

RESEARCH ARTICLE

Unified signature cumulants and generalized Magnus expansions

Peter K. Friz ^{1,2}, Paul P. Hager ³ and Nikolas Tapia ^{1,2}

¹Institut für Mathematik, Technische Universität Berlin, Str. des 17. Juni 136, Berlin 10586, Germany;
E-mail: friz@math.tu-berlin.de.

²Weierstraß-Institut für Angewandte Analysis und Stochastik, Mohrenstr. 39, Berlin 10117, Germany;
E-mail: tapia@wias-berlin.de.

³Institut für Mathematik, Humboldt Universität zu Berlin, Unter den Linden 6, Berlin 10099, Germany;
E-mail: paul.hager@hu-berlin.de.

Received: 5 February 2021; **Revised:** 4 March 2022; **Accepted:** 24 March 2022

Keywords: Signatures, Lévy processes, Markov processes, stochastic Volterra processes, universal signature relations for semimartingales, moment-cumulant relations, characteristic functions, diamond product, Magnus expansion

2020 Mathematics subject classification: *Primary* – 60L10; *Secondary* – 60L90, 60E10, 60G44, 60G48, 60G51, 60J76

Abstract

The signature of a path can be described as its full non-commutative exponential. Following T. Lyons, we regard its expectation, the *expected signature*, as a path space analogue of the classical moment generating function. The logarithm thereof, taken in the tensor algebra, defines the *signature cumulant*. We establish a universal functional relation in a general semimartingale context. Our work exhibits the importance of Magnus expansions in the algorithmic problem of computing expected signature cumulants and further offers a far-reaching generalization of recent results on characteristic exponents dubbed diamond and cumulant expansions with motivations ranging from financial mathematics to statistical physics. From an affine semimartingale perspective, the functional relation may be interpreted as a type of generalized Riccati equation.

Contents

1	Introduction and main results	2
2	Preliminaries	6
2.1	The tensor algebra and tensor series	6
2.2	The outer tensor algebra	8
2.3	Some quotients of the tensor algebra	8
2.4	Semimartingales	9
2.5	Diamond products	12
2.6	Generalized signatures	13
3	Expected signatures and signature cumulants	14
3.1	Definitions and existence	14
3.2	Moments and cumulants	15
4	Main results	17
4.1	Functional equation for signature cumulants	17
4.2	Recursive formulas for signature cumulants	18

5	Two special cases	21
5.1	Variations on Hausdorff, Magnus and Baker–Campbell–Hausdorff	21
5.2	Diamond relations for multivariate cumulants	23
5.3	Remark on tree representation	25
6	Applications	26
6.1	Brownian and stopped Brownian signature cumulants	26
6.2	Lévy and diffusion processes	30
6.3	Affine Volterra processes	40
7	Proofs	41
7.1	Proof of Theorem 3.2	45
7.2	Proof of Theorem 4.1	46

1. Introduction and main results

Write $\mathcal{T} := T((\mathbb{R}^d)) = \Pi_{k \geq 0}(\mathbb{R}^d)^{\otimes k}$ for the tensor series over \mathbb{R}^d , equipped with a concatenation product, elements of which are written indifferently as

$$\mathbf{x} = (\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots) \equiv \mathbf{x}^{(0)} + \mathbf{x}^{(1)} + \mathbf{x}^{(2)} + \dots$$

The affine subspace $\overline{\mathcal{T}}_0$ (respectively, \mathcal{T}_1) with scalar component $\mathbf{x}^{(0)} = 0$ (respectively, $= 1$) has a natural Lie algebra (respectively, formal Lie group) structure, with the comutator Lie-bracket $[\mathbf{y}, \mathbf{x}] = \mathbf{y}\mathbf{x} - \mathbf{x}\mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \overline{\mathcal{T}}_0$, where \mathbf{xy} stands for the concatenation product.¹

Let further $\mathcal{S} = \mathcal{S}(\mathbb{R}^d)$, respectively, $\mathcal{S}^c = \mathcal{S}^c(\mathbb{R}^d)$, denote the class of càdlàg,² respectively, continuous, d -dimensional semimartingales on some filtered probability space $(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. We recall that this is the somewhat decisive class of stochastic processes that allows for a reasonable, stochastic integration theory. Classic texts include [59, 43] (continuous semimartingale) and [36, 57] (càdlàg semimartingales); a concise introduction for readers with no background in stochastic analysis is [2], Chapter 1.³ (Readers with no background in probability may also focus in a first reading on deterministic semimartingales; these are precisely càdlàg paths of finite variation on compacts.) Following Lyons [45], for a continuous semimartingale $X \in \mathcal{S}^c$, the *signature* given as the formal sum of iterated Stratonovich integrals,

$$\text{Sig}(X)_{s,t} = 1 + X_{s,t} + \int_s^t X_{s,u} \circ dX_u + \int_s^t \left(\int_s^{u_1} X_{s,u_2} \circ dX_{u_2} \right) \circ dX_{u_1} + \dots$$

for $0 \leq s \leq t$ defines a random element in \mathcal{T}_1 and, as a process, a formal \mathcal{T}_1 -valued semimartingale. By regarding the d -dimensional semimartingale X as a $\overline{\mathcal{T}}_0$ -valued semimartingale ($X \leftrightarrow \mathbf{X} = (0, X, 0, \dots)$), we see that the signature of $X \in \mathcal{S}^c$ satisfies the Stratonovich stochastic differential equation

$$dS = S \circ d\mathbf{X}. \tag{1.1}$$

In the general case of $\mathbf{X} \in \mathcal{S}(\overline{\mathcal{T}}_0)$ with possibly nontrivial higher-order semimartingale components $\mathbf{X} = (0, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots)$, the solution to equation (1.1) is also known as the Lie group valued stochastic exponential (or development), with classical references [51, 31]; the càdlàg case [20] is consistent with the geometric or Marcus [49, 50, 41, 4, 25] interpretation of equation (1.1)⁴ with jump behavior $S_t = e^{\Delta \mathbf{X}_t} S_{t-}$. From a stochastic differential geometry point of view, one aims for an intrinsic understanding of equation (1.1) valid for arbitrary Lie groups. For instance, if \mathbf{X} takes values in any sub Lie algebra $\mathcal{L} \subset \overline{\mathcal{T}}_0$, then S takes values in the group $\mathcal{G} = \exp \mathcal{L}$. In case of a d -dimensional semimartingale X , the

¹Also known as a tensor product; see Section 2.1 for a precise definition.

²Continuous from the right, limits from the left.

³At the time of writing, available as a free sample chapter at <https://www.princeton.edu/~yacine/research>.

⁴Diamond notation for Marcus SDEs, $dS = S \circ d\mathbf{X}$ (see also [4]) will not be used here to avoid notational clash with the diamond product, introduced and studied in [3] and [23], respectively.

minimal choice is the free Lie algebra $\text{Lie}(\mathbb{R}^d)$ spanned by \mathbb{R}^d (see, for example, [58]), and the resulting Lie algebra structure of iterated integrals (both in the smooth and Stratonovich semimartingale case) is well-known. The extrinsic linear ambient space $\mathcal{T} \supset \exp \mathcal{L}$ will be important to us. Indeed, writing $S_t = \text{Sig}(\mathbf{X})_{0,t}$ for the (unique, global) \mathcal{T}_1 -valued solution of equation (1.1) driven by \mathcal{T}_0 -valued \mathbf{X} , started at $S_0 = 1$, we define, whenever $\text{Sig}(\mathbf{X})_{0,T}$ is (componentwise) integrable, the *expected signature* and *signature cumulants* (SigCum)

$$\boldsymbol{\mu}(T) := \mathbb{E}(\text{Sig}(\mathbf{X})_{0,T}) \in \mathcal{T}_1, \quad \boldsymbol{\kappa}(T) := \log \boldsymbol{\mu}(T) \in \mathcal{T}_0.$$

Already when \mathbf{X} is deterministic, and sufficiently regular to make equation (1.1) meaningful, this leads to an interesting (ordinary differential) equation for $\boldsymbol{\kappa}$ with accompanying (Magnus) expansion, well understood as an effective computational tool [34, 6]. The importance of the stochastic case $\mathbf{X} = \mathbf{X}(\omega)$, with expectation and logarithm thereof, was developed by Lyons and coauthors; see [45] and references therein, with a variety of applications, ranging from machine learning to numerical algorithms on Wiener space known as cubature [47]; signature cumulants were named and first studied in their own right in [7]. The joint appearance of cumulants and Magnus type-expansion is also seen in non-commutative probability [10], although the methods and aims appear quite different.⁵

In the special case of $d = 1$ and $\mathbf{X} = (0, X, 0, \dots)$, where X is a scalar semimartingale, $\boldsymbol{\mu}(T)$ and $\boldsymbol{\kappa}(T)$ are nothing but the sequence of moments and cumulants of the real valued random variable $X_T - X_0$. When $d > 1$, the expected signature / cumulants provides an effective way to describe the process X on $[0, T]$; see [44, 45, 13]. The question arises how to compute them. If one takes \mathbf{X} as d -dimensional Brownian motion, the signature cumulant $\boldsymbol{\kappa}(T)$ equals $(T/2)\mathbf{I}_d$, where \mathbf{I}_d is the identity 2-tensor over \mathbb{R}^d . This is known as *Fawcett’s formula* [47, 24]. Loosely speaking, and postponing precise definitions, our main result is a vast generalization of Fawcett’s formula.

Theorem 1.1 (FunctEqu \mathcal{S} -SigCum). *For sufficiently integrable $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0)$, the (time- t) conditional signature cumulants $\boldsymbol{\kappa}_t(T) \equiv \boldsymbol{\kappa}_t := \log \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T})$ is the unique solution of the functional equation*

$$\begin{aligned} \boldsymbol{\kappa}_t(T) = \mathbb{E}_t \left\{ \int_{(t,T]} H(\text{ad } \boldsymbol{\kappa}_{u-})(d\mathbf{X}_u) + \frac{1}{2} \int_t^T H(\text{ad } \boldsymbol{\kappa}_{u-})(d\langle \mathbf{X}^c \rangle_u) \right. \\ \left. + \frac{1}{2} \int_t^T H(\text{ad } \boldsymbol{\kappa}_{u-}) \circ Q(\text{ad } \boldsymbol{\kappa}_{u-})(d\|\boldsymbol{\kappa}, \boldsymbol{\kappa}\|_u^c) + \int_t^T H(\text{ad } \boldsymbol{\kappa}_{u-}) \circ (\text{Id} \odot G(\text{ad } \boldsymbol{\kappa}_{u-}))(d\|\mathbf{X}, \boldsymbol{\kappa}\|_u^c) \right. \\ \left. + \sum_{t < u \leq T} \left(H(\text{ad } \boldsymbol{\kappa}_{u-}) \left(\exp(\Delta \mathbf{X}_u) \exp(\boldsymbol{\kappa}_u) \exp(-\boldsymbol{\kappa}_{u-}) - 1 - \Delta \mathbf{X}_u \right) - \Delta \boldsymbol{\kappa}_u \right) \right\}, \end{aligned} \tag{1.2}$$

where all integrals are understood in an Itô- and Riemann–Stieltjes sense, respectively,⁶ and $\text{ad } \mathbf{x} = [\mathbf{x}, \cdot] : \mathcal{T}_0 \rightarrow \mathcal{T}_0$ denotes the adjointed operator associated to $\mathbf{x} \in \mathcal{T}_0$. The functions H, G, Q are defined in equation (4.1) below; see also Section 2 for further notation.

As displayed in Figures 1 and 2, this theorem has an avalanche of consequences on which we now comment.

- Equation (1.2) allows to compute $\boldsymbol{\kappa}^{(n)} \in (\mathbb{R}^d)^{\otimes n}$ as a function of $\boldsymbol{\kappa}^{(1)}, \dots, \boldsymbol{\kappa}^{(n-1)}$. (This remark applies *mutatis mutandis* to all special cases seen as vertices in Figure 1.) The resulting expansions, displayed in Figure 2, are of computational interest. In particular, our approach allows us, in some special cases, either to derive closed expressions for the conditional cumulant series $\boldsymbol{\kappa}_t$ or to characterize it as the unique solution of a certain parabolic PDE (see Section 6). In any case, this provides a means of computation that can in principle be more efficient than the naïve Monte Carlo approach. Even when such a concrete form of $\boldsymbol{\kappa}_t$ is not available, the recursive nature of the expressions for each homogeneous component can be useful in some numerical scenarios.

⁵We thank the participants of *Cumulants in Stochastic Analysis*, TU Berlin (online), February 2021, for many discussions.
⁶Here, \circ denotes composition, not to be confused with Stratonovich integration $\circ d\mathbf{X}$.

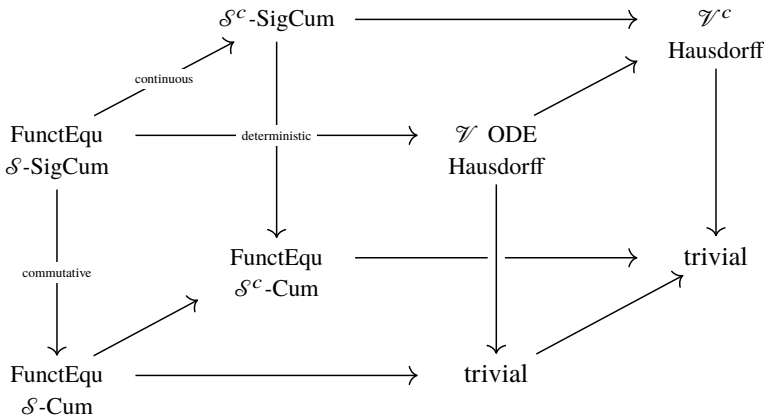


Figure 1. FunctEqu \mathcal{S} -SigCum (Theorem 4.1) and implications. \mathcal{S} (respectively, \mathcal{S}^c) stands for general (respectively, continuous) semimartingales and \mathcal{V} (respectively, \mathcal{V}^c) stands for finite variation (respectively, finite variation and continuous) processes.

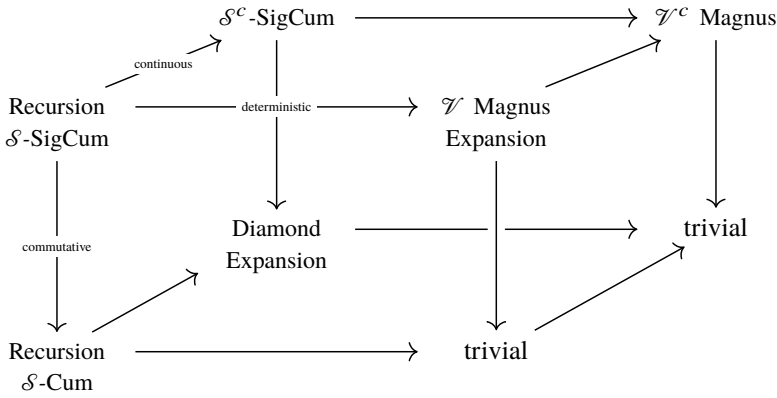


Figure 2. Computational consequence: accompanying recursions.

- o The most classical consequence of equation (1.2) appears when \mathbf{X} is a deterministic continuous semimartingale: that is, in particular, the components of \mathbf{X} are continuous paths of finite variation, which also covers the absolutely continuous case with integrable componentwise derivative $\dot{\mathbf{X}}$. In this case all bracket terms and the final jump-sum disappear. What remains is a classical differential equation due to [32], here in backward form:

$$-d\kappa_t(T) = H(\text{ad } \kappa_t)d\mathbf{X}_t. \tag{1.3}$$

The accompanying expansion is then precisely Magnus expansion [48, 33, 34, 6]. By taking \mathbf{X} continuous and piecewise linear on two adjacent intervals, say $[0, 1) \cup [1, 2)$, one obtains the Baker–Campbell–Hausdorff formula (see, e.g., [52, Theorem 5.5])

$$\begin{aligned} \kappa_0(2) &= \log(\exp(\mathbf{x}_1) \exp(\mathbf{x}_2)) =: \text{BCH}(\mathbf{x}_1, \mathbf{x}_2) \\ &= \mathbf{x}_2 + \int_0^1 \Psi(\exp(\text{ad } t\mathbf{x}_1) \circ \exp(\text{ad } \mathbf{x}_2))(\mathbf{x}_1) dt, \end{aligned} \tag{1.4}$$

with

$$\Psi(z) := \frac{\ln(z)}{z-1} = \sum_{n \geq 0} \frac{(-1)^n}{n+1} (z-1)^n.$$

It is also instructive to let \mathbf{X} piecewise constant on these intervals, with $\Delta \mathbf{X}_1 = \mathbf{x}_1, \Delta \mathbf{X}_2 = \mathbf{x}_2$, in which case equation (1.2) reduces to the first equality in equation (1.4). Such jump variations of the Magnus expansion are discussed in Section 5.1.

- Write $\pi_{\text{Sym}} : \mathcal{T} \rightarrow \mathcal{S}$ for the canonical projection to the extended symmetric algebra \mathcal{S} , the linear space identified with symmetric tensor series, and define the \mathcal{S} -valued semimartingale $\hat{\mathbf{X}} := \pi_{\text{Sym}}(\mathbf{X})$ and symmetric signature cumulants $\hat{\kappa}(T) := \log(\mathbb{E} . (\text{Sig}(\hat{\mathbf{X}})_{\cdot, T})) = \pi_{\text{Sym}}(\kappa(T))$ (see Section 2.3.1 for more detail). Then equation (1.2), in its projected and commutative form becomes (see also Section 5.2)

$$\begin{aligned} \text{FuncEqu } \mathcal{S}\text{-Cum: } \hat{\kappa}_t(T) &= \mathbb{E}_t \left\{ \hat{\mathbf{X}}_{t, T} + \frac{1}{2} \langle (\hat{\mathbf{X}} + \hat{\kappa})^c \rangle_{t, T} \right. \\ &\quad \left. + \sum_{t < u \leq T} \left(\exp \left(\Delta \hat{\mathbf{X}}_u + \Delta \hat{\kappa}_u \right) - 1 - (\Delta \hat{\mathbf{X}}_u + \Delta \hat{\kappa}_u) \right) \right\}, \end{aligned} \tag{1.5}$$

where $\exp : \mathcal{S}_0 \mapsto \mathcal{S}_1$ is defined by the usual power series. First-level tensors are trivially symmetric, and therefore equation (1.5) applies to the a \mathbb{R}^d -valued semimartingale X via the canonical embedding $\hat{\mathbf{X}} = (0, X, 0, \dots) \in \mathcal{S}(\mathcal{S}_0)$. More interestingly, the case $\hat{\mathbf{X}} = (0, aX, b\langle X \rangle, 0, \dots)$ for a d -dimensional continuous martingale X can be seen to underlie the expansions of [23], which improves and unifies previous results [42, 3] treating $(a, b) = (1, 0)$ and $(a, b) = (1, -1/2)$, respectively. Following Gatheral and coworkers, equation (1.5) and subsequent expansions involve ‘diamond’ products of semimartingales, given, whenever well-defined, by

$$(A \diamond B)_t(T) := \mathbb{E}_t(\langle A^c, B^c \rangle_{t, T}).$$

We note that equation (1.5) induces recursive formulae for cumulants, dubbed Diamond expansions in Figure 2, previously discussed in [42, 3, 23, 28], together with a range of applications, from quantitative finance (including rough volatility models [1, 29]) to statistical physics: in [42], the authors rely on such formulae to compute the cumulants function of log-correlated Gaussian fields, more precisely approximations thereof, that underlies the Sine-Gordon model, which is a key ingredient in their renormalization procedure.

With regard to the existing (‘commutative’) literature, our algebraic setup is ideally suited to work under finite moment assumptions; we are able to deal with jumps, not treated in [42, 3]. Equation (1.5), the commutative shadow of equation (1.2), should be compared with Riccati’s ordinary differential equation from affine process theory [19, 16, 40]. A systematic comparison would lead us too far astray from our main object of study; nevertheless, we illustrate the connection in Remark 6.9. Of course, our results, in particular equations (1.2) and (1.5), are not restricted to affine semimartingales. In turn, expected signatures and cumulants - and subsequently all our statements above these - require moments, which is not required for the Riccati evolution of the characteristic function of affine processes. Of recent interest, explicit diamond expansions have been obtained for ‘rough affine’ processes, non-Markov by nature, with a cumulant generating function characterized by Riccati Volterra equations; see [1, 29, 23]. It is remarkable that analytic tractability remains intact when one passes to path space and considers signature cumulants, as we illustrate in Section 6.3.

- Finally, we mention *Signature-SDEs* [5], tractable classes of stochastic differential equations that can be studied from an infinite dimensional affine and polynomial perspective [18]. Calibration of such models hinges on the efficient computation of expected signatures, which is the very purpose of this paper.

We conclude this introduction with some remarks on *convergence*. As explained, this work contains generalizations of cumulant type recursions, previously studied in [3, 42, 23], the interest therein being the algorithmic computation of cumulants. Basic facts of analytic functions show that *classical* moment- and cumulant-generating functions, for random variables with finite moments of all orders, have a radius of convergence $\rho \geq 0$, directly related to growth of the corresponding sequence. Convergence, in the sense $\rho > 0$, implies that the moment problem is well-posed. That is, the moments (equivalently: cumulants) determine the law of the underlying random variable. (See also [23] for a related discussion in the context of diamond expansions.) The point of view taken here is to work directly on this space of sequences, which is even more natural in the non-commutative setting, as already seen in the deterministic setting of [48]. While *convergence* of expected signatures or signature cumulants series is not directly an interesting question,⁷ understanding their growth most certainly is: in a celebrated paper [13], it was shown that under a growth condition of the expected signature, the ‘expected signature problem’ is well-posed; that is, the expected signature (equivalently: signature cumulants) determines the law of the random signature. With this in mind, it is conceivable that Theorem 1.1 will prove useful toward controlling the growth of signature cumulants (and hence expected signatures).

2. Preliminaries

2.1. The tensor algebra and tensor series

Denote by $T(\mathbb{R}^d)$ the tensor algebra over \mathbb{R}^d : that is,

$$T(\mathbb{R}^d) := \bigoplus_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k},$$

elements of which are *finite* sums (also known as tensor polynomials) of the form

$$\mathbf{x} = \sum_{k \geq 0} \mathbf{x}^{(k)} = \sum_{w \in \mathcal{W}^d} \mathbf{x}^w e_w \tag{2.1}$$

with $\mathbf{x}^{(k)} \in (\mathbb{R}^d)^{\otimes k}$, $\mathbf{x}^w \in \mathbb{R}$ and linear basis vectors $e_w := e_{i_1} \cdots e_{i_k} \in (\mathbb{R}^d)^{\otimes k}$, where w ranges over all words $w = i_1 \cdots i_k \in \mathcal{W}_d$ over the alphabet $\{1, \dots, d\}$. Note $\mathbf{x}^{(k)} = \sum_{|w|=k} \mathbf{x}^w e_w$, where $|w|$ denotes the length a word w . The element $e_\emptyset = 1 \in (\mathbb{R}^d)^{\otimes 0} \cong \mathbb{R}$ is the neutral element of the concatenation (also known as a tensor) product, which is obtained by linear extension of $e_w e_{w'} = e_{ww'}$, where $ww' \in \mathcal{W}_d$ denotes concatenation of two words. We thus have, for $\mathbf{x}, \mathbf{y} \in T(\mathbb{R}^d)$,

$$\mathbf{x}\mathbf{y} = \sum_{k \geq 0} \sum_{\ell=0}^k \mathbf{x}^{(\ell)} \mathbf{y}^{(k-\ell)} = \sum_{w \in \mathcal{W}^d} \left(\sum_{w_1 w_2 = w} \mathbf{x}^{w_1} \mathbf{y}^{w_2} \right) e_w \in T(\mathbb{R}^d).$$

This extends naturally to *infinite* sums, also known as tensor series, elements of the ‘completed’ tensor algebra

$$\mathcal{T} := T((\mathbb{R}^d)) := \prod_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k},$$

which are written as in equation (2.1), but now as formal infinite sums with identical notation and multiplication rules; the resulting algebra \mathcal{T} obviously extends $T(\mathbb{R}^d)$. For any $n \in \mathbb{N}_{\geq 1}$, define the projection to tensor levels by

$$\pi_n : \mathcal{T} \rightarrow (\mathbb{R}^d)^{\otimes n}, \quad \mathbf{x} \mapsto \mathbf{x}^{(n)}.$$

⁷All such series, seen as sequences of their partial sums, ‘converge’ if one equips the space of tensor series with the minimal topology that renders all finite-dimensional projections continuous.

Denote by \mathcal{T}_0 and \mathcal{T}_1 the subspaces of tensor series starting with 0 and 1, respectively; that is, $\mathbf{x} \in \mathcal{T}_0$ (respectively, \mathcal{T}_1) if and only if $\mathbf{x}^0 = 0$ (respectively, $\mathbf{x}^0 = 1$). Restricted to \mathcal{T}_0 and \mathcal{T}_1 , respectively, the exponential and logarithm in \mathcal{T} , defined by the usual series,

$$\begin{aligned} \exp: \mathcal{T}_0 &\rightarrow \mathcal{T}_1, & \mathbf{x} &\mapsto \exp(\mathbf{x}) := 1 + \sum_{k=1}^{\infty} \frac{1}{k!} (\mathbf{x})^k, \\ \log: \mathcal{T}_1 &\rightarrow \mathcal{T}_0, & 1 + \mathbf{x} &\mapsto \log(1 + \mathbf{x}) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (\mathbf{x})^k, \end{aligned}$$

are globally defined and inverse to each other. The vector space \mathcal{T}_0 becomes a Lie algebra with the commutator bracket

$$[\mathbf{x}, \mathbf{y}] := \mathbf{xy} - \mathbf{yx}, \quad \mathbf{x}, \mathbf{y} \in \mathcal{T}_0.$$

Define the adjointed operator associated to an Lie-algebra element $\mathbf{y} \in \mathcal{T}_0$ by

$$\text{ad } \mathbf{y}: \mathcal{T}_0 \rightarrow \mathcal{T}_0, \quad \mathbf{x} \mapsto [\mathbf{y}, \mathbf{x}].$$

The exponential image $\mathcal{T}_1 = \exp(\mathcal{T}_0)$ is a Lie group, at least formally so. We refrain from equipping the infinite-dimensional \mathcal{T}_1 with a differentiable structure, not necessary in view of the ‘locally finite’ nature of the group law $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{xy}$.

Let $(a_k)_{k \geq 1}$ be a sequence of real numbers and $\mathbf{x} \in \mathcal{T}_0$. Then we can always define a linear operator on \mathcal{T}_0 by

$$\left[\sum_{k \geq 0} a_k (\text{ad } \mathbf{x})^k \right]: \mathcal{T}_0 \rightarrow \mathcal{T}_0, \quad \mathbf{y} \mapsto \sum_{k \geq 0} a_k (\text{ad } \mathbf{x})^k (\mathbf{y}),$$

where $(\text{ad } \mathbf{x})^0 = \text{Id}$ is the identity operator and $(\text{ad } \mathbf{x})^n = \text{ad } \mathbf{x} \circ (\text{ad } \mathbf{x})^{n-1}$ for any $n \in \mathbb{N}_{\geq 1}$. Indeed, there is no convergence issue due to the graded structure, as can be seen by projecting to some tensor level $n \in \mathbb{N}_{\geq 1}$

$$\begin{aligned} \pi_n \left(\sum_{k \geq 0} a_k (\text{ad } \mathbf{x})^k (\mathbf{y}) \right) &= \sum_{k=0}^{n-1} a_k \pi_n ((\text{ad } \mathbf{x})^k (\mathbf{y})) \\ &= a_0 \mathbf{y}^{(n)} + \sum_{k=1}^{n-1} a_k \sum_{\|\ell\|=n, |\ell|=k+1} (\text{ad } \mathbf{x}^{(l_2)} \circ \dots \circ \text{ad } \mathbf{x}^{(l_{k+1})}) (\mathbf{y}^{(l_1)}), \end{aligned} \tag{2.2}$$

where the inner summation in the right-hand side is over a finite set of multi-indices $\ell = (l_1, \dots, l_{k+1}) \in (\mathbb{N}_{\geq 1})^{k+1}$, where $|\ell| := k + 1$ and $\|\ell\| := l_1 + \dots + l_{k+1}$. In the following we will simply write $(\text{ad } \mathbf{x} \text{ ad } \mathbf{y}) \equiv (\text{ad } \mathbf{x} \circ \text{ad } \mathbf{y})$ for the composition of adjoint operators. Further, when $\ell = (l_1)$ is a multi-index of length one, we will use the notation $(\text{ad } \mathbf{x}^{(l_2)} \dots \text{ad } \mathbf{x}^{(l_{k+1})}) \equiv \text{Id}$. Note also that the iteration of adjoint operations can be explicitly expanded in terms of left- and right-multiplication, as follows:

$$\text{ad } \mathbf{y}^{(l_2)} \dots \text{ad } \mathbf{y}^{(l_{k+1})} (\mathbf{x}_u^{(l_1)}) = \sum_{I \dot{\cup} J = \{1, \dots, k\}} (-1)^{|J|} \left(\prod_{i \in I} \mathbf{y}^{(l_{i+1})} \right) \mathbf{x}_u^{(l_1)} \left(\prod_{j \in J} \mathbf{y}^{(l_{j+1})} \right). \tag{2.3}$$

For a word $w \in \mathcal{W}_d$ with $|w| > 0$, we define the directional derivative for a function $f: \mathcal{T} \rightarrow \mathbb{R}$ by

$$(\partial_w f)(\mathbf{x}) := \partial_t (f(\mathbf{x} + t e_w)) \Big|_{t=0},$$

for any $\mathbf{x} \in \mathcal{T}$ such that the right-hand derivative exists.

2.2. The outer tensor algebra

Denote by $m : T(\mathbb{R}^d) \otimes T(\mathbb{R}^d) \rightarrow T(\mathbb{R}^d)$ the multiplication (concatenation) map of the tensor algebra. Note that m is linear and, due to the non-commutativity of the tensor product, not symmetric. The map can naturally be extended to linear map $m : T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d) \rightarrow \mathcal{T}$, where $T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d)$ is the following graded algebra:

$$T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d) := \prod_{n=0}^{\infty} \left(\bigoplus_{i=0}^n (\mathbb{R}^d)^{\otimes i} \otimes (\mathbb{R}^d)^{\otimes (n-i)} \right).$$

Note that there is the following natural linear embedding

$$\mathcal{T} \otimes \mathcal{T} \hookrightarrow T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d), \quad \mathbf{x} \otimes \mathbf{y} \mapsto \sum_{n=0}^{\infty} \left(\sum_{i=0}^n \mathbf{x}^{(i)} \otimes \mathbf{y}^{(n-i)} \right).$$

We will of course refrain from explicitly denoting the embedding and simply regard $\mathbf{x} \otimes \mathbf{y}$ as an element in $T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d)$. We emphasize that here \otimes does *not* denote the (inner) tensor product in \mathcal{T} , for which we did not reserve a symbol, but it denotes another (outer) tensor product. We can lift two linear maps $g, f : \mathcal{T} \rightarrow \mathcal{T}$ to a linear map $g \circ f : T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d) \rightarrow \mathcal{T}$ defined by

$$g \circ f := m \circ (g \otimes f).$$

In particular, for all $\mathbf{x}, \mathbf{y} \in \mathcal{T}$, it holds that

$$(g \circ f)(\mathbf{x} \otimes \mathbf{y}) = g(\mathbf{x})f(\mathbf{y}) \in \mathcal{T}.$$

2.3. Some quotients of the tensor algebra

2.3.1. The symmetric tensor algebra

The *symmetric algebra* over \mathbb{R}^d , denoted by $S(\mathbb{R}^d)$, is the quotient of $T(\mathbb{R}^d)$ by the two-sided ideal I generated by $\{xy - yx : x, y \in \mathbb{R}^d\}$. A linear basis of $S(\mathbb{R}^d)$ is then given by $\{\hat{e}_v\}$ over non-decreasing words, $v = (i_1, \dots, i_n) \in \hat{\mathcal{W}}_d$, with $1 \leq i_1 \leq \dots \leq i_n \leq d, n \geq 0$. Every $\hat{\mathbf{x}} \in S(\mathbb{R}^d)$ can be written as a finite sum,

$$\hat{\mathbf{x}} = \sum_{v \in \hat{\mathcal{W}}_d} \hat{\mathbf{x}}^v \hat{e}_v,$$

and we have an immediate identification with polynomials in d commuting indeterminates. The canonical projection

$$\pi_{\text{Sym}} : T(\mathbb{R}^d) \twoheadrightarrow S(\mathbb{R}^d), \quad \mathbf{x} \mapsto \sum_{w \in \mathcal{W}_d} \mathbf{x}^w \hat{e}_{\hat{w}}, \tag{2.4}$$

where $\hat{w} \in \hat{\mathcal{W}}_d$ denotes the non-decreasing reordering of the letters of the word $w \in \mathcal{W}_d$, is an algebra epimorphism, which extends to an epimorphism $\pi_{\text{Sym}} : \mathcal{T} \twoheadrightarrow \mathcal{S}$, where $\mathcal{S} = S(\widehat{(\mathbb{R}^d)})$ is the algebra completion, identifiable as a formal series in d non-commuting (respectively, commuting) indeterminates. As a vector space, \mathcal{S} can be identified with *symmetric* formal tensor series. Denote by \mathcal{S}_0 and \mathcal{S}_1 the affine space given by those $\hat{\mathbf{x}} \in \mathcal{S}$ with $\hat{\mathbf{x}}^0 = 0$ and $\hat{\mathbf{x}}^0 = 1$, respectively. The usual power series in \mathcal{S} define $\widehat{\text{exp}} : \mathcal{S}_0 \rightarrow \mathcal{S}_1$ with inverse $\widehat{\text{log}} : \mathcal{S}_1 \rightarrow \mathcal{S}_0$, and we have

$$\begin{aligned} \pi_{\text{Sym}} \exp(\mathbf{x} + \mathbf{y}) &= \widehat{\text{exp}}(\hat{\mathbf{x}}) \widehat{\text{exp}}(\hat{\mathbf{y}}), \quad \mathbf{x}, \mathbf{y} \in \mathcal{T}_0 \\ \pi_{\text{Sym}} \log(\mathbf{xy}) &= \widehat{\text{log}}(\hat{\mathbf{x}}) + \widehat{\text{log}}(\hat{\mathbf{y}}), \quad \mathbf{x}, \mathbf{y} \in \mathcal{T}_1. \end{aligned}$$

We shall abuse notation in what follows and write \exp (respectively, \log), instead of $\widehat{\exp}$ (respectively, $\widehat{\log}$).

2.3.2. The (step- n) truncated tensor algebra

For $n \in \mathbb{N}$, the subspace

$$\mathcal{I}_n := \prod_{k=n+1}^{\infty} (\mathbb{R}^d)^{\otimes k}$$

is a two-sided ideal of \mathcal{T} . Therefore, the quotient space $\mathcal{T}/\mathcal{I}_n$ has a natural algebra structure. We denote the *projection map* by $\pi_{(0,n)}$. We can identify $\mathcal{T}/\mathcal{I}_n$ with

$$\mathcal{T}^n := \bigoplus_{k=0}^n (\mathbb{R}^d)^{\otimes k},$$

equipped with truncated tensor product

$$\mathbf{x}\mathbf{y} = \sum_{k=0}^n \sum_{\ell_1+\ell_2=k} \mathbf{x}^{(\ell_1)}\mathbf{y}^{(\ell_2)} = \sum_{w \in \mathcal{V}^d, |w| \leq n} \left(\sum_{w_1 w_2 = w} \mathbf{x}^{w_1}\mathbf{y}^{w_2} \right) e_w \in \mathcal{T}^n.$$

The sequence of algebras $(\mathcal{T}^n : n \geq 0)$ forms an inverse system with limit \mathcal{T} . There are also canonical inclusions $\mathcal{T}^k \hookrightarrow \mathcal{T}^n$ for $k \leq n$; in fact, this forms a direct system with limit $T(\mathbb{R}^d)$. The usual power series in \mathcal{T}^n define $\exp_n : \mathcal{T}_0^n \rightarrow \mathcal{T}_1^n$ with inverse $\log_n : \mathcal{T}_1^n \rightarrow \mathcal{T}_0^n$; we may again abuse notation and write \exp and \log when no confusion arises. (In the proof section Section 7.2, we will stick to the \exp_n notation to emphasis the presence of truncation.) As before, \mathcal{T}_0^n has a natural Lie algebra structure, and \mathcal{T}_1^n (now finite dimensional) is a *bona fide* Lie group.

We equip $T(\mathbb{R}^d)$ with the norm

$$|a|_T(\mathbb{R}^d) := \max_{k \in \mathbb{N}} |a^{(k)}|_{(\mathbb{R}^d)^{\otimes k}},$$

where $|\cdot|_{(\mathbb{R}^d)^{\otimes k}}$ is the Euclidean norm on $(\mathbb{R}^d)^{\otimes k} \cong \mathbb{R}^{d^k}$, which makes it a Banach space. The same norm makes sense in \mathcal{T}^n , and since the definition is consistent in the sense that $|a|_{\mathcal{T}^k} = |a|_{\mathcal{T}^n}$ for any $a \in \mathcal{T}^n$ and $k \geq n$, and $|a|_{\mathcal{T}^n} = |a|_{(\mathbb{R}^d)^{\otimes n}}$ for any $a \in (\mathbb{R}^d)^{\otimes n}$. We will drop the index whenever it is possible and write simply $|a|$.

2.4. Semimartingales

Let \mathcal{D} be the space of adapted càdlàg process $X : \Omega \times [0, T) \rightarrow \mathbb{R}$ with $T \in (0, \infty]$ defined on some filtered probability space $(\Omega, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$. The space of *semimartingales* \mathcal{S} is given by the processes $X \in \mathcal{D}$ that can be decomposed as

$$X_t = X_0 + M_t + A_t,$$

where $M \in \mathcal{M}_{loc}$ is a càdlàg local martingale and $A \in \mathcal{V}$ is a càdlàg adapted process of locally bounded variation, both started at zero. Recall that every $X \in \mathcal{S}$ has a well-defined continuous local martingale part denoted by $X^c \in \mathcal{M}_{loc}^c$. The quadratic variation process of X is then given by

$$[X]_t = \langle X^c \rangle_t + \sum_{0 < u \leq t} (\Delta X_u)^2, \quad 0 \leq t \leq T,$$

where $\langle \cdot \rangle$ denotes the (predictable) quadratic variation of a continuous semimartingale. Covariation square, respectively, angle brackets $[X, Y]$ and $\langle X^c, Y^c \rangle$, for another real-valued semimartingale Y , are defined by polarization. For $q \in [1, \infty)$, write $\mathcal{L}^q = L^q(\Omega, \mathcal{F}, \mathbb{P})$; then a Banach space $\mathcal{H}^q \subset \mathcal{S}$ is given by those $X \in \mathcal{S}$ with $X_0 = 0$ and

$$\|X\|_{\mathcal{H}^q} := \inf_{X=M+A} \left\| [M]_T^{1/2} + \int_0^T |dA_s| \right\|_{\mathcal{L}^q} < \infty.$$

Note that for a local martingale $M \in \mathcal{M}_{\text{loc}}$, it holds that (see [57, Ch. V, p. 245])

$$\|M\|_{\mathcal{H}^q} = \left\| [M]_T^{1/2} \right\|_{\mathcal{L}^q}.$$

For a process $X \in \mathcal{D}$, we define

$$\|X\|_{\mathcal{S}^q} := \left\| \sup_{0 \leq t \leq T} |X_t| \right\|_{\mathcal{L}^q}$$

and define the space $\mathcal{S}^q \subset \mathcal{S}$ of semimartingales $X \in \mathcal{S}$ such that $\|X\|_{\mathcal{S}^q} < \infty$. Note that there exists a constant $c_q > 0$ depending on q such that (see [57, Ch. V, Theorem 2])

$$\|X\|_{\mathcal{S}^q} \leq c_q \|X\|_{\mathcal{H}^q}. \tag{2.5}$$

We view d -dimensional semimartingales, $X = \sum_{i=1}^d X^i e_i \in \mathcal{S}(\mathbb{R}^d)$, as special cases of tensor series valued semimartingales $\mathcal{S}(\mathcal{T})$ of the form

$$\mathbf{X} = \sum_{w \in \mathcal{W}_d} \mathbf{X}^w e_w$$

with each component \mathbf{X}^w a real-valued semimartingale. This extends mutatis mutandis to the spaces of \mathcal{T} -valued adapted càdlàg processes $\mathcal{D}(\mathcal{T})$, martingales $\mathcal{M}(\mathcal{T})$ and adapted càdlàg processes with finite variation $\mathcal{V}(\mathcal{T})$. Note also that we typically deal with \mathcal{T}_0 -valued semimartingales, which amounts to having only words with length $|w| \geq 1$. Standard notions such as continuous local martingale \mathbf{X}^c and jump process $\Delta \mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-}$ are defined componentwise.

Brackets: Now let \mathbf{X} and \mathbf{Y} be \mathcal{T} -valued semimartingales. We define the (non-commutative) *outer quadratic covariation bracket* of \mathbf{X} and \mathbf{Y} by

$$[\mathbf{X}, \mathbf{Y}]_t := \sum_{w_1, w_2 \in \mathcal{W}_d} [\mathbf{X}^{w_1}, \mathbf{Y}^{w_2}]_t e_{w_1} \otimes e_{w_2} \in T(\mathbb{R}^d) \overline{\otimes} T(\mathbb{R}^d).$$

Similarly, define the (non-commutative) *inner quadratic covariation bracket* by

$$[\mathbf{X}, \mathbf{Y}]_t := m([\mathbf{X}, \mathbf{Y}]) = \sum_{w \in \mathcal{W}_d} \left(\sum_{w_1 w_2 = w} [\mathbf{X}^{w_1}, \mathbf{Y}^{w_2}]_t \right) e_w \in \mathcal{T};$$

for *continuous* \mathcal{T} -valued semimartingales \mathbf{X}, \mathbf{Y} , this coincides with the predictable quadratic covariation

$$\langle \mathbf{X}, \mathbf{Y} \rangle_t := \sum_{w \in \mathcal{W}_d} \left(\sum_{w_1 w_2 = w} \langle \mathbf{X}^{w_1}, \mathbf{Y}^{w_2} \rangle_t \right) e_w \in \mathcal{T}.$$

As usual, we may write $[\mathbf{X}] \equiv [\mathbf{X}, \mathbf{X}]$ and $\langle \mathbf{X} \rangle \equiv \langle \mathbf{X}, \mathbf{X} \rangle$.

\mathcal{H} -spaces: The definition of \mathcal{H}^q -norm naturally extends to tensor valued martingales. More precisely, for $\mathbf{X}^{(n)} \in \mathcal{S}((\mathbb{R}^d)^{\otimes n})$ with $n \in \mathbb{N}_{\geq 1}$ and $q \in [1, \infty)$, we define

$$\|\mathbf{X}^{(n)}\|_{\mathcal{H}^q} := \|\mathbf{X}^{(n)}\|_{\mathcal{H}^q((\mathbb{R}^d)^{\otimes n})} := \inf_{\mathbf{X}^{(n)} = \mathbf{M} + \mathbf{A}} \left\| |\mathbf{M}|_T^{1/2} + |\mathbf{A}|_{1\text{-var}; [0; T]} \right\|_{\mathcal{L}^q},$$

where the infimum is taken over all possible decompositions $\mathbf{X}^{(n)} = \mathbf{M} + \mathbf{A}$ with $\mathbf{M} \in \mathcal{M}_{\text{loc}}((\mathbb{R}^d)^{\otimes n})$ and $\mathbf{A} \in \mathcal{V}((\mathbb{R}^d)^{\otimes n})$, where

$$|\mathbf{A}|_{1\text{-var}; [0; T]} := \sup_{0 \leq t_1 \leq \dots \leq t_k \leq T} \sum_{i=1}^k |\mathbf{A}_{t_{i+1}} - \mathbf{A}_{t_i}| \leq \sum_{w \in \mathcal{W}_d, |w|=n} \int_0^T |dA_s^w|,$$

with the supremum taken over all partitions of the interval $[0, T]$. One may readily check that

$$\|\mathbf{X}^{(n)}\|_{\mathcal{H}^q} \leq \sum_{w \in \mathcal{W}_d, |w|=n} \|X^w\|_{\mathcal{H}^q}; \quad \text{and for } \mathbf{X}^{(n)} \in \mathcal{M}_{\text{loc}} : \|\mathbf{X}^{(n)}\|_{\mathcal{H}^q} = \left\| \left[\mathbf{X}^{(n)} \right]_T \right\|_{\mathcal{L}^q}.$$

Further define the following subspace $\mathcal{H}^{q,N} \subset \mathcal{S}(\mathcal{T}_0^N)$ of homogeneously q -integrable semimartingales

$$\mathcal{H}^{q,N} := \left\{ \mathbf{X} \in \mathcal{S}(\mathcal{T}_0^N) \mid \mathbf{X}_0 = 0, \|\mathbf{X}\|_{\mathcal{H}^{q,N}} < \infty \right\},$$

where for any $\mathbf{X} \in \mathcal{S}(\mathcal{T}^N)$, we define

$$\|\mathbf{X}\|_{\mathcal{H}^{q,N}} := \sum_{n=1}^N \left(\|\mathbf{X}^{(n)}\|_{\mathcal{H}^{qN/n}} \right)^{1/n}.$$

Note that $\|\cdot\|_{\mathcal{H}^{q,N}}$ is sub-additive and positive definite on $\mathcal{H}^{q,N}$, and it is *homogeneous* under dilation in the sense that

$$\|\delta_\lambda \mathbf{X}\|_{\mathcal{H}^{q,N}} = |\lambda| \|\mathbf{X}\|_{\mathcal{H}^{q,N}}, \quad \delta_\lambda \mathbf{X} := (\mathbf{X}^{(0)}, \lambda \mathbf{X}^{(1)}, \dots, \lambda^N \mathbf{X}^{(N)}), \quad \lambda \in \mathbb{R}.$$

We also introduce the following subspace of $\mathcal{S}(\mathcal{T})$:

$$\mathcal{H}^{\infty-}(\mathcal{T}) := \{ \mathbf{X} \in \mathcal{S}(\mathcal{T}) : \mathbf{X}^w \in \mathcal{H}^q, \forall 1 \leq q < \infty, w \in \mathcal{W}_d \}.$$

Note that if $\mathbf{X} \in \mathcal{S}(\mathcal{T})$ such that $\|\mathbf{X}^{(0,N)}\|_{\mathcal{H}^{1,N}} < \infty$ for all $N \in \mathbb{N}_{\geq 1}$, then it also holds that $\mathbf{X} \in \mathcal{H}^{\infty-}(\mathcal{T})$.

Stochastic integrals: We are now going to introduce a notation for the stochastic integration with respect to tensor valued semimartingales. Denote by $\mathcal{L}(\mathcal{T}; \mathcal{T}) = \{f : \mathcal{T} \rightarrow \mathcal{T} \mid f \text{ is linear.}\}$ the space of endomorphism on \mathcal{T} , and let $\mathbf{F} : \Omega \times [0, T] \rightarrow \mathcal{L}(\mathcal{T}; \mathcal{T})$ with $(t, \omega) \mapsto \mathbf{F}_t(\omega; \cdot)$ such that it holds

$$(\mathbf{F}_t(\mathbf{x}))_{0 \leq t \leq T} \in \mathcal{D}(\mathcal{T}), \quad \text{for all } \mathbf{x} \in \mathcal{T} \tag{2.6}$$

$$\text{and } \mathbf{F}_t(\omega; \mathcal{I}_n) \subset \mathcal{I}_n, \quad \text{for all } n \in \mathbb{N}, (\omega, t) \in \Omega \times [0, T], \tag{2.7}$$

where $\mathcal{I}_n \subset \mathcal{T}$ was introduced in Section 2.3.2, consisting of series with tensors of level $n + 1$ and higher. In this case, we can define the stochastic Itô integral (and then analogously the Stratonovich/Marcus integral) of \mathbf{F} with respect to $\mathbf{X} \in \mathcal{S}(\mathcal{T})$ by

$$\int_{(0, \cdot]} \mathbf{F}_{t-}(d\mathbf{X}_t) := \sum_{w \in \mathcal{W}_d} \sum_{v \in \mathcal{W}_d, |v| \leq |w|} \int_{(0, \cdot]} \mathbf{F}_{t-}(e_v)^w d\mathbf{X}_t^v e_w \in \mathcal{S}(\mathcal{T}). \tag{2.8}$$

For example, let $\mathbf{Y}, \mathbf{Z} \in \mathcal{D}(\mathcal{T})$, and define $\mathbf{F} := \mathbf{Y} \text{Id} \mathbf{Z}$: that is, $\mathbf{F}_t(\mathbf{x}) = \mathbf{Y}_t \mathbf{x} \mathbf{Z}_t$, the concatenation product from the left and right, for all $\mathbf{x} \in \mathcal{T}$. Then we see that \mathbf{F} indeed satisfies the conditions in equations (2.6) and (2.7), and we have

$$\int_{(0, \cdot]} (\mathbf{Y}_{t-} \text{Id} \mathbf{Z}_{t-})(d\mathbf{X}_t) = \int_{(0, \cdot]} \mathbf{Y}_{t-} d\mathbf{X}_t \mathbf{Z}_{t-} = \sum_{w \in \mathcal{W}_d} \left(\sum_{w_1 w_2 w_3 = w} \int_{(0, \cdot]} \mathbf{Z}_{t-}^{w_1} \mathbf{Y}_{t-}^{w_3} d\mathbf{X}_t^{w_2} \right) e_w. \tag{2.9}$$

Another important example is given by $\mathbf{F} = (\text{ad } \mathbf{Y})^k$ for any $\mathbf{Y} \in \mathcal{D}(\mathcal{T}_0)$ and $k \in \mathbb{N}$. Indeed, we immediately see that \mathbf{F} satisfies the condition in equation (2.7); and recalling from equation (2.3) that the iteration of adjoint operations can be expanded in terms of left- and right-multiplication, we also see that \mathbf{F} satisfies equation (2.6). More generally, let $(a_k)_{k=0}^\infty \subset \mathbb{R}$, and let $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0)$; then the following integral

$$\int_{(0, \cdot]} \left[\sum_{k=0}^\infty a_k (\text{ad } \mathbf{Y}_{t-})^k \right] (d\mathbf{X}_t) = \sum_{n=1}^\infty \sum_{k=0}^{n-1} \sum_{\|\ell\|=n, |\ell|=k+1} a_k \int_{(0, \cdot]} \text{ad } \mathbf{Y}_{t-}^{(\ell_2)} \cdots \text{ad } \mathbf{Y}_{t-}^{(\ell_{k+1})} (d\mathbf{X}_t^{(\ell_1)}) \tag{2.10}$$

is well defined in the sense of equation (2.9). The definition of the integral with integrands of the form $\mathbf{F} : \Omega \times [0, T] \rightarrow \mathcal{L}(T(\mathbb{R}^d) \otimes T(\mathbb{R}^d); \mathcal{T})$ with respect to processes $\mathbf{X} \in \mathcal{S}(T(\mathbb{R}^d) \otimes T(\mathbb{R}^d))$ is completely analogous.

Quotient algebras: All of this extends in a straightforward way to the case of semimartingales in the quotient algebra of Section 2.3: that is, symmetric and truncated algebra, in particular given \mathbf{X} and \mathbf{Y} in $\mathcal{S}(\mathcal{S})$ have well-defined continuous local martingale parts denoted by $\mathbf{X}^c, \mathbf{Y}^c$, respectively, with *inner* (predictable) quadratic covariation given by

$$\langle \mathbf{X}^c, \mathbf{Y}^c \rangle = \sum_{w_1, w_2 \in \widehat{\mathcal{W}}_d} \langle \mathbf{X}^{w_1, c}, \mathbf{Y}^{w_2, c} \rangle \hat{e}_{w_1} \hat{e}_{w_2}.$$

Write \mathcal{S}^N for the truncated symmetric algebra, linearly spanned by $\{\hat{e}_w : w \in \widehat{\mathcal{W}}_d, |w| \leq N\}$ and \mathcal{S}_0^N for those elements with zero scalar entry. In complete analogy with non-commutative setting discussed above, we then write $\widehat{\mathcal{H}}^{q, N} \subset \mathcal{S}(\mathcal{S}_0^N)$ for the corresponding space of homogeneously q -integrable semimartingales.

2.5. Diamond products

We extend the notion of the *diamond product* introduced in [3] for continuous scalar semimartingales to our setting. Denote by \mathbb{E}_t the conditional expectation with respect to the sigma algebra \mathcal{F}_t . Then we have the following:

Definition 2.1. For \mathbf{X} and \mathbf{Y} in $\mathcal{S}(\mathcal{T})$, define

$$(\mathbf{X} \diamond \mathbf{Y})_t(T) := \mathbb{E}_t(\langle \mathbf{X}^c, \mathbf{Y}^c \rangle_{t, T}) = \sum_{w \in \mathcal{W}_d} \left(\sum_{w_1 w_2 = w} (\mathbf{X}^{w_1} \diamond \mathbf{Y}^{w_2})_t(T) \right) e_w \in \mathcal{T}$$

whenever the \mathcal{T} -valued quadratic covariation that appears on the right-hand side is integrable. Similar to the previous section, we also define an *outer diamond*, for $\mathbf{X}, \mathbf{Y} \in \mathcal{T}$, by

$$(\mathbf{X} \blacklozenge \mathbf{Y})_t(T) := \mathbb{E}_t(\llbracket \mathbf{X}^c, \mathbf{Y}^c \rrbracket_{t, T}) = \sum_{w_1, w_2 \in \mathcal{W}_d} (\mathbf{X}^{w_1} \diamond \mathbf{Y}^{w_2})_t(T) e_{w_1} \otimes e_{w_2} \in T(\mathbb{R}^d) \otimes T(\mathbb{R}^d).$$

This definition extends immediately to semimartingales with values in the quotient algebras of Section 2.3. In particular, given $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ in $\mathcal{S}(\mathcal{S})$, we have

$$(\hat{\mathbf{X}} \diamond \hat{\mathbf{Y}})_t(T) := \mathbb{E}_t(\langle \hat{\mathbf{X}}^c, \hat{\mathbf{Y}}^c \rangle_{t,T}) = \sum_{w_1, w_2 \in \widehat{\mathcal{W}}_d} (\hat{\mathbf{X}}^{w_1} \diamond \hat{\mathbf{Y}}^{w_2})_t(T) \hat{e}_{w_1} \hat{e}_{w_2} \in \mathcal{S},$$

where the last expression is given in terms of diamond products of scalar semimartingales.

Lemma 2.2. *Let $p, q, r \in [1, \infty)$ such that $1/p + 1/q + 1/r < 1$, and let $X \in \mathcal{M}_{\text{loc}}^c((\mathbb{R}^d)^{\otimes l})$, $Y \in \mathcal{M}_{\text{loc}}^c((\mathbb{R}^d)^{\otimes m})$, and $Z \in \mathcal{D}((\mathbb{R}^d)^{\otimes n})$ with $l, m, n \in \mathbb{N}$, such that $\|X\|_{\mathcal{H}^p}, \|Y\|_{\mathcal{H}^q}, \|Z\|_{\mathcal{S}^r} < \infty$. Then it holds that for all $0 \leq t \leq T$*

$$\mathbb{E}_t \left(\int_t^T Z_{u-} d(X \diamond Y)_u(T) \right) = -\mathbb{E}_t \left(\int_t^T Z_{u-} d\langle X, Y \rangle_u \right).$$

Proof. Using the Kunita-Watanabe inequality (Lemma 7.1) we see that the expectation on the right-hand side is well defined. Further note that it follows from Emery’s inequality (Lemma 7.3) and Doob’s maximal inequality that the local martingale

$$\int_0^\cdot Z_{u-} d(\mathbb{E}_u \langle X, Y \rangle_T)$$

is a true martingale. Recall the definition of the diamond product, and observe that the difference of the left- and right-hand sides of the above equation is a conditional expectation of a martingale increment and is hence zero. □

2.6. Generalized signatures

We now give the precise meaning of equation (1.1): that is, $d\mathbf{S} = \mathbf{S} \circ d\mathbf{X}$, or component-wise, for every word $w \in \mathcal{W}_d$,

$$d\mathbf{S}^w = \sum_{w_1 w_2 = w} \mathbf{S}^{w_1} \circ d\mathbf{X}^{w_2},$$

where the driving noise \mathbf{X} is a \mathcal{T}_0 -valued semimartingale, so that $\mathbf{X}^0 \equiv 0$. Following [49, 50, 20, 41, 25, 8], the integral meaning of this equation, started at time s from $\mathbf{s} \in \mathcal{T}_1$, for times $t \geq s$, is given by

$$\mathbf{S}_t = \mathbf{s} + \int_{(s,t]} \mathbf{S}_{u-} d\mathbf{X}_u + \frac{1}{2} \int_s^t \mathbf{S}_{u-} d\langle \mathbf{X}^c \rangle_u + \sum_{s < u \leq t} \mathbf{S}_{u-} (\exp(\Delta \mathbf{X}_u) - 1 - \Delta \mathbf{X}_u), \tag{2.11}$$

leaving the component-wise version to the reader. We have

Proposition 2.3. *Let $\mathbf{s} \in \mathcal{T}_1$ and suppose \mathbf{X} takes values in \mathcal{T}_0 . For every $s \geq 0$ and $\mathbf{s} \in \mathcal{T}_1$, equation (2.11) has a unique global solution on \mathcal{T}_1 starting from $\mathbf{S}_s = \mathbf{s}$.*

Proof. Note that \mathbf{S} solves equation (2.11) iff $\mathbf{s}^{-1}\mathbf{S}$ solves the same equation started from $1 \in \mathcal{T}_1$. We may thus take $\mathbf{s} = 1$ without loss of generality. The graded structure of our problem, and more precisely that $\mathbf{X} = (0, X, \mathbb{X}, \dots)$ in equation (2.11) has no scalar component, shows that the (necessarily) unique solution is given explicitly by iterated integration, as may be seen explicitly when writing out $\mathbf{S}^{(0)} \equiv 1$, $\mathbf{S}_t^{(1)} = \int_s^t dX = X_{s,t} \in \mathbb{R}^d$,

$$\mathbf{S}_t^{(2)} = \int_{(s,t]} \mathbf{S}_{u-}^{(1)} dX_u + \mathbb{X}_t - \mathbb{X}_s + \frac{1}{2} \langle X^c \rangle_{s,t} + \frac{1}{2} \sum_{s < u \leq t} (\Delta X_u)^2 \in (\mathbb{R}^d)^{\otimes 2},$$

and so on. (In particular, we do not need to rely on abstract existence, uniqueness results for Marcus SDEs [41] or Lie group stochastic exponentials [31].) □

Definition 2.4. Let \mathbf{X} be a \mathcal{T}_0 -valued semimartingale defined on some interval $[s, t]$. Then

$$\text{Sig}(\mathbf{X}|_{[s,t]}) \equiv \text{Sig}(\mathbf{X})_{s,t}$$

is defined to be the unique solution to equation (2.11) on $[s, t]$, such that $\text{Sig}(\mathbf{X})_{s,s} = 1$.

The following can be seen as a (generalized) Chen relation.

Lemma 2.5. Let \mathbf{X} be a \mathcal{T}_0 -valued semimartingale on $[0, T]$ and $0 \leq s \leq t \leq u \leq T$. Then the following identity holds with probability one, for all such s, t, u :

$$\text{Sig}(\mathbf{X})_{s,t} \text{Sig}(\mathbf{X})_{t,u} = \text{Sig}(\mathbf{X})_{s,u}. \tag{2.12}$$

Proof. Call $\Phi_{t \leftarrow s} \mathbf{s} := \mathbf{S}_t$ the solution to equation (2.11) at time $t \geq s$, started from $\mathbf{S}_s = \mathbf{s}$. By uniqueness of the solution flow, we have $\Phi_{u \leftarrow t} \circ \Phi_{t \leftarrow s} = \Phi_{u \leftarrow s}$. It now suffices to remark that, thanks to the multiplicative structure of equation (2.11), we have $\Phi_{t \leftarrow s} \mathbf{s} = \text{Sig}(\mathbf{X})_{s,t}$. □

3. Expected signatures and signature cumulants

3.1. Definitions and existence

Throughout this section, let $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0)$ be defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$. Recall that \mathbb{E}_t denotes the conditional expectation with respect to the sigma algebra \mathcal{F}_t . When $\mathbb{E}(|\text{Sig}(\mathbf{X})_{0,t}^w|) < \infty$ for all $0 \leq t \leq T$ and all words $w \in \mathcal{W}_d$, then the (conditional) expected signature

$$\mu_t(T) := \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}) = \sum_{w \in \mathcal{W}_d} \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}^w) e_w \in \mathcal{T}_1, \quad 0 \leq t \leq T,$$

is well defined. In this case, we can also define the (conditional) signature cumulant of \mathbf{X} by

$$\kappa_t(T) := \log(\mu_t(T)) \in \mathcal{T}_0, \quad 0 \leq t \leq T.$$

An important observation is the following:

Lemma 3.1. Given $\mathbb{E}(|\text{Sig}(\mathbf{X})_{0,t}^w|) < \infty$ for all $0 \leq t \leq T$ and words $w \in \mathcal{W}_d$, then $\mu(T) \in \mathcal{S}(\mathcal{T}_1)$ and $\kappa(T) \in \mathcal{S}(\mathcal{T}_0)$.

Proof. It follows from the relation in equation (2.12) that

$$\mu_t(T) = \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}) = \mathbb{E}_t(\text{Sig}(\mathbf{X})_{0,t}^{-1} \text{Sig}(\mathbf{X})_{0,T}) = \text{Sig}(\mathbf{X})_{0,t}^{-1} \mathbb{E}_t(\text{Sig}(\mathbf{X})_{0,T}).$$

Therefore, projecting to the tensor components, we have

$$\mu_t(T)^w = \sum_{w_1 w_2 = w} (-1)^{|w_1|} \text{Sig}(\mathbf{X})_{0,t}^{w_1} \mathbb{E}_t(\text{Sig}(\mathbf{X})_{0,T}^{w_2}), \quad 0 \leq t \leq T, \quad w \in \mathcal{W}_d.$$

Since $(\text{Sig}(\mathbf{X})_{0,t}^w)_{0 \leq t \leq T}$ and $(\mathbb{E}_t(\text{Sig}(\mathbf{X})_{0,T}^w))_{0 \leq t \leq T}$ are semimartingales (the latter in fact a martingale), it follows from Itô's product rule that $\mu^w(T)$ is also a semimartingale for all words $w \in \mathcal{W}_d$, hence $\mu(T) \in \mathcal{S}(\mathcal{T}_1)$. Further recall that $\kappa(T) = \log(\mu(T))$, and therefore it follows from the definition of the logarithm on \mathcal{T}_1 that each component $\kappa(T)^w$ with $w \in \mathcal{W}_d$ is a polynomial of $(\mu(T)^v)_{v \in \mathcal{W}_d, |v| \leq |w|}$. Hence it follows again by Itô's product rule that $\kappa(T) \in \mathcal{S}(\mathcal{T}_0)$. □

It is of strong interest to have a more explicit necessary condition for the existence of the expected signature. The following theorem, the proof of which can be found in Section 7.1, yields such a criterion.

Theorem 3.2. Let $q \in [1, \infty)$ and $N \in \mathbb{N}_{\geq 1}$; then there exist two constants $c, C > 0$ depending only on d, N and q , such that for all $\mathbf{X} \in \mathcal{H}^{q,N}$

$$c \|\mathbf{X}\|_{\mathcal{H}^{q,N}} \leq \|\text{Sig}(\mathbf{X})_{0,\cdot}\|_{\mathcal{H}^{q,N}} \leq C \|\mathbf{X}\|_{\mathcal{H}^{q,N}}.$$

In particular, if $\mathbf{X} \in \mathcal{H}^{\infty-}(\mathcal{T}_0)$, then $\text{Sig}(\mathbf{X})_{0,\cdot} \in \mathcal{H}^{\infty-}(\mathcal{T}_1)$, and the expected signature exists.

Remark 3.3. Let $\mathbf{X} = (0, M, 0, \dots, 0)$, where $M \in \mathcal{M}(\mathbb{R}^d)$ is a martingale; then

$$\|\mathbf{X}\|_{\mathcal{H}^{q,N}} = \|M\|_{\mathcal{H}^{qN}} = \left\| | [M]_T |^{1/2} \right\|_{\mathcal{L}^{qN}},$$

and we see that the above estimate implies that

$$\max_{n=1,\dots,N} \left\| \text{Sig}(\mathbf{X})_{0,\cdot}^{(n)} \right\|_{\mathcal{S}^{qN/n}}^{1/n} \leq C \|M\|_{\mathcal{H}^{qN}}.$$

This estimate is already known and follows from the Burkholder-Davis-Gundy inequality for enhanced martingales, which was first proved in the continuous case in [26] and for the general case in [12].

Remark 3.4. When $q > 1$, the above estimate also holds true when the signature $\text{Sig}(\mathbf{X})_{0,\cdot}$ is replaced by the conditional expected signature $\mu(T)$ or the conditional signature cumulant $\kappa(T)$. This will be seen in the proof of Theorem 4.1 below (more precisely in Claim 7.12).

3.2. Moments and cumulants

We quickly discuss the development of a symmetric algebra valued semimartingale, more precisely, $\hat{\mathbf{X}} \in \mathcal{S}(\mathcal{S}_0)$, in the group \mathcal{S}_1 . That is, we consider

$$d\hat{\mathbf{S}} = \hat{\mathbf{S}} \circ d\hat{\mathbf{X}}. \tag{3.1}$$

It is immediate (validity of chain rule) that the unique solution to this equation, at time $t \geq s$, started at $\hat{\mathbf{S}}_s = \hat{\mathbf{s}} \in \mathcal{S}_1$ is given by

$$\hat{\mathbf{S}}_t := \exp(\hat{\mathbf{X}}_t - \hat{\mathbf{X}}_s) \hat{\mathbf{s}} \in \mathcal{S}_1,$$

and we also write $\hat{\mathbf{S}}_{s,t} = \exp(\hat{\mathbf{X}}_t - \hat{\mathbf{X}}_s)$ for this solution started at time s from $1 \in \mathcal{S}_1$. The relation to signatures is as follows. Recall that the π_{Sym} denotes the canonical projection from \mathcal{T} to \mathcal{S} .

Proposition 3.5. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{S}(\mathcal{T})$, and define $\hat{\mathbf{X}} := \pi_{\text{Sym}}(\mathbf{X})$ and $\hat{\mathbf{Y}} := \pi_{\text{Sym}}(\mathbf{Y})$. Then it holds that

(i) $\hat{\mathbf{X}}, \hat{\mathbf{Y}} \in \mathcal{S}(\mathcal{S})$, and for the indefinite Itô integral, we have in the sense of indistinguishable processes

$$\pi_{\text{Sym}} \int \mathbf{X} d\mathbf{Y} = \int \hat{\mathbf{X}} d\hat{\mathbf{Y}}, \tag{3.2}$$

(ii) $\hat{\mathbf{S}} := \pi_{\text{Sym}} \text{Sig}(\mathbf{X})_{s,\cdot}$ solves equation (3.1) started at time s from $1 \in \mathcal{S}_1$ and driven by $\hat{\mathbf{X}}$. In particular,

$$\hat{\mathbf{S}}_{s,t} = \exp(\hat{\mathbf{X}}_t - \hat{\mathbf{X}}_s)$$

for all $0 \leq s \leq t \leq T$.

Proof. (i) That the projections $\hat{\mathbf{X}}, \hat{\mathbf{Y}}$ define \mathcal{S} -valued semimartingales follows from the componentwise definition and the fact that the canonical projection is linear. In particular, the right-hand side of equation (3.2) is well defined. To show equation (3.2), we apply the canonical projection π_{Sym} to both sides of

equation (2.9) after choosing $Z_t \equiv \mathbf{1}$; and using the explicit action of π_{Sym} on basis tensors, we obtain the identity

$$\pi_{\text{Sym}} \int \mathbf{X} d\mathbf{Y} = \sum_{w \in \mathcal{W}^d} \left(\sum_{uv=w} \int \mathbf{X}^u d\mathbf{Y}^v \right) \hat{e}_w = \int \hat{\mathbf{X}} d\hat{\mathbf{Y}}$$

by equation (2.4). Part (ii) is then immediate. □

Assuming componentwise integrability, we then define *symmetric moments* and *cumulants* of the \mathcal{S} -valued semimartingale $\hat{\mathbf{X}}$ by

$$\begin{aligned} \hat{\mu}_t(T) &:= \mathbb{E}_t \left(\exp(\hat{\mathbf{X}}_T - \hat{\mathbf{X}}_t) \right) = \sum_{v \in \mathcal{V}_d} \mathbb{E}_t \left(\exp(\hat{\mathbf{X}}_T - \hat{\mathbf{X}}_t)^v \right) \hat{e}_v \in \mathcal{S}_1, \\ \hat{\kappa}_t(T) &:= \log(\hat{\mu}_t(T)) \in \mathcal{S}_0, \end{aligned}$$

for $0 \leq t \leq T$. If $\hat{\mathbf{X}} = \pi_{\text{Sym}}(\mathbf{X})$, for $\mathbf{X} \in \mathcal{S}(T)$, with expected signature and signature cumulants $\mu(T)$ and $\kappa(T)$, it is then clear that the symmetric moments and cumulants of $\hat{\mathbf{X}}$ are obtained by projection,

$$\hat{\mu}(T) = \pi_{\text{Sym}}(\mu(T)), \quad \hat{\kappa}(T) = \pi_{\text{Sym}}(\kappa(T)).$$

Example 3.6. Let X be an \mathbb{R}^d -valued martingale in $\mathcal{H}^{\infty-}$, and $\hat{\mathbf{X}}_t := \sum_{i=1}^d X_t^i \hat{e}_i$. Then

$$\hat{\mu}_t(T) = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbb{E}_t (X_T - X_t)^n = 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{i_1, \dots, i_n=1}^d \mathbb{E}_t \left[(X_T^{i_1} - X_t^{i_1}) \cdots (X_T^{i_n} - X_t^{i_n}) \right] \hat{e}_{\overline{i_1 \dots i_n}}$$

consists of the (time- t conditional) multivariate moments of $X_T - X_t \in \mathbb{R}^d$. Here, the series on the right-hand side is understood in the formal sense. It readily follows, also noted in [7, Example 3.3], that $\hat{\kappa}_t(T) = \log(\hat{\mu}_t(T))$ consists precisely of the multivariate cumulants of $X_T - X_t$. Note that the symmetric moments and cumulants of the scaled process aX , $a \in \mathbb{R}$, are precisely given by $\delta_a \hat{\mu}$ and $\delta_a \hat{\kappa}$, where the linear dilation map is defined by $\delta_a : \hat{e}_w \mapsto a^{|w|} \hat{e}_w$. The situation is similar for $a \cdot X = (a_1 X^1, \dots, a_d X^d)$, $a \in \mathbb{R}^d$, but now with $\delta_a : \hat{e}_w \mapsto a^w \hat{e}_w^{|w|}$ with $a^w = a_1^{n_1} \cdots a_d^{n_d}$, where n_i denotes the multiplicity of the letter $i \in \{1, \dots, d\}$ in the word w .

We next consider linear combinations, $\hat{\mathbf{X}} = aX + b\langle X \rangle$, for general pairs $a, b \in \mathbb{R}$, having already dealt with $b = 0$. The special case $b = -a^2/2$, by scaling there is no loss in generality to take $(a, b) = (1, -1/2)$, yields a (at least formally) familiar exponential martingale identity.

Example 3.7. Let X be an \mathbb{R}^d -valued martingale in $\mathcal{H}^{\infty-}$, and define

$$\hat{\mathbf{X}}_t := \sum_{i=1}^d X_t^i \hat{e}_i - \frac{1}{2} \sum_{1 \leq i \leq j \leq d} \langle X^i, X^j \rangle_t \hat{e}_{ij}.$$

In this case, we have trivial symmetric cumulants $\hat{\kappa}_t(T) = 0$ for all $0 \leq t \leq T$. Indeed, Itô’s formula shows that $t \mapsto \exp(\hat{\mathbf{X}}_t)$ is an \mathcal{S}_1 -valued martingale, so that

$$\hat{\mu}_t(T) = \mathbb{E}_t \exp(\hat{\mathbf{X}}_T - \hat{\mathbf{X}}_t) = \exp(-\hat{\mathbf{X}}_t) \mathbb{E}_t \exp(\hat{\mathbf{X}}_T) = 1.$$

While the symmetric cumulants of the last example carries no information, it suffices to work with

$$\hat{\mathbf{X}} = \sum_{i=1}^d a^i X^i \hat{e}_i + \sum_{1 \leq i \leq j \leq d} b_{ijk} \langle X^j, X^k \rangle \hat{e}_{ij},$$

in which case $\hat{\mu} = \hat{\mu}(a, b), \hat{\kappa} = \hat{\kappa}(a, b)$ contains full information of the joint moments of X and its quadratic variation process. A recursion of these was constructed as a diamond expansion in [23].

4. Main results

4.1. Functional equation for signature cumulants

Let $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0)$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T < \infty}, \mathbb{P})$ satisfying the usual conditions. For all $\mathbf{x} \in \mathcal{T}_0$ (or \mathcal{T}_0^N), define the following operators, with Bernoulli numbers $(B_k)_{k \geq 0} = (1, -\frac{1}{2}, \frac{1}{6}, \dots)$,

$$\begin{aligned}
 G(\text{ad } \mathbf{x}) &= \sum_{k=0}^{\infty} \frac{(\text{ad } \mathbf{x})^k}{(k+1)!}, & Q(\text{ad } \mathbf{x}) &= \sum_{m,n=0}^{\infty} 2 \frac{(\text{ad } \mathbf{x})^n \odot (\text{ad } \mathbf{x})^m}{(n+1)!(m)!(n+m+2)}, \\
 H(\text{ad } \mathbf{x}) &:= \sum_{k=0}^{\infty} \frac{B_k}{k!} (\text{ad } \mathbf{x})^k,
 \end{aligned}
 \tag{4.1}$$

noting $G(z) = (\exp(z) - 1)/z, H(z) = G^{-1}(z) = z/(\exp(z) - 1)$. Our main result is the following.

Theorem 4.1. *Let $\mathbf{X} \in \mathcal{H}^{\infty-}(\mathcal{T}_0)$; then the signature cumulant $\kappa = \kappa(T) = (\log \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}))_{0 \leq t \leq T}$ is the unique solution (up to indistinguishably) of the following functional equation: for all $0 \leq t \leq T$*

$$\begin{aligned}
 0 &= \mathbb{E}_t \left\{ \mathbf{X}_{t,T} + \frac{1}{2} \langle \mathbf{X}^c \rangle_{t,T} + \int_{(t,T]} G(\text{ad } \kappa_{u-})(d\kappa_u) + \frac{1}{2} \int_t^T Q(\text{ad } \kappa_{u-})(d\llbracket \kappa^c, \kappa^c \rrbracket_u) \right. \\
 &\quad + \int_t^T (\text{Id} \odot G(\text{ad } \kappa_{u-}))(d\llbracket \mathbf{X}^c, \kappa^c \rrbracket_u) \\
 &\quad \left. + \sum_{t < u \leq T} \left(\exp(\Delta \mathbf{X}_u) \exp(\kappa_u) \exp(-\kappa_{u-}) - 1 - \Delta \mathbf{X}_u - G(\text{ad } \kappa_{u-})(\Delta \kappa_u) \right) \right\}.
 \end{aligned}
 \tag{4.2}$$

Equivalently, $\kappa = \kappa(T)$ is the unique solution to

$$\begin{aligned}
 \kappa_t &= \mathbb{E}_t \left\{ \int_{(t,T]} H(\text{ad } \kappa_{u-})(d\mathbf{X}_u) + \frac{1}{2} \int_t^T H(\text{ad } \kappa_{u-})(d\langle \mathbf{X}^c \rangle_u) \right. \\
 &\quad + \frac{1}{2} \int_t^T H(\text{ad } \kappa_{u-}) \circ Q(\text{ad } \kappa_{u-})(d\llbracket \kappa^c, \kappa^c \rrbracket_u) \\
 &\quad + \int_t^T H(\text{ad } \kappa_{u-}) \circ (\text{Id} \odot G(\text{ad } \kappa_{u-}))(d\llbracket \mathbf{X}^c, \kappa^c \rrbracket_u) \\
 &\quad \left. + \sum_{t < u \leq T} \left(H(\text{ad } \kappa_{u-}) \left(\exp(\Delta \mathbf{X}_u) \exp(\kappa_u) \exp(-\kappa_{u-}) - 1 - \Delta \mathbf{X}_u \right) - \Delta \kappa_u \right) \right\}.
 \end{aligned}
 \tag{4.3}$$

Furthermore, if $\mathbf{X} \in \mathcal{H}^{1,N}$ for some $N \in \mathbb{N}_{\geq 1}$, then the identities in equations (4.2) and (4.3) still hold true for the truncated signature cumulant $\kappa := (\log \mathbb{E}_t(\text{Sig}(\mathbf{X}^{(0,N)})_{t,T}))_{0 \leq t \leq T}$.

Proof. We postpone the proof of the fact that κ satisfies equations (4.2) and (4.3) to section Section 7.2. The uniqueness part of the statement can be easily seen as follows. Regarding equation (4.2), we first note that it holds

$$\mathbb{E}_t \left\{ \int_{(t,T]} G(\text{ad } \kappa_{u-})(d\kappa_u) \right\} = \mathbb{E}_t \left\{ \int_{(t,T]} (G(\text{ad } \kappa_{u-}) - \text{Id})(d\kappa_u) \right\} - \kappa_t, \quad 0 \leq t \leq T,$$

where we have used that $\kappa_T \equiv 0$ (and the fact that the conditional expectation is well defined, which is shown in the first part of the proof). Hence, after subtracting the identity from G , we can bring κ_t to the left-hand side in equation (4.2). This identity is an equality of tensor series in \mathcal{T}_0 and can be projected to yield an equality for each tensor level of the series. As presented in more detail in the following subsection, we see that projecting the latter equation to tensor level, say $n \in \mathbb{N}_{\geq 1}$, the right-hand side only depends on $\kappa^{(k)}$ for $k < n$, hence giving an explicit representation $\kappa^{(n)}$ in terms of \mathbf{X} and strictly lower tensor levels of κ . Therefore equation (4.2) characterizes κ up to a modification and then due to right-continuity up to indistinguishably. The same argument applies to equation (4.3), referring to the following subsections for details on the recursion. \square

Diamond formulation: The functional equations given in Theorem 4.1 above can be phrased in terms of the diamond product between \mathcal{T}_0 -valued semimartingales. Writing $\mathbf{J}_t(T) = \sum_{t < u \leq T} (\dots)$ for the last (jump) sum in equation (4.2), this equation can be written, thanks to Lemma 2.2, which applies just the same with outer diamonds,

$$\begin{aligned} & \frac{1}{2}(\mathbf{X} \diamond \mathbf{X})_t(T) + \mathbb{E}_t \left\{ \mathbf{X}_{t,T} + \int_{(t,T]} G(\text{ad } \kappa_{u-})(d\kappa_u) + \mathbf{J}_t(T) \right\} \\ &= \mathbb{E}_t \left\{ \frac{1}{2} \int_t^T Q(\text{ad } \kappa_{u-}) d(\kappa \diamond \kappa)_u(T) + \int_t^T (\text{Id} \odot G(\text{ad } \kappa_{u-})) d(\mathbf{X} \diamond \kappa)_u(T) \right\} \end{aligned}$$

and a similar form may be given for equation (4.3). While one may or may not prefer this equation to equation (4.2), diamonds become very natural in $d = 1$ (or upon projection to the symmetric algebra; see also Section 5.2). In this case, $G = \text{Id}$, $Q = \text{Id} \odot \text{Id}$; and with identities of the form

$$\int_t^T (\text{Id} \odot \text{Id}) d(\mathbf{X} \diamond \mathbf{Y})_u(T) = (\mathbf{X} \diamond \mathbf{Y})_u(T) \Big|_{u=t}^T = -(\mathbf{X} \diamond \mathbf{Y})_t(T),$$

some simple rearrangement, using bilinearity of the diamond product, gives

$$\kappa_t(T) = \mathbb{E}_t \{ \mathbf{X}_{t,T} \} + \frac{1}{2} ((\mathbf{X} + \kappa) \diamond (\mathbf{X} + \kappa))_t(T) + \mathbb{E}_t \{ \mathbf{J}_t(T) \}. \tag{4.4}$$

If we further impose martingality and continuity, we arrive at

$$\kappa_t(T) = \frac{1}{2} ((\mathbf{X} + \kappa) \diamond (\mathbf{X} + \kappa))_t(T).$$

4.2. Recursive formulas for signature cumulants

Theorem 4.1 allows for an iterative computation of signature cumulants, trivially started from

$$\kappa_t^{(1)} = \mu_t^{(1)} = \mathbb{E}_t \left(\mathbf{X}_{t,T}^{(1)} \right).$$

The second signature cumulant, obtained from Theorem 4.1 or first principles, reads

$$\begin{aligned} \kappa_t^{(2)} = & \mathbb{E}_t \left\{ \mathbf{X}_{t,T}^{(2)} + \frac{1}{2} \left\langle \mathbf{X}^{(1)c} \right\rangle_{t,T} + \frac{1}{2} \int_{(t,T]} \left[\kappa_{u-}^{(1)}, d\kappa_u^{(1)} \right] + \frac{1}{2} \left\langle \kappa^{(1)c} \right\rangle_{t,T} + \left\langle \mathbf{X}^{(1)c}, \kappa^{(1)c} \right\rangle_{t,T} \right. \\ & \left. + \sum_{t < u \leq T} \left(\frac{1}{2} (\Delta \mathbf{X}_u^{(1)})^2 + \Delta \mathbf{X}_u^{(1)} \Delta \kappa_u^{(1)} + \frac{1}{2} (\Delta \kappa_u^{(1)})^2 \right) \right\}. \end{aligned}$$

For instance, consider the special case with vanishing higher-order components, $\mathbf{X}^{(i)} \equiv 0$, for $i \neq 1$, and $\mathbf{X} = \mathbf{X}^{(1)} \equiv M$, a d -dimensional continuous square-integrable martingale. In this case, $\kappa^{(1)} = \mu^{(1)} \equiv 0$,

and from the very definition of the logarithm relating κ and μ , we have $\kappa^{(2)} = \mu^{(2)} - \frac{1}{2}\mu^{(1)}\mu^{(1)} = \mu^{(2)}$. It then follows from Stratonovich-Ito correction that

$$\kappa_t^{(2)} = \mathbb{E}_t \int_t^T (M_u - M_t) \circ dM_u = \frac{1}{2} \mathbb{E}_t \langle M \rangle_{t,T} = \frac{1}{2} \mathbb{E}_t \left\langle \mathbf{X}^{(1)} \right\rangle_{t,T},$$

which is indeed a (very) special case of the general expression for $\kappa^{(2)}$. We now treat general higher-order signature cumulants.

Corollary 4.2. *Let $\mathbf{X} \in \mathcal{H}^{1,N}$ for some $N \in \mathbb{N}_{\geq 1}$; then we have*

$$\kappa_t^{(1)} = \mathbb{E}_t \left(\mathbf{X}_{t,T}^{(1)} \right),$$

for all $0 \leq t \leq T$ and for $n \in \{2, \dots, N\}$, we have recursively (the right-hand side only depends on $\kappa^{(j)}, j < n$)

$$\begin{aligned} \kappa_t^{(n)} &= \mathbb{E}_t \left(\mathbf{X}_{t,T}^{(n)} \right) + \frac{1}{2} \sum_{k=1}^{n-1} \mathbb{E}_t \left(\left\langle \mathbf{X}^{(k)c}, \mathbf{X}^{(n-k)c} \right\rangle_{t,T} \right) \\ &\quad + \sum_{|\ell| \geq 2, \|\ell\|=n} \mathbb{E}_t \left(\text{Mag}(\kappa; \ell)_{t,T} + \text{Qua}(\kappa; \ell)_{t,T} + \text{Cov}(\mathbf{X}, \kappa; \ell)_{t,T} + \text{Jmp}(\mathbf{X}, \kappa; \ell)_{t,T} \right) \end{aligned} \quad (4.5)$$

with $\ell = (l_1, \dots, l_k), l_i \in \mathbb{N}_{\geq 1}, |\ell| := k \in \mathbb{N}_{\geq 1}, \|\ell\| := l_1 + \dots + l_k$ and

$$\begin{aligned} \text{Mag}(\kappa; l_1, \dots, l_k)_{t,T} &= \frac{1}{k!} \int_{(t,T]} \text{ad } \kappa_{u-}^{(l_2)} \cdots \text{ad } \kappa_{u-}^{(l_k)} (d\kappa_u^{(l_1)}) \\ \text{Qua}(\kappa; l_1, \dots, l_k)_{t,T} &= \frac{1}{k!} \sum_{m=2}^k \binom{k-1}{m-1} \\ &\quad \times \int_t^T \left(\text{ad } \kappa_{u-}^{(l_3)} \cdots \text{ad } \kappa_{u-}^{(l_m)} \odot \text{ad } \kappa_{u-}^{(l_{m+1})} \cdots \text{ad } \kappa_{u-}^{(l_k)} \right) \left(d \left\| \kappa^{(l_1)c}, \kappa^{(l_2)c} \right\|_u \right) \\ \text{Cov}(\mathbf{X}, \kappa; l_1, \dots, l_k)_{t,T} &= \frac{1}{(k-1)!} \int_t^T \left(\text{Id} \odot \text{ad } \kappa_{u-}^{(l_3)} \cdots \text{ad } \kappa_{u-}^{(l_k)} \right) \left(d \left\| \mathbf{X}^{(l_1)c}, \kappa^{(l_2)c} \right\|_u \right) \\ \text{Jmp}(\mathbf{X}, \kappa; l_1, \dots, l_k)_{t,T} &= \sum_{t < u \leq T} \sum_{0 \leq m \leq j \leq k} \left((-1)^{k-j} \frac{\Delta \mathbf{X}_u^{(l_1)} \cdots \Delta \mathbf{X}_u^{(l_m)} \kappa_u^{(l_{m+1})} \cdots \kappa_u^{(l_j)} \kappa_{u-}^{(l_{j+1})} \cdots \kappa_{u-}^{(l_k)}}{m!(m-j)!(k-j)!} \right) \\ &\quad - \frac{1}{k!} \text{ad } \kappa_{u-}^{(l_2)} \cdots \text{ad } \kappa_{u-}^{(l_k)} \left(\Delta \kappa_u^{(l_1)} \right). \end{aligned}$$

Proof. As in the proof of Theorem 4.1 above, in equation (4.2), we can separate the identity from G and bring the resulting κ_t to the left-hand side. Projecting to the tensor level n , and using that the projection can be interchanged with taking the expectation, we obtain the following equation

$$\kappa_t^{(n)} = \mathbb{E}_t \left\{ \mathbf{X}_{t,T}^{(n)} + \frac{1}{2} \langle \mathbf{X}^c \rangle_{t,T}^{(n)} + \mathbf{Y}_{t,T}^{(n)} + \mathbf{V}_{t,T}^{(n)} + \mathbf{C}_{t,T}^{(n)} + \mathbf{J}_{t,T}^{(n)} \right\},$$

for all $0 \leq t \leq T$, where $\mathbf{Y} \in \mathcal{S}(\mathcal{T}_0^N)$ and $\mathbf{V}, \mathbf{C}, \mathbf{J} \in \mathcal{V}(\mathcal{T}_0^N)$ are defined in equation (7.9). Note that we can take the conditional expectation of each term separately, as $\mathbf{X}^{(n)}$ has sufficient integrability by the assumption, and the integrability of the remaining terms is shown in the Proof of Theorem 4.1 below, more precisely in equation (7.27). The recursion then follows by spelling out the explicit composition for each of the terms appearing in the above equation. For the quadratic variation term, we may easily

verify from the bilinearity of the covariation bracket that it holds

$$\langle \mathbf{X}^c \rangle_{t,T}^{(n)} = \pi_n \sum_{k,j=1}^n \langle \mathbf{X}^{(k)c}, \mathbf{X}^{(j)c} \rangle_{t,T} = \sum_{k=1}^n \langle \mathbf{X}^{(k)c}, \mathbf{X}^{(n-k)c} \rangle_{t,T}.$$

We can also take the conditional expectation of each term in the above right-hand side separately, as the integrability of these terms follows from the assumptions on \mathbf{X} and Lemma 7.1. The composition of the terms $\mathbf{Y}^{(n)}$, $\mathbf{V}^{(n)}$ and $\mathbf{C}^{(n)}$ follows from the explicit form of the stochastic Itô integral of a power series of adjoint operations with respect to a tensor valued semimartingale in Section 2.4, more specifically in equation (2.10). We will demonstrate this for the term $\mathbf{Y}^{(n)}$ in more detail:

$$\begin{aligned} \mathbf{Y}_{t,T}^{(n)} &= \pi_n \int_{(t,T]} (G(\text{ad } \kappa_{u-}) - \text{Id})(d\kappa_u) \\ &= \pi_n \int_{(t,T]} \left[\sum_{k=1}^{\infty} \frac{1}{(k+1)!} (\text{ad } \kappa_{u-})^k \right] (d\kappa_u) \\ &= \sum_{k=1}^{n-1} \frac{1}{(k+1)!} \sum_{\|\ell\|=n, |\ell|=k+1} \int_{(t,T]} \text{ad } \kappa_{u-}^{(\ell_2)} \cdots \text{ad } \kappa_{u-}^{(\ell_{k+1})} (d\kappa_u^{(\ell_1)}) = \sum_{\|\ell\|=n, |\ell| \geq 2} \text{Mag}(\kappa; \ell)_{t,T} \end{aligned}$$

for all $0 \leq t \leq T$. The composition of the terms $\mathbf{V}^{(n)}$ and $\mathbf{C}^{(n)}$, respectively, in terms of $\text{Qua}(\mathbf{X}, \kappa)$ and $\text{Cov}(\mathbf{X}, \kappa)$ follows analogously. It then remains to show that the term $\mathbf{J}^{(n)}$ can be composed in terms of $\text{Jmp}(\mathbf{X}, \kappa)$, which is, however, a simple combinatorial exercise. \square

We obtain another recursion for the signature cumulants from projecting the functional equation (4.3). Note that, apart from the first two levels, it is far from trivial to see that the following recursion is equivalent to the recursion in Corollary 4.2.

Corollary 4.3. *Let $\mathbf{X} \in \mathcal{H}^{1,N}$ for some $N \in \mathbb{N}_{\geq 1}$; then we have*

$$\begin{aligned} \kappa_t^{(n)} &= \mathbb{E}_t \left(\mathbf{X}_{t,T}^{(n)} \right) + \sum_{|\ell| \geq 2, \|\ell\|=n} \mathbb{E}_t \left(\text{HMag}^1(\mathbf{X}, \kappa; \ell)_{t,T} + \frac{1}{2} \text{HMag}^2(\mathbf{X}, \kappa; \ell)_{t,T} + \text{HQua}(\kappa; \ell)_{t,T} \right. \\ &\quad \left. + \text{HCov}(\mathbf{X}, \kappa; \ell)_{t,T} + \text{HJmp}(\mathbf{X}, \kappa; \ell)_{t,T} \right) \end{aligned} \tag{4.6}$$

with $\ell = (l_1, \dots, l_k)$, $l_i \geq 1$, $|\ell| = k$, $\|\ell\| = l_1 + \dots + l_k$ and

$$\begin{aligned} \text{HMag}^1(\mathbf{X}, \kappa; l_1, \dots, l_k)_{t,T} &= \frac{B_{k-1}}{(k-1)!} \int_{(t,T]} \text{ad } \kappa_{u-}^{(l_2)} \cdots \text{ad } \kappa_{u-}^{(l_k)} (d\mathbf{X}_u^{(l_1)}) \\ \text{HMag}^2(\mathbf{X}, \kappa; l_1, \dots, l_k)_{t,T} &= \frac{B_{k-2}}{(k-2)!} \int_t^T \text{ad } \kappa_{u-}^{(l_3)} \cdots \text{ad } \kappa_{u-}^{(l_k)} \left(d \left\langle \mathbf{X}^{(l_1)c}, \mathbf{X}^{(l_2)c} \right\rangle_u \right) \\ \text{HQua}(\kappa; l_1, \dots, l_k)_{t,T} &= \int_t^T \sum_{j=2}^k \frac{B_{k-j}}{(k-j)!} \text{ad } \kappa_{u-}^{(l_{j+1})} \cdots \text{ad } \kappa_{u-}^{(l_k)} (d\text{Qua}(\kappa; l_1, \dots, l_j)_u) \\ \text{HCov}(\mathbf{X}, \kappa; l_1, \dots, l_{k+1})_{t,T} &= \int_t^T \sum_{j=1}^k \frac{B_{k-j}}{(k-j)!} \text{ad } \kappa_{u-}^{(l_{j+1})} \cdots \text{ad } \kappa_{u-}^{(l_k)} (d\text{Cov}(\mathbf{X}, \kappa; l_1, \dots, l_j)_u) \end{aligned}$$

$$\begin{aligned} & \text{HJmp}(\mathbf{X}, \boldsymbol{\kappa}; l_1, \dots, l_k)_{t,T} \\ &= \sum_{t < u \leq T} \sum_{i=1}^k \frac{B_{k-i}}{(k-i)!} \text{ad } \boldsymbol{\kappa}_{u-}^{(l_{i+1})} \cdots \text{ad } \boldsymbol{\kappa}_{u-}^{(l_k)} \\ & \quad \times \left(\mathbf{1}_{\{i=1\}} \Delta \boldsymbol{\kappa}_{u-}^{(l_1)} + \mathbf{1}_{\{i>1\}} \sum_{0 \leq m \leq j \leq i} (-1)^{i-j} \frac{\Delta \mathbf{X}_u^{(l_1)} \cdots \Delta \mathbf{X}_u^{(l_m)} \boldsymbol{\kappa}_u^{(l_{m+1})} \cdots \boldsymbol{\kappa}_u^{(l_j)} \boldsymbol{\kappa}_{u-}^{(l_{j+1})} \cdots \boldsymbol{\kappa}_{u-}^{(l_i)}}{m!(m-j)!(i-j)!} \right). \end{aligned}$$

Proof. The recursion follows from projecting equation (4.3) to each tensor level, analogously to the way that the recursion of Corollary 4.2 follows from equation (4.2) (see the proof of Corollary 4.2). □

Diamonds. All recursions here can be rewritten in terms of diamonds. In a first step, by definition, the second term in Corollary 4.2 can be rewritten as

$$\frac{1}{2} \sum_{k=1}^n (\mathbf{X}^{(k)} \diamond \mathbf{X}^{(n-k)})_t(T).$$

Thanks to Lemma 2.2, we may also write

$$\begin{aligned} & \mathbb{E}_t \text{Qua}(\boldsymbol{\kappa}; \ell)_{t,T} \\ &= -\mathbb{E}_t \left\{ \frac{1}{k!} \sum_{m=2}^k \binom{k-1}{m-1} \int_t^T \left(\text{ad } \boldsymbol{\kappa}_{u-}^{(\ell_3)} \cdots \text{ad } \boldsymbol{\kappa}_{u-}^{(\ell_m)} \odot \text{ad } \boldsymbol{\kappa}_{u-}^{(\ell_{m+1})} \cdots \text{ad } \boldsymbol{\kappa}_{u-}^{(\ell_k)} \right) \left(d(\boldsymbol{\kappa}^{(\ell_1)} \blacklozenge \boldsymbol{\kappa}^{(\ell_2)})_u(T) \right) \right\}. \end{aligned}$$

Similarly,

$$\mathbb{E}_t \text{Cov}(\mathbf{X}, \boldsymbol{\kappa}; \ell)_{t,T} = -\mathbb{E}_t \left\{ \frac{1}{(k-1)!} \int_t^T \left(\text{Id} \odot \text{ad } \boldsymbol{\kappa}_{u-}^{(\ell_3)} \cdots \text{ad } \boldsymbol{\kappa}_{u-}^{(\ell_k)} \right) \left(d(\mathbf{X}^{(\ell_1)} \blacklozenge \boldsymbol{\kappa}^{(\ell_2)})_u(T) \right) \right\}.$$

Inserting these expressions into Equation (4.6), we may obtain a ‘diamond’ form of the recursions in H form.

When $d = 1$ (or in the projection onto the symmetric algebra; see also Section 5.2), the recursions take a particularly simple form, since $\text{ad } \mathbf{x} \equiv 0$ for all $\mathbf{x} \in \mathcal{T}_0$, for $d = 1$ a commutative algebra. Equation (4.5) then becomes

$$\boldsymbol{\kappa}_t^{(n)}(T) = \mathbb{E}_t \left(\mathbf{X}_{t,T}^{(n)} \right) + \frac{1}{2} \sum_{k=1}^{n-1} \left((\mathbf{X}^{(k)} + \boldsymbol{\kappa}^{(k)}) \diamond (\mathbf{X}^{(n-k)} + \boldsymbol{\kappa}^{(n-k)}) \right)_t(T) + \mathbb{E}_t \left(\mathbf{J}_t^{(n)}(T) \right),$$

where $\mathbf{J}_t^{(n)}(T) = \sum_{|\ell| \geq 2, \|\ell\| = n} \text{Jmp}(\mathbf{X}, \boldsymbol{\kappa}; \ell)_{t,T}$ contains the n th tensor component of the jump contribution. The above diamond recursion can also be obtained by projecting the functional relation (4.4) to the n th tensor level. We shall revisit this in a multivariate setting and comment on related works in Section 5.2.

5. Two special cases

5.1. Variations on Hausdorff, Magnus and Baker–Campbell–Hausdorff

We now consider a deterministic driver \mathbf{X} of finite variation. This includes the case when \mathbf{X} is absolutely continuous, in which case we recover, up to a harmless time reversal, $t \leftrightarrow T - t$, Hausdorff’s ODE and the classical Magnus expansion for the solution to a linear ODE in a Lie group [32, 48, 11, 33]. Our extension with regard to discontinuities seems to be new and somewhat unifies Hausdorff’s equation with multivariate Baker–Campbell–Hausdorff integral formulas.

Theorem 5.1. Let $\mathbf{X} \in \mathcal{V}(\mathcal{T}_0)$, and more specifically $\mathbf{X}: [0, T] \rightarrow \mathcal{T}_0$ deterministic, càdlàg of bounded variation. The log-signature $\Omega_t = \Omega_t(T) := \log(\text{Sig}(\mathbf{X})_{t,T})$ satisfies the integral equation

$$\Omega_t(T) = \int_t^T H(\text{ad } \Omega_{u-})(d\mathbf{X}_u^c) + \sum_{t < u \leq T} \int_0^1 \Psi(\exp(\text{ad } \theta \Delta \mathbf{X}_u) \circ \exp(\text{ad } \Omega_u))(\Delta \mathbf{X}_u) d\theta, \quad (5.1)$$

with $\Psi(z) := H(\log z) = \log z / (z - 1)$ as in the introduction. The sum in equation (5.1) is absolutely convergent, over (at most countably many) jump times of \mathbf{X} , and vanishes when $\mathbf{X} \equiv \mathbf{X}^c$, in which case equation (1.4) reduces to Hausdorff's ODE.

(i) The accompanying Jump Magnus expansion becomes $\Omega_t^{(1)}(T) = \mathbf{X}_{t,T}^{(1)}$ followed by

$$\Omega_t^{(n)}(T) = \mathbf{X}_{t,T}^{(n)} + \sum_{|\ell| \geq 2, \|\ell\| = n} \left(\text{HMag}^1(\mathbf{X}, \Omega; \ell)_{t,T} + \text{HJmp}(\mathbf{X}, \Omega; \ell)_{t,T} \right)$$

where the right-hand side only depends on $\Omega^{(k)}$, $k < n$.

(ii) If $\mathbf{X} \in \mathcal{V}(V)$ for some linear subspace $V \subset \mathcal{T}_0 = T_0((\mathbb{R}^d))$, it follows that, for all $t \in [0, T]$,

$$\Omega_t(T) \in \mathcal{L} := \text{Lie}(V) \subset \mathcal{T}_0, \quad \text{Sig}(\mathbf{X})_{t,T} \in \exp(\mathcal{L}) \subset \mathcal{T}_1,$$

and we say that $\Omega_t(T)$ is Lie in V . In case $V = \mathbb{R}^d$, one speaks of (free) Lie series; see also [45, Def. 6.2].

Proof. Since we are in a purely deterministic setting, the signature cumulant coincides with the log-signature $\kappa_t(T) = \Omega_t(T)$, and Theorem 4.1 applies without any expectation and angle brackets.

Using $\Delta \Omega_u = \Omega_u - \Omega_{u-} = \Omega_u - \log(e^{\Delta \mathbf{X}_u} e^{\Omega_{u-}})$, we see that

$$\begin{aligned} \Omega_t(T) &= \int_t^T H(\text{ad } \Omega_{u-})(d\mathbf{X}_u^c) - \sum_{t < u \leq T} \Delta \Omega_u \\ &= \int_t^T H(\text{ad } \Omega_{u-})(d\mathbf{X}_u^c) - \sum_{t < u \leq T} (\Omega_u - \text{BCH}(\Delta \mathbf{X}_u, \Omega_u)) \\ &= \int_t^T H(\text{ad } \Omega_{u-})(d\mathbf{X}_u^c) + \sum_{t < u \leq T} \int_0^1 \Psi(\exp(\theta \text{ad } \Delta \mathbf{X}_u) \circ \exp(\text{ad } \Omega_u))(\Delta \mathbf{X}_u) d\theta, \end{aligned}$$

where we used the identity

$$\text{BCH}(\mathbf{x}_1, \mathbf{x}_2) - \mathbf{x}_2 = \log(\exp(\mathbf{x}_1) \exp(\mathbf{x}_2)) - \mathbf{x}_2 = \int_0^1 \Psi(\exp(\theta \text{ad } \mathbf{x}_1) \circ \exp(\text{ad } \mathbf{x}_2))(\mathbf{x}_1) d\theta. \quad (5.2)$$

□

Remark 5.2 (Baker–Campbell–Hausdorff). The identity equation (5.2) is well-known but also easy to obtain *en passant*, thereby rendering the above proof self-contained. We treat directly the n -fold case. Given $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{T}_0$, one defines a continuous piecewise affine linear path $(\mathbf{X}_t : 0 \leq t \leq n)$ with $\mathbf{X}_i - \mathbf{X}_{i-1} = \mathbf{x}_i$. Then $\text{Sig}(\mathbf{X}|_{[i-1,i]}) = \text{Sig}(\mathbf{X})_{i-1,i} = \exp(\mathbf{x}_i)$ and by Lemma 2.5 have

$$\Omega_0 = \log(\exp(\mathbf{x}_1) \cdots \exp(\mathbf{x}_n)) =: \text{BCH}(\mathbf{x}_1, \dots, \mathbf{x}_n).$$

A computation based on Theorem 5.1, but now applied without jumps, reveals the general form

$$\begin{aligned} \text{BCH}(\mathbf{x}_1, \dots, \mathbf{x}_n) &= \mathbf{x}_n + \sum_{k=1}^{n-1} \int_0^1 \Psi(\exp(\theta \text{ad } \mathbf{x}_k) \circ \exp(\text{ad } \mathbf{x}_{k+1}) \circ \dots \circ \exp(\text{ad } \mathbf{x}_n))(\mathbf{x}_k) \, d\theta \\ &= \sum_i \mathbf{x}_i + \frac{1}{2} \sum_{i < j} [\mathbf{x}_i, \mathbf{x}_j] + \frac{1}{12} \sum_{i < j} ([\mathbf{x}_i, [\mathbf{x}_i, \mathbf{x}_j]] + [\mathbf{x}_j, [\mathbf{x}_j, \mathbf{x}_i]]) \\ &\quad + \frac{1}{6} \sum_{i < j < k} ([\mathbf{x}_i, [\mathbf{x}_j, \mathbf{x}_k]] - [\mathbf{x}_k, [\mathbf{x}_i, \mathbf{x}_j]]) - \frac{1}{24} \sum_{i < j} [\mathbf{x}_i, [\mathbf{x}_j, [\mathbf{x}_i, \mathbf{x}_j]]] \dots \end{aligned}$$

The flexibility of our Theorem 5.1 is then nicely illustrated by the fact that this n -fold BCH formula is an immediate consequence of equation (5.1), applied to a *piecewise constant càdlàg* path $(\mathbf{X}_t : 0 \leq t \leq n)$ with $\mathbf{X}_i - \mathbf{X}_{i-1} \equiv \mathbf{x}_i$ on $[i - 1, i)$.

5.2. Diamond relations for multivariate cumulants

As in Section 2.3, we write \mathcal{S} for the symmetric algebra over \mathbb{R}^d , and $\mathcal{S}_0, \mathcal{S}_1$ for those elements with scalar component 0, 1, respectively. Recall the exponential map $\exp : \mathcal{S}_0 \rightarrow \mathcal{S}_1$ with global defined inverse \log . Following Definition 2.1, the diamond product for \mathcal{S}_0 -valued semimartingales $\hat{\mathbf{X}}, \hat{\mathbf{Y}}$ is another \mathcal{S}_0 -valued semimartingale given by

$$(\hat{\mathbf{X}} \diamond \hat{\mathbf{Y}})_t(T) = \mathbb{E}_t(\langle \hat{\mathbf{X}}^c, \hat{\mathbf{Y}}^c \rangle_{t,T}) = \sum (\mathbb{E}_t \langle \hat{\mathbf{X}}^{w_1}, \hat{\mathbf{Y}}^{w_2} \rangle_{t,T}) \hat{e}_{w_1} \hat{e}_{w_2},$$

with summation over all $w_1, w_2 \in \widehat{\mathcal{W}}_d$, provided all brackets are integrable. This trivially adapts to \mathcal{S}^N -valued semimartingales, $N \in \mathbb{N}_{\geq 1}$, in which case all words have length less equal N ; the summation is restricted accordingly to $|w_1| + |w_2| \leq N$.

Theorem 5.3. (i) Let $\Xi = (0, \Xi^{(1)}, \Xi^{(2)}, \dots)$ be an \mathcal{F}_T -measurable random variable with values in $\mathcal{S}_0(\mathbb{R}^d)$, componentwise in $\mathcal{L}^{\infty-}$. Then

$$\mathbb{K}_t(T) := \log \mathbb{E}_t \exp(\Xi)$$

satisfies the following functional equation, for all $0 \leq t \leq T$,

$$\mathbb{K}_t(T) = \mathbb{E}_t \Xi + \frac{1}{2} (\mathbb{K} \diamond \mathbb{K})_t(T) + \mathbb{J}_t(T) \tag{5.3}$$

with jump component,

$$\mathbb{J}_t(T) = \mathbb{E}_t \left(\sum_{t < u \leq T} (e^{\Delta \mathbb{K}_u} - 1 - \Delta \mathbb{K}_u) \right) = \mathbb{E}_t \left(\sum_{t < u \leq T} \left(\frac{1}{2!} (\Delta \mathbb{K}_u)^2 + \frac{1}{3!} (\Delta \mathbb{K}_u)^3 + \dots \right) \right).$$

Furthermore, if $N \in \mathbb{N}_{\geq 1}$, and $\Xi = (\Xi^{(1)}, \dots, \Xi^{(N)})$ is \mathcal{F}_T -measurable with graded integrability condition

$$\|\Xi^{(n)}\|_{\mathcal{L}^{N/n}} < \infty, \quad n = 1, \dots, N, \tag{5.4}$$

then the identity equation (5.3) holds for the cumulants up to level N : that is, for $\mathbb{K}^{(0,N)} := \log(\mathbb{E}_t \exp(\Xi^{(0,N)}))$ with values in $\mathcal{S}_0^{(N)}(\mathbb{R}^d)$.

Remark 5.4. Identity equation (5.3) is reminiscent to the quadratic form of the generalized Riccati equations for affine jump diffusions. The relation will be presented more explicitly in Remark 6.9 of

Section 6.2.2, when the involved processes are assumed to have a Markov structure and the functional signature cumulant equation reduces to a PIDE system. The framework described here, however, requires neither Markov nor affine structure. We will show in Section 6.3 that such computations are also possible in the fully non-commutative setting: that is, to obtain signature cumulants of affine Volterra processes.

Proof. We first observe that since $\Xi \in \mathcal{L}^{\infty-}$, by Doob’s maximal inequality and the BDG inequality, we have that $\hat{\mathbf{X}}_t := \mathbb{E}_t \Xi$ is a martingale in $\mathcal{H}^{\infty-}(\mathcal{S}_0)$. In particular, thanks to Theorem 3.2, the signature moments are well defined. According to Section 3.2, the signature is then given by

$$\text{Sig}(\hat{\mathbf{X}})_{t,T} = \exp(\Xi - \mathbb{E}_t \Xi),$$

hence $\hat{\kappa}_t(T) = \mathbb{K}_t(T) - \hat{\mathbf{X}}_t$.

Projecting equation (4.3) onto the symmetric algebra yields

$$\begin{aligned} \hat{\kappa}_t(T) &= \mathbb{E}_t \left\{ \hat{\mathbf{X}}_{t,T} + \frac{1}{2} \langle \hat{\mathbf{X}}^c \rangle_{t,T} + \frac{1}{2} \langle \kappa(T)^c \rangle_{t,T} + \langle \hat{\mathbf{X}}^c, \kappa(T)^c \rangle_{t,T} \right. \\ &\quad \left. + \sum_{t < u \leq T} \left(e^{\Delta \hat{\mathbf{X}}_u + \Delta \kappa_u(T)} - 1 - \Delta \hat{\mathbf{X}}_u - \Delta \kappa_u(T) \right) \right\} \\ &= \mathbb{E}_t \left\{ \Xi + \frac{1}{2} \langle \mathbb{K}(T)^c \rangle_{t,T} + \sum_{t < u \leq T} \left(e^{\Delta \mathbb{K}_u(T)} - 1 - \Delta \mathbb{K}_u(T) \right) \right\} - \hat{\mathbf{X}}_t, \end{aligned}$$

and equation (5.3) follows upon recalling that $(\mathbb{K} \diamond \mathbb{K})_t(T) = \mathbb{E}_t \langle \mathbb{K}(T)^c \rangle_{t,T}$. The proof of the truncated version is left to the reader. □

As a corollary, we provide a general view on recent results of [3, 42, 23]. Note that we also include jump terms in our recursion.

Corollary 5.5. *The conditional multivariate cumulants $(\mathbb{K}_t)_{0 \leq t \leq T}$ of a random variable Ξ with values in $\mathcal{S}_0(\mathbb{R}^d)$, componentwise in $\mathcal{L}^{\infty-}$ satisfy the recursion*

$$\mathbb{K}_t^{(1)} = \mathbb{E}_t(\Xi^{(1)}) \quad \text{and} \quad \mathbb{K}_t^{(n)} = \mathbb{E}_t(\Xi^{(n)}) + \frac{1}{2} \sum_{k=1}^n \left(\mathbb{K}^{(k)} \diamond \mathbb{K}^{(n-k)} \right)_t(T) + \mathbb{J}_t^{(n)}(T) \quad \text{for } n \geq 2, \tag{5.5}$$

with

$$\mathbb{J}_t^{(n)}(T) = \mathbb{E}_t \left(\sum_{t < u \leq T} \sum_{k=2}^n \frac{1}{k!} \sum_{\|\ell\|=n, |\ell|=k} \Delta \mathbb{K}_u^{(\ell_1)}(T) \cdots \Delta \mathbb{K}_u^{(\ell_k)}(T) \right).$$

The analogous statement holds true in the N-truncated setting: that is, as a recursion for $n = 1, \dots, N$ under the condition in equation (5.4).

Example 5.6 (Continuous setting). In case of an absence of jumps and higher-order information (i.e., $\mathbb{J} \equiv 0, \Xi^{(2)} = \Xi^{(3)} = \dots \equiv 0$), this type of cumulant recursion appears in [42] and under optimal integrability conditions $\Xi^{(1)}$ with finite N .th moments [23]. (This requires a localization argument that is avoided here by directly working in the correct algebraic structure.)

Example 5.7 (Discrete filtration). As the opposite of the previous continuous example, we consider a purely discrete situation, starting from a discretely filtered probability space with filtration $(\mathcal{F}_t : t = 0, 1, \dots, T \in \mathbb{N})$. For Ξ as in Corollary 5.5, a discrete martingale is defined by $\mathbb{E}_t \exp(\Xi)$, which we may regard as a càdlàg semimartingale with respect to $\mathcal{F}_t := \mathcal{F}_{[t]}$, and similar for $\mathbb{K}_t(T) =$

$\log \mathbb{E}_t \exp(\Xi) \in \mathcal{S}_0$: that is, the conditional cumulants of Ξ . Clearly, the continuous martingale part of $\mathbb{K}(T)$ vanishes, as does any diamond product with $\mathbb{K}(T)$. What remains is the functional equation

$$\mathbb{K}_t(T) = \mathbb{E}_t(\Xi) + \mathbb{J}_t(T) = \mathbb{E}_t(\Xi) + \mathbb{E}_t\left(\sum_{u=t+1}^T (\exp(\Delta \mathbb{K}_u) - 1 - \Delta \mathbb{K}_u)\right).$$

As before, the resulting expansions are of interest. On the first level, trivially, $\mathbb{K}_t^{(1)} = \mathbb{E}_t(\Xi^{(1)})$, whereas on the second level we see

$$\mathbb{K}_t^{(2)}(T) = \mathbb{E}_t(\Xi^{(2)}) + \mathbb{E}_t\left(\sum_{u=t+1}^T (\mathbb{E}_u(\Xi^{(1)}) - \mathbb{E}_{u-1}(\Xi^{(1)}))^2\right),$$

which one can recognize, in case $\Xi^{(2)} = 0$ as an energy identity for the discrete square-integrable martingale $\ell_u := \mathbb{E}_u \Xi^{(1)}$. Going further in the recursion yields increasingly non-obvious relations. Taking $\Xi^{(2)} = \Xi^{(3)} = \dots \equiv 0$ for notational simplicity gives

$$\mathbb{K}_t^{(3)}(T) = \mathbb{E}_t\left(\sum_{u=t+1}^T (\ell_u - \ell_{u-1})^3 + 3(\ell_u - \ell_{u-1})\{\mathbb{E}_u \kappa(\ell, \ell)_{u,T} - \mathbb{E}_{u-1} \kappa(\ell, \ell)_{u-1,T}\}\right).$$

It is interesting to note that related identities have appeared in the statistics literature under the name *Bartlett identities*; see also Mykland [53] and the references therein.

5.3. Remark on tree representation

As illustrated in the previous section, in the case where $d = 1$, or when projecting onto the symmetric algebra, our functional equation takes a particularly simple form (see Theorem 5.3). If one further specializes the situation, in particular discards all jumps, we are from an algebraic perspective in the setting of Friz, Gatheral and Radoičić [23], which give a tree series expansion of cumulants using binary trees. This representation follows from the fact that the diamond product of semimartingales is commutative but not associative. As an example (with notations taken from Section 5.2), in case of a one-dimensional continuous martingale, the first terms are

$$\mathbb{K}_t(T) = \bullet + \frac{1}{2} \text{tree}_1 + \frac{1}{2} \text{tree}_2 + \frac{1}{2} \text{tree}_3 + \frac{1}{8} \text{tree}_4 + \dots$$

This expansion is organized (graded) in terms of the number of leaves in each tree, and each leaf represents the underlying martingale.

In the deterministic case, tree expansions are also known for the Magnus expansion [33] and the BCH formula [9]. These expansions are also, in terms of binary trees, *different* from the ones above as they are required to be *non-planar* to account for the non-commutativity of the Lie algebra. As an example (with the notations of Section 5.1), we have

$$\Omega_t(T) = \bullet + \frac{1}{2} \text{tree}_1 + \frac{1}{12} \text{tree}_2 + \frac{1}{4} \text{tree}_3 + \dots$$

In this expansion, the nodes represent the underlying vector field and edges represent integration and application of the Lie bracket, coming from the ad operator.

Since our functional equation and the associated recursion puts both contexts into a single common framework. We suspect that our general recursion, Corollary 4.2 and thereafter, allows for a sophisticated tree representation, at least in absence of jumps, and propose to return to this question in future work.

6. Applications

6.1. Brownian and stopped Brownian signature cumulants

6.1.1. Time dependent Brownian motion

Let B be a m -dimensional standard Brownian motion defined on a portability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the canonical filtration $(\mathcal{F}_t)_{t \geq 0}$, and define the continuous (Gaussian) martingale $X = (X_t)_{0 \leq t \leq T}$ by

$$X_t = \int_0^t \sigma(u) dB_u, \quad 0 \leq t \leq T,$$

with $\sigma \in L^2([0, T], \mathbb{R}^{m \times d})$. The quadratic variation of X is finite and deterministic, and therefore we immediately see that the integrability condition $\mathbf{X} = (0, X, 0, \dots) \in \mathcal{H}^{\infty-}$ is trivially satisfied, and thus Theorem 4.1 applies. The *Brownian signature cumulants* $\kappa_t(T) = \log(\mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}))$ satisfies the functional equation, with $\mathbf{a}(t) := \sigma(t)\sigma(t)^T \in \text{Sym}(\mathbb{R}^d \otimes \mathbb{R}^d)$,

$$\kappa_t(T) = \frac{1}{2} \int_t^T H(\text{ad } \kappa_u(T))(\mathbf{a}(u))du, \quad 0 \leq t \leq T. \tag{6.1}$$

Therefore the tensor levels are precisely given by the Magnus expansion, starting with

$$\kappa_t^{(1)}(T) = 0, \quad \kappa_t^{(2)}(T) = \frac{1}{2} \int_t^T \mathbf{a}(u)du,$$

and the general term

$$\begin{aligned} \kappa_t^{(2n-1)}(T) &\equiv 0, \quad \kappa_t^{(2n)}(T) = \frac{1}{2} \sum_{|\ell| \geq 2, \|\ell\|=2n} \text{HMag}^2(\mathbf{X}, \kappa; \ell)_{t,T} \\ &= \frac{1}{2} \sum_{\|\ell\|=n-1} \frac{B_k}{k!} \int_t^T \text{ad } \kappa_u^{(2 \cdot l_1)} \dots \text{ad } \kappa_u^{(2 \cdot l_k)}(\mathbf{a}(u))du. \end{aligned}$$

Note that $\kappa_t(T)$ is Lie in $\text{Sym}(\mathbb{R}^d \otimes \mathbb{R}^d) \subset \mathcal{T}_0$, but, in general, not a Lie series. In the special case $X = B$, that is, $m = d$ and identity matrix $\sigma = \mathbf{I}_d = \sum_{i=1}^d e_{ii} \in \text{Sym}(\mathbb{R}^d \otimes \mathbb{R}^d)$, all commutators vanish, and we obtain what is known as *Fawcett's formula* [21, 24]:

$$\kappa_t(T) = \frac{1}{2}(T - t)\mathbf{I}_d.$$

Example 6.1. Consider B^1, B^2 two Brownian motions on the filtered space $(\Omega, \mathcal{F}, \mathbb{P})$, with correlation $d\langle B^1, B^2 \rangle_t = \rho dt$ for some fixed constant $\rho \in [-1, 1]$. Suppose that $K^1, K^2: [0, \infty)^2 \rightarrow \mathbb{R}$ are two kernels such that $K^i(t, \cdot) \in L^2([0, t])$ for all $t \in [0, T]$, and set

$$X_t^i := X_0^i + \int_0^t K^i(t, s) dB_s^i, \quad i = 1, 2$$

for some fixed initial values X_0^1, X_0^2 . Note that neither process is a semimartingale in general. However, for each $T > 0$, the process $\xi_t^i(T) := \mathbb{E}_t[X_T^i]$ is a martingale, and we have

$$\xi_t^i(T) = X_0^i + \int_0^t K^i(T, s) dB_s^i,$$

that is, (ξ^1, ξ^2) is a time-dependent Brownian motion as defined above. In particular, one sees that

$$\mathbf{a}(t) = \begin{pmatrix} \int_0^t K^1(T, u)^2 du & \rho \int_0^t K^1(T, u)K^2(T, u) du \\ \rho \int_0^t K^1(T, u)K^2(T, u) du & \int_0^t K^2(T, u)^2 du \end{pmatrix}.$$

Equation (6.1) and the paragraph below it then give an explicit recursive formula for the signature cumulants, the first of which are given by

$$\begin{aligned} \kappa_t^{(1)}(T) &= 0, \\ \kappa_t^{(2)}(T) &= \frac{1}{2} \begin{pmatrix} \int_t^T \int_0^u K^1(T, r)^2 dr du & \rho \int_t^T \int_0^u K^1(T, r)K^2(T, r) dr du \\ \rho \int_t^T \int_0^u K^1(T, r)K^2(T, r) dr du & \int_t^T \int_0^u K^2(T, r)^2 dr du \end{pmatrix}, \\ \kappa_t^{(3)}(T) &= 0, \\ \kappa_t^{(4)}(T) &= \frac{1}{2} \sum_{i,j,i',j'=1}^2 \left[\int_t^T \int_u^T \left(\mathbf{a}^{ij}(u)\mathbf{a}^{i'j'}(r) - \mathbf{a}^{i'j'}(u)\mathbf{a}^{ij}(r) \right) dr du \right] e_{ij i' j'}. \end{aligned}$$

We notice that in the particular case when $K^1 = K^2 \equiv K$, the matrix \mathbf{a} has the form

$$\mathbf{a}(t) = \int_0^t K(T, u)^2 du \times \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Therefore, we have $\mathbf{a}(t) \otimes \mathbf{a}(t') - \mathbf{a}(t') \otimes \mathbf{a}(t) = 0$ for any $t, t' \in [0, T]$. Hence, in this case, our recursion shows that for any $\rho \in [-1, 1]$,

$$\kappa_t^{(1)}(T) = 0, \quad \kappa_t^{(2)}(T) = \frac{1}{2} \int_t^T \int_0^u K(T, r)^2 dr du \times \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

and $\kappa_t^{(n)}(T) = 0$ for all $0 \leq t \leq T$ and $n \geq 3$.

6.1.2. Brownian motion up to the first exit time from a domain

Let $B = (B_t)_{t \geq 0}$ be a d -dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a possibly random starting value B_0 . Assume also that there is a family of probability measures $\{\mathbb{P}^x\}_{x \in \mathbb{R}^d}$ on (Ω, \mathcal{F}) such that $\mathbb{P}^x(B_0 = x) = 1$, and denote by \mathbb{E}^x the expectation with respect to \mathbb{P}^x . We define the canonical Brownian filtration⁸ by $(\mathcal{F}_t)_{t \geq 0} = (\mathcal{F}_t^B)_{t \geq 0}$. Further let $\Gamma \subset \mathbb{R}^d$ be a bounded domain, and define the stopping time τ_Γ of the first exit of B from the domain Γ : that is,

$$\tau_\Gamma = \inf\{t \geq 0 \mid B_t \in \Gamma^c\}.$$

In [46], Lyons–Ni exhibit an infinite system of partial differential equations for the expected signature of the Brownian motion until the exit time as a functional of the starting point. The following result can be seen as the corresponding result for the signature cumulant, which follows directly from the expansion in Theorem 1.1. Recall that a boundary point $x \in \partial\Gamma$ is called *regular* if and only if

$$\mathbb{P}^x(\inf\{t > 0 \mid B_t \in \Gamma^c\} = 0) = 1. \tag{6.2}$$

The domain Γ is called *regular* if all points on the boundary are regular. For example, domains with smooth boundary are regular; and see [38, Section 4.2.C] for a further characterization of regularity.

⁸The smallest right-continuous, \mathbb{P} and $\{\mathbb{P}^x\}$ -complete filtration in (Ω, \mathcal{F}) that the Brownian motion B is adapted to.

Corollary 6.2. *Let $\Gamma \subset \mathbb{R}^d$ be a regular domain, such that*

$$\sup_{x \in \Gamma} \mathbb{E}^x(\tau_\Gamma^n) < \infty, \quad n \in \mathbb{N}_{\geq 1}. \tag{6.3}$$

The signature cumulant $\kappa_t = \log(\mathbb{E}(\text{Sig}(B)_{t \wedge \tau_\Gamma, \tau_\Gamma}))$ of the Brownian motion B up to the first exit from the domain Γ has the following form

$$\kappa_t = \mathbf{1}_{\{t < \tau_\Gamma\}} \mathbf{F}(B_t), \quad t \geq 0,$$

where $\mathbf{F} = \sum_{|w| \geq 2} e_w F^w$ with $F^w \in C^0(\bar{\Gamma}, \mathbb{R}) \cap C^2(\Gamma, \mathbb{R})$ is the unique bounded classical solution to the elliptic PDE

$$-\Delta \mathbf{F}(x) = \sum_{i=1}^d H(\text{ad } \mathbf{F}(x)) \left(e_{ii} + Q(\text{ad } \mathbf{F}(x)) (\partial_i \mathbf{F}(x))^{\otimes 2} + 2e_i G(\text{ad } \mathbf{F}(x)) (\partial_i \mathbf{F}(x)) \right), \tag{6.4}$$

for all $x \in \Gamma$ with the boundary condition $\mathbf{F}|_{\partial\Gamma} \equiv 0$.

Proof. Define the martingale $\mathbf{X} = ((0, B_{t \wedge \tau_\Gamma}, 0, \dots))_{t \geq 0} \in \mathcal{S}(\mathcal{T}_0)$, and note that $|\langle \mathbf{X} \rangle_\infty| = \mathbf{I}_d |_{\tau_\Gamma}$. It then follows from the integrability of τ_Γ that $\mathbf{X} \in \mathcal{H}^{\infty-}(\mathcal{T}_0)$ and thus by Theorem 3.2 that $(\text{Sig}(\mathbf{X})_{0,t})_{t \geq 0} \in \mathcal{H}(\mathcal{T}_1)^{\infty-}$. This implies that the signature cumulant $\kappa_t(T) := \log(\mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}))$ is well defined for all $0 \leq t \leq T < \infty$, and furthermore under (component-wise) application of the dominated convergence theorem that it holds

$$\kappa_t = \lim_{T \rightarrow \infty} \kappa_t(T) = \lim_{T \rightarrow \infty} \log(\mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T})) = \log(\mathbb{E}_t(\text{Sig}(B)_{t \wedge \tau_\Gamma, \tau_\Gamma})), \quad t \geq 0.$$

Again by $\mathbf{X} \in \mathcal{H}^{\infty-}(\mathcal{T}_0)$, it follows that Theorem 1.1 applies to the martingale $(\mathbf{X}_t)_{0 \leq t \leq T}$ for any $T > 0$ and therefore $\kappa(T)$ satisfies the functional equation (4.3). It follows from the Itô’s representation theorem [59, Theorem 3.4] that all local martingales with respect to the Brownian filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ are continuous, and therefore it is easy to see that also $\kappa(T) \in \mathcal{S}^c(\mathcal{T}_0)$. Therefore equation (4.3) simplifies to the following equation

$$\begin{aligned} \kappa_t(T) = \mathbf{1}_{\{t < \tau_\Gamma\}} \mathbb{E}_t \left\{ \frac{1}{2} \int_t^{\tau_\Gamma \wedge T} H(\text{ad } \kappa_u) (\mathbf{I}_d) du + \int_t^{\tau_\Gamma \wedge T} \frac{1}{2} H(\text{ad } \kappa_u) \circ Q(\text{ad } \kappa_u) (d\llbracket \kappa, \kappa \rrbracket_u) \right. \\ \left. + \int_t^{\tau_\Gamma \wedge T} H(\text{ad } \kappa_u) \circ (\text{Id} \odot G(\text{ad } \kappa_u)) (d\llbracket X, \kappa \rrbracket_u) \right\}, \tag{6.5} \end{aligned}$$

where we have already used the martingality of \mathbf{X} and the explicit form of the quadratic variation $\langle \mathbf{X} \rangle_t = \mathbf{I}_d(t \wedge \tau_\Gamma)$ with $\mathbf{I}_d = \sum_{i=1}^d e_{ii} \in (\mathbb{R}^d)^{\otimes 2}$. It follows that $\kappa^{(1)} \equiv \kappa(T)^{(1)} \equiv 0$, and for the second level, we have from the integrability of τ_Γ and the strong Markov property of Brownian motion that

$$\kappa_t^{(2)} = \frac{1}{2} \mathbf{I}_d \lim_{T \rightarrow \infty} \mathbb{E}_t(\mathbf{1}_{\{t < \tau_\Gamma\}} (\tau_\Gamma \wedge T - t)) = \frac{1}{2} \mathbf{I}_d \mathbf{1}_{\{t < \tau_\Gamma\}} \mathbb{E}^x(\tau_\Gamma)|_{x=B_t}, \quad t \geq 0.$$

Now note that the function $u(x) := \mathbb{E}^x(\tau_\Gamma)$ for $x \in \Gamma$ is in $C^0(\bar{\Gamma}, \mathbb{R}) \cap C^2(\Gamma, \mathbb{R})$ and solves the Poisson equation $-(1/2)\Delta u = g$ with boundary condition $u|_{\partial\Gamma} = 0$ and data $g \equiv 1$. Indeed, since Γ is regular and g is bounded and differentiable, this follows from Theorem 9.3.3 (and the remark thereafter) in [55]. Moreover, from the assumption in equation (6.3), we immediately see that u is bounded on $\bar{\Gamma}$, and it follows from Theorem 9.3.2 in [55] that u is the unique bounded classical such solution. Thus we have shown that the statement holds true up to the second tensor level with $\mathbf{F}^{(1)} \equiv 0$ and $\mathbf{F}^{(2)}(u) = \frac{1}{2} \mathbf{I}_d u$ under the usual notation $\mathbf{F}^{(n)} = \sum_{|w|=n} e_w F^w$.

Now assume that the statement of the corollary holds true up to the tensor level $(N - 1)$ for some $N \geq 3$. Then, for any $n, k < N$, we have by applying Itô's formula

$$\left[\kappa^{(n)}, \kappa^{(k)} \right]_t = \sum_{i=1}^d \int_0^{t \wedge \tau_\Gamma} (\partial_i \mathbf{F}^{(n)}(B_u)) \otimes (\partial_i \mathbf{F}^{(k)}(B_u)) \, du, \quad t \geq 0,$$

and

$$\left[\mathbf{X}, \kappa^{(n)} \right]_t = \sum_{i=1}^d \int_0^{t \wedge \tau_\Gamma} e_i \otimes (\partial_i \mathbf{F}^{(n)}(B_u)) \, du, \quad t \geq 0.$$

Further define the function $\mathbf{G}^{(N)}$ by the projection under π_N of the right-hand side of equation (6.4) multiplied by the factor $1/2$. Then applying Theorem 4.1 to $\mathbf{X}^{(0,N)}$ on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}^x)$, we see that it follows from the estimate in equation (7.30) that there exists a constant $c > 0$ such that

$$\sup_{x \in \Gamma} \mathbb{E}^x \left\{ \int_0^{\tau_\Gamma} |\mathbf{G}^{(N)}(B_u)| \, du \right\} \leq c \sup_{x \in \Gamma} \|\mathbf{X}^{(0,N)}\|_{\mathcal{H}^{1,N}(\mathbb{P}^x)} = c \sup_{x \in \Gamma} \mathbb{E}_x(\tau_\Gamma^N) < \infty.$$

Therefore it follows, from projecting equation (6.5) to level N and using the dominated convergence theorem to pass to the $T \rightarrow \infty$ limit, that $\kappa^{(N)}$ is of the form

$$\kappa_t^{(N)} = \mathbf{1}_{\{t < \tau_\Gamma\}} \mathbf{F}^{(N)}(B_t) \quad \text{with} \quad \mathbf{F}^{(N)}(x) := \mathbb{E}^x \left\{ \int_0^{\tau_\Gamma} \mathbf{G}^{(N)}(B_u) \, du \right\}, \quad x \in \bar{\Gamma}.$$

Furthermore, by the assumption, it also holds that $G^w \in C^1(\Gamma)$ for all $w \in \mathcal{W}_d, |w| = N$. Therefore we can conclude again with Theorem 9.3.3 in [55] that $F^w \in C^0(\bar{\Gamma}, \mathbb{R}) \cap C^2(\bar{\Gamma}, \mathbb{R})$ solves the Poisson equation with data $g = G^w$ for all words w with $|w| = N$. The statement then follows by induction. \square

Example 6.3. For $n \in \{1, \dots, d\}$, let \mathbb{D}^n be the open unit ball in \mathbb{R}^n , and define the (regular) domain $\Gamma = \mathbb{D}^n \times \mathbb{R}^{d-n} \subset \mathbb{R}^d$. Further note that it holds

$$\tau_\Gamma = \inf\{t \geq 0 \mid B_t \notin \Gamma\} = \inf\{t \geq 0 \mid |(B_t^1, \dots, B_t^n)| \geq 1\}.$$

Hence we readily see that τ_Γ satisfies the condition in equation (6.3). Applying Corollary 6.2, it follows that the signature cumulant of the Brownian motion B up to the exit of the domain Γ is of the form $\kappa_t = \mathbf{1}_{\{t < \tau_\Gamma\}} \mathbf{F}(B_t)$, where \mathbf{F} satisfies the PDE (6.4). Recall that $\mathbf{F}^{(1)} \equiv 0$ and projecting to the second level we see that

$$-\Delta \mathbf{F}^{(2)}(x) = \mathbf{I}_d, \quad x \in \Gamma; \quad \mathbf{F}^{(2)}|_{\partial\Gamma} \equiv 0.$$

The unique bounded solution of the above Poisson equation is given by

$$\mathbf{F}^{(2)}(x) = \frac{1}{2} \mathbf{I}_d \left(1 - \sum_{i=1}^n x_i^2 \right), \quad x \in \Gamma.$$

More generally, we see that the Poisson equation $\Delta u = -g$ on Γ with zero boundary condition, where $g: \Gamma \rightarrow \mathbb{R}$ is a polynomial in the first n -variables, has a unique bounded solution u , which is also a polynomial of the first n -variables of degree $\deg(u) = \deg(g) + 2$ and has the factor $(1 - \sum_{i=1}^n x_i^2)$ (see Lemma 3.10 in [46]). Hence it follows inductively that each component of $\mathbf{F}^{(n)}$ is a polynomial of degree n with the factor $(1 - \sum_{i=1}^n x_i^2)$. The precise coefficients of the polynomial can be obtained as the solution to a system of linear equations recursively derived from the forcing term in (6.4). This is similar to [46, Theorem 3.5]; however, we note that a direct conversion of the latter result for the

expected signature to signature cumulants is not trivially seen to yield the same recursion and requires combinatorial relations as studied in [7].

6.2. Lévy and diffusion processes

Let $X \in \mathcal{S}(\mathbb{R}^d)$, and throughout this section assume that the filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ is generated by X . Denote by ε_a the Dirac measure at point $a \in \mathbb{R}^d$; the random measure μ^X associated to the jumps of X is an integer-valued random measure of the form

$$\mu^X(\omega; dt, dx) := \sum_{s \geq 0} \mathbf{1}_{\{\Delta X_s(\omega) \neq 0\}} \varepsilon_{(s, \Delta X_s(\omega))}(dt, dx).$$

There is a version of the predictable compensator of μ^X , denoted by ν , such that the \mathbb{R}^d -valued semimartingale X is quasi-left continuous if and only if $\nu(\omega, \{t\} \times \mathbb{R}^d) = 0$ for all $\omega \in \Omega$; see [36, Corollary II.1.19]. In general, ν satisfies $(|x|^2 \wedge 1) * \nu \in \mathcal{A}_{loc}$: that is, locally of integrable variation. The semimartingale X admits a *canonical representation* (using the usual notation for stochastic integrals with respect to random measures as introduced, for example, in [36, II.1])

$$X = X_0 + B(h) + X^c + (x - h(x)) * \mu^X + h(x) * (\mu^X - \nu), \tag{6.6}$$

where $h(x) = x \mathbf{1}_{|x| \leq 1}$ is a truncation function (other choice are possible). Here $B(h)$ is a predictable \mathbb{R}^d -valued process with components in \mathcal{V} and X^c is the continuous martingale part of X .

Denote by C the predictable $\mathbb{R}^d \otimes \mathbb{R}^d$ -valued covariation process defined as $C^{ij} := \langle X^{i,c}, X^{j,c} \rangle$. Then the triplet $(B(h), C, \nu)$ is called the *triplet of predictable characteristics* of X (or simply the *characteristics* of X). In many cases of interest, including the case of Lévy and diffusion processes discussed in the subsection below, we have differential characteristics (b, c, K) such that

$$dB_t = b_t(\omega)dt, \quad dC_t = c_t(\omega)dt, \quad \nu(dt, dx) = K_t(dx; \omega)dt,$$

where b is a d -dimensional predictable process, c is a predictable process taking values in the set of symmetric non-negative definite $d \times d$ -matrices and K is a transition kernel from $(\Omega \times \mathbb{R}_+, \mathcal{B}^d)$ into $(\mathbb{R}^d, \mathcal{B}^d)$. We call such a process *Itô semimartingale* and the triplet (b, c, K) its *differential* (or *local characteristics*). This extends mutatis mutandis to an \mathcal{T}_0^N (and then \mathcal{T}_0) valued semimartingale \mathbf{X} , with local characteristics $(\mathbf{b}, \mathbf{c}, \mathbf{K})$.

While every Itô semimartingale is quasi-left continuous, it is in general not true that κ is continuous (with the notable exception of time-inhomogeneous Lévy processes discussed below), and therefore there is no significant simplification of the functional equation (4.3) in these general terms. The following example illustrates this point in more detail.

Example 6.4. Take $X \in \mathcal{S}(\mathbb{R}^d)$ and then $d = 1$, so that we are effectively in the symmetric setting. In this case $\exp(\kappa_t(T)) = \mathbb{E}_t(\exp(X_T - X_t))$, in the power series sense of enlisting all moments with factorial factors. These can also be obtained by taking higher-order derivatives at $u = 0$ of $\mathbb{E}_t(\exp(u(X_T - X_t)))$, now with the classical calculus interpretation of the exponential. The important class of affine models satisfies

$$\mathbb{E}_t(\exp(uX_T - uX_t)) = \exp(\phi(T - t, u) + (\Psi(T - t, u) - u)X_t)$$

In the Levy-case, we have the trivial situation $\Psi(\cdot, u) \equiv u$, but otherwise (ϕ, Ψ) solve (generalized) Riccati equations and are in particular continuous in $T - t$. We see that, in non-trivial situations, the log of $\mathbb{E}_t(\exp(uX_T - uX_t))$ and any of its derivatives will jump when X jumps. In particular, $\kappa_t(T)$ will not be continuous in t , even if X is quasi-left continuous: that is, when $\Delta X_\tau = 0$ a.s. for all predictable

times τ (see [36, p. 22]). Let us note in this context that, in the general non-commutative setting and directly from definition of κ ,

$$\exp(\kappa_{t-}) = \mathbb{E}_{t-}(\exp(\Delta X_t) \exp(\kappa_t)) = \mathbb{E}_{t-}(\exp \kappa_t),$$

where the second equality holds true under the assumption of quasi-left continuity of X . If we assume for a moment $\mathcal{F}_{t-} = \mathcal{F}_t$, then we could conclude that $\kappa_{t-} = \kappa_t$ and hence (right-continuity is clear) that κ_t is continuous in t . Since we know that this fails beyond Lévy processes, such left continuity of filtrations is not a good assumption, at least not beyond Lévy processes.

6.2.1. The case of time-inhomogeneous Lévy processes

We consider now a d -dimensional time-inhomogeneous Lévy processes of the form

$$X_t = \int_0^t \mathbf{b}(u) \, du + \int_0^t \sigma(u) \, dB_u + \int_{(0,t]} \int_{|x| \leq 1} x (\mu^X - \nu)(ds, dx) + \int_{(0,t]} \int_{|x| > 1} x \mu^X(ds, dx), \tag{6.7}$$

for all $0 \leq t \leq T$, with $b \in L^1([0, T], \mathbb{R}^d)$, $\sigma \in L^2([0, T], \mathbb{R}^{m \times d})$, B a d -dimensional Brownian motion and μ^X is an independent inhomogeneous Poisson random measure with the intensity measure ν on $[0, T] \times \mathbb{R}^d$, such that $\nu(dt, dx) = K_t(dx)dt$ with Lévy measures K_t : that is, $K_t(\{0\}) = 0$, and

$$\int_0^T \int_{\mathbb{R}^d} (|x|^2 \wedge 1) K_t(dx)dt < \infty,$$

and measurability of $t \mapsto K_t(A) \in [0, \infty]$, any measurable $A \subset \mathbb{R}^d$. Consider further the condition

$$\int_0^T \int_{\mathbb{R}^d} |x|^N \mathbf{1}_{|x| > 1} K_t(dx) < \infty \tag{6.8}$$

for some integer $N \in \mathbb{N}_{\geq 1}$. The Brownian case in equation (6.1) then generalizes as follows.

Corollary 6.5. *Let X be an inhomogenous Lévy process of the form in equation (6.7), such that the family of Lévy measures $\{K_t\}_{t>0}$ satisfy the moment condition in equation (6.8) for all $N \in \mathbb{N}_{\geq 1}$. Then $X \in \mathcal{H}^{\infty-}(\mathbb{R}^d)$, and the signature cumulant $\kappa_t := \log(\mathbb{E}_t(\text{Sig}(X)_{t,T}))$ satisfies the following integral equation*

$$\kappa_t = \int_t^T H(\text{ad } \kappa_u)(\eta(u)) \, du, \quad 0 < t \leq T, \tag{6.9}$$

with $\mathbf{a}(t) = \sigma(t)\sigma(t)^T \in \mathbb{R}^d \otimes \mathbb{R}^d \subset \mathcal{T}_0$ and

$$\eta(t) := \mathbf{b}(t) + \frac{1}{2} \mathbf{a}(t) + \int_{\mathbb{R}^d} (\exp(\mathbf{x}) - 1 - \mathbf{x} \mathbf{1}_{|x| \leq 1}) K_t(dx) \in \mathcal{T}_0, \tag{6.10}$$

where $\mathbf{x} = (0, x, 0, \dots) \in \mathcal{T}_0$ for $x \in \mathbb{R}^d$. In case the Lévy measures $\{K_t\}_{t>0}$ satisfy the condition in equation (6.8) only up to some finite level $N \in \mathbb{N}_{\geq 1}$, we have $X \in \mathcal{H}^N$, and the identity equation (6.9) holds for the truncated signature cumulant in \mathcal{T}_0^N .

Remark 6.6. Corollary 6.5 extends a main result of [25], where a Lévy-Kintchin type formula was obtained for the expected signature of time homogeneous Lévy processes with constant triplet $(\mathbf{b}, \mathbf{a}, K)$.

Now this is an immediate consequence of equation (6.9), with all commutators vanishing in time-homogeneous case and explicit solution

$$\kappa_t(T) = (T - t) \left(\mathbf{b} + \frac{1}{2} \mathbf{a} + \int_{\mathbb{R}^d} (\exp(\mathbf{x}) - 1 - \mathbf{x} \mathbf{1}_{|x| \leq 1}) K(\mathrm{d}x) \right).$$

Proof. Assume that the Lévy measures $\{K_t\}_{t>0}$ satisfy the condition in equation (6.8) for some $N \in \mathbb{N}_{\geq 1}$. We will first show that $X \in \mathcal{H}^N(\mathbb{R}^d)$. Note that the decomposition in equation (6.7) naturally yields a semimartingale decomposition $X = M + A$, where the local martingale M and the adapted bounded variation process A are defined by

$$M = \int_0^\cdot \sigma(u) \mathrm{d}B_u + (x \mathbf{1}_{|x| \leq 1}) * (\mu^X - \nu), \quad A = \int_0^\cdot \mathbf{b}(u) \mathrm{d}u + (x \mathbf{1}_{|x| > 1}) * \mu^X.$$

Regarding the integrability of the 1-variation of A , we note that it holds

$$|A|_{1\text{-var}; [0, T]} = \int_0^T |\mathbf{b}(u)| \mathrm{d}u + (|x| \mathbf{1}_{|x| > 1}) * \mu_T^X.$$

Define the increasing, piecewise constant process $V := (|x| \mathbf{1}_{|x| > 1}) * \mu^X$. Since b is deterministic and integrable over the interval $[0, T]$, it suffices to show that V_T has finite N th moment. To this end, note that it holds

$$\mathbb{E}(V_T) = \int_0^T \int_{|x| > 1} |x| K_t(\mathrm{d}x) \mathrm{d}t < \infty.$$

Further, it holds that for any $n \in \{1, \dots, N\}$ that

$$V_T^n = \sum_{0 < t \leq T} (V_t^n - V_{t-}^n) = \sum_{0 < t \leq T} \sum_{k=0}^{n-1} \binom{n}{k} V_{s-}^k (\Delta V_s)^{n-k}$$

and by definition $\Delta V_t = |\Delta X_t| \mathbf{1}_{|\Delta X_t| > 1}$. Now let $n = 2$ and $k \in \{0, \dots, n - 1\}$; then we have

$$\begin{aligned} \mathbb{E} \left(\sum_{0 < t \leq T} V_{t-}^k (\Delta V_t)^{n-k} \right) &= \mathbb{E} \left(\int_0^T \int_{|x| > 1} V_{t-}^k |x|^{n-k} K_t(\mathrm{d}x) \mathrm{d}t \right) \\ &\leq \mathbb{E} \left(V_T^k \right) \int_0^T \int_{|x| > 1} |x|^{n-k} K_t(\mathrm{d}x) \mathrm{d}t < \infty. \end{aligned}$$

It then follows inductively that $\mathbb{E}(V_T^n)$ is finite for all $n = 1, \dots, N$ and hence that the 1-variation of A has finite N th moment.

Concerning the integrability of the quadratic variation of M , let $w \in \{1, \dots, d\}$; then it is well known that (see, e.g., [36, Ch. II Theorem 1.33])

$$\langle (x^w \mathbf{1}_{|x| \leq 1}) * (\mu^X - \nu) \rangle_T = (x^w)^2 \mathbf{1}_{|x| \leq 1} * \nu_T,$$

where $\langle M \rangle$ denotes the dual predictable projection (or compensator) of $[M]$. Further, using that the compensated martingale $(x \mathbf{1}_{|x| \leq 1}) * (\mu^X - \nu)$ is orthogonal to continuous martingales, we have

$$\langle M^w \rangle_T = \int_0^T a^{ww}(t) \mathrm{d}t + \int_0^T \int_{|x| \leq 1} (x^w)^2 K_t(\mathrm{d}x) \mathrm{d}t < \infty.$$

Now let $q \in [1, \infty)$; then from Theorem 8.2.20 in [14], we have the following estimation

$$\mathbb{E}([M^w]_T^q) \leq c \mathbb{E} \left(\langle M^w \rangle_T^q + \sup_{0 \leq t \leq T} (\Delta M_t^w)^{2q} \right) \leq c (\langle M^w \rangle_T^q + 1) < \infty,$$

where $c > 0$ is a constant depending on q .

We have shown that $\mathbf{X} = (0, X, 0, \dots, 0) \in \mathcal{X}^{1,N}$, and it follows from Theorem 4.1 that the signature cumulant $\kappa_t = \log(\mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}))$ satisfies the functional equation (4.3). On the other hand, it follows from the condition in equation (6.8) that η in equation (6.10) is well defined. Now define $\tilde{\kappa} = (\tilde{\kappa}_t)_{0 \leq t \leq T}$ by the identity equation (6.9). Noting that $\tilde{\kappa}$ is deterministic and has absolutely continuous components, it is easy to see that $\tilde{\kappa}$ also satisfies the functional equation (4.3) for the semimartingale X . It thus follows that κ and $\tilde{\kappa}$ are identical. □

6.2.2. Markov jump diffusions

The generator of a general Markov jump diffusion is of the form

$$\mathcal{L}f(x) = \sum_i \mathbf{b}^i(x) \partial_i f(x) + \sum_{i,j} \mathbf{a}^{ij}(x) \partial_i \partial_j f(x) + \int_{\mathbb{R}^d} \left(f(x+y) - f(x) - \mathbf{1}_{|y| \leq 1} \sum_i y^i \partial_i f(x) \right) K(x, dy), \tag{6.11}$$

where the summations are over $i, j \in \{1, \dots, d\}$, $\mathbf{b}: \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\mathbf{a}: \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ (symmetric, positive definite) and K is a Borel transition kernel from \mathbb{R}^d into \mathbb{R}^d with $K(\cdot, \{0\}) \equiv 0$. Denote by $\text{Lip}(E)$ the space of real valued, bounded and globally Lipschitz continuous functions over a Banach space E . We pose the following further assumptions on the generator:

- (A1) $\mathbf{b}^i, \mathbf{a}^{ij} \in \text{Lip}(\mathbb{R}^d)$ for all $i, j = 1, \dots, d$,
- (A2) There exists a constant $c > 0$ such that it holds

$$\sum_{ij} \mathbf{a}^{ij}(x) \xi^i \xi^j > c |\xi|^2, \quad \text{for all } x, \xi \in \mathbb{R}^d.$$

- (A3) There exists a bounded measure m on \mathbb{R}^d and a measurable function $\delta: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, such that the kernel K is of the form

$$K(x, A) = \int_{\mathbb{R}^d} \mathbf{1}_{A \setminus \{0\}}(\delta(x, z)) m(dz), \quad \text{for all measurable } A \subset \mathbb{R}^d \text{ and } x \in \mathbb{R}^d.$$

Further, we assume that there exists a function $\rho: \mathbb{R}^d \rightarrow \mathbb{R}_+$, such that

$$\int_{\mathbb{R}^d} (\rho(z)^2 \vee \rho(z)^n) m(dz) < \infty, \quad \text{for all } n \in \mathbb{N},$$

and for all $x, x', z \in \mathbb{R}^d$, it holds that

$$\begin{aligned} |\delta(x, z)| &\leq \rho(z), \\ |h \circ \delta(x, z) - h \circ \delta(x', z)| &\leq \rho(z) |x - x'|, \\ |h' \circ \delta(x, z) - h' \circ \delta(x', z)| &\leq \rho(z)^2 |x - x'|, \end{aligned}$$

where $h(y) = y \mathbf{1}_{|y| \leq 1}$ and $h'(y) = y - h(y)$.

Note that under the above assumptions,⁹ the martingale problem associated to the generator \mathcal{L} admits a unique solution (see [35, Theorem 13.58]). More precisely, for any probability measure η on \mathbb{R}^d ,

⁹In fact, as seen in [35, Theorem 13.58], the existence of a unique solution to the martingale problem holds also under fairly weaker assumptions than (A1)–(A3), where Lipschitz continuity can essentially be relaxed to continuity and the moment

there exists a unique measure \mathbb{P} on the canonical Skorohod space $(\Omega, \mathcal{F}, (\mathcal{F}_t))$, such that the canonical process X satisfies $\mathbb{P}(X_0 \in A) = \eta(A)$ for all measurable $A \subset \mathbb{R}^d$, and the process

$$M_t := f(X_t) - f(X_0) - \int_0^t \mathcal{L}f(X_s)ds \quad 0 \leq t \leq T,$$

is a local martingale under \mathbb{P} for all $f \in C_b^2(\mathbb{R}^d)$, the space of real-valued, bounded and twice continuously differentiable functions, with bounded first and second-order partial derivatives. In general, we may call a semimartingale a Markov jump diffusion associated to the generator \mathcal{L} if its law coincides with \mathbb{P} : that is, it solves the martingale problem associated to \mathcal{L} . The martingale problem can be equivalently formulated in terms of semimartingale characteristics (see [35, XIII.3] and [36, III.2.c]). In particular, under the given assumptions, it holds that \mathbb{P} is a solution to the martingale problem if and only if the canonical processes has the following semimartingale characteristics (see [35, Theorem 13.58])

$$d\mathbf{B}_t = \mathbf{b}(X_{t-})dt, \quad d\mathbf{C}_t = \mathbf{a}(X_{t-})dt, \quad \nu(dt, dx) = dtK(X_{t-}, dx).$$

The extensions to Markov processes with characteristics of the form $(\mathbf{b}(t, x), \mathbf{a}(t, x), K(t, x, dy))$ with associated local Lévy generators [60] is mostly notational. Apart from already presented references, the reader may also consult the abundant literature in the continuous case (e.g., [61], [59, VII.2] or [38, 5.4]).

The expected signature of a Markov jump diffusion X was seen in [25] to satisfy a system of linear partial integro-differential equations (PIDEs); the continuous case was already presented [54] with a corresponding system of PDEs. The passage to signature cumulants amounts to taking the logarithm, which represents a non-commutative Cole–Hopf transform with resulting quadratic non-linearity, if viewed as a \mathcal{T}_1 -valued PIDE, resolved - thanks to the graded structure - into a system of linear PIDEs. In Corollary 6.7 below we will show how this PIDE can be derived directly from our Theorem 4.1.

Since we are mainly interested in exposing the algebraic structure of this PIDE system and we want to avoid dwelling much further on the solution theory of PIDEs associated to the operator \mathcal{L} , we pose one further assumption:

(A4) For any $f \in \text{Lip}([0, T] \times \mathbb{R}^d)$ the Cauchy problem with zero terminal condition

$$\begin{aligned} [\partial_t + \mathcal{L}]u(t, x) &= f(t, x), & (t, x) \in [0, T] \times \mathbb{R}^d, \\ u(T, x) &= 0, & x \in \mathbb{R}^d, \end{aligned} \tag{6.12}$$

has a unique solution u in the space¹⁰

$$C_b^{1,2}([0, T] \times \mathbb{R}^d) := \{u \in C^{1,2}([0, T] \times \mathbb{R}^d) : u, \partial_{x_1}u, \dots, \partial_{x_d}u \in \text{Lip}([0, T] \times \mathbb{R}^d)\},$$

where $C^{1,2}([0, T] \times \mathbb{R}^d)$ is the space of real valued functions on $[0, T] \times \mathbb{R}^d$ that are once continuously differentiable in t and twice continuously differentiable in x .

A result similar to the above assumption (A4) based on conditions similar to (A1)-(A3) can be found [56, Proposition 5.3]. The main difference is that in the latter result only a linear growth condition on the forcing f is assumed and hence only a growth property of the solution is obtained. Note that the proof in [56] is based on a viscosity solution approach, which then allows to recast the PIDE in terms of a PDE with modified drift and forcing terms. Hence the bounds on the solution of the Cauchy problem and its derivatives should follow as a consequence of the stronger assumptions on the forcing and classical estimates as they can be found in [22, Sec. 9.4].

We now present the main result of this section.

conditions can be replaced to the integrability of the kernel against $1 \wedge |y|^2$. Note, however, that the representation of the kernel as in assumption (A3), with the corresponding Lipschitz conditions on δ , is standard when constructing a jump diffusion as a strong solution to an SDE driven by a Wiener processes and a Poisson random measure (see, e.g., [36, XIII.3]).

¹⁰Note that there is not much difference in defining these spaces with half-open time intervals $[0, T)$, as the global Lipschitz assumptions always allow for a continuous extension.

Corollary 6.7. Assume that \mathcal{L} in equation (6.11) satisfies (A1)–(A4), and let X be a d -dimensional Markov jump diffusion with this generator. Then $\mathbf{X} = (0, X, 0, \dots) \in \mathcal{H}^{\infty-}$, and the signature cumulant is of the form

$$\kappa_t(T) = \mathbf{v}(t, X_t; T) = \mathbf{v}(t, X_t),$$

where $\mathbf{v} = \sum_w \mathbf{v}^w e_w$ is the unique solution with $\mathbf{v}^w \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ for all $w \in \mathcal{W}_d$ of the following partial integro-differential equation

$$\begin{aligned} -[\partial_t + \mathcal{L}]\mathbf{v} &= H(\text{ad } \mathbf{v}) \left\{ \mathbf{b} + \frac{1}{2} \mathbf{a} + \frac{1}{2} \sum_{i,j} \mathbf{a}^{ij} Q(\text{ad } \mathbf{v})(\partial_i \mathbf{v} \otimes \partial_j \mathbf{v}) + \sum_{i,j} \mathbf{a}^{ij} e_j G(\text{ad } \mathbf{v}) \partial_i \mathbf{v} \right\} \\ &+ \int_{\mathbb{R}^d} \left\{ H(\text{ad } \mathbf{v}) \left(\exp(\mathbf{y}) \exp(\mathbf{v} \circ \tau_y) \exp(-\mathbf{v}) - 1 - \mathbf{y} \mathbf{1}_{|y| \leq 1} \right) - (\mathbf{v} \circ \tau_y - \mathbf{v}) \right\} K(\cdot, dy), \end{aligned} \tag{6.13}$$

on $[0, T] \times \mathbb{R}^d$ with terminal condition $\mathbf{v}(T, \cdot) \equiv 0$, where $\mathbf{y} := (0, y, 0, \dots) \in \mathcal{T}_0$ and $\tau_y(t, x) = (t, x + y)$.

Remark 6.8. It is instructive to look at the one-dimensional case. As seen several times before, all adjoint operators vanish due to commutativity in this case. With the more classical notation $\partial_x = \partial_1 \partial_{xx} = \partial_{11}$, $\varepsilon = e_1$, $\mathbf{b} = b\varepsilon$, $\mathbf{a} = a\varepsilon^2$ and with $\mathbf{v}_\varepsilon(t, x) := \varepsilon x + \mathbf{v}(t, x)$, we see that the PIDE (6.13) then simplifies to

$$-[\partial_t + \mathcal{L}]\mathbf{v}_\varepsilon = \frac{1}{2} a (\partial_x \mathbf{v}_\varepsilon)^2 + \int_{\mathbb{R}} \left\{ \exp(\mathbf{v}_\varepsilon \circ \tau_y - \mathbf{v}_\varepsilon) - 1 - (\mathbf{v}_\varepsilon \circ \tau_y - \mathbf{v}_\varepsilon) \right\} K(\cdot, dy). \tag{6.14}$$

Introducing the Cole-Hopf transformation $\mathbf{u}_\varepsilon(t, x) = \exp(\mathbf{v}_\varepsilon(t, x))$, multiplying equation (6.14) with \mathbf{u}_ε and considering the differential relations

$$\partial_t \mathbf{u}_\varepsilon = \mathbf{u}_\varepsilon \partial_t \mathbf{v}_\varepsilon, \quad \partial_x \mathbf{u}_\varepsilon = \mathbf{u}_\varepsilon \partial_x \mathbf{v}_\varepsilon, \quad \partial_{xx} \mathbf{u}_\varepsilon = \mathbf{u}_\varepsilon (\partial_x \mathbf{v}_\varepsilon)^2 + \mathbf{u}_\varepsilon \partial_{xx} \mathbf{v}_\varepsilon,$$

we obtain the following PIDE for \mathbf{u}_ε :

$$-\mathbf{u}_\varepsilon (\mathcal{L} \mathbf{u}_\varepsilon) = 0 \quad \Leftrightarrow \quad \mathcal{L} \mathbf{u}_\varepsilon = 0,$$

where the equivalence follows from the invertability of \mathbf{u}_ε in $\mathbb{R}[[\varepsilon]] \cong T(\mathbb{R})$. Conversely, the signature simplifies in the one-dimensional case to the exponential of the path increment, and therefore we have

$$\mathbf{u}_\varepsilon(t, X_t) = \exp(\mathbf{v}_\varepsilon(t, X_t)) = \exp(\varepsilon X_t) \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}) = \mathbb{E}_t(\exp(\varepsilon X_T)).$$

Hence, the above PIDE for \mathbf{u}_ε is already expected from the Feynman-Kac representation (see, e.g., [15, Proposition 12.5]). The expansion $\mathbf{v}_\varepsilon = \varepsilon v^1 + \varepsilon^2 v^2 + \dots$, with $\mathbf{v}_\varepsilon^{(i)} = \varepsilon^i v^i$, is related to the *Wild expansion* used in Hairer’s approach for solving the KPZ equation [30]. Indeed, as explained in [23, Section 4.5] for the continuous case, each term v^i can be composed as a sum of solutions to linear PDEs, indexed by binary trees with i leaves. The jump case follows similarly; however, the root joining operation corresponds to the creation of a different forcing term, which apart from the carré du champ (compare with [23, Remark 4.1]), includes terms derived from the integral in equation (6.14) (the combinatorial relations from breaking apart the exponential are analogous to those in Corollary 5.5).

Remark 6.9. We are still in the one-dimensional case and use notation from the previous remark. Assume now that the underlying Markov jump diffusion X takes values in the domain $D = \mathbb{R}$ or $D = \mathbb{R}_{\geq 0}$ and has affine differential characteristics of the form

$$b(x) = \beta_0 + \beta_1 x, \quad a(x) = \alpha_0 + \alpha_1 x, \quad K(x, dy) = (\lambda_0 + \lambda_1 x) m(dy),$$

for all $x \in D$ with $\beta_0 \in D$, $\beta_1 \in \mathbb{R}$, $\alpha_0, \alpha_1, \lambda_0, \lambda_1 \geq 0$, and m is a suitably integrable measure with $\text{supp}(m) \subset D$, where $\alpha_0 = 0$ if $D = \mathbb{R}_{\geq 0}$ and $\lambda_1 = \alpha_1 = 0$ if $D = \mathbb{R}$. In this case, \mathbf{b} and \mathbf{a} clearly do not satisfy the conditions (A1) and (A2) in general. Nevertheless, we can start by considering the PIDE (6.14) and use the following affine ansatz

$$\mathbf{v}_\varepsilon(t, x) = \Phi(T - t) + \Psi(T - t)x, \quad \Phi = \sum_{i=1}^\infty \phi^i \varepsilon^i, \quad \Psi = \sum_{i=1}^\infty \psi^i \varepsilon^i$$

for differentiable functions $\phi^i, \psi^i : [0, T] \rightarrow \mathbb{R}$ with $\psi^1(0) = 1$ and $\phi^i(0) = \psi^{i+1}(0) = 0$ for all $i \in \mathbb{N}_{\geq 1}$, which, after equating coefficients in x , yields the following formal Riccati ODE

$$\begin{aligned} \dot{\Psi} &= \beta_1 \Psi + \frac{1}{2} \alpha_1 (\Psi)^2 + \lambda_1 \int_{\mathbb{R}} \exp(\Psi y) m(dy) \\ \dot{\Phi} &= \beta_0 \Psi + \frac{1}{2} \alpha_0 (\Psi)^2 + \lambda_0 \int_{\mathbb{R}} \exp(\Psi y) m(dy). \end{aligned}$$

Equating coefficients in ε this formal quadratic ODE turns into a system of linear ODEs

$$\begin{aligned} \dot{\psi}_i &= \beta_1 \psi_i + \frac{1}{2} \alpha_1 \sum_{k=1}^{i-1} \psi_k \psi_{i-k} + \lambda_1 \sum_{\|\ell\|=i} \sum_{|\ell|=k} \sum_{k=2}^i \frac{m_k}{k!} \psi_{l_1} \cdots \psi_{l_k}, \\ \dot{\phi}_i &= \beta_0 \psi_i + \frac{1}{2} \alpha_0 \sum_{k=1}^{i-1} \psi_k \psi_{i-k} + \lambda_0 \sum_{\|\ell\|=i} \sum_{|\ell|=k} \sum_{k=2}^i \frac{m_k}{k!} \psi_{l_1} \cdots \psi_{l_k}, \quad i = 1, 2, \dots, \end{aligned}$$

where m_k is the k th moment of m ; that can easily be solved explicitly. The resulting formulas relate to the cumulant formulas obtained in the more general setting of polynomial processes in [17]. We do not claim to provide any new results on the topic of affine processes, and therefore we also do not relate the above formal steps back to the cumulants of the process, which would require a more careful analysis. An interesting open question from this perspective is how to fully employ the affine structure in the non-commutative setting. In the next section, we present a result in this direction applied to affine Volterra processes.

Proof of Corollary 6.7. We can easily verify the boundedness of \mathbf{b} and \mathbf{a} , and the moment condition in assumption (A3) implies that $\mathbf{X} = (0, X, 0, \dots) \in \mathcal{H}^{\infty-}$ (compare also with the proof of Corollary 6.5). It then follows from Theorem 4.1 that $\kappa(T) = (\mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}))_{0 \leq t \leq T}$ is the unique solution to the functional equation (4.3).

Next we are going to discuss the existence and uniqueness of a solution to the PIDE. Therefore, note that projecting equation (6.13) to any tensor level $w \in \mathcal{W}_d$, we obtain a Cauchy problem of the form in equation (6.12), where the forcing term f is given by the projection of the right-hand side in equation (6.13) to the tensor component w . Therefore the existence of a unique solution $\mathbf{v}^w \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ follows by assumption (A4), provided we can show that the corresponding forcing term is in $\text{Lip}([0, T] \times \mathbb{R}^d)$. We prove the latter assertion inductively: projecting the right-hand side of equation (6.13) to the tensor component $i = \{1, \dots, d\}$, we obtain the following forcing term

$$x \mapsto \mathbf{b}^i(x) + \int_{|y|>1} y^i K(x, dy) = \mathbf{b}^i(x) + \int_{\mathbb{R}^d} (h' \circ \delta(x, z))^i m(dz).$$

Recall that $\mathbf{b}^i \in \text{Lip}([0, T] \times \mathbb{R}^d)$ by assumption (A1). Regarding the integral term, we see that boundedness and the global Lipschitz continuity is a consequence of assumption (A3). Hence it follows that the above forcing term is indeed in $\text{Lip}([0, T] \times \mathbb{R}^d)$, which closes the base of the induction.

As the induction claim, assume that the projection of the right-hand side of equation (6.13) to all tensor components $w' \in \mathcal{W}_d$ with $|w'| \leq n$ for some $n \in \mathbb{N}_{\geq 1}$ yields a function in $\text{Lip}([0, T] \times \mathbb{R}^d)$ and

hence, by assumption (A4), the corresponding Cauchy problem with generator \mathcal{L} has unique solutions $\mathbf{v}^{w'} \in C^{1,2}([0, T] \times \mathbb{R}^d)$.

For the induction step, let $w \in \mathcal{W}_d$ with $|w| = n + 1$ be arbitrary. Note that the right-hand side of equation (6.13) consists of two main terms, the second one of which is given by an integral. Projecting the first term to the tensor components yields a linear combination of (finite) products of the functions \mathbf{a}^{ij} , $\mathbf{v}^{w'}$ and $\partial_i \mathbf{v}^{w'}$ with $|w'| \leq n$, $i \in \{1, \dots, d\}$. By assumption (A1) and the induction hypothesis, each of these functions is in $\text{Lip}([0, T] \times \mathbb{R}^d)$ and therefore their products and linear combinations also are (clearly $\text{Lip}([0, T] \times \mathbb{R}^d)$ forms an algebra). Hence we are left with proving that the projection of the integral term in the right-hand side of equation (6.13) is also in $\text{Lip}([0, T] \times \mathbb{R}^d)$. Using a Taylor expansion and the derivatives of the exponential map from Lemma 7.5 (compare also with Lemma 7.9), we obtain the following identity:

$$\begin{aligned} & H(\text{ad } \mathbf{v}) \left(\exp(\mathbf{y}) \exp(\mathbf{v} \circ \tau_{\mathbf{y}}) \exp(-\mathbf{v}) - 1 - \mathbf{y} \mathbf{1}_{|y| \leq 1} \right) - (\mathbf{v} \circ \tau_{\mathbf{y}} - \mathbf{v}) \\ &= H(\text{ad } \mathbf{v}) \int_0^1 \exp(\theta \mathbf{y}) \left((\mathbf{y})^2 + \mathbf{y} \sum_i y^i G(\text{ad } \mathbf{v} \circ \tau_{\theta \mathbf{y}}) (\partial_i \mathbf{v} \circ \tau_{\theta \mathbf{y}}) \right. \\ &\quad \left. + \sum_{i,j} y^i y^j Q(\text{ad } \mathbf{v} \circ \tau_{\theta \mathbf{y}}) (\partial_i \mathbf{v} \circ \tau_{\theta \mathbf{y}} \otimes \partial_j \mathbf{v} \circ \tau_{\theta \mathbf{y}}) \right) \exp(\mathbf{v} \circ \tau_{\theta \mathbf{y}}) \exp(\mathbf{v}) (1 - \theta) d\theta. \end{aligned}$$

Hence, we see from the above right-hand side and the induction claim that the projection of the integral term in the forcing of equation (6.13) to the tensor component w is a linear combination of functions of the form

$$(t, x) \mapsto f(t, x) := \int_{\mathbb{R}^d} g(t, x) p(y) \left(\int_0^1 q(y\theta) h(t, x + \theta y) (1 - \theta) d\theta \right) K(x, dy),$$

where $g, h \in \text{Lip}([0, T] \times \mathbb{R}^d)$, and where $p(y)$ and $q(y)$ are homogeneous polynomials in (y^1, \dots, y^d) with $\text{deg}(p) = 2$, $0 \leq \text{deg}(q) \leq n - 1$ and $q \neq 0$. It follows from assumption (A3) that the above function and hence any linear combination of functions of the same form are in $\text{Lip}([0, T] \times \mathbb{R}^d)$. Indeed, let $(t, x), (t', x') \in [0, T] \times \mathbb{R}^d$; then we have

$$\begin{aligned} & |f(t, x) - f(t', x')| \\ & \leq \int_{\mathbb{R}^d} |g(t, x) - g(t', x')| p(y) q(y) \left| \left(\int_0^1 \theta^{\text{deg}(q)} |h(t, x + \theta y)| (1 - \theta) d\theta \right) K(x, dy) \right. \\ & \quad \left. + \int_{\mathbb{R}^d} |g(t', x')| p(y) q(y) \left| \left(\int_0^1 \theta^{\text{deg}(q)} |h(t, x + \theta y) - h(t', x' + \theta y)| (1 - \theta) d\theta \right) K(x, dy) \right. \right. \\ & \quad \left. \left. + \left| \int_{\mathbb{R}^d} g(t', x') p(y) q(y) \left(\int_0^1 \theta^{\text{deg}(q)} h(t', x' + \theta y) (1 - \theta) d\theta \right) (K(x, dy) - K(x', dy)) \right| \right| \\ & \leq c(|t - t'| + |x - x'|) \int_{\mathbb{R}^d} |p(y) q(y)| K(x, dy) + c' \int_{\mathbb{R}^d} |(pq)(\delta(x, z)) - (pq)(\delta(x', z))| m(dz), \end{aligned}$$

where $c, c' > 0$ are constants (which will change from line to line in what follows). Regarding the first integral term in the last line above, note that for all $y \in \mathbb{R}^d$, it holds that

$$|p(y) q(y)| \leq c(|y|^2 \vee |y|^{2+\text{deg}(q)}) \leq c(|y|^2 \vee |y|^{n+1}).$$

It then follows by assumption (A3) that for all $x \in \mathbb{R}^d$, we have

$$\begin{aligned} \int_{\mathbb{R}^d} |p(y)q(y)|K(x, dy) &\leq c \int_{\mathbb{R}^d} (|\delta(x, z)|^2 \vee |\delta(x, z)|^{n+1})m(dz) \\ &\leq c \int_{\mathbb{R}^d} (|\rho(z)|^2 \vee |\rho(z)|^{n+1})m(dz) < \infty. \end{aligned}$$

Regarding the second integral term, we further note that for all $y, y' \in \mathbb{R}^d$, it holds

$$|p(y)q(y) - p(y')q(y')| \leq c''|y - y'| \sup_{\substack{\theta \in [0,1] \\ z=(1-\theta)y+\theta y'}} \sum_i |\partial_{y_i}(pq)(z)| \leq c''|y - y'| \max_{z \in \{y, y'\}} (|z| \vee |z|^n).$$

Therefore, again by assumption (A3), we have

$$\begin{aligned} \int_{\mathbb{R}^d} |(pq)(\delta(x, z)) - (pq)(\delta(x', z))|m(dz) &\leq c' \int_{\mathbb{R}^d} |\delta(x, z) - \delta(x', z)|(|\rho(z)| \vee |\rho(z)|^n)m(dz) \\ &\leq c'|x - x'| \int_{\mathbb{R}^d} (|\rho(z)| \vee |\rho(z)|^2)(|\rho(z)| \vee |\rho(z)|^n)m(dz) \\ &= c'|x - x'| \int_{\mathbb{R}^d} (|\rho(z)|^2 \vee |\rho(z)|^{n+2})m(dz) \\ &\leq c''|x - x'|, \end{aligned}$$

where $c'' > 0$ is another constant. This finishes the proof of the assertion that $f \in \text{Lip}([0, T] \times \mathbb{R}^d)$ and hence also the proof of the induction step. We have thus concluded the proof about the existence and uniqueness of a solution to the PIDE.

Now define $\tilde{\kappa} \in \mathcal{S}(\mathcal{T}_0)$ by $\tilde{\kappa}_t := \mathbf{v}(t, X_t)$ for all $0 \leq t \leq T$, and note that $\tilde{\kappa}_{t-} = \mathbf{v}(t, X_{t-})$. We are going to show that $\tilde{\kappa}$ also satisfies the functional equation (4.3). Since X solves the martingale problem with generator \mathcal{L} and \mathbf{v} is sufficiently regular, an application of Itô’s formula yields that the processes

$$\mathbf{v}(t, X_t) - \mathbf{v}(0, X_0) - \int_0^t [\partial_t + \mathcal{L}]\mathbf{v}(u, X_{u-})du, \quad 0 \leq t \leq T,$$

is a local martingale: that is, in $\mathcal{M}_{\text{loc}}(\mathcal{T}_0)$. However, as $\mathbf{v}^w \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$, it follows that the above process is bounded and therefore also a true martingale: that is, in $\mathcal{M}(\mathcal{T}_0)$. Hence, using further the terminal condition $\mathbf{v}(T, \cdot) \equiv 0$, we obtain the identity

$$\tilde{\kappa}_t = -\mathbb{E}_t(\mathbf{v}(T, X_T) - \mathbf{v}(t, X_t)) = \mathbb{E}_t\left(-\int_t^T [\partial_t + \mathcal{L}]\mathbf{v}(u, X_{u-})du\right). \tag{6.15}$$

On the other hand, we can plug $\tilde{\kappa}$ into the right-hand side of equation (4.3). We then obtain for the first integral inside the conditional expectation

$$\int_{(0,t]} H(\text{ad } \tilde{\kappa}_{u-})(d\mathbf{X}_u) = \int_{(0,t]} H(\text{ad } \tilde{\kappa}_{u-})(d\mathbf{B}_u + d\mathbf{X}_u^c) + \mathbf{W} * (\mu^X - \nu)_t + \overline{\mathbf{W}} * \mu_t^X,$$

where μ^X denotes the random measure associated with the jumps of X and

$$\mathbf{W}_t(y) := H(\text{ad } \tilde{\kappa}_{t-})(h(y)), \quad \text{and} \quad \overline{\mathbf{W}}_t(y) := H(\text{ad } \tilde{\kappa}_{t-})(y - h(y)),$$

for all $0 \leq t \leq T$, $y \in \mathbb{R}^d$ and where $h(y) = \mathbf{1}_{|y| \leq 1}$ is the usual truncation function. Similarly, we have

$$\sum_{0 < u \leq t} \left\{ H(\text{ad } \tilde{\kappa}_{u-}) \left(\exp(\Delta \mathbf{X}_u) \exp(\tilde{\kappa}_u) \exp(-\tilde{\kappa}_{u-}) - 1 - \Delta \mathbf{X}_u \right) - \Delta \tilde{\kappa}_u \right\} = \mathbf{J} * \mu_t^X,$$

where $0 \leq t \leq T$ and for $y \in \mathbb{R}^d$ with $\mathbf{y} = (0, y, 0, \dots) \in \mathcal{T}_0$

$$\mathbf{J}_t(\mathbf{y}) := \left\{ H(\text{ad } \mathbf{v}) \left(\exp(\mathbf{y}) \exp(\mathbf{v} \circ \tau_y) \exp(-\mathbf{v}) - 1 - \mathbf{y} \right) - (\mathbf{v} \circ \tau_y - \mathbf{v}) \right\} (t, X_{t-}).$$

Finally, for the quadratic variation terms with respect to continuous parts, we have

$$\begin{aligned} \mathbf{U}_t &:= \int_0^t H(\text{ad } \tilde{\kappa}_{u-}) \left\{ d\langle \mathbf{X}^c \rangle_u + (\text{Id} \odot G(\text{ad } \tilde{\kappa}_{u-})) (d\llbracket \mathbf{X}^c, \tilde{\kappa}^c \rrbracket_u) + Q(\text{ad } \tilde{\kappa}_{u-}) (d\llbracket \tilde{\kappa}^c, \tilde{\kappa}^c \rrbracket_u) \right\} \\ &= \sum_{i,j} \int_0^t \mathbf{a}^{ij} H(\text{ad } \mathbf{v}) \left\{ e_{ij} + (\text{Id} \odot G(\text{ad } \mathbf{v})) (e_i \otimes \partial_j \mathbf{v}) + Q(\text{ad } \mathbf{v}) (\partial_i \mathbf{v} \otimes \partial_j \mathbf{v}) \right\} (t, X_{u-}) du \\ &=: \int_0^t H(\text{ad } \mathbf{v}) (\mathbf{u}(u, X_{u-})) du. \end{aligned}$$

Provided that we can show the following integrability property holds for all words $w \in \mathcal{W}_d$

$$\mathbb{E} \left(\int_0^T |\{H(\text{ad } \tilde{\kappa}_{u-})(\mathbf{b}(X_{u-}) + \mathbf{u}(u, X_{u-}))\}^w| du + (|\mathbf{W}^w|^2 + |\overline{\mathbf{W}}^w| + |\mathbf{J}^w|) * \nu_T \right) < \infty, \tag{6.16}$$

it follows that

$$\begin{aligned} &\mathbb{E}_t \left\{ \int_{(t,T]} H(\text{ad } \tilde{\kappa}_{u-}) (d\mathbf{X}_u) + \mathbf{U}_{t,T} + \mathbf{J} * \mu_{t,T}^X \right\} \\ &= \mathbb{E}_t \left\{ \int_t^T H(\text{ad } \tilde{\kappa}_{u-}) (d\mathbf{B}_u) + \mathbf{U}_{t,T} + (\mathbf{J} + \overline{\mathbf{W}}) * \nu_{t,T} \right\} \\ &= \mathbb{E}_t \left\{ \int_{(t,T]} \left(H(\text{ad } \mathbf{v}) (\mathbf{b}(X_{u-}) + \mathbf{u}(u, X_{u-})) + \int_{\mathbb{R}^d} (\mathbf{J}_u(y) + \overline{\mathbf{W}}_u(y)) K(X_{u-}, dy) \right) du \right\} \\ &= \mathbb{E}_t \left(- \int_t^T [\partial_t + \mathcal{L}] \mathbf{v}(u, X_{u-}) du \right), \end{aligned}$$

where in the last line we have used \mathbf{v} , which satisfies the PIDE. Since the above left-hand side is precisely the right-hand side of the functional equation (4.3), it follows together with equation (6.15) that $\tilde{\kappa}$ satisfies the functional equation (4.3).

Note that in case the integrability condition in equation (6.16) is satisfied for all words $w \in \mathcal{W}_d$ with $|w| \leq n$ for some length $n \in \mathbb{N}_{\geq 1}$, it follows that the above equality holds up to the projection with $\pi_{(0,n)}$. For words with $|w| = 1$, the condition in equation (6.16) is an immediate consequence of $\mathbf{X} \in \mathcal{H}^{\infty-}$. It then follows inductively by the same arguments as in the proof of Claim 7.12 that equation (6.16) is indeed satisfied for all words $w \in \mathcal{W}_d$.

Since $\kappa(T)$ is the unique solution to equation (4.3), it then follows that $\tilde{\kappa} \equiv \kappa(T)$. □

6.3. Affine Volterra processes

For $i = 1, 2$, let K^i be an integration kernel such that $K^i(t, \cdot) \in L^2([0, t])$ for all $0 \leq t \leq T$, and let V^i be the solution to the Volterra integral equation

$$V_t^i = V_0^i + \int_0^t K^i(t, s) \sqrt{V_s^i} dW_s^i, \quad 0 \leq t \leq T,$$

with $V_0^i > 0$, where W^1 and W^2 are uncorrelated standard Brownian motions which generate the filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$. Note that in general, V^i is not a semimartingale. In particular, this is not the case when K^i is a power-law kernel of the form $K(t, s) \sim (t - s)^{H-1/2}$ for some $H \in (0, 1/2)$, which is the prototype of a rough affine volatility model (see, e.g., [39]). However, a martingale $\xi^i(T)$ is naturally associated to V^i by

$$\xi_t^i(T) = \mathbb{E}_t(V_T^i), \quad 0 \leq t \leq T.$$

In the financial context, $\xi^i(T)$ is the central object of a forward variance model (see, e.g., [29]). It was seen in [23] that the iterated diamond products of $\xi^1(T)$ are of a particularly simple form and easily translated to a system of convolutional Riccati equations of the type studied in [1] for the cumulant generating function. We are interested in the signature cumulant of the two-dimensional martingale $X = (\xi^1(T), \xi^2(T))$.

Corollary 6.10. *It holds that $\mathbf{X} = (0, \xi^1(T)e_1 + \xi^2(T)e_2, 0, \dots) \in \mathcal{H}^{\infty-}$ and the signature cumulant $\kappa_t(T) = \log \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T})$ is the unique solution to the functional equation: for all $0 \leq t \leq T$*

$$\begin{aligned} \kappa_t(T) = -\mathbb{E}_t \left(\frac{1}{2} \sum_{i=1,2} \int_t^T H(\text{ad } \kappa_u)(e_{ii}) K^i(T, u)^2 V_u^i du + \frac{1}{2} \int_t^T H(\text{ad } \kappa_u) \circ Q(\text{ad } \kappa_u)(d(\kappa \diamond \kappa)_u(T)) \right. \\ \left. + \sum_{i=1,2} \int_t^T H(\text{ad } \kappa_u) \{ e_i G(\text{ad } \kappa_u)(d(\xi^i(T) \diamond \kappa)_u(T)) \} \right). \end{aligned}$$

Proof. Regarding the integrability statement, it suffices to check that V_T^i has moments of all order for $i = 1, 2$. This is indeed the case, and we refer to [1, Lemma 3.1] for a proof. Hence we can apply Theorem 4.1, and we see that κ satisfies the functional equation (4.3). As described in Section 4.1, this equation can be reformulated with brackets replaced by diamonds. Further note that, due to the continuity, jump terms vanish and, due to the martingality, the Itô integrals with respect to \mathbf{X} have zero expectation. The final step to arrive at the above form of the functional equation is to calculate the brackets $\langle \xi^i(T), \xi^j(T) \rangle$. From the definition $\xi^i(T)$ and V^i , we have for all $0 \leq t \leq T$

$$\xi_t^i(T) = \mathbb{E}_t \left(V_0^i + \int_0^t K^i(T, s) \sqrt{V_s^i} dW_s^i + \int_t^T K^i(t, s) \sqrt{V_s^i} dW_s^i \right) = V_0^i + \int_0^t K^i(T, s) \sqrt{V_s^i} dW_s^i.$$

Therefore, due to the independence, we have $\langle \xi^1(T), \xi^2(T) \rangle = 0$; and for the square bracket, we have $d\langle \xi^i(T), \xi^i(T) \rangle_t = K^i(T, t)^2 V_t^i dt$. □

The recursion for the signature cumulants from Corollary 4.3 is easily simplified in analogy to the above corollary. In the rest of this section, we are going to demonstrate explicit calculations for the first four levels. Clearly, due to the martingality, the first-level signature cumulants are identically zero

$\kappa^{(1)}(T) \equiv 0$. In the second level, we start to observe the type of simplifications that appear due to the affine structure

$$\begin{aligned} \kappa_t^{(2)}(T) &= \frac{1}{2} \sum_{i=1,2} e_{ii}(\xi^i(T) \diamond \xi^i(T))_t(T) = \frac{1}{2} \sum_{i=1,2} e_{ii} \mathbb{E}_t \left(\int_t^T K^i(T, u)^2 V_u^i du \right) \\ &= \frac{1}{2} \sum_{i=1,2} e_{ii} \int_t^T K^i(T, u)^2 \xi_t^i(u) du, \end{aligned}$$

where $\xi_t^i(u) = \mathbb{E}_t(V_u^i)$ for all $0 \leq t \leq u \leq T$. The third level is of the same form

$$\begin{aligned} \kappa_t^{(3)}(T) &= \frac{1}{2} \sum_{i=1,2} e_i(\xi^i(T) \diamond \kappa^{(2)}(T))_t(T) \\ &= \frac{1}{2} \sum_{i=1,2} e_{iii} \int_t^T \left(\int_u^T K^i(T, s)^2 K^i(T, u) K^i(s, u) ds \right) \xi_t^i(u) du, \end{aligned}$$

where we have used that for any suitable $h : [0, T] \rightarrow \mathbb{R}$, it holds that for all $0 \leq t \leq T$

$$\int_t^T h(u) \xi_t^i(u) du = \int_0^T h(u) V_0^i du - \int_0^t h(u) V_u^i du + \int_0^t \left(\int_u^T h(s) K^i(s, u) ds \right) \sqrt{V_u^i} dW_u^i.$$

The fourth level starts to reveal some of the structure that is not visible in the commutative setting

$$\kappa_t^{(4)}(T) = \sum_{i=1,2} \left\{ \frac{1}{8} [e_{\bar{i}\bar{i}}, e_{ii}] \int_t^T \left(\int_u^T K^{\bar{i}}(T, s) \xi_t^{\bar{i}}(s) ds \right) K^i(T, u)^2 \xi_t^i(u) du + e_{iiii} \int_t^T h^i(T, u) \xi_t^i(u) du \right\},$$

where $\{i, \bar{i}\} = \{1, 2\}$ and h^i is defined by

$$\begin{aligned} h^i(T, u) &= \frac{1}{8} \left(\int_u^T K^i(T, s)^2 K^i(u, s) ds \right)^2 \\ &\quad + \frac{1}{2} \int_u^T \left(\int_s^T K^i(T, r)^2 K^i(T, s) K^i(r, s) dr \right) K^i(T, s) K^i(s, u) ds, \quad 0 \leq u \leq T. \end{aligned}$$

7. Proofs

For ease of notation, we introduce a norm on the space of tensor valued finite variation process, which could have been introduced in Section 2.4 but was not needed until now. Let $q \in [1, \infty)$ and $\mathbb{A} \in \mathcal{V}((\mathbb{R}^d)^{\otimes n})$ for some $n \in \mathbb{N}_{\geq 1}$; then we define

$$\|\mathbb{A}\|_{\mathcal{V}^q} := \|\mathbb{A}\|_{\mathcal{V}^q((\mathbb{R}^d)^{\otimes n})} := \|\|\mathbb{A}\|_{1\text{-var}; [0, T]}\|_{\mathcal{L}^q}.$$

It is easy to see that it holds $\|\mathbb{A}\|_{\mathcal{V}^q} \leq \|\mathbb{A}\|_{\mathcal{V}^q}$, and this inequality can be strict.

Further, for an element $\mathbb{A} \in T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d)$, we introduce the following notation

$$\mathbb{A} = \sum_{w_1, w_2 \in \mathcal{W}_d} \mathbb{A}^{w_1, w_2} e_{w_1} \otimes e_{w_2}, \quad \mathbb{A}^{w_1, w_2} \in \mathbb{R},$$

and for $l_1, l_2 \in \mathbb{N}_{\geq 1}$

$$\mathbb{A}^{(l_1, l_2)} = \sum_{|w_1|=l_1, |w_2|=l_2} e_{w_1 w_2} \otimes \mathbb{A}^{w_1, w_2} \in (\mathbb{R}^d)^{\otimes l_1} \otimes (\mathbb{R}^d)^{\otimes l_2} \subset T(\mathbb{R}^d) \bar{\otimes} T(\mathbb{R}^d).$$

Next we will prove two well-known lemmas translated to the setting of tensor valued semimartingales.

Lemma 7.1 (Kunita-Watanabe inequality). *Let $\mathbf{X} \in \mathcal{S}((\mathbb{R}^d)^{\otimes n})$ and $\mathbf{Y} \in \mathcal{S}((\mathbb{R}^d)^{\otimes m})$; then the following estimate holds a.s.*

$$|\langle \mathbf{X}^c, \mathbf{Y}^c \rangle|_{1\text{-var};[0,T]} + \sum_{0 < t \leq T} |\Delta \mathbf{X}_t \Delta \mathbf{Y}_t| \leq \sum_{|w_1|=n} \sqrt{[\mathbf{X}^{w_1}]_T} \sum_{|w_2|=m} \sqrt{[\mathbf{Y}^{w_2}]_T} \leq c \sqrt{[\mathbf{X}]_T} \sqrt{[\mathbf{Y}]_T},$$

where $c > 0$ is a constant that only depends on d, m and n .

Proof. From the definition of the quadratic variation of tensor valued semimartingales in Section 2.4, we have

$$\begin{aligned} |\langle \mathbf{X}^c, \mathbf{Y}^c \rangle|_{1\text{-var};[0,T]} + \sum_{0 < s \leq T} |\Delta \mathbf{X}_s \Delta \mathbf{Y}_s| &\leq \sum_{|w_1|=n, |w_2|=m} \int_0^T |d\langle \mathbf{X}^{w_1 c}, \mathbf{Y}^{w_2 c} \rangle_s| + \sum_{0 < s \leq T} |\Delta \mathbf{X}_s^{w_1} \Delta \mathbf{Y}_s^{w_2}| \\ &\leq \sum_{|w_1|=n, |w_2|=m} \sqrt{[\mathbf{X}^{w_1}]_T} \sqrt{[\mathbf{Y}^{w_2}]_T} \\ &\leq d^{(n+m)/2} \sqrt{\sum_{|w_1|=n} [\mathbf{X}^{w_1}]_T} \sqrt{\sum_{|w_2|=m} [\mathbf{Y}^{w_2}]_T} \\ &\leq d^{n+m} |\mathbf{X}]_T ||\mathbf{Y}]_T|, \end{aligned}$$

where the first estimate follows from the triangle inequality, the second estimate from the (scalar) Kunita-Watanabe inequality [57, Ch. II, Theorem 25] and the last two estimates follow from the standard estimate between the 1-norm and the 2-norm on $(\mathbb{R}^d)^{\otimes m} \cong \mathbb{R}^{d^m}$. \square

In order to prove the next well-known lemma (Emery’s inequality), we need the following technical lemma.

Lemma 7.2. *Let $\mathbf{A} \in \mathcal{V}((\mathbb{R}^d)^{\otimes n})$, $\mathbf{Y} \in \mathcal{D}((\mathbb{R}^d)^{\otimes l})$, $\mathbf{Z} \in \mathcal{D}((\mathbb{R}^d)^{\otimes m})$; then it holds that*

$$\left| \int_{(0, \cdot]} \mathbf{Y}_{s-} d\mathbf{A}_s \mathbf{Z}_{s-} \right|_{1\text{-var};[0,T]} \leq \int_{(0,T]} |\mathbf{Y}_{s-} \mathbf{Z}_{s-}| |d\mathbf{A}_s|,$$

where the integration with respect to $|d\mathbf{A}|$ denotes the integration with respect to the increasing one-dimensional path $(|\mathbf{A}|_{1\text{-var};[0,t]})_{0 \leq t \leq T}$. Further, let $\mathbf{Y}' \in \mathcal{D}((\mathbb{R}^d)^{\otimes l'})$ and $\mathbf{Z}' \in \mathcal{D}((\mathbb{R}^d)^{\otimes m'})$, and let $(\mathbf{A}_t)_{0 \leq t \leq T}$ be a process taking values in $(\mathbb{R}^d)^{\otimes n} \otimes (\mathbb{R}^d)^{\otimes n'}$ such that $\mathbf{A}^{w_1, w_2} \in \mathcal{V}$ for all $w_1, w_2 \in \mathcal{W}_d$ with $|w_1| = n$ and $|w_2| = n'$. Then it holds that

$$\left| \int_{(0, \cdot]} (\mathbf{Y}_{s-} \text{Id} \mathbf{Y}'_{s-}) \odot (\mathbf{Z}_{s-} \text{Id} \mathbf{Z}'_{s-}) (d\mathbf{A}_s) \right|_{1\text{-var};[0,T]} \leq \int_{(0,T]} |\mathbf{Y} \mathbf{Y}' \mathbf{Z} \mathbf{Z}'|_{s-} |dm(\mathbf{A})_s|,$$

where $(\mathbf{Y} \text{Id} \mathbf{Y}')(\mathbf{A}) = \mathbf{Y} \mathbf{A} \mathbf{Y}'$ is the left- respectively right-multiplication by \mathbf{Y} , respectively \mathbf{Y}' .

Proof. Let $0 \leq s \leq t \leq T$; then it holds that

$$\left| \int_{(s,t]} \mathbf{Y}_u d\mathbf{A}_u \mathbf{Z}_u \right| \leq \int_{(s,t]} |\mathbf{Y}_u \mathbf{Z}_u| |d\mathbf{A}_u|.$$

Indeed, as follows, for example, from [62, Theorem on Stieltjes integrability], we can approximate the integral in the left-hand side by Riemann sums. Then for a partition $(t_i)_{i=1, \dots, k}$ of the interval $[s, t]$, we have

$$\left| \sum_{i=1}^{k-1} \mathbf{Y}_{t_i-} (\mathbf{A}_{t_{i+1}} - \mathbf{A}_{t_i}) \mathbf{Z}_{t_i-} \right| \leq \sum_{i=1}^{k-1} |\mathbf{Y}_{t_i-} (\mathbf{A}_{t_{i+1}} - \mathbf{A}_{t_i}) \mathbf{Z}_{t_i-}| \leq \sum_{i=1}^{k-1} |\mathbf{Y}_{t_i-} \mathbf{Z}_{t_i-}| |\mathbf{A}_{t_{i+1}} - \mathbf{A}_{t_i}|,$$

where the last inequality follows from the fact that for homogeneous tensors $\mathbf{x} \in (\mathbb{R}^d)^{\otimes m}$ and $\mathbf{y} \in (\mathbb{R}^d)^{\otimes n}$, it holds that $|\mathbf{xy}| = |\mathbf{yx}| \leq |\mathbf{x}||\mathbf{y}|$. Regarding the 1-variation, we then have

$$\begin{aligned} \left| \int_{(0, \cdot]} \mathbf{Y}_{s-} d\mathbf{A}_s \mathbf{Z}_{s-} \right|_{1\text{-var}; [0, T]} &= \sup_{0 \leq t_1 \leq \dots \leq t_k \leq T} \sum_{i=1}^k \left| \int_{(t_i, t_{i+1}]} \mathbf{Y}_{s-} d\mathbf{A}_s \mathbf{Z}_{s-} \right| \\ &\leq \sup_{0 \leq t_1 \leq \dots \leq t_k \leq T} \sum_{i=1}^k \int_{(t_i, t_{i+1}]} |\mathbf{Y}_{s-} \mathbf{Z}_{s-}| |d\mathbf{A}_s| \\ &= \int_{(0, T]} |\mathbf{Y}_{s-} \mathbf{Z}_{s-}| |d\mathbf{A}_s|. \end{aligned}$$

Regarding the second statement, we see that for any $0 \leq s \leq t \leq T$, we have

$$\left| \int_{(s, t]} (\mathbf{Y}_{u-} \text{Id} \mathbf{Y}'_{u-}) \odot (\mathbf{Z}_{u-} \text{Id} \mathbf{Z}'_{u-}) (d\mathbb{A}_u) \right| \leq \int_{(s, t]} |\mathbf{Y} \mathbf{Y}' \mathbf{Z} \mathbf{Z}'|_{u-} |dm(\mathbb{A})_u|.$$

Indeed, we approximate the integral in the right-hand side again by a Riemann sum. Then for a partition $(t_i)_{i=1, \dots, k}$ of the interval $[s, t]$, we have

$$\begin{aligned} &\left| \sum_{i=1}^{k-1} (\mathbf{Y}_{t_i-} \text{Id} \mathbf{Y}'_{t_i-}) \odot (\mathbf{Z}_{t_i-} \text{Id} \mathbf{Z}'_{t_i-}) (\mathbb{A}_{t_{i+1}} - \mathbb{A}_{t_i}) \right| \\ &\leq \sum_{i=1}^{k-1} \left| \sum_{|w_1|=m, |w_2|=m'} \mathbf{Y}_{t_i-} e_{w_1} \mathbf{Y}'_{t_i-} \mathbf{Z}_{t_i-} e_{w_2} \mathbf{Z}'_{t_i-} (\mathbb{A}_{t_{i+1}} - \mathbb{A}_{t_i}) \right| \\ &\leq \sum_{i=1}^{k-1} |\mathbf{Y}_{t_i-} \mathbf{Y}'_{t_i-} \mathbf{Z}_{t_i-} \mathbf{Z}'_{t_i-}| |m(\mathbb{A}_{t_{i+1}}) - m(\mathbb{A}_{t_i})|, \end{aligned}$$

where the last inequality follows from the definition of the norm on (homogeneous) tensors and the definition of the multiplication map m . We conclude analogously to the proof of the first statement. \square

Lemma 7.3 (Emery’s inequality). *Let $\mathbf{X} \in \mathcal{S}((\mathbb{R}^d)^{\otimes n})$, $\mathbf{Y} \in \mathcal{D}((\mathbb{R}^d)^{\otimes l})$ and $\mathbf{Z} \in \mathcal{D}((\mathbb{R}^d)^{\otimes m})$; then for $p, q \in [1, \infty)$ and $1/r = 1/p + 1/q$, it holds that*

$$\left\| \int_{(0, \cdot]} \mathbf{Y}_{s-} d\mathbf{X}_s \mathbf{Z}_{s-} \right\|_{\mathcal{H}^r((\mathbb{R}^d)^{\otimes (l+n+m)}} \leq c \|\mathbf{Y} \mathbf{Z}\|_{\mathcal{S}^q((\mathbb{R}^d)^{\otimes (l+m)})} \|\mathbf{X}\|_{\mathcal{H}^p((\mathbb{R}^d)^{\otimes n})},$$

where $c > 0$ is a constant that only depends on d and m .

Proof. Let $\mathbf{X} = \mathbf{X}_0 + \mathbf{M} + \mathbf{A}$ be a semimartingale decomposition with $\mathbf{M}_0 = \mathbf{A}_0 = 0$. Then it follows by definition of the \mathcal{H}^r -norm and the above Lemma 7.2

$$\begin{aligned} \left\| \int_{(0, \cdot]} \mathbf{Y}_{s-} d\mathbf{X}_s \mathbf{Z}_{s-} \right\|_{\mathcal{H}^r} &\leq \left\| \int_{(0, T]} (\mathbf{Y}_{s-} \text{Id} \mathbf{Z}_{s-})^{\odot 2} d[\mathbf{M}, \mathbf{M}]_s \right\|^{1/2} + \left\| \int_{(0, \cdot]} \mathbf{Y}_{s-} d\mathbf{A}_s \mathbf{Z}_{s-} \right\|_{1\text{-var}; [0, T]} \Big\|_{\mathcal{L}^r} \\ &\leq \left\| \int_{(0, T]} |(\mathbf{Y}_{s-} \mathbf{Z}_{s-})^2| |d[\mathbf{M}]_s| \right\|^{1/2} + \left\| \int_{(0, T]} |\mathbf{Y}_{s-} \mathbf{Z}_{s-}| |d\mathbf{A}_s| \right\|_{\mathcal{L}^r} \\ &\leq \left\| \sup_{0 \leq s \leq T} |\mathbf{Y}_s \mathbf{Z}_s| \left(\|\mathbf{M}\|_{1\text{-var}; [0, T]}^{1/2} + \|\mathbf{A}\|_{1\text{-var}; [0, T]} \right) \right\|_{\mathcal{L}^r} \\ &\leq c \|\mathbf{Y} \mathbf{Z}\|_{\mathcal{S}^q} \left(\|\mathbf{M}\|_T + \|\mathbf{A}\|_{1\text{-var}; [0, T]} \right) \Big\|_{\mathcal{L}^p}, \end{aligned}$$

where we have used the generalized Hölder inequality and the Kunita-Watanabe inequality (Lemma 7.1) to get to the last line. Taking the infimum of over all semimartingale decomposition $\mathbf{M} + \mathbf{A}$ yields the statement. □

The following technical lemma will be used in the proof of both Theorem 3.2 and Theorem 4.1.

Lemma 7.4. *Let $\mathbf{X}, \mathbf{Y} \in \mathcal{S}(\mathcal{T}^N)$, $N \in \mathbb{N}_{\geq 1}$, $q \in [1, \infty)$, and assume that there exists a constant $c > 0$ such that*

$$\|\mathbf{Y}^{(n)}\|_{\mathcal{H}^{qN/n}} \leq c \sum_{\|\ell\|=n} \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}} \cdots \|\mathbf{X}^{(l_j)}\|_{\mathcal{H}^{qN/l_j}}, \quad n = 1, \dots, N,$$

where the summation is over $\ell = (l_1, \dots, l_j) \in (\mathbb{N}_{\geq 1})^j$, $j \in \mathbb{N}_{\geq 1}$, $\|\ell\| = l_1 + \dots + l_j$. Then there exists a constant $C > 0$, depending only on c and N , such that

$$\|\mathbf{Y}\|_{\mathcal{H}^{q,N}} \leq C \|\mathbf{X}\|_{\mathcal{H}^{q,N}}.$$

Proof. Note that for any $n \in \{1, \dots, N\}$, it holds that

$$\begin{aligned} \left(\sum_{\|\ell\|=n} \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}} \cdots \|\mathbf{X}^{(l_j)}\|_{\mathcal{H}^{qN/l_j}} \right)^{1/n} &\leq \sum_{\|\ell\|=n} \left(\|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}}^{1/l_1} \right)^{l_1/n} \cdots \left(\|\mathbf{X}^{(l_j)}\|_{\mathcal{H}^{qN/l_j}}^{1/l_j} \right)^{l_j/n} \\ &\leq \sum_{\|\ell\|=n} \left(\frac{l_1}{n} \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}}^{1/l_1} + \cdots + \frac{l_j}{n} \|\mathbf{X}^{(l_j)}\|_{\mathcal{H}^{qN/l_j}}^{1/l_j} \right) \\ &\leq c_n \sum_{i=1}^n \|\mathbf{X}^{(i)}\|_{\mathcal{H}^{qN/i}}^{1/i}, \end{aligned}$$

where $c_n > 0$ is a constant depending only on n , and the second inequality follows from Young’s inequality for products. Hence by the above estimate and the assumption we have

$$\|\mathbf{Y}\|_{\mathcal{H}^{q,N}} = \sum_{n=1}^N \|\mathbf{Y}^{(n)}\|_{\mathcal{H}^{qN/n}}^{1/n} \leq c^{1/n} c_n \sum_{n=1}^N \sum_{i=1}^n \|\mathbf{X}^{(i)}\|_{\mathcal{H}^{qN/i}}^{1/i} \leq C \|\mathbf{X}\|_{\mathcal{H}^{q,N}},$$

where $C > 0$ is a constant depending only on c and N . □

7.1. Proof of Theorem 3.2

Proof. Denote by $\mathbf{S} = (\text{Sig}(\mathbf{X})_{0,t})_{0 \leq t \leq T}$ the signature process. We will first prove the upper inequality: that is, that there exists a constant $C > 0$ depending only on d, N and q such that

$$\|\mathbf{S}\|_{\mathcal{H}^{q,N}} \leq C \|\mathbf{X}\|_{\mathcal{H}^{q,N}}. \tag{7.1}$$

According to Lemma 7.4, it is sufficient to show that for all $n \in \{1, \dots, N\}$, it holds that

$$c_n \|\mathbf{S}^{(n)}\|_{\mathcal{H}^{qN/n}} \leq \sum_{\|\ell\|=n} \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}} \cdots \|\mathbf{X}^{(l_j)}\|_{\mathcal{H}^{qN/l_j}} =: \rho_{\mathbf{X}}^n, \tag{7.2}$$

where $c_n > 0$ is a constant (depending only on q, d and n). In the inequality above (and the rest of the proof), the summation is over $\ell = (l_1, \dots, l_j) \in (\mathbb{N}_{\geq 1})^j, j \in \mathbb{N}_{\geq 1}$ with $\|\ell\| = l_1 + \dots + l_j = n$; in particular, the variable j is reserved for the length $|\ell|$ of the multiindex ℓ inside such summations. Note that $\|\ell\| = n$ implies that $l_i \leq n$ for all $i = 1, \dots, j$. It can easily be seen from the definition of the quantities $(\rho_{\mathbf{X}}^1, \dots, \rho_{\mathbf{X}}^N)$ in equation (7.2) that they satisfy the following (‘cascading’) property for all $n = 1, \dots, N$

$$\rho_{\mathbf{X}}^n \leq \sum_{\|\ell\|=n} \rho_{\mathbf{X}}^{l_j} \|\mathbf{X}^{(l_{j-1})}\|_{\mathcal{H}^{qN/l_{j-1}}} \cdots \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}} \leq c'' \sum_{\|\ell\|=n} \rho_{\mathbf{X}}^{l_j} \cdots \rho_{\mathbf{X}}^{l_1} \leq c' \rho_{\mathbf{X}}^n, \tag{7.3}$$

where c' and c'' are constants depending only on n . We are going to prove equation (7.2) inductively.

For $n = 1$, we have $\mathbf{S}^{(1)} = \mathbf{X}^{(1)} - \mathbf{X}_0^{(1)} = \mathbf{X}^{(1)} \in \mathcal{H}^{qN}$, and therefore the estimate follows immediately. Now, assume that equation (7.2) holds for all tensor levels up to some level $n - 1$ with $n \in \{2, \dots, N\}$. We will denote by $c', c'' > 0$ constants that only depend on n, d and q . Then we have from equation (2.11)

$$\begin{aligned} \mathbf{S}_t^{(n)} &= \sum_{\|\ell\|=n, |\ell| \leq 2} \int_0^t \mathbf{S}_{u-}^{(l_2)} d\mathbf{X}_u^{(l_1)} + \sum_{\|\ell\|=n, 2 \leq |\ell| \leq 3} \frac{1}{2} \int_0^t \mathbf{S}_{u-}^{(l_3)} d\langle \mathbf{X}^{(l_1)c}, \mathbf{X}^{(l_2)c} \rangle_u \\ &+ \sum_{\|\ell\|=n, |\ell| \geq 2} \sum_{0 \leq u \leq t} \mathbf{S}_{u-}^{(l_j)} \frac{\Delta \mathbf{X}_u^{(l_{j-1})} \cdots \Delta \mathbf{X}_u^{(l_1)}}{(j-1)!}. \end{aligned}$$

For the first term in the above right-hand side, we have by Emery’s inequality (Lemma 7.3) the following estimate

$$\left\| \sum_{\|\ell\|=n, |\ell| \leq 2} \int_0^t \mathbf{S}_{u-}^{(l_2)} d\mathbf{X}_u^{(l_1)} \right\|_{\mathcal{H}^{qN/n}} \leq c' \sum_{\|\ell\|=n, |\ell| \leq 2} \|\mathbf{S}^{(l_2)}\|_{\mathcal{S}^{qN/l_2}} \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_2}} \leq c'' \rho_{\mathbf{X}}^n,$$

where the last inequality follows from the induction claim and equation (7.3). Further, from the Kunita-Watanabe inequality (Lemma 7.1) and the generalized Hölder inequality, it follows that for all $l_1, l_2 \in \mathbb{N}_{\geq 1}$ with $l_1 + l_2 \leq n$, we have

$$\left\| \langle \mathbf{X}^{(l_2)c}, \mathbf{X}^{(l_1)c} \rangle \right\|_{\mathcal{H}^{qN/(l_1+l_2)}} \leq c' \|\mathbf{X}^{(l_2)}\|_{\mathcal{H}^{qN/l_2}} \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}}.$$

Then we have again by Emery’s inequality, the induction base and equation (7.3) that it holds

$$\left\| \sum_{\|\ell\|=n, 2 \leq |\ell| \leq 3} \int_0^t \mathbf{S}_{u-}^{(l_3)} d\langle \mathbf{X}^{(l_2)c}, \mathbf{X}^{(l_1)c} \rangle_u \right\|_{\mathcal{H}^{qN/n}} \leq c' \rho_{\mathbf{X}}^n.$$

Finally, we have for the summation term

$$\begin{aligned} & \left\| \sum_{\|\ell\|=n, |\ell|\geq 2} \sum_{0\leq u\leq t} \mathbf{S}_{u-}^{(l_j)} \frac{\Delta \mathbf{X}_u^{(l_2)} \cdots \Delta \mathbf{X}_u^{(l_k)}}{(j-1)!} \right\|_{\mathcal{H}^{qN/n}} \\ & \leq \sum_{\|\ell\|=n, |\ell|\geq 2} \left\| \sum_{0\leq u\leq t} \left| \mathbf{S}_{u-}^{(l_k)} \right| \left\| \Delta \mathbf{X}_u^{(l_{k-1})} \right\| \cdots \left\| \Delta \mathbf{X}_u^{(l_3)} \right\| \left\| \Delta \mathbf{X}_u^{(l_1)} \Delta \mathbf{X}_u^{(l_2)} \right\| \right\|_{\mathcal{L}^{qN/|w|}} \\ & \leq c' \sum_{\|\ell\|=n, |\ell|\geq 2} \left\| \mathbf{S}^{(l_k)} \right\|_{\mathcal{S}_\infty^{qN/l_1}} \left\| \mathbf{X}^{(l_{k-1})} \right\|_{\mathcal{S}_\infty^{qN/l_{k-1}}} \cdots \left\| \mathbf{X}^{(l_3)} \right\|_{\mathcal{S}_\infty^{qN/l_3}} \left\| \sum_{0\leq u\leq t} \left\| \Delta \mathbf{X}_u^{(l_1)} \Delta \mathbf{X}_u^{(l_2)} \right\| \right\|_{\mathcal{L}^{qN/(l_1+l_2)}} \\ & \leq c'' \rho_{\mathbf{X}}^n \end{aligned}$$

where the last inequality follows again by the Kunita-Watanabe inequality, the induction basis and equation (7.3). Thus we have shown that equation (7.2) holds for all $n \in \{1, \dots, N\}$.

Now we will prove the lower inequality: that is, that there exists a constant $c > 0$ depending only on d, N and q such that

$$c \|\mathbf{X}\|_{\mathcal{H}^{q,N}} \leq \|\mathbf{S}\|_{\mathcal{H}^{q,N}}. \tag{7.4}$$

Therefore define $\bar{\mathbf{X}}^n := (0, \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}, 0, \dots, 0) \in \mathcal{H}^{q,N}$, and note that it holds

$$\|\bar{\mathbf{X}}^1\|_{\mathcal{H}^{q,N}} = \|\mathbf{X}^{(1)}\|_{\mathcal{H}^{qN}} = \|\mathbf{S}^{(1)}\|_{\mathcal{H}^{qN}} \leq \|\mathbf{S}\|_{\mathcal{H}^{q,N}}.$$

Now assume that it holds

$$\|\bar{\mathbf{X}}^{n-1}\|_{\mathcal{H}^{q,N}} \leq c' \|\mathbf{S}\|_{\mathcal{H}^{q,N}}$$

for some $n \in \{1, \dots, N\}$. It follows from the definition of the signature that

$$\mathbf{S}_t^{(n)} = \mathbf{X}_{0,t}^{(n)} + \text{Sig}(\bar{\mathbf{X}}^{n-1})_{0,t}^{(n)}, \tag{7.5}$$

and further, we have from the upper bound in equation (7.1), which was already proven above, that

$$\|\text{Sig}(\bar{\mathbf{X}}^{n-1})_{0,\cdot}\|_{\mathcal{H}^{q,N}} \leq C \|\bar{\mathbf{X}}^{n-1}\|_{\mathcal{H}^{q,N}} \leq Cc' \|\mathbf{S}\|_{\mathcal{H}^{q,N}}. \tag{7.6}$$

Then we have

$$\begin{aligned} \|\bar{\mathbf{X}}^n\|_{\mathcal{H}^{q,N}} &= \|\bar{\mathbf{X}}^{n-1}\|_{\mathcal{H}^{q,N}} + \left\| \mathbf{X}^{(n)} \right\|_{\mathcal{H}^{qN/n}}^{1/n} \\ &\leq c' \|\mathbf{S}\|_{\mathcal{H}^{q,N}} + \left\| \mathbf{S}^{(n)} \right\|_{\mathcal{H}^{qN/n}}^{1/n} + \left\| \text{Sig}(\bar{\mathbf{X}}^{n-1})_{0,\cdot}^{(n)} \right\|_{\mathcal{H}^{qN/n}}^{1/n} \\ &\leq c' \|\mathbf{S}\|_{\mathcal{H}^{q,N}} + \|\mathbf{S}\|_{\mathcal{H}^{q,N}} + \|\text{Sig}(\bar{\mathbf{X}}^{n-1})_{0,\cdot}\|_{\mathcal{H}^{q,N}} \\ &\leq c'' \|\mathbf{S}\|_{\mathcal{H}^{q,N}}, \end{aligned}$$

where we have used equation (7.5) in the second line and equation (7.6) in the last line. Therefore, noting that $\bar{\mathbf{X}}^N = \mathbf{X}$, the inequality in equation (7.4) follows by induction. \square

7.2. Proof of Theorem 4.1

We prepare the proof of Theorem 4.1 with a few more lemmas. Recall from Section 2.3.2 that $\exp_N : \mathcal{T}_0^N \rightarrow \mathcal{T}_1^N$ denotes the exponential map in the truncated tensor algebra \mathcal{T}^N for some $N \in \mathbb{N}_{\geq 1}$, which is given by the power series

$$\begin{aligned} \exp_N(\mathbf{x}) &= 1 + \sum_{k=1}^N \frac{1}{k!} (\mathbf{x})^k \\ &= 1 + \sum_{k=1}^N \sum_{\|\ell\| \leq N, |\ell|=k} \frac{1}{k!} \mathbf{x}^{(\ell_1)} \dots \mathbf{x}^{(\ell_k)} \in \mathcal{T}_1^N, \quad \mathbf{x} \in \mathcal{T}_0, \end{aligned}$$

where all concatenation products are understood in the truncated tensor algebra \mathcal{T}_1^N , and the summation is over all $\ell = (\ell_1, \dots, \ell_k) \in (\mathbb{N}_{\geq 1})^k$, $k \in \mathbb{N}_{\geq 1}$ with $|\ell| = k$ and $\|\ell\| = \ell_1 + \dots + \ell_k$. Also recall that $\log_N : \mathcal{T}_1 \rightarrow \mathcal{T}_0$ denotes the logarithm in the truncated tensor algebra \mathcal{T}^N , which is analogously given by the corresponding log-power series.

Lemma 7.5. *Let $N \in \mathbb{N}_{\geq 1}$; then we have the following directional derivatives of the truncated exponential map $\exp_N : \mathcal{T}_0^N \rightarrow \mathcal{T}_1^N$*

$$\begin{aligned} (\partial_w \exp_N)(\mathbf{x}) &= G(\text{ad } \mathbf{x})(e_w) \exp_N(\mathbf{x}) = \exp_N(\mathbf{x}) G(-\text{ad } \mathbf{x})(e_w), \quad \mathbf{x} \in \mathcal{T}_0^N, \\ (\partial_w \partial_{w'} \exp_N)(\mathbf{x}) &= \tilde{Q}(\text{ad } \mathbf{x})(e_w \otimes e_{w'}) \exp_N(\mathbf{x}), \quad \mathbf{x} \in \mathcal{T}_0^N, \end{aligned}$$

for all words $w, w' \in \mathcal{W}_d$ with $1 \leq |w|, |w'| \leq N$, where G is defined in equation (4.1) and for

$$\begin{aligned} \tilde{Q}(\text{ad } \mathbf{x})(a \otimes b) &= G(\text{ad } \mathbf{x})(b) G(\text{ad } \mathbf{x})(a) + \int_0^1 \tau [G(\tau \text{ad } \mathbf{x})(a), e^{\tau \text{ad } \mathbf{x}}(b)] \, d\tau \\ &= \sum_{n,m=0}^N \frac{(\text{ad } \mathbf{x})^n(b)}{(n+1)!} \frac{(\text{ad } \mathbf{x})^m(a)}{(m+1)!} + \sum_{n,m=0}^N \frac{[(\text{ad } \mathbf{x})^n(a), (\text{ad } \mathbf{x})^m(b)]}{(n+m+2)(n+1)!m!}, \quad \mathbf{x}, a, b \in \mathcal{T}_0^N. \end{aligned}$$

Proof. For all $w \in \mathcal{W}_d$ with $0 \leq |w| \leq N$ and $\mathbf{x} \in \mathcal{T}_0^N$, the expression $\exp_N(\mathbf{x})^w$ is a polynomial in the tensor components $(\mathbf{x}^v)_{1 \leq |v| \leq |w|}$. Therefore the map $\exp_N : \mathcal{T}_1^N \rightarrow \mathcal{T}_0^N$ is smooth, and in particular the first and second-order partial derivatives exist in all directions. For a proof of the explicit form of the first-order partial derivatives, we refer to [27, Lemma 7.23]. For the second-order derivatives, we follow the proof of [37, Lemma A.1]. Therefore, let $\mathbf{x} \in \mathcal{T}_0^N$ and w, w' arbitrary with $1 \leq |w|, |w'| \leq N$. Then we have by the definition of the partial derivatives in \mathcal{T}_0^N and the product rule

$$\begin{aligned} \partial_w (\partial_{w'} \exp_N(\mathbf{x})) &= \frac{d}{dt} \left(G(\text{ad } \mathbf{x} + te_w)(e_{w'}) \exp_N(\mathbf{x} + te_w) \right) \Big|_{t=0} \\ &= \frac{d}{dt} G(\text{ad } \mathbf{x} + te_w)(e_{w'}) \Big|_{t=0} \exp_N(\mathbf{x}) + G(\text{ad } \mathbf{x})(e_{w'}) G(\text{ad } \mathbf{x})(e_w) \exp_N(\mathbf{x}). \end{aligned}$$

From [27, Lemma 7.22], it holds that $\exp_N(\text{ad } \mathbf{x})(y) = \exp_N(\mathbf{x})y \exp_N(-\mathbf{x})$ for all $\mathbf{x}, y \in \mathcal{T}_0^N$, and it follows further by representing G in integral form that

$$\begin{aligned} \frac{d}{dt} G(\text{ad } \mathbf{x} + te_w)(e_{w'}) \Big|_{t=0} &= \frac{d}{dt} \left(\int_0^1 \exp_N(\tau \text{ad } \mathbf{x} + te_w)(e_{w'}) \, d\tau \right) \Big|_{t=0} \\ &= \int_0^1 \frac{d}{dt} \left(\exp_N(\tau(\mathbf{x} + te_w)) e_{w'} \exp_N(-\tau(\mathbf{x} + te_w)) \right) \Big|_{t=0} \, d\tau \\ &= \int_0^1 \tau G(\text{ad } \tau \mathbf{x})(e_w) \exp_N(\tau \mathbf{x}) e_{w'} \exp_N(-\tau \mathbf{x}) \, d\tau \\ &\quad - \int_0^1 \tau \exp_N(\tau \mathbf{x}) e_{w'} \exp_N(-\tau \mathbf{x}) G(\text{ad } \tau \mathbf{x})(e_w) \, d\tau \\ &= \int_0^1 \tau [G(\tau \text{ad } \mathbf{x})(e_w), \exp_N(\tau \text{ad } \mathbf{x})(e_{w'})] \, d\tau. \end{aligned}$$

Then the proof is finished after noting that

$$\begin{aligned} \int_0^1 \tau G(\tau \operatorname{ad} \mathbf{x})(e_w) \exp(\tau \operatorname{ad} \mathbf{x})(e_{w'}) \, d\tau &= \int_0^1 \tau \sum_{n,m=0}^N \frac{(\tau \operatorname{ad} \mathbf{x})^n(e_w)}{(n+1)!} \frac{(\tau \operatorname{ad} \mathbf{x})^m(e_{w'})}{m!} \, d\tau \\ &= \int_0^1 \sum_{n,m=0}^N \frac{(\operatorname{ad} \mathbf{x})^n(e_w)}{(n+1)!} \frac{(\operatorname{ad} \mathbf{x})^m(e_{w'})}{m!} \tau^{1+m+n} \, d\tau \\ &= \sum_{n,m=0}^N \frac{(\operatorname{ad} \mathbf{x})^n(e_w)(\operatorname{ad} \mathbf{x})^m(e_{w'})}{(n+1)!m!(n+m+2)}. \end{aligned}$$

□

Note that the operator Q defined in equation (4.1) differs from the operator \tilde{Q} defined above. However, we have the following:

Lemma 7.6. *Let $N \in \mathbb{N}_{\geq 1}$ and $\mathbf{x} \in \mathcal{T}_0^N$; then it holds that*

$$\tilde{Q}(\operatorname{ad} \mathbf{x})(\mathbb{A}) = Q(\operatorname{ad} \mathbf{x})(\mathbb{A}),$$

for all $\mathbb{A} \in \mathcal{T}_0^N \otimes \mathcal{T}_0^N$ with symmetric coefficients $\mathbb{A}^{w_1, w_2} = \mathbb{A}^{w_2, w_1}$ for all $w_1, w_2 \in \mathcal{W}_d$.

Proof. Let $N \in \mathbb{N}_{\geq 1}$ and $\mathbf{x} \in \mathcal{T}_0^N$ be arbitrary. Then from the bilinearity of $\tilde{Q}(\operatorname{ad} \mathbf{x})$ and the symmetry of \mathbb{A} , we have, with summation over all words w_1, w_2 with $1 \leq |w_1|, |w_2| \leq N$,

$$\begin{aligned} \tilde{Q}(\operatorname{ad} \mathbf{x})(\mathbb{A}) &= \sum_{w_1, w_2} \mathbb{A}^{w_1, w_2} \sum_{n,m=0}^N \left(\frac{(\operatorname{ad} \mathbf{x})^n(e_{w_2})}{(n+1)!} \frac{(\operatorname{ad} \mathbf{x})^m(e_{w_1})}{(m+1)!} + \frac{[(\operatorname{ad} \mathbf{x})^n(e_{w_1}), (\operatorname{ad} \mathbf{x})^m(e_{w_2})]1}{(n+m+2)(n+1)!m!} \right) \\ &= \sum_{w_1, w_2} \mathbb{A}^{w_1, w_2} \left(\sum_{n,m=0}^N \frac{(\operatorname{ad} \mathbf{x})^n(e_{w_2})}{(n+1)!} \frac{(\operatorname{ad} \mathbf{x})^m(e_{w_1})}{(m+1)!} \right. \\ &\quad \left. + \frac{(\operatorname{ad} \mathbf{x})^n(e_{w_1})(\operatorname{ad} \mathbf{x})^m(e_{w_2}) - (\operatorname{ad} \mathbf{x})^m(e_{w_2})(\operatorname{ad} \mathbf{x})^n(e_{w_1})}{(n+m+2)(n+1)!m!} \right) \\ &= \sum_{w_1, w_2} \mathbb{A}^{w_1, w_2} \left(\sum_{n,m=0}^N \frac{(\operatorname{ad} \mathbf{x})^n(e_{w_1})}{(n+1)!} \frac{(\operatorname{ad} \mathbf{x})^m(e_{w_2})}{(m+1)!} \right. \\ &\quad \left. + \frac{(\operatorname{ad} \mathbf{x})^n(e_{w_1})(\operatorname{ad} \mathbf{x})^m(e_{w_2})}{(n+m+2)(n+1)!m!} - \frac{(\operatorname{ad} \mathbf{x})^n(e_{w_1})(\operatorname{ad} \mathbf{x})^m(e_{w_2})}{(m+n+2)(m+1)!n!} \right) \\ &= \sum_{w_1, w_2} \mathbb{A}^{w_1, w_2} \sum_{n,m=0}^N (2m+2) \frac{(\operatorname{ad} \mathbf{x})^n(e_{w_1})(\operatorname{ad} \mathbf{x})^m(e_{w_2})}{(n+1)!(m+1)!(n+m+2)}(\mathbb{A}) = Q(\operatorname{ad} \mathbf{x})(\mathbb{A}). \end{aligned}$$

□

The following two applications of Itô’s formula in the non-commutative setting will be a key ingredient in the proof of Theorem 4.1.

Lemma 7.7 (Itô’s product rule). *Let $\mathbf{X}, \mathbf{Y} \in \mathcal{S}(\mathcal{T}_1^N)$ for some $N \in \mathbb{N}_{\geq 1}$; then it holds that*

$$\mathbf{X}_t \mathbf{Y}_t - \mathbf{X}_0 \mathbf{Y}_0 = \int_{(0,t]} (d\mathbf{X}_u) \mathbf{Y}_u + \int_{(0,t]} \mathbf{X}_u (d\mathbf{Y}_u) + m(\llbracket \mathbf{X}, \mathbf{Y} \rrbracket_{0,T}), \quad 0 \leq t \leq T.$$

Proof. The statement is an immediate consequence of the one-dimensional Itô’s product rule for càdlàg semimartingales (e.g., [57, Ch. II, Corollary 2]) and the definition of the outer bracket and the multiplication map in Section 2.4. □

Lemma 7.8. Let $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0^N)$ for some $N \in \mathbb{N}_{\geq 1}$; then it holds that

$$\begin{aligned} & \exp_N(\mathbf{X}_t) - \exp_N(\mathbf{X}_0) \\ &= \int_{(0,t]} G(\text{ad } \mathbf{X}_{u-})(d\mathbf{X}_u) \exp_N(\mathbf{X}_{u-}) + \int_0^t Q(\text{ad } \mathbf{X}_{u-})(d\llbracket \mathbf{X}^c, \mathbf{X}^c \rrbracket_u) \exp_N(\mathbf{X}_{u-}) \\ &+ \sum_{0 < u \leq t} \left(\exp_N(\mathbf{X}_u) - \exp_N(\mathbf{X}_{u-}) - G(\text{ad } \mathbf{X}_{u-})(\Delta \mathbf{X}_u) \exp_N(\mathbf{X}_{u-}) \right), \end{aligned}$$

for all $0 \leq t \leq T$.

Proof. As discussed in the proof of Lemma 7.5, it is clear that the map $\exp_N : \mathcal{T}_0^N \rightarrow \mathcal{T}_1^N$ is smooth. Further, \mathcal{T}_0^N is isomorphic to \mathbb{R}^D with $D = d + \dots + d^N$, and we can apply the multidimensional Itô’s formula for càdlàg semimartingales (e.g., [57, Ch. II, Theorem 33]) to obtain

$$\begin{aligned} \exp_N(\mathbf{X}_t) - \exp_N(\mathbf{X}_0) &= \sum_{1 \leq |w| \leq N} \int_{(0,t]} (\partial_w \exp_N)(\mathbf{X}_{u-}) d\mathbf{X}_u^w \\ &+ \frac{1}{2} \sum_{1 \leq |w_1|, |w_2| \leq N} \int_0^t (\partial_{w_1} \partial_{w_2} \exp_N)(\mathbf{X}_{u-}) d\langle \mathbf{X}^{w_1 c}, \mathbf{X}^{w_2 c} \rangle_u \\ &+ \sum_{0 < u \leq t} \left(\exp_N(\mathbf{X}_u) - \exp_N(\mathbf{X}_{u-}) - \sum_{1 \leq |w| \leq N} (\partial_w \exp_N)(\mathbf{X}_{u-})(\Delta \mathbf{X}_u^w) \right) \end{aligned}$$

for all $0 \leq t \leq T$. From Lemma 7.5 we then have for the first integral term

$$\begin{aligned} \sum_{1 \leq |w| \leq N} \int_{(0,t]} (\partial_w \exp_N)(\mathbf{X}_{u-}) d\mathbf{X}_u^w &= \sum_{1 \leq |w| \leq N} \int_{(0,t]} G(\text{ad } \mathbf{X}_{u-})(e_w) \exp_N(\mathbf{X}_{u-}) d\mathbf{X}_u^w \\ &= \int_{(0,t]} G(\text{ad } \mathbf{X}_{u-})(d\mathbf{X}_u) \exp_N(\mathbf{X}_{u-}), \end{aligned}$$

and analogously

$$\sum_{1 \leq |w| \leq N} (\partial_w \exp_N)(\mathbf{X}_{u-})(\Delta \mathbf{X}_u^w) = \sum_{1 \leq |w| \leq N} G(\text{ad } \mathbf{X}_{u-})(\Delta \mathbf{X}_u) \exp_N(\mathbf{X}_{u-}).$$

Moreover, from Lemma 7.8 and the definition of the outer bracket in Section 2.4,

$$\sum_{1 \leq |w_1|, |w_2| \leq N} \int_{(0,t]} (\partial_{w_1} \partial_{w_2} \exp_N)(\mathbf{X}_{u-}) d\langle \mathbf{X}^{w_1 c}, \mathbf{X}^{w_2 c} \rangle_u = \int_0^t \tilde{Q}(\text{ad } \mathbf{X}_{u-})(d\llbracket \mathbf{X}^c, \mathbf{X}^c \rrbracket_u) \exp_N(\mathbf{X}_{u-}).$$

Finally, the outer bracket $\llbracket \mathbf{X}^c, \mathbf{X}^c \rrbracket_t \in \mathcal{T}_0^N \otimes \mathcal{T}_0^N$ is symmetric in the sense of Lemma 7.6, and therefore we can replace \tilde{Q} with Q in the above identity. □

Lemma 7.9. Let $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0)$, and let $\mathbf{A} \in \mathcal{V}(\mathcal{T}_0)$. For all $k \in \mathbb{N}_{\geq 1}$ and $\ell = (l_1, \dots, l_k) \in (\mathbb{N}_{\geq 1})^k$, it holds that

$$\left| \int_0^t \left(\text{ad } \mathbf{X}_u^{(l_2)} \dots \text{ad } \mathbf{X}_u^{(l_k)} \right) \left(d\mathbf{A}_u^{(l_1)} \right) \right| \leq 2^{k-1} \int_0^t \left| \mathbf{X}_u^{(l_2)} \dots \mathbf{X}_u^{(l_k)} \right| \left| d\mathbf{A}_u^{(l_1)} \right|,$$

for all $0 \leq t \leq T$. Furthermore, let $(\mathbb{A}_t)_{0 \leq t \leq T}$ be a process taking values in $\mathcal{T}_0 \otimes \mathcal{T}_0$ such that $\mathbb{A}^{w_1, w_2} \in \mathcal{V}$ for all $w_1, w_2 \in \mathcal{W}_d$. Then it holds that for all $0 \leq t \leq T$

$$\left| \int_0^t \left(\text{ad } \mathbf{X}_u^{(l_3)} \cdots \text{ad } \mathbf{X}_u^{(l_m)} \odot \text{ad } \mathbf{X}_u^{(l_{m+1})} \cdots \text{ad } \mathbf{X}_u^{(l_k)} \right) \left(d\mathbb{A}_u^{(l_1, l_2)} \right) \right| \leq 2^{k-2} \int_0^t \left| \mathbf{X}_u^{(l_3)} \cdots \mathbf{X}_u^{(l_k)} \right| \left| dm \left(\mathbb{A}^{(l_1, l_2)} \right) \right|.$$

Proof. Recall from equation (2.3) that we expand iterated adjoined operations into a sum of left- and right tensor multiplications and apply Lemma 7.2. Note again that for homogeneous tensors \mathbf{x} and \mathbf{y} , it holds that $|\mathbf{x}\mathbf{y}| = |\mathbf{y}\mathbf{x}|$. Therefore the statement follows by counting the terms in the expansion. \square

Lemma 7.10. Let $N \in \mathbb{N}_{\geq 1}$, $\Delta \mathbf{x}, \mathbf{y}, \Delta \mathbf{y} \in \mathcal{T}_0^N$, and define the function

$$f: [0, 1] \times [0, 1] \rightarrow \mathcal{T}_1^N, \quad (s, t) \mapsto f(s, t) = \exp_N(s\Delta \mathbf{x}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y}).$$

Then $f(0, 0) = 1$, and the first-order partial derivatives of f at $(s, t) = (0, 0)$ are given by

$$(\partial_s f)|_{(s,t)=(0,0)} = \Delta \mathbf{x}, \quad (\partial_t f)|_{(s,t)=(0,0)} = G(\text{ad } \mathbf{y})(\Delta \mathbf{y}).$$

Further, the following explicit bound for the second-order partial derivatives holds

$$\sup_{0 \leq s, t \leq 1} \left| (\nabla^2 f^{(n)})|_{(s,t)} \right| \leq c_n \sum_{\|\ell\|=n, |\ell| \geq 2} \left(\left| \Delta \mathbf{x}^{(l_1)} \right| + \left| \Delta \mathbf{y}^{(l_1)} \right| \right) \left(\left| \Delta \mathbf{x}^{(l_2)} \right| + \left| \Delta \mathbf{y}^{(l_2)} \right| \right) z_{l_3} \cdots z_{l_n},$$

for all $n \in \{2, \dots, N\}$, where $c_n > 0$ is a constant depending only on n , $\ell = (l_1, \dots, l_k) \in (\mathbb{N}_{\geq 1})^k$ with $|\ell| = k$ and $z_l := \max\{|\Delta \mathbf{x}^{(l)}|, |\mathbf{y}^{(l)}|, |(\mathbf{y} + \Delta \mathbf{y})^{(l)}|\}$ for all $l \in \{1, \dots, N-2\}$.

Proof. The tensor components of $f(s, t)$ are polynomial in s and t , and it follows that f is smooth. From Lemma 7.5, we have that the first-order partial derivatives of f are given by

$$\begin{aligned} (\partial_s f)|_{(s,t)} &= G(\text{ad } s\Delta \mathbf{x})(\Delta \mathbf{x}) \exp_N(s\Delta \mathbf{x}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y}), \\ &= \Delta \mathbf{x} \exp_N(s\Delta \mathbf{x}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y}), \\ (\partial_t f)|_{(s,t)} &= \exp_N(s\Delta \mathbf{x}) G(\text{ad } \mathbf{y} + t\Delta \mathbf{y})(\Delta \mathbf{y}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y}). \end{aligned}$$

Evaluating at $s = t = 0$, we obtain the first result. Further, by Lemma 7.5 the second-order derivatives are given by

$$\begin{aligned} (\partial_{ss} f)|_{(s,t)} &= (\Delta \mathbf{x})^2 \exp_N(s\Delta \mathbf{x}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y}), \\ (\partial_{st} f)|_{(s,t)} &= \Delta \mathbf{x} \exp_N(s\Delta \mathbf{x}) G(\text{ad } \mathbf{y} + t\Delta \mathbf{y})(\Delta \mathbf{y}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y}), \\ (\partial_{tt} f)|_{(s,t)} &= \exp_N(s\Delta \mathbf{x}) Q(\text{ad } \mathbf{y} + t\Delta \mathbf{y})((\Delta \mathbf{y})^{\otimes 2}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y}). \end{aligned}$$

Now let $n \in \{2, \dots, N\}$. Then it follows from above and Lemma 7.9 that we can bound the second-order derivatives as follows

$$\begin{aligned} &\sup_{0 \leq s, t \leq 1} (\partial_{ss} f^{(n)})|_{(s,t)} \\ &\leq \sup_{0 \leq s, t \leq 1} \left| \pi_n(Q(\text{ad } s\Delta \mathbf{x})((\Delta \mathbf{x})^{\otimes 2}) \exp_N(s\Delta \mathbf{x}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y})) \right| \\ &\leq c'_n \sum_{\|\ell\|=n, |\ell| \geq 2} |\Delta \mathbf{x}^{(l_1)}| |\Delta \mathbf{x}^{(l_2)}| z_{l_3} \cdots z_{l_n}, \end{aligned}$$

$$\begin{aligned} & \sup_{0 \leq s, t \leq 1} (\partial_{tt} f^{(n)})|_{(s,t)} \\ & \leq \sup_{0 \leq s, t \leq 1} \left| \pi_n(\exp_N(s\Delta \mathbf{x})Q(\text{ad } \mathbf{y} + t\Delta \mathbf{y})((\Delta \mathbf{y})^{\otimes 2}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y})) \right| \\ & \leq c''_n \sum_{\|\ell\|=n, |\ell| \geq 2} |\Delta \mathbf{y}^{(\ell_1)}| |\Delta \mathbf{y}^{(\ell_2)}|_{z_{l_3} \cdots z_{l_k}}, \end{aligned}$$

and

$$\begin{aligned} & \sup_{0 \leq s, t \leq 1} (\partial_{st} f^{(n)})|_{(s,t)} \\ & \leq \sup_{0 \leq s, t \leq 1} \left| \pi_n(G(\text{ad } s\Delta \mathbf{x})(\Delta \mathbf{x}) \exp_N(s\Delta \mathbf{x})G(\text{ad } \mathbf{y} + t\Delta \mathbf{y}) \exp_N(\mathbf{y} + t\Delta \mathbf{y}) \exp_N(-\mathbf{y})) \right| \\ & \leq c'''_n \sum_{\|\ell\|=n, |\ell| \geq 2} |\Delta \mathbf{x}^{(\ell_1)}| |\Delta \mathbf{y}^{(\ell_2)}|_{z_{l_3} \cdots z_{l_k}}, \end{aligned}$$

where $c'_n, c''_n, c'''_n > 0$ are constants depending only on n , and the second statement of the lemma follows. □

Lemma 7.11. For all $N \in \mathbb{N}_{\geq 1}$ and all $\mathbf{x} \in \mathcal{T}_0^N$, it holds that

$$H(\text{ad } \mathbf{x}) \circ G(\text{ad } \mathbf{x}) = \text{Id},$$

where G and H are defined in equation (4.1). Hence, the identity also holds for all $x \in \mathcal{T}_0$.

Proof. Recall the exponent generating function of the Bernoulli numbers, for z near 0,

$$H(z) = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n = \frac{z}{e^z - 1}, \quad G(z) = \sum_{k=0}^{\infty} \frac{1}{(k+1)!} z^k = \frac{e^z - 1}{z}.$$

Therefore $H(z)G(z) \equiv 1$ identically for all z in a neighbourhood of zero. Repeated differentiation in z then yields the following property of the Bernoulli numbers

$$\sum_{k=0}^n \frac{B_k}{k!} \frac{1}{(n-k+1)!} = 0, \quad n \in \mathbb{N}_{\geq 1}.$$

Hence the statement of the lemma follows by projecting $H(\text{ad } \mathbf{x}) \circ G(\text{ad } \mathbf{x})$ to each tensor level. □

We are now ready to give the

Proof of Theorem 4.1. Note that $\pi_{(0,N)} \text{Sig}(\mathbf{X}) = \text{Sig}(\mathbf{X}^{(0,N)})$ for any $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0)$ and all truncation levels $N \in \mathbb{N}_{\geq 1}$. Therefore it suffices to show that the identities in equations (4.2) and (4.3) hold for the signature cumulant of an arbitrary \mathcal{T}_0^N -valued semimartingale $\mathbf{X} \in \mathcal{S}(\mathcal{T}_0^N)$ that satisfies the integrability condition $\|\mathbf{X}\|_{\mathcal{S}^{1,N}} < \infty$. Recall from Theorem 3.2 that this implies that $\|\text{Sig}(\mathbf{X})\|_{\mathcal{S}^{1,N}} < \infty$, and thus the truncated signature cumulant $\kappa = (\log_N \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}))_{0 \leq t \leq T} \in \mathcal{S}(\mathcal{T}_0^N)$ is well defined. Throughout the proof, we will use the symbol \lesssim to denote an inequality that holds up to a multiplication of the right-hand side by a constant that may depend only on d and N .

Recall the definition of the signature in the Marcus sense from Section 2.6. Projecting equation (2.11) to the truncated tensor algebra, we see that the signature process $S = (\text{Sig}(\mathbf{X})_{0,t})_{0 \leq t \leq T} \in \mathcal{S}(\mathcal{T}_1^N)$ satisfies the integral equation

$$S_t = 1 + \int_{(0,t]} S_{u-} d\mathbf{X}_u + \frac{1}{2} \int_0^t S_u d\langle \mathbf{X}^c \rangle_u + \sum_{0 < u \leq t} S_{u-} (\exp_N(\Delta \mathbf{X}_u) - 1 - \Delta \mathbf{X}_u), \quad (7.7)$$

for $0 \leq t \leq T$. Then by Chen’s relation in equation (2.12), we have

$$\mathbb{E}_t(S_T \exp_N(\boldsymbol{\kappa}_T)) = \mathbb{E}_t(\text{Sig}(\mathbf{X})_{0,T}) = S_t \mathbb{E}_t(\text{Sig}(\mathbf{X})_{t,T}) = S_t \exp_N(\boldsymbol{\kappa}_t), \quad 0 \leq t \leq T.$$

It then follows from the above identity and the integrability of S_T that the process $S \exp_N(\boldsymbol{\kappa})$ is a \mathcal{T}_1^N -valued martingale in the sense of Section 2.4. On the other hand, we have by applying Itô’s product rule in Lemma 7.7

$$\begin{aligned} S_t \exp_N(\boldsymbol{\kappa}_t) - 1 &= \int_{(0,t]} (dS_u) \exp_N(\boldsymbol{\kappa}_{u-}) + \int_{(0,t]} S_{u-} (d \exp_N(\boldsymbol{\kappa}_u)) + m\left(\llbracket S^c, \exp_N(\boldsymbol{\kappa})^c \rrbracket_{0,t}\right) \\ &\quad + \sum_{0 < u \leq t} \Delta S_u \Delta \exp_N(\boldsymbol{\kappa}_u). \end{aligned}$$

Further, by applying the Itô’s rule for the exponential map from Lemma 7.8 to the \mathcal{T}_0^N -valued semi-martingale $\boldsymbol{\kappa}$ and using equation (7.7), we have the following form of the continuous covariation term

$$\begin{aligned} m\left(\llbracket S^c, \exp_N(\boldsymbol{\kappa}^c) \rrbracket_{0,t}\right) &= m\left(\left\llbracket \int_{(0,\cdot]} S_{u-} d\mathbf{X}_u^c, \int_{(0,\cdot]} G(\text{ad } \boldsymbol{\kappa}_{u-})(d\boldsymbol{\kappa}_u^c) \exp_N(\boldsymbol{\kappa}_{u-}) \right\rrbracket_{0,t}\right) \\ &= \int_{(0,t]} S_{u-} (\text{Id} \odot G(\text{ad } \boldsymbol{\kappa}_{u-})) (d\llbracket \mathbf{X}^c, \boldsymbol{\kappa}^c \rrbracket_u) \exp_N(\boldsymbol{\kappa}_{u-}) \end{aligned}$$

and for the jump covariation term

$$\sum_{0 < u \leq t} \Delta S_u \Delta \exp_N(\boldsymbol{\kappa}_u) = \sum_{0 < u \leq t} S_{u-} (\exp_N(\Delta \mathbf{X}_u) - 1) (\exp_N(\boldsymbol{\kappa}_u) \exp_N(-\boldsymbol{\kappa}_{u-}) - 1) \exp_N(\boldsymbol{\kappa}_{u-}).$$

From the above identities and again with Lemma 7.8 and in equation (7.7), we have

$$S_t \exp_N(\boldsymbol{\kappa}_t) - 1 = \int_{(0,t]} S_{u-} d(\mathbf{L}_u + \boldsymbol{\kappa}_u) \exp_N(\boldsymbol{\kappa}_{u-}), \quad 0 \leq t \leq T, \tag{7.8}$$

where $\mathbf{L} \in \mathcal{S}(\mathcal{T}_0^N)$ is defined by

$$\begin{aligned} \mathbf{L}_t &= \mathbf{X}_t + \frac{1}{2} \langle \mathbf{X}^c \rangle_t + \sum_{0 < u \leq t} (\exp_N(\Delta \mathbf{X}_u) - 1 - \Delta \mathbf{X}_u) + \int_{(0,t]} (G - \text{Id})(\text{ad } \boldsymbol{\kappa}_{u-})(d\boldsymbol{\kappa}_u) \\ &\quad + \int_0^t \frac{1}{2} \mathcal{Q}(\text{ad } \boldsymbol{\kappa}_{u-})(d\llbracket \boldsymbol{\kappa}^c, \boldsymbol{\kappa}^c \rrbracket_u) + \sum_{0 < u \leq t} (\exp_N(\boldsymbol{\kappa}_u) \exp_N(-\boldsymbol{\kappa}_{u-}) - 1 - G(\text{ad } \boldsymbol{\kappa}_{u-})(\Delta \boldsymbol{\kappa}_u)) \\ &\quad + \int_{(0,t]} (\text{Id} \odot G(\text{ad } \boldsymbol{\kappa}_{u-})) (d\llbracket \mathbf{X}^c, \boldsymbol{\kappa}^c \rrbracket_u) \\ &\quad + \sum_{0 < u \leq t} (\exp_N(\Delta \mathbf{X}_u) - 1) (\exp_N(\boldsymbol{\kappa}_u) \exp_N(-\boldsymbol{\kappa}_{u-}) - 1) \\ &= \mathbf{X}_t + \frac{1}{2} \langle \mathbf{X}^c \rangle_t + \mathbf{Y}_t + \mathbf{V}_t + \mathbf{C}_t + \mathbf{J}_t, \end{aligned}$$

with $\mathbf{Y} \in \mathcal{S}(\mathcal{T}_0^N)$, $\mathbf{V}, \mathbf{C}, \mathbf{J} \in \mathcal{V}(\mathcal{T}_0^N)$ given by

$$\begin{aligned} \mathbf{Y}_t &= \int_{(0,t]} (G - \text{Id})(\text{ad } \kappa_{u-})(d\kappa_u), \\ \mathbf{V}_t &= \frac{1}{2} \int_0^t \mathcal{Q}(\text{ad } \kappa_{u-})(d[\kappa^c, \kappa^c]_u), \\ \mathbf{C}_t &= \int_{(0,t]} (\text{Id} \odot G(\text{ad } \kappa_{u-}))(d[\mathbf{X}^c, \kappa^c]_u), \\ \mathbf{J}_t &= \sum_{0 < u \leq t} (\exp_N(\Delta \mathbf{X}_u) \exp_N(\kappa_u) \exp_N(-\kappa_{u-}) - 1 - \Delta \mathbf{X}_u - G(\text{ad } \kappa_{u-})(\Delta \kappa_u)). \end{aligned} \tag{7.9}$$

Note that we have explicitly separated the identity operator Id from G in the above definition of \mathbf{L} .

Since the left-hand side in equation (7.8) is a martingale, and since S_t and $\exp(\kappa_t)$ have the multiplicative left- respectively right-inverse S_t^{-1} and $\exp(-\kappa_t)$, respectively, for all $0 \leq t \leq T$, it follows that $\mathbf{L} + \kappa$ is a \mathcal{T}_0^N -valued local martingale. Let $(\tau_k)_{k \geq 1}$ be a sequence of increasing stopping times with $\tau_k \rightarrow T$ a.s. for $k \rightarrow \infty$, such that the stopped process $(\mathbf{L}_{t \wedge \tau_k} + \kappa_{t \wedge \tau_k})_{0 \leq t \leq T}$ is a true martingale. Then we have

$$\kappa_{t \wedge \tau_k, T \wedge \tau_k} = \mathbb{E}_t \{ \mathbf{L}_{t \wedge \tau_k, T \wedge \tau_k} \}, \quad 0 \leq t \leq T, \quad k \in \mathbb{N}_{\geq 1}. \tag{7.10}$$

The estimate in equation (7.11) below shows that \mathbf{L} has sufficient integrability in order to use the dominated convergence theorem to pass to the $k \rightarrow \infty$ limit in the above identity equation (7.10), which yields precisely the identity equation (4.2) (recall that $\kappa_T = 0$) and hence concludes the first part of the proof of Theorem 4.1.

Claim 7.12. *It holds that*

$$\|\mathbf{L}\|_{\mathcal{H}^{1,N}} \lesssim \|\mathbf{X}\|_{\mathcal{H}^{1,N}}. \tag{7.11}$$

Proof of Claim 7.12. According to Lemma 7.4, it suffices to show that for all $n \in \{1, \dots, N\}$, it holds that

$$\|\mathbf{L}^{(n)}\|_{\mathcal{H}^{qN/n}} \lesssim \sum_{\|\ell\|=n} \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{qN/l_1}} \cdots \|\mathbf{X}^{(l_j)}\|_{\mathcal{H}^{qN/l_j}} =: \rho_{\mathbf{X}}^n,$$

where the summation above (and in the rest of the proof) is over multi-indices $\ell = (l_1, \dots, l_j) \in (\mathbb{N}_{\geq 1})^j$, with $|\ell| = j$ and $\|\ell\| = l_1 + \dots + l_j$. For $\mathbf{M} \in \mathcal{M}_{\text{loc}}(\mathcal{T}_0^N)$ and $\mathbf{A} \in \mathcal{V}(\mathcal{T}_0^N)$ define

$$\rho_{\mathbf{M}, \mathbf{A}}^n := \sum_{\|\ell\|=n} \zeta^{l_1/N}(\mathbf{M}^{(l_1)}, \mathbf{A}^{(l_1)}) \cdots \zeta^{l_j/N}(\mathbf{M}^{(l_j)}, \mathbf{A}^{(l_j)}) < \infty, \quad n = 1, \dots, N,$$

where for any $q \in [1, \infty)$

$$\zeta^q(\mathbf{M}^{(l)}, \mathbf{A}^{(l)}) := \left\| \left| [\mathbf{M}^{(l)}]_T \right|^{1/2} + \left| \mathbf{A}^{(l)} \right|_{1\text{-var}; [0, T]} \right\|_{\mathcal{L}^q}.$$

Note that it holds

$$\rho_{\mathbf{M}, \mathbf{A}}^n \leq \sum_{\|\ell\|=n} \left(\rho_{\mathbf{M}, \mathbf{A}}^{l_1} \cdots \rho_{\mathbf{M}, \mathbf{A}}^{l_j} \right) \lesssim \rho_{\mathbf{M}, \mathbf{A}}^n. \tag{7.12}$$

Furthermore, it follows from the definition of the \mathcal{H}^q -norm that

$$\rho_{\mathbf{X}}^n = \inf_{\mathbf{X}=\mathbf{M}+\mathbf{A}} \rho_{\mathbf{M},\mathbf{A}}^n, \tag{7.13}$$

where the infimum is taken over all semimartingale decomposition of \mathbf{X} .

Now fix $\mathbf{M} \in \mathcal{M}_{\text{loc}}(\mathcal{T}_0^N)$ and $\mathbf{A} \in \mathcal{V}(\mathcal{T}_0^N)$ arbitrarily, such that $\mathbf{X} = \mathbf{M} + \mathbf{A}$ and $\rho_{\mathbf{M},\mathbf{A}}^n < \infty$ for all $n \in \{1, \dots, N\}$ (such a decomposition always exists since $\mathbf{X} \in \mathcal{H}^{1,N}$). In particular, it holds that \mathbf{M} is a true martingale. Next we will prove the following.

Claim 7.13. *For all $n \in \{1, \dots, N\}$, it holds that*

$$\|\mathbf{L}^{(n)}\|_{\mathcal{H}^{N/n}} \lesssim \rho_{\mathbf{M},\mathbf{A}}^n, \tag{7.14}$$

and further, there exists a semimartingale decomposition $\kappa^{(n)} = \kappa_0^{(n)} + \mathbf{m}^{(n)} + \mathbf{a}^{(n)}$, with $\mathbf{m}^{(n)} \in \mathcal{M}((\mathbb{R}^d)^{\otimes n})$ and $\mathbf{a}^{(n)} \in \mathcal{V}((\mathbb{R}^d)^{\otimes n})$ such that $\|\mathbf{a}^{(n)}\|_{\mathcal{H}^{N/n}} < \rho_{\mathbf{M},\mathbf{A}}^n$; and in case $n \leq N - 1$, it holds that

$$\zeta^{n/N}(\mathbf{m}^{(n)}, \mathbf{a}^{(n)}) \lesssim \rho_{\mathbf{M},\mathbf{A}}^n. \tag{7.15}$$

Proof of Claim 7.13. We are going to prove inductively over $n \in \{1, \dots, N\}$. Let $n = 1$, and note that we have $\mathbf{L}^{(1)} = \mathbf{X}^{(1)}$, and therefore

$$\|\mathbf{L}^{(1)}\|_{\mathcal{H}^N} \leq \zeta^N(\mathbf{M}^{(1)}, \mathbf{A}^{(1)}) = \rho_{\mathbf{M},\mathbf{A}}^1.$$

Using that $\mathbf{M}^{(1)}$ is a martingale, we can identify a semimartingale decomposition of $\kappa^{(1)}$ by

$$\mathbf{m}_t^{(1)} := \mathbb{E}_t(\mathbf{A}_T^{(1)}) - \mathbb{E}(\mathbf{A}_T^{(1)}), \quad \mathbf{a}_t^{(1)} := -\mathbf{A}_t, \quad 0 \leq t \leq T.$$

In case $N \geq 2$, we further have from the BDG-inequality and Doob’s maximal inequality that

$$\|\mathbf{m}^{(1)}\|_{\mathcal{H}^N} \lesssim \|\mathbf{m}^{(1)}\|_{\mathcal{S}^N} \lesssim \|\mathbf{m}_T^{(1)}\|_{\mathcal{L}^N} = \|\mathbb{E}(\mathbf{A}_T^{(1)}) - \mathbf{A}_T^{(1)}\|_{\mathcal{L}^N} \lesssim \|\mathbf{A}^{(1)}\|_{\mathcal{H}^{N/n}} \lesssim \rho_{\mathbf{M},\mathbf{A}}^1$$

and this shows the second part of the induction claim.

Now assume that $N \geq 2$ and that the induction claim in equations (7.14) and (7.15) holds true up to level $n - 1$ for some $n \in \{2, \dots, N\}$. Note that $\mathbf{L}^{(n)}$ has the following decomposition:

$$\mathbf{L}^{(n)} = \left\{ \mathbf{M}^{(n)} + \mathbf{N}^{(n)} \right\} + \left\{ \mathbf{A}^{(n)} + \frac{1}{2} \langle \mathbf{X}^c \rangle^{(n)} + \mathbf{B}^{(n)} + \mathbf{V}^{(n)} + \mathbf{C}^{(n)} + \mathbf{J}^{(n)} \right\}, \tag{7.16}$$

where $\mathbf{N}^{(n)} \in \mathcal{M}_{\text{loc}}((\mathbb{R}^d)^{\otimes n})$ and $\mathbf{B}^{(n)} \in \mathcal{V}((\mathbb{R}^d)^{\otimes n})$ is a decomposition of $\mathbf{Y}^{(n)} \in \mathcal{S}((\mathbb{R}^d)^{\otimes n})$ defined by

$$\begin{aligned} \mathbf{N}^{(n)} &= \pi_n \int_{(0,t]} (G - \text{Id})(\text{ad } \kappa_{u-})(d\bar{\mathbf{m}}_u), \\ \mathbf{B}^{(n)} &= \pi_n \int_{(0,t]} (G - \text{Id})(\text{ad } \kappa_{u-})(d\bar{\mathbf{a}}_u) \end{aligned}$$

with $\bar{\mathbf{a}} = \pi_{(0,N)}(\mathbf{a}^{(1)} + \dots + \mathbf{a}^{(n-1)}) \in \mathcal{V}(\mathcal{T}_0^N)$ and $\bar{\mathbf{m}} = \pi_{(0,N)}(\mathbf{m}^{(1)} + \dots + \mathbf{m}^{(n-1)}) \in \mathcal{M}(\mathcal{T}_0^N)$.

From Lemma 7.1 and the generalized Hölder inequality, we have

$$\|\langle \mathbf{X}^c \rangle^{(n)}\|_{\mathcal{H}^{N/n}} \leq \sum_{i=1}^n \|\langle \mathbf{M}^{(i)c}, \mathbf{M}^{(n-i)c} \rangle\|_{\mathcal{H}^{N/n}} \lesssim \sum_{i=1}^n \|\mathbf{M}^{(i)}\|_{\mathcal{H}^{N/i}} \|\mathbf{M}^{(n-i)}\|_{\mathcal{H}^{N/(n-i)}} \lesssim \rho_{\mathbf{M},\mathbf{A}}^n. \tag{7.17}$$

It follows from equation (7.15) and the induction basis that for all $l \in \{1, \dots, n - 1\}$, it holds that

$$\left\| \boldsymbol{\kappa}^{(l)} \right\|_{\mathcal{H}^{N/l}} = \left\| \boldsymbol{\kappa}^{(l)} - \boldsymbol{\kappa}_0^{(l)} \right\|_{\mathcal{H}^{N/l}} \lesssim \rho_{\mathbf{M}, \mathbf{A}}^l, \tag{7.18}$$

and further that

$$\kappa_T^{l*} := \sup_{0 \leq t \leq T} |\boldsymbol{\kappa}_t^{(l)}|, \quad \left\| \boldsymbol{\kappa}^{(l)} \right\|_{\mathcal{S}^{N/l}} = \left\| \kappa_T^{l*} \right\|_{\mathcal{L}^{N/l}} \leq |\kappa_0^{(l)}| + \left\| \boldsymbol{\kappa}^{(l)} - \boldsymbol{\kappa}_0^{(l)} \right\|_{\mathcal{S}^{N/l}} \lesssim \rho_{\mathbf{M}, \mathbf{A}}^l. \tag{7.19}$$

From the definition and linearity of $Q(\text{ad } \mathbf{x})$ ($\mathbf{x} \in \mathcal{T}_0^N$), Lemmas 7.1 and 7.9, we have the following estimate:

$$\begin{aligned} & \left| \mathbf{V}^{(n)} \right|_{1\text{-var}; [0, T]} \\ & \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \sum_{m=2}^j \left| \int_0^{\cdot} \left(\text{ad } \boldsymbol{\kappa}_{u^-}^{(l_3)} \dots \text{ad } \boldsymbol{\kappa}_{u^-}^{(l_m)} \odot \text{ad } \boldsymbol{\kappa}_{u^-}^{(l_{m+1})} \dots \text{ad } \boldsymbol{\kappa}_{u^-}^{(l_j)} \right) \left(d \left[\mathbf{m}^{(l_1)c}, \mathbf{m}^{(l_2)c} \right]_u \right) \right|_{1\text{-var}; [0, T]} \\ & \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \int_0^T \left| \boldsymbol{\kappa}_{u^-}^{(l_3)} \dots \boldsymbol{\kappa}_{u^-}^{(l_j)} \right| d \left\langle \mathbf{m}^{(l_1)c}, \mathbf{m}^{(l_2)c} \right\rangle_u \\ & \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \kappa_T^{l_3*} \dots \kappa_T^{l_j*} \sqrt{[\mathbf{m}^{(l_1)}]_T} \sqrt{[\mathbf{m}^{(l_2)}]_T}. \end{aligned}$$

It then follows from the generalized Hölder inequality

$$\begin{aligned} \left\| \mathbf{V}^{(n)} \right\|_{\mathcal{V}^{N/n}} & \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \left\| \boldsymbol{\kappa}_T^{(l_3)} \right\|_{\mathcal{S}^{N/l_3}} \dots \left\| \boldsymbol{\kappa}_T^{(l_j)} \right\|_{\mathcal{S}^{N/l_j}} \left\| \mathbf{m}^{(l_1)} \right\|_{\mathcal{H}^{N/l_1}} \left\| \mathbf{m}^{(l_2)} \right\|_{\mathcal{H}^{N/l_2}} \\ & \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \rho_{\mathbf{M}, \mathbf{A}}^{l_1} \dots \rho_{\mathbf{M}, \mathbf{A}}^{l_j} \\ & \lesssim \rho_{\mathbf{M}, \mathbf{A}}^n, \end{aligned} \tag{7.20}$$

where the second inequality follows from the induction basis and the estimates in equations (7.19) and (7.15), noting that $\|\ell\| = n$ and $|\ell| \geq 2$ implies that $l_1, \dots, l_j \leq n - 1$, and the third inequality follows from equation (7.12). From similar arguments, we see that the following two estimates also hold:

$$\begin{aligned} \left\| \mathbf{C}^{(n)} \right\|_{\mathcal{V}^{N/n}} & \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \left\| \int_0^{\cdot} \left(\text{Id} \odot \text{ad } \boldsymbol{\kappa}_{u^-}^{(l_3)} \dots \text{ad } \boldsymbol{\kappa}_{u^-}^{(l_j)} \right) \left(d \left[\mathbf{M}^{(l_1)c}, \mathbf{m}^{(l_2)c} \right]_u \right) \right\|_{\mathcal{V}^{N/n}} \\ & \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \left\| \boldsymbol{\kappa}_T^{(l_3)} \right\|_{\mathcal{S}^{N/l_3}} \dots \left\| \boldsymbol{\kappa}_T^{(l_j)} \right\|_{\mathcal{S}^{N/l_j}} \left\| \mathbf{M}^{(l_1)} \right\|_{\mathcal{H}^{N/l_1}} \left\| \mathbf{m}^{(l_2)} \right\|_{\mathcal{H}^{N/l_2}} \\ & \lesssim \rho_{\mathbf{M}, \mathbf{A}}^n \end{aligned} \tag{7.21}$$

and

$$\left\| \mathbf{B}^{(n)} \right\|_{\mathcal{V}^{N/n}} \lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \left\| \boldsymbol{\kappa}_T^{(l_2)} \right\|_{\mathcal{S}^{N/l_2}} \dots \left\| \boldsymbol{\kappa}_T^{(l_j)} \right\|_{\mathcal{S}^{N/l_j}} \left\| \mathbf{a}^{(l_1)} \right\|_{\mathcal{V}^{N/l_1}} \lesssim \rho_{\mathbf{M}, \mathbf{A}}^n. \tag{7.22}$$

For the local martingale $\mathbf{N}^{(n)}$, we use Lemmas 7.1 and 7.9 to estimate its quadratic variation as follows:

$$\begin{aligned} \left\| \left[\mathbf{N}^{(n)} \right]_T \right\| &= \left\| \left[\sum_{\|\ell\|=n, |\ell| \geq 2} \int_{(0, \cdot]} \frac{1}{k!} \left(\text{ad } \kappa_{u-}^{(l_2)} \cdots \text{ad } \kappa_{u-}^{(l_k)} \right) \left(d\mathbf{m}_u^{(l_1)} \right) \right]_T \right\| \\ &\lesssim \sum_{\|\ell\|=2n, |\ell| \geq 4} \left\| \int_{(0, T]} \left(\text{ad } \kappa_{u-}^{(l_2)} \cdots \text{ad } \kappa_{u-}^{(l_m)} \odot \text{ad } \kappa_{u-}^{(l_{m+1})} \cdots \text{ad } \kappa_{u-}^{(l_k)} \right) \left(d \left[\mathbf{m}^{(l_1)}, \mathbf{m}^{(l_2)} \right]_u \right) \right\| \\ &\lesssim \sum_{\|\ell\|=2n, |\ell| \geq 4} \kappa_T^{l_3^*} \cdots \kappa_T^{l_j^*} \sqrt{[\mathbf{m}^{(l_1)}]_T} \sqrt{[\mathbf{m}^{(l_2)}]_T}. \end{aligned}$$

Then it follows once again by the generalized Hölder inequality and the induction basis that

$$\left\| \mathbf{N}^{(n)} \right\|_{\mathcal{H}^{N/n}} \lesssim \rho_{\mathbf{M}, \mathbf{A}}^n. \tag{7.23}$$

Finally, let us treat the term $\mathbf{J}^{(n)}$. First define

$$Z^l := \sup_{0 < u \leq T} \left(\max \left\{ \left| \Delta \mathbf{X}_u^{(l)} \right|, \left| \kappa_{u-}^{(l)} \right|, \left| \kappa_u^{(l)} \right| \right\} \right), \quad l = \{1, \dots, n-1\}.$$

And from equation (7.19), it follows that for all $l = \{1, \dots, n-1\}$, it holds that

$$\|Z^l\|_{\mathcal{L}^{N/l}} \leq 2 \|\mathbf{X}^{(l)}\|_{\mathcal{S}^{N/l}} + \|\kappa_T^{(l)}\|_{\mathcal{S}^{N/l}} \leq 2 \|\mathbf{X}^{(l)}\|_{\mathcal{H}^{N/l}} + \|\kappa_T^{(l)}\|_{\mathcal{S}^{N/l}} \lesssim \rho_{\mathbf{M}, \mathbf{A}}^{l, N}. \tag{7.24}$$

Then by Taylor’s theorem and Lemma 7.10, we have

$$\begin{aligned} \left| \mathbf{J}^{(n)} \right|_{1\text{-var}; [0, T]} &= \sum_{0 < u \leq T} \left| \pi_n \left(\exp_N(\Delta \mathbf{X}_u) \exp_N(\kappa_u) \exp_N(-\kappa_{u-}) - 1 - \Delta \mathbf{X}_u - G(\text{ad } \kappa_{u-})(\Delta \kappa_u) \right) \right| \\ &\lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} Z^{l_3} \cdots Z^{l_j} \sum_{0 < u \leq T} \left(\left| \Delta \mathbf{X}_u^{(l_1)} \right| + \left| \Delta \kappa_u^{(l_1)} \right| \right) \left(\left| \Delta \mathbf{X}_u^{(l_2)} \right| + \left| \Delta \kappa_u^{(l_2)} \right| \right) \\ &\lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} Z^{l_3} \cdots Z^{l_j} \left(\sqrt{[\mathbf{X}^{(l_1)}]_T} + \sqrt{[\kappa^{(l_1)}]_T} \right) \left(\sqrt{[\mathbf{X}^{(l_2)}]_T} + \sqrt{[\kappa^{(l_2)}]_T} \right). \end{aligned}$$

Hence it follows by the generalized Hölder inequality that

$$\begin{aligned} \left\| \mathbf{J}^{(n)} \right\|_{\mathcal{H}^{N/n}} &\lesssim \sum_{\|\ell\|=n, |\ell| \geq 2} \|Z^{l_3}\|_{\mathcal{L}^{N/l_3}} \cdots \|Z^{l_j}\|_{\mathcal{L}^{N/l_j}} \left(\left\| \mathbf{X}^{(l_1)} \right\|_{\mathcal{H}^{N/l_1}} + \left\| \kappa^{(l_1)} \right\|_{\mathcal{H}^{N/l_1}} \right) \\ &\quad \cdot \left(\left\| \mathbf{X}^{(l_2)} \right\|_{\mathcal{H}^{N/l_2}} + \left\| \kappa^{(l_2)} \right\|_{\mathcal{H}^{N/l_2}} \right) \\ &\lesssim \rho_{\mathbf{M}, \mathbf{A}}^n, \end{aligned} \tag{7.25}$$

where the last estimate follows from equations (7.24), (7.18) and (7.12).

Summarizing the estimates in equations (7.20), (7.21), (7.22), (7.23) and (7.25), we have

$$\begin{aligned} \left\| \mathbf{L}^{(n)} \right\|_{\mathcal{H}^{N/n}} &\lesssim \left\| \mathbf{M}^{(n)} \right\|_{\mathcal{H}^{N/n}} + \left\| \mathbf{N}^{(n)} \right\|_{\mathcal{H}^{N/n}} + \left\| \mathbf{A}^{(n)} \right\|_{\mathcal{H}^{N/n}} + \left\| \langle \mathbf{X}^c \rangle^{(n)} \right\|_{\mathcal{H}^{N/n}} \\ &\quad + \left\| \mathbf{B}^{(n)} \right\|_{\mathcal{H}^{N/n}} + \left\| \mathbf{V}^{(n)} \right\|_{\mathcal{H}^{N/n}} + \left\| \mathbf{C}^{(n)} \right\|_{\mathcal{H}^{N/n}} + \left\| \mathbf{J}^{(n)} \right\|_{\mathcal{H}^{N/n}} \\ &\lesssim \rho_{\mathbf{M}, \mathbf{A}}^n, \end{aligned} \tag{7.26}$$

which proves the first part of the induction claim in equation (7.14). Then it follows from dominated convergence theorem that projecting equation (7.10) to the tensor level n and passing to the $k \rightarrow \infty$ limit yields

$$\kappa_t^{(n)} = \mathbb{E}_t(\mathbf{L}_{T,t}^{(n)}), \quad 0 \leq t \leq T.$$

Since $\mathbf{M}^{(n)}$ and $\mathbf{N}^{(n)}$ are true martingales (for the latter, this follows from equation (7.23)), we are able to identify a decomposition $\kappa^{(n)} = \kappa_0^{(n)} + \mathbf{m}^{(n)} + \mathbf{a}^{(n)}$ by

$$\begin{aligned} \mathbf{a}^{(n)} &= -\left\{ \mathbf{A}^{(n)} + \frac{1}{2} \langle \mathbf{X}^c \rangle^{(n)} + \mathbf{B}^{(n)} + \mathbf{V}^{(n)} + \mathbf{C}^{(n)} + \mathbf{J}^{(n)} \right\} \\ \mathbf{m}_t^{(n)} &= \mathbb{E}(\mathbf{a}_T^{(n)}) - \mathbb{E}_t(\mathbf{a}_T^{(n)}), \quad 0 \leq t \leq T. \end{aligned}$$

Again from the estimates in equations (7.20), (7.21), (7.23) and (7.25), it follows that

$$\|\mathbf{a}^{(n)}\|_{\mathcal{F}^{N/n}} \lesssim \rho_{\mathbf{M}, \mathbf{A}}^n,$$

and in case $n \leq N - 1$, it follows from the BDG-inequality and Doob’s maximal inequality that

$$\|\mathbf{m}^{(n)}\|_{\mathcal{F}^{N/n}} \lesssim \|\mathbf{m}^{(n)}\|_{\mathcal{S}^{N/n}} \lesssim \|\mathbf{m}_T^{(n)}\|_{\mathcal{L}^{N/n}} = \|\mathbb{E}(\mathbf{a}_T^{(n)}) - \mathbf{a}_T^{(n)}\|_{\mathcal{L}^{N/n}} \lesssim \|\mathbf{a}^{(n)}\|_{\mathcal{F}^{N/n}} \lesssim \rho_{\mathbf{M}, \mathbf{A}}^n,$$

which proves the second part of the induction claim in equation (7.15). □

The estimate in equation (7.11) immediately follows from equations (7.13) and (7.14), which finishes the proof of Claim 7.13. □

Note that since $\mathbf{Y} = \mathbf{N} + \mathbf{B}$ and $\langle \mathbf{X}^c \rangle, \mathbf{V}, \mathbf{C}, \mathbf{J}$ are independent of the decomposition $\mathbf{X} = \mathbf{M} + \mathbf{A}$, it follows from taking the infimum over all such decompositions in the inequality in equation (7.26) that

$$\|\langle \mathbf{X}^c \rangle^{(n)}\|_{\mathcal{F}^{N/n}} + \|\mathbf{Y}^{(n)}\|_{\mathcal{F}^{N/n}} + \|\mathbf{V}^{(n)}\|_{\mathcal{F}^{N/n}} + \|\mathbf{C}^{(n)}\|_{\mathcal{F}^{N/n}} + \|\mathbf{J}^{(n)}\|_{\mathcal{F}^{N/n}} \lesssim \rho_{\mathbf{X}}^n \tag{7.27}$$

for all $n \in \{1, \dots, N\}$. The same argument applies to κ and the estimate in equation (7.18) and we obtain

$$\|\kappa^{(n)}\|_{\mathcal{F}^{N/n}} \lesssim \rho_{\mathbf{X}}^n, \tag{7.28}$$

for all $n \in \{1, \dots, N - 1\}$.

Next we are going to show that κ satisfies the functional equation (4.3). Recall that $\mathbf{L} + \kappa \in \mathcal{M}_{\text{loc}}(\mathcal{T}_0^N)$. From Lemma 7.11, we have the following equality

$$\int_{(0,t]} H(\text{ad } \kappa_{u-})(d(\mathbf{L}_u + \kappa_u)) = \kappa_t - \kappa_0 + \tilde{\mathbf{L}}_t$$

for all $0 \leq t \leq T$, where

$$\tilde{\mathbf{L}}_t = \int_{(0,t]} H(\text{ad } \kappa_{u-}) \left\{ d\mathbf{X}_u + \frac{1}{2} d\langle \mathbf{X}^c \rangle_u + d\mathbf{V}_u + d\mathbf{C}_u + d\mathbf{J}_u \right\}. \tag{7.29}$$

From Lemma 7.3 (Emery’s inequality) and the estimates in equations (7.27) and (7.28), it follows

$$\begin{aligned} \|\tilde{\mathbf{L}}^{(n)}\|_{\mathcal{H}^{N/n}} &\lesssim \sum_{\|\ell\|=n} \|\kappa^{(l_2)}\|_{\mathcal{S}^{N/l_2}} \cdots \|\kappa^{(l_j)}\|_{\mathcal{S}^{N/l_j}} \left\{ \|\mathbf{X}^{(l_1)}\|_{\mathcal{H}^{N/l_1}} + \|\langle \mathbf{X}^c \rangle^{(l_1)}\|_{\mathcal{H}^{N/l_1}} \right. \\ &\quad \left. + \|\mathbf{V}^{(l_1)}\|_{\mathcal{H}^{N/l_1}} + \|\mathbf{C}^{(l_1)}\|_{\mathcal{H}^{N/l_1}} + \|\mathbf{J}^{(l_1)}\|_{\mathcal{H}^{N/l_1}} \right\} \\ &\lesssim \rho_{\mathbf{X}}^n. \end{aligned}$$

Hence, by Lemma 7.4, it holds that

$$\|\tilde{\mathbf{L}}\|_{\mathcal{H}^{1,N}} \lesssim \|\mathbf{X}\|_{\mathcal{H}^{1,N}}. \tag{7.30}$$

Now note that we have already shown in Claim 7.13 that $\kappa = \kappa_0 + \mathbf{m} + \mathbf{a}$, where $\mathbf{m} \in \mathcal{M}(\mathcal{T}_0^N)$ and $\mathbf{a} \in \mathcal{V}(\mathcal{T}_0^N)$, which satisfies that $\|\mathbf{a}^{(n)}\|_{\mathcal{H}^{N/n}} < \infty$ for all $n \in \{1, \dots, N\}$. Together with the above estimate, it then follows that $\kappa + \tilde{\mathbf{L}}$ is indeed a true martingale and therefore

$$\kappa_t = \mathbb{E}\left(\tilde{\mathbf{L}}_{T,t}\right), \quad 0 \leq t \leq T,$$

which is precisely the identity equation (4.3). □

Acknowledgments. The authors thank the anonymous referees for the careful reading.

Funding statement. PKF has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 683164) and the DFG Research Unit FOR 2402. PH and NT have been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689).

Conflicts of Interest. None

References

[1] E. Abi Jaber, M. Larsson, and S. Pulido, *Affine volterra processes*, *Ann. Appl. Probab.* **29** (2019), no. 5, 3155–3200.
 [2] Y. Ait-Sahalia and J. Jacod, *High-frequency financial econometrics*, Princeton University Press, 2014.
 [3] E. Alos, J. Gatheral, and R. Radovićić, *Exponentiation of conditional expectations under stochastic volatility*, *Quantitative Finance; SSRN* (2017) **20** (2020), no. 1, 13–27.
 [4] D. Applebaum, *Lévy processes and stochastic calculus*, Cambridge University Press, 2009.
 [5] I. P. Arribas, C. Salvi, and L. Szpruch, *Sig-sdes model for quantitative finance*, 2020, [arXiv:2006.00218](https://arxiv.org/abs/2006.00218) [q-fin.CP], 1st ACM International Conference on AI in Finance (ICAIF 2020).
 [6] S. Blanes, F. Casas, J. Oteo, and J. Ros, *The magnus expansion and some of its applications*, *Phys. Rep.* **470** (2009), no. 5-6, 151–238.
 [7] P. Bonnier and H. Oberhauser, *Signature cumulants, ordered partitions, and independence of stochastic processes*, *Bernoulli* **26** (2020), no. 4, 2727–2757.
 [8] Y. Bruned, C. Curry, and K. Ebrahimi-Fard, *Quasi-shuffle algebras and renormalisation of rough differential equations*, *Bulletin of the London Mathematical Society* **52** (2020), no. 1, 43–63.
 [9] F. Casas and A. Murua, *An efficient algorithm for computing the Baker-Campbell-Hausdorff series and some of its applications*, *J. Math. Phys.* **50** (2009), no. 3, 033513, 23.
 [10] A. Celestino, K. Ebrahimi-Fard, F. Patras, and D. Perales, *Cumulant–cumulant relations in free probability theory from Magnus’ expansion*, (2021).
 [11] K.-T. Chen, *Iterated integrals and exponential homomorphisms†*, *Proc. London Math. Soc.* **s3-4** (1954), no. 1, 502–512.
 [12] I. Chevyrev and P. K. Friz, *Canonical rdes and general semimartingales as rough paths*, *Ann. Probab.* **47** (2019), no. 1, 420–463.
 [13] I. Chevyrev and T. Lyons, *Characteristic functions of measures on geometric rough paths*, *Ann. Probab.* **44** (2016), no. 6, 4049–4082.
 [14] S. Cohen and R. J. Elliott, *Stochastic calculus and applications*, 2nd ed., Birkhäuser, Basel, 2015.
 [15] R. Cont and P. Tankov, *Financial modelling with jump processes*, Financial Mathematics Series, Chapman & Hall/CRC, 2004.

- [16] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann, *Affine processes on positive semidefinite matrices*, *The Annals of Applied Probability* **21** (2011), no. 2, 397–463.
- [17] C. Cuchiero, M. Keller-Ressel, and J. Teichmann, *Polynomial processes and their applications to mathematical finance*, *Finance and Stochastics* **16** (2009).
- [18] C. Cuchiero, S. Svaluto-Ferro, and J. Teichmann, *Signature sdes from an affine and polynomial perspective*, 2021, In preparation.
- [19] D. Duffie, D. Filipović, W. Schachermayer, et al., *Affine processes and applications in finance*, *Ann. Appl. Probab.* **13** (2003), no. 3, 984–1053.
- [20] A. Estrade, *Exponentielle stochastique et intégrale multiplicative discontinues*, *Ann. Inst. Henri Poincaré Probab. Stat.* **28** (1992), no. 1, 107–129.
- [21] T. Fawcett, *Problems in stochastic analysis: connections between rough paths and non-commutative harmonic analysis*, Ph.D. thesis, University of Oxford, 2002.
- [22] A. Friedman, *Partial differential equations of parabolic type*, R.E. Krieger Publishing Company, 1983.
- [23] P. K. Friz, J. Gatheral, and R. Radoičić, *Forests, cumulants, martingales*, to appear in *The Annals of Probability* (2022).
- [24] P. K. Friz and M. Hairer, *A course on rough paths*, 2nd ed., Universitext, Springer International Publishing, 2020.
- [25] P. K. Friz and A. Shekhar, *General rough integration, Lévy rough paths and a Lévy–Kintchine-type formula*, *Ann. Probab.* **45** (2017), no. 4, 2707–2765.
- [26] P. K. Friz and N. B. Victoir, *The burkholder-davis-gundy inequality for enhanced martingales*, *Lecture Notes in Mathematics* **1934** (2006).
- [27] P. K. Friz and N. B. Victoir, *Multidimensional stochastic processes as rough paths: Theory and applications*, *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, 2010.
- [28] M. Fukasawa and K. Matsushita, *Realized cumulants for martingales*, *Electronic Communications in Probability* **26** (2021), 1–10.
- [29] J. Gatheral and M. Keller-Ressel, *Affine forward variance models*, *Finance Stoch.* **23** (2019), no. 3, 501–533.
- [30] M. Hairer, *Solving the kpz equation*, *Annals of Mathematics* **178** (2013), no. 2, 559–664.
- [31] M. Hakim-Dowek and D. Lépingle, *L'exponentielle stochastique des groupes de Lie*, Séminaire de Probabilités XX 1984/85, Springer, 1986, pp. 352–374.
- [32] F. Hausdorff, *Die symbolische Exponentialformel in der Gruppentheorie*, *Ber. Verh. Kgl. Sächs. Ges. Wiss. Leipzig., Math.-phys. Kl.* **58** (1906), 19–48.
- [33] A. Iserles and S. P. Nørsett, *On the solution of linear differential equations in lie groups*, *Philos. Trans. Roy. Soc. A* **357** (1999), no. 1754, 983–1019.
- [34] A. Iserles, H. Munthe-Kaas, S. Nørsett, and A. Zanna, *Lie-group methods*, *Acta numerica* (2005).
- [35] J. Jacod, *Calcul stochastique et problèmes de martingales*, *Lecture Notes in Mathematics*, vol. **714**, Springer Berlin Heidelberg, Berlin, Heidelberg, 1979 (eng).
- [36] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes.*, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, no. **488**, Springer Berlin, 2003.
- [37] K. Kamm, S. Pagliarani, and A. Pascucci, *The stochastic magnus expansion*, *Journal of Scientific Computing* **89** (2021), 56. <https://doi.org/10.1007/s10915-021-01633-6>.
- [38] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, 2 ed., *Graduate Texts in Mathematics*, vol. **113**, Springer, New York, NY, 1998.
- [39] M. Keller-Ressel, M. Larsson, and S. Pulido, *Affine Rough Models*, arXiv e-prints (2018), [arXiv:1812.08486](https://arxiv.org/abs/1812.08486).
- [40] M. Keller-Ressel, W. Schachermayer, and J. Teichmann, *Affine processes are regular*, *Probab. Theory Related Fields* **151** (2011), no. 3–4, 591–611.
- [41] T. G. Kurtz, E. Pardoux, and P. Protter, *Stratonovich stochastic differential equations driven by general semimartingales*, *Ann. Inst. Henri Poincaré Probab. Stat.* **31** (1995), no. 2, 351–377.
- [42] H. Lacoïn, R. Rhodes, and V. Vargas, *A probabilistic approach of ultraviolet renormalisation in the boundary sine-gordon model*, to appear in *Probability Theory and Related Fields* (2022).
- [43] J.-F. Le Gall, *Brownian motion, martingales, and stochastic calculus*, Springer, 2016.
- [44] Y. LeJan and Z. Qian, *Stratonovich's signatures of brownian motion determine brownian sample paths*, *Probability Theory and Related Fields* **157** (2011).
- [45] T. Lyons, *Rough paths, signatures and the modelling of functions on streams*, *Proceedings of the International Congress of Mathematicians—Seoul 2014*. Vol. **IV**, Kyung Moon Sa, Seoul, 2014, pp. 163–184.
- [46] T. Lyons and H. Ni, *Expected signature of brownian motion up to the first exit time from a bounded domain*, *Ann. Probab.* **43** (2015), no. 5, 2729–2762.
- [47] T. Lyons and N. Victoir, *Cubature on wiener space*, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **460** (2004), no. 2041, 169–198.
- [48] W. Magnus, *On the exponential solution of differential equations for a linear operator*, *Commun. Pure Appl. Math.* **7** (1954), no. 4, 649–673.
- [49] S. I. Marcus, *Modeling and analysis of stochastic differential equations driven by point processes*, *IEEE Trans. Inform. Theory* **24** (1978), no. 2, 164–172.

- [50] S. I. Marcus, *Modeling and approximation of stochastic differential equations driven by semimartingales*, *Stochastics* **4** (1981), no. 3, 223–245.
- [51] H. P. McKean, *Stochastic integrals*, AMS Chelsea Publishing Series, no. **353**, American Mathematical Society, 1969.
- [52] W. Miller, Jr., *Symmetry groups and their applications*, Pure and Applied Mathematics, vol. **50**, Academic Press, New York-London, 1972.
- [53] P. A. Mykland, *Bartlett type identities for martingales*, *Ann. Statist.* **22** (1994), no. 1, 21–38.
- [54] H. Ni, *The expected signature of a stochastic process*, Ph.D. thesis, University of Oxford, 2012.
- [55] B. Øksendal, *Stochastic differential equations: An introduction with applications*, 6 ed., Springer Berlin / Heidelberg, Berlin, Heidelberg, 2014 (eng).
- [56] H. Pham, *Optimal stopping of controlled jump diffusion processes: A viscosity solution approach*, *J. Math. Syst. Est. Control* **8** (1998), 1–27.
- [57] P. E. Protter, *Stochastic integration and differential equations*, 2 ed., Stochastic Modelling and Applied Probability, Springer-Verlag Berlin Heidelberg, 2005.
- [58] C. Reutenauer, *Free lie algebras*, *Handbook of Algebra*, vol. **3**, Elsevier, 2003, pp. 887–903.
- [59] D. Revuz and M. Yor, *Continuous martingales and brownian motion*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 2004.
- [60] D. W. Stroock, *Diffusion processes associated with lévy generators*, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **32** (1975), no. 3, 209–244.
- [61] D. W. Stroock and S. R. S. Varadhan, *Diffusion processes with continuous coefficients, i*, *Communications on Pure and Applied Mathematics* **22** (1969), no. 3, 345–400.
- [62] L. C. Young, *An inequality of the hölder type, connected with stieljes integration*, *Acta Math.* **67** (1936), 251–282.