On the Wiener Chaos Expansion of the Signature of a Gaussian Process

Thomas Cass[1] and Emilio Ferrucci[*2]

[1]Department of Mathematics, Imperial College London
[2]Mathematical Institute, University of Oxford

December 14, 2023

Abstract

We compute the Wiener chaos decomposition of the signature for a class of Gaussian processes, which contains fractional Brownian motion (fBm) with Hurst parameter $H \in (1/4, 1)$. At level 0, our result yields an expression for the expected signature of such processes, which determines their law [13]. In particular, this formula simultaneously extends both the one for $1/2 < H$-fBm [3] and the one for Brownian motion ($H = 1/2$) [16], to the general case $H > 1/4$, thereby resolving an established open problem. Other processes studied include continuous and centred Gaussian semimartingales.

Acknowledgements

Introduction

The signature of a path $X \colon [0, T] \to \mathbb{R}^d$,

$$\mathbb{S}(X)_{0T} \coloneqq \sum_{n=0}^{\infty} \int_{0 < u_1 < \ldots < u_n < T} \mathrm{d}X_{u_1} \otimes \cdots \otimes \mathrm{d}X_{u_n} \quad \in T((\mathbb{R}^d)), \tag{1}$$

is a series of tensors which, up to "retracings", determines the image of X [22, 6]. The probabilistic counterpart to this result states that, in many cases of interest, the law of a stochastic process is determined by its expected signature [13], which is therefore seen to play a role for processes analogous to that of moments for random variables.

The best-known example of an explicit formula for the expected signature of a stochastic process occurs in the case of Brownian motion: calling $\{e_1, \ldots, e_d\}$ the canonical basis of $\mathbb{R}^d$, we have

$$\mathbb{E}\mathbb{S}(X)_{st} = \exp\left(\frac{t-s}{2} \sum_{\gamma=1}^{d} e_\gamma^{\otimes 2}\right) = \sum_{n=0}^{\infty} \frac{(t-s)^n}{2^n n!} \sum_{\gamma_1, \ldots, \gamma_n = 1}^{d} e_{\gamma_1}^{\otimes 2} \otimes \cdots \otimes e_{\gamma_n}^{\otimes 2}. \tag{2}$$

---

[*]Corresponding author: Emilio.RossiFerrucci@maths.ox.ac.uk

This identity was first shown by [16, 31], and later proved in a variety of different ways [2, 20]. The expected signature of Brownian motion has also been studied in the case in which the process is stopped upon hitting the boundary of a domain [29, 5, 27].

In [3] the authors derive an integral expression for the expected signature of fractional Brownian motion (fBm) with Hurst parameter $H \in (1/2, 1)$. This result was extended in [7, 4] to a more general class of Gaussian Volterra processes with sample paths that are more regular than Brownian motion, with the formula for the expected signature written in terms of the Volterra kernel. The method used involves a piecewise-linear interpolation of the paths of the process X, which reduces the calculation to that of a sum of mixed Gaussian moments, to which Wick's theorem applies, followed by a convergence argument. The expression in [3] does not, however, yield the correct prediction for the case of Brownian motion $H = 1/2$. When $H < 1/2$ it involves integrals that do not converge at all, and new ideas are needed to obtain a formula. On a technical level, the reason for these differences can be seen by considering the expression for the expected signature of a scalar $1/2 < H$-fBm X at level 2: calling $R(s, t) := \mathbb{E}[X_s X_t]$ the covariance function of X, the formula states that

$$\mathbb{E}\mathbb{S}(X)_{st}^{(2)} = \int_{s<u<v<t} R(\mathrm{d}u, \mathrm{d}v) = H(2H-1) \int_{s<u<v<t} (v-u)^{2H-2} \mathrm{d}u \mathrm{d}v. \tag{3}$$

Integrating either of the two variables generates an evaluation $(v-u)^{2H-1}|_{u=v}$, which is only finite when $H > 1/2$ and indeterminate when $H = 1/2$. In fact, approximating X with a sequence of piecewise linear processes $(X^\ell)_{\ell \in \mathbb{N}}$ one obtains a sequence of integrals (actually finite sums) $\int_{s<u<v<t} \mathbb{E}[\dot{X}_u^\ell \dot{X}_v^\ell] \mathrm{d}u \mathrm{d}v$ which converges to the above double integral when $H > 1/2$, to $(t-s)/2$ when $H = 1/2$ (as predicted by (2)), and continues to converge to $(t-s)^{2H}/2$ for $1/4 < H \leq 1/2$. When $H \leq 1/4$ the iterated integrals (in particular the Lévy area) of smooth approximations of X do not converge in mean square, and other techniques (e.g. [36]) must be relied upon to define a rough path, and hence a signature. These rough paths present a number of differences with the canonical one defined for $H > 1/4$, and are therefore not considered in this paper.
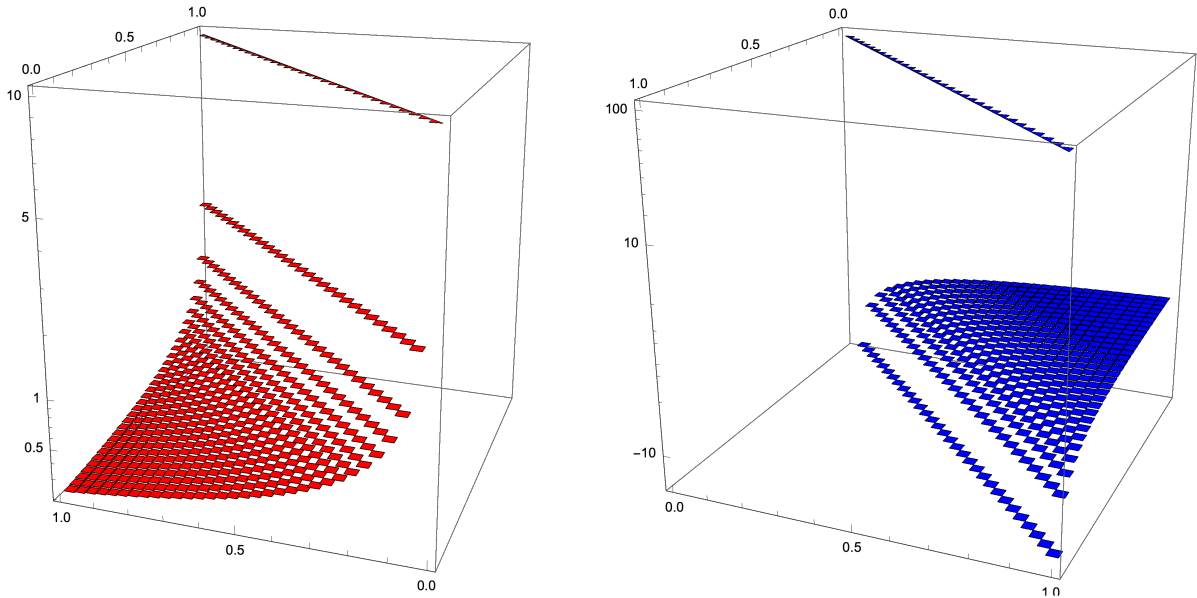

Figure 1: Here we compare the two behaviours, corresponding to $H > 1/2$ and $H < 1/2$, of $\int_{0<u<v<1} \mathbb{E}[\dot{X}_u^\ell \dot{X}_v^\ell] \mathrm{d}u \mathrm{d}v$ with X^ℓ the sequence piecewise linear interpolations of X on a partition. On the left we have chosen $H = 2/3$, and the sequence of integrals converges to a finite improper integral, whereas on the right $H = 1/3$ and the on- and off-diagonal contributions diverge to opposite infinities. (The plots are oriented in different ways and the z-axis is rescaled, both for improved visibility.) This graphic has been created using Wolfram Mathematica.

What is needed to obtain a formula for the expected signature that also works in the case of negatively-correlated increments $1/4 < H < 1/2$ is a way of expressing the indeterminacy "$\infty - \infty$" explained in Figure 1. The trick for doing this is simple to describe: integrate out the first variable in (3) and, calling $R(t) := R(t,t)$ the variance function of X, note that for $H > 1/2$ we have

$$\int_{s<u<v<t} \partial_{12}R(u,v)\mathrm{d}u\mathrm{d}v = \int_s^t \left[\partial_2 R(v,v) - \partial_2 R(s,v)\right]\mathrm{d}v = \int_s^t \left[\tfrac{1}{2}R'(v) - \partial_2 R(s,v)\right]\mathrm{d}v. \quad (4)$$

We have replaced $\partial_2 R(v,v)$ with $\tfrac{1}{2}R'(v)$, which can be done by symmetry of R:

$$R'(v) = \frac{\mathrm{d}}{\mathrm{d}v}R(v,v) = \partial_1 R(v,v) + \partial_2 R(v,v) = 2\partial_2 R(v,v). \quad (5)$$

This is relevant to the case of $(1/4, 1/2) \ni H$-fBm since, while $\partial_2 R(v,v)$ or $\partial_1 R(v,v)$ is the infinite evaluation discussed earlier, the last integral in (4) is perfectly well defined. These integrands can be chained together on simplices, e.g. $\int_{s<u<v<t}[\tfrac{1}{2}R'(u) - \partial_2 R(s,u)][\tfrac{1}{2}R'(v) - \partial_2 R(u,v)]\mathrm{d}u\mathrm{d}v$, and combined with the other types of integrand $\partial_{12}R(w,z)$, to yield a formula that is very similar to that of [3], but continues to be convergent for $1/4 < H < 1/2$ and agrees with (2) for $H = 1/2$.

Showing that the formula obtained by such substitution actually coincides with the expected signature for X in a broad class of Gaussian processes — essentially those Gaussian rough paths introduced in [15, 30, 19] with the imposition of a few additional smoothness and regularity requirements on the (co)variance function — is the main focus of this paper. In fact, our main result will prove a formula for the full Wiener chaos expansion of $\mathcal{S}(X)$, the 0th level of which is the expectation. As far as we know, the expression for the positive chaos projections of the signature is not to be found in the literature even in the classical case of Brownian motion. While the expression of the positive levels of Wiener chaos is very similar in spirit to that of the 0th, it requires us to use some Malliavin calculus in the setting of 1-parameter Gaussian processes, and results in technical complications in the proof of convergence. The main additional ingredients needed are Stroock's formula for the m-th Wiener chaos projection and a novel definition of multiple Wiener integral of a function. For the latter, it should be noted that while multivariate, deterministic integrands for Gaussian noise naturally live in a certain Hilbert space (which for fBm can be identified with a Sobolev space), we are interested in integrating functions of multiple times, i.e. $\int_{[0,T]^m} f(t_1,\ldots,t_m)\mathrm{d}X_{t_1}^{\gamma_1}\cdots\mathrm{d}X_{t_m}^{\gamma_m}$ in a Skorokhod-type sense: this is achieved by approximating f with elementary integrands, and showing independence of the approximation. Computing the Wiener chaos projections of the signature of a Gaussian process X has the benefit of expressing $\mathcal{S}(X)$ as a sum of terms that are orthogonal in L^2, something that has the potential to be used for various types of numerical calculations, e.g. estimates of Euler expansions for Gaussian rough differential equations. It should be mentioned that, while (in the cases considered) the expected signature already determines the law of X and therefore that of the Wiener chaos projections of $\mathcal{S}(X)$, it does not appear obvious how one may obtain the latter from the former directly. While fBm is the main example of a process for which our calculation is novel, we briefly also consider centred, continuous Gaussian semimartingales, such as the Brownian bridge returning to the origin and centred Ornstein-Uhlenbeck processes with deterministic initial condition.

As in the main reference article [3], the technique that underlies our proof is piecewise-linear approximation of X. The arguments needed to prove the result are however much more involved, for three essential reasons. First is the fact that we must perform and justify the substitution (4), which requires novel arguments for convergence; even proving finiteness of the integrals in the main formula requires more sophisticated bounds in the $1/4 < H < 1/2$ case than it does in the $H > 1/2$ case (see Figure 2 for the simplest example of an observation that must be made when $H < 1/2$). Second is that Malliavin derivatives are involved for positive levels of the Wiener chaos and third is that our arguments must accommodate a wider class of Gaussian processes. While the substitution (4) may seem very natural, it does not emerge obviously from the proof that we have given here, and must instead be guessed in advance. Indeed, it is worth mentioning that the way in which we first derived the statement of the main result involved an entirely different approach, which made use of the Skorokhod-rough integral conversion formula [10, 11], applied recursively to the RDE for the signature.
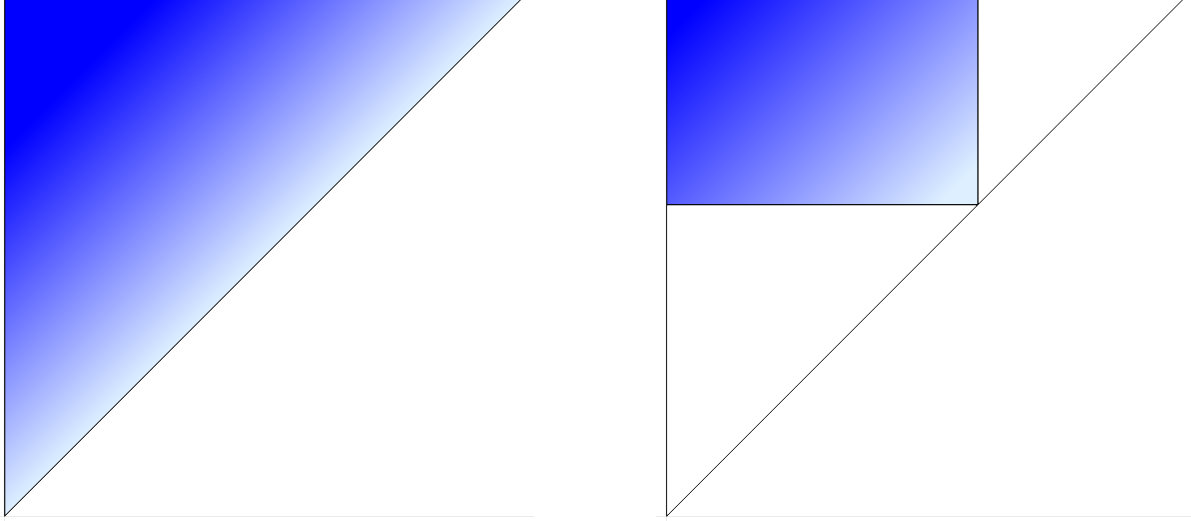
3

Figure 2: A graphic representing the contour plot of $(t - s)^{2H-2}$ on $\{0 < s < t < 1\}$ (on the left) and $\{0 < s < u < t < 1\}$ with $u \in [0, 1]$ fixed (on the right): the integral of the former is improper on the whole diagonal, while that of the later only at a point: when $0 < H < 1/2$, only the latter converges. This graphic has been created using Wolfram Mathematica.

The outline of this proof can be found in the second named author's PhD thesis [17, Ch. 5]. While this approach has the drawback of generating further technical problems, reason for which it is not the one presented here, it has the advantage of leading up constructively to the main formula.

This paper is organised as follows: in Section 1 we briefly introduce the class of Gaussian processes considered and the Malliavin calculus framework for them; we then use this language to identify functions as multiple Wiener integrands. In Section 2 we state the main result Theorem 2.3 and discuss a few consequences and examples that follow; in Section 3 we prove the main result; in Conclusions and further directions we outline some aspects that could be tackled in further research. Finally, it should be mentioned that in [3], in addition to the expected signature of $1/2 < H$-fBm, the authors also compute the expected signature at levels 2 and 4 for $1/4 < H$-fBm in a manner that does not obviously generalise to different processes or higher levels; while not necessary in our proofs, it is sensible to verify that our main result agrees with this calculation: this check is performed in Appendix A.

1 Background on Malliavin calculus for Gaussian processes

In this section we introduce the class of Gaussian processes to which this paper applies, establish some notation, and give a brief overview of the tools of Malliavin calculus that are necessary in the proof of the main result. We follow [35, 34] for the general Malliavin calculus framework, [23] for its aspects that pertain to Gaussian processes indexed by a time parameter, and [9, 10, 11] for aspects regarding the rough path lifts of such processes.

Throughout this paper we will be working with a Gaussian process with i.i.d. components $X \colon \Omega \times [0, T] \to \mathbb{R}^d$ where $\Omega = C([0, T], \mathbb{R}^d)$, $X_t(\omega) := \omega(t)$, $\mathcal{F}_t := \sigma(X_s : 0 \leq s \leq t)$. We assume X to be centred, i.e. $\mathbb{E}X \equiv 0$, and for it to have deterministic initial condition $X_0 = 0$. We will write $X_{st} := X_t - X_s$ for the increments of X. By Gaussianity, the probability measure $\mathbb{P}$ on Ω is characterised by the covariance function of X

$$R \colon [0, T]^2 \to \mathbb{R}^d \otimes \mathbb{R}^d, \quad R(s, t) := \mathbb{E}[X_s \otimes X_t]. \tag{6}$$

We will denote $R(\,\cdot\,)$ the variance function of X, i.e. $R(t) := R(t, t)$. The independence hypothesis implies that R is a diagonal matrix $R^{\alpha\beta} = \delta^{\alpha\beta} R^{\alpha\alpha}$, and the fact that they are identically distributed, $R^{\alpha\beta} = \delta^{\alpha\beta} R^{11}$

4

will be determined by a single scalar function, which by abuse of notation we will also call R. Although our results can be conjectured to continue to hold in the case in which the components are not identically distributed, our proof will make essential use of this assumption. We define

$$R(\Delta(s,t)) := R(t) - R(s)$$
$$R(\Delta(s,t), v) := R(t,v) - R(s,v) = \mathbb{E}[X_{st} \otimes X_v]$$
$$R(\Delta(s,t), \Delta(u,v)) := R(t,v) + R(s,u) - R(t,u) - R(s,v) = \mathbb{E}[X_{st} \otimes X_{uv}]$$

(7)

for $u, v, s, t \in [0, T]$. Note that $R(\Delta(s,t)) \neq R(\Delta(s,t), \Delta(s,t))$.

We assume X and R satisfy the conditions that make it possible to consider the *signature* of X, $\mathcal{S}(X)$, defined by the limit in L^2 of Stieltjes iterated integrals of smooth or piecewise-linear approximations of X, and carry out Malliavin calculus: these are existence of rough path lift and complementary Cameron-Martin regularity [9, Conditions 2] and non-degeneracy of R [9, Conditions 3]. More elementary conditions that imply these may be found, for instance, in [10, 11]. The expected signatures of such processes characterise their law, i.e. if Y is any other process with a well-defined signature $\mathcal{S}(Y)_{0T}$ (as a $\mathcal{G}(\mathbb{R}^d)$-valued random variable) and $\mathbb{E}\mathcal{S}(X)_{0T} = \mathbb{E}\mathcal{S}(Y)_{0T}$, then X and Y are equal in law: see [13, Example 6.7], a consequence, among other things, of the greedy estimate [12]. We refer the reader to [14] for a treatment of the theory in the case of more general processes, whose expected signatures may not directly characterise the law of the process.

We will denote $\mathcal{S}^N(X)$ the signature of X truncated at level N (i.e. its projection onto $\bigoplus_{n=0}^{N}(\mathbb{R}^d)^{\otimes n}$) and $\mathcal{S}(X)^{(n)}$ the n-th level of the signature (i.e. its projection onto $(\mathbb{R}^d)^{\otimes n}$). The signature of a process, as that of a path, satisfies two important algebraic relations. The first is the Chen identity, namely that $\mathcal{S}(X)_{su} \otimes \mathcal{S}(X)_{ut} = \mathcal{S}(X)_{st}$. The second is the shuffle identity: letting $\{e_1, \ldots, e_d\}$ denote the canonical basis of $\mathbb{R}^d$, and using coordinate notation, i.e. $S^{\gamma_1 \ldots \gamma_n} := \langle e_{\gamma_1} \otimes \cdots \otimes e_{\gamma_n}, S \rangle$ for $S \in T((\mathbb{R}^d))$ and $\gamma_1, \ldots, \gamma_n \in [d] := \{1, \ldots, d\}$ (and extending linearly), for $0 \leq s \leq t \leq T$ it holds that

$$\mathcal{S}(X)_{st}^{\alpha_1 \ldots \alpha_m} \mathcal{S}(X)_{st}^{\beta_1 \ldots \beta_n} = \mathcal{S}(X)_{st}^{(\alpha_1 \ldots \alpha_m) \shuffle (\beta_1 \ldots \beta_n)}$$

(8)

where $\shuffle$ denotes "shuffling" the tuples $\alpha_1 \ldots \alpha_m$ and $\beta_1 \ldots \beta_n$, i.e. summing over all ways of permuting their concatenation $\alpha_1 \ldots \alpha_m \beta_1 \ldots \beta_n$ whilst preserving the order of each. For further details see, for example, [28].

In addition to the standard conditions on R, we will have to assume a certain amount of smoothness of R together with bounds on its derivatives; the reasons for such hypotheses will be made clear in due course. We assume $R(\cdot, \cdot)$ is C^2 on the open simplex $\Delta[s, t] := \{0 < s < t < T\}$ and continuous on $[0, T]^2$, and that $R(\cdot)$ is C^1 on $(0, T)$. The lack of smoothness assumptions of $R(\cdot, \cdot)$ on the diagonal $\{s = t\}$ is crucial for the inclusion of $(1/4, 1/2] \ni H$-fBm, which does not even have first partial derivatives on it. Furthermore, we assume there exists an $H \in (0, 1)$ with the property that the sample paths of X are either H-Hölder, or are K-Hölder for all $K < H$; for fBm H will coincide with the Hurst parameter, but the letter H will be used for more general processes to denote the Hölder exponent/supremum of exponents. This the rough path above X will be of finite $1/H$-variation or of finite p-variation for all $p > 1/H$.

We also need some quantitative estimates on the derivatives of R. Here and throughout the paper, the constant of proportionality implied by the use of $\lesssim$ may only depend on T, H and other general characteristics of the process X. We require

$$|\partial_{12}R(s,t)| \lesssim (t-s)^{2H-2} \qquad \text{on } \Delta[0, T]$$

(9)

$$\left|\tfrac{1}{2}R'(t) - \partial_2 R(s,t)\right| \lesssim (t-s)^{2H-1} \qquad \text{on } [0, T]^2 \setminus \{s = t\}$$

(10)

$$|R'(t)| \lesssim t^{2H-1} \qquad \text{on } (0, T)$$

(11)

where ∂_2 denotes partial differentiation w.r.t. the second component and ∂_{12} denotes second-order mixed partial differentiation. Since R is not smooth on the diagonal, the following estimate for on-diagonal square increments of the covariance function, which already appeared in [15], must be required separately:

$$R(\Delta(s,t), \Delta(s,t)) \lesssim (t-s)^{2H}$$

(12)

We move onto the treatment of Malliavin calculus for X. We let $\mathcal{H}$ be the Hilbert space given by the completion of the following $\mathbb{R}$-linear span of *elementary* functions $[0,T] \to \mathbb{R}^d$, or equivalently $[0,T] \times [d] \to \mathbb{R}$:

$$\mathcal{E} := \mathrm{span}_{\mathbb{R}}\{\mathbb{1}_{[0,t)}^{\gamma} \mid t \in [0,T],\ \gamma = 1,\ldots,d\} \tag{13}$$

w.r.t. the inner product

$$\langle \mathbb{1}_{[0,s)}^{\alpha}, \mathbb{1}_{[0,t)}^{\beta} \rangle_{\mathcal{H}} := R^{\alpha\beta}(s,t)\,. \tag{14}$$

Because of independence of components, $\mathcal{H}$ is equal to an orthogonal direct sum $\mathcal{H}^1 \oplus \ldots \oplus \mathcal{H}^d$, and because of equal distribution the direct summands are all equal. Elements of $\mathcal{H}$ should be viewed as admissible deterministic integrands for $\mathrm{d}X$, which are represented as Cauchy sequences of elementary integrands in $\mathcal{E}$. This framework allows us to view the process as an isometry

$$X : \mathcal{H} \to L^2\Omega, \quad \mathbb{1}_{[0,t)}^{\gamma} \mapsto X_t^{\gamma} \tag{15}$$

often called an isonormal Gaussian process.

The *multiple Wiener integral*

$$\delta^m : \mathcal{H}^{\odot m} \to L^2\Omega \tag{16}$$

is the operator defined by the adjoint property (which more generally characterises the divergence operator, when required on random arguments f)

$$\forall Z \in \mathbb{D}^{m,2} \quad E[Z\delta^m(f)] = \mathbb{E}[\langle \mathcal{D}^m Z, f\rangle_{\mathcal{H}^{\otimes m}}] \tag{17}$$

where

$$\mathcal{D}^m : \mathbb{D}^{m,2} \to L^2(\Omega, \mathcal{H}^{\odot m}) \tag{18}$$

is the mth Malliavin derivative, defined as

$$\mathcal{D}^m f(X_{t_1}^{\gamma_1}, \ldots, X_{t_n}^{\gamma_n}) := \sum_{k_1,\ldots,k_m=1}^{n} \partial_{k_1,\ldots,k_m} f(X_{t_1}^{\gamma_1}, \ldots, X_{t_n}^{\gamma_n}) \mathbb{1}_{[0,t_{k_1})}^{\gamma_{k_1}} \otimes \cdots \otimes \mathbb{1}_{[0,t_{k_n})}^{\gamma_{k_n}} \tag{19}$$

for $f \in C^\infty(\mathbb{R}^n)$ with derivatives (including the 0th) of polynomial growth, and extended as a closed operator to a certain domain $\mathbb{D}^{m,2}$. $\mathcal{H}^{\odot m}$ denotes the subspace of $\mathcal{H}^{\otimes m}$ (the tensor product taken in the category of Hilbert spaces) of symmetric tensors. $\mathcal{D}^m$ takes a square-integrable random variable and returns a random element of $\mathcal{H}^{\odot m}$, which in case of membership to $\mathcal{E}^{\odot m}$ (or otherwise a function member of $\mathcal{H}^{\odot m}$ in the sense of Definition 1.1 below) will be a function of m (time,index) pairs. Note that, while δ is symmetric in the sense that it is left invariant by permuting (time,index) jointly, it is not symmetric if only time variables or indices are permuted (e.g. it is possible to use δ to define a Lévy area — see Example 2.6 below). When $\mathcal{D}^m Z$ is a function, as in the case (19), we denote its evaluation on m (time,index) pairs $\mathcal{D}_{(u_1,\gamma_1),\ldots,(u_m,\gamma_m)}Z$; occasionally it may make more sense to suppress the indices in the notation, in which case we can just write $\mathcal{D}_{u_1,\ldots,u_m}Z$. We may extend δ to a map $\delta^m := \mathcal{H}^{\otimes m} \to L^2\Omega$ by pre-composing with symmetrisation, and we have for $f,g \in \mathcal{H}^{\otimes m}$

$$\mathbb{E}[\delta^m(f)\delta^n(g)] = \delta^{mn} \sum_{\sigma \in \mathfrak{S}_m} \langle f, \sigma_* g\rangle_{\mathcal{H}^{\otimes m}}\,. \tag{20}$$

This implies that multiple Wiener integration defines an isometry

$$\delta^\bullet : \bigoplus_{m=0}^{\infty} \mathcal{H}^{\odot m} \xrightarrow{\cong} L^2\Omega \tag{21}$$

where the source is given the degree-wise rescaled inner product $(f,g) \mapsto m!\langle f,g\rangle_{\mathcal{H}^{\otimes m}}$ for f,g of the same degree and zero otherwise, and Ω is endowed with the sigma algebra generated by the process $X_{t \in [0,T]}$. The image

of the m-th Wiener integral operator, the space of the random variables $\delta^m(f)$ with f ranging in $\mathcal{H}^{\odot m}$, is called the m-th *Wiener chaos* of X. We denote it $\mathcal{W}^m$ and the m-th Wiener chaos projection $w^m \colon L^2\Omega \twoheadrightarrow \mathcal{W}^m$. Note that $w^0 = \mathbb{E}$ with values in $\mathcal{W}^0 = \mathbb{R}$, while $\mathcal{W}^1$ is given by linear functionals of X. We thus have the Wiener chaos decomposition $L^2\Omega = \bigoplus_{m=0}^\infty \mathcal{W}^m$ which means it is possible to represent any random variable in $L^2\Omega$ (measurable w.r.t. to the sigma-algebra generated by X) as an L^2-absolutely convergent series

$$L^2\Omega \ni Z = \sum_{m=0}^\infty w^m Z, \qquad \|Z\|_{L^2}^2 = \sum_{m=0}^\infty \|w^m Z\|_{L^2}^2 = \sum_{m=0}^\infty m! \|f^m\|_{\mathcal{H}^{\otimes m}}^2 \qquad (22)$$

where $f^m = (\delta^m)^{-1} \circ w^m(Z)$. The map $(\delta^m)^{-1} \circ w^m$ admits an expression in terms of the Malliavin derivative: this is *Stroock's formula*, which states that for $Z \in \mathbb{D}^{m,2}$

$$(\delta^m)^{-1} \circ w^m(Z) = \frac{1}{m!} \mathbb{E}[\mathcal{D}^m Z]. \qquad (23)$$

As a consequence, if $Z \in \mathbb{D}^{\infty,2} := \bigcap_{m=0}^\infty \mathbb{D}^{m,2}$ we can write its Wiener chaos decomposition as the series

$$Z = \sum_{m=0}^\infty \frac{1}{m!} \delta^m \mathbb{E}[\mathcal{D}^m Z]. \qquad (24)$$

We continue calling elements of $\mathcal{E}^{\otimes m}$ elementary functions, in light of the fact that they can be identified with functions $([0,T] \times [d])^m \to \mathbb{R}$ by the mapping

$$\mathbb{1}_{[0,t_1)}^{\gamma_1} \otimes \cdots \otimes \mathbb{1}_{[0,t_m)}^{\gamma_m} \mapsto \mathbb{1}_{[0,t_1) \times \cdots \times [0,t_m)}^{\gamma_1,\ldots,\gamma_m}. \qquad (25)$$

This is the map given by the product of the Kronecker deltas $\delta^{\gamma_1} \cdots \delta^{\gamma_m}$ and the indicator function on the m-cube $[0,t_1) \times \cdots \times [0,t_m)$, each δ paired with the respective time variable. Since $\mathcal{E}^{\otimes m}$ is dense in $\mathcal{H}^{\otimes m}$, elements of the latter may be identified as equivalence classes of Cauchy sequences in $\mathcal{E}^{\otimes m}$. While $\mathcal{H}^{\otimes m}$ is not, in general, a space of functions, it is possible to uniquely associate elements of $\mathcal{H}^{\otimes m}$ to certain measurable functions $([0,T] \times [d])^m \to \mathbb{R}$ as follows:

Definition 1.1 (Functions as elements of $\mathcal{H}^{\otimes m}$)**.** For a function $f \colon ([0,T] \times [d])^m \to \mathbb{R}$ we will write $f \in \mathcal{H}^{\otimes m}$ if there exist a Cauchy sequence $(f_n)_n \subset \mathcal{E}^{\otimes m}$, uniformly bounded as a sequence of functions (according to the identification (25)), with $f_n \to f$ a.e. In this case we will say that f *represents* $\lim f_n \in \mathcal{H}^{\otimes m}$. If f represents $\phi, \psi \in \mathcal{H}^{\otimes m}$ then $\phi = \psi$: this is an immediate consequence of the following

Lemma 1.2. *Let $(f_n)_n$ be as in the above definition with $f = 0$. Then $f_n \to 0$ in $\mathcal{H}^{\otimes m}$.*

Proof. Let

$$f_n = \sum_{\gamma_1,\ldots,\gamma_m=1}^d f_{n;\gamma_1,\ldots,\gamma_m} \mathbb{1}^{\gamma_1,\ldots,\gamma_m}$$

with $f_{n;\gamma_1,\ldots,\gamma_m} \colon [0,T]^m \to \mathbb{R}$. Then $f_n \to 0$ a.e. if and only if $f_{n;\gamma_1,\ldots,\gamma_m} \to 0$ a.e. for each $(\gamma_1,\ldots,\gamma_m) \in [d]^m$. Keeping in mind that $\mathcal{H} \cong (\mathcal{H}^1)^{\oplus d}$ we may therefore assume $d = 1$ and suppress indices. Following [23, p.588], we test the sequence with elementary functions: letting $f_n = \sum_{s_1^n,\ldots,s_m^n} f_n^{s_1^n,\ldots,s_m^n} \mathbb{1}_{[0,s_1^n) \times \ldots \times [0,s_m^n)}$ and $\mathcal{E}^{\otimes m} \ni g = \sum_{t_1,\ldots,t_m} g^{t_1,\ldots,t_m} \mathbb{1}_{[0,t_1) \times \ldots \times [0,t_m)}$ with $f_n^{s_1^n,\ldots,s_m^n}, g^{t_1,\ldots,t_m} \in \mathbb{R}$ uniformly bounded (and the sums finite) we have that

$$\langle f_n, g \rangle_{\mathcal{H}^{\otimes m}}$$
$$= \sum_{\substack{s_1^n,\ldots,s_m^n \\ t_1,\ldots,t_m}} f_n^{s_1^n,\ldots,s_m^n} g^{t_1,\ldots,t_m} R(s_1^n, t_1) \cdots R(s_m^n, t_m)$$
$$= \sum_{t_1,\ldots,t_m} g^{t_1,\ldots,t_m} \int_{((0,T]\setminus\{t_1\}) \times \cdots \times ((0,T]\setminus\{t_m\})} f_n(s_1,\ldots,s_m) \partial_1 R(s_1, t_1) \cdots \partial_1 R(s_m, t_m) \mathrm{d}s_1 \cdots \mathrm{d}s_m.$$

(10) and (11) imply that the integrands are absolutely and uniformly bounded by $[|t_1 - s_1|^{2H-1} \vee s_1^{2H-1}] \cdots [|t_m - s_m|^{2H-1} \vee s_m^{2H-1}]$ (up to a constant), which is integrable on $((0, T] \setminus \{t_1\}) \times \cdots \times ((0, T] \setminus \{t_m\})$. By dominated convergence $\langle \phi, g \rangle_{\mathcal{H}^{\otimes m}} = \lim \langle f_n, g \rangle_{\mathcal{H}^{\otimes m}} = 0$, where $\phi := \lim f_n$ in $\mathcal{H}^{\otimes m}$, and $\phi = 0$ follows from the fact that g ranges in a dense set. ∎

In light of the aforementioned non-degeneracy condition on X, we also expect the converse to hold: if $\phi \in \mathcal{H}^{\otimes m}$ is represented by the functions f, g in the above sense, then $f = g$ a.e. An example of a degenerate stochastic process, for which this property would not hold, is given by taking any process X and concatenating it with itself path by path; the resulting covariance function R would be invariant under transposing the intervals $[0, T]$ and $[T, 2T]$. We also note that, in specific cases, it is possible to describe $\mathcal{H}$ explicitly: if X is a fractional Brownian motion with Hurst parameter $H \in (0, 1)$, the identity on $\mathcal{E}$ induces an isomorphism between $\mathcal{H}$ and the Sobolev space $W^{1/2-H,2}$ [24], which is a space of functions for $H \in (0, 1/2]$ but not for $H \in (1/2, 1)$.

We will mostly be considering Wiener integrals on simplices, which has the effect of quotienting out symmetry of the operator δ^m. We will often resort to integral notation, e.g. if $\mathbb{1}_{\Delta[s,t]}^{\alpha\beta} \in \mathcal{H}^{\otimes 2}$ (the function that maps $((u, \gamma), (v, \delta)) \mapsto \delta^{\alpha\gamma}\delta^{\beta\delta}\mathbb{1}_{s<u<v<t}$) in the sense of Definition 1.1, we will write $\delta^2(\mathbb{1}_{\Delta[s,t]}^{\alpha\beta}) =: \int_{s<u<v<t} \delta X_u^\alpha \delta X_v^\beta$ to be the limit in L^2 of $\delta^2(f_n)$. Wiener integrals of elements of $\mathcal{E}^{\otimes m}$, on the other hand, can be computed explicitly by using the adjoint property (17): for example, it can be checked that

$$\delta^2(\mathbb{1}_{[s,t)}^\alpha \otimes \mathbb{1}_{[u,v)}^\beta) = X_{st}^\alpha X_{uv}^\beta - R^{\alpha\beta}(\Delta(s, t), \Delta(u, v)).$$

The more general formula involves multivariate analogues of the Hermite polynomials (see [34, §2.7.2] and [17, p.244]). When X is a Gaussian martingale (but not necessarily if it is only a semimartingale), multiple Wiener integration on the simplex coincides with iterated Wiener-Itô integration.

2 The main result, some consequences

We begin this section with some more notation. We denote $[n] := \{1, \ldots, n\}$ the set with n elements. We will be concerned with iterated integrals on the n-simplex $\Delta^n[s,t] := \{(u_1, \ldots, u_n) \mid s < u_1 < \ldots < u_n < t\}$. Because such integrals will involve the covariance function, integration variables will sometimes come in pairs. For $m, n \in \mathbb{N}$ we denote $\mathcal{P}_m^n$ the collection of partitions of subsets of $[n]$ of cardinality $n - m$ into sets of cardinality 2. Note that this means $\mathcal{P}_m^n = \varnothing$ whenever $n \neq m \pmod 2$ or $m > n$, but $\mathcal{P}_n^n$ has precisely one element, $\varnothing$: the empty set admits the empty collection of subsets as a partition, which vacuously belongs to $\mathcal{P}_n^n$. For example, $Q := \{\{1, 4\}, \{3, 8\}, \{5, 6\}\} \in \mathcal{P}_2^8$ viewed as a partition of the set $\{1, 3, 4, 5, 6, 8\} \subseteq [8]$. For $P \in \mathcal{P}_m^n$ we will denote $\overline{P} := [n] \setminus \cup P$ (in the partition of the above example, $\overline{Q} = \{2, 7\}$).

It will convenient to use graphical notation to denote such objects, and for reasons that will become apparent shortly, for a pair $\{i, j\}$ with $i \leq j$ we will distinguish between the consecutive case $j = i + 1$ and the non-consecutive one $j > i + 1$. The partition $Q \in \mathcal{P}_2^8$ above is represented by

$$Q = \overbrace{} \quad . \tag{26}$$

We will refer to such graphics as *diagrams*. We have drawn one node for each $i \in [n]$ that is not paired with a consecutive integer, and one node for each consecutive pair (in this case only $\{5, 6\}$); when counting nodes, a node corresponding to such a pair should be thought as having double weight. In our example, the 5th node actually counts for positions 5 and 6. With this convention, for each non-consecutive pair $\{i, j\}$ we have drawn an *arc* connecting the two nodes of positions i and j, and for each node corresponding to a consecutive pair we have drawn a *line* going upwards. Nodes that do not have a line or arc entering them correspond to elements of $\overline{P}$, and we will call them *single*. Note that, by construction, there is never an arc between two consecutive nodes: this will be critical for convergence of the associated integrals described below. In the next sec-

tion, we will be particularly concerned with maximal *sequences of consecutive pairings*, i.e. collections of pairings $\{k, k+1\}, \ldots, \{k+l, k+l+1\} \in P$ with $l \geq 0$ and s.t. $\{k-2, k-1\}, \{k+l+2, k+l+3\} \notin P$.

Now, given $P \in \mathcal{P}_m^n$, $0 \leq s \leq t \leq T$ and $\gamma_1, \ldots, \gamma_n \in [d]$ we associate to it a continuous function $P_{st}^{\gamma_1, \ldots, \gamma_n} \colon \Delta^m[s,t] \times [d]^m \to \mathbb{R}$ by integrating over as many variables as there are non-single nodes in the diagram that represents P: call this number, which equals twice the number of non-consecutive pairs in P plus the number of consecutive ones, $\#P$. This explains our choice for the above notation: each node either corresponds to an integration variable or to a free variable, i.e. a variable of which $P_{st}^{\gamma_1, \ldots, \gamma_n}$ is a function. We use the shorthands

$$R(\mathrm{d}u_i, \mathrm{d}u_j) := \partial_{12} R(u_i, u_j) \mathrm{d}u_i \mathrm{d}u_j$$
$$\tfrac{1}{2} R(\mathrm{d}u_{h+1}) - R(u_{h-1}, \mathrm{d}u_{h+1}) := \left[\tfrac{1}{2} R'(u_{h+1}) - \partial_2 R(u_{h-1}, u_{h+1})\right] \mathrm{d}u_{h+1},$$

(27)

and the former will only be used when $j > i + 1$. Crucially, we are defining the second case as $\tfrac{1}{2} R(\mathrm{d}u_{h+1}) - R(u_{h-1}, \mathrm{d}u_{h+1})$, not as $R(u_{h+1}, \mathrm{d}u_{h+1}) - R(u_{h-1}, \mathrm{d}u_{h+1})$, since this would be ill-defined in many cases (including $1/2 > H$-fBm) because $R(\cdot, \cdot)$ may not admit partial derivatives on the diagonal. On the other hand, we are assuming that the variance function $R(\cdot)$ is differentiable.

Definition 2.1 ($P_{st}^{\gamma_1, \ldots, \gamma_n}$). For $\gamma_1, \ldots, \gamma_n \in [d]$, $0 \leq s \leq t \leq T$ and $P \in \mathcal{P}_m^n$ define

$$P_{st}^{\gamma_1, \ldots, \gamma_n}(u_k \mid k \in \overline{P}) := \prod_{k \in \overline{P}} \mathbb{1}^{\gamma_k} \cdot \int_{\Delta^{\#P}[s,t]} \prod_{\substack{\{i,j\} \in P \\ |j-i|>1}} R^{\gamma_i \gamma_j}(\mathrm{d}u_i, \mathrm{d}u_j)$$

$$\cdot \prod_{\{h,h+1\} \in P} \left[\tfrac{1}{2} R^{\gamma_h \gamma_{h+1}}(\mathrm{d}u_{h+1}) - R^{\gamma_h \gamma_{h+1}}(u_{h-1}, \mathrm{d}u_{h+1})\right]$$

(28)

as a function $([0,T] \times [d])^m \to \mathbb{R}$ extended with the value 0 outside $\Delta^m[s,t]$.

The variables u_k with $k \in \overline{P}$ are supplied as arguments, so in fact this is an integral over a disjoint union of up to $m+1$ simplices (fewer if some of the elements of $\overline{P}$ are consecutive). The kth index in $[d]^n$ is given as argument to $\mathbb{1}^{\gamma_k}$ as a Kronecker delta: this means that $P_{st}^{\gamma_1, \ldots, \gamma_n}$ vanishes on all but one element of $[d]^m$. The reason why we still consider $P_{st}^{\gamma_1, \ldots, \gamma_n}$ as a function on $[d]^m$ is that this is necessary to view it as an element of $\mathcal{H}^{\otimes m}$; nevertheless, when the indices are fixed it will sometimes be convenient to just think of it as a function of m times. If $m = 0$, $P_{st}^{\gamma_1, \ldots, \gamma_n}$ is just a real number.

Remark 2.2. The presence of the second type of integrand in Definition 2.1 is the reason for the smoothness assumptions on the variance and covariance functions, which are not to be found in most of the literature on these topics: this is because it would be difficult to define integrals such as $\int_{s<u<v<t} \left[\tfrac{1}{2} R(\mathrm{d}u) - R(s, \mathrm{d}u)\right] \left[\tfrac{1}{2} R(\mathrm{d}v) - R(u, \mathrm{d}v)\right]$ as iterated Young integrals, without taking derivatives, since the variable u in its undifferentiated form appears after the integrator $\tfrac{1}{2} R(\mathrm{d}u) - R(s, \mathrm{d}u)$; this of course is no longer an issue under our smoothness hypotheses, thanks to which the above integral is defined as the Lebesgue integral on the simplex $\int_{s<u<v<t} \left[\tfrac{1}{2} R'(u) - \partial_2 R(s, u)\right] \left[\tfrac{1}{2} R'(v) - \partial_2 R(u, v)\right] \mathrm{d}u \mathrm{d}v$.

When P is represented by a diagram, we will decorate the nodes with labels. For example, the integral associated to (26) with labelling $\alpha, \ldots, \vartheta$ is given by

$$\left(\underset{\alpha \ \beta \ \gamma \ \delta \ \varepsilon\zeta \ \eta \ \vartheta}{\overbrace{\bullet \ \bullet \ \bullet \ \bullet \ \bullet \ \bullet \ \bullet \ \bullet}} \right)_{st}$$

$$= \mathbb{1}^{\beta\eta} \int_{\Delta^7[s,t]} R^{\alpha\delta}(\mathrm{d}u_1, \mathrm{d}u_4) R^{\gamma\vartheta}(\mathrm{d}u_3, \mathrm{d}u_8) \left[\tfrac{1}{2} R^{\varepsilon\zeta}(\mathrm{d}u_6) - R^{\varepsilon\zeta}(u_4, \mathrm{d}u_6)\right]$$

$$= \mathbb{1}^{\beta\eta}_{\Delta^2[s,t]}(u_2, u_6) \int_{\substack{s<u_1<u_2 \\ u_2<u_3<u_4<u_6<u_7 \\ u_7<u_8<t}} \partial_{12} R^{\alpha\delta}(u_1, u_4) \partial_{12} R^{\gamma\vartheta}(u_3, u_8) \left[\tfrac{1}{2} R^{\varepsilon\zeta\prime}(u_6) - \partial_2 R^{\varepsilon\zeta}(u_4, u_6)\right]$$
$$\mathrm{d}u_1 \mathrm{d}u_3 \mathrm{d}u_4 \mathrm{d}u_6 \mathrm{d}u_8.$$

This is viewed as a function of the variables u_2, u_6 ranging on the simplex $\Delta^2[s,t]$, each paired with an index variable, which must respectively be equal to β, η for the expression not to vanish. The variable u_5 has been skipped, since it is the first term in the consecutive pair $\{5,6\}$. We will show that integrals defined in this fashion are a.e. limits of Cauchy sequences in $\mathcal{E}^{\otimes m}$, which therefore uniquely represent elements of $\mathcal{H}^{\otimes m}$ according to Definition 1.1. When taking multiple Wiener integrals of them, the indices corresponding to the nodes that represent free variables will become the coordinate processes that are being integrated against, e.g.

$$\delta^2 (\overset{\frown}{\underset{\alpha\ \ \beta\ \ \gamma\ \ \delta\ \ \varepsilon\zeta\ \ \eta\ \ \vartheta}{\cdot\ \ \cdot\ \ \cdot\ \ \cdot\ \ \cdot\cdot\ \ \downarrow\ \ \cdot}})_{st}$$

$$= \int_{s<u_2<u_7<t} \left[\int_{\substack{s<u_1<u_2 \\ u_2<u_3<u_4<u_6<u_7 \\ u_7<u_8<t}} \partial_{12}R^{\alpha\delta}(u_1,u_4)\partial_{12}R^{\gamma\vartheta}(u_3,u_8)\left[\tfrac{1}{2}R'^{\varepsilon\zeta}(u_6)-\partial_2 R^{\varepsilon\zeta}(u_4,u_6)\right] \right] \delta X^{\beta}_{u_2}\delta X^{\eta}_{u_7}.$$

We are now ready to state the main theorem.

Theorem 2.3 (Wiener chaos expansion of the signature of a Gaussian process).
Given $m, n \in \mathbb{N}$, $P \in \mathcal{P}^n_m$, $\gamma_1, \dots, \gamma_n \in [d]$, $0 \le s \le t \le T$, it holds that $P^{\gamma_1,\dots,\gamma_n}_{st} \in \mathcal{H}^{\otimes m}$ in the sense of Definition 1.1, and the mth Wiener chaos projection of the signature of X is given by

$$w^m \mathcal{S}(X)^{\gamma_1,\dots,\gamma_n}_{st} = \sum_{P \in \mathcal{P}^n_m} \delta^m P^{\gamma_1,\dots,\gamma_n}_{st}. \tag{29}$$

In particular, notice that $w^m \mathcal{S}(X)^{\gamma_1,\dots,\gamma_n}_{st}$ can only be non-zero when $m \le n$ and $m \equiv n \pmod 2$. The most important case of this result is when $m = 0$:

Corollary 2.4 (Expected signature of a Gaussian process). *With notation as above, we have*

$$\mathbb{E}\mathcal{S}(X)^{\gamma_1,\dots,\gamma_n}_{st} = \sum_{P \in \mathcal{P}^n_0} P^{\gamma_1,\dots,\gamma_n}_{st}. \tag{30}$$

Remark 2.5 (Eliminating variables). While convergence rules out always considering integrands of the first type in (27) (which would mean allowing diagrams with arcs between consecutive nodes), one may wonder whether it is possible to only consider integrands of the second type, i.e. by integrating out one variable per pair and thus simplifying the presentation of the formula. This, however, is not possible in general, because of the additional constraint that requires two consecutive variables not to be both integrated out (for the expression to make sense as an integral). It is not difficult to see, for example, that in the following diagram

$$\overset{\frown}{\underset{}{\cdot\overset{\frown}{\cdot\ \cdot\ \cdot}\cdot}}$$

at most two variables can be integrated out (unless the remaining integral can be solved or simplified analytically). Luckily, the only case in which it is necessary for convergence to integrate out certain variables (as specified in the second case of (27)), is when there are consecutive pairs: this is always possible, even when more than one pair in a row is consecutive, since we may always pick the first variable to integrate out (as done here — one could equivalently have chosen the second). Of course, there is always some number of additional variables that can be eliminated, but we do not immediately see a way of doing this in a maximal way that is canonical.

Example 2.6 (The Wiener chaos decomposition of $\mathcal{S}^3(X)_{st}$). We give the explicit expression for the Wiener chaos expansion of the signature truncated at level 3. These terms are especially significant, considering that they are the ones that define the rough path when $1/4 < H \le 1/3$: higher signature terms can be derived in a pathwise fashion by Lyons's extension theorem without involving probability. We represent each signature

term as a sum of their Wiener chaos projections in ascending order; in particular the sum of all non-random terms constitutes the expectation of the left hand side.

$$\mathbb{S}(X)_{st}^{\varnothing} = \varnothing_{st} = 1, \quad \mathbb{S}(X)_{st}^{\gamma} = \delta(\underset{\gamma}{\bullet})_{st} = \delta(\mathbb{1}_{[s,t)}^{\gamma}) = X_{st}^{\gamma}$$

$$\mathbb{S}(X)_{st}^{\alpha\beta} = \underset{\alpha\beta}{\downarrow} + \delta^2(\underset{\alpha}{\bullet}\ \underset{\beta}{\bullet})$$

$$= \frac{R^{\alpha\beta}(s) + R^{\alpha\beta}(t)}{2} - R^{\alpha\beta}(s,t) + \int_{s<u<v<t} \delta X_u^{\alpha} \delta X_v^{\beta}$$

$$\mathbb{S}(X)_{st}^{\alpha\beta\gamma} = \delta(\underset{\alpha\beta}{\downarrow}\ \bullet)_{st} + \delta(\bullet\ \underset{\alpha\ \beta\gamma}{\downarrow})_{st} + \delta(\underset{\alpha\ \beta\ \gamma}{\frown})_{st} + \delta^3(\underset{\alpha\ \beta\ \gamma}{\bullet\ \bullet\ \bullet})_{st}$$

$$= \int_s^t \left(\frac{R^{\alpha\beta}(s) + R^{\alpha\beta}(u)}{2} - R^{\alpha\beta}(s,u) \right) \delta X_u^{\gamma} + \int_s^t \left(\frac{R^{\beta\gamma}(u) + R^{\beta\gamma}(t)}{2} - R^{\beta\gamma}(u,t) \right) \delta X_u^{\alpha}$$

$$+ \int_s^t R^{\alpha\gamma}(\Delta(s,u), \Delta(u,t)) \delta X_u^{\beta} + \int_{s<u<v<w<t} \delta X_u^{\alpha} \delta X_v^{\beta} \delta X_w^{\gamma}$$

In particular, notice how the expected signature of level 2 is given by the difference between the average of the variances and the covariance:

$$\mathbb{E}\mathbb{S}(X)_{st}^{\alpha\beta} = \frac{R^{\alpha\beta}(s) + R^{\alpha\beta}(t)}{2} - R^{\alpha\beta}(s,t) \tag{31}$$

and that the statement that "the Itô and Stratonovich Lévy areas are equal" carries over to the Gaussian Wiener-rough setting, in the sense that

$$\frac{1}{2}\big(\mathbb{S}(X)_{st}^{\alpha\beta} - \mathbb{S}(X)_{st}^{\beta\alpha}\big) = \frac{1}{2}\int_{s<u<v<t} \delta X_u^{\alpha}\delta X_v^{\beta} - \delta X_u^{\beta}\delta X_v^{\alpha} \tag{32}$$

by symmetry of the covariance function.

Example 2.7 ($\mathbb{E}\mathbb{S}(X)^{(4)}$). Corollary 2.4 at level 4 is given by

$$\mathbb{E}\mathbb{S}(X)_{st}^{\alpha\beta\gamma\delta} = (\underset{\alpha\beta\ \gamma\delta}{\downarrow\ \downarrow})_{st} + (\underset{\alpha\ \beta\gamma\ \delta}{\frown\downarrow})_{st} + (\underset{\alpha\ \beta\ \gamma\ \delta}{\frown})_{st}$$

$$= \int_{s<u<v<t} \left[\tfrac{1}{2}R^{\alpha\beta}(\mathrm{d}u) - R^{\alpha\beta}(s,\mathrm{d}u)\right]\left[\tfrac{1}{2}R^{\gamma\delta}(\mathrm{d}v) - R^{\gamma\delta}(u,\mathrm{d}v)\right]$$

$$+ \int_{s<u<v<w<t} R^{\alpha\delta}(\mathrm{d}u,\mathrm{d}w)\left[\tfrac{1}{2}R^{\beta\gamma}(\mathrm{d}v) - R^{\beta\gamma}(u,\mathrm{d}v)\right] \tag{33}$$

$$+ \int_{s<u<v<w<z<t} R^{\alpha\gamma}(\mathrm{d}u,\mathrm{d}w)R^{\beta\delta}(\mathrm{d}v,\mathrm{d}z) .$$

Using a clever transformation, [3, Theorem 34] are able to compute $\mathbb{E}\mathbb{S}(X)_{01}^{(2)}$ and $\mathbb{E}\mathbb{S}(X)_{01}^{(4)}$ for $1/4 < H$-fBm. Their formulae are specific to the cases $n = 2, 4$ and X a fBm, and are quite different to those given by Theorem 2.3. That the two coincide is immediate at level 2 by (31), and in Appendix A we perform this check at level 4.

The following example shows how Theorem 2.3 has the potential to generate insight into numerics of numerical schemes for rough differential equations driven by Gaussian signals.

Example 2.8 (Itô-Taylor expansions for solutions to RDEs driven by Gaussian signals). Assume

$$\mathrm{d}Y = V(Y)\mathrm{d}\boldsymbol{X}, \qquad Y_0 = y_0$$

is an RDE (rough differential equation) driven by the Gaussian rough path $\boldsymbol{X}$ (defined by the first 1, 2 or 3 levels of $\mathcal{S}(X)$, depending on how rough X is). Proceeding formally, and denoting by $V_{\gamma_1} \cdots V_{\gamma_n}$ composition of vector fields (and using Einstein notation), we can then expand the solution Y as

$$
\begin{aligned}
Y_t &= \sum_{n=0}^{\infty} V_{\gamma_1} \cdots V_{\gamma_n}(y_0) \mathcal{S}(X)_{0t}^{\gamma_1,\ldots,\gamma_n} \\
&= \sum_{n=0}^{\infty} V_{\gamma_1} \cdots V_{\gamma_n}(y_0) \sum_{\substack{0 \le m \le n \\ m \equiv n \bmod 2}} w^m \mathcal{S}(X)_{0t}^{\gamma_1,\ldots,\gamma_n} \\
&= \sum_{n=0}^{\infty} V_{\gamma_1} \cdots V_{\gamma_n}(y_0) \sum_{\substack{0 \le m \le n \\ m \equiv n \bmod 2}} \sum_{P \in \mathcal{P}_m^n} \delta^m P_{0t}^{\gamma_1,\ldots,\gamma_n} \\
&= \sum_{m=0}^{\infty} \sum_{\substack{n \ge m \\ n \equiv m \bmod 2}}^{\infty} V_{\gamma_1} \cdots V_{\gamma_n}(y_0) \sum_{P \in \mathcal{P}_m^n} \delta^m P_{0t}^{\gamma_1,\ldots,\gamma_n} .
\end{aligned}
$$

The expansion on the first line can be viewed as the extension to the Gaussian case of Stratonovich-Taylor series, the one on the last line can be viewed as that of Itô-Taylor series [25]. The latter has the advantage that its terms fit in well with the Wiener chaos decomposition of Y_t, although it should be observed that $w^m Y_t$ is represented as an infinite series, namely the second sum in the last line above. Also, this expansion cannot be expected to coincide with the Wiener chaos decomposition of Y_t if it is performed at times other than 0, with $Y_0 = y_0$ deterministic. This is because, unless X is a martingale, the Wiener chaos isometries will not hold conditionally on $\mathcal{F}_s$.

Remark 2.9 (Stationarity and joint stationarity of increments). X is stationary if and only if we may write

$$R(s,t) = \overline{R}(t-s) \tag{34}$$

for some function $\overline{R} \colon [0,T] \to \mathbb{R}^{d \times d}$. In this case we have

$$
\begin{aligned}
&\partial_{12} R(s,t) = -\overline{R}''(t-s), \quad R'(t) = 0, \quad \partial_2 R(s,t) = \overline{R}'(t-s) \\
\implies \quad &\tfrac{1}{2} R'(t) - \partial_2 R(s,t) = -\overline{R}'(t-s) .
\end{aligned}
\tag{35}
$$

An example of a centred stationary Gaussian process is the stationary Ornstein-Uhlenbeck process $e^{-t/2} W_{e^t}$ where W is a Brownian motion and $t \in [0,T]$: its covariance function is $R(s,t) = e^{-(t-s)/2}$ for $s \le t$. This process however, strictly speaking, is not among those considered here, as it has random initial condition.

There is a much weaker property that results in a similar simplification. We will say that a stochastic process X has *jointly stationary increments* if for all $s_1 \le t_1, \ldots, s_n \le t_n$ the distribution of the random vector of increments $(X_{s_1 t_1}, \ldots, X_{s_n t_n})$ only depends on the differences $t_1 - s_1, \ldots, t_n - s_n$ and $s_2 - s_1, \ldots, s_n - s_{n-1}$ (if $n = 1$ the latter condition vanishes, and ordinary stationarity of increments is recovered). If X is Gaussian this need only be required for $n = 2$, and if it holds we may write

$$R(\Delta(s,u), \Delta(t,v)) = \mathbb{E}[X_{su} \otimes X_{tv}] = \widehat{R}(u-s, v-t, t-s) \tag{36}$$

for some function $\widehat{R} \colon [0,T]^3 \to \mathbb{R}^{d \times d}$. This property is satisfied by fBm, since if H is the Hurst parameter we have

$$R(\Delta(s,u), \Delta(t,v))$$

$$= \frac{1}{2}\left[(t-u)^{2H} + (v-s)^{2H} - (t-s)^{2H} - (v-u)^{2H}\right]$$
$$= \frac{1}{2}\left[\left((t-s)-(u-s)\right)^{2H} + \left((v-t)+(t-s)\right)^{2H} - \left(t-s\right)^{2H} - \left((v-t)+(t-s)-(u-s)\right)^{2H}\right].$$

If X has jointly stationary increments

$$\partial_{12}R(s,t) = \lim_{\substack{u\to s\\v\to t}} \frac{R(\Delta(s,u),\Delta(t,v))}{(v-t)(u-s)} = \partial_{12}\widehat{R}(0,0,t-s). \tag{37}$$

Although similar simplifications are not available for $\partial_2 R(s,t)$ and $R'(t)$ individually (as they are in the stationary case), they are for their difference: indeed, using that $R(\,\cdot\,,0) \equiv 0$, we have

$$\tfrac{1}{2}R(t+h) - \tfrac{1}{2}R(t) - \big(R(s,t+h) - R(s,t)\big)$$
$$= \tfrac{1}{2}\big[R(\Delta(s,t),\Delta(t,t+h)) + R(\Delta(s,t+h),\Delta(t,t+h))\big]$$

which implies

$$\tfrac{1}{2}R'(t) - \partial_2 R(s,t) = \tfrac{1}{2}\partial_h|_{h=0}\big[\widehat{R}(t-s,h,t-s) + \widehat{R}(t+h-s,h,t-s)\big]$$
$$= \tfrac{1}{2}\partial_1\widehat{R}(t-s,0,t-s) + \partial_2\widehat{R}(t-s,0,t-s).$$

We therefore conclude that joint stationarity of increments, though a much more general property than stationarity, results in the same simplifications that are of relevance to Theorem 2.3, namely that $\partial_{12}R(s,t)$ and $\tfrac{1}{2}R'(t) - \partial_2 R(s,t)$ only depend on $t-s$. This can be of aid in simplifying the expression of the integrals in the formula for $w^m S(X)$, since it is possible to perform substitutions of the form $v_{ij} = u_j - u_i$. It does not, however, guarantee that these integrals become analytically solvable, as simple examples show (e.g. the integral $\int_0^1 v^{2H-1}(1-v)^{2H-1}\mathrm{d}v$ appearing in Appendix A).

We now consider a few examples of Gaussian processes to which our results apply; in all cases, X will have i.i.d. components, and we will use R to denote the scalar covariance function of each component. Arguably the most important example of a stochastic process for which the signature has not yet been computed is fractional Brownian motion in the regime of negatively-correlated increments:

Example 2.10 ($(1/4,1/2) \ni H$-fBm). *Fractional Brownian motion* with *Hurst parameter* $H \in (0,1)$ (H-fBm), introduced in [32], is a scalar centred Gaussian process with covariance function

$$R(s,t) = \frac{1}{2}(t^{2H} + s^{2H} - (t-s)^{2H}), \quad s \leq t. \tag{38}$$

It is not a semimartingale unless $H = 1/2$, in which case it is Brownian motion. Here we consider the case $H \in (1/4,1/2)$: this is well known to satisfy the preliminary hypotheses required in Section 1, and the smoothness conditions and bounds are simple to verify. Indeed, the integrands of interest for the formula of Theorem 2.3 are given by ($s \leq t$)

$$\partial_{12}R(s,t) = H(2H-1)(t-s)^{2H-2}$$
$$\tfrac{1}{2}R'(t) - \partial_2 R(s,t) = H(t-s)^{2H-1}. \tag{39}$$

As predicted by Remark 2.9, these both are functions of $t-s$.

Remark 2.11 ($(1/2,1) \ni H$-fBm, [3]). If $R(\,\cdot\,,\,\cdot\,)$ is once differentiable on the diagonal, then

$$R'(t) = \frac{\mathrm{d}}{\mathrm{d}t}R(t,t) = 2\partial_2 R(t,t)$$

and we have

$$\int_s^t \left[\tfrac{1}{2}R'(v) - \partial_2 R(s,v)\right]\mathrm{d}v = \int_s^t \left[\partial_2 R(v,v) - \partial_2 R(s,v)\right]\mathrm{d}v = \int_{s<u<v<t} \partial_{12} R(u,v)\mathrm{d}u\mathrm{d}v\,.$$

By performing this substitution in Corollary 2.4 for the case of $1/2 < H$-fBm (this means always applying the first case in (27), i.e. allowing arcs between consecutive nodes, which replace lines), we recover the formula of [3, Theorem 31] (note that the symmetry factor — meant to factor out permutations of pairings and transpositions within each pair — is not present in our case, since we are summing over pairings and not permutations). Other examples of processes in a similar regularity regime are those Gaussian Volterra processes with strictly regular kernels considered in [7].

The following is another example of a fractional, non-semimartingale process.

Example 2.12 (The Riemann-Liouville process). Another centred continuous Gaussian process, originally introduced in [26] and subsequently in [32], is the *Riemann-Liouville process* with Hurst parameter $H \in (0,1)$ (sometimes called "type-II fBm"), is a centred Gaussian process with covariance function [33, p.116-117]

$$R(s,t) \underset{s<t}{=} \frac{1}{2}\left[t^{2H} + s^{2H} - \underbrace{2H(t-s)^{2H}\left(\frac{1}{2H} + \int_0^{s/(t-s)} \left((1+u)^{H-1/2} - u^{H-1/2}\right)^2\mathrm{d}u\right)}_{=R(\Delta(s,t),\Delta(s,t))}\right]\,. \tag{40}$$

Like fBm, this process specifies to Brownian motion when $H = 1/2$ and is otherwise not a semimartingale. Their main difference between the two is that fBm has jointly stationary increments while for the Riemann-Liouville process not even single increments are stationary. We were not able to find a satisfactory expression for the derivatives of the covariance function of this process, and thus were not able to determine whether (for $H > 1/4$) it satisfies the conditions necessary for applying Theorem 2.3. However, we believe that examples such as this provide strong motivation for not confining our study to fBm and to allow for more general processes.

Another important restriction of the main result is the following case:

Remark 2.13 (Gaussian martingales, [16]). When X is a continuous Gaussian martingale, its quadratic variation coincides with its variance function (as can be seen by the fact that $X_t^2 - R(t)$ is a martingale). The Dubins-Schwarz theorem then implies that X can be represented as the deterministically-reparametrised Brownian motion $W_{R(t)}$. Assuming equal distribution of components, we can use this and the formula for the expected signature of Brownian motion (2) to compute

$$\mathbb{E}\mathbb{S}(X)_{st}^{\gamma_1,\dots,\gamma_{2n}} = \frac{R(\Delta(s,t))^n}{2^n n!}\delta^{\gamma_1\gamma_2}\cdots\delta^{\gamma_{2n-1}\gamma_{2n}}\,. \tag{41}$$

Since by martingality $\partial_{12}R(s,t) = 0 = \partial_2 R(s,t)$ on $s < t$, Theorem 2.3 reduces to a sum of iterated integrals that only involve $\tfrac{1}{2}R'$, which coincides with the above formula.

We conclude with two examples of centred, continuous Gaussian semimartingales which are not martingales and do not have stationary increments.

Example 2.14 (Brownian bridge returning to the origin). The Brownian Bridge returning to the origin at time T is a process whose law is given by disintegrating the Wiener measure on the event $W_T = 0$, where W is a d-dimensional Brownian motion starting at the origin. It can be written either as

$$X_t = W_t - \frac{t}{T}W_T, \qquad t \in [0,T]$$

or adaptedly as

$$X_t = (T-t)\int_0^t \frac{\mathrm{d}W_s}{T-s}, \qquad t \in [0,T)$$

(and $X_T = 0$). Its covariance function is given by

$$R(s,t) = s\left(1 - \frac{t}{T}\right), \qquad s \le t \tag{42}$$

and the integrands of interest are thus

$$\partial_{12} R(s,t) = -\frac{1}{T}$$
$$\tfrac{1}{2} R'(t) - \partial_2 R(s,t) = \frac{1}{2} - \frac{t-s}{T}. \tag{43}$$

It should be mentioned that X, as a process defined on $[0, T]$, fails the non-degeneracy condition [9, p.2125]. This is, however, not a problem, as we can view it as defined on the interval $[0, T - \varepsilon]$ and obtain the signature terms $\mathcal{S}(X)_{sT}$ through a limiting argument. The bounds of (9), which in this example and the one below only involve linear terms, are easily checked (and indeed the first is not even sharp). Note that the iterated integrals of (43) can all be solved explicitly as polynomials.

Example 2.15 (Centred Ornstein-Uhlenbeck processes started at 0). We consider an Ornstein-Uhlenbeck process with zero mean and deterministic initial condition, given by the Wiener-Itô integral

$$X_t = \sigma \int_0^t e^{-\theta(t-u)} \mathrm{d}W_u$$

with $\sigma, \theta \in (0, +\infty)$. Its covariance function is given by

$$R(s,t) = \frac{\sigma^2}{2\theta}\left(e^{-\theta(t-s)} - e^{-\theta(s+t)}\right), \qquad s \le t \tag{44}$$

and $\partial_{12} R(\mathrm{d}s, \mathrm{d}t)$, $\tfrac{1}{2} R'(t) - \partial_2 R(s,t)$ can be computed directly. Once again, all conditions are satisfied (see [9, p2138]).

3 Proof of the main result

Recall that we are using $\lesssim$ to denote inequalities whose constant of proportionality may only depend on T, H and other properties of a fixed process X. Since most of the arguments presented in this section only concern bounds and convergence, we will suppress indices (i.e. treat the scalar case) most of the time, so as not to clutter the notation. Given $P \in \mathcal{P}_m^n$, denote $|P|_{st}$ the function $\Delta^m[s,t] \to \mathbb{R}$ defined analogously to Definition 2.1, but replacing each integrand $\partial_{12} R(u,v)$ with $(v-u)^{2H-2}$ and each integrand $\tfrac{1}{2} R'(v) - \partial_2 R(u,v)$ with $(v-u)^{2H-1}$. For example, if Q is the diagram of (26)

$$|Q|_{st} = \mathbb{1}_{\Delta^2[s,t]}(u_2, u_7) \int_{\substack{s < u_1 < u_2 \\ u_2 < u_3 < u_4 < u_6 < u_7 \\ u_7 < u_8 < t}} (u_4 - u_1)^{2H-2}(u_8 - u_3)^{2H-2}(u_6 - u_4)^{2H-1} \mathrm{d}u_1 \mathrm{d}u_3 \mathrm{d}u_4 \mathrm{d}u_6 \mathrm{d}u_8.$$

The following proposition guarantees that all the integrals considered in the main theorem are convergent.

Proposition 3.1 (Finite improper integrals). *For $m \le n$ and $P \in \mathcal{P}_m^n$*

$$|P|_{st} \lesssim (t-s)^{(n-m)H} \tag{45}$$

uniformly over $\Delta^m[s,t]$.

Proof. We proceed by induction on $n - m$. When P only has single nodes ($m = n$) the statement is trivial. We will proceed by considering several cases for the last node in P; the simplest of these occurs when it is single: the statement follows immediately from the inductive hypothesis. For the next case, we will need the following bound:

$$\int_{\Delta^n[s,t]} (u_1 - s)^{2H-1} \cdots (u_n - u_{n-1})^{2H-1} \mathrm{d}u_1 \cdots \mathrm{d}u_n$$

$$\lesssim \int_{\Delta^{n-1}[s,t]} (u_1 - s)^{2H-1} \cdots (u_{n-1} - u_{n-2})^{2H-1} (t - u_{n-1})^{2H} \mathrm{d}u_1 \cdots \mathrm{d}u_{n-1}$$

$$\leq (t-s)^{2H} \int_{\Delta^{n-1}[s,t]} (u_1 - s)^{2H-1} \cdots (u_{n-1} - u_{n-2})^{2H-1} \mathrm{d}u_1 \cdots \mathrm{d}u_{n-1}$$

$$\lesssim \cdots \lesssim (t-s)^{2nH}.$$

For a diagram C whose last node is the right endpoint of an arc, using the bound above we have

$$|C \underbrace{\downarrow \cdots \downarrow}_{n}|_{st}$$

$$= \int_{\Delta^{n+1}[s,t]} |C|'_{su_0} (u_1 - u_0)^{2H-1} \cdots (u_n - u_{n-1})^{2H-1} \mathrm{d}u_0 \cdots \mathrm{d}u_n$$

$$= \int_s^t |C|'_{su_0} \int_{\Delta^n[u_0,t]} (u_1 - u_0)^{2H-1} \cdots (u_n - u_{n-1})^{2H-1} \mathrm{d}u_1 \cdots \mathrm{d}u_n \ \mathrm{d}u_0$$

$$\lesssim \int_s^t |C|'_{su_0} (t - u_0)^{2nH} \mathrm{d}u_0$$

$$\leq (t-s)^{2nH} \int_s^t |C|'_{su_0} \mathrm{d}u_0$$

$$\leq (t-s)^{2nH} |C|_{st}$$

where $|C|'_{su_0}$ equals the integral representing $|C|_{su_0}$ with the only difference that we are not integrating w.r.t. the variable u_0 in $(u_0 - r)^{2H-2}$, which represents the arc that terminates at the last node of C. Similarly, if the last node in C is single, we have

$$|C \underbrace{\downarrow \cdots \downarrow}_{n}|_{st}$$

$$= |C|_{su_0} \int_{\Delta^n[u_0,t]} (u_1 - u_0)^{2H-1} \cdots (u_n - u_{n-1})^{2H-1} \mathrm{d}u_1 \cdots \mathrm{d}u_n$$

$$\lesssim |C|_{su_0} (t-s)^{2nH}$$

$$\leq |C|_{st} (t-s)^{2nH}$$

where C is not differentiated since it terminates in a node representing a free variable, u_0. We now consider arcs: assume there are i arcs/lines within A, j within B, and that there are k arcs between nodes in A and nodes in B (collectively represented below by the dashed arc). Let A° and B° denote the diagrams given by eliminating such arcs from A and B: the nodes that have become single as a result now represent free variables, which we

call $w_1, \ldots, w_k, z_1, \ldots, z_k$. We first consider the case in which $j > 0$:

$$|\overset{\frown}{A \centerdot B \centerdot}|_{st}$$

$$= \int_{s<u<v<t} \int_{\Delta^k[s,u]\times\Delta^k[u,v]} |A^\circ|_{su} |B^\circ|_{uv} (z_1 - w_1)^{2H-2} \cdots (z_k - w_k)^{2H-2} \mathrm{d}w_1 \mathrm{d}z_1 \cdots \mathrm{d}w_k \mathrm{d}z_k$$
$$(v-u)^{2H-2} \mathrm{d}u \mathrm{d}v$$

$$\lesssim \int_{s<u<v<t} \int_{\Delta^k[s,u]\times\Delta^k[u,v]} (u-s)^{2iH} (v-u)^{2jH} (z_1 - w_1)^{2H-2} \cdots (z_k - w_k)^{2H-2} \mathrm{d}w_1 \mathrm{d}z_1 \cdots \mathrm{d}w_k \mathrm{d}z_k$$
$$(v-u)^{2H-2} \mathrm{d}u \mathrm{d}v$$

$$\leq (t-s)^{2iH} \int_{s<u<v<t} (v-u)^{2jH+2H-2} \int_{[s,u]^k\times[u,v]^k} (z_1 - w_1)^{2H-2} \cdots (z_k - w_k)^{2H-2} \mathrm{d}w_1 \mathrm{d}z_1 \cdots$$
$$\cdots \mathrm{d}w_k \mathrm{d}z_k \, \mathrm{d}u \mathrm{d}v$$

$$\lesssim (t-s)^{2iH} \int_{s<u<v<t} (v-u)^{2(j+1)H-2} |(v-u)^{2H} + (u-s)^{2H} - (v-s)^{2H}|^k \mathrm{d}u \mathrm{d}v$$

$$\lesssim (t-s)^{2(i+k)H} \int_{s<u<v<t} (v-u)^{2(j+1)H-2} \mathrm{d}u \mathrm{d}v$$

$$\leq (t-s)^{2(i+j+k)H}$$

where we have used $2H(j+1) - 1 \geq 4H - 1 > 0$ since $H > 1/4$. Note that the absolute values in the third-last expression can be removed by separately considering the cases $H > 1/2$ and $H < 1/2$. Assume instead $j = 0$: this means B must contain at least one node that is either single or paired with a node in A; it cannot be that $B = \varnothing$ or the diagram would contain an arc between two consecutive nodes, which is ruled out. The case in which there is a node in B which is single (see Figure 2) does not require $H > 1/4$: letting r denote the free variable represented by such a node, and proceeding similarly to the above, we have

$$|\overset{\frown}{A \centerdot B \centerdot}|_{st}$$

$$= \int_{s<u<r<v<t} \int_{\Delta^k[s,u]\times\Delta^k[u,v]} |A^\circ|_{su} (z_1 - w_1)^{2H-2} \cdots (z_k - w_k)^{2H-2} \mathrm{d}w_1 \mathrm{d}z_1 \cdots \mathrm{d}w_k \mathrm{d}z_k$$
$$(v-u)^{2H-2} \mathrm{d}u \mathrm{d}v$$

$$\lesssim (t-s)^{2(i+k)H} \int_{s<u<r<v<t} (v-u)^{2H-2} \mathrm{d}u \mathrm{d}v$$

$$\lesssim (t-s)^{2(i+k)H} |(t-r)^{2H} + (r-s)^{2H} - (t-s)^{2H}|$$

$$\lesssim (t-s)^{2(i+k+1)H} \, .$$

Finally, consider the case in which $j = 0$ and $k > 0$ (and B may have no single nodes):

$$|\overset{\frown}{A \centerdot B \centerdot}|_{st}$$

$$\lesssim (t-s)^{2iH} \int_{s<u<v<t} \int_{[s,u]^k\times\Delta^k[u,v]} (z_1 - w_1)^{2H-2} \cdots (z_k - w_k)^{2H-2} \mathrm{d}w_1 \mathrm{d}z_1 \cdots \mathrm{d}w_k \mathrm{d}z_k$$
$$(v-u)^{2H-2} \mathrm{d}u \mathrm{d}v$$

$$\lesssim (t-s)^{2iH} \int_{s<u<z_1<\ldots<z_k<t} |(z_1 - u)^{2H-1} - (z_1 - s)^{2H-1}| \cdots |(z_k - u)^{2H-1} - (z_k - s)^{2H-1}|$$
$$|(t-u)^{2H-1} - (z_k - u)^{2H-1}| \mathrm{d}u \mathrm{d}z_1 \cdots \mathrm{d}z_k$$

$$\lesssim (t-s)^{2(i+k-1)H} \int_{s<u<z_k<t} |(z_k - u)^{2H-1} - (z_k - s)^{2H-1}| |(t-u)^{2H-1} - (z_k - u)^{2H-1}| \mathrm{d}u \mathrm{d}z_k \, .$$

Once again, the absolute values distinguish between $H \lessgtr 1/2$. Expanding the product, we observe that three of the integrals feature products of different terms, each to the power of $2H - 1$: in these, at least one of z_k or u only appears once, which means this variable may be integrated out and the resulting term bounded (up to a constant) by $(t - s)^{2H}$, with the remaining integral solved similarly. The fourth integral instead is $\int_{s<u<z<t}(z - u)^{4H-2}\mathrm{d}u\mathrm{d}z$ which is finite again thanks to $H > 1/4$. This shows that we have $\lesssim (t - s)^{2(i+k+1)H}$ in the above expression and concludes the proof. ∎

Remark 3.2 (Modified $|P|$). We have stated the previous proposition under in the most natural manner; in particular note how, in the prototypical case of fBm, the integrals $|P|_{st}$ are multiples of P_{st}. We will, however, additionally need a slightly modified version of this result, in which the definition of $|P|$ is changed as follows: maximal sequences

$$\int_{\Delta^k[u,v]}(w_1 - u)^{2H-1}\cdots(w_k - w_{k-1})^{2H-1}\mathrm{d}w_1\cdots\mathrm{d}w_k$$

occurring in the middle of the expression for $|P|$, are replaced with their bound $(v-u)^{2kH}$, and each integrand $(v-u)^{2H-2}$ is replaced with $((v-u) \wedge 1/2)^{2H-2}$. That the statement continues despite these modifications to hold is obvious for the first, and for the second it follows from the facts that all integrals are still convergent (by the same proof) and the $1/2$ can be replaced with $1/2 \wedge T$ and absorbed in the constant of proportionality.

Just like in [3], we approximate X piecewise linearly. Let X^ℓ be a sequence of piecewise linear approximations of X along partitions π_ℓ on $[0, T]$ with step size that vanishes as $\ell \to \infty$. It will be helpful to assume that the intervals in the mesh π_ℓ all have the same length ϱ_ℓ; this simplifying assumption can be made because it is only necessary to show convergence along *a* sequence of such approximations, since it is known that the limit does not depend on the particular choice of π_ℓ (or indeed on the type of piecewise smooth approximation in a broad class of these) [21, Ch. 15]. For $t \in [0, T]$ we will write t_ℓ^- and t_ℓ^+ to respectively denote the endpoints a and b of the interval of π_ℓ s.t. $t \in [a, b)$. Explicitly, X^ℓ and its piecewise-defined derivative are given by

$$\begin{aligned} X_t^\ell &= X_{t_\ell^-} + \varrho_\ell^{-1}(t - t_\ell^-)X_{t_\ell^- t_\ell^+} \\ \dot{X}_t^\ell &= \varrho_\ell^{-1}X_{t_\ell^- t_\ell^+} \end{aligned} \tag{46}$$

where, as usual, $X_{ab} := X_b - X_a$ denotes the increment. In order to use Stroock's formula (23), we will be considering Malliavin derivatives of the signature of the piecewise-linear interpolations of X,

$$\mathbb{S}(X^\ell)_{st}^{\gamma_1,\dots,\gamma_n} = \int_{\Delta^n[s,t]}\dot{X}_{u_1}^{\ell;\gamma_1}\cdots\dot{X}_{u_n}^{\ell;\gamma_n}\mathrm{d}u_1\cdots\mathrm{d}u_n\,,$$

which in turn requires us to consider those of the single factors:

$$\mathcal{D}_v\dot{X}_u^{\ell;\gamma} = \varrho_\ell^{-1}\mathbb{1}_{[u_\ell^-,u_\ell^+)}^\gamma(v) = \varrho_\ell^{-1}\mathbb{1}_{[v_\ell^-,v_\ell^+)}^\gamma(u)\,. \tag{47}$$

For $P \in \mathcal{P}_m^n$, we provide a discretised analogue to Definition 2.1:

Definition 3.3 ($P_{st}^{\ell;\gamma_1,\dots,\gamma_n}$)**.** For $\gamma_1,\dots,\gamma_n \in [d]$, $0 \le s \le t \le T$ and $P \in \mathcal{P}_m^n$ define

$$P_{st}^{\ell;\gamma_1,\dots,\gamma_n}(v_k \mid k \in \overline{P}) := \int_{\Delta^n[s,t]}\prod_{\{i,j\}\in P}\mathbb{E}[\dot{X}_{u_i}^{\ell;\gamma_i}\dot{X}_{u_j}^{\ell;\gamma_j}]\mathrm{d}u_i\mathrm{d}u_j \cdot \prod_{k\in\overline{P}}\varrho_\ell^{-1}\mathbb{1}_{[v_{k;\ell}^-,v_{k;\ell}^+)}^{\gamma_k}(u_k)\mathrm{d}u_k\,. \tag{48}$$

as an element of $\mathcal{E}^{\otimes m}$, whose arguments are given to the functions $\mathbb{1}_{[u_{k;\ell}^-u_{k;\ell}^+)}^{\gamma_k}$ with $k \in \overline{P}$.

Note how the above definition, unlike Definition 2.1 does not distinguish between consecutive and non-consecutive pairings: this will only become important in the limit. Moreover, we are integrating over all n variables, including the u_k with $k \in \overline{P}$: this is because the time arguments of the function, v_k, are supplied separately, with the respective index variables supplied as arguments to δ_{γ_k}, $k \in \overline{P}$. The functions P_{st}^{ℓ} are summands in the expression of which we want to compute the limit:

Lemma 3.4 (Expected Malliavin derivatives of signature approximations)**.**

$$\mathbb{E}\mathcal{D}^m \mathbb{S}(X^\ell)_{st}^{\gamma_1,\dots,\gamma_n} = m! \sum_{P \in \mathcal{P}_m^n} P_{st}^{\ell;\gamma_1,\dots,\gamma_n} \in \mathcal{E}^{\odot m} \tag{49}$$

Proof. This is a consequence of (46), (47), the (iterated) Leibniz rule for the Malliavin derivative and Wick's formula for the mixed moments of a Gaussian vector (as it was already used in [2, Theorem 31]). The details are a matter of simple combinatorics; in particular note how, instead of summing over $m!$ terms corresponding to the ways of permuting the m derivatives (for a fixed $P \in \mathcal{P}_m^n$), we are only including the term corresponding to the identity permutation and multiplying by $m!$, which identifies the same element of $\mathcal{E}^{\otimes m}$ up to symmetry. ■

In order to prove convergence, it is unfortunately not possible to argue by dominated convergence applied to Definition 3.3: this is because the factors in the integrand given by consecutive pairings $\mathbb{E}[\dot{X}_{u_i}^{\ell;\gamma_i} \dot{X}_{u_{i+1}}^{\ell;\gamma_{i+1}}]$ converge to non-integrable functions (e.g. $(v - u)^{2H-2}$ on $\Delta^2[s,t]$ for fBm) and the ones corresponding to Malliavin derivatives $\varrho_\ell^{-1} \mathbb{1}_{[v_{k;\ell}^-, v_{k;\ell}^+)}^{\gamma_k}(u_k) \mathrm{d}u_k$ do not converge at all (in fact they converge, as distributions, to Dirac deltas δ_{v_k}). The reason that convergence holds is that all these quantities are integrated. To successfully exploit this, we will write each integral P_{st}^{ℓ} as a nested integral, distinguishing between the three types of integrands:

$$\int (\text{non-consecutive pairings}) \int (\text{Malliavin derivatives}) \prod_{\substack{\text{maximal} \\ \text{sequences}}} \int (\text{consecutive pairings}) . \tag{50}$$

The outer integral contains the product of all terms $\mathbb{E}[\dot{X}_{u_i}^{\ell;\gamma_i} \dot{X}_{u_j}^{\ell;\gamma_j}]$ with $|j - i| > 1$. These are multiplied with the second integral, which integrates all factors coming from Malliavin derivatives. Finally, we partition the remaining integrands $\mathbb{E}[\dot{X}_{u_h}^{\ell;\gamma_h} \dot{X}_{u_{h+1}}^{\ell;\gamma_{h+1}}]$ into maximal sequences and integrate each individually: these integrals are integrands in the second integral, alongside the Malliavin derivatives. The operations of exchanging the order of integrals are all justified by Fubini's theorem, considering that all integrals are actually finite sums. We illustrate all of this with a simple example: consider the diagram (suppressing indices)

$$P := \overset{\frown}{\cdot \cdot \cdot \cdot} \in \mathcal{P}_2^6 .$$

According to Definition 3.3, we have

$$P_{st}^{\ell}(v_1, v_2) = \int_{\Delta^6[s,t]} \varrho_\ell^{-2} \mathbb{1}_{[v_{2;\ell}^-, v_{2;\ell}^+)}(u_2) \mathbb{1}_{[v_{3;\ell}^-, v_{3;\ell}^+)}(u_3) \mathbb{E}[\dot{X}_{u_1}^{\ell} \dot{X}_{u_6}^{\ell}] \mathbb{E}[\dot{X}_{u_4}^{\ell} \dot{X}_{u_5}^{\ell}] \mathrm{d}u_1 \mathrm{d}u_4 \mathrm{d}u_5 \mathrm{d}u_6 .$$

Re-organising this expression as described in (50) we obtain

$$\int_{s < u_1 < u_6 < t} \mathbb{E}[\dot{X}_{u_1}^{\ell} \dot{X}_{u_6}^{\ell}]$$
$$\left[\int_{u_1 < u_2 < u_3 < u_6} \varrho_\ell^{-2} \mathbb{1}_{[v_{2;\ell}^-, v_{2;\ell}^+)}(u_2) \mathbb{1}_{[v_{3;\ell}^-, v_{3;\ell}^+)}(u_3) \right.$$
$$\left. \left[\int_{u_3 < u_4 < u_5 < u_6} \mathbb{E}[\dot{X}_{u_4}^{\ell} \dot{X}_{u_5}^{\ell}] \mathrm{d}u_4 \mathrm{d}u_5 \right] \mathrm{d}u_2 \mathrm{d}u_3 \right] \mathrm{d}u_1 \mathrm{d}u_6 .$$

Note that the domain of integration of the innermost integral can be described in terms of variables of the two outer integrals: this extends to the case in which there is more than one maximal sequence, by maximality, and is crucial for the factorisation into integrals over maximal sequences to be possible.

The reason for the nested rewriting of (50) is that it will be possible to show convergence of the integrals over maximal sequences, then by a separate argument infer the convergence of the middle integral, and finally by dominated convergence conclude that the outer integrals converge. We preface the proof of convergence with a few lemmas; the first of these considers the case of a single consecutive pairing, and will form the base case of an induction that handles maximal sequences of arbitrary length.

Lemma 3.5 (One consecutive pairing).

$$\lim_{\ell\to\infty}\int_{s<u<v<t}\mathbb{E}[\dot{X}_u^\ell\dot{X}_v^\ell]\mathrm{d}u\mathrm{d}v=\frac{1}{2}\mathbb{E}[X_{st}^2]=\int_s^t\big[\tfrac{1}{2}R'(v)-\partial_2R(s,v)\big]\mathrm{d}v$$

and the convergents are uniformly bounded by $\lesssim (t-s)^{2H}$.

Proof. Considering that $\dot{X}^\ell$ is a piecewise-constant, and that the integral on the right is therefore a finite sum, we can write

$$\int_{s<u<v<t}\mathbb{E}[\dot{X}_u^\ell\dot{X}_v^\ell]\mathrm{d}u\mathrm{d}v=\mathbb{E}\int_{s<u<v<t}\dot{X}_u^\ell\dot{X}_v^\ell\mathrm{d}u\mathrm{d}v$$

$$=\frac{1}{2}\mathbb{E}[(X_{st}^\ell)^2]$$

$$\xrightarrow{\ell\to\infty}\frac{1}{2}\mathbb{E}[X_{st}^2]$$

$$=\frac{1}{2}R(\Delta(s,t),\Delta(s,t))$$

$$=\frac{R(s)+R(t)}{2}-R(s,t)$$

$$=\int_s^t\big[\tfrac{1}{2}R'(v)-\partial_2R(s,v)\big]\mathrm{d}v$$

where we have used that $(X_{st}^\ell)^2\xrightarrow{\ell\to\infty}X_{st}^2$ in L^2. For the second statement, we rely on the first two identities above and distinguish between the cases $s_\ell^-=t_\ell^-$ and $s_\ell^-<t_\ell^-$: in the former we have, using (12)

$$|\mathbb{E}[(X_{st}^\ell)^2]|=|\mathbb{E}[\varrho_\ell^{-2}(t-s)^2X_{s_\ell^-s_\ell^+}^2]|\lesssim\varrho_\ell^{2H-2}(t-s)^2\le(t-s)^{2H}$$

since

$$\Big(\frac{t-s}{\varrho_\ell}\Big)^{2-2H}\le1$$

by $H<1$ and $t-s\le\varrho_\ell$. Let now $s_\ell^-<t_\ell^-$:

$$|\mathbb{E}[(X_{st}^\ell)^2]|=|\mathbb{E}[(X_{ss_\ell^+}^\ell+X_{s_\ell^+t_\ell^-}^\ell+X_{t_\ell^-t}^\ell)^2]|$$

$$=\Big|\mathbb{E}\Big[\big(\varrho^{-1}(s_\ell^+-s)X_{ss_\ell^+}^\ell+X_{s_\ell^+t_\ell^-}+\varrho^{-1}(t-t_\ell^-)X_{t_\ell^-t}\big)^2\Big]\Big|$$

$$\lesssim\varrho^{-2}(s_\ell^+-s)^2\mathbb{E}[X_{ss_\ell^+}^2]+\mathbb{E}[X_{s_\ell^+t_\ell^-}^2]+\varrho^{-2}(t-t_\ell^-)^2\mathbb{E}[X_{t_\ell^-t}^2]$$

$$\lesssim(s_\ell^+-s)^{2H}+(t_\ell^--s_\ell^+)^{2H}+(t-t_\ell^-)^{2H}$$

$$\lesssim(t-s)^{2H}$$

by ℓ^2-Jensen's inequality, the previous case, and again (12). ∎

The case of several consecutive pairings is more difficult to handle, and in Proposition 3.9 convergence of these terms will be bootstrapped from terms that only contain shorter sequences of consecutive pairings, and the above single case, by means of an inductive argument. It is worth remarking that the plausible strategy of handling these integrands together with the others by integrating only one of the variables fails:

Remark 3.6 (Lack of convergence of $\mathbb{E}[X_{uv}^{\ell} \dot{X}_v^{\ell}]$). One way of dealing with sequences of consecutive pairings is by rewriting them as

$$
\int_{s < u_1 < v_1 < \ldots < u_n < v_n < t} \mathbb{E}[\dot{X}_{u_1}^{\ell} \dot{X}_{v_1}^{\ell}] \cdots \mathbb{E}[\dot{X}_{u_n}^{\ell} \dot{X}_{v_n}^{\ell}] \mathrm{d}u_1 \mathrm{d}v_1 \cdots \mathrm{d}u_n \mathrm{d}v_n
$$
$$
= \int_{\Delta^n[s,t]} \mathbb{E}[X_{sv_1}^{\ell} \dot{X}_{v_1}^{\ell}] \mathbb{E}[X_{v_1 v_2}^{\ell} \dot{X}_{v_2}^{\ell}] \cdots \mathbb{E}[X_{v_{n-1} v_k}^{\ell} \dot{X}_{v_n}^{\ell}] \mathrm{d}v_1 \cdots \mathrm{d}u_n \,. \tag{51}
$$

This has the benefit of expressing the convergents as integrals over n, and not $2n$, variables. The problem with this strategy is that it does not hold that $\mathbb{E}[X_{uv}^{\ell} \dot{X}_v^{\ell}] \xrightarrow{\ell \to \infty} \frac{1}{2} R'(v) - \partial_2 R(u, v)$: a simple calculation reveals

$$
\mathbb{E}[X_{uv}^{\ell} \dot{X}_v^{\ell}]
$$
$$
= \varrho_\ell^{-1} \Big[\big(1 - \varrho_\ell^{-1}(v - v_\ell^-)\big) R(v_\ell^-, \Delta(v_\ell^-, v_\ell^+)) + \varrho_\ell^{-1}(v - v_\ell^-) R(v_\ell^+, \Delta(v_\ell^-, v_\ell^+)) \Big]
$$
$$
- \varrho_\ell^{-1} \Big[\big(1 - \varrho_\ell^{-1}(u - u_\ell^-)\big) R(u_\ell^-, \Delta(v_\ell^-, v_\ell^+)) + \varrho_\ell^{-1}(u - u_\ell^-) R(u_\ell^+, \Delta(v_\ell^-, v_\ell^+)) \Big] \,.
$$

While the second term converges to $\partial_2 R(u, v)$ (e.g. by the intermediate value theorem applied on the interval $[v_\ell^-, v_\ell^+]$), the first does not converge in general. To see why, it suffices to take X to be Brownian motion and π_ℓ to by a diadic sequence: the first term on the right above is then equal to $\varrho_\ell^{-1}(v - v_\ell^-)$ which is indeterminate in view of the fact that for v in a set of full Lebesgue measure its decimal expansion contains infinitely many 00's and 11's. The fractional case with $H < 1/2$ appears even worse behaved, i.e. divergent in a possibly indeterminate fashion.

We now move outward in (50) and prove a lemma that will guarantee convergence of the middle integral, conditional on the convergence of the inner ones.

Lemma 3.7. *Let $f_\ell \colon [0, T]^m \to \mathbb{R}$ be a uniformly bounded sequence of functions that are continuous and piecewise smooth on the mesh π_ℓ. Assume that f_ℓ converges to $f \colon [0, T]^m \to \mathbb{R}$ uniformly. Then*

$$
\int_{\Delta^m[s,t]} f_\ell(u_1, \ldots, u_m) \varrho_\ell^{-m} \prod_{k=1}^m \mathbb{1}_{[v_{k;\ell}^-, v_{k;\ell}^+)}(u_k) \mathrm{d}u_1, \ldots, \mathrm{d}u_m \xrightarrow{\ell \to \infty} \mathbb{1}_{\Delta^m[s,t]}(v_1, \ldots, v_m) f(v_1, \ldots, v_m)
$$

where the convergence is a.e. in the variables $(v_1, \ldots, v_m) \in [0, T]^m$. Moreover, the convergents are uniformly bounded by $\sup_\ell \| f_\ell \|_\infty$.

Proof. The second statement holds by uniform boundedness of f_ℓ and the fact that

$$
\int_{\Delta^m[s,t]} \varrho_\ell^{-m} \prod_{k=1}^m \mathbb{1}_{[v_{k;\ell}^-, v_{k;\ell}^+)}(u_k) \mathrm{d}u_1, \ldots, \mathrm{d}u_n \leq 1 \,.
$$

We will prove pointwise convergence on the subset

$$
[0, T]_*^m := \{ (v_1, \ldots, v_m) \in [0, T]^m \mid v_i \neq v_j \text{ for } i \neq j \}
$$

of $[0, T]^m$ of full Lebesgue measure. For $(v_1, \ldots, v_m) \in [0, T]_*^m$, we may, without loss of generality, start the sequence when ℓ is already large enough so that $[v_{i;\ell}^-, v_{i;\ell}^+) \cap [v_{j;\ell}^-, v_{j;\ell}^+) = \varnothing$ for $i \neq j$, where we are including

$v_0 := s$ and $v_{m+1} := t$ in this requirement. By the mean value theorem applied individually to each u_k, there exist $w_{k;\ell} \in (v_{k;\ell}^-, v_{k;\ell}^+)$ s.t.

$$\int_{\Delta^m[s,t]} f_\ell(u_1, \ldots, u_m) \varrho_\ell^{-m} \prod_{k=1}^m \mathbb{1}_{[v_{k;\ell}^-, v_{k;\ell}^+]}(u_k) du_1, \ldots, du_m = \mathbb{1}_{\Delta^m[s,t]}(v_1, \ldots, v_m) f_\ell(w_{1;\ell}, \ldots, w_{m;\ell})$$

and

$$
\begin{aligned}
&|\mathbb{1}_{\Delta^m[s,t]}(v_1, \ldots, v_m) f_\ell(w_{1;\ell}, \ldots, w_{m;\ell}) - \mathbb{1}_{\Delta^m[s,t]}(v_1, \ldots, v_m) f(v_1, \ldots, v_m)| \\
&\leq \mathbb{1}_{\Delta^m[s,t]}(v_1, \ldots, v_m) \big[|f_\ell(w_{1;\ell}, \ldots, w_{m;\ell}) - f(w_{1;\ell}, \ldots, w_{m;\ell})| \\
&\quad + |f(w_{1;\ell}, \ldots, w_{m;\ell}) - f(v_1, \ldots, v_m)| \big] \\
&\leq \mathbb{1}_{\Delta^m[s,t]}(v_1, \ldots, v_m) \big[\|f_\ell - f\|_\infty + \omega_{(v_1,\ldots,v_m)}^f(\varrho_\ell) \big]
\end{aligned}
$$

where $\omega_{(v_1,\ldots,v_m)}^f$ is the modulus of continuity of f at the point $(v_1, \ldots, v_m)$. Both summands on the right hand side above vanish in the limit of $\ell \to \infty$, the first by uniform convergence and the second by continuity of the uniform limit of continuous functions. ∎

The next two results constitute the core of our argument. They both rely on the same induction used to reduce the length of consecutive pairings, the base case of which is provided by Lemma 3.5. To illustrate it at level 4, letting Y be a stochastic process (which below will be taken to be X^ℓ and X) we have for $\alpha \neq \beta$

$$2\mathbb{E}\mathbb{S}(Y)_{st}^{\alpha\alpha\beta\beta} = \mathbb{E}\mathbb{S}(Y)_{st}^{\alpha\alpha} \cdot \mathbb{E}\mathbb{S}(Y)_{st}^{\beta\beta} - 2\mathbb{E}\mathbb{S}(Y)_{st}^{\alpha\beta\alpha\beta} - 2\mathbb{E}\mathbb{S}(Y)_{st}^{\alpha\beta\beta\alpha}$$

by the shuffle property (8), using identical distribution of components to group together $2\mathbb{E}\mathbb{S}(Y)_{st}^{\alpha\alpha\beta\beta} = \mathbb{E}\mathbb{S}(Y)_{st}^{\alpha\alpha\beta\beta} + \mathbb{E}\mathbb{S}(Y)_{st}^{\beta\beta\alpha\alpha}$ (and similar on the right hand side), and using independence of components to write $\mathbb{E}[\mathbb{S}(Y)_{st}^{\alpha\alpha}\mathbb{S}(Y)_{st}^{\beta\beta}] = \mathbb{E}\mathbb{S}(Y)_{st}^{\alpha\alpha} \cdot \mathbb{E}\mathbb{S}(Y)_{st}^{\beta\beta}$. While the left hand side contains a sequence of two consecutive pairs, only sequences of consecutive pairs of length one appear on the right.

Lemma 3.8 (Dominating function). *For $P \in \mathcal{P}_m^n$ it holds that the integrand of the outermost integral of P_{st}^ℓ expressed in the nested form (50), is absolutely bounded by an integrable function, uniformly in ℓ and on $\Delta^m[s,t]$, so that $|P_{st}^\ell| \lesssim (t-s)^{2(n-m)H'}$ for any $1/4 < H' < H$.*

Proof. We begin by bounding expectations corresponding to non-consecutive pairings. As done in the proof of [3, Theorem 31], we now consider the terms

$$\mathbb{E}[\dot{X}_u^\ell \dot{X}_v^\ell] = \varrho_\ell^{-2} R(\Delta(u_\ell^-, u_\ell^+), \Delta(v_\ell^-, v_\ell^+))$$

in three different cases: for $u_\ell^- = v_\ell^-$

$$|\mathbb{E}[\dot{X}_u^\ell \dot{X}_v^\ell]| \lesssim \varrho_\ell^{2H-2} \leq (v-u)^{2H-2}.$$

By Cauchy-Schwarz the same estimate as above holds in the case $u_\ell^+ = v_\ell^-$, with a constant in the second inequality given by the fact that $v - u \leq 2\varrho_\ell$. Let $u_\ell^+ < v_\ell^-$: we have, by (9) and for any H' as in the statement

$$
\begin{aligned}
|\mathbb{E}[\dot{X}_u^\ell \dot{X}_v^\ell]| = \varrho_\ell^{-2} \left| \int_{[u_\ell^-, u_\ell^+] \times [v_\ell^-, v_\ell^+]} \partial_{12} R(u,v) du dv \right| \\
\lesssim \varrho_\ell^{-2} \int_{[u_\ell^-, u_\ell^+] \times [v_\ell^-, v_\ell^+]} (v-u)^{2H-2} du dv \\
\leq (v_\ell^- - u_\ell^+)^{2H-2} \\
\lesssim ((v_\ell^+ - u_\ell^-) \wedge 1/2)^{2H'-2} \\
\leq ((v-u) \wedge 1/2)^{2H'-2}
\end{aligned}
$$

In the second-last inequality we have used that there exists some L s.t. for all $\ell \geq L$

$$\vartheta^{2H-2} \leq (\vartheta + 2\varrho_\ell)^{2H'-2}$$

for all $\vartheta \in [\varrho_\ell, 1/2]$.

We now consider terms corresponding to maximal sequences of consecutive pairings, i.e.

$$\int_{s<u_1<v_1<...<u_k<v_k<t} \mathbb{E}[\dot{X}^\ell_{u_1}\dot{X}^\ell_{v_1}]\cdots\mathbb{E}[\dot{X}^\ell_{u_k}\dot{X}^\ell_{v_k}]\mathrm{d}u_1\mathrm{d}v_1\cdots\mathrm{d}u_k\mathrm{d}v_k\,. \tag{52}$$

It is always possible (e.g. by Kolmogorov's extension theorem) to add independent components to X. With this in mind, by Wick's theorem we may write the above integral as $\mathbb{E}\mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_k\alpha_k}_{st}$ with $\alpha_i \neq \alpha_j$ for all $i \neq j$. By the shuffle identity (8) we have

$$
\begin{aligned}
&\sum_{h=0}^n \mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_h\alpha_h\beta\beta\alpha_{h+1}\alpha_{h+1}...\alpha_k\alpha_k}_{st}\\
&= \mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_k\alpha_k}_{st}\mathcal{S}(X^\ell)^{\beta\beta}_{st} - \sum_{0\leq i<j\leq k} \mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_i\alpha_i\beta\alpha_{i+1}\alpha_{i+1}...\alpha_j\alpha_j\beta\alpha_{j+1}\alpha_{j+1}...\alpha_k\alpha_k}_{st}\\
&\quad - \sum_{0\leq i<j\leq k} \mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_i\alpha_i\beta\alpha_{i+1}\alpha_{i+1}...\alpha_{j+1}\beta\alpha_{j+1}...\alpha_k\alpha_k}_{st}\\
&\quad - \sum_{0\leq i<j\leq k} \mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_i\beta\alpha_i...\alpha_j\alpha_j\beta\alpha_{j+1}\alpha_{j+1}...\alpha_k\alpha_k}_{st}\\
&\quad - \sum_{0\leq i<j\leq k} \mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_i\beta\alpha_i...\alpha_j\beta\alpha_j...\alpha_k\alpha_k}_{st}\\
&\quad - \sum_{h=0}^k \left(\mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_h\beta\alpha_h\beta\alpha_{h+1}\alpha_{h+1}...\alpha_k\alpha_k}_{st} + \mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_{h-1}\alpha_{h-1}\beta\alpha_h\beta\alpha_h...\alpha_k\alpha_k}_{st}\right).
\end{aligned}
\tag{53}
$$

When shuffling we have separated the cases in which all $\alpha_h\alpha_h$ and $\beta\beta$ occur as consecutive pairs, from those in which at least one such pair is separated. We now take expectations: note that both independence and equal distribution of components are used.

$$(k+1)\mathbb{E}\mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_k\alpha_k\beta\beta}_{st} = \mathbb{E}\mathcal{S}(X^\ell)^{\alpha_1\alpha_1...\alpha_k\alpha_k}_{st} \cdot \tfrac{1}{2}\mathbb{E}[(X^\ell)^2_{st}] - \sum_Q Q^\ell_{st} \tag{54}$$

where we are summing over a finite number of diagrams Q whose longest sequence of consecutive pairings contains k pairs or fewer.

We now prove the statement in the case in which P has no single nodes, by induction on n. For $n = 0$, $P^\ell_{st} = \varnothing^\ell_{st} \equiv 1$ there is nothing to show. Let $n \geq 1$, and assume we have rewritten the integral according to (50) (where the middle integral may be skipped, since there are no Malliavin derivatives). If P is not given by a sequence of $n/2$ consecutive pairs, all maximal sequences of consecutive pairs in P consist of fewer than n pairs, and that thus the inductive hypothesis applies to them: this means that for each such sequence Q with k pairs, $|Q^\ell_{uv}| \lesssim (v-u)^{2kH'}$. Using the bounds for the first two types of integrand derived in the first part of this proof, the statement for P then follows from Proposition 3.1 applied in the modified case of Remark 3.2 and with exponent H'. Assume now $n = 2(k+1)$ and let P be given by the diagram consisting of $k+1$ consecutive pairs: the only thing needed to conclude the induction is the bound. This follows from (54) thanks to the inductive hypothesis and the boundedness statement of Lemma 3.5.

Finally, we consider the general case in which P may have single nodes. This follows again by writing P^ℓ_{st} in nested form, bounding terms corresponding to non-consecutive pairings as done above, and bounding the middle integral in (50) thanks to the boundedness statement of Lemma 3.7. When invoking this lemma, f_ℓ is

going to be a product of terms of the form (52) (with the extrema s and t replaced with variables u_i and u_j already integrated in the outer or middle integral), which as already proved is bounded by $\lesssim (t-s)^{2H'k}$: this yields the required bound overall. ∎

Proposition 3.9 (Convergence). *The functions $[0,T]^m \to \mathbb{R}$ of Definition 3.3 individually converge a.e. to those of Definition 2.1: for $P \in \mathcal{P}_m^n$ it holds that*

$$P_{st}^\ell \xrightarrow{\ell \to \infty} P_{st}. \tag{55}$$

Moreover $|P_{st}| \lesssim |P|_{st}$ (the integrals of Proposition 3.1) uniformly on $\Delta^m[s,t]$.

Proof. The inequality is an absolute estimate of P_{st} using (9) and (10). The structure of the proof of the first statement closely mirrors that of the previous lemma: we first consider the case in which P does not have single nodes. For $u_\ell^- < v_\ell^-$

$$\mathbb{E}[\dot{X}_u^\ell \dot{X}_v^\ell] = \varrho_\ell^{-2} R(\Delta(u_\ell^-, u_\ell^+), \Delta(v_\ell^-, v_\ell^+)) = \partial_{12} R(\overline{u}, \overline{v})$$

for some $\overline{u} \in (u_\ell^-, u_\ell^+), \overline{v} \in (v_\ell^-, v_\ell^+)$, by the intermediate value theorem applied twice. Pointwise convergence $\mathbb{E}[\dot{X}_u^\ell \dot{X}_v^\ell] \to \partial_{12} R(u,v)$ then holds by continuity of $\partial_{12} R$ and thanks to the fact that for any $u < v$ there exists L s.t. $u_\ell^- < v_\ell^-$ for all $\ell \geq L$. This takes care of convergence of terms corresponding to non-consecutive pairings (of course, the same holds for consecutive pairings, but is not useful since $\partial_{12} R(u,v)$ may not be integrable in this case).

We now proceed by induction on n. For $n = 0$ there is nothing to prove, so let $n \geq 1$ and first consider the case in which P is not given by a sequence of $n/2$ consecutive pairs: the statement follows by dominated convergence applied to the outer integral in (50), by the above and the inductive hypothesis applied to sequences of consecutive nodes of length less than n, in conjunction with Lemma 3.8. Let now $n = 2(k+1)$ and let P be given by the diagram consisting of $k+1$ consecutive pairs: recalling the argument (and indexing notation) of the previous proof, we have $P_{st}^\ell = \mathbb{E}\mathcal{S}(X^\ell)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_k \alpha_k \beta \beta}$, which is convergent since $\mathcal{S}(X^\ell)_{st} \to \mathcal{S}(X)_{st}$ in L^2. By same calculation of (53) applied to X instead of to X^ℓ, and taking expectations

$$
\begin{aligned}
&(k+1) \lim_{\ell \to \infty} P_{st}^\ell \\
&= (k+1) \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_k \alpha_k \beta \beta} \\
&= \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_k \alpha_k} \cdot \tfrac{1}{2} \mathbb{E}[(X)_{st}^2] - \sum_{0 \leq i < j \leq k} \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_i \alpha_i \beta \alpha_{i+1} \alpha_{i+1} \ldots \alpha_j \alpha_j \beta \alpha_{j+1} \alpha_{j+1} \ldots \alpha_k \alpha_k} \\
&\quad - \sum_{0 \leq i < j \leq k} \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_i \alpha_i \beta \alpha_{i+1} \alpha_{i+1} \ldots \alpha_{j+1} \beta \alpha_{j+1} \ldots \alpha_k \alpha_k} \\
&\quad - \sum_{0 \leq i < j \leq k} \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_i \beta \alpha_i \ldots \alpha_j \alpha_j \beta \alpha_{j+1} \alpha_{j+1} \ldots \alpha_k \alpha_k} \\
&\quad - \sum_{0 \leq i < j \leq k} \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_i \beta \alpha_i \ldots \alpha_j \beta \alpha_j \ldots \alpha_k \alpha_k} \\
&\quad - \sum_{h=0}^{k} \left(\mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_h \beta \alpha_h \beta \alpha_{h+1} \alpha_{h+1} \ldots \alpha_k \alpha_k} + \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_{h-1} \alpha_{h-1} \beta \alpha_h \beta \alpha_h \ldots \alpha_k \alpha_k} \right).
\end{aligned}
\tag{56}
$$

We now expand the product: by the inductive hypothesis and Lemma 3.5, and using Fubini's theorem we have (setting $u_0 = s = w_0$)

$$
\begin{aligned}
&\mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1 \alpha_1 \ldots \alpha_k \alpha_k} \cdot \mathbb{E}[X_{st}^2] \\
&= \int_{s < u_1 < \ldots < u_k < t} \left[\tfrac{1}{2} R'(u_1) - \partial_2 R(u_0, u_1) \right] \cdots \left[\tfrac{1}{2} R'(u_k) - \partial_2 R(u_{k-1}, u_k) \right] \mathrm{d}u_1 \cdots \mathrm{d}u_k
\end{aligned}
$$

$$\cdot \int_s^t \left[\tfrac{1}{2}R'(v) - \partial_2 R(s,v)\right] \mathrm{d}v$$

$$= \int_{\substack{s<u_1<...<u_k<t \\ s<v<t}} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right] \cdots \left[\tfrac{1}{2}R'(u_k) - \partial_2 R(u_{k-1},u_k)\right]$$
$$\cdot \left[\tfrac{1}{2}R'(v) - \partial_2 R(s,v)\right] \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}v$$

$$= \sum_{j=0}^k \int_{s<u_1<...<u_j<v<u_{j+1}<...<u_k<t} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right] \cdots \left[\tfrac{1}{2}R'(u_k) - \partial_2 R(u_{k-1},u_k)\right]$$
$$\cdot \left[\tfrac{1}{2}R'(v) - \partial_2 R(s,v)\right] \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}v \,.$$

Note that the use of Fubini's theorem is justified in view of (10) applied to absolutely bound each integral above, and Proposition 3.1. Writing

$$\partial_2 R(\Delta(x,y),z) := \partial_2 R(y,z) - \partial_2 R(x,z) = \int_x^y \partial_{12} R(w,z) \mathrm{d}w \,,$$

we expand each summand:

$$\int_{s<u_1<...<u_j<v<u_{j+1}<...<u_k<t} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right] \cdots \left[\tfrac{1}{2}R'(u_k) - \partial_2 R(u_{k-1},u_k)\right]$$
$$\cdot \left[\tfrac{1}{2}R'(v) - \partial_2 R(s,v)\right] \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}v$$

$$= \int_{s<u_1<...<u_j<v<u_{j+1}<...<u_k<t} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right] \cdots \left[\tfrac{1}{2}R'(u_j) - \partial_2 R(u_{j-1},u_j)\right]$$
$$\cdot \left[\tfrac{1}{2}R'(v) - \partial_2 R(u_j,v) + \sum_{i=0}^{j-1} \partial_2 R(\Delta(u_i,u_{i+1}),v)\right]$$
$$\cdot \left[\tfrac{1}{2}R'(u_{j+1}) - \partial_2 R(v,u_{j+1}) + \partial_2 R(\Delta(u_j,v),u_{j+1})\right]$$
$$\cdot \left[\tfrac{1}{2}R'(u_{j+2}) - \partial_2 R(u_{j+1},u_{j+2})\right] \cdots \left[\tfrac{1}{2}R'(u_k) - \partial_2 R(u_{k-1},u_k)\right] \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}v$$

$$= \int_{s<u_1<...<u_j<v<u_{j+1}<...<u_k<t} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right] \cdots \left[\tfrac{1}{2}R'(v) - \partial_2 R(u_j,v)\right]$$
$$\cdot \left[\tfrac{1}{2}R'(u_{j+1}) - \partial_2 R(u_{j+1},v)\right] \cdots \left[\tfrac{1}{2}R'(u_k) - \partial_2 R(u_{k-1},u_k)\right] \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}v$$

$$+ \int_{s<u_1<...<u_j<r<v<u_{j+1}<...<u_k<t} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right] \cdots \left[\tfrac{1}{2}R'(u_j) - \partial_2 R(u_{j-1},u_j)\right]$$
$$\cdot \left[\tfrac{1}{2}R'(v) - \partial_2 R(r,v) + \partial_2 R(\Delta(u_j,r),v)\right] \partial_{12} R(r,u_{j+1}) \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}r \mathrm{d}v$$

$$+ \sum_{i=0}^{j-1} \int_{s<u_1<...<u_i<q<u_{i+1}<...<u_j<v<u_{j+1}<...<u_k<t} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right]$$
$$\cdots \partial_{12} R(q,v) \left[\tfrac{1}{2}R'(u_{i+1}) - \partial_2 R(q,u_{i+1}) + \partial_2 R(\Delta(u_i,q),u_{i+1})\right] \cdots \left[\tfrac{1}{2}R'(u_{j+1}) - \partial_2 R(v,u_{j+1})\right]$$
$$\cdots \left[\tfrac{1}{2}R'(u_k) - \partial_2 R(u_{k-1},u_k)\right] \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}q \mathrm{d}v$$

$$+ \sum_{i=0}^{j-1} \int_{s<u_1<...<u_i<q<u_{i+1}<...<u_j<r<v<u_{j+1}<...<u_k<t} \left[\tfrac{1}{2}R'(u_1) - \partial_2 R(u_0,u_1)\right]$$
$$\cdots \partial_{12} R(q,v) \left[\tfrac{1}{2}R'(u_{i+1}) - \partial_2 R(q,u_{i+1}) + \partial_2 R(\Delta(u_i,q),u_{i+1})\right] \cdots \partial_{12} R(r,u_{j+1})$$
$$\cdots \left[\tfrac{1}{2}R'(u_k) - \partial_2 R(u_{k-1},u_k)\right] \mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}q \mathrm{d}r \mathrm{d}v$$

$$= (\underbrace{\downarrow \ldots \downarrow}_{k+1})_{st} + (\underbrace{\downarrow \ldots \downarrow}_{j} \,\overgroup{\downarrow\downarrow}\, \underbrace{\downarrow \ldots \downarrow}_{k-j-1})_{st} + (\underbrace{\downarrow \ldots \downarrow}_{j} \,\overgroup{\downarrow\downarrow}\, \underbrace{\downarrow \ldots \downarrow}_{k-j-1})_{st}$$

$$+ \sum_{i=0}^{j-1} (\underbrace{\downarrow \ldots \downarrow}_{i} \,\bullet\, \underbrace{\overbrace{\downarrow \ldots \downarrow}}_{j-i}\, \bullet\, \underbrace{\downarrow \ldots \downarrow}_{k-j})_{st} + \sum_{i=0}^{j-1} (\underbrace{\downarrow \ldots \downarrow}_{i} \,\bullet\, \underbrace{\overbrace{\downarrow \ldots \downarrow}}_{j-i-1}\, \underbrace{\downarrow \ldots \downarrow}_{k-j})_{st}$$

$$+ \sum_{i=0}^{j-1} (\underbrace{\downarrow \ldots \downarrow}_{i} \; \overbrace{\underbrace{\downarrow \ldots \downarrow}_{j-i}}^{} \; \overparen{} \; \underbrace{\downarrow \ldots \downarrow}_{k-j-1})_{st} + \sum_{i=0}^{j-1} (\underbrace{\downarrow \ldots \downarrow}_{i} \; \overparen{} \; \overbrace{\underbrace{\downarrow \ldots \downarrow}_{j-i-1}}^{} \; \overparen{} \; \underbrace{\downarrow \ldots \downarrow}_{k-j-2})_{st}$$

$$= (k+1)P_{st} + \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1\alpha_1\ldots\alpha_j\alpha_j\beta\alpha_{j+1}\alpha_{j+1}\beta\alpha_{j+2}\alpha_{j+2}\ldots\alpha_k\alpha_k} + \mathbb{E}\mathcal{S}(X)_{st}^{\alpha_1\alpha_1\ldots\alpha_j\alpha_j\beta\alpha_{j+1}\beta\alpha_{j+1}\ldots\alpha_k\alpha_k}$$

$$+ \sum_{i=0}^{j-1}\mathcal{S}(X)_{st}^{\alpha_1\alpha_1\ldots\alpha_i\alpha_i\beta\alpha_{i+1}\alpha_{i+1}\ldots\alpha_j\alpha_j\beta\alpha_{j+1}\alpha_{j+1}\ldots\alpha_k\alpha_k} + \sum_{i=0}^{j-1}\mathcal{S}(X)_{st}^{\alpha_1\alpha_1\ldots\alpha_{i+1}\beta\alpha_{i+1}\ldots\alpha_j\alpha_j\beta\alpha_{j+1}\alpha_{j+1}\ldots\alpha_k\alpha_k}$$

$$+ \sum_{i=0}^{j-1}\mathcal{S}(X)_{st}^{\alpha_1\alpha_1\ldots\alpha_i\alpha_i\beta\alpha_{i+1}\alpha_{i+1}\ldots\alpha_{j+1}\beta\alpha_{j+1}\ldots\alpha_k\alpha_k} + \sum_{i=0}^{j-1}\mathcal{S}(X)_{st}^{\alpha_1\alpha_1\ldots\alpha_{i+1}\beta\alpha_{i+1}\ldots\alpha_{j+1}\beta\alpha_{j+1}\ldots\alpha_k\alpha_k} \;.$$

It now follows by substitution into the sum $\sum_{j=0}^{k}$ and simplifying in (56) that $\lim_\ell P_{st}^\ell = P_{st}$.

Finally, we consider diagrams that contain single nodes. In order to invoke Lemma 3.7 we must argue that $f_\ell \to f$ uniformly (uniform boundedness holds by the previous lemma). This again follows from the fact that f_ℓ can be written as a product of expected signatures of X^ℓ, each of which converges uniformly in ℓ as a function of its extrema: recalling the notations for truncation and projection introduced in Section 1 and the definition of inhomogeneous p-variation distance [21, §8.1.2], we have

$$\sup_{u<v} |\mathbb{E}\mathcal{S}(X)_{uv}^{(n)} - \mathbb{E}\mathcal{S}(X^\ell)_{uv}^{(n)}| \leq \mathbb{E}\sup_{u<v} |\mathcal{S}(X)_{uv}^{(n)} - \mathcal{S}(X^\ell)_{uv}^{(n)}|$$

$$\leq \|\rho_{p\text{-var}}(\mathcal{S}^{\lfloor p \rfloor}(X^\ell), \mathcal{S}^{\lfloor p \rfloor}(X))\|_{L^1}$$

$$\xrightarrow{\ell\to\infty} 0$$

for $p > (1/H) \vee n$, where we have used [18, Theorem 1]. The statement now follows once again by dominated convergence and Fubini's theorem. ∎

We are ready to put it all together:

Proof of Theorem 2.3.

$$w^m\mathcal{S}(X)_{st}^{\gamma_1,\ldots,\gamma_n} = \frac{1}{m!}\delta^m\big(\mathbb{E}\mathcal{D}^m\mathcal{S}(X)_{st}^{\gamma_1,\ldots,\gamma_n}\big) \tag{57}$$

$$= \frac{1}{m!}\delta^m \lim_{\ell\to\infty}\big(\mathbb{E}\mathcal{D}^m\mathcal{S}(X^\ell)_{st}^{\gamma_1,\ldots,\gamma_n}\big) \tag{58}$$

$$= \delta^m \sum_{P\in\mathcal{P}_m^n} \lim_{\ell\to\infty} P_{st}^{\ell;\gamma_1,\ldots,\gamma_n} \tag{59}$$

$$= \sum_{P\in\mathcal{P}_m^n} \delta^m P_{st}^{\gamma_1,\ldots,\gamma_n} \tag{60}$$

In (57) we have used Stroock's formula (23), which is possible since $\mathcal{S}(X)_{st}^{\gamma_1,\ldots,\gamma_n} \in \mathbb{D}^{\infty,2}$: this is because $\mathcal{S}(X^\ell)_{st} \to \mathcal{S}(X)_{st}$ in $\bigoplus_{k\leq n}\mathscr{W}^k$ which is closed in $L^2\Omega$. In (58) we have used that convergence of $\mathcal{S}(X^\ell)_{st}$ actually holds in $\mathbb{D}^{\infty,2}$, since the norm of $\mathbb{D}^{\infty,2}$ is dominated by the L^2 norm in bounded Wiener chaos [35, Proposition 1.2.2]. (59) uses Lemma 3.4 and (60) is just the statement (required by our definition of membership of a function to $\mathcal{H}^{\otimes m}$ Definition 1.1) that $P_{st}^{\ell;\gamma_1,\ldots,\gamma_n}$ converges a.e. boundedly to P_{st}^ℓ, which holds by Proposition 3.9 and Lemma 3.8. As argued in the previous two proofs, $P_{st}^{\gamma_1,\ldots,\gamma_n}$ can always be expressed as the expected signature evaluated on a word, up to augmenting X with independent copies of itself: this can be used to infer that each $P_{st}^{\gamma_1,\ldots,\gamma_n}$ — not just their sum — belongs to $\mathbb{D}^{\infty,2}(\mathcal{H}^{\otimes m})$. This concludes the proof of the main result. ∎

Conclusions and further directions

By providing a single formula for the expected signature of fractional Brownian motion that holds for any Hurst parameter $H \in (1/4, 1)$, this article closes a gap in the literature left open by [3]. Along the way, we have had opportunity to consider numerous other aspects of our computation, such as similar formulae for higher levels of the Wiener chaos expansion of the signature, and other examples of Gaussian processes.

We believe this work recommends a variety of applications and further investigations. First and foremost, it would be interesting to write stochastic Taylor expansions as suggested by Example 2.8, under precise conditions on the vector fields, and by providing bounds on the mean square error. Making this calculation rigorous and providing precise asymptotic estimates such as those in [37] would be an interesting result, which could be applied to approximation problems for Gaussian RDEs on manifolds such as those considered in [1] for SDEs (although for this precise problem, the joint signature $\mathbb{S}(X, t)$ would have to be considered). A further step would involve proving conditional versions of the results in this paper, which would make it possible to estimate the error generated by multiple steps in an Euler scheme.

The fact that (e.g. for fBm) the integral $\mathbb{E}\mathbb{S}(X)_{st}^{\alpha_1\alpha_1\cdots\alpha_k\alpha_k}$ with $\alpha_i \neq \alpha_j$ is actually convergent for any $H > 0$ raises the question of whether something can be said about the sequence $\mathbb{S}(X^\ell)_{st}^{\alpha_1\alpha_1\cdots\alpha_k\alpha_k}$, i.e. by considering the particular word on which $\mathbb{S}(X^\ell)$, which is not convergent in probability for $H \leq 1/4$, is evaluated.

It would be interesting to express the expected signature of a Gaussian process as the exponential of a formal series of tensors, thus computing its signature cumulants [8]: this is how the expected signature of Brownian motion (2) is usually presented (with the series a finite sum), but the analogous formulation for Gaussian processes that are not martingales appears more difficult to write down.

A more computational goal, though not one that appears trivial, is to explicitly compute Theorem 2.3 for certain semimartingales, such as the Brownian bridge, for which the integrals are all analytically solvable. An interesting question is how the relationship between Brownian motion and Brownian bridge is reflected by their expected signatures. It would also be helpful to see whether similar formulae to ours can be made available for non-centred Gaussian processes, e.g. general Ornstein-Uhlenbeck processes. Finally, it would be interesting to try to apply the main theorem to the Riemann-Liouville process Example 2.12.

A Equivalence with [3] for the expected signature of fBm at level 4

In [3, Theorem 34] the authors check that the explicit integral expressions for $\mathbb{E}\mathbb{S}(X)_{01}^{(n)}$ with $n = 2, 4$, previously derived for X a fractional Brownian motion of Hurst parameter $H \in (1/2, 1)$, continue to be valid for $H \in (1/4, 1)$. This is done by performing transformations on $\mathbb{E}\mathbb{S}(X^\ell)_{01}$ before passing to the limit. This calculation is specific to levels 2 and 4 and to fBm, and for this reason the expression for the expected signature is not immediately comparable to that obtained as a special case of Theorem 2.3. We devote this appendix to checking that the two agree.

Level 2 is easy to check, since (31) reduces to $\delta^{\alpha\beta}/2$. Referring to Example 2.7, we consider level 4 using (39): starting with the first integral above, we have

$$\int_{0<u<v<1} \left[\tfrac{1}{2}R(\mathrm{d}u) - R(s, \mathrm{d}u)\right]\left[\tfrac{1}{2}R(\mathrm{d}v) - R(u, \mathrm{d}v)\right]\mathrm{d}u\mathrm{d}v$$

$$= H^2 \int_{0<u<v<1} u^{2H-1}(v-u)^{2H-1}\mathrm{d}u\mathrm{d}v$$

$$= H^2 \int_0^1 u^{2H-1}\left[\frac{(v-u)^{2H-1}}{2H}\right]_{u=0}^1 \mathrm{d}u$$

$$= \frac{H^2}{2}\int_0^1 u^{2H-1}\left[\frac{(v-u)^{2H}}{2H}\right]_{v=u}^1 \mathrm{d}u$$

27

$$= \frac{H}{2} \int_0^1 u^{2H-1}(1-u)^{2H} \mathrm{d}u$$

$$= \frac{H}{4} \int_0^1 u^{2H-1}(1-u)^{2H-1} \mathrm{d}u$$

where the last identity can be verified by showing that the difference of the two integrands is odd about the point $u = 1/2$, which in turn is seen by observing that

$$\frac{H}{4} u^{2H-1}(1-u)^{2H-1} - \frac{H}{2} u^{2H-1}(1-u)^{2H} + \frac{H}{4}(1-u)^{2H-1}u^{2H-1} - \frac{H}{2}(1-u)^{2H-1}u^{2H}$$

has zero derivative and vanishes at $u = 1/2$. This shows equality with [3, coefficient of the first term of Γ_H^2 in Corollary 33]. We proceed with the second integral in Example 2.7:

$$\int_{s<u<v<w<t} R(\mathrm{d}u, \mathrm{d}w)\left[\tfrac{1}{2}R(\mathrm{d}v) - R(u, \mathrm{d}v)\right] \mathrm{d}u \mathrm{d}v \mathrm{d}w$$

$$= H^2(2H-1) \int_{0<u<v<w<1} (w-u)^{2H-2}(v-u)^{2H-1} \mathrm{d}u \mathrm{d}v \mathrm{d}w$$

$$= H^2 \int_{0<u<v<1} \left[(1-u)^{2H-1}(v-u)^{2H-1} - (v-u)^{4H-2}\right] \mathrm{d}u \mathrm{d}v$$

$$= \left(\frac{H}{2} - \frac{H^2}{4H-1}\right) \int_0^1 (1-u)^{4H-1} \mathrm{d}u$$

$$= \frac{2H-1}{8(4H-1)}.$$

For the third integral we have

$$\int_{0<u<v<w<z<1} R(\mathrm{d}u, \mathrm{d}w)R(\mathrm{d}v, \mathrm{d}z)$$

$$= H^2(2H-1)^2 \int_{0<u<v<w<z<1} (w-u)^{2H-2}(z-v)^{2H-2} \mathrm{d}u \mathrm{d}v \mathrm{d}w \mathrm{d}z$$

$$= H^2(2H-1) \int_{0<u<v<z<1} \left[(z-u)^{2H-1}(z-v)^{2H-2} - (v-u)^{2H-1}(z-v)^{2H-2}\right] \mathrm{d}u \mathrm{d}v \mathrm{d}z$$

$$= \frac{H(2H-1)}{2} \int_{0<v<z<1} \left[z^{2H}(z-v)^{2H-2} - (z-v)^{4H-2} - v^{2H}(z-v)^{2H-2}\right] \mathrm{d}v \mathrm{d}z$$

$$= \frac{H(2H-1)}{2} \int_{0<v<z<1} (z^{2H} - v^{2H})(z-v)^{2H-2} \mathrm{d}v \mathrm{d}z - \frac{H(2H-1)}{4H-1} \int_0^1 (1-v)^{4H-1} \mathrm{d}v$$

$$= \frac{H(2H-1)}{2} \int_{0<v<z<1} (z^{2H} - v^{2H})(z-v)^{2H-2} \mathrm{d}v \mathrm{d}z - \frac{2H-1}{4(4H-1)}$$

$$= \frac{H}{2} \int_0^1 (1-v^{2H})(1-v)^{2H-1} \mathrm{d}v - H^2 \int_{0<v<z<1} z^{2H-1}(z-v)^{2H-1} - \frac{2H-1}{4(4H-1)}$$

$$= \frac{H}{4(4H-1)} - \frac{H}{4} \int_0^1 v^{2H-1}(1-v)^{2H-1} \mathrm{d}v.$$

In the integration by parts we have used that $\lim_{z \to v^+} (z^{2H} - v^{2H})(z-v)^{2H-1} = 0$ which can be shown by using that for $1/4 < H < 1/2$

$$0 \le (z^{2H} - v^{2H})(z-v)^{2H-1} \le (z-v)^{4H-1} \xrightarrow{z \to v^+} 0$$

since $z^{2H} - v^{2H} < (z-v)^{2H}$ for $0 < v < z$ and $H < 1/2$. In the last identity we have used a similar symmetry argument as the one used in the first calculation, solved trivial integrals and rearranged terms. Note how this calculation would have been simpler if $H \ge 1/2$ since it would not have been necessary to integrate by parts to avoid integrating $(z-v)^{2H-2}$ (cf. [3, Lemma 32]).

Data availability statement

The research presented in this manuscript is theoretical in nature and does not refer to empirical data. All necessary information is contained in the cited references.

Author contribution statement

Thomas Cass and Emilio Ferrucci are both authors of this paper. Emilio Ferrucci is the corresponding author.

Conflict of interest statement

The authors declare no conflicts of interest.

References

[1] J. Armstrong, D. Brigo, and E. Rossi Ferrucci. Optimal approximation of SDEs on submanifolds: the Itô-vector and Itô-jet projections. *Proc. Lond. Math. Soc. (3)*, 119(1):176–213, 2019.

[2] F. Baudoin. *An Introduction to the Geometry of Stochastic Flows*. Published by Imperial College Press and distributed by World Scientific Publishing Co., 2004.

[3] F. Baudoin and L. Coutin. Operators associated with a stochastic differential equation driven by fractional Brownian motions. *Stochastic Processes and their Applications*, 117(5):550 – 574, 2007.

[4] H. Boedihardjo. *Signatures of Gaussian processes and SLE curves*. PhD thesis, University of Oxford, 2014.

[5] H. Boedihardjo, J. Diehl, M. Mezzarobba, and H. Ni. The expected signature of brownian motion stopped on the boundary of a circle has finite radius of convergence. *Bulletin of the London Mathematical Society*, 53(1):285–299, 2021.

[6] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. The signature of a rough path: Uniqueness. *Advances in mathematics (New York. 1965)*, 293:720–737, 2016.

[7] H. Boedihardjo, A. Papavasiliou, and Z. Qian. Expected signature of gaussian processes with strictly regular kernels. arXiv:1304.4930, 2013. https://arxiv.org/abs/1304.4930.

[8] P. Bonnier and H. Oberhauser. Signature cumulants, ordered partitions, and independence of stochastic processes. 2020.

[9] T. Cass and P. Friz. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)*, 171(3):2115–2141, 2010.

[10] T. Cass and N. Lim. A Stratonovich-Skorohod integral formula for Gaussian rough paths. *Ann. Probab.*, 47(1):1–60, 2019.

[11] T. Cass and N. Lim. Skorohod and rough integration for stochastic differential equations driven by Volterra processes. *L'Institut Henri Poincare, Annales B: Probabilites et Statistiques*, 2021.

[12] T. Cass, C. Litterer, and T. Lyons. Integrability and tail estimates for gaussian rough differential equations. *The Annals of probability*, 41(4):3026–3050, 2013.

[13] I. Chevyrev and T. Lyons. Characteristic functions of measures on geometric rough paths. *The Annals of probability*, 44(6):4049–4082, 2016.

[14] I. Chevyrev and H. Oberhauser. Signature moments to characterize laws of stochastic processes. 2022.

[15] L. Coutin and Z. Qian. Stochastic analysis, rough path analysis and fractional brownian motions. *Probability theory and related fields*, 122(1):108–140, 2002.

[16] T. Fawcett. *Problems in stochastic analysis: connections between rough paths and non-commutative harmonic analysis*. PhD thesis, University of Oxford, 2003.

[17] E. R. Ferrucci. *Rough path perspectives on the Itô-Stratonovich dilemma*. PhD thesis, Imperial College London, 2022. https://spiral.imperial.ac.uk/handle/10044/1/96036.

[18] P. Friz and S. Riedel. Convergence rates for the full Gaussian rough paths. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(1):154 – 194, 2014.

[19] P. Friz and N. Victoir. Differential equations driven by Gaussian signals. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 46(2):369 – 413, 2010.

[20] P. K. Friz and A. Shekhar. General rough integration, Lévy rough paths and a Lévy–Kintchine-type formula. *The Annals of Probability*, 45(4):2707 – 2765, 2017.

[21] P. K. Friz and N. B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[22] B. Hambly and T. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of mathematics*, 171(1):109–167, 2010.

[23] S. T. Huang and S. Cambanis. Stochastic and multiple wiener integrals for gaussian processes. *The Annals of probability*, 6(4):585–614, 1978.

[24] M. Jolis. On the Wiener integral with respect to the fractional Brownian motion on an interval. *J. Math. Anal. Appl.*, 330(2):1115–1127, 2007.

[25] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[26] P. Lévy. *Random Functions: General Theory with Special Reference to Laplacian Random Functions*. University of California publications in statistics. University of California Press, 1953.

[27] S. Li and H. Ni. Expected signature of stopped brownian motion on d-dimensional $C^{2,\alpha}$-domains has finite radius of convergence everywhere: $2 \leq d \leq 8$. *Journal of Functional Analysis*, 282(12):109447, 2022.

[28] T. Lyons, M. Caruana, and T. Lévy. Differential equations driven by rough paths ecole d'eté de probabilités de saint-flour xxxiv-2004, 2007.

[29] T. Lyons and H. Ni. Expected signature of Brownian motion up to the first exit time from a bounded domain. *The Annals of Probability*, 43(5):2729 – 2762, 2015.

[30] T. Lyons and Z. Qian. *System control and rough paths*. Oxford mathematical monographs. Clarendon Press, Oxford, 2002.

[31] T. Lyons and N. Victoir. Cubature on Wiener space. volume 460, pages 169–198. 2004. Stochastic analysis with applications to mathematical finance.

[32] B. B. Mandelbrot and J. W. V. Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[33] D. Marinucci and P. M. Robinson. Alternative forms of fractional brownian motion. *Journal of Statistical Planning and Inference*, 80(1):111 – 122, 1999.

[34] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus: from Stein's method to universality*. Cambridge tracts in mathematics; 192. Cambridge University Press, Cambridge [England]; New York, 2012.

[35] D. Nualart. *The Malliavin Calculus and Related Topics*. Probability and Its Applications. 2nd ed. edition, 2006.

[36] D. Nualart and S. Tindel. A construction of the rough path above fractional Brownian motion using Volterra's representation. *The Annals of Probability*, 39(3):1061 – 1096, 2011.

[37] R. Passeggeri. On the signature and cubature of the fractional brownian motion for h>12. *Stochastic Processes and their Applications*, 130(3):1226–1257, 2020.