{where } a \odot b = (a \otimes b) \circ \Delta,$$

where $\Delta : \mathcal{CF}_D \to \mathcal{CF}_D \otimes \mathcal{CF}_D$ or $\Delta : \mathcal{AF}_D \to \mathcal{AF}_D \otimes \mathcal{AF}_D$ is the $\mathbb{R}$-linear deshuffle coproduct.

Proof We check that $\langle \pi \cdot \mu, \eta \rangle_\sigma = \langle \pi \otimes \mu, \Delta(\eta) \rangle_\sigma$ for $\pi, \mu, \eta \in \mathcal{CF}_D$ or $\mathcal{AF}_D$. $\square$

Let the map $\Phi : (\mathcal{CF}_D, \curvearrowright) \to (\mathcal{AF}_D, \curvearrowright)$ be a commutative D-algebra morphism that "forgets" the clumping, for example,

$$\Phi(\overset{\circ}{\bullet} \cdot \overset{\circ}{\bullet}) = \Phi(\bullet \cdot \overset{\circ}{\bullet}) = \overset{\circ}{\bullet} \bullet \overset{\circ}{\bullet}.$$

We define $\Phi^* : \mathcal{AF}_D \to \mathcal{CF}_D$ as $(\Phi \circ \delta_\sigma)(a) = \delta_\sigma(a \circ \Phi^*)$ with $a \in \mathcal{CF}_D^*$, in particular, Φ^* is the adjoint of Φ with respect to the $\langle -, - \rangle_\sigma$ inner product. Let us consider the exponential maps

$$\exp^{\cdot} : \mathcal{AT}_D \to \mathcal{AF}_D, \quad \exp^{\odot} : \mathcal{AT}_D^* \to \mathcal{AF}_D^*,$$
$$\exp_C^{\cdot} : \mathcal{AT}_D \to \mathcal{CF}_D, \quad \exp_C^{\odot} : \mathcal{AT}_D^* \to \mathcal{CF}_D^*.$$

By using Lemma 2.15, we obtain for $a_0 \in \mathcal{AT}_D^*$ the following identities

$$\exp^{\cdot}(\delta_\sigma(a_0)) = \delta_\sigma(\exp^{\odot}(a_0)), \quad \exp_C^{\cdot}(\delta_\sigma(a_0)) = \delta_\sigma(\exp_C^{\odot}(a_0)),$$

with the functionals $\exp^{\odot}(a_0)$ and $\exp_C^{\odot}(a_0)$ being characterized in Propositon 2.16.

Proposition 2.16 *Let $a_0 \in \mathcal{AT}^*$ be an infinitisimal character and let*

$$a_e := \exp^{\odot}(a_0), \quad a_c := \exp_C^{\odot}(a_0).$$

Then, $a_c \in \mathcal{CF}_D^*$ is a character of $(\mathcal{CF}_D, \cdot)$ and $a_e = a_c \circ \Phi^*$ with $a_e \in \mathcal{AF}_D^*$.

Proof We prove that $a_c := \exp_C^{\odot}(a_0)$ is a character of $\mathcal{CF}_D$ by considering a decorated clumped forest $\pi = \tau_1 \cdots \tau_n \in \mathcal{CF}_D$ where $\tau_i \in \mathcal{AT}_D$. Then,

$$\exp_C^{\odot}(a_0)(\pi) = \frac{1}{n!} \cdot (a_0 \otimes \cdots \otimes a_0)(\Delta^{n-1}(\pi))$$
$$= \frac{1}{n!} \cdot \sum_{\sigma \in S_n} a_0(\tau_{\sigma(1)}) \cdots a_0(\tau_{\sigma(n)}) = a_0(\tau_1) \cdots a_0(\tau_n).$$

We prove $a_e = a_c \circ \Phi^*$ by using the identity $\exp^{\cdot}(\delta_\sigma(a_0)) = \Phi\big(\exp_C^{\cdot}(\delta_\sigma(a_0))\big)$. $\square$

We can define a commutative D-algebra morphism $F_D : (\mathcal{CF}_D, \curlywedge) \to (S_{\mathcal{C}^\infty(\mathbb{R}^d)}(\mathcal{X}_n), -[-])$ as $F_D := F_D \circ \Phi$ where we use the same notation for the morphism over clumped and aromatic forests. This way, we obtain S-series over decorated clumped forests. We note that, following the definition of Φ^*, any S-series $S(a)$ with $a \in \mathcal{CF}_D^*$ is identical to the S-series over decorated aromatic forests $S(a \circ \Phi^*)$. Moreover, given any functional $a \in \mathcal{AF}_D^*$, there exists a functional $a_C \in \mathcal{CF}_D^*$ such that $a = a_C \circ \Phi^*$, since Φ^* is injective. A possible definition of a_C is

$$a_C(\pi) = \frac{1}{n^m} a(\Phi(\pi)), \quad \text{for } \pi \in \mathcal{CF}_D,$$

where n is the number of rooted components and m is the number of aromas.

We note that a tree $\tau \in \mathcal{T}_D$ induces a map $\tau' : \mathcal{A}_D \to \mathcal{AT}_D$ with $\tau'(\omega) = \omega\tau$ for $\omega \in \mathcal{A}_D$. We can extend $(-)'$ to $\mathcal{F}_D$ in two possible ways, that is, for a $\pi \in \mathcal{F}_D$, we have two maps,

$$\pi' : \mathcal{A}_D \to \mathcal{AF}_D, \quad \pi'' : \mathcal{A}_D \to \mathcal{CF}_D,$$

defined as

$$(\pi \cdot \mu)'(\omega) = \omega\pi\mu, \quad (\pi \cdot \mu)''(\omega) = (\pi'' \odot \mu'')(\omega),$$

with $\pi'' \odot \mu'' := \cdot \circ (\pi'' \otimes \mu'') \circ \Delta$ where Δ is a deshuffle coproduct over $\mathcal{A}_D$. For example,

$$(\bullet\ \mathbf{I})'(\mathbf{O}) = \mathbf{O}\bullet\ \mathbf{I}, \quad (\bullet\ \mathbf{I})''(\mathbf{O}) = \mathbf{O}\bullet\cdot\mathbf{I} + \bullet\cdot\mathbf{O}\mathbf{I}.$$

The following result presents a convenient method to compute Φ^*.

Proposition 2.17 *Given a decorated forest $\pi \in \mathcal{F}_D$ and $\omega \in \mathcal{A}_D$, we have the following identity*

$$\Phi^*(\pi'(\omega)) = \pi''(\omega).$$

Proof We take the adjoints of π' and π'' with respect to the $\langle -, - \rangle_\sigma$ inner product and denote them by π'^* and π''^*. Then, $\pi'^*(\omega\pi) = \sigma(\pi)\omega$ where $\omega \in \mathcal{A}_D$ and $\pi \in \mathcal{F}_D$. To prove the statement, we have to show that $\pi''^*(\pi_\omega) = \sigma(\pi)\omega$ where $\pi_\omega \in \mathcal{CF}_D$ is a decorated clumped forest that occurs as a term in $\pi''(\omega)$, that is, we need to show that $\pi'^*(\Phi(\pi_\omega)) = \pi''^*(\pi_\omega)$.

We note that for $\tau \in \mathcal{T}_D, \tau''^*(\omega) = \tau'^*(\omega) = \sigma(\tau)\omega$ and the statement is true. We use an inductive assumption and Lemma 2.15 to obtain

$$(\pi \cdot \mu)''^*(\eta_\omega) = (\cdot \circ (\pi''^* \otimes \mu''^*) \circ \Delta)(\eta_\omega) = \sigma(\pi)\sigma(\mu)|A|\omega,$$

where A is the set of $\eta_{\omega,(1)} \otimes \eta_{\omega,(2)}$ such that the rooted components of $\eta_{\omega,(1)}$ and $\eta_{\omega,(2)}$ are isomorphic to π and μ, respectively. We note that $\sigma(\pi)\sigma(\mu)|A| = \sigma(\pi \cdot \mu)$ and the proof is finished. $\square$

For example, let $\mathbf{O}\ \rule{0pt}{0pt}\ \bullet\ \mathbf{I} \in \mathcal{AF}_D$, then,

$$\Phi^*(\mathbf{O}\ \rule{0pt}{0pt}\ \bullet\ \mathbf{I}) = \mathbf{O}\ \rule{0pt}{0pt}\ \bullet\cdot\mathbf{I} + \rule{0pt}{0pt}\ \bullet\cdot\mathbf{O}\mathbf{I} + \mathbf{O}\bullet\cdot\ \rule{0pt}{0pt}\ \mathbf{I} + \bullet\cdot\mathbf{O}\ \rule{0pt}{0pt}\ \mathbf{I}.$$

We note that Propositions 2.16 and 2.17 imply the following identity

$$a_e(\omega\pi) = a_c(\pi''(\omega)), \quad \text{for } \pi \in \mathcal{F}_D, \quad \omega \in \mathcal{A}_D,$$

a version of which is presented in [9] for classical aromatic forests.

2.4 Substitution law for decorated aromatic S-series

We introduce the decorated insertion product $\blacktriangleright_d\colon \mathcal{AT}_D \otimes \mathcal{AT}_D \to \mathcal{AT}_D$ which inserts the aromatic tree from the left operand into the vertices decorated by $d \in D$ of the right operand in all possible ways. For example, we recall that $\mathrel{\vcenter{\hbox{$\circlearrowleft$}}}\,{\vcenter{\hbox{$\mathrm{tree}$}}} = \mathrm{div}(\circ)(\circ \curvearrowright \circ) \curvearrowright \bullet$, then,

$$\mathrel{\vcenter{\hbox{$\circlearrowleft$}}}\, \blacktriangleright_\circ\, \mathrel{\vcenter{\hbox{$\circlearrowleft$}}}\, = \mathrm{div}(\mathrel{\vcenter{\hbox{$\circlearrowleft$}}})(\circ \curvearrowright \circ) \curvearrowright \bullet + \mathrm{div}(\circ)(\mathrel{\vcenter{\hbox{$\circlearrowleft$}}} \curvearrowright \circ) \curvearrowright \bullet + \mathrm{div}(\circ)(\circ \curvearrowright \mathrel{\vcenter{\hbox{$\circlearrowleft$}}}) \curvearrowright \bullet$$

$$= \mathrel{\vcenter{\hbox{$\circlearrowleft$}}} + \mathrel{\vcenter{\hbox{$\circlearrowleft$}}} + \mathrel{\vcenter{\hbox{$\circlearrowleft$}}} + \mathrel{\vcenter{\hbox{$\circlearrowleft$}}} + \mathrel{\vcenter{\hbox{$\circlearrowleft$}}} + \mathrel{\vcenter{\hbox{$\circlearrowleft$}}} + \mathrel{\vcenter{\hbox{$\circlearrowleft$}}}.$$

The product $\blacktriangleright_d$ generalizes the insertion product studied in [80, 66, 81]. The family $(\mathcal{AT}_D, (\blacktriangleright_d)_{d \in D})$ is a multi-pre-Lie algebra [13, 34], that is, for $\tau, \gamma, \nu \in \mathcal{AT}_D$ and $d, e \in D$ we have

$$\tau \blacktriangleright_d (\gamma \blacktriangleright_e \nu) - (\tau \blacktriangleright_d \gamma) \blacktriangleright_e \nu = \gamma \blacktriangleright_e (\tau \blacktriangleright_d \nu) - (\gamma \blacktriangleright_e \tau) \blacktriangleright_d \nu.$$

Let $\mathcal{AT}_D^{\oplus D} := \mathcal{AT}_D \otimes \mathbb{R}[D] = \bigoplus_{d \in D} \mathcal{AT}_D \iota_d$ where ι_d for $d \in D$ form the basis of $\mathbb{R}[D]$ and let us define the action $\blacktriangleright\colon \mathcal{AT}_D^{\oplus D} \otimes \mathcal{AT}_D \to \mathcal{AT}_D$ by

$$\tau \iota_d \blacktriangleright \gamma = \tau \blacktriangleright_d \gamma.$$

Let us consider $S(\mathcal{AT}_D^{\oplus D})$ which becomes $\mathcal{CF}_D^{\otimes D} := \bigotimes_{d \in D} \mathcal{CF}_D \iota_d$ after we assume the identity $\pi \iota_d \cdot \mu \iota_d = (\pi \cdot \mu)\iota_d$. We check that the Guin-Oudom process for multi-pre-Lie products [34, Thm. 2.4] is well-defined and use it to define

$$\blacktriangleright\colon \mathcal{CF}_D^{\otimes D} \otimes \mathcal{CF}_D \to \mathcal{CF}_D.$$

For example, let $D := \{\bullet, \circ\}$, then, $(\mathrel{\vcenter{\hbox{$\iota_\circ$}}}) \cdot (_\circ {}_\circ \iota_\bullet) \blacktriangleright \mathrel{\vcenter{\hbox{Y}}} = 2\mathrel{\vcenter{\hbox{Y}}} + 4\mathrel{\vcenter{\hbox{Y}}} + 2\mathrel{\vcenter{\hbox{Y}}}$. We note that for $\pi, \mu \in \mathcal{CF}_D$, we have $\pi \iota_d \blacktriangleright \mu = 0$ if the number $|\pi|_{AT}$ of aromatic trees in π is greater than the number $|\mu|_d$ of vertices decorated by d in μ. We define a substitution action $\rhd\colon \mathcal{CF}_D^{\otimes D} \otimes \mathcal{CF}_D \to \mathcal{CF}_D$ in the following way

$$(\otimes_{d \in D} \pi_d \iota_d) \rhd \mu := \begin{cases} (\otimes_{d \in D} \pi_d \iota_d) \blacktriangleright \mu, & \text{if } |\pi_d|_{AT} = |\mu|_d \text{ for all } d \in D, \\ 0, & \text{otherwise.} \end{cases}$$

The substitution action substitutes all vertices in the right operand by the aromatic trees from the left operand. We recall that CF_D is a free D-algebra generated by the set D, therefore, given a map $\varphi\colon D \to \mathcal{AT}_D$, there exists a unique morphism $A_\varphi\colon \mathcal{CF}_D \to \mathcal{CF}_D$. The morphism A_φ can be written using the substitution action.

Lemma 2.18 *Given a morphism $A_\varphi\colon \mathcal{CF}_D \to \mathcal{CF}_D$ that acts on generators D as φ, we have*

$$A_\varphi(\pi) = \otimes_{d \in D} \dot{\exp}_C(\varphi(d)\iota_d) \rhd \pi.$$

Proof Assume $\varphi(d) = \tau_d \in \mathcal{AT}_D$, then,

$$\otimes_{d \in D} \dot{\exp}_C(\tau_d \iota_d) \rhd \pi = \frac{1}{\prod_{d \in D} |\pi|_d!} (\otimes_{d \in D} \tau_d^{|\pi|_d} \iota_d) \blacktriangleright \pi = A_\varphi(\pi),$$

where we use the definition of $\blacktriangleright$. $\qquad\qquad\square$

Let $a^{\otimes D} := \otimes_{d\in D} a_d \iota_d$ with $a_d \in \mathcal{CF}_D^*$ be a functional over $\mathcal{CF}_D^{\otimes D}$ defined as

$$a^{\otimes D}(\otimes_{d\in D}\pi_d\iota_d) = \prod_{d\in D} a_d(\pi_d).$$

For example, let $D = \{\bullet, \circ\}$, then, $(a\iota_\bullet b\iota_\circ)(\overset{\bullet}{\vee}\iota_\bullet \cdot {}_\bullet \overset{\bullet}{\downarrow}\iota_\circ) = a(\overset{\bullet}{\vee}{}_\bullet)b({}_\bullet\overset{\bullet}{\downarrow})$. We define a coaction $\Delta_{CEM} : \mathcal{CF}_D \to \mathcal{CF}_D^{\otimes D} \otimes \mathcal{CF}_D$ and show in Proposition 2.20 that it is dual to the substitution action up to the symmetry coefficient, that is, we have

$$\delta_\sigma(b^{\otimes D} \star a) = \delta_\sigma(b^{\otimes D}) \triangleright \delta_\sigma(a) \quad \text{where } b^{\otimes D} \star a = (b^{\otimes D} \otimes a) \circ \Delta_{CEM},$$

for $a \in \mathcal{CF}_D^*$, $b^{\otimes D} \in \mathcal{CF}_D^{\otimes D*}$, and where we denote $\sigma^{\otimes D}$ by σ for simplicity. The coaction Δ_{CEM} is a generalization of the Calaque-Ebrahimi-Fard-Manchon (CEM) coproduct from [20].

Definition 2.19 *Define the CEM coaction* $\Delta_{CEM} : \mathcal{CF}_D \to \mathcal{CF}_D^{\otimes D} \otimes \mathcal{CF}_D$ *as*

$$\Delta_{CEM}(\pi) = \sum_{\otimes_{d\in D}p_d\iota_d} (\otimes_{d\in D}p_d\iota_d) \otimes \pi/_{\otimes_{d\in D}p_d\iota_d}, \quad \text{for } \pi \in \mathcal{CF}_D,$$

where the sum is over all monomials of disjoint decorated clumped subforests $p_d\iota_d$ that partition π, and $\pi/_{\otimes_{d\in D}p_d\iota_d}$ denotes the decorated clumped forest obtained by contracting the decorated aromatic trees of $p_d\iota_d$ into vertices decorated by d.

For example, let $D = \{\bullet, \circ\}$, then,

$$\Delta_{CEM}(\bullet) = {}_\bullet\iota_\bullet \otimes {}_\bullet + {}_\bullet\iota_\circ \otimes {}_\circ,$$

$$\Delta_{CEM}(\overset{\bullet}{\downarrow}) = \overset{\bullet}{\downarrow}\iota_\bullet \otimes {}_\bullet + \overset{\bullet}{\downarrow}\iota_\circ \otimes {}_\circ + {}_\bullet {}_\circ\iota_\bullet \otimes \overset{\bullet}{\downarrow} + {}_\bullet {}_\circ\iota_\circ \otimes \overset{\circ}{\downarrow}$$

$$\quad\quad {}_\bullet\iota_\bullet \cdot {}_\circ\iota_\circ \otimes \overset{\bullet}{\downarrow} + {}_\bullet\iota_\circ \cdot {}_\circ\iota_\bullet \otimes \overset{\circ}{\downarrow},$$

$$\Delta_{CEM}(\overset{\bullet}{\vee}) = \overset{\bullet}{\vee}\iota_\bullet \otimes {}_\bullet + \overset{\bullet}{\downarrow} {}_\circ\iota_\bullet \otimes \overset{\bullet}{\downarrow} + \overset{\bullet}{\downarrow} {}_\bullet\iota_\bullet \otimes \overset{\bullet}{\downarrow} + {}_\bullet {}_\bullet {}_\circ\iota_\bullet \otimes \overset{\bullet}{\vee} +$$

$$\overset{\bullet}{\vee}\iota_\circ \otimes {}_\circ + \overset{\bullet}{\downarrow} {}_\circ\iota_\circ \otimes \overset{\circ}{\downarrow} + \overset{\bullet}{\downarrow} {}_\bullet\iota_\circ \otimes \overset{\circ}{\downarrow} + {}_\bullet {}_\bullet {}_\circ\iota_\circ \otimes \overset{\circ}{\vee} +$$

$$\overset{\bullet}{\downarrow}\iota_\bullet \cdot {}_\circ\iota_\circ \otimes \overset{\bullet}{\downarrow} + \overset{\bullet}{\downarrow}\iota_\circ \cdot {}_\circ\iota_\bullet \otimes \overset{\circ}{\downarrow} + \overset{\bullet}{\downarrow}\iota_\bullet \cdot {}_\circ\iota_\circ \otimes \overset{\bullet}{\downarrow} + \overset{\bullet}{\downarrow}\iota_\circ \cdot {}_\bullet\iota_\bullet \otimes \overset{\circ}{\downarrow} +$$

$$\bullet\iota_\circ \cdot {}_\bullet {}_\circ\iota_\bullet \otimes \overset{\bullet}{\vee} + {}_\bullet\iota_\circ \cdot {}_\bullet {}_\circ\iota_\bullet \otimes \overset{\bullet}{\vee} + {}_\bullet {}_\bullet\iota_\bullet \cdot {}_\circ\iota_\circ \otimes \overset{\bullet}{\vee} +$$

$$\bullet\iota_\bullet \cdot {}_\bullet {}_\circ\iota_\circ \otimes \overset{\circ}{\vee} + {}_\bullet\iota_\bullet \cdot {}_\bullet {}_\circ\iota_\circ \otimes \overset{\circ}{\vee} + {}_\bullet {}_\bullet\iota_\circ \cdot {}_\circ\iota_\bullet \otimes \overset{\circ}{\vee}.$$

Proposition 2.20 *Let* $a \in \mathcal{CF}_D^*$, $b^{\otimes D} \in \mathcal{CF}_D^{\otimes D*}$, *and* $b^{\otimes D} \star a = (b^{\otimes D} \otimes a) \circ \Delta_{CEM}$, *then,*

$$\delta_\sigma(b^{\otimes D} \star a) = \delta_\sigma(b^{\otimes D}) \triangleright \delta_\sigma(a).$$

Proof We prove the statement in two steps. First, we prove the statement over the space of ordered clumped forests. Ordered clumped forest are forests in which all vertices are totally ordered. Next, we describe the relationships between the product and coproduct over ordered and non-ordered clumped forests and use them to finish the proof.

1. The space $\mathcal{CF}_{DO}$ is defined by assigning every decorated clumped forest π a total order over the vertices. We note that a forest $\pi \in \mathcal{CF}_D$ corresponds to $|\pi|!$ ordered forests in $\mathcal{CF}_{DO}$. The symmetry of any element of $\mathcal{CF}_{DO}$ is equal to 1. Let $\mathcal{CF}_{DO}^{\otimes D} := \bigotimes_{d\in D}\mathcal{CF}_{DO}\iota_d$.

We define $\triangleright^O : \mathcal{CF}_{DO}^{\otimes D} \otimes \mathcal{CF}_D \to \mathcal{CF}_{DO}$ and $\Delta_{CEM}^O : \mathcal{CF}_{DO} \to \mathcal{CF}_{DO}^{\otimes D} \otimes \mathcal{CF}_D$ to be the natural extensions of $\triangleright$ and Δ_{CEM}. The map $\triangleright^O$ substitutes the vertices decorated by d by the trees

of $\pi_d \iota_d$ and chooses a total order between the vertices of π_d and π_l with $d \neq l$ in all possible ways. We note that there are $\left(\sum_{d \in D} |\pi_d| \right)! / \prod_{d \in D} |\pi_d|!$ ways to choose a total order between the vertices of π_d and π_l with $d \neq l$. Let $\pi \in \mathcal{CF}_{DO}^{\otimes D}$ and $\mu \in \mathcal{CF}_D$, then,

$$\pi \rhd \mu = \sum_{\eta \in \mathcal{CF}_{DO}} N(\pi, \mu, \eta) \eta,$$

where $N(\pi, \mu, \eta)$ is the number of ways to substitute the vertices of μ by trees in π to obtain η. We note that there are $|\operatorname{Aut}(\mu)|$ ways to substitute the vertices of μ to obtain the same ordered forest η, therefore, $N(\pi, \mu, \eta) = \sigma(\mu)$. The symmetry is 1 for an ordered clumped forest, and, since all terms in $\Delta_{CEM}^O(\pi)$ have coefficient 1, we obtain

$$\delta_\sigma^O(b^{\otimes D}) \rhd^O \delta_\sigma(a) = \delta_\sigma^O(b^{\otimes D} \star^O a), \quad \text{where } b^{\otimes D} \star^O a = (b^{\otimes D} \otimes a) \circ \Delta_{CEM}^O,$$

for $b^{\otimes D} \in \mathcal{CF}_{DO}^{\otimes D*}$ and $a \in \mathcal{CF}_D^*$. The image of δ_{DO}^O is a formal sum over ordered clumped forests.

2. We define the map φ that forgets the ordering of the vertices and let

$$\hat{\varphi}(\pi) := \frac{\varphi(\pi)}{|\pi|!}, \quad \varphi^{\otimes D}(\otimes_{d \in D} \pi_d \iota_d) = \otimes_{d \in D} \varphi(\pi_d) \iota_d,$$

where $\varphi, \hat{\varphi} : \mathcal{CF}_{DO} \to \mathcal{CF}_D$ and $\varphi^{\otimes D}, \hat{\varphi}^{\otimes D} : \mathcal{CF}_{DO}^{\otimes D} \to \mathcal{CF}_D^{\otimes D}$. We have the following properties

$$\hat{\varphi} \circ \rhd^O = \rhd \circ (\hat{\varphi}^{\otimes D} \otimes \operatorname{id}), \quad (\varphi^{\otimes D} \otimes \operatorname{id}) \circ \Delta_{CEM}^O = \Delta_{CEM} \circ \varphi.$$

We also note that for a functional $b^{\otimes D} \in \mathcal{CF}_D^{\otimes D*}$, $\delta_\sigma(b^{\otimes D}) = \hat{\varphi}^{\otimes D}(\delta_\sigma^O(b^{\otimes D} \circ \varphi^{\otimes D}))$, therefore,

$$
\begin{aligned}
\delta_\sigma(b^{\otimes D}) \rhd \delta_\sigma(a) &= \hat{\varphi}^{\otimes D}\big(\delta_\sigma^O(b^{\otimes D} \circ \varphi^{\otimes D})\big) \rhd \delta_\sigma(a) \\
&= \hat{\varphi}\big(\delta_\sigma^O(b^{\otimes D} \circ \varphi^{\otimes D}) \rhd^O \delta_\sigma(a)\big) \\
&= \hat{\varphi}\big(\delta_\sigma^O((b^{\otimes D} \circ \varphi^{\otimes D}) \star^O a)\big) \\
&= \hat{\varphi}\big(\delta_\sigma^O((b^{\otimes D} \star a) \circ \varphi)\big) = \delta_\sigma(b^{\otimes D} \star a).
\end{aligned}
$$

This proves that δ_σ is an algebra morphism. $\qquad\square$

We extend the insertion action $\blacktriangleright: \mathcal{AT}_D^{\oplus D} \otimes \mathcal{AT}_D \to \mathcal{AT}_D$ to the insertion product $\blacktriangleright: \mathcal{AT}_D^{\oplus D} \otimes \mathcal{AT}_D^{\oplus D} \to \mathcal{AT}_D^{\oplus D}$ by

$$\tau \iota_d \blacktriangleright \gamma \iota_l = (\tau \blacktriangleright_d \gamma) \iota_l.$$

We recall that $\mathcal{CF}_D^{\otimes D} = S(\mathcal{AT}_D^{\oplus D})$ which allows us to use the Guin-Oudom process and define $\blacktriangleright: \mathcal{CF}_D^{\otimes D} \otimes \mathcal{CF}_D^{\otimes D} \to \mathcal{CF}_D^{\otimes D}$ and the substitution product $\rhd: \mathcal{CF}_D^{\otimes D} \otimes \mathcal{CF}_D^{\otimes D} \to \mathcal{CF}_D^{\otimes D}$. The algebra $(\mathcal{CF}_D^{\otimes D}, \rhd)$ is endowed with the deshuffle coproduct Δ of $S(\mathcal{AT}_D^{\oplus D})$. We present a generalization of the result obtained in [20] for classical forests.

Theorem 2.21 *The algebra $B_\rhd := (\mathcal{CF}_D^{\otimes D}, \rhd)$ endowed with the deshuffle coproduct Δ is a bialgebra dual to the CEM bialgebra $B_{CEM} := (\mathcal{CF}_D^{\otimes D}, \cdot, \Delta_{CEM})$ with respect to the $\langle -, - \rangle_\sigma$ inner product. Moreover, $H_{CEM} := B_{CEM} / \langle (1 - \overset{d}{\bullet}) \iota_d : d \in D \rangle$ is a Hopf algebra dual to the Grossman-Larson Hopf algebra $H_\rhd$, which is an appropriate sub-bialgebra of $B_\rhd$ and a universal enveloping algebra of $(\mathcal{AT}_D^{\oplus D}, \blacktriangleright)$.*

Proof Let us start by proving that $B_\triangleright$ is a bialgebra. The unit is given by $u := \otimes_{d\in D}\exp_C^{\cdot}(\overset{\mathrm{d}}{\underset{\bullet}{}}\iota_d)$ and counit is given by $\epsilon := \mathbf{1}^* \in \mathcal{CF}_D^{\otimes D*}$. It is straightforward to check that the counit is compatible with the product and unit. We use the property $\Delta(\exp_C^{\cdot}(\tau)) = \exp_C^{\cdot}(\tau) \otimes \exp_C^{\cdot}(\tau)$ to check that the unit is compatible with the coproduct, that is, $\Delta(u) = u \otimes u$. We use

$$\pi \triangleright (- \cdot -) = \sum_{(\pi)} (\pi_{(1)} \triangleright -) \cdot (\pi_{(2)} \triangleright -)$$

and the associativity of $\triangleright$ to obtain the identity

$$\sum_{(\pi),(\mu)} (\pi_{(1)} \triangleright \mu_{(1)} \triangleright -)(\pi_{(2)} \triangleright \mu_{(2)} \triangleright -) = \sum_{(\pi\triangleright\mu)} ((\pi \triangleright \mu)_{(1)} \triangleright -)((\pi \triangleright \mu)_{(2)} \triangleright -).$$

Therefore, the compatibility of the product and coproduct is necessary. This proves that $B_\triangleright$ is a bialgebra. Lemma 2.15 and Proposition 2.20 show that B_{CEM} is the dual bialgebra of $B_\triangleright$.

Similarly to [20], we obtain a Hopf algebra once we take the quotient of B_{CEM} by the ideal $\langle(\mathbf{1} - \overset{\mathrm{d}}{\underset{\bullet}{}})\iota_d : d \in D\rangle$. The Hopf algebra H_{CEM} thus obtained is dual to the Hopf algebra $H_\triangleright$ which is isomorphic to the Grossman-Larson Hopf algebra obtained using Guin-Oudom process applied to the insertion pre-Lie product $\blacktriangleright$. The elements of $H_\triangleright$ have the form

$$\otimes_{d\in D}\exp_C^{\cdot}(\overset{\mathrm{d}}{\underset{\bullet}{}}\iota_d) \cdot \pi, \quad \text{for } \pi \in \mathcal{CF}_D^{\otimes D}.$$

Following Guin-Oudom, Grossman-Larson algebra is a universal enveloping algebra of the respective pre-Lie algebra. $\square$

Theorem 2.22 (Substitution law) *Let $a \in \mathcal{CF}_D^*$ and let $\{b_{0,d} \in \mathcal{AT}_D^* : d \in D\}$ be a set of infinitesimal characters. Let us consider a map $\varphi : D \to \mathcal{AT}_D$ with $\varphi(d) = \delta_\sigma(b_{0,d})$, let $F_D(d) = f_d$ for $d \in D$, and let $S_\varphi = F_D \circ A_\varphi \circ \delta_\sigma$, then,*

$$S_\varphi(a) = S(b_c^{\otimes D} \star a),$$

where $b_c^{\otimes D} = \otimes_{d\in D}b_{c,d}\iota_d$ with $b_{c,d}$ being a character of $\mathcal{CF}_D$ that extends $b_{0,d}$.

Proof We use Lemma 2.18 and Proposition 2.16 to write A_φ as

$$A_\varphi(\pi) = \delta_\sigma(b_c^{\otimes D}) \triangleright \pi, \quad \text{where } b_c^{\otimes D} = \otimes_{d\in D}b_{c,d}\iota_d,$$

with $b_{c,d}$ being a character of $\mathcal{CF}_D$ that extends $b_{0,d}$. We use the fact that $S_\varphi = F_D \circ A_\varphi \circ \delta_\sigma$ and Proposition 2.20 to obtain

$$S_\varphi(a) = F_D\big(\delta_\sigma(b_c^{\otimes D}) \triangleright \delta_\sigma(a)\big) = F_D\big(\delta_\sigma(b_c^{\otimes D} \star a)\big) = S(b_c^{\otimes D} \star a).$$

This finishes the proof. $\square$

Analogously to [64], due to the relation between the D-algebra homomorphism $A_\varphi : \mathcal{CF}_D \to \mathcal{CF}_D$ and the corresponding map $b_c^{\otimes D}\star : \mathcal{CF}_D^* \to \mathcal{CF}_D^*$, we obtain Corollary 2.23.

Corollary 2.23 (Cointeraction) *There is the following cointeraction between the substitution law and the composition law:*

$$a_c^{\otimes D} \star (b * c) = (a_c^{\otimes D} \star b) * (a_c^{\otimes D} \star c),$$

for $a_c^{\otimes D} = \otimes_{d\in D}a_{c,d}\iota_d$ where $a_{c,d}$ are characters of $\mathcal{CF}_D$ and $b, c \in \mathcal{CF}_D^*$.

17

Proof We prove the statement by performing the following computation with $\varphi(d) = \delta_\sigma(a_{0,d})$,

$$\delta_\sigma(a_c^{\otimes D} \star (b * c)) = A_\varphi(\delta_\sigma(b * c)) = A_\varphi(\delta_\sigma(b)) \diamond A_\varphi(\delta_\sigma(c)) = \delta_\sigma((a_c^{\otimes D} \star b) * (a_c^{\otimes D} \star c)).$$

Since δ_σ is an isomorphism, the statement is proved. $\qquad\square$

We extend the definition of $\blacktriangleright$ to decorated aromatic forests as $\blacktriangleright \colon \mathcal{CF}_D^{\otimes D} \otimes \mathcal{AF}_D \to \mathcal{AF}_D$ and note that Φ is a $\mathcal{CF}_D^{\otimes D}$-module morphism, that is,

$$\Phi(\pi \blacktriangleright \mu) = \pi \blacktriangleright \Phi(\mu).$$

We extend the substitution product $\vartriangleright \colon \mathcal{CF}_D^{\otimes D} \otimes \mathcal{AF}_D \to \mathcal{AF}_D$ and $\Delta_{CEM} \colon \mathcal{AF}_D \to \mathcal{CF}_D^{\otimes D} \otimes \mathcal{AF}_D$ and use Proposition 2.20 to see that Φ^* is a $\mathcal{CF}_D^{\otimes D}$-comodule morphism, that is,

$$\Delta_{CEM} \circ \Phi^* = (\mathrm{id} \otimes \Phi^*) \circ \Delta_{CEM}.$$

Using the freeness of the tracial commutative D-algebra $(\mathcal{AF}_D, \curvearrowright)$, we note that for every map $\varphi \colon D \to \mathcal{AT}_D$, there exists a corresponding $A_\varphi \colon \mathcal{AF}_D \to \mathcal{AF}_D$ with the following property

$$\Phi \circ A_\varphi = A_\varphi \circ \Phi,$$

where we use the same notation for the A_φ over decorated clumped and aromatic forests. We describe the substitution law for S-series over decorated aromatic forests by using the fact that any functional over decorated aromatic forests $a \in \mathcal{AF}_D^*$ can be written as an image of the map Φ^*, that is, there exists a functional $a_C \in \mathcal{CF}_D^*$ such that $a = a_C \circ \Phi^*$.

Theorem 2.24 *Using the notation from Theorem 2.22 and given $a \in \mathcal{AF}_D$, we have*

$$S_\varphi(a) = S(b_c^{\otimes D} \star a),$$

with convolution product $\star$ with respect to $\Delta_{CEM} \colon \mathcal{AF}_D \to \mathcal{CF}_D^{\otimes D} \otimes \mathcal{AF}_D$.

Proof We note that given a functional $a \in \mathcal{AF}_D^*$, we can define a functional $a_C \in \mathcal{CF}_D^*$ such that $a = a_C \circ \Phi^*$, moreover, $S(a) = S(a_C)$. The statement follows from the following identities,

$$S_\varphi(a) = S_\varphi(a_C) = S(b_c^{\otimes D} \star a_C) = S((b_c^{\otimes D} \star a_C) \circ \Phi^*) = S(b_c^{\otimes D} \star (a_C \circ \Phi^*)) = S(b_c^{\otimes D} \star a),$$

with $b_c^{\otimes D} := \otimes_{d \in D} b_{c,d} \iota_d$. $\qquad\square$

Corollary 2.23 is extended to functionals over decorated aromatic forests in a similar fashion.

3 The Hopf algebra structures of exotic aromatic S-series

Exotic aromatic forests are a specific class of decorated aromatic forests that naturally appears in the study of numerical methods for solving stochastic differential equations with additive noise [53, 54, 49]. They are defined in the following way.

Definition 3.1 *An exotic aromatic forest is a decorated aromatic forest $(\pi, \alpha) \in AF_D$ with the decorations $D = \{\bullet\} \cup \mathbb{N}$, $\mathbb{N} = \{1, 2, 3, \dots\}$, that follows the following rules. All vertices of π are black, except the leaves that can be either black or numbered. If a natural number is used as a decoration, then it must decorate two leaves, that is, $|\alpha^{-1}(n)| \in \{0, 2\}$ for any $n \in \mathbb{N}$. Two*

exotic aromatic forests (π_1, α_1) and (π_2, α_2) are considered to be identical if $\pi_1 = \pi_2$ and there exists a map $\varphi \in S_{\mathbb{N}}$ such that $\alpha_1|_{\mathbb{N}} = \varphi \circ \alpha_2|_{\mathbb{N}}$. The pair of numbered leaves that correspond to the same number is called a liana, gathered in the set $L \subset V \times V/(x,y) = (y,x)$. The order of an exotic aromatic forest π is the following, where $V_\bullet$ denotes the number of black vertices of π,

$$|\pi| = |V_\bullet| + |L| - |S|,$$

where $S \subset V \times V/(x,y) = (y,x)$ is the set of stolons.

The exotic aromatic trees of order up to two are the following. Note that the order does not coincide with the number of black vertices in general.
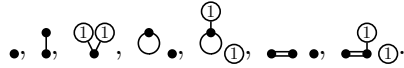
Exotic aromatic forests with one root are called exotic aromatic trees and form a set denoted by EAT. Exotic aromatic forests without a root are called exotic aromas and form a set denoted by EA. The corresponding vector spaces are denoted by $\mathcal{EAT}$ and $\mathcal{EA}$, respectively. We refer to Appendix A for further examples. We emphasize that the order of a forest is never negative. Note also that the following exotic aromatic forests are identical:

We extend the results of Section 2 to the space $\mathcal{EAF}$ of exotic aromatic forests. To do so, we build the commutative tracial D-algebra $(\mathcal{EAF}, \curvearrowright)$ of exotic aromatic forests using a commutative tracial D-algebra $(\mathcal{AF}_{\bullet\mathbb{N}}, \curvearrowright)$ of decorated aromatic forests.

We define the D-algebra $(\mathcal{AF}_{\bullet\mathbb{N}}, \curvearrowright)$ to be a vector space spanned by aromatic forests with black vertices and vertices decorated by natural numbers $\mathbb{N}$. The decoration is denoted by $\alpha : V(\pi) \to \{\bullet\} \cup \mathbb{N}$ and we assume that every number decorates an even number of vertices, that is, $\alpha^{-1}(n)$ is even for $n \in \mathbb{N}$, and a vertex decorated by a number cannot have an incoming edge. We note that our assumptions on the structure of $\pi \in \mathcal{AF}_{\bullet\mathbb{N}}$ are compatible with the commutative tracial D-algebra structure. For example,
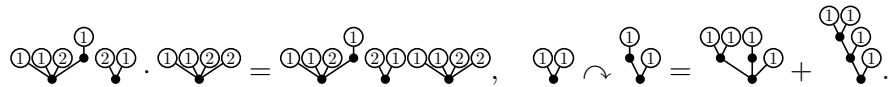
We can obtain the D-algebra $(\mathcal{AF}_{\bullet\mathbb{N}}, \curvearrowright)$ by choosing an appropriate sub-D-algebra and taking the quotient over an appropriate ideal of the free D-algebra. We define the commutative tracial D-algebra of exotic forests $(\mathcal{EAF}, \curvearrowright)$ as a sub-D-algebra of $(\overline{\mathcal{AF}}_{\bullet\mathbb{N}}, \curvearrowright)$ spanned by the elements

$$(\pi, \alpha_e) := \sum_{\alpha \in P(\alpha_e)} (\pi, \alpha) \ \in \ \overline{\mathcal{AF}}_{\bullet\mathbb{N}}, \tag{2}$$

with $|\alpha_e^{-1}(n)| \in \{0, 2\}$ for $n \in \mathbb{N}$. $P(\alpha_e)$ is the set of decorations α with $\alpha^{-1}(\bullet) = \alpha_e^{-1}(\bullet)$ and

$$\alpha(v_1) = \alpha(v_2) \quad \text{if} \quad \alpha_e(v_1) = \alpha_e(v_2), \quad \text{for } v_1, v_2 \in V(\pi).$$

We note that the exotic aromatic forests constructed in this way agree with Definition 3.1.

The D-algebra $(\mathcal{EAF}, \curvearrowright)$ of exotic aromatic forests is graded by the number of roots. An exotic aromatic forest is connected if it cannot be written as a concatenation of non-trivial exotic

aromatic forests. This notion of connectedness coincides with the one found in [52]. We note that due to the pairings of the number vertices (that are also called *lianas*), exotic aromatic forests can contain connected components which have more than one root, which is a major difference with the standard Butcher trees and forests. For example, the following exotic aromatic forests are connected:

$$\text{(forest diagrams)}.$$

We consider the coalgebra $(\mathcal{EAF}, \Delta_{\mathcal{EA}})$ with $\mathcal{EA}$-linear deshuffle coproduct $\Delta_{\mathcal{EA}}$ whose $\mathcal{EA}$-module of primitive elements $\mathrm{Prim}(\mathcal{EAF}, \Delta_{\mathcal{EA}})$ is spanned by the connected exotic aromatic forests, for example,

$$\Delta_{\mathcal{EA}}(\text{forest}) = \text{forest} \otimes \mathbf{1} + \mathbf{1} \otimes \text{forest}.$$

The algebraic structures related to D-algebras are studied in Section 2.1. We note that the space of exotic aromatic forests $\mathcal{EAF}$ can be defined as the symmetric algebra $S_{\mathcal{EA}}(\mathrm{Prim}(\mathcal{EAF}))$ over the ring of exotic aromas. We use the analysis of Section 2.3 to define two possible extentions of the concept of clumped forests to the exotic context,

$$\mathcal{CEF} := S(\mathrm{Prim}(\mathcal{EAF})), \quad \mathcal{CEF}^1 := S(\mathcal{EAT}).$$

We note that both symmetric algebras are over the base field $\mathbb{R}$, meaning that the exotic aromas are attached to the rooted components. We recall that $\mathcal{EAT}$ is the space of exotic aromatic forests with one root and $\mathcal{CEF}^1 \subset \mathcal{CEF}$. Some elements of $\mathcal{CEF}$ are

$$\text{(forest diagrams)}, \quad \text{(forest diagrams)}.$$

We note that $\text{(forest)} \neq \text{(forest)}$ in $\mathcal{CEF}$. Some elements of $\mathcal{CEF}^1$ are

$$\text{(forest diagrams)}, \quad \text{(forest diagrams)}.$$

Let $\Phi : \mathcal{CEF} \to \mathcal{EAF}$ denote the map that forgets the "clumping", that is, $\ker \Phi = \langle \omega\tau \cdot \gamma \cdot \pi - \tau \cdot \omega\gamma \cdot \pi \ : \ \omega \in \mathcal{EA}, \tau, \gamma \in \mathcal{EAT}, \pi \in \mathcal{CEF} \rangle$. The results of Section 2 yield the following structure on exotic aromatic and exotic clumped forests.

Theorem 3.2 *The space of exotic aromatic forests forms a Grossman-Larson Hopf algebroid*

$$(\mathcal{EAF}, \mathbf{1}, \diamond, \epsilon_{\mathcal{EA}}, \Delta_{\mathcal{EA}}, S_\diamond),$$

and the space of clumped exotic forests forms a Grossman-Larson Hopf algebra

$$(\mathcal{CEF}, \mathbf{1}, \diamond, \epsilon, \Delta, S_\diamond^C).$$

Moreover, $\Phi : \mathcal{CEF} \to \mathcal{EAF}$ is a surjective algebra morphism.

Proof We note that, according to the definition (2), $\mathcal{EAF}$ is obtained by taking a sub-D-algebra of $\overline{\mathcal{AF}}_{\bullet\mathbb{N}}$, therefore, we can use the analysys from Section 2.1 to build a $\mathcal{A}/_\mathbb{R}$-bialgebra $(\mathcal{EAF}, \mathbf{1}, \diamond, \epsilon_{\mathcal{EA}}, \Delta_{\mathcal{EA}})$. We obtain a bialgebra $(\mathcal{CEF}, \mathbf{1}, \diamond, \epsilon, \Delta)$ by constructing $\mathcal{CEF}$ as a sub-D-algebra of D-algebra $\overline{\mathcal{CF}}_{\bullet\mathbb{N}}$ which is a special case of $\mathcal{CF}_D$ (Section 2.3).

Following [61], the primitive elements $\mathrm{Prim}(\mathcal{EAF})$ are endowed with a pre-Lie product:

$$\pi_1 \tilde{\curvearrowright} \pi_2 := \pi_1 \diamond \pi_2 - \pi_1 \cdot \pi_2, \quad \text{for } \pi_1, \pi_2 \in \mathrm{Prim}(\mathcal{EAF}).$$

We use the pre-Lie product $\tilde{\curvearrowright}$ to define the antipode $S_\diamond$ for $\mathcal{EAF}$ by replacing all instances of $\curvearrowright$ in Proposition 2.8 by the product $\tilde{\curvearrowright}$, all instances of $\tau \in \mathcal{T}_D$ by $\tau \in \mathrm{Prim}(\mathcal{EF}, \Delta)$, and $\omega \in \mathcal{A}_D$ by $\omega \in \mathcal{EA}$ where $\mathcal{EF}$ denotes the space of exotic forests, that is, exotic aromatic forests without aromas. The antipode $S_\diamond^C$ for $\mathcal{CEF}$ is obtained the same way, but the trees $\tau \in \mathcal{T}_D$ are replaced by $\tau \in \mathrm{Prim}(\mathcal{EAF})$ and the identities (i) are ignored. This proves that we have a Hopf algebroid and Hopf algebra structures over $\mathcal{EAF}$ and $\mathcal{CEF}$, respectively. We note that $\Phi(\pi \diamond \mu) = \Phi(\pi) \diamond \Phi(\mu)$ and $\Phi(\mathbf{1}) = \mathbf{1}$, so Φ is a surjective algebra morphism. $\qquad\square$

To obtain exotic aromatic S-series, we define the elementary differential map by

$$F_{\mathcal{E}}(\pi)[\phi] = \sum_{\substack{i_v = 1, \ldots, d \\ v \in V}} \delta_{I_{S \cup L}} \left(\prod_{v \in V_\bullet} f_{I_{\Pi(v)}}^{i_v} \right) \phi_{I_R},$$

where R is the set of roots, $\Pi(v)$ is the set of predecessors of v, and $\delta_{I_{S \cup L}}$ identifies the indices of the lianas and stolons, that is, $i_x = i_y$ if $(x, y) \in S \cup L$. For instance, we have

$$\pi = \overset{\textcircled{1}}{\underset{\quad}{\circ}}\; \underset{\textcircled{2}}{\bullet}\!\!\overset{\textcircled{2}\textcircled{1}}{\bullet}\!\!, \quad F_{\mathcal{E}}(\pi)[\phi] = \sum_{i,j,s,l_1,l_2=1}^{d} f_{il_1}^i f^s f_{l_2}^s f_{l_1}^j \phi_{jl_2}.$$

We can obtain $F_{\mathcal{E}}$ by using the freeness of the D-algebra $\mathcal{AF}_{\bullet\mathbb{N}}$ and restricting to $\mathcal{EAF}$. Given a coefficient map $a \in \mathcal{EAF}^*$, an exotic aromatic S-series is the following formal series (see Definition 2.11),

$$S^h(a) := \sum_{\pi \in EAF} h^{|\pi|} \frac{a(\pi)}{\sigma(\pi)} F_{\mathcal{E}}(\pi), \quad S(a) = S^1(a),$$

where the symmetry coefficient $\sigma(\pi)$ counts the number of graph automorphisms leaving π unchanged. The S-series formalism allows us to rewrite the tedious combinatorics of numerical analysis in terms of simple graph operations that do not involve the dimension of the problem.

The composition and substitution laws of exotic aromatic S-series are detailed in the following two results, whose proofs are straightforwardly derived from the analysis of Section 2. We present concrete applications of these results in numerical analysis in Section 4 and several examples for order up to three in Appendix A.

Theorem 3.3 (Composition law) *[12] Let $S(a)$ and $S(b)$ be two exotic aromatic S-series and let ϕ be a test function. Then,*

$$S(a)[S(b)[\phi]] = S(a * b)[\phi], \quad \text{with } a * b = (a \otimes b) \circ \Delta_{CK},$$

with Δ_{CK} being the Connes-Kreimer coproduct over exotic aromatic forests.

Example 3.4 *The composition law is computed similarly to the deterministic context (see [24, Thm. 3.1]): the Connes-Kreimer coproduct of an exotic aromatic forest π (Definition 2.12) still is the linear combination of the $(\pi \setminus \pi_0) \otimes \pi_0$, where π_0 is a rooted subforest of π and $\pi \setminus \pi_0$ contains the remaining components. The major difference with the deterministic case is that*

one must make sure that the paired number vertices are not separated across different sides of the tensor product. We find, for instance,

$$\Delta_{CK}(\text{\includegraphics[height=1em]{a}}) = \mathbf{1} \otimes \text{\includegraphics[height=1em]{b}} + \text{\includegraphics[height=1em]{c}} \otimes \text{\includegraphics[height=1em]{d}} + \text{\includegraphics[height=1em]{e}} \otimes \text{\includegraphics[height=1em]{f}} + \text{\includegraphics[height=1em]{g}} \otimes \text{\includegraphics[height=1em]{h}} +$$

$$\text{\includegraphics[height=1em]{i}} \otimes \text{\includegraphics[height=1em]{j}} + \text{\includegraphics[height=1em]{k}} \otimes \text{\includegraphics[height=1em]{l}} + \text{\includegraphics[height=1em]{m}} \otimes \text{\includegraphics[height=1em]{n}} + \text{\includegraphics[height=1em]{o}} \otimes \mathbf{1},$$

$$\Delta_{CK}(\text{\includegraphics[height=1em]{p}}) = \mathbf{1} \otimes \text{\includegraphics[height=1em]{q}} + \text{\includegraphics[height=1em]{r}} \otimes \text{\includegraphics[height=1em]{s}} + \text{\includegraphics[height=1em]{t}} \otimes \text{\includegraphics[height=1em]{u}} + \text{\includegraphics[height=1em]{v}} \otimes \text{\includegraphics[height=1em]{w}} +$$

$$\text{\includegraphics[height=1em]{x}} \otimes \text{\includegraphics[height=1em]{y}} + \text{\includegraphics[height=1em]{z}} \otimes \text{\includegraphics[height=1em]{aa}} + \text{\includegraphics[height=1em]{ab}} \otimes \mathbf{1}.$$

We present in Appendix A the values of Δ_{CK} over all forests in $\mathrm{Prim}(\mathcal{EAF})$ up to order 3.

Given an infinitesimal functional $b_0 \in \mathcal{EAT}^*$, its exponential $\exp_C^{\odot}(b_0) \in \mathcal{CEF}^{1*}$ is a character of $\mathcal{CEF}^1$. We consider a particular case of the substitution law from Theorem 2.22 in which the black vertices are substituted by $\delta_\sigma(b_0) \in \overline{\mathcal{EAT}}$ while the numbered vertices are not changed, that is, a number vertex $\text{\scriptsize ⓝ}$ is substituted by $\text{\scriptsize ⓝ}$ for $n \in \mathbb{N}$. This allows us to simplify the definition of the CEM coaction.

Definition 3.5 *Let the CEM coaction $\Delta_{CEM} : \mathcal{EAF} \to \mathcal{CEF}^1 \otimes \mathcal{EAF}$ over exotic aromatic forests be defined as*

$$\Delta_{CEM}(\pi) := \sum_{p \subset \pi} p \otimes \pi/_p,$$

where the sum is over all clumped exotic subforests $p \in \mathcal{CEF}^1$ that cover all black vertices and $\pi/_p$ is the exotic aromatic forest obtained by contracting the exotic aromatic trees of p into black vertices. If the forest $\pi \in \mathcal{EAF}$ doesn't have valid subforests $p \in \mathcal{CEF}^1$, then $\Delta_{CEM}(\pi) = \mathbf{1} \otimes \pi$. For details see the proof of Theorem 3.7.

Example 3.6 *The substitution on exotic aromatic forests has additional difficulties compared to the deterministic context. While the coaction Δ_{CEM} works with a similar idea to the partitions of trees as in [24, Thm. 3.2], it ensures that paired vertices stay in the same component and it sums over all possible combinations of clumped exotic forests in $\mathcal{CEF}^1$ (and not exotic aromatic forests) in the left side of the tensor product. For instance, we have*

$$\Delta_{CEM}(\text{\includegraphics[height=1em]{ac}}) = \text{\includegraphics[height=1em]{ad}} \otimes \text{\includegraphics[height=1em]{ae}} + \text{\includegraphics[height=1em]{af}} \otimes \text{\includegraphics[height=1em]{ag}} + \text{\includegraphics[height=1em]{ah}} \otimes \text{\includegraphics[height=1em]{ai}} + \text{\includegraphics[height=1em]{aj}} \otimes \text{\includegraphics[height=1em]{ak}}$$

$$+ \text{\includegraphics[height=1em]{al}} \otimes \text{\includegraphics[height=1em]{am}} + \text{\includegraphics[height=1em]{an}} \otimes \text{\includegraphics[height=1em]{ao}} + \text{\includegraphics[height=1em]{ap}} \otimes \text{\includegraphics[height=1em]{aq}} + \text{\includegraphics[height=1em]{ar}} \otimes \text{\includegraphics[height=1em]{as}},$$

$$\Delta_{CEM}(\text{\includegraphics[height=1em]{at}}) = \text{\includegraphics[height=1em]{au}} \otimes \text{\includegraphics[height=1em]{av}} + \text{\includegraphics[height=1em]{aw}} \otimes \text{\includegraphics[height=1em]{ax}} + \text{\includegraphics[height=1em]{ay}} \otimes \text{\includegraphics[height=1em]{az}} + \text{\includegraphics[height=1em]{ba}} \otimes \text{\includegraphics[height=1em]{bb}}$$

$$+ \text{\includegraphics[height=1em]{bc}} \otimes \text{\includegraphics[height=1em]{bd}} + \text{\includegraphics[height=1em]{be}} \otimes \text{\includegraphics[height=1em]{bf}}.$$

We present in Appendix A the values of Δ_{CEM} over the elements of $\mathcal{EA}$ and $\mathrm{Prim}(\mathcal{EAF})$ up to order 3.

Theorem 3.7 *The space of clumped exotic forests $\mathcal{CEF}$ forms a Hopf algebra*

$$(\mathcal{CEF}, \mathbf{1}, \cdot, \mathbf{1}^*, \Delta_{CEM}, S),$$

22

where Δ_{CEM} is the coproduct extended from $\mathrm{Prim}(\mathcal{EAF})$ to $\mathcal{CEF}$ by respecting the concatenation product. Moreover, $\Phi^* : \mathcal{EAF} \to \mathcal{CEF}$ is a $\mathcal{CEF}^1$-comodule morphism where Φ^* is the adjoint of Φ, that is, $\Phi \circ \delta_\sigma = \delta_\sigma \circ \Phi^*$.

Proof We use the fact that the space $\mathcal{CEF}$ can be defined analogously to $\mathcal{EAF}$ using (2) such that $\mathcal{CEF} \subset \overline{\mathcal{CF}_D}$ with $D = \{\bullet\} \cup \mathbb{N}$. We recall that, following Theorem 2.21, $B_{CEM} := (\mathcal{CF}_D^{\otimes D}, \cdot, \Delta_{CEM})$ is a bialgebra which becomes a Hopf algebra $H_{CEM} := (\mathcal{CF}_D, \cdot, \Delta_{CEM})$ once we take the quotient of B_{CEM} by the ideal $\langle (\mathbf{1} - {}_\bullet)\iota_\bullet \rangle + \mathcal{J}$ defined as

$$\mathcal{J} := \langle (\mathbf{1} - {}_{\Bbbk})\iota_{\Bbbk}, \pi\iota_{\Bbbk} \ : \ \pi \notin \{\mathbf{1}, {}_{\Bbbk}\}, k \in \mathbb{N} \rangle.$$

The obtained coproduct of H_{CEM} can now be described as

$$\Delta_{CEM}(\pi) = \sum_{p \subset \pi} p \otimes \pi/_p,$$

where the sum is over all clumped subforests $p \in \mathcal{CF}_D$ that cover all black vertices and $\pi/_p$ is the clumped forest obtained by contracting the aromatic trees of p into black vertices. If the forest $\pi \in \mathcal{CF}_D$ doesn't have valid subforests $p \in \mathcal{CF}_D$, then $\Delta_{CEM}(\pi) = \mathbf{1} \otimes \pi$. We note that the construction of $\mathcal{CEF}$ with the CEM Hopf algebra structure as a quotient and subspace of $\overline{\mathcal{CF}_D}$ must be dual to the one seen in (2). We obtain the Hopf algebra

$$(\mathcal{CEF}, \mathbf{1}, \cdot, \mathbf{1}^*, \Delta_{CEM}, S),$$

with $\Delta_{CEM} : \mathcal{CEF} \to \mathcal{CEF} \otimes \mathcal{CEF}$. We recall the analysis of Section 2.4 and note that since $\exp_C^{\odot}(b_0)$ is a character of $\mathcal{CEF}^1$, we consider $\Delta_{CEM} : \mathcal{CEF} \to \mathcal{CEF}^1 \otimes \mathcal{CEF}$. We refer to the discussion from Section 2.4 for showing that Φ^* is a $\mathcal{CEF}^1$-comodule morphism. $\qquad\square$

We use the discussion from Section 2.4 to obtain Theorem 3.8.

Theorem 3.8 (Substitution law) *Let $a \in \mathcal{EAF}^*$, $b_0 \in \mathcal{EAT}^*$, then,*

$$S_\varphi(a) = S(b_c \star a), \quad \text{with } b_c \star a = (b_c \otimes a) \circ \Delta_{CEM},$$

where b_c is the character of $\mathcal{CEF}^1$ that extends b_0 and Δ_{CEM} is defined in Definition 3.5.

We can simplify the computation of the substitution law using the map Φ and Theorem 3.7. Given a functional $a \in \mathcal{EAF}^*$, we use the discussion from Section 2.3 to define a functional $a_C \in \mathcal{CEF}$ over clumped exotic forests as

$$a_C(\pi) := \frac{1}{n^m} a(\Phi(\pi)),$$

where n is the number of rooted components and m is the number of aromas. We have the property $a = a_C \circ \Phi^*$ which we use to compute the substitution law as follows

$$b_c \star a = b_c \star (a_C \circ \Phi^*) = (b_c \star a_C) \circ \Phi^*.$$

The substitution law $b_c \star a_C$ over $\mathcal{CEF}$ is easier to compute since Δ_{CEM} over $\mathcal{CEF}$ respects concatenation. Therefore, computing the values of Δ_{CEM} on $\mathrm{Prim}(\mathcal{EAF})$ is enough to obtain its values over all clumped exotic forests.

Example 3.9 *Let us compute the CEM coproduct over $\mathcal{EAF}$ using the comodule morphism Φ^*.*

$$(\Delta_{CEM} \circ \Phi^*)(\text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet) = \Delta_{CEM}(\text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\cdot\bullet) + \Delta_{CEM}(^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\cdot\text{⟲}_{\bullet})$$

$$= \Delta_{CEM}(\text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}})\hat{\cdot}\Delta_{CEM}(\bullet) + \Delta_{CEM}(^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}})\hat{\cdot}\Delta_{CEM}(\text{⟲}_{\bullet}),$$

where $(\pi_{(1)} \otimes \pi_{(2)})\hat{\cdot}(\mu_{(1)} \otimes \mu_{(2)}) = (\pi_{(1)} \cdot \mu_{(1)}) \otimes (\pi_{(2)} \cdot \mu_{(2)})$ *and*

$$\Delta_{CEM}(\text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}) = \bullet\,\bullet \otimes \text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}} + \text{⟲}_{\bullet} \otimes {}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}, \qquad\qquad \Delta_{CEM}(\bullet) = \bullet \otimes \bullet,$$

$$\Delta_{CEM}(\text{⟲}_{\bullet}) = \bullet\,\bullet \otimes \text{⟲}_{\bullet} + \text{⟲}_{\bullet} \otimes \bullet, \qquad\qquad \Delta_{CEM}(^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}) = \bullet \otimes {}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}.$$

Therefore, we have

$$(\Delta_{CEM} \circ \Phi^*)(\text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet) = \bullet\,\bullet\,\bullet \otimes \text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\cdot\bullet + \text{⟲}_{\bullet}\cdot\bullet \otimes {}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet + \bullet\,\bullet\,\bullet \otimes {}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\cdot\text{⟲}_{\bullet} + \text{⟲}_{\bullet}\cdot\bullet \otimes {}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet$$

$$= \bullet\,\bullet\,\bullet \otimes \Phi^*(\text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet) + 2\text{⟲}_{\bullet}\cdot\bullet \otimes \Phi^*(^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet),$$

which agrees with a direct computation that gives

$$\Delta_{CEM}(\text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet) = \bullet\,\bullet\,\bullet \otimes \text{⟳}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet + 2\text{⟲}_{\bullet}\cdot\bullet \otimes {}^{\text{Ⓘ}}_{\bullet}{}_{\text{Ⓘ}}\bullet.$$

Remark 3.10 *The Butcher forests and their extensions have been successfully applied to rough paths with the construction of branched rough paths [37] (see also [40, 35]), planarly branched rough paths [73, 26, 31], and aromatic rough paths [57]. In this context, the substitution law corresponds to the translation of rough paths, and the composition multi-index [71, 52] is closely linked to the fertility [44]. The creation of exotic aromatic or exotic clumped rough paths is natural and is matter for future work.*

4 Applications in stochastic numerical analysis

In the numerical analysis of ordinary differential equations, the composition and substitution of B-series have a variety of applications, including the derivation of order conditions, the composition of numerical methods, backward error analysis, or high-order integration based on modified equations (see [24] and references therein). The aromatic B-series appear naturally in the study of volume-preserving integrators and we refer to [25, 43, 9] for details. The stolons appear when studying projection methods on embedded manifolds, while the lianas are used in stochastic numerical analysis (see [49]). In this section, we focus on the applications of the composition and substitution laws for the numerical approximation of stochastic differential equations. As described in the original references [53, 54, 12], exotic aromatic B-series and S-series are a crucial calculation tool for the high-order approximation of SDEs in the weak context and for the invariant measure. There are several major differences with the deterministic context, especially on the concept of backward error analysis. Our analysis gives, in particular, new explicit expressions as exotic B-series of the modified vector field for the backward error analysis and for using modified equations for the invariant measure in $\mathbb{R}^d$ (see Theorem 4.10 and Theorem 4.14). We start this section with an algebraic reformulation of the different approaches for the approximation of SDEs with additive noise with exotic aromatic S-series. We then present applications of the composition and substitution laws for the numerical integration of SDEs.

4.1 Stochastic order theory with exotic aromatic S-series

We consider stochastic differential equations with additive noise on the d-dimensional torus $\mathcal{M} = \mathbb{T}^d$ or on smooth compact manifolds $\mathcal{M} = \{x \in \mathbb{R}^d, \zeta(x) = 0\}$ of codimension 1 with a smooth constraint $\zeta \colon \mathbb{R}^d \to \mathbb{R}$ of the form

$$dX(t) = \Pi_{\mathcal{M}}(X(t))f(X(t))dt + \Pi_{\mathcal{M}}(X(t)) \circ dW(t), \quad X(0) = x \in \mathcal{M}, \tag{3}$$

where x is assumed deterministic for simplicity, $\Pi_{\mathcal{M}} \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ is the orthogonal projection on the tangent bundle of the manifold $\mathcal{M}$, f is a smooth Lipschitz vector field, W is a standard d-dimensional Brownian motion on a probability space equipped with a filtration $(\mathcal{F}_t)_{t>0}$ and fulfilling the standard assumptions. Dynamics of the form (3) include overdamped Langevin dynamics [58] when $f = -\nabla V$ for V a smooth potential. If $\mathcal{M} = \mathbb{T}^d$, the orthogonal projection is $\Pi_{\mathcal{M}} = I_d$ and equation (3) reduces to

$$dX(t) = f(X(t))dt + dW(t).$$

There are three main types of approximations of (3). A strong approximation focuses on approaching $X(t)$ for a given realisation of W. A weak approximation approaches averages of functionals of the solution $\mathbb{E}[\phi(X(t))]$ where $\phi \in \mathcal{C}^\infty(\mathcal{N}, \mathbb{R})$ is a smooth test function defined on an open neighbourhood $\mathcal{N}$ of $\mathcal{M}$. An approximation for the invariant measure approximates averages of functionals at the equilibrium for ergodic systems. In this paper, we focus on the algebraic calculation underlying the high order approximation of (3) in the weak sense and for the invariant measure. We give the definitions of order in terms of S-series directly and we refer to the textbooks [69, 58, 3] for the detailed analysis on non-compact manifolds.

The generator of equation (3) is given in $\mathbb{T}^d$, for ϕ a test function, by the following linear combination of the primitive elements of order one:

$$\delta_\sigma(l) = \bullet + \frac{1}{2}①①, \quad \mathcal{L}\phi = S(l)[\phi] = \phi'f + \frac{1}{2}\Delta\phi,$$

where we recall that δ_σ is the isomorphism given by

$$\delta_\sigma(a) = \sum_{\pi \in \mathcal{E}\mathcal{A}\mathcal{F}} \frac{a(\pi)}{\sigma(\pi)}\pi.$$

In the manifold case, we add a new decoration $\circ$ to $\mathcal{E}\mathcal{A}\mathcal{F}$ that represents the map $n = \nabla\zeta \colon \mathbb{R}^d \to \mathbb{R}^d$ and we extend the elementary differential with $F_{\mathcal{E}}(\circ) = n$. The definition of the S-series is

$$S^h(a) = \sum_{\pi \in \mathcal{E}\mathcal{A}\mathcal{F}} h^{|\pi|} \frac{a(\pi)}{\sigma(\pi)\,|n|^{|V_\circ|}} F_{\mathcal{E}}(\pi), \quad |\pi| = |V_\bullet| + \frac{|V_\circ|}{2} + |L| - |S|,$$

and the generator becomes

$$\delta_\sigma(l) = \bullet - \circ\!\!\!\rightarrow\!\bullet \circ - \frac{1}{2}\bigcirc\!\circ + \frac{1}{2}\circ\!\!\!-\!\!\!\circ \circ + \frac{1}{2}①① - \frac{1}{2}\circ \circ,$$

$$\mathcal{L}\phi = S(l)[\phi] = \phi'f - |n|^{-2}\langle n, f\rangle\phi'n - \frac{1}{2}\,|n|^{-2}\operatorname{div}(n)\phi'n + \frac{1}{2}\,|n|^{-4}\langle n, n'n\rangle\phi'n$$

$$+ \frac{1}{2}\Delta\phi - \frac{1}{2}\,|n|^{-2}\,\phi''(n, n).$$

The quantity of interest in the weak context is $u(t, x) = \mathbb{E}[\phi(X(t))|X(0) = x]$. It satisfies the following formal expansion in a neighbourhood of $\mathcal{M}$, derived from the backward Kolmogorov equation (see, for instance, [41, 32, 3, 46, 47, 49]),

$$u(x, h) = \exp(h\mathcal{L})[\phi](x) = S^h(e)[\phi](x), \quad e = \exp^*(l) := \sum_{n=0}^{\infty} \frac{1}{n!} l^{*n},$$

where we recall that $*$ is the composition law (see Theorem 3.3). The first terms of e in $\mathbb{T}^d$ are

$$\delta_\sigma(e) = \mathbf{1} + \bullet + \frac{1}{2}①① + \frac{1}{2}\begin{smallmatrix}\bullet\\\bullet\end{smallmatrix} + \frac{1}{2}\bullet\,\bullet + \frac{1}{2}\bullet①① + \frac{1}{4}①\begin{smallmatrix}①\\\end{smallmatrix} + \frac{1}{2}\begin{smallmatrix}\bullet\\\bullet\end{smallmatrix}① + \frac{1}{8}①①②② + \cdots \tag{4}$$

In opposition to the deterministic formalism of Butcher trees, the S-series of the exact flow is not the exponential of a combination of trees, but the exponential of a combination of forests, as the primitive elements of exotic forests do not reduce to exotic trees:

$$\mathcal{ET} \subsetneq \mathrm{Prim}(\mathcal{EF}), \quad \mathcal{T} = \mathrm{Prim}(\mathcal{F}).$$

This major difference with deterministic geometric numerical integration makes, in particular, backward error analysis much more tedious in the stochastic case, as we rely on an additional non-trivial operation, similar to the integration by parts, that transforms primitive elements into exotic trees. In the manifold case, the first terms of $S(e)$ can be found in [54]. Consider now a one-step integrator for solving (3) of the form

$$X_{n+1} = \Phi_h(X_n), \quad X_0 = x \in \mathcal{M}, \tag{5}$$

where h is the stepsize of the method and the dependence in random variables is omitted for clarity. We assume that the method (5) is an exotic aromatic S-series method, that is, the numerical analogue of $u(x, h)$ satisfies the following formal Talay-Tubaro expansion [84] in a neighbourhood of $\mathcal{M}$:

$$\mathbb{E}[\phi(X_1)|X_0 = x] = S^h(a)[\phi](x).$$

For instance, the following class of stochastic Runge-Kutta integrators presented in [49] satisfy this assumption naturally in $\mathbb{T}^d$:

$$Y_n^i = X_n + h\sum_{j=1}^s a_{i,j} f(Y_n^j) + \sqrt{h} d_i \xi_n, \quad i = 1, \ldots, s, \tag{6}$$

$$X_{n+1} = X_n + h\sum_{i=1}^s b_i f(Y_n^i) + \sqrt{h}\xi_n,$$

where we use one Gaussian vector $\xi_n \sim \mathcal{N}(0, I_d)$ at each step for simplicity. Thanks to [12, Prop. 4.3] (see also [38, 77]), the exotic S-series $S(a)$ of methods of the form (6) is given by the following coefficient map,

$$\tau = \begin{smallmatrix} ①② \\ \text{j} \end{smallmatrix}\begin{smallmatrix} \text{l} \ \text{m} \\ \bullet \end{smallmatrix} \qquad a(\tau) = \sum_{i,j,k} b_i a_{i,j} d_j^2 d_i b_k a_{k,l} a_{k,m}.$$

The coefficient maps a of the exact flow and of stochastic Runge-Kutta methods (6) are naturally characters over $\mathcal{EF}$ equipped with the concatenation product, that is, a satisfies

$$a(\pi_1 \cdot \pi_2) = a(\pi_1) a(\pi_2), \quad \pi_1, \pi_2 \in \mathcal{EF}.$$

26

We denote $\mathrm{Char}(\mathcal{S})$ the set of characters over a subset $\mathcal{S}$ of $\mathcal{EF}$ equipped with the concatenation product.

The standard definition of weak order [84, 69] rewrites in the following way with exotic aromatic S-series.

Definition 4.1 *An integrator* (5) *with S-series $S^h(a)$ has (at least) weak order p for solving* (3) *if $a - e \in \mathcal{EAF}^*_{\geq p+1}$, i.e., if a and e coincide on all exotic aromatic forests of order up to p. An integrator of at least weak order one is called consistent.*

Example 4.2 *Consider for instance the Euler-Maruyama method in $\mathbb{T}^d$:*

$$X_{n+1} = X_n + hf(X_n) + \sqrt{h}\xi_n, \quad \xi_n \sim \mathcal{N}(0, I_d). \tag{7}$$

The exotic aromatic S-series of the Euler-Maruyama method is given by

$$\exp^{\cdot}(\delta_\sigma(l)) = \delta_\sigma(\exp^{\odot}(l)) = \mathbf{1} + {\scriptstyle\bullet} + \frac{1}{2}\text{⟨⟩} + \frac{1}{2}{\scriptstyle\bullet}\,{\scriptstyle\bullet} + \frac{1}{2}{\scriptstyle\bullet}\,\text{⟨⟩} + \frac{1}{8}\text{⟨⟩⟨⟩} + \cdots, \tag{8}$$

where the convolution product is $a \odot b := (a \otimes b) \circ \Delta$ with Δ the deshuffle coproduct over exotic aromatic forests (see Section 2.2). Comparing the S-series (4) and (8) shows that, under the standard regularity assumptions [84, 69, 3], the method is of weak order one.

The overdamped Langevin dynamics (3) with $f = -\nabla V$ are ergodic (see, for instance, [58, 30, 3]), that is, there exists a unique invariant measure $d\mu_\infty$ on $\mathcal{M}$ that has a probability density ρ_∞ with respect to $d\sigma_\mathcal{M}$, the canonical measure on $\mathcal{M}$ induced by the Euclidean metric of $\mathbb{R}^d$, such that for all test functions ϕ,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \phi(X(t))dt = \int_\mathcal{M} \phi(x)d\mu_\infty(x) \quad \text{almost surely}, \quad \rho_\infty(x) \propto \exp\left(-2V(x)\right).$$

The invariant measure is the unique solution to $\mathcal{L}^* d\mu_\infty = 0$. We take over the equivalence relation $\sim$ on exotic aromatic S-series and their dual defined in [53, 54, 12]. This relation is called the integration by parts as

$$a \sim b \quad \Rightarrow \quad \int_\mathcal{M} S^h(a)[\phi](x)d\mu_\infty(x) = \int_\mathcal{M} S^h(b)[\phi](x)d\mu_\infty(x),$$

and involves detaching and grafting back edges. The integration by parts allows one to transform primitive elements into trees. We give examples and refer to [53, 54, 12] for the detailed definition:

$$\text{⟨⟩} \sim -2{\scriptstyle\bullet}, \quad \text{(fig)} \sim -\text{(fig)} - 2\text{(fig)}, \quad \text{(fig)} \sim -\text{(fig)} - \text{(fig)} - 2\text{(fig)}, \quad \text{(fig)} \sim -\text{(fig)} - 2\text{(fig)}.$$

The approximation for the invariant measure is defined algebraically in the following way. We shall see that it can also be understood as the stochastic extension of pseudo-volume-preserving approximations.

Definition 4.3 *A consistent integrator* (5) *with S-series $S^h(a)$ has order p for sampling the invariant measure of* (3) *if $a - \delta_{\mathbf{1}} \sim \mathcal{EAF}^*_{\geq p+1}$.*

An integrator of weak order p immediately has at least order p for the invariant measure as $e \sim \delta_{\mathbf{1}}$, but there exist methods with high order for the invariant measure and weak order one (see, for instance, for the underdamped and overdamped Langevin equation [11, 55, 56, 3, 4, 53, 54]). An important result for our analysis in the $\mathbb{T}^d$ case with $f = -\nabla V$ is in [12, Thm. 5.8]. We mention that the assumption of this result is naturally satisfied for the stochastic Runge-Kutta integrators (6).

Proposition 4.4 ([53, 12]) *There exists an algorithm that transforms a character $a_c \in \mathrm{Char}(\mathcal{EF})$ over exotic forests into an equivalent character $a_c \circ A^* \in \mathrm{Char}(S(\mathcal{ET}))/\sim$ over the symmetric algebra spanned by exotic trees. This defines a map $A(.) = . \circ A^* \colon \mathrm{Char}(\mathcal{EF}) \to \mathrm{Char}(S(\mathcal{ET}))/\sim$ such that $a_c \sim A(a_c)$. Moreover, A naturally induces a map $A \colon \mathrm{Prim}(\mathcal{EF})^* \to \mathcal{ET}^*/\sim$ satisfying $A|_{\mathcal{ET}^*} = id$.*

The precise algorithm for computing A is omitted for simplicity as it can be found in [53, 12]. The operator A is called IBP+ELI in [12] and it shares similarities with the horizontal homotopy operator presented in [51] (see also [5, 6]). The map A extends to the manifold case (see [54]), and adapting Proposition 4.4 in this context is matter for future work.

Remark 4.5 *The output of A is defined up to linear combinations of exotic trees that vanish by integration by parts. The precise description of the kernel of the integration by parts relation $\sim$ is an important open question for the high order integration for the invariant measure. The first occurrence of a non-trivial element in the kernel appears for order four. Here is one such element:*

4.2   The composition law in stochastic numerical analysis

The use of the composition rule of exotic aromatic S-series in stochastic numerical analysis first appears in [53, 54, 49] (see also [29]), without the Hopf algebra formalism.

Proposition 4.6 (Composition of integrators) *Consider two independent integrators Φ_h^1 and Φ_h^2 with exotic S-series $S^h(a_1)$ and $S^h(a_2)$, then the composition of Φ_h^1 and Φ_h^2 has the following S-series*

$$\mathbb{E}[\phi((\Phi_h^2 \circ \Phi_h^1)(x))] = S^h(a_1 * a_2)[\phi](x).$$

Example 4.7 *Consider the explicit and implicit Euler methods in $\mathbb{T}^d$*

$$\Phi_h^1(x) = x + hf(x) + \sqrt{h}\xi^1, \quad \Phi_h^2(x) = x + hf(\Phi_2(x)) + \sqrt{h}\xi^2,$$

with independent random variables ξ_1, ξ_2. Then, a calculation yields that the composed method Φ has weak order 2 for solving equation (3):

$$\mathbb{E}[\phi(\Phi_h(x))] = S^h\left(\mathbf{1} + l + \frac{l^{*2}}{2} + \dots\right)[\phi](x), \quad \Phi_h = \Phi_{h/2}^2 \circ \Phi_{h/2}^1.$$

The composed method Φ coincides in law with the stochastic trapezoidal method

$$\Phi_h(x) = x + h\frac{f(x) + f(\Phi_h(x))}{2} + \sqrt{h}\xi.$$

In the spirit of effective order for ODEs [16], the postprocessing idea is used in [85] for the high-order sampling of the invariant measure of (3) in $\mathbb{T}^d$ (or $\mathbb{R}^d$). The approach with trees was then introduced in [53, Sec. 5.3], where order conditions are presented for stochastic Runge-Kutta methods and postprocessors. Consider an integrator Φ for the invariant measure. After applying Φ through the entire time interval, we apply a correction $\overline{\Phi}$ at the very last step. If $\overline{\Phi}$ is chosen carefully, the postprocessed integrator has a higher order than the original method Φ. This methodology yields a costless way to improve the order of a method. Extending the use of postprocessors to the manifold case is a matter for future work.

Proposition 4.8 (Postprocessed integrators) *Consider an integrator with S-series $S^h(a)$ of order p for the invariant measure of* (3), *and a correction with S-series of the form*

$$S^h(\overline{a}) = \mathbf{1} + h\alpha_1 l + \cdots + h^{p-1}\alpha_{p-1}l^{*p-1} + \ldots,$$

for some constants $\alpha_k \in \mathbb{R}$. Then, the postprocessed method has order $p + 1$ for the invariant measure of (3) *in $\mathbb{T}^d$ if the following condition is satisfied:*

$$a + [l, \overline{a}] \sim \mathcal{EF}^*_{\geq p+1}, \quad [a, b] := a * b - b * a.$$

Example 4.9 *The following scheme, first introduced in [55], has weak order one and order two for the invariant measure,*

$$\Phi_h(x) = x + hf(x + \frac{\sqrt{h}}{2}\xi) + \sqrt{h}\xi, \quad \overline{\Phi}_h(x) = x + \frac{\sqrt{h}}{2}\xi.$$

Further examples can be found in [85, 1, 53].

4.3 The substitution law in stochastic numerical analysis

The main applications of the substitution law of standard B-series are backward error analysis and modified equations for ODEs. Adapting backward error analysis and modified equations in the stochastic context is challenging [83] and it is an active field of research [86, 2, 30, 3, 46, 47, 59, 36, 52]. It was in particular proven in [30, 3] that backward error analysis rewrites nicely in the context of the invariant measure. We use the substitution law and the integration by parts operation to show that the modified vector field writes as an exotic S-series, for which we provide an explicit expression. For simplicity, we work in $\mathbb{T}^d$ and discuss the manifold case at the end of the subsection.

The goal of backward error analysis is to find a modified vector field written formally as an exotic aromatic B-series,

$$h\widetilde{f} = B^h(b) = hf + h^2 f_1 + h^3 f_2 + \ldots, \quad b: \mathcal{ET} \to \mathbb{R}, \quad b(\bullet) = 1,$$

where the invariant measure of the ergodic integrator (5) with S-series $S(a)$ coincides with the invariant measure of the modified dynamics:

$$d\widetilde{X}(t) = \Pi_{\mathcal{M}}(\widetilde{X}(t))\widetilde{f}(\widetilde{X}(t))dt + \Pi_{\mathcal{M}}(\widetilde{X}(t)) \circ dW(t), \quad \widetilde{X}(0) = \widetilde{X}_0 \in \mathcal{M}. \tag{9}$$

The coefficient map b is the solution to the substitution $b \star e \sim a$. It is known [30, 3] that there exists a modified vector field for large classes of methods, such as stochastic Runge-Kutta methods, but it is not unique and there is no reason in general why such a map could be written as an exotic B-series. A geometric justification of the importance of writing the modified vector

field as an exotic aromatic B-series is given in [52]: it ensures that $\widetilde{f}$ is invariant with respect to orthogonal changes of coordinates, which is a natural property in the stochastic context. We provide here a simple algebraic criterion, satisfied by large classes of methods, for the description of integrators that have a modified vector field in the form of an exotic B-series. We also give an explicit expression of the modified vector field $\widetilde{f}$.

Theorem 4.10 (Backward error analysis) *Consider a consistent method with the exotic S-series $S^h(a)$ for solving equation (3) with $f = -\nabla V$. Assume that a is a character of $(\mathcal{EF}, \cdot)$. Then, there exists a modified vector field $h\widetilde{f} = B^h(b)$ that can be written as an exotic B-series with a coefficient map $b \colon \mathcal{ET} \to \mathbb{R}$ satisfying $b(\bullet) = 1$, $b \star e \sim a$, and given by*

$$b = \delta_\bullet + A\Big(\sum_{k=0}^\infty (-1)^k A_{\tilde{\star}e}^k (a - e) \Big),$$

where $A_{\tilde{\star}e}(x) = A(x)\tilde{\star}e$ and for all $\tau \in ET$ such that $|\tau| > 1$, we define

$$\tilde{\Delta}_{CEM}(\tau) = \Delta_{CEM}(\tau) - \bullet \otimes \tau - \tau \otimes \bullet, \quad and \quad b_{n-1}\tilde{\star}e = (b_{n-1,c} \otimes e) \circ \tilde{\Delta}_{CEM}.$$

Proof After initializing $b_0 = \delta_\bullet$, we construct recursively the coefficient map sequence (b_n) by

$$b_n = b_{n-1} + A(a - b_{n-1} \star e), \quad b_{n-1} \star e = (b_{n-1,c} \otimes e) \circ \Delta_{CEM},$$

with $b_{n-1,c} := \exp^\odot(b_{n-1})$ and the coproduct $\Delta_{CEM} : \mathcal{EF} \to S(\mathcal{ET}) \otimes \mathcal{EF}$. Assume $b_{n-1}(\bullet) = 1$, since $A^*(\bullet) = \bullet - \text{⊙⊙}$ and the method associated to a is consistent, we find

$$b_n(\bullet) = b_{n-1}(\bullet) + (a - b_{n-1} \star e)(\bullet - \text{⊙⊙}) = 1 + a(\bullet) - a(\text{⊙⊙}) - e(\bullet) + e(\text{⊙⊙}) = 1,$$

thus we obtain that $b_n(\bullet) = 1$ for any n by induction. For all $\tau \in \mathcal{ET}$ such that $|\tau| > 1$, using the reduced coproduct yields

$$\begin{aligned} b_n(\tau) &= b_{n-1}(\tau) + A(a) - (b_{n-1} \otimes e)(A^*(\tau) \otimes \bullet) - (b_{n-1} \otimes e)(\bullet \otimes A^*(\tau)) - A(b_{n-1}\tilde{\star}e) \\ &= b_{n-1}(\tau) - A(b_{n-1}) + A(a - e - b_{n-1}\tilde{\star}e) \\ &= A(a - e + b_{n-1}\tilde{\star}e), \end{aligned}$$

where we recall that the only exotic tree τ for which $|\tau| \le 1$ is $\tau = \bullet$ and we used $A(b_{n-1}) = b_{n-1}$. The first values are

$$\begin{aligned} b_1 &= \delta_\bullet + A(a - e), \\ b_2 &= \delta_\bullet + A(a - e) - A(A(a - e)\tilde{\star}e), \\ &\vdots \\ b_n &= \delta_\bullet + A\Big(\sum_{k=0}^{n-1} (-1)^k A_{\tilde{\star}e}^k (a - e) \Big), \quad \text{where } A_{\tilde{\star}e}(x) = A(x)\tilde{\star}e. \end{aligned}$$

Since $|A^*(\tau)|_e \le |\tau|_e$ where $|\tau|_e$ is the number of edges of τ, and $\tilde{\Delta}_{CEM}^n(\tau) = 0$ if $n \ge |\tau|_e$, we have for all $n \ge |\tau|_e$,

$$b_n(\tau) = \delta_\bullet + A\Big(\sum_{k=0}^{|\tau|_e - 1} (-1)^k A_{\tilde{\star}e}^k (a - e) \Big)(\tau),$$

so that the sequence (b_n) converges to the desired coefficient map b by stationarity. $\qquad\square$

Example 4.11 *The first terms of the modified vector field $h\widetilde{f} = B^h(b)$ given by Theorem 4.10 for the Euler-Maruyama method (7) are*

$$B^h(b) = h_\bullet + \frac{h^2}{2}\text{\normalsize↑} + \frac{h^2}{4}\text{\normalsize⚉} - \frac{h^2}{2}\text{\normalsize↑} + \frac{h^2}{12}\text{\normalsize∀} - \frac{h^2}{4}\text{\normalsize∀} - \frac{h^2}{12}\text{\normalsize∀} + \frac{h^2}{12}\text{\normalsize⚉} + \frac{h^2}{12}\text{\normalsize⚉} + \dots$$

Note that removing the stochastic terms (that are, the trees with lianas) does not yield the modified vector field for the Euler method with the standard deterministic backward error analysis (see [38, Chap. IX]). Indeed, the approximation for the invariant measure is a purely stochastic concept as it relies on ergodic dynamics, so that removing the terms with noise does not make sense.

Remark 4.12 *Similarly to the deterministic context [38, Chap. IX], the properties of the integrator are observed directly on its modified vector field. In particular, if $h\widetilde{f} = B^h(b)$ and $b - \delta_\bullet \sim \mathcal{ET}^*_{>p}$, then the integrator has at least order p for the invariant measure. This gives an equivalent definition of the order for the invariant measure based on the modified vector field, as observed in [82, 4]. In particular, a method is exact for the invariant measure if its modified vector field satisfies*

$$\text{div}(\widetilde{f} - f) + (f, \widetilde{f} - f) = 0. \tag{10}$$

Equation (10) is analogous to the one satisfied by the modified vector field of volume-preserving aromatic B-series methods for ODEs, whose understanding is already an important unsolved problem of deterministic geometric numerical integration (see, for instance, the works [25, 43, 71, 9, 10, 51]). Moreover, following Proposition 2.1, the operations div and $(f, .)$ applied to exotic B-series generate the cyclic and the stolonic aromas, so that the exotic aromatic formalism is the natural tool for the study of exotic aromatic S-series methods that preserve the invariant measure exactly. This calls for future works in the spirit of [51, 50].

Remark 4.13 *Following [30], the exotic B-series $h\widetilde{f} = B^h(b)$ translates into a formal expansion of the invariant measure $d\mu^h = \rho^h d\sigma_{\mathcal{M}}$ of the integrator by solving:*

$$\rho^h = \rho_\infty + h\rho_1 + h^2\rho_2 + \dots, \quad \widetilde{\mathcal{L}}^*\rho^h = 0, \quad \widetilde{\mathcal{L}} = \mathcal{L} + (\widetilde{f} - f)\nabla.$$

For the first few terms, we find (see [30] for the detailed expansion),

$$\mathcal{L}^*\rho_1 = -\text{div}(f_1\rho_\infty), \quad \mathcal{L}^*\rho_2 = -\text{div}(f_1\rho_1) - \text{div}(f_2\rho_\infty).$$

Consider now a consistent exotic aromatic B-series method with S-series $S(a)$. Similar to backward error analysis, we are interested in finding a modified vector field $h\widetilde{f} = B^h(b)$ with $b \in \mathcal{ET}^*$ and $b(\bullet) = 1$ such that $b \star a \sim \delta_{\mathbf{1}}$, that is, the integrator applied to the modified equation (9) is exact. This technique allows in particular to increase the order of a numerical method when the partial derivatives of f are not costly to evaluate (see, for instance, in the deterministic setting [21, 38, 23]). A general expansion of the modified vector field is presented in the $\mathbb{T}^d$ case in [3, 53], but it is not an exotic B-series in general and it is not unique. Following [12], we propose a simple criterion to obtain the existence of a modified vector field in the form of an exotic B-series in the context of $\mathbb{T}^d$, for which we also provide an explicit expression.

Theorem 4.14 (Modified equations) *Consider a consistent method with the exotic S-series $S^h(a)$ for solving equation (3) with $f = -\nabla V$. Assume that a is a character of $(\mathcal{EF}, \cdot)$. Then,*

there exists a modified vector field $h\widetilde{f} = B^h(b)$ that can be written as an exotic B-series with the coefficient map $b: \mathcal{ET} \to \mathbb{R}$ satisfying $b(\bullet) = 1$, $b \star a \sim \delta_{\mathbf{1}}$, and given by

$$b = \delta_\bullet - A\big(\sum_{k=0}^{\infty}(-1)^k A_{\tilde{\star}a}^k(a)\big).$$

To prove Theorem 4.14, we use a similar reasoning as in Theorem 4.10.

Proof We introduce the sequence (in the spirit of the works [3, 53])

$$b_n = b_{n-1} - A(b_{n-1} \star a), \quad b_0 = \delta_\bullet.$$

Similarly to the proof of Theorem 4.10, we show by induction that $b_n(\bullet) = 1$ and we obtain $b_n = -A(a + b_{n-1}\tilde{\star}a)$ using the reduced coproduct. The sequence (b_n) takes the following first values

$$\begin{aligned}
b_1 &= \delta_\bullet - A(a), \\
b_2 &= \delta_\bullet - A(a) + A(A(a)\tilde{\star}a), \\
&\vdots \\
b_n &= \delta_\bullet - A\big(\sum_{k=0}^{n-1}(-1)^k A_{\tilde{\star}a}^k(a)\big), \quad \text{where } A_{\tilde{\star}a}(x) = A(x)\tilde{\star}a,
\end{aligned}$$

and (b_n) converges to $b = \delta_\bullet - A\big(\sum_{k=0}^{\infty}(-1)^k A_{\tilde{\star}a}^k(a)\big)$ by stationarity. $\qquad\square$

Any stochastic Runge-Kutta method (6) has a coefficient map that is a character, so that Theorem 4.14 applies and there exists a modified vector field that can be written as an exotic B-series. We refer to [53, Sec. 5.1] for examples in $\mathbb{T}^d$. In the manifold case, the most popular integrators for solving constrained SDEs in the weak sense or for sampling the invariant measure are projection methods (see, for instance, the textbook [58, Chap. 3]). One could also use stochastic Lie-group methods, with a generalisation of the exotic S-series formalism, but this is out of the scope of the present paper and we leave it for future work. In [54], a new class of Runge-Kutta projection methods was introduced for the high-order approximation of constrained overdamped Langevin processes in codimension one. Theorems 4.10 and 4.14 do not extend straightforwardly to the manifold case as there does not exist an equivalent of Proposition 4.4 with general exotic aromatic S-series. However, one can adapt the algorithm described in the proof of Theorem 4.14, taking over the integration by parts relation $\sim$ described in [54]. Define the projection operator $\Pi: \mathcal{EAT} \to \mathcal{EAT}$ by $\Pi\tau = \tau - \tau \blacktriangleright_\bullet (\!\circ\!\!-\!\!\bullet\ \circ)$, with $\blacktriangleright_\bullet$ the decorated insertion product (see Section 2.4), and its dual Π^*. Define $b_0 = \delta_\bullet$. For the induction, at the step n, if $A(b_{n-1} \star a)$ is well-defined and if $\Pi^* A(b_{n-1} \star a) = A(b_{n-1} \star a)$, compute

$$b_n = b_{n-1} - A(b_{n-1} \star a).$$

Then, applying the integrator to the equation (9) with the modified vector field $B_{h,f}(b_n)$ yields a method of order $n + 1$ for the invariant measure of the original problem (3). The condition with Π^* ensures that $B^h(b_n)$ is a vector field on $\mathcal{M}$. We present an example on the sphere and we leave the extension of Theorems 4.10 and 4.14 to the manifold case for future works.

Example 4.15 *For the sake of simplicity, we consider the unit sphere $\mathcal{M} = \{x \in \mathbb{R}^d, |x|^2 = 1\}$, with $n(x) = x$. The constrained Euler scheme with implicit projection direction:*

$$X_{n+1} = X_n + hf(X_{n+1}) + \sqrt{h}\xi_n + \lambda_n X_{n+1}, \quad |X_{n+1}| = 1.$$

has order 2 for sampling the invariant measure of (3) *when applied to the modified equation* (9) *with the modified vector field*

$$\widetilde{f} = f + h\bigg[-\frac{1}{2}f'f - \frac{1}{4}\Delta f + \frac{3}{4}f - \frac{1}{4}\operatorname{div}(n)f - \frac{1}{2}(n,f)f - \frac{1}{4}f'n - \frac{1}{4}\operatorname{div}(n)f'n$$

$$- \frac{1}{2}(n,f)f'n - \frac{1}{4}f''(n,n) + \frac{1}{2}(n,f)(n,f'n)n + \frac{1}{4}(n,f''(n,n))n - \frac{1}{4}\operatorname{div}(f)'nn$$

$$- \frac{1}{2}(n,f'f)n + \frac{1}{4}(n,f'n)n + \frac{1}{4}\operatorname{div}(n)(n,f'n)n + \frac{1}{2}(n,f)^2n - \frac{3}{4}(n,f)n + \frac{1}{4}\operatorname{div}(n)(n,f)n \bigg].$$

References

[1] A. Abdulle, I. Almuslimani, and G. Vilmart. Optimal explicit stabilized integrator of weak order 1 for stiff and ergodic stochastic differential equations. *SIAM/ASA J. Uncertain. Quantif.*, 6(2):937–964, 2018.

[2] A. Abdulle, D. Cohen, G. Vilmart, and K. C. Zygalakis. High weak order methods for stochastic differential equations based on modified equations. *SIAM J. Sci. Comput.*, 34(3):A1800–A1823, 2012.

[3] A. Abdulle, G. Vilmart, and K. C. Zygalakis. High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM J. Numer. Anal.*, 52(4):1600–1622, 2014.

[4] A. Abdulle, G. Vilmart, and K. C. Zygalakis. Long time accuracy of Lie-Trotter splitting methods for Langevin dynamics. *SIAM J. Numer. Anal.*, 53(1):1–16, 2015.

[5] I. M. Anderson. The variational bicomplex. Unpublished, 1989.

[6] I. M. Anderson. Introduction to the variational bicomplex. In *Mathematical aspects of classical field theory (Seattle, WA, 1991)*, volume 132 of *Contemp. Math.*, pages 51–73. Amer. Math. Soc., Providence, RI, 1992.

[7] I. Angiono, C. Galindo, and L. Vendramin. Hopf braces and Yang-Baxter operators. *Proc. Amer. Math. Soc.*, 145(5):1981–1995, 2017.

[8] S. Anmarkrud and A. Kværnø. Order conditions for stochastic Runge-Kutta methods preserving quadratic invariants of Stratonovich SDEs. *J. Comput. Appl. Math.*, 316:40–46, 2017.

[9] G. Bogfjellmo. Algebraic structure of aromatic B-series. *J. Comput. Dyn.*, 6(2):199–222, 2019.

[10] G. Bogfjellmo, E. Celledoni, R. McLachlan, B. Owren, and R. Quispel. Using aromas to search for preserved measures and integrals in Kahan's method. *Submitted*, 2022.

[11] N. Bou-Rabee and H. Owhadi. Long-run accuracy of variational integrators in the stochastic context. *SIAM J. Numer. Anal.*, 48(1):278–297, 2010.

[12] E. Bronasco. Clumped forests for the description of the Hopf algebra structures of exotic aromatic series and their applications in stochastic numerical analysis. *Submitted*, 2022.

[13] Y. Bruned and D. Manchon. Algebraic deformation for (S)PDEs. *J. Math. Soc. Japan*, 75(2):485–526, 2023.

[14] K. Burrage and P. M. Burrage. High strong order explicit Runge-Kutta methods for stochastic ordinary differential equations. *Appl. Numer. Math.*, 22(1-3):81–101, 1996. Special issue celebrating the centenary of Runge-Kutta methods.

[15] K. Burrage and P. M. Burrage. Order conditions of stochastic Runge-Kutta methods by B-series. *SIAM J. Numer. Anal.*, 38(5):1626–1646, 2000.

[16] J. C. Butcher. The effective order of Runge-Kutta methods. In *Conf. on Numerical Solution of Differential Equations (Dundee, 1969)*, pages 133–139. Springer, Berlin, 1969.

[17] J. C. Butcher. An algebraic theory of integration methods. *Math. Comp.*, 26:79–106, 1972.

[18] J. C. Butcher. *Numerical methods for ordinary differential equations.* John Wiley & Sons, Ltd., Chichester, third edition, 2016.

[19] J. C. Butcher. *B-series: algebraic analysis of numerical methods.* Springer, 2021.

[20] D. Calaque, K. Ebrahimi-Fard, and D. Manchon. Two interacting Hopf algebras of trees: a Hopf-algebraic approach to composition and substitution of B-series. *Adv. in Appl. Math.*, 47(2):282–308, 2011.

[21] M. P. Calvo, A. Murua, and J. M. Sanz-Serna. Modified equations for ODEs. In *Chaotic numerics (Geelong, 1993)*, volume 172 of *Contemp. Math.*, pages 63–74. Amer. Math. Soc., Providence, RI, 1994.

[22] P. Chartier, E. Hairer, and G. Vilmart. A substitution law for B-series vector fields. Research Report RR-5498, INRIA, 2005.

[23] P. Chartier, E. Hairer, and G. Vilmart. Numerical integrators based on modified differential equations. *Math. Comp.*, 76(260):1941–1953, 2007.

[24] P. Chartier, E. Hairer, and G. Vilmart. Algebraic structures of B-series. *Found. Comput. Math.*, 10(4):407–427, 2010.

[25] P. Chartier and A. Murua. Preserving first integrals and volume forms of additively split systems. *IMA J. Numer. Anal.*, 27(2):381–405, 2007.

[26] C. Curry, K. Ebrahimi-Fard, D. Manchon, and H. Z. Munthe-Kaas. Planarly branched rough paths and rough differential equations on homogeneous spaces. *J. Differential Equations*, 269(11):9740–9782, 2020.

[27] K. Debrabant. Runge-Kutta methods for third order weak approximation of SDEs with multidimensional additive noise. *BIT Numer. Math.*, 50(3):541–558, 2010.

[28] K. Debrabant and A. Kværnø. B-series analysis of stochastic Runge-Kutta methods that use an iterative scheme to compute their internal stage values. *SIAM J. Numer. Anal.*, 47(1):181–203, 2008/09.

[29] K. Debrabant and A. Kværnø. Composition of stochastic B-series with applications to implicit Taylor methods. *Appl. Numer. Math.*, 61(4):501–511, 2011.

[30] A. Debussche and E. Faou. Weak backward error analysis for SDEs. *SIAM J. Numer. Anal.*, 50(3):1735–1752, 2012.

[31] K. Ebrahimi-Fard and L. Rahm. Primitive elements in the Munthe-Kaas-Wright Hopf algebra. *Submitted*, 2023.

[32] E. Faou and T. Lelièvre. Conservative stochastic differential equations: mathematical and numerical analysis. *Math. Comp.*, 78(268):2047–2074, 2009.

[33] G. Fløystad, D. Manchon, and H. Z. Munthe-Kaas. The universal pre-Lie-Rinehart algebras of aromatic trees. In *Geometric and harmonic analysis on homogeneous spaces and applications*, volume 366 of *Springer Proc. Math. Stat.*, pages 137–159. Springer, Cham, [2021] ©2021.

[34] L. Foissy. Algebraic structures on typed decorated rooted trees. *SIGMA Symmetry Integrability Geom. Methods Appl.*, 17:Paper No. 086, 28, 2021.

[35] P. K. Friz and M. Hairer. *A course on rough paths.* Universitext. Springer, Cham, [2020] ©2020. With an introduction to regularity structures, Second edition of [3289027].

[36] S. D. Giovacchino, D. J. Higham, and K. Zygalakis. Backward error analysis and the qualitative behaviour of stochastic optimization algorithms: Application to stochastic coordinate descent. *Submitted*, 2023.

[37] M. Gubinelli. Ramification of rough paths. *J. Differential Equations*, 248(4):693–721, 2010.

[38] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006. Structure-preserving algorithms for ordinary differential equations.

[39] E. Hairer and G. Wanner. On the Butcher group and general multi-value methods. *Computing (Arch. Elektron. Rechnen)*, 13(1):1–15, 1974.

[40] M. Hairer and D. Kelly. Geometric versus non-geometric rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(1):207–251, 2015.

[41] R. Z. Hasminskii. *Stochastic stability of differential equations*, volume 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn—Germantown, Md., 1980. Translated from the Russian by D. Louvish.

[42] M. E. Hoffman. Combinatorics of rooted trees and Hopf algebras. *Trans. Amer. Math. Soc.*, 355(9):3795–3811, 2003.

[43] A. Iserles, G. R. W. Quispel, and P. S. P. Tse. B-series methods cannot be volume-preserving. *BIT Numer. Math.*, 47(2):351–378, 2007.

[44] J.-D. Jacques and L. Zambotti. Post-Lie algebras of derivations and regularity structures. *Submitted*, arXiv:2306.02484, 2023.

[45] Y. Komori, T. Mitsui, and H. Sugiura. Rooted tree analysis of the order conditions of ROW-type scheme for stochastic differential equations. *BIT Numer. Math.*, 37(1):43–66, 1997.

[46] M. Kopec. Weak backward error analysis for Langevin process. *BIT Numer. Math.*, 55(4):1057–1103, 2015.

[47] M. Kopec. Weak backward error analysis for overdamped Langevin processes. *IMA J. Numer. Anal.*, 35(2):583–614, 2015.

[48] D. Küpper, A. Kværnø, and A. Rößler. A Runge-Kutta method for index 1 stochastic differential-algebraic equations with scalar noise. *BIT Numer. Math.*, 52(2):437–455, 2012.

[49] A. Laurent. *Algebraic Tools and Multiscale Methods for the Numerical Integration of Stochastic Evolutionary Problems.* PhD thesis, University of Geneva, 2021.

[50] A. Laurent. The Lie derivative and Noether's theorem on the aromatic bicomplex for the study of volume-preserving numerical integrators. *J. Comput. Dyn.*, 11(1):10–22, 2024.

[51] A. Laurent, R. I. McLachlan, H. Z. Munthe-Kaas, and O. Verdier. The aromatic bicomplex for the description of divergence-free aromatic forms and volume-preserving integrators. *Forum Math. Sigma*, 11:Paper No. e69, 2023.

[52] A. Laurent and H. Z. Munthe-Kaas. The universal equivariance properties of exotic aromatic B-series. *Submitted*, arXiv:2305.10993, 2023.

[53] A. Laurent and G. Vilmart. Exotic aromatic B-series for the study of long time integrators for a class of ergodic SDEs. *Math. Comp.*, 89(321):169–202, 2020.

[54] A. Laurent and G. Vilmart. Order conditions for sampling the invariant measure of ergodic stochastic differential equations on manifolds. *Found. Comput. Math.*, 22(3):649–695, 2022.

[55] B. Leimkuhler and C. Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Appl. Math. Res. Express. AMRX*, 2013(1):34–56, 2013.

[56] B. Leimkuhler, C. Matthews, and G. Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.*, 36(1):13–79, 2016.

[57] A. Lejay. Constructing general rough differential equations through flow approximations. *Electron. J. Probab.*, 27:Paper No. 7, 24, 2022.

[58] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations*. Imperial College Press, London, 2010. A mathematical perspective.

[59] Q. Li, C. Tai, and W. E. Stochastic modified equations and dynamics of stochastic gradient algorithms I: mathematical foundations. *J. Mach. Learn. Res.*, 20:Paper No. 40, 47, 2019.

[60] Y. Li, Y. Sheng, and R. Tang. Post-Hopf algebras, relative Rota-Baxter operators and solutions of the Yang-Baxter equation. *Journal of Noncommutative Geometry*, 2023.

[61] J.-L. Loday and M. Ronco. Combinatorial Hopf algebras. In *Quanta of maths*, volume 11 of *Clay Math. Proc.*, pages 347–383. Amer. Math. Soc., Providence, RI, 2010.

[62] J.-H. Lu. Hopf algebroids and quantum groupoids. *Internat. J. Math.*, 7(1):47–70, 1996.

[63] A. Lundervold and H. Munthe-Kaas. Hopf algebras of formal diffeomorphisms and numerical integration on manifolds. In *Combinatorics and physics*, volume 539 of *Contemp. Math.*, pages 295–324. Amer. Math. Soc., Providence, RI, 2011.

[64] A. Lundervold and H. Munthe-Kaas. Backward error analysis and the substitution law for Lie group integrators. *Found. Comput. Math.*, 13(2):161–186, 2013.

[65] A. Lundervold and H. Z. Munthe-Kaas. On algebraic structures of numerical integration on vector spaces and manifolds. In *Faà di Bruno Hopf algebras, Dyson-Schwinger equations, and Lie-Butcher series*, volume 21 of *IRMA Lect. Math. Theor. Phys.*, pages 219–263. Eur. Math. Soc., Zürich, 2015.

[66] D. Manchon and A. Saïdi. Lois pré-Lie en interaction. *Comm. Algebra*, 39(10):3662–3680, 2011.

[67] R. I. McLachlan, K. Modin, H. Munthe-Kaas, and O. Verdier. B-series methods are exactly the affine equivariant methods. *Numer. Math.*, 133(3):599–622, 2016.

[68] R. I. McLachlan, K. Modin, H. Munthe-Kaas, and O. Verdier. Butcher series: a story of rooted trees and numerical methods for evolution equations. *Asia Pac. Math. Newsl.*, 7(1):1–11, 2017.

[69] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[70] I. Moerdijk and J. Mrčun. On the universal enveloping algebra of a Lie algebroid. *Proc. Amer. Math. Soc.*, 138(9):3135–3145, 2010.

[71] H. Munthe-Kaas and O. Verdier. Aromatic Butcher series. *Found. Comput. Math.*, 16(1):183–215, 2016.

[72] H. Z. Munthe-Kaas and W. M. Wright. On the Hopf algebraic structure of Lie group integrators. *Found. Comput. Math.*, 8(2):227–257, 2008.

[73] H. Z. Munthe-Kaas and W. M. Wright. On the Hopf algebraic structure of Lie group integrators. *Found. Comput. Math.*, 8(2):227–257, 2008.

[74] J.-M. Oudom and D. Guin. On the Lie enveloping algebra of a pre-Lie algebra. *J. K-Theory*, 2(1):147–167, 2008.

[75] L. Rahm. An operadic approach to substitution in Lie-Butcher series. *Forum Math. Sigma*, 10:Paper No. e20, 29, 2022.

[76] A. Rößler. Stochastic Taylor expansions for the expectation of functionals of diffusion processes. *Stochastic Anal. Appl.*, 22(6):1553–1576, 2004.

[77] A. Rößler. Rooted tree analysis for order conditions of stochastic Runge-Kutta methods for the weak approximation of stochastic differential equations. *Stoch. Anal. Appl.*, 24(1):97–134, 2006.

[78] A. Rößler. Runge-Kutta methods for Itô stochastic differential equations with scalar noise. *BIT Numer. Math.*, 46(1):97–110, 2006.

[79] A. Rößler. Stochastic Taylor expansions for functionals of diffusion processes. *Stoch. Anal. Appl.*, 28(3):415–429, 2010.

[80] A. Saïdi. On a pre-Lie algebra defined by insertion of rooted trees. *Lett. Math. Phys.*, 92(2):181–196, 2010.

[81] A. Saïdi. *Algèbres de Hopf d'arbres et structures pré-Lie*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II, 2011.

[82] J. M. Sanz-Serna. Markov chain Monte Carlo and numerical differential equations. In *Current challenges in stability issues for numerical differential equations*, volume 2082 of *Lecture Notes in Math.*, pages 39–88. Springer, Cham, 2014.

[83] T. Shardlow. Modified equations for stochastic differential equations. *BIT Numer. Math.*, 46(1):111–125, 2006.

[84] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.

[85] G. Vilmart. Postprocessed integrators for the high order integration of ergodic SDEs. *SIAM J. Sci. Comput.*, 37(1):A201–A220, 2015.

[86] K. C. Zygalakis. On the existence and the applications of modified equations for stochastic differential equations. *SIAM J. Sci. Comput.*, 33(1):102–130, 2011.

Appendices

A   Examples for forests of order up to three

We present in Tables 1 and 2 the Connes-Kreimer coproduct and the substitution law of all primitive exotic aromatic forests $\mathrm{Prim}(\mathcal{EAF}, \Delta_{\mathcal{EA}})$ of order up to three. To compute the Connes-Kreimer coproduct efficiently, we use Remark 2.7 which implies the property

$$\Delta_{CK}(\pi_1 \cdot \pi_2) = \Delta_{CK}(\pi_1) \hat{\cdot} \Delta_{CK}(\pi_2),$$

where $(x \otimes y) \hat{\cdot} (u \cdot v) = (x \cdot u) \otimes (y \cdot v)$ and $\pi_1, \pi_2 \in \mathcal{AF}_D$. Therefore, the values of the Connes-Kreimer coproduct on primitive exotic aromatic forests are enough to compute the values for general exotic aromatic forests in a straightforward way.

Table 1: Connes-Kreimer coproduct for all primitive exotic aromatic forests up to order three

π	$\Delta_{CK}(\pi)$

(Table content consists of graphical aromatic-forest diagrams in both columns; the mathematical structure of each row expresses the Connes-Kreimer coproduct $\Delta_{CK}(\pi) = \mathbf{1}\otimes\pi + \cdots + \pi\otimes\mathbf{1}$ for the corresponding primitive exotic aromatic forest.)

We refer to the discussion from Section 3 for an efficient way to compute the CEM coaction. The computations can be checked using Proposition 2.20 with the property

$$\langle \Delta_{CEM}(\pi), \pi_1 \otimes \pi_2 \rangle = \frac{\sigma(\pi)\langle \pi_1 \triangleright \pi_2, \pi \rangle}{\sigma(\pi_1)\sigma(\pi_2)}.$$

Table 2: CEM coaction for all exotic aromas and primitive exotic aromatic forests up to order three

π	$\Delta_{CEM}(\pi)$

40