Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2014, Vol. 50, No. 1, 154–194
DOI: 10.1214/12-AIHP507

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Convergence rates for the full Gaussian rough paths

Peter Friz[a,b,1] and Sebastian Riedel[c,2]

[a] *Fakultät II, Institut für Mathematik, TU Berlin, MA 7-2, Strasse des 17. Juni 136, 10623 Berlin, Germany*
[b] *Weierstrass Institut für Angewandte Analysis und Stochastik, Mohrenstrasse 39, 10117 Berlin, Germany. E-mail: friz@math.tu-berlin.de*
[c] *Fakultät II, Institut für Mathematik, TU Berlin, MA 7-4, Straße des 17. Juni 136, 10623 Berlin, Germany. E-mail: riedel@math.tu-berlin.de*

Received 5 August 2011; revised 23 April 2012; accepted 22 June 2012

Abstract. Under the key assumption of finite ρ-variation, $\rho \in [1, 2)$, of the covariance of the underlying Gaussian process, sharp a.s. convergence rates for approximations of Gaussian rough paths are established. When applied to Brownian resp. fractional Brownian motion (fBM), $\rho = 1$ resp. $\rho = 1/(2H)$, we recover and extend the respective results of (*Trans. Amer. Math. Soc.* **361** (2009) 2689–2718) and (*Ann. Inst. Henri Poincasé Probab. Stat.* **48** (2012) 518–550). In particular, we establish an a.s. rate $k^{-(1/\rho-1/2-\varepsilon)}$, any $\varepsilon > 0$, for Wong–Zakai and Milstein-type approximations with mesh-size $1/k$. When applied to fBM this answers a conjecture in the afore-mentioned references.

Résumé. Nous établissons des vitesses fines de convergence presque sûre pour les approximations des chemins rugueux Gaussiens, sous l'hypothèse que la fonction de covariance du processus Gaussien sous-jacent ait une ρ-variation finie, $\rho \in [1, 2)$. Dans le cas du mouvement Brownien, respectivement du Brownien fractionnaire (fBM), pour lesquels $\rho = 1$ resp. $\rho = 1/(2H)$, ce résultat généralise les résultats respectifs de (*Trans. Amer. Math. Soc.* **361** (2009) 2689–2718) et (*Ann. Inst. Henri Poincasé Probab. Stat.* **48** (2012) 518–550).

Notamment, nous établissons le taux de convergence presque sure $k^{-(1/\rho-1/2-\varepsilon)}$, tout $\varepsilon > 0$, pour les approximations de Wong–Zakai et de type Milstein avec pas de discrétisation $1/k$. Dans le cas du fBM, ce résultat résout une conjecture posée par les références ci-dessus.

MSC: 60H35; 60H10; 60G15; 65C30

Keywords: Gaussian processes; Rough paths; Numerical schemes; Rates of convergence

1. Introduction

Recall that *rough path theory* [7,14,15] is a general framework that allows to establish existence, uniqueness and stability of differential equations driven by multi-dimensional continuous signals $x : [0, T] \to \mathbb{R}^d$ of low regularity. Formally, a *rough differential equation (RDE)* is of the form

$$dy_t = \sum_{i=1}^{d} V_i(y_t)\, dx_t^i \equiv V(y_t)\, dx_t; \qquad y_0 \in \mathbb{R}^e, \tag{1.1}$$

[1] Supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement nr. 258237.
[2] Supported by an IRTG (Berlin–Zurich) PhD-scholarship and a scholarship from the Berlin Mathematical School (BMS).

where $(V_i)_{i=1,\dots,d}$ is a family of vector fields in $\mathbb{R}^e$. When x has finite p-variation, $p<2$, such differential equations can be handled by Young integration theory. Of course, this point of view does not allow to handle differential equations driven by Brownian motion, indeed

$$\sup_{D\subset[0,T]}\sum_{t_i\in D}|B_{t_{i+1}}-B_{t_i}|^2=+\infty \quad \text{a.s.},$$

leave alone differential equations driven by stochastic processes with less sample path regularity than Brownian motion (such as fractional Brownian motion (fBM) with Hurst parameter $H<1/2$). Lyons' key insight was that low regularity of x, say p-variation or $1/p$-Hölder for some $p\in[1,\infty)$, can be compensated by including "enough" higher order information of x such as all increments

$$\mathbf{x}^n_{s,t} \equiv \int_{s<t_1<\dots<t_n<t} \mathrm{d}x_{t_1}\otimes\cdots\otimes \mathrm{d}x_{t_n} \tag{1.2}$$

$$\equiv \sum_{1\le i_1,\dots,i_n\le d}\left(\int_{s<t_1<\cdots<t_n<t}\mathrm{d}x^{i_1}_{t_1}\cdots \mathrm{d}x^{i_n}_{t_n}\right)e_{i_1}\otimes\cdots\otimes e_{i_n}\in\left(\mathbb{R}^d\right)^{\otimes n}, \tag{1.3}$$

where "enough" means $n\le[p]$ ($\{e_1,\dots,e_d\}$ denotes just the usual Euclidean basis in $\mathbb{R}^d$ here). Subject to some generalized p-variation (or $1/p$-Hölder) regularity, the ensemble $(\mathbf{x}^1,\dots,\mathbf{x}^{[p]})$ then constitutes what is known as a rough path.[3] In particular, no higher order information is necessary in the Young case; whereas the regime relevant for Brownian motion requires second order – or level 2 – information ("Lévy's area"), and so on. Note that the iterated integral on the r.h.s. of (1.2) is not – in general – a well-defined Riemann–Stieltjes integral. Instead one typically proceeds by mollification – given a multi-dimensional sample path $x=X(\omega)$, consider piecewise linear approximations or convolution with a smooth kernel, compute the iterated integrals and then pass, if possible, to a limit in probability. Following this strategy one can often construct a "canonical" enhancement of some stochastic process to a (random) rough path. Stochastic integration and differential equations are then discussed in a (rough) pathwise fashion; even in the complete absence of a semi-martingale structure.

It should be emphasized that rough path theory was – from the very beginning – closely related to higher order Euler schemes. Let $D=\{0=t_0<\cdots<t_{\#D-1}=1\}$ be a partition of the unit interval.[4] Considering the solution y of (1.1), the step-N Euler approximation $y^{\mathrm{Euler}^N;D}$ is given by

$$\begin{aligned}
y_0^{\mathrm{Euler}^N;D} &= y_0,\\
y_{t_{j+1}}^{\mathrm{Euler}^N;D} &= y_{t_j}^{\mathrm{Euler}^N;D} + V_i\big(y_{t_j}^{\mathrm{Euler}^N;D}\big)\mathbf{x}^i_{t_j,t_{j+1}} + \mathcal{V}_{i_1}V_{i_2}\big(y_{t_j}^{\mathrm{Euler}^N;D}\big)\mathbf{x}^{i_1,i_2}_{t_j,t_{j+1}}\\
&\quad + \cdots + \mathcal{V}_{i_1}\cdots\mathcal{V}_{i_{N-1}}V_{i_N}\big(y_{t_j}^{\mathrm{Euler}^N;D}\big)\mathbf{x}^{i_1,\dots,i_N}_{t_j,t_{j+1}}
\end{aligned}$$

at the points $t_j\in D$ where we use the Einstein summation convention, $\mathcal{V}_i$ stands for the differential operator $\sum_{k=1}^e V_i^k\partial_{x_k}$ and $\mathbf{x}^{i_1,\dots,i_n}_{s,t}=\int_{s<t_1<\cdots<t_n<t}\mathrm{d}x^{i_1}_{t_1}\cdots\mathrm{d}x^{i_n}_{t_n}$. An extension of the work of A. M. Davie (cf. [3,7]) shows that the step-N Euler scheme[5] for an RDE driven by a $1/p$-Hölder rough path with step size $1/k$ (i.e. $D=D_k=\{\frac{j}{k}\colon j=0,\dots,k\}$) and $N\ge[p]$ will converge with rate $\mathrm{O}(\frac{1}{k})^{(N+1)/p-1}$. Of course, in a probabilistic context, simulation of the iterated (stochastic) integrals $\mathbf{x}^n_{t_j,t_{j+1}}$ is not an easy matter. A natural simplification of the step-N Euler scheme thus amounts to replace in each step

$$\left\{\mathbf{x}^n_{t_j,t_{j+1}}\colon n\in\{1,\dots,N\}\right\}\leftrightarrow\left\{\frac{1}{n!}\left(\mathbf{x}^1_{t_j,t_{j+1}}\right)^{\otimes n}\colon n\in\{1,\dots,N\}\right\}$$

---

[3] A basic theorem of rough path theory asserts that further iterated integrals up to any level $N\ge[p]$, i.e.

$$S_N(\mathbf{x}):=\left(\mathbf{x}^n\colon n\in\{1,\dots,N\}\right)$$

are then deterministically determined and the map $\mathbf{x}\mapsto S_N(\mathbf{x})$, known as Lyons lift, is continuous in rough path metrics.

[4] A general time horizon $[0,T]$ is handled by trivial reparametrization of time.

[5] ... which one would call Milstein scheme when $N=2$...

which leads to the *simplified* step-N Euler scheme

$$y_0^{\mathrm{sEuler}^N;D} = y_0,$$

$$y_{t_{j+1}}^{\mathrm{sEuler}^N;D} = y_{t_j}^{\mathrm{sEuler}^N;D} + V_i\big(y_{t_j}^{\mathrm{sEuler}^N;D}\big)\mathbf{x}_{t_j,t_{j+1}}^i + \frac{1}{2}\mathcal{V}_{i_1}V_{i_2}\big(y_{t_j}^{\mathrm{sEuler}^N;D}\big)\mathbf{x}_{t_j,t_{j+1}}^{i_1}\mathbf{x}_{t_j,t_{j+1}}^{i_2}$$
$$+ \cdots + \frac{1}{N!}\mathcal{V}_{i_1}\cdots\mathcal{V}_{i_{N-1}}V_{i_N}\big(y_{t_j}^{\mathrm{sEuler}^N;D}\big)\mathbf{x}_{t_j,t_{j+1}}^{i_1}\cdots\mathbf{x}_{t_j,t_{j+1}}^{i_N}.$$

Since $\mathbf{x}_{t_j,t_{j+1}}^1 = X_{t_j,t_{j+1}}(\omega) = X_{t_{j+1}}(\omega) - X_{t_j}(\omega)$ this is precisely the effect in replacing the underlying sample path segment of X by its piecewise linear approximation, i.e.

$$\big\{X_t(\omega)\colon t\in[t_j,t_{j+1}]\big\} \leftrightarrow \left\{X_{t_j}(\omega) + \frac{t-t_j}{t_{j+1}-t_j}X_{t_j,t_{j+1}}(\omega)\colon t\in[t_j,t_{j+1}]\right\}.$$

Therefore, as pointed out in [4] in the level $N=2$ Hölder rough path context, it is immediate that a Wong–Zakai type result, i.e. a.s. convergence of $y^{(k)}\to y$ for $k\to\infty$ where $y^{(k)}$ solves

$$\mathrm{d}y_t^{(k)} = V\big(y_t^{(k)}\big)\,\mathrm{d}x_t^{(k)};\quad y_0^{(k)} = y_0\in\mathbb{R}^e$$

and $x^{(k)}$ is the piecewise linear approximation of x at the points $(t_j)_{j=0}^k = D_k$, i.e.

$$x_t^{(k)} = x_{t_j} + \frac{t-t_j}{t_{j+1}-t_j}x_{t_j,t_{j+1}}\quad \text{if } t\in[t_j,t_{j+1}],\, t_j\in D_k,$$

leads to the convergence of the simplified (and implementable!) step-N Euler scheme.

While Wong–Zakai type results in rough path metrics are available for large classes of stochastic processes [7], Chapters 13, 14, 15, 16 our focus here is on *Gaussian* processes which can be enhanced to rough paths. This problem was first discussed in [2] where it was shown in particular that piecewise linear approximation to fBM are convergent in p-variation rough path metric if and only if $H>1/4$. A practical (and essentially sharp) structural condition for the covariance, namely finite ρ-variation based on rectangular increments for some $\rho<2$ of the underlying Gaussian process was given in [6] and allowed for a unified and detailed analysis of the resulting class of Gaussian rough paths. This framework has since proven useful in a variety of different applications ranging from non-Markovian Hörmander theory [1] to non-linear PDEs perturbed by space–time white-noise [10]. Of course, fractional Brownian motion can also be handled in this framework (for $H>1/4$) and we shall make no attempt to survey its numerous applications in engineering, finance and other fields.

Before describing our main result, let us recall in more detail some aspects of Gaussian rough path theory (e.g. [6], [7], Chapter 15, [8]). The basic object is a centred, continuous Gaussian process with sample paths $X(\omega) = (X^1(\omega),\dots,X^d(\omega)):[0,1]\to\mathbb{R}^d$ where X^i and X^j are independent for $i\neq j$. The law of this process is determined by $R_X:[0,1]^2\to\mathbb{R}^{d\times d}$, the covariance function, given by

$$R_X(s,t) = \operatorname{diag}\big(E\big(X_s^1X_t^1\big),\dots,E\big(X_s^dX_t^d\big)\big).$$

We need:

Definition 1. *Let $f=f(s,t)$ be a function from $[0,1]^2$ into a normed space; for $s\le t$, $u\le v$ we define rectangular increments as*

$$f\begin{pmatrix} s,t\\ u,v\end{pmatrix} = f(t,v) - f(t,u) - f(s,v) + f(s,u).$$

For $\rho \geq 1$ we then set

$$V_\rho\big(f, [s,t]\times[u,v]\big) = \left(\sup_{\substack{D\subset[s,t]\\ \tilde{D}\subset[u,v]}} \sum_{\substack{t_i\in D\\ \tilde{t}_j\in\tilde{D}}} \left| f\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1}\end{pmatrix}\right|^\rho \right)^{1/\rho},$$

where the supremum is taken over all partitions D and $\tilde{D}$ of the intervals $[s,t]$ resp. $[u,v]$. If $V_\rho(f,[0,1]^2) < \infty$ we say that f has finite $(2D)$ ρ-variation.

The main result in this context (see e.g. [7], Theorem 15.33, [8]) now asserts that if there exists $\rho < 2$ such that $V_\rho(R_X,[0,1]^2) < \infty$ then X lifts to an *enhanced Gaussian process* $\mathbf{X}$ with sample paths in the p-variation rough path space $C^{0,p\text{-}var}([0,1], G^{[p]}(\mathbb{R}^d))$, any $p \in (2\rho, 4)$. (This and other notations are introduced in Section 2.) This lift is "natural" in the sense that for a large class of smooth approximations $X^{(k)}$ of X (say piecewise linear, mollifier, Karhunen–Loeve) the corresponding iterated integrals of $X^{(k)}$ converge (in probability) to $\mathbf{X}$ with respect to the p-variation rough path metric. (We recall from [7] that $\rho_{p\text{-var}}$, the so-called inhomogeneous p-variation metric for $G^N(\mathbb{R}^d)$-valued paths, is called p-variation rough path metric when $[p] = N$; the Itō–Lyons map enjoys local Lipschitz regularity in this p-variation rough path metric.) Moreover, this condition is sharp; indeed fBM falls into this framework with $\rho = 1/(2H)$ and we known that piecewise-linear approximations to Lévy's area diverge when $H = 1/4$.

Our main result (cf. Theorem 5), when applied to (mesh-size $1/k$) piecewise linear approximations $X^{(k)}$ of X, reads as follows.

Theorem 1. *Let $X = (X^1,\dots,X^d) : [0,1] \to \mathbb{R}^d$ be a centred Gaussian process on a probability space $(\Omega, \mathcal{F}, P)$ with continuous sample paths where X^i and X^j are independent for $i \neq j$. Assume that the covariance R_X has finite ρ-variation for $\rho \in [1,2)$ and $K \geq V_\rho(R_X,[0,1]^2)$. Then there is an enhanced Gaussian process $\mathbf{X}$ with sample paths a.s. in $C^{0,p\text{-}var}([0,1], G^{[p]}(\mathbb{R}^d))$ for any $p \in (2\rho, 4)$ and*

$$\left|\rho_{p\text{-}var}\big(S_{[p]}\big(X^{(k)}\big), \mathbf{X}\big)\right|_{L^r} \to 0$$

for $k \to \infty$ and every $r \geq 1$ ($|\cdot|_{L^r}$ denotes just the usual $L^r(P)$-norm for real valued random variables here). Moreover, for any $\gamma > \rho$ such that $\frac{1}{\gamma} + \frac{1}{\rho} > 1$ and any $q > 2\gamma$ and $N \in \mathbb{N}$ there is a constant $C = C(q,\rho,\gamma,K,N)$ such that

$$\left|\rho_{q\text{-}var}\big(S_N\big(X^{(k)}\big), S_N(\mathbf{X})\big)\right|_{L^r} \leq C r^{N/2} \sup_{0\leq t\leq 1} \left|X_t^{(k)} - X_t\right|_{L^2}^{1-\rho/\gamma}$$

holds for every $k \in \mathbb{N}$.

As an immediate consequence we obtain (essentially) sharp a.s. convergence rates for Wong–Zakai approximations and the simplified step-3 Euler scheme.

Corollary 1. *Consider a RDE with C^∞-bounded vector fields driven by a Gaussian Hölder rough path $\mathbf{X}$. Then mesh-size $1/k$ Wong–Zakai approximations (i.e. solutions of ODEs driven by $X^{(k)}$) converge uniformly with a.s. rate $k^{-(1/\rho-1/2-\varepsilon)}$, any $\varepsilon > 0$, to the RDE solution. The same rate is valid for the simplified (and implementable) step-3 Euler scheme.*

Proof. See Corollary 8 and Corollary 9. □

Several remarks are in order.

- Rough path analysis usually dictates that $N = 2$ (resp. $N = 3$) levels need to be considered when $\rho \in [1, 3/2)$ resp. $\rho \in [3/2, 2)$. Interestingly, the situation for the Wong–Zakai error is quite different here – referring to Theorem 1, when $\rho = 1$ we can and will take γ arbitrarily large in order to obtain the optimal convergence rate. Since $\rho_{q\text{-var}}$ is a rough path metric only in the case $N = [q] \geq [2\gamma]$, we see that we need to consider all levels N which is what Theorem 1 allows us to do. On the other hand, as ρ approaches 2, there is not so much room left for taking $\gamma > \rho$. Even so, we can always find γ with $[\gamma] = 2$ such that $1/\gamma + 1/\rho > 1$. Picking $q > 2\gamma$ small enough shows that we need $N = [q] = 4$.
- The assumption of C^∞-bounded vector fields in the corollary was for simplicity only. In the proof we employ local Lipschitz continuity of the Itō–Lyons map for q-variation rough paths (involving $N = [q]$ levels). As is well-known, this requires $\mathrm{Lip}^{q+\varepsilon}$-regularity of the vector fields.[6] Curiously again, we need C^∞-bounded vector fields when $\rho = 1$ but only $\mathrm{Lip}^{4+\varepsilon}$ as ρ approaches the critical value 2.
- Brownian motion falls in this framework with $\rho = 1$. While the a.s. (Wong–Zakai) rate $k^{-(1/2-\varepsilon)}$ is part of the folklore of the subject (e.g. [9]) the C^∞-boundedness assumption appears unnecessarily strong. Our explanation here is that our rates are *universal* (i.e. valid away from one universal null-set, not dependent on starting points, coefficients etc). In particular, the (Wong–Zakai) rates are valid on the level of stochastic flows of diffeomorphisms; we previously discussed these issues in the Brownian context in [5].
- A surprising aspect appears in the proof of Theorem 1. The strategy is to give sharp estimates for the levels $n = 1, \dots, 4$ first, then performing an induction similar to the one used in Lyon's extension theorem [14] for the higher levels. This is in contrast to the usual considerations of level 1 to 3 only (without level 4!) which is typical for Gaussian rough paths. (Recall that we deal with Gaussian processes which have sample paths of finite p-variation, $p \in (2\rho, 4)$, hence $[p] \leq 3$ which indicates that we would need to control the first 3 levels only before using the extension theorem.)
- Although Theorem 1 was stated here for (step-size $1/k$) piecewise linear approximations $\{X^{(k)}\}$, the estimate holds in great generality for (Gaussian) approximations whose covariance satisfies a uniform ρ-variation bound. The statements of Theorem 5 and Theorem 6 reflect this generality.
- Wong–Zakai rates for the Brownian rough path (level 2) were first discussed in [12]. They prove that Wong–Zakai approximations converge (in γ-Hölder metric) with rate $k^{-(1/2-\gamma-\varepsilon)}$ (in fact, a logarithmic sharpening thereof without ε) provided $\gamma \in (1/3, 1/2)$. This restriction on γ is serious (for they fully rely on "level 2" rough path theory); in particular, the best "uniform" Wong–Zakai convergence rate implied is $k^{-(1/2-1/3-\varepsilon)} = k^{-(1/6-\varepsilon)}$ leaving a significant gap to the well-known Brownian a.s. Wong–Zakai rate.
- Wong–Zakai (and Milstein) rates for the fractional Brownian rough path (level 2 only, Hurst parameter $H > 1/3$) were first discussed in [4]. They prove that Wong–Zakai approximations converge (in γ-Hölder metric) with rate $k^{-(H-\gamma-\varepsilon)}$ (again, in fact, a logarithmic sharpening thereof without ε) provided $\gamma \in (1/3, H)$. Again, the restriction on γ is serious and the best "uniform" Wong–Zakai convergence rate – and the resulting rate for the Milstein scheme – is $k^{-(H-1/3-\varepsilon)}$. This should be compared to the rate $k^{-(2H-1/2-\varepsilon)}$ obtained from our corollary. In fact, this rate was conjectured in [4] and is sharp as may be seen from a precise result concerning Levy's stochastic area for fBM, see [16].

The remainder of the article is structured as follows: In Section 2, we repeat the basic notions of (Gaussian) rough paths theory. Section 3 recalls the connection between the shuffle algebra and iterated integrals. In particular, we will use the shuffle structure to see that in order to show the desired estimates, we can concentrate on some iterated integrals which somehow generate all the others. Our main tool for showing L^2 estimates on the lower levels is multidimensional Young integration which we present in Section 4. The main work, namely showing the desired L^2-estimates for the difference of high-order iterated integrals, is done in Section 5. After some preliminary lemmas in Section 5.1, we show the estimates for the lower levels, namely for $n = 1, 2, 3, 4$ in Section 5.2 , then give an induction argument in Section 5.3 for the higher levels $n > 4$. Section 6 contains our main result, namely sharp a.s. convergence rates for a class of Wong–Zakai approximations, including piecewise-linear and mollifier approximations. We further show in Section 6.3 how to use these results in order to obtain sharp convergence rates for the simplified Euler scheme.

---

[6] ... in the sense of E. Stein; cf. [7,15] for instance.

2. Notations and basic definitions

For $N \in \mathbb{N}$ we define

$$T^N(\mathbb{R}^d) = \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d \otimes \mathbb{R}^d) \oplus \cdots \oplus (\mathbb{R}^d)^{\otimes N} = \bigoplus_{n=0}^{N} (\mathbb{R}^d)^{\otimes n}$$

and write $\pi_n : T^N(\mathbb{R}^d) \to (\mathbb{R}^d)^{\otimes n}$ for the projection on the nth Tensor level. It is clear that $T^N(\mathbb{R}^d)$ is a (finite-dimensional) vector space. For elements $g, h \in T^N(\mathbb{R}^d)$, we define $g \otimes h \in T^N(\mathbb{R}^d)$ by

$$\pi_n(g \otimes h) = \sum_{i=0}^{n} \pi_{n-i}(g) \otimes \pi_i(h).$$

One can easily check that $(T^N(\mathbb{R}^d), +, \otimes)$ is an associative algebra with unit element $\mathbf{1} = \exp(0) = 1 + 0 + 0 + \cdots + 0$. We call it the *truncated tensor algebra of level* N. A norm is defined by

$$|g|_{T^N(\mathbb{R}^d)} = \max_{n=0,\ldots,N} \left|\pi_n(g)\right|$$

which turns $T^N(\mathbb{R}^d)$ into a Banach space.

For $s < t$, we define

$$\Delta^n_{s,t} = \left\{(u_1, \ldots, u_n) \in [s,t]^n; u_1 < \cdots < u_n\right\}$$

which is the n-simplex on the square $[s,t]^n$. We will use $\Delta = \Delta^2_{0,1}$ for the 2-simplex over $[0,1]^2$. A continuous map $\mathbf{x} : \Delta \to T^N(\mathbb{R}^d)$ is called *multiplicative functional* if for all $s < u < t$ one has $\mathbf{x}_{s,t} = \mathbf{x}_{s,u} \otimes \mathbf{x}_{u,t}$. For a path $x = (x^1, \ldots, x^d) : [0,1] \to \mathbb{R}^d$ and $s < t$, we will use the notation $x_{s,t} = x_t - x_s$. If x has finite variation, we define its nth iterated integral by

$$\begin{aligned}\mathbf{x}^n_{s,t} &= \int_{\Delta^n_{s,t}} \mathrm{d}x \otimes \cdots \otimes \mathrm{d}x \\ &= \sum_{1 \le i_1, \ldots, i_n \le d} \int_{\Delta^n_{s,t}} \mathrm{d}x^{i_1} \cdots \mathrm{d}x^{i_n} e_{i_1} \otimes \cdots \otimes e_{i_n} \in (\mathbb{R}^d)^{\otimes n},\end{aligned}$$

where $\{e_1, \ldots, e_d\}$ denotes the Euclidean basis in $\mathbb{R}^d$ and $(s,t) \in \Delta$. The canonical lift $S_N(x) : \Delta \to T^N(\mathbb{R}^d)$ is defined by

$$\pi_n(S_N(x)_{s,t}) = \begin{cases} \mathbf{x}^n_{s,t}, & \text{if } n \in \{1, \ldots, N\}, \\ 1, & \text{if } n = 0. \end{cases}$$

It is well know (as a consequence of Chen's theorem) that $S_N(x)$ is a multiplicative functional. Actually, one can show that $S_N(x)$ takes values in the smaller set $G^N(\mathbb{R}^d) \subset T^N(\mathbb{R}^d)$ defined by

$$G^N(\mathbb{R}^d) = \left\{S_N(x)_{0,1} : x \in C^{1\text{-}var}([0,1], \mathbb{R}^d)\right\}$$

which is still a group with $\otimes$. If $\mathbf{x}, \mathbf{y} : \Delta \to T^N(\mathbb{R}^d)$ are multiplicative functionals and $p \ge 1$ we set

$$\rho_{p\text{-}var}(\mathbf{x}, \mathbf{y}) := \max_{n=1,\ldots,N} \sup_{(t_i) \in [0,1]} \left(\sum_i \left|\mathbf{x}^n_{t_i,t_{i+1}} - \mathbf{y}^n_{t_i,t_{i+1}}\right|^{p/n}\right)^{n/p}.$$

This generalizes the p-variation distance induced by the usual p-variation semi-norm

$$|x|_{p\text{-}var;[s,t]} = \left(\sup_{(t_i) \subset [s,t]} \sum_i |x_{t_{i+1}} - x_{t_i}|^p\right)^{1/p}$$

for paths $x:[0,1]\to\mathbb{R}^d$. The Lie group $G^N(\mathbb{R}^d)$ admits a natural norm $\|\cdot\|$, called the *Carnot-Caratheodory norm* (cf. [7], Chapter 7). If $\mathbf{x}:\Delta\to G^N(\mathbb{R}^d)$, we set

$$\|\mathbf{x}\|_{p\text{-}var;[s,t]}=\left(\sup_{(t_i)\subset[s,t]}\sum_i\|\mathbf{x}_{t_i,t_{i+1}}\|^p\right)^{1/p}.$$

Definition 2. *The space $C_o^{0,p\text{-}var}([0,1],G^N(\mathbb{R}^d))$ is defined as the set of continuous paths $\mathbf{x}:\Delta\to G^N(\mathbb{R}^d)$ for which there exists a sequence of smooth paths $x_k:[0,1]\to\mathbb{R}^d$ such that $\rho_{p\text{-}var}(\mathbf{x},S_N(x_k))\to 0$ for $k\to\infty$. If $N=[p]=\max\{n\in\mathbb{N}\colon n<p\}$ we call this the space of (geometric) p-rough paths.*

It is clear by definition that every p-rough path is also a multiplicative functional. By Lyon's First Theorem (or Extension Theorem, see [14], Theorem 2.2.1 or [7], Theorem 9.5) every p-rough path $\mathbf{x}$ has a unique lift to a path in $G^N(\mathbb{R}^d)$ for $N\ge[p]$. We denote this lift by $S_N(\mathbf{x})$ and call it the *Lyons lift*. For a p-rough path $\mathbf{x}$, we will also use the notation

$$\mathbf{x}^n_{s,t}=\pi_n\big(S_N(\mathbf{x})_{s,t}\big)$$

for $N\ge n$. Note that this is consistent with our former definition in the case where x had finite variation. We will always use small letters for paths x and capital letters for stochastic processes X. The same notation introduced here will also be used for stochastic processes.

Definition 3. *A function $\omega:\Delta\to\mathbb{R}^+$ is called a $(1D)$ control if it is continuous and superadditive, i.e. if for all $s<u<t$ one has*

$$\omega(s,u)+\omega(u,t)\le\omega(s,t).$$

If $x:[0,1]\to\mathbb{R}^d$ is a continuous path with finite p-variation, one can show that

$$(s,t)\mapsto V_p\big(x,[s,t]\big)^p:=|x|^p_{p\text{-}var;[s,t]}$$

is continuous and superadditive, hence defines a $1D$-control function. Unfortunately, this is not the case for higher dimensions. Recall Definition 1. If $f:[0,1]^2\to\mathbb{R}$ has finite p-variation,

$$(s,t),(u,v)\mapsto V_p\big(f,[s,t]\times[u,v]\big)^p$$

in general fails to be superadditive (cf. [8]). Therefore, we will need a second definition. If $A=[s,t]\times[u,v]$ is a rectangle in $[0,1]^2$, we will use the notation $f(A):=f\left(\begin{smallmatrix}s & t\\ u & v\end{smallmatrix}\right)$. We call two rectangles *essentially disjoint* if their intersection is empty or degenerate. A partition Π of a rectangle $R\subset[0,1]^2$ is a finite set of essentially disjoint rectangles whose union is R. The family of all such partitions is denoted by $\mathcal{P}(R)$.

Definition 4. *A function $\omega:\Delta\times\Delta\to\mathbb{R}^+$ is called a $(2D)$ control if it is continuous, zero on degenerate rectangles and super-additive in the sense that for all rectangles $R\subset[0,1]^2$,*

$$\sum_{i=1}^n\omega(R_i)\le\omega(R)$$

whenever $\{R_i:i=1,\dots,n\}\in\mathcal{P}(R)$. ω is called symmetric if $\omega([s,t]\times[u,v])=\omega([u,v]\times[s,t])$ holds for all $s<t$ and $u<v$. If $f:[0,1]^2\to B$ is a continuous function, we say that its p-variation is controlled by ω if $|f(R)|^p\le\omega(R)$ holds for all rectangles $R\subset[0,1]^2$.

It is easy to see that if ω is a $2D$ control, $(s,t)\mapsto\omega([s,t]^2)$ defines a $1D$-control.

Definition 5. *For* $f : [0,1]^2 \to \mathbb{R}$, $R \subset [0,1]^2$ *a rectangle and* $p \geq 1$ *we define*

$$|f|_{p\text{-}var;R} := \sup_{\Pi \in \mathcal{P}(R)} \left(\sum_{A \in \Pi} |f(A)|^p \right)^{1/p}.$$

If $|f|_{p\text{-}var;[0,1]^2} < \infty$ *we say that* f *has finite controlled* p*-variation.*

The difference of $2D$ p-variation introduced in Definition 1 and *controlled* p-variation is that in the former, one only takes the supremum over grid-like partitions whereas in the latter, one takes the supremum over all partitions of the rectangle. By superadditivity, the existence of a control ω which controls the p-variation of f implies that f has finite controlled p-variation and $|f|_{p\text{-}var;R} \leq \omega(R)^{1/p}$. In this case, we can always assume w.l.o.g. that ω is symmetric, otherwise we just substitute ω by its symmetrization ω_{sym} given by

$$\omega_{\text{sym}}\big([s,t] \times [u,v]\big) = \omega\big([s,t] \times [u,v]\big) + \omega\big([u,v] \times [s,t]\big).$$

The connection between finite variation and finite controlled p-variation is summarized in the following theorem.

Theorem 2. *Let* $f : [0,1]^2 \to \mathbb{R}$ *be continuous and* $R \subset [0,1]^2$ *be a rectangle.*

(1) *We have*

$$V_1(f,R) = |f|_{1\text{-}var;R}.$$

(2) *For any* $p \geq 1$ *and* $\varepsilon > 0$ *there is a constant* $C = C(p,\varepsilon)$ *such that*

$$\frac{1}{C}|f|_{(p+\varepsilon)\text{-}var;R} \leq V_{p\text{-}var}(f,R) \leq |f|_{p\text{-}var;R}.$$

(3) *If* f *has finite controlled* p*-variation, then*

$$R \mapsto |f|^p_{p\text{-}var;R}$$

is a $2D$*-control. In particular, there exists a* $2D$*-control* ω *such that for all rectangles* $R \subset [0,1]^2$ *we have* $|f(R)|^p \leq \omega(R)$*, i.e.* ω *controls the* p*-variation of* f*.*

Proof. [8], Theorem 1. □

In the following, unless mentioned otherwise, X will always be a Gaussian process as in Theorem 1 and $\mathbf{X}$ denotes the natural Gaussian rough path. We will need the following Proposition:

Proposition 1. *Let* X *be as in Theorem* 1 *and assume that* ω *controls the* ρ*-variation of the covariance of* X, $\rho \in [1,2)$*. Then for every* $n \in \mathbb{N}$ *there is a constant* $C(n) = C(n,\rho)$ *such that*

$$\big|\mathbf{X}^n_{s,t}\big|_{L^2} \leq C(n)\omega\big([s,t]^2\big)^{n/(2\rho)}$$

for any $s < t$.

Proof. For $n = 1,2,3$ this is proven in [7], Proposition 15.28. For $n \geq 4$ and fixed $s < t$, we set $\tilde{X}_\tau := \frac{1}{\omega([s,t]^2)^{1/(2\rho)}} X_{s+\tau(t-s)}$. Then $|R_{\tilde{X}}|^\rho_{\rho\text{-}var;[0,1]} \leq 1 =: K$ and by the standard (deterministic) estimates for the Lyons lift,

$$\frac{|\mathbf{X}^n_{s,t}|^{1/n}}{\omega([s,t]^2)^{1/(2\rho)}} \leq c_1 \big\| S_n(\tilde{\mathbf{X}}) \big\|_{p\text{-}var;[0,1]} \leq c_2(n,p) \|\tilde{\mathbf{X}}\|_{p\text{-}var;[0,1]}$$

for any $p \in (2\rho, 4)$. Now we take the L^2-norm on both sides. From [7], Theorem 15.33, we know that $|\|\tilde{\mathbf{X}}\|_{p\text{-}var;[0,1]}|_{L^2}$ is bounded by a constant only depending on p, ρ and K which shows the claim.

Alternatively (and more in the spirit of the forthcoming arguments), one performs an induction similar (but easier) as in the proof of Proposition 8. □

3. Iterated integrals and the shuffle algebra

Let $x = (x^1, \dots, x^d) : [0,1] \to \mathbb{R}^d$ be a path of finite variation. Forming finite linear combinations of iterated integrals of the form

$$\int_{\Delta_{0,1}^n} \mathrm{d}x^{i_1} \cdots \mathrm{d}x^{i_n}, \quad i_1, \dots, i_n \in \{1, \dots, d\}, n \in \mathbb{N}$$

defines a vector space over $\mathbb{R}$. In this section, we will see that this vector space is also an algebra where the product is given simply by taking the usual multiplication. Moreover, we will describe precisely how the product of two iterated integrals looks like.

3.1. *The shuffle algebra*

Let A be a set which we will call from now on the alphabet. In the following, we will only consider the finite alphabet $A = \{a, b, \dots\} = \{a_1, a_2, \dots, a_d\} = \{1, \dots, d\}$. We denote by A^* the set of words composed by the letters of A, hence $w = a_{i_1} a_{i_2} \cdots a_{i_n}$, $a_{i_j} \in A$. The empty word is denoted by e. A^+ is the set of non-empty words. The length of the word is denoted by $|w|$ and $|w|_a$ denotes the number of occurrences of the letter a. We denote by $\mathbb{R}\langle A\rangle$ the vector space of noncommutative polynomials on A over $\mathbb{R}$, hence every $P \in \mathbb{R}\langle A\rangle$ is a linear combination of words in A^* with coefficients in $\mathbb{R}$. (P, w) denotes the coefficient in P of the word w. Hence every polynomial P can be written as

$$P = \sum_{w \in A^*} (P, w) w$$

and the sum is finite since the (P, w) are non-zero only for a finite set of words w. We define the degree of P as

$$\deg(P) = \max\bigl\{|w|; (P, w) \neq 0\bigr\}.$$

A polynomial is called *homogeneous* if all monomials have the same degree. We want to define a product on $\mathbb{R}\langle A\rangle$. Since a polynomial is determined by its coefficients on each word, we can define the product PQ of P and Q by

$$(PQ, w) = \sum_{w = uv} (P, u)(Q, v).$$

Note that this definition coincides with the usual multiplication in a (noncommutative) polynomial ring. We call this product the *concatenation product* and the algebra $\mathbb{R}\langle A\rangle$ endowed with this product the *concatenation algebra*.

There is another product on $\mathbb{R}\langle A\rangle$ which will be of special interest for us. We need some notation first. Given a word $w = a_{i_1} a_{i_2} \cdots a_{i_n}$ and a subsequence $U = (j_1, j_2, \dots, j_k)$ of $(i_1, \dots, i_n)$, we denote by $w(U)$ the word $a_{j_1} a_{j_2} \cdots a_{j_k}$ and we call $w(U)$ a *subword* of w. If w, u, v are words and if w has length n, we denote by $\binom{w}{u \;\; v}$ the number of subsequences U of $(1, \dots, n)$ such that $w(U) = u$ and $w(U^c) = v$.

Definition 6. *The* (*homogeneous*) *polynomial*

$$u * v = \sum_{w \in A^*} \binom{w}{u \quad v} w$$

is called the shuffle product of u and v. By linearity we extend it to a product on $\mathbb{R}\langle A\rangle$.

In order to proof our main result, we want to use some sort of induction over the length of the words. Therefore, the following definition will be useful.

Definition 7. *If U is a set of words of the same length, we call a subset $\{w_1, \ldots, w_k\}$ of U a generating set for U if for every word $w \in U$ there is a polynomial R and real numbers $\lambda_1, \ldots, \lambda_k$ such that*

$$w = \sum_{j=1}^{k} \lambda_j w_j + R,$$

where R is of the form $R = \sum_{u,v \in A^+} \mu_{u,v} u * v$ for real numbers $\mu_{u,v}$.

Definition 8. *We say that a word w is composed by $a_1^{n_1}, \ldots, a_d^{n_d}$ if $w \in \{a_1, \ldots, a_d\}^*$ and $|w|_{a_i} = n_i$ for $i = 1, \ldots, d$, hence every letter appears in the word with the given multiplicity.*

The aim now is to find a (possibly small) generating set for the set of all words composed by some given letters. The next definition introduces a special class of words which will be important for us.

Definition 9. *Let A be totally ordered and put on A^* the alphabetical order. If w is a word such that whenever $w = uv$ for $u, v \in A^+$ one has $u < v$, then w is called a Lyndon word.*

Proposition 2.

(1) *For the set $\{$words composed by $a, a, b\}$ a generating set is given by $\{aab\}$.*
(2) *For the set $\{$words composed by $a, a, a, b\}$ a generating set is given by $\{aaab\}$.*
(3) *For the set $\{$words composed by $a, a, b, b\}$ a generating set is given by $\{aabb\}$.*
(4) *For the set $\{$words composed by $a, a, b, c\}$ a generating set is given by $\{aabc, aacb, baac\}$.*

Proof. Consider the alphabet $A = \{a, b, c\}$. We choose the order $a < b < c$. A general theorem states that every word w has a unique decreasing factorization into Lyndon words, i.e. $w = l_1^{i_1} \cdots l_k^{i_k}$ where $l_1 > \cdots > l_k$ are Lyndon words and $i_1, \ldots, i_k \geq 1$ (see [17], Theorem 5.1 and Corollary 4.7), and the formula

$$\frac{1}{i_1! \cdots i_k!} l_1^{*i_1} * \cdots * l_k^{*i_k} = w + \sum_{u < w} \alpha_u u$$

holds, where α_u are some natural integers (see again [17], Theorem 6.1). By repeatedly applying this formula for the words in the sum on the right hand side, it follows that a generating set for each of the sets in (1) to (4) is given exactly by the Lyndon words composed by these letters. One can easily show that indeed aab, $aaab$ and $aabb$ are the only Lyndon words composed by the corresponding letters. The Lyndon words composed by a, a, b, c are $\{aabc, abac, aacb\}$ which therefore is a generating set for $\{$words composed by $a, a, b, c\}$. From the shuffle identity

$$abac = baac + aabc + aacb - b * aac$$

it follows that also $\{aabc, aacb, baac\}$ generates this set. □

3.2. *The connection to iterated integrals*

Let $x = (x^1, \ldots, x^d) : [0, 1] \to \mathbb{R}^d$ be a path of finite variation and fix $s < t \in [0, 1]$. For a word $w = (a_{i_1} \cdots a_{i_n}) \in A^*$, $A = \{1, \ldots, d\}$ we define

$$\mathbf{x}^w = \begin{cases} \int_{\Delta_{s,t}^n} \mathrm{d}x^{i_1} \cdots \mathrm{d}x^{i_n}, & \text{if } w \in A^+, \\ 1, & \text{if } w = e. \end{cases}$$

Let $(\mathbb{R}\langle A \rangle, +, *)$ be the shuffle algebra over the alphabet A. We define a map $\Phi : \mathbb{R}\langle A \rangle \to \mathbb{R}$ by $\Phi(w) = \mathbf{x}_{s,t}^w$ and extend it linearly to polynomials $P \in \mathbb{R}\langle A \rangle$. The key observation is the following:

Theorem 3. *Φ is an algebra homomorphism from the shuffle algebra $(\mathbb{R}\langle A\rangle, +, *)$ to $(\mathbb{R}, +, \cdot)$.*

Proof. [17], Corollary 3.5. □

The next proposition shows that we can restrict ourselves in showing the desired estimates only for the iterated integrals which generate the others.

Proposition 3. *Let $(X, Y) = (X^1, Y^1, \dots, X^d, Y^d)$ be a Gaussian process on $[0, 1]$ with paths of finite variation. Let $A = \{1, \dots, d\}$ be the alphabet, let U be a set of words of length n and $V = \{w_1, \dots, w_k\}$ be a generating set for U. Let ω be a control, $\rho, \gamma \geq 1$ constants and $s < t \in [0, 1]$. Assume that there are constants $C = C(|w|)$ such that*

$$\left|\mathbf{X}^w_{s,t}\right|_{L^2} \leq C\big(|w|\big)\omega(s,t)^{|w|/(2\rho)} \quad \text{and} \quad \left|\mathbf{Y}^w_{s,t}\right|_{L^2} \leq C\big(|w|\big)\omega(s,t)^{|w|/(2\rho)}$$

holds for every word $w \in A^*$ with $|w| \leq n-1$. Assume also that for some $\varepsilon > 0$

$$\left|\mathbf{X}^w_{s,t} - \mathbf{Y}^w_{s,t}\right|_{L^2} \leq C\big(|w|\big)\varepsilon\omega(s,t)^{1/(2\gamma)}\omega(s,t)^{(|w|-1)/(2\rho)}$$

holds for every word w with $|w| \leq n-1$ and $w \in V$. Then there is a constant $\tilde{C}$ which depends on the constants C, on n and on d such that

$$\left|\mathbf{X}^w_{s,t} - \mathbf{Y}^w_{s,t}\right|_{L^2} \leq \tilde{C}\varepsilon\omega(s,t)^{1/(2\gamma)}\omega(s,t)^{(n-1)/(2\rho)}$$

holds for every $w \in U$.

Remark 1. *We could account for the factor $\omega(s,t)^{1/(2\gamma)}$ in ε here but the present form is how we shall use this proposition later on.*

Proof. Consider a copy $\bar{A}$ of A. If $a \in A$, we denote by $\bar{a}$ the corresponding letter in $\bar{A}$. If $w = a_{i_1} \cdots a_{i_n} \in A^*$, we define $\bar{w} = \bar{a}_{i_1} \cdots \bar{a}_{i_n} \in A^*$ and in the same way we define $\bar{P} \in \mathbb{R}\langle\bar{A}\rangle$ for $P \in \mathbb{R}\langle A\rangle$. Now we consider $\mathbb{R}\langle A \dot{\cup} \bar{A}\rangle$ equipped with the usual shuffle product. Define $\Psi : \mathbb{R}\langle A \dot{\cup} \bar{A}\rangle \to \mathbb{R}$ by

$$\Psi(w) = \int_{\Delta^n_{s,t}} \mathrm{d}Z^{b_{i_1}} \cdots \mathrm{d}Z^{b_{i_n}}$$

for a word $w = b_{i_1} \cdots b_{i_n}$ where

$$Z^{b_j} = \begin{cases} X^{a_j}, & \text{for } b_j = a_j, \\ Y^{\bar{a}_j}, & \text{for } b_j = \bar{a}_j \end{cases}$$

and extend this definition linearly. By Theorem 3, we know that Ψ is an algebra homomorphism. Take $w \in U$. By assumption, we know that there is a vector $\lambda = (\lambda_1, \dots, \lambda_k)$ such that

$$w - \bar{w} = \sum_{j=1}^{k} \lambda_j (w_j - \bar{w}_j) + R - \bar{R},$$

where R is of the form $R = \sum_{u,v \in A^+, |u|+|v|=n} \mu_{u,v} u * v$ with real numbers $\mu_{u,v}$. Applying Ψ and taking the L^2 norm yields

$$\begin{aligned}\left|\mathbf{X}^w_{s,t} - \mathbf{Y}^w_{s,t}\right|_{L^2} &\leq \sum_{l=1}^{k} |\lambda_j| \left|\mathbf{X}^{w_j}_{s,t} - \mathbf{Y}^{w_j}_{s,t}\right|_{L^2} + \left|\Psi(R - \bar{R})\right|_{L^2} \\ &\leq c_1 \varepsilon\omega(s,t)^{1/(2\gamma)}\omega(s,t)^{(n-1)/(2\rho)} + \left|\Psi(R - \bar{R})\right|_{L^2}.\end{aligned}$$

Now,

$$R - \bar{R} = \sum_{u,v} \mu_{u,v}(u * v - \bar{u} * \bar{v}) = \sum_{u,v} \mu_{u,v}(u - \bar{u}) * v + \mu_{u,v}\bar{u} * (v - \bar{v}).$$

Applying Ψ and taking the L^2 norm gives then

$$\begin{aligned}
\left|\Psi(R - \bar{R})\right|_{L^2} &\le \sum_{u,v} |\mu_{u,v}| \left|\left(\mathbf{X}^u_{s,t} - \mathbf{Y}^u_{s,t}\right)\mathbf{X}^v_{s,t}\right|_{L^2} + |\mu_{u,v}|\left|\mathbf{Y}^u_{s,t}\left(\mathbf{X}^v_{s,t} - \mathbf{Y}^v_{s,t}\right)\right|_{L^2} \\
&\le \sum_{u,v} c_2\left(\left|\mathbf{X}^u_{s,t} - \mathbf{Y}^u_{s,t}\right|_{L^2}\left|\mathbf{X}^v_{s,t}\right|_{L^2} + \left|\mathbf{Y}^u_{s,t}\right|_{L^2}\left|\mathbf{X}^v_{s,t} - \mathbf{Y}^v_{s,t}\right|_{L^2}\right) \\
&\le \sum_{u,v} c_3 \varepsilon \omega(s,t)^{1/(2\gamma)} \omega(s,t)^{(|v|+|u|-1)/(2\rho)} \\
&\le c_4 \varepsilon \omega(s,t)^{1/(2\gamma)} \omega(s,t)^{(n-1)/(2\rho)},
\end{aligned}$$

where we used equivalence of L^q-norms in the Wiener Chaos (cf. [7, Proposition 15.19 and Theorem D.8]). Putting all together shows the assertion. □

4. Multidimensional Young-integration and grid-controls

Let $f : [0,1]^n \to \mathbb{R}$ be a continuous function. If $s_1 < t_1, \dots, s_n < t_n$ and $u_1, \dots, u_n$ are elements in $[0,1]$, we make the following recursive definition:

$$f\begin{pmatrix} s_1, t_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} := f\begin{pmatrix} t_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} - f\begin{pmatrix} s_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \quad \text{and}$$

$$f\begin{pmatrix} s_1, t_1 \\ \vdots \\ s_{k-1}, t_{k-1} \\ s_k, t_k \\ u_{k+1} \\ \vdots \\ u_n \end{pmatrix} := f\begin{pmatrix} s_1, t_1 \\ \vdots \\ s_{k-1}, t_{k-1} \\ t_k \\ u_{k+1} \\ \vdots \\ u_n \end{pmatrix} - f\begin{pmatrix} s_1, t_1 \\ \vdots \\ s_{k-1}, t_{k-1} \\ s_k \\ u_{k+1} \\ \vdots \\ u_n \end{pmatrix}.$$

We will also use the simpler notation

$$f(R) = f\begin{pmatrix} s_1, t_1 \\ \vdots \\ s_n, t_n \end{pmatrix}$$

for the rectangle $R = [s_1, t_1] \times \cdots \times [s_n, t_n] \subset [0,1]^n$. Note that for $n = 2$ this is consistent with our initial definition of $f\binom{s_1,t_1}{s_2,t_2}$. If $f, g : [0,1]^n \to \mathbb{R}$ are continuous functions, the n-dimensional Young-integral is defined by

$$\int_{[s_1,t_1]\times\cdots\times[s_n,t_n]} f(x_1,\dots,x_n)\,\mathrm{d}g(x_1,\dots,x_n) := \lim_{|D_1|,\dots,|D_n|\to 0} \sum_{\substack{(t^1_{i_1})\subset D_1 \\ \vdots \\ (t^n_{i_n})\subset D_n}} f\left(t^1_{i_1},\dots,t^n_{i_n}\right) g\begin{pmatrix} t^1_{i_1}, t^1_{i_1+1} \\ \vdots \\ t^n_{i_n}, t^n_{i_n+1} \end{pmatrix}$$

if this limit exists. Take $p \geq 1$. The n-dimensional p-variation of f is defined by

$$V_p\big(f,[s_1,t_1]\times\cdots\times[s_n,t_n]\big) = \left(\sup_{\substack{D_1\subset[s_1,t_1]\\ \vdots \\ D_n\subset[s_n,t_n]}} \sum_{\substack{(t^1_{i_1})\subset D_1\\ \vdots \\ (t^n_{i_n})\subset D_n}} \left| f\begin{pmatrix} t^1_{i_1}, t^1_{i_1+1} \\ \vdots \\ t^n_{i_n}, t^n_{i_n+1}\end{pmatrix}\right|^p \right)^{1/p}$$

and if $V_p(f,[0,1]^n) < \infty$ we say that f has finite (n-dimensional) p-variation. The fundamental theorem is the following:

Theorem 4. *Assume that f has finite p-variation and g finite q-variation where $\frac{1}{p}+\frac{1}{q}>1$. Then the joint Young-integral below exists and there is a constant $C = C(p,q)$ such that*

$$\left| \int_{[s_1,t_1]\times\cdots\times[s_n,t_n]} f\begin{pmatrix} s_1,u_1 \\ \vdots \\ s_n,u_n\end{pmatrix} \mathrm{d}g(u_1,\ldots,u_n)\right| \\ \leq C V_p\big(f,[s_1,t_1]\times\cdots\times[s_n,t_n]\big) V_q\big(g,[s_1,t_1]\times\cdots\times[s_n,t_n]\big).$$

Proof. [18], Theorem 1.2(c). □

We will mainly consider the case $n=2$, but we will also need $n=3$ and 4 later on. In particular, the discussion of level $n=4$ will require us to work with $4D$ grid control functions which we now introduce. With no extra complication we make the following general definition.

Definition 10 (n-dimensional grid control). *A map $\tilde{\omega}\colon \underbrace{\Delta\times\cdots\times\Delta}_{n\text{-times}} \to \mathbb{R}^+$ is called a n-D grid-control if it is continuous and partially super-additive, i.e. for all $(s_1,t_1),\ldots,(s_n,t_n)\in\Delta$ and $s_i<u_i<t_i$ we have*

$$\tilde{\omega}\big([s_1,t_1]\times\cdots\times[s_i,u_i]\times\cdots\times[s_n,t_n]\big) + \tilde{\omega}\big([s_1,t_1]\times\cdots\times[u_i,t_i]\times\cdots\times[s_n,t_n]\big) \\ \leq \tilde{\omega}\big([s_1,t_1]\times\cdots\times[s_i,t_i]\times\cdots\times[s_n,t_n]\big)$$

for every $i=1,\ldots,n$. $\tilde{\omega}$ is called symmetric if

$$\tilde{\omega}\big([s_1,t_1]\times\cdots\times[s_n,t_n]\big) = \tilde{\omega}\big([s_{\sigma(1)},t_{\sigma(1)}]\times\cdots\times[s_{\sigma(n)},t_{\sigma(n)}]\big)$$

holds for every $\sigma\in S_n$.

The point of this definition is that $|f(A)|^p \leq \tilde{\omega}(A)$ for every rectangle $A\subset[0,1]^n$ implies that $V_p(f,R)^p \leq \tilde{\omega}(R)$ for every rectangle $R\subset[0,1]^n$. Note that a 2D control in the sense of Definition 4 is automatically a $2D$ grid-control. The following immediate properties will be used in Section 5.2.3 with $m=n=2$.

Lemma 1.

(1) *The restriction of a $(m+n)$-dimensional grid-control to m arguments is a m-dimensional grid-control.*
(2) *The product of a m- and a n-dimensional grid-control is a $(m+n)$-dimensional grid-control.*

4.1. *Iterated* $2D$*-integrals*

In the 1-dimensional case, the classical Young-theory allows to define iterated integrals of functions with finite p-variation where $p < 2$. There, the superadditivity of $(s,t) \mapsto |\cdot|^p_{p\text{-}var;[s,t]}$ played an essential role. We will see that Theorem 2 can be used to define and estimate iterated $2D$-integrals. This will play an important role in Section 5 when we estimate the L^2-norm of iterated integrals of Gaussian processes.

Lemma 2. *Let* $f, g : [0,1]^2 \to \mathbb{R}$ *be continuous where* f *has finite* p*-variation and* g *finite controlled* q*-variation with* $p^{-1} + q^{-1} > 1$. *Let* $(s,t) \in \Delta$ *and assume that* $f(s,\cdot) = f(\cdot,s) = 0$. *Define* $\Phi : [s,t]^2 \to \mathbb{R}$ *by*

$$\Phi(u,v) = \int_{[s,u]\times[s,v]} f \, dg.$$

Then there is a constant $C = C(p,q)$ *such that*

$$V_{q\text{-}var}\big(\Phi;[s,t]^2\big) \le C(p,q) V_{p\text{-}var}\big(f;[s,t]^2\big) |g|_{q\text{-}var;[s,t]^2}.$$

Proof. Let $t_i < t_{i+1}$ and $\tilde{t}_j < \tilde{t}_{j+1}$. Then,

$$\Phi\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1} \end{pmatrix} = \int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f \, dg.$$

Now let $t_i < u < t_{i+1}$ and $\tilde{t}_j < v < \tilde{t}_{j+1}$. Then one has

$$f\begin{pmatrix} t_i, u \\ \tilde{t}_j, v \end{pmatrix} = f(u,v) - f(t_i,v) - f(u,\tilde{t}_j) + f(t_i,\tilde{t}_j).$$

Therefore,

$$\begin{aligned}\left|\Phi\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1} \end{pmatrix}\right| \le{}& \left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f\begin{pmatrix} t_i, u \\ \tilde{t}_j, v \end{pmatrix} dg(u,v)\right| + \left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f(t_i,v)\, dg(u,v)\right| \\ &+ \left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f(u,\tilde{t}_j)\, dg(u,v)\right| + \left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f(t_i,\tilde{t}_j)\, dg(u,v)\right|.\end{aligned}$$

For the first integral we use Young $2D$-estimates to see that

$$\begin{aligned}&\left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f\begin{pmatrix} t_i, u \\ \tilde{t}_j, v \end{pmatrix} dg(u,v)\right| \\ &\quad \le c_1(p,q) V_p\big(f,[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]\big) V_q\big(g,[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]\big) \\ &\quad \le c_1(p,q) V_p\big(f,[s,t]^2\big) |g|_{q\text{-}var;[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]}.\end{aligned}$$

For the second, one has by a Young $1D$-estimate

$$\begin{aligned}\left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f(t_i,v)\, dg(u,v)\right| &= \left|\int_{[\tilde{t}_j,\tilde{t}_{j+1}]} f(t_i,v)\, d\big(g(t_{i+1},v) - g(t_i,v)\big)\right| \\ &\le c_2 \sup_{u\in[s,t]} \big|f(u,\cdot)\big|_{p\text{-}var;[s,t]} |g|_{q\text{-}var;[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]}.\end{aligned}$$

Similarly,

$$\left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f(u,\tilde{t}_j)\, dg(u,v)\right| \le c_2 \sup_{v\in[s,t]} \big|f(\cdot,v)\big|_{p\text{-}var;[s,t]} |g|_{q\text{-}var;[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]}.$$

Finally,

$$\left|\int_{[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} f(t_i,\tilde{t}_j)\,\mathrm{d}g(u,v)\right| = \left|f(t_i,\tilde{t}_j)\right|\left|g\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1}\end{pmatrix}\right| \le |f|_{\infty;[s,t]}|g|_{q\text{-}var;[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]}.$$

Putting all together, we get

$$\begin{aligned}&\left|\Phi\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1}\end{pmatrix}\right|^q \\ &\quad \le c_3\Big(V_p\big(f,[s,t]^2\big) + \sup_{u\in[s,t]}\big|f(u,\cdot)\big|_{p\text{-}var;[s,t]} + \sup_{v\in[s,t]}\big|f(\cdot,v)\big|_{p\text{-}var;[s,t]} + |f|_{\infty;[s,t]}\Big)^q \\ &\qquad \times |g|^q_{q\text{-}var;[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]}.\end{aligned}$$

Take a partition $D\subset[s,t]$ and $u\in[s,t]$. Then

$$\sum_{t_i\in D}\big|f(u,t_{i+1})-f(u,t_i)\big|^p = \sum_{t_i\in D}\left|f\begin{pmatrix} s, u \\ t_i, t_{i+1}\end{pmatrix}\right|^p \le V_p\big(f,[s,t]^2\big)^p$$

and hence

$$\sup_{u\in[s,t]}\big|f(u,\cdot)\big|_{p\text{-}var;[s,t]} \le V_p\big(f,[s,t]^2\big).$$

The same way one obtains

$$\sup_{v\in[s,t]}\big|f(\cdot,v)\big|_{p\text{-}var;[s,t]} \le V_p\big(f,[s,t]^2\big).$$

Finally, for $u,v\in[s,t]$,

$$\big|f(u,v)\big| = \left|f\begin{pmatrix} s, u \\ s, v\end{pmatrix}\right| \le V_p\big(f,[s,t]^2\big)$$

and therefore $|f|_{\infty;[s,t]} \le V_p(f,[s,t]^2)$. Putting everything together, we end up with

$$\left|\Phi\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1}\end{pmatrix}\right|^q \le c_4 V_p\big(f,[s,t]^2\big)^q |g|^q_{q\text{-}var;[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]}.$$

Hence for every partition $D,\tilde{D}\subset[s,t]$ one gets, using superadditivity of $|g|^q_{q\text{-}var}$,

$$\begin{aligned}\sum_{t_i\in D,\tilde{t}_j\in\tilde{D}}\left|\Phi\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1}\end{pmatrix}\right|^q &\le c_4 V_p\big(f,[s,t]^2\big)^q \sum_{t_i\in D,\tilde{t}_j\in\tilde{D}} |g|^q_{q\text{-}var;[t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]} \\ &\le c_4 V_p\big(f,[s,t]^2\big)^q |g|^q_{q\text{-}var;[s,t]^2}.\end{aligned}$$

Passing to the supremum over all partitions shows the assertion. □

This lemma allows us to define iterated $2D$-integrals. Let $f,g_1,\dots,g_n:[0,1]^2\to\mathbb{R}$. An iterated $2D$-integral is given by $\int_{\Delta^1_{s,t}\times\Delta^1_{s',t'}} f\,\mathrm{d}g_1 = \int_{[s,t]\times[s',t']} f(u,v)\,\mathrm{d}g_1(u,v)$ for $n=1$ and recursively defined by

$$\int_{\Delta^n_{s,t}\times\Delta^n_{s',t'}} f\,\mathrm{d}g_1\cdots\mathrm{d}g_n := \int_{[s,t]\times[s',t']}\left(\int_{\Delta^{n-1}_{s,u}\times\Delta^{n-1}_{s',v}} f\,\mathrm{d}g_1\cdots\mathrm{d}g_{n-1}\right)\mathrm{d}g_n(u,v)$$

for $n\ge 2$.

Proposition 4. *Let $f, g_1, g_2, \ldots : [0,1]^2 \to \mathbb{R}$ and $p, q_1, q_2, \ldots$ be real numbers such that $p^{-1} + q_1^{-1} > 1$ and $q_i^{-1} + q_{i+1}^{-1} > 1$ for every $i \geq 1$. Assume that f has finite p-variation and g_i has finite q_i-variation for $i = 1, 2, \ldots$ and that for $(s,t) \in \Delta$ we have $f(s,\cdot) = f(\cdot,s) = 0$. Then for every $n \in \mathbb{N}$ there is a constant $C = C(p, q_1, \ldots, q_n)$ such that*

$$\left| \int_{\Delta_{s,t}^n \times \Delta_{s,t}^n} f \, \mathrm{d}g_1 \cdots \mathrm{d}g_n \right| \leq C V_p\bigl(f, [s,t]^2\bigr) V_{q_1}\bigl(g_1, [s,t]^2\bigr) \cdots V_{q_n}\bigl(g_n, [s,t]^2\bigr).$$

Proof. Define $\Phi^{(n)}(u,v) = \int_{\Delta_{s,u}^n \times \Delta_{s,v}^n} f \, \mathrm{d}g_1 \cdots \mathrm{d}g_n$. We will show a stronger result; namely that for every $n \in \mathbb{N}$ and $q_n' > q_n$ there is a constant $C = C(p, q_1, \ldots, q_n, q_n')$ such that

$$V_{q_n'}\bigl(\Phi^{(n)}, [s,t]^2\bigr) \leq C V_p\bigl(f, [s,t]^2\bigr) V_{q_1}\bigl(g_1, [s,t]^2\bigr) \cdots V_{q_n}\bigl(g_n, [s,t]^2\bigr).$$

To do so, let $\tilde{q}_1, \tilde{q}_2, \ldots$ be a sequence of real numbers such that $\tilde{q}_j > q_j$ and $\frac{1}{\tilde{q}_{j-1}} + \frac{1}{\tilde{q}_j} > 1$ for every $j = 1, 2, \ldots$ where we set $\tilde{q}_0 = p$. We make an induction over n. For $n = 1$, we have $\tilde{q}_1 > q_1$ and $\frac{1}{p} + \frac{1}{\tilde{q}_1} > 1$, hence from Theorem 2 we know that g_1 has finite controlled $\tilde{q}_1$-variation and Lemma 2 gives us

$$V_{\tilde{q}_1}\bigl(\Phi^{(1)}; [s,t]^2\bigr) \leq c_1 V_p\bigl(f; [s,t]^2\bigr) |g_1|_{\tilde{q}_1; [s,t]^2} \leq c_2 V_p\bigl(f; [s,t]^2\bigr) V_{q_1}\bigl(g_1; [s,t]^2\bigr).$$

W.l.o.g, we may assume that $q_1' > \tilde{q}_1 > q_1$, otherwise we choose $\tilde{q}_1$ smaller in the beginning. From $V_{q_1'}(\Phi^{(1)}; [s,t]^2) \leq V_{\tilde{q}_1}(\Phi^{(1)}; [s,t]^2)$ the assertion follows for $n = 1$. Now take $n \in \mathbb{N}$. Note that

$$\Phi^{(n)}(u,v) = \int_{[s,u] \times [s,v]} \Phi^{(n-1)} \, \mathrm{d}g_n$$

and clearly $\Phi^{(n-1)}(s,\cdot) = \Phi^{(n-1)}(\cdot,s) = 0$. We can use Lemma 2 again to see that

$$\begin{aligned} V_{\tilde{q}_n}\bigl(\Phi^{(n)}, [s,t]^2\bigr) &\leq c_3 V_{\tilde{q}_{n-1}}\bigl(\Phi^{(n-1)}; [s,t]^2\bigr) |g_n|_{\tilde{q}_n\text{-}var; [s,t]^2} \\ &\leq c_4 V_{\tilde{q}_{n-1}}\bigl(\Phi^{(n-1)}; [s,t]^2\bigr) V_{q_n}\bigl(g_n; [s,t]^2\bigr). \end{aligned}$$

Using our induction hypothesis shows the result for $\tilde{q}_n$. By choosing $\tilde{q}_n$ smaller in the beginning if necessary, we may assume that $q_n' > \tilde{q}_n$ and the assertion follows. □

5. The main estimates

In the following section, $(X, Y) = (X^1, Y^1, \ldots, X^d, Y^d)$ will always denote a centred continuous Gaussian process where (X^i, Y^i) and (X^j, Y^j) are independent for $i \neq j$. We will also assume that the ρ-variation of $R_{(X,Y)}$ is finite for a $\rho < 2$ and controlled by a symmetric $2D$-control ω (this in particular implies that the ρ-variation of R_X, R_Y and R_{X-Y} is controlled by ω, see [7], Section 15.3.2). Let $\gamma > \rho$ such that $\frac{1}{\rho} + \frac{1}{\gamma} > 1$. The aim of this section is to show that for every $n \in \mathbb{N}$ there are constants $C(n)$ such that[7]

$$\bigl|\mathbf{X}_{s,t}^n - \mathbf{Y}_{s,t}^n\bigr|_{L^2((\mathbb{R}^d)^{\otimes n})} \leq C(n) \varepsilon \omega\bigl([s,t]^2\bigr)^{1/(2\gamma)} \omega\bigl([s,t]^2\bigr)^{(n-1)/(2\rho)} \quad \text{for every } s < t, \tag{5.1}$$

where $\varepsilon^2 = V_\infty(R_{X-Y}, [s,t]^2)^{1-\rho/\gamma}$ (see Definition 11 below for the exact definition of V_∞). Equivalently, we might show (5.1) coordinate-wise, i.e. proving that the same estimate holds for $|\mathbf{X}^w - \mathbf{Y}^w|_{L^2(\mathbb{R})}$ for every word w formed by the alphabet $A = \{1, \ldots, d\}$. In some special cases, i.e. if a word w has a very simple structure, we can do this directly using multidimensional Young integration. This is done in Section 5.1. Section 5.2 shows (5.1) for $n = 1, 2, 3, 4$

[7] We prefer to write it in this notation instead of writing $\omega([s,t]^2)^{1/(2\gamma)+(n-1)/(2\rho)}$ to emphasize the different roles of the two terms. The first term will play no particular role and just comes from interpolation whereas the second one will be crucial when doing the induction step from lower to higher levels in Proposition 8.

coordinate-wise, using the shuffle algebra structure for iterated integrals and multidimensional Young integration. In Section 5.3, we show (5.1) coordinate-free for all $n > 4$, using an induction argument very similar to the one Lyon's used for proving the Extension Theorem (cf. [14]).

We start with giving a 2-dimensional analogue for the one-dimensional interpolation inequality.

Definition 11. *If $f : [0,1]^2 \to B$ is a continuous function in a Banach space and $(s,t) \times (u,v) \in \Delta \times \Delta$ we set*

$$V_\infty\big(f, [s,t] \times [u,v]\big) = \sup_{A \subset [s,t]\times[u,v]} \big|f(A)\big|.$$

Lemma 3. *For $\gamma > \rho \geq 1$ we have the interpolation inequality*

$$V_{\gamma\text{-}var}\big(f, [s,t] \times [u,v]\big) \leq V_\infty\big(f, [s,t] \times [u,v]\big)^{1-\rho/\gamma} V_{\rho\text{-}var}\big(f, [s,t] \times [u,v]\big)^{\rho/\gamma}$$

for all $(s,t), (u,v) \in \Delta$.

Proof. Exactly as $1D$-interpolation, see [7], Proposition 5.5. □

5.1. *Some special cases*

If $Z : [0,1] \to \mathbb{R}$ is a process with smooth sample paths, we will use the notation

$$\mathbf{Z}^{(n)}_{s,t} = \int_{\Delta^n_{s,t}} \mathrm{d}Z \cdots \mathrm{d}Z$$

for $s < t$.

Lemma 4. *Let $X : [0,1] \to \mathbb{R}$ be a centred Gaussian process with continuous paths of finite variation and assume that the ρ-variation of the covariance R_X is controlled by a $2D$-control ω. For fixed $s < t$, define*

$$f(u,v) = E\big(\mathbf{X}^{(n)}_{s,u}\mathbf{X}^{(n)}_{s,v}\big).$$

Then there is a constant $C = C(\rho, n)$ such that

$$V_\rho\big(f, [s,t]^2\big) \leq C\omega\big([s,t]^2\big)^{n/\rho}.$$

Proof. Let $t_i < t_{i+1}$, $\tilde{t}_j < \tilde{t}_{j+1}$. Then

$$f\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1} \end{pmatrix} = E\big(\big(\mathbf{X}^{(n)}_{s,t_{i+1}} - \mathbf{X}^{(n)}_{s,t_i}\big)\big(\mathbf{X}^{(n)}_{s,\tilde{t}_{j+1}} - \mathbf{X}^{(n)}_{s,\tilde{t}_j}\big)\big).$$

We know that $\mathbf{X}^{(n)} = \frac{(X)^n}{n!}$. From the identity

$$b^n - a^n = (b-a)\big(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}\big)$$

we deduce that

$$f\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1} \end{pmatrix} = \frac{1}{(n!)^2} \sum_{k,l=0}^{n-1} E\big(X_{t_i,t_{i+1}} X_{\tilde{t}_j,\tilde{t}_{j+1}} (X_{s,t_{i+1}})^{n-1-k} (X_{s,t_i})^k (X_{s,\tilde{t}_{j+1}})^{n-1-l} (X_{s,\tilde{t}_j})^l\big).$$

We want to apply Wick's formula now (cf. [13], Theorem 1.28). If $Z, \tilde{Z} \in \{X_{s,t_{i+1}}, X_{s,t_i}, X_{s,\tilde{t}_{j+1}}, X_{s,\tilde{t}_j}\}$ we know that

$$\big|E(X_{t_i,t_{i+1}} Z)\big|^\rho \leq \omega\big([t_i, t_{i+1}] \times [s,t]\big),$$

$$\big|E(X_{t_i,t_{i+1}} X_{\tilde{t}_j,\tilde{t}_{j+1}})\big|^\rho \leq \omega\big([t_i, t_{i+1}] \times [\tilde{t}_j, \tilde{t}_{j+1}]\big),$$

$$\big|E(Z\tilde{Z})\big|^\rho \leq \omega\big([s,t]^2\big)$$

and the same holds for $X_{\tilde{t}_j,\tilde{t}_{j+1}}$. Now take two partitions $D, \tilde{D} \in [0,1]$. Then, by Wick's formula and the estimates above,

$$\begin{aligned}\sum_{t_i\in D,\tilde{t}_j\in\tilde{D}} \left| f\begin{pmatrix} t_i, t_{i+1} \\ \tilde{t}_j, \tilde{t}_{j+1}\end{pmatrix}\right|^\rho &\le c_1(\rho,n)\omega\big([s,t]^2\big)^{n-2} \sum_{t_i\in D,\tilde{t}_j\in\tilde{D}} \omega\big([t_i,t_{i+1}]\times[s,t]\big)\omega\big([\tilde{t}_j,\tilde{t}_{j+1}]\times[s,t]\big) \\ &\quad + c_2(\rho,n)\omega\big([s,t]^2\big)^{n-1} \sum_{t_i\in D,\tilde{t}_j\in\tilde{D}} \omega\big([t_i,t_{i+1}]\times[\tilde{t}_j,\tilde{t}_{j+1}]\big) \\ &\le c_3\omega\big([s,t]^2\big)^n. \end{aligned}$$

□

Lemma 5. *Let (X,Y) be a centred Gaussian process in $\mathbb{R}^2$ with continuous paths of finite variation. Assume that the ρ-variation of $R_{(X,Y)}$ is controlled by a $2D$-control ω for $\rho < 2$ and take $\gamma > \rho$. Then for every $n \in \mathbb{N}$ there is a constant $C = C(n)$ such that*

$$\big|\mathbf{X}^{(n)}_{s,t} - \mathbf{Y}^{(n)}_{s,t}\big|_{L^2} \le C(n)\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{(n-1)/(2\rho)}$$

for any $s < t$ where $\varepsilon^2 = V_\infty(R_{X-Y},[s,t]^2)^{1-\rho/\gamma}$.

Proof. By induction. For $n = 1$ we simply have from Lemma 3

$$\begin{aligned}|X_{s,t} - Y_{s,t}|^2_{L^2} &= E\big[(X_{s,t}-Y_{s,t})(X_{s,t}-Y_{s,t})\big] \le V_{\gamma\text{-}var}\big(R_{X-Y},[s,t]^2\big) \\ &\le \varepsilon^2 V_{\rho\text{-}var}\big(R_{X-Y},[s,t]^2\big)^{\rho/\gamma} \le \varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma}.\end{aligned}$$

For $n \in \mathbb{N}$ we use the identity

$$\mathbf{X}^{(n)}_{s,t} - \mathbf{Y}^{(n)}_{s,t} = \frac{1}{n}\big(X_{s,t}\mathbf{X}^{(n-1)}_{s,t} - Y_{s,t}\mathbf{Y}^{(n-1)}_{s,t}\big)$$

and hence

$$\begin{aligned}\big|\mathbf{X}^{(n)}_{s,t} - \mathbf{Y}^{(n)}_{s,t}\big|_{L^2} &\le c_1\big(|X_{s,t}-Y_{s,t}|_{L^2}\big|\mathbf{X}^{(n-1)}_{s,t}\big|_{L^2} + \big|\mathbf{X}^{(n-1)}_{s,t} - \mathbf{Y}^{(n-1)}_{s,t}\big|_{L^2}|Y_{s,t}|_{L^2}\big) \\ &\le c_2\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{(n-1)/(2\rho)}.\end{aligned}$$

□

Assume that (Z^1, Z^2) is a centred, continuous Gaussian process in $\mathbb{R}^2$ with smooth sample paths and that both components are independent. Then (at least formally, cf. [6]),

$$\left|\int_0^1 Z^1_{0,u}\,\mathrm{d}Z^2_u\right|^2_{L^2} = E\left[\left(\int_0^1 Z^1_{0,u}\,\mathrm{d}Z^2_u\right)^2\right] = E\left[\int_{[0,1]^2} Z^1_{0,u}Z^1_{0,v}\,\mathrm{d}Z^2\,\mathrm{d}Z^2_v\right] \tag{5.2}$$

$$= \int_{[0,1]^2} E\big[Z^1_{0,u}Z^1_{0,v}\big]\,\mathrm{d}E\big[Z^2_u Z^2_v\big] = \int_{[0,1]^2} R_{Z^1}\begin{pmatrix} 0 & \cdot \\ 0 & \cdot\end{pmatrix}\mathrm{d}R_{Z^2}, \tag{5.3}$$

where the integrals in the second row are $2D$ Young-integrals (to make this rigorous, one uses that the integrals are a.s. limits of Riemann sums and that a.s. convergence implies convergence in L^1 in the (inhomogeneous) Wiener chaos). These kinds of computations together with our estimates for $2D$ Young-integrals will be heavily used from now on.

Lemma 6. *Let $(X,Y) = (X^1,Y^1,\ldots,X^d,Y^d)$ be a centred Gaussian process with continuous paths of finite variation where $(X^i,Y^i)$and (X^j,Y^j) are independent for $i \neq j$. Assume that the ρ-variation of $R_{(X,Y)}$ is controlled by a $2D$-control ω for $\rho < 2$. Let w be a word of the form $w = i_1\cdots i_n$ where $i_1,\ldots,i_n \in \{1,\ldots,d\}$ are all distinct. Take $\gamma > \rho$ such that $\frac{1}{\rho}+\frac{1}{\gamma} > 1$. Then there is a constant $C = C(\rho,\gamma,n)$ such that*

$$\big|\mathbf{X}^w_{s,t} - \mathbf{Y}^w_{s,t}\big|_{L^2} \le C(n)\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{(n-1)/(2\rho)}$$

for any $s < t$ *where* $\varepsilon^2 = V_\infty(R_{X-Y}, [s,t]^2)^{1-\rho/\gamma}$.

Proof. By the triangle inequality,

$$\left|\mathbf{X}^w_{s,t} - \mathbf{Y}^w_{s,t}\right|_{L^2} = \left|\int_{\Delta^n_{s,t}} \mathrm{d}X^{i_1}\cdots \mathrm{d}X^{i_n} - \int_{\Delta^n_{s,t}} \mathrm{d}Y^{i_1}\cdots \mathrm{d}Y^{i_n}\right|_{L^2}$$
$$\le \sum_{k=1}^n \left|\int_{\Delta^n_{s,t}} \mathrm{d}Y^{i_1}\cdots \mathrm{d}Y^{i_{k-1}}\,\mathrm{d}\big(X^{i_k} - Y^{i_k}\big)\,\mathrm{d}X^{i_{k+1}}\cdots \mathrm{d}X^{i_n}\right|_{L^2}.$$

From independence, Proposition 4 and Lemma 3

$$\left|\int_{\Delta^n_{s,t}} dY^{i_1}\cdots \mathrm{d}Y^{i_{k-1}}\,\mathrm{d}\big(X^{i_k} - Y^{i_k}\big)\,\mathrm{d}X^{i_{k+1}}\cdots \mathrm{d}X^{i_n}\right|^2_{L^2}$$
$$= \int_{\Delta^n_{s,t}\times\Delta^n_{s,t}} \mathrm{d}R_{Y^{i_1}}\cdots \mathrm{d}R_{Y^{i_{k-1}}}\,\mathrm{d}R_{X^{i_k}-Y^{i_k}}\,\mathrm{d}R_{X^{i_{k+1}}}\cdots \mathrm{d}R_{X^{i_n}}$$
$$\le c_1 V_\rho\big(R_{Y^{i_1}}, [s,t]^2\big)\cdots V_\rho\big(R_{Y^{i_{k-1}}}, [s,t]^2\big) V_\gamma\big(R_{X^{i_k}-Y^{i_k}}, [s,t]^2\big)$$
$$\times V_\rho\big(R_{X^{i_{k+1}}}, [s,t]^2\big)\cdots V_\rho\big(R_{X^{i_n}}, [s,t]^2\big)$$
$$\le c_1 V_\gamma\big(R_{X-Y}, [s,t]^2\big)\omega\big([s,t]^2\big)^{(n-1)/\rho} \le c_1\varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma}\omega\big([s,t]^2\big)^{(n-1)/\rho}.$$

The first inequality above is an immediate generalization of the calculations made in (5.2) and (5.3). Note that the respective random terms are not only pairwise but mutually independent here since we are dealing with a Gaussian process (X,Y). Interchanging the limits is allowed since convergence in probability implies convergence in L^p, any $p > 0$, in the Wiener chaos. □

5.2. *Lower levels*

5.2.1. $n = 1, 2$

Proposition 5. *Let* (X,Y), ω, ρ *and* γ *as in Lemma 6. Then there are constants* $C(1), C(2)$ *which depend on* ρ *and* γ *such that*

$$\left|\mathbf{X}^n_{s,t} - \mathbf{Y}^n_{s,t}\right|_{L^2} \le C(n)\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{(n-1)/(2\rho)}$$

holds for $n = 1, 2$ *and every* $(s,t) \in \Delta$ *where* $\varepsilon^2 = V_\infty(R_{X-Y}, [s,t]^2)^{1-\rho/\gamma}$.

Proof. The coordinate-wise estimates are just special cases of Lemma 5 and Lemma 6. □

5.2.2. $n = 3$

Proposition 6. *Let* (X,Y), ω, ρ *and* γ *as in Lemma 6. Then there is a constant* $C(3)$ *which depends on* ρ *and* γ *such that*

$$\left|\mathbf{X}^3_{s,t} - \mathbf{Y}^3_{s,t}\right|_{L^2} \le C(3)\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{2/(2\rho)}$$

holds for every $(s,t) \in \Delta$ *where* $\varepsilon^2 = V_\infty(R_{X-Y}, [s,t]^2)^{1-\rho/\gamma}$.

Proof. We have to show the estimate for $\mathbf{X}^{i,j,k} - \mathbf{Y}^{i,j,k}$ where $i,j,k \in \{1,\dots,d\}$. From Proposition 3 and 2 it follows that it is enough to show the estimate for $\mathbf{X}^w - \mathbf{Y}^w$ where

$$w \in \big\{iii, ijk, iij\colon\, i,j,k \in \{1,\dots,d\} \text{ distinct}\big\}.$$

The cases $w = iii$ and $w = ijk$ are special cases of Lemma 5 and Lemma 6. The rest of this section is devoted to show the estimate for $w = iij$. □

Lemma 7. *Let $(X, Y) : [0, 1] \to \mathbb{R}^2$ be a centred Gaussian process and consider*

$$f(u, v) = E\big((X_u - Y_u)X_v\big).$$

Assume that the ρ-variation of $R_{(X,Y)}$ is controlled by a 2D-control ω where $\rho \geq 1$. Let $s < t$ and consider a rectangle $[\sigma, \tau] \times [\sigma', \tau'] \subset [s, t]^2$. Let $\gamma > \rho$. Then

$$V_{\gamma\text{-}var}\big(f, [\sigma, \tau] \times \big[\sigma', \tau'\big]\big) \leq \varepsilon \omega\big([s, t]^2\big)^{1/2(1/\rho - 1/\gamma)} \omega\big([\sigma, \tau] \times \big[\sigma', \tau'\big]\big)^{1/\gamma},$$

where $\varepsilon^2 = V_\infty(R_{X-Y}, [s, t]^2)^{1-\rho/\gamma}$.

Proof. Let $u < v$ and $u' < v' \in [s, t]$. Then

$$\begin{aligned}\big|E\big((X_{u,v} - Y_{u,v})X_{u',v'}\big)\big| &\leq |X_{u,v} - Y_{u,v}|_{L^2} |X_{u',v'}|_{L^2} \\ &\leq V_\infty\big(R_{X-Y}, [s, t]^2\big)^{1/2} V_{\rho\text{-}var}\big(R_{(X,Y)}, [s, t]^2\big)^{1/2}\end{aligned}$$

and hence

$$\sup_{u<v, u'<v'} \big|E\big((X_{u,v} - Y_{u,v})X_{u',v'}\big)\big| \leq V_\infty\big(R_{X-Y}, [s, t]^2\big)^{1/2} \omega\big([s, t]^2\big)^{1/(2\rho)}.$$

Now take a partition D of $[\sigma, \tau]$ and a partition $\tilde{D}$ of $[\sigma', \tau']$. Then

$$\begin{aligned}&\sum_{t_i \in D, \tilde{t}_j \in \tilde{D}} \big|E\big((X_{t_i,t_{i+1}} - Y_{t_i,t_{i+1}})X_{\tilde{t}_j,\tilde{t}_{j+1}}\big)\big|^\gamma \\ &\quad \leq \sup_{u<v, u'<v'} \big|E\big((X_{u,v} - Y_{u,v})X_{u',v'}\big)\big|^{\gamma-\rho} \sum_{t_i \in D, \tilde{t}_j \in \tilde{D}} \big|E\big((X_{t_i,t_{i+1}} - Y_{t_i,t_{i+1}})X_{\tilde{t}_j,\tilde{t}_{j+1}}\big)\big|^\rho \\ &\quad \leq V_\infty\big(R_{X-Y}, [s, t]^2\big)^{1/2(\gamma-\rho)} \omega\big([s, t]^2\big)^{1/2(\gamma/\rho - 1)} \omega\big([\sigma, \tau] \times \big[\sigma', \tau'\big]\big)\end{aligned}$$

and taking the supremum over all partitions shows the result. □

Lemma 8. *Let $(X, Y) : [0, 1] \to \mathbb{R}^2$ be a centred Gaussian process with continuous paths of finite variation. Assume that the ρ-variation of $R_{(X,Y)}$ is controlled by a 2D-control ω where $\rho \geq 1$. Consider the function*

$$g(u, v) = E\big[\big(\mathbf{X}^{(2)}_{s,u} - \mathbf{Y}^{(2)}_{s,u}\big)\big(\mathbf{X}^{(2)}_{s,v} - \mathbf{Y}^{(2)}_{s,v}\big)\big].$$

Then for every $\gamma > \rho$ there is a constant $C = C(\rho, \gamma)$ such that

$$V_{\gamma\text{-}var}\big(g, [s, t]^2\big) \leq C\varepsilon^2 \omega\big([s, t]^2\big)^{1/\gamma + 1/\rho}$$

holds for every $(s, t) \in \Delta$ where $\varepsilon^2 = V_\infty(R_{X-Y}, [s, t]^2)^{1-\rho/\gamma}$.

Proof. Let $u < v$ and $u' < v'$. Then

$$\begin{aligned}g\begin{pmatrix} u, v \\ u', v' \end{pmatrix} &= E\big[\big(\big(\mathbf{X}^{(2)}_{s,v} - \mathbf{X}^{(2)}_{s,u}\big) - \big(\mathbf{Y}^{(2)}_{s,v} - \mathbf{Y}^{(2)}_{s,u}\big)\big)\big(\big(\mathbf{X}^{(2)}_{s,v'} - \mathbf{X}^{(2)}_{s,u'}\big) - \big(\mathbf{Y}^{(2)}_{s,v'} - \mathbf{Y}^{(2)}_{s,u'}\big)\big)\big] \\ &= \frac{1}{2^2} E\big[\big(\big(X^2_{s,v} - X^2_{s,u}\big) - \big(Y^2_{s,v} - Y^2_{s,u}\big)\big)\big(\big(X^2_{s,v'} - X^2_{s,u'}\big) - \big(Y^2_{s,v'} - Y^2_{s,u'}\big)\big)\big].\end{aligned}$$

Now,

$$\begin{aligned}\left(X_{s,v}^2 - X_{s,u}^2\right) - \left(Y_{s,v}^2 - Y_{s,u}^2\right) &= X_{u,v}(X_{s,u} + X_{s,v}) - Y_{u,v}(Y_{s,u} + Y_{s,v}) \\ &= X_{u,v}(X_{s,u} - Y_{s,u}) + (X_{u,v} - Y_{u,v})Y_{s,u} \\ &\quad + X_{u,v}(X_{s,v} - Y_{s,v}) + (X_{u,v} - Y_{u,v})Y_{s,v}.\end{aligned}$$

The same way one gets

$$\begin{aligned}\left(X_{s,v'}^2 - X_{s,u'}^2\right) - \left(Y_{s,v'}^2 - Y_{s,u'}^2\right) &= X_{u',v'}(X_{s,u'} - Y_{s,u'}) + (X_{u',v'} - Y_{u',v'})Y_{s,u'} \\ &\quad + X_{u',v'}(X_{s,v'} - Y_{s,v'}) + (X_{u',v'} - Y_{u',v'})Y_{s,v'}.\end{aligned}$$

Now we expand the product of both sums and take expectation. For the first term we obtain, using the Wick formula and Lemma 7,

$$\begin{aligned}&\left|E\left(X_{u,v}(X_{s,u} - Y_{s,u})X_{u',v'}(X_{s,u'} - Y_{s,u'})\right)\right| \\ &\quad \le \left|E(X_{u,v}X_{u',v'})E\left[(X_{s,u} - Y_{s,u})(X_{s,u'} - Y_{s,u'})\right]\right| \\ &\qquad + \left|E\left[X_{u,v}(X_{s,u'} - Y_{s,u'})\right]E\left[X_{u',v'}(X_{s,u} - Y_{s,u})\right]\right| \\ &\qquad + \left|E\left[X_{u',v'}(X_{s,u'} - Y_{s,u'})\right]E\left[X_{u,v}(X_{s,u} - Y_{s,u})\right]\right| \\ &\quad \le V_{\rho\text{-}var}\left(R_{(X,Y)}, [u,v] \times \left[u',v'\right]\right)V_{\gamma\text{-}var}\left(R_{X-Y}, [s,t]^2\right) \\ &\qquad + 2V_{\gamma\text{-}var}\left(R_{(X,X-Y)}, [u,v] \times [s,t]\right)V_{\gamma\text{-}var}\left(R_{(X,X-Y)}, \left[u',v'\right] \times [s,t]\right) \\ &\quad \le \varepsilon^2 \omega\left([u,v] \times \left[u',v'\right]\right)^{1/\rho} \omega\left([s,t]^2\right)^{1/\gamma} \\ &\qquad + 2\varepsilon^2 \omega\left([s,t]^2\right)^{1/\rho - 1/\gamma} \omega\left([u,v] \times [s,t]\right)^{1/\gamma} \omega\left(\left[u',v'\right] \times [s,t]\right)^{1/\gamma}.\end{aligned}$$

Now take two partitions D, $\tilde{D}$ of $[s,t]$. With our calculations above,

$$\begin{aligned}&\sum_{t_i \in D, \tilde{t}_j \in \tilde{D}} \left|E\left(X_{t_i,t_{i+1}}(X_{s,t_i} - Y_{s,t_i})X_{\tilde{t}_j,\tilde{t}_{j+1}}(X_{s,\tilde{t}_j} - Y_{s,\tilde{t}_j})\right)\right|^\gamma \\ &\quad \le c_1 \varepsilon^{2\gamma} \omega\left([s,t]^2\right) \sum_{t_i \in D, \tilde{t}_j \in \tilde{D}} \omega\left([t_i, t_{i+1}] \times [\tilde{t}_j, \tilde{t}_{j+1}]\right)^{\gamma/\rho} \\ &\qquad + c_2 \varepsilon^{2\gamma} \omega\left([s,t]^2\right)^{\gamma/\rho - 1} \sum_{t_i \in D, \tilde{t}_j \in \tilde{D}} \omega\left([t_i, t_{i+1}] \times [s,t]\right) \omega\left([\tilde{t}_j, \tilde{t}_{j+1}] \times [s,t]\right) \\ &\quad \le c_3 \varepsilon^{2\gamma} \left(\omega\left([s,t]^2\right) \omega\left([s,t]^2\right)^{\gamma/\rho} + \omega\left([s,t]^2\right)^{\gamma/\rho - 1} \omega\left([s,t]^2\right)^2\right).\end{aligned}$$

The other terms are treated exactly the same way. Taking the supremum over all partitions shows the result. □

The next corollary completes the proof of Proposition 6.

Corollary 2. *Let (X, Y), ω, ρ and γ as in Lemma 6. Then there is a constant $C = C(\rho, \gamma)$ such that*

$$\left|\mathbf{X}_{s,t}^{i,i,j} - \mathbf{Y}_{s,t}^{i,i,j}\right|_{L^2} \le C\varepsilon\omega\left([s,t]^2\right)^{1/(2\gamma)} \omega\left([s,t]^2\right)^{2/(2\rho)}$$

holds for every $(s,t) \in \Delta$ and $i \ne j$ where $\varepsilon^2 = V_\infty(R_{X-Y}, [s,t]^2)^{1 - \rho/\gamma}$.

Proof. From the triangle inequality,

$$|\mathbf{X}^{i,i,j}_{s,t} - \mathbf{Y}^{i,i,j}_{s,t}|_{L^2} \le \left| \int_{[s,t]} \big(\mathbf{X}^{i,i}_{s,u} - \mathbf{Y}^{i,i}_{s,u}\big)\, \mathrm{d}Y^j_u \right|_{L^2} + \left| \int_{[s,t]} \mathbf{Y}^{i,i}_{s,u}\, \mathrm{d}\big(X^j - Y^j\big)_u \right|_{L^2}.$$

For the first integral, we use independence to move the expectation inside the integral as seen in the proof of Lemma 6, then we use $2D$ Young integration and Lemma 8 to obtain the desired estimate. The second integral is estimated in the same way using Lemma 4. □

5.2.3. $n = 4$

Proposition 7. *Let (X, Y), ω, ρ and γ as in Lemma 6. Then there is a constant $C(4)$ which depends on ρ and γ such that*

$$|\mathbf{X}^4_{s,t} - \mathbf{Y}^4_{s,t}|_{L^2} \le C(4)\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{3/(2\rho)}$$

holds for every $(s,t) \in \Delta$ where $\varepsilon^2 = V_\infty(R_{X-Y}, [s,t]^2)^{1-\rho/\gamma}$.

Proof. From Propositions 3 and 2 one sees that it is enough to show the estimate for $\mathbf{X}^w - \mathbf{Y}^w$ where

$$w \in \big\{iiii, ijkl, iijj, iiij, iijk, jiik\colon i,j,k,l \in \{1,\dots,d\} \text{ distinct}\big\}.$$

The cases $w = iiii$ and $w = ijkl$ are special cases of Lemma 5 and Lemma 6. Hence it remains to show the estimate for

$$w \in \big\{iijj, iiij, iijk, jiik\colon i,j,k \in \{1,\dots,d\} \text{ pairwise distinct}\big\}.$$

This is the content of the remaining section. □

Lemma 9. *Let (X, Y), ω, ρ and γ as in Lemma 6. Then there is a constant $C = C(\rho,\gamma)$ such that*

$$|\mathbf{X}^{i,i,j,k}_{s,t} - \mathbf{Y}^{i,i,j,k}_{s,t}|_{L^2} \le C\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{3/(2\rho)}$$

holds for every $(s,t) \in \Delta$ where i,j,k are distinct and $\varepsilon^2 = V_\infty(R_{X-Y}, [s,t]^2)^{1-\rho/\gamma}$.

Proof. From the triangle inequality,

$$\begin{aligned}
&|\mathbf{X}^{i,i,j,k}_{s,t} - \mathbf{Y}^{i,i,j,k}_{s,t}|_{L^2} \\
&\quad = \left| \int_{\{s<u<v<t\}} \mathbf{X}^{i,i}_{s,u}\, \mathrm{d}X^j_u\, \mathrm{d}X^k_v - \int_{\{s<u<v<t\}} \mathbf{Y}^{i,i}_{s,u}\, \mathrm{d}Y^j_u\, \mathrm{d}Y^k_v \right|_{L^2} \\
&\quad \le \left| \int_{\{s<u<v<t\}} \big(\mathbf{X}^{i,i}_{s,u} - \mathbf{Y}^{i,i}_{s,u}\big)\, \mathrm{d}X^j_u\, \mathrm{d}X^k_v \right|_{L^2} + \left| \int_{\{s<u<v<t\}} \mathbf{Y}^{i,i}_{s,u}\, \mathrm{d}\big(X^j - Y^j\big)_u\, \mathrm{d}X^k_v \right|_{L^2} \\
&\qquad + \left| \int_{\{s<u<v<t\}} \mathbf{Y}^{i,i}_{s,u}\, \mathrm{d}Y^j_u\, \mathrm{d}\big(X^k - Y^k\big)_v \right|_{L^2}.
\end{aligned}$$

For the first integral, we use Proposition 4 and Lemma 8 to obtain

$$\begin{aligned}
\left| \int_{\{s<u<v<t\}} \big(\mathbf{X}^{i,i}_{s,u} - \mathbf{Y}^{i,i}_{s,u}\big)\, \mathrm{d}X^j_u\, \mathrm{d}X^k_v \right|^2_{L^2} &= \int_{\Delta^2_{s,t}\times\Delta^2_{s,t}} E\big[\big(\mathbf{X}^{i,i}_{s,\cdot} - \mathbf{Y}^{i,i}_{s,\cdot}\big)\big(\mathbf{X}^{i,i}_{s,\cdot} - \mathbf{Y}^{i,i}_{s,\cdot}\big)\big]\, \mathrm{d}R_{X^j}\, \mathrm{d}R_{X^k} \\
&\le c_1\varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma+1/\rho}\omega\big([s,t]^2\big)^{2/\rho}.
\end{aligned}$$

For the other two integrals we also use Proposition 4 together with Lemma 4 to obtain the same estimate. □

Lemma 10. *Let $(X, Y) : [0, 1] \to \mathbb{R}^2$ be a centred Gaussian process with continuous paths of finite variation. Assume that the ρ-variation of $R_{(X,Y)}$ is controlled by a $2D$-control ω where $\rho \geq 1$. Consider the function*

$$g(u, v) = E\big[\big(\mathbf{X}^{(3)}_{s,u} - \mathbf{Y}^{(3)}_{s,u}\big)\big(\mathbf{X}^{(3)}_{s,v} - \mathbf{Y}^{(3)}_{s,v}\big)\big].$$

Then for every $\gamma > \rho$ there is a constant $C = C(\rho, \gamma)$ such that

$$V_{\gamma\text{-}var}\big(g, [s, t]^2\big) \leq C\varepsilon^2 \omega\big([s, t]^2\big)^{1/\gamma + 2/\rho}$$

holds for every $(s, t) \in \Delta$ where $\varepsilon^2 = V_\infty(R_{X-Y}, [s, t]^2)^{(1-\rho/\gamma)}$.

Proof. Similar to the one of Lemma 8 applying again Wick's formula. □

Corollary 3. *Let (X, Y), ω, ρ and γ as in Lemma 6. Then there is a constant $C = C(\rho, \gamma)$ such that*

$$\big|\mathbf{X}^{i,i,i,j}_{s,t} - \mathbf{Y}^{i,i,i,j}_{s,t}\big|_{L^2} \leq C\varepsilon \omega\big([s, t]^2\big)^{1/(2\gamma)} \omega\big([s, t]^2\big)^{3/(2\rho)}$$

holds for every $(s, t) \in \Delta$ and $i \neq j$ where $\varepsilon^2 = V_\infty(R_{X-Y}, [s, t]^2)^{(1-\rho/\gamma)}$.

Proof. The triangle inequality gives

$$\begin{aligned}\big|\mathbf{X}^{i,i,i,j}_{s,t} - \mathbf{Y}^{i,i,i,j}_{s,t}\big|_{L^2} &= \left|\int_{[s,t]} \mathbf{X}^{i,i,i}_{s,u}\, \mathrm{d}X^j_u - \int_{[s,t]} \mathbf{Y}^{i,i,i}_{s,u}\, \mathrm{d}Y^j_u\right| \\ &\leq \left|\int_{[s,t]} \big(\mathbf{X}^{i,i,i}_{s,u} - \mathbf{Y}^{i,i,i}_{s,u}\big)\, \mathrm{d}X^j_u\right|_{L^2} + \left|\int_{[s,t]} \mathbf{Y}^{i,i,i}_{s,u}\, \mathrm{d}\big(X^j - Y^j\big)_u\right|_{L^2}.\end{aligned}$$

For the first integral, we move the expectation inside the integral, use $2D$ Young integration and Lemma 10 to conclude the estimate. The second integral is estimated the same way applying Lemma 4. □

It remains to show the estimates for $\mathbf{X}^w - \mathbf{Y}^w$ where $w \in \{iijj, jiik\}$. We need to be a bit careful here for the following reason: It is clear that $\mathbf{X}^{i,i,j}_{0,1} = \int_{[0,1]} \mathbf{X}^{i,i}_u\, \mathrm{d}X^j_u$. One might expect that also $\mathbf{X}^{j,i,i}_{0,1} = \int_{[0,1]} X^j_u\, \mathrm{d}\mathbf{X}^{i,i}_u$ holds, but this is not true in general. Indeed, just take $f(u) = g(u) = u$. Then

$$\int_0^1 f(u)\, \mathrm{d}\bigg(\int_0^u g(v)\, \mathrm{d}g(v)\bigg) = \frac{1}{2}\int_0^1 u\, \mathrm{d}\big(u^2\big) = \int_0^1 u^2\, \mathrm{d}u = \frac{1}{3}$$

but

$$\int_{\Delta^2_{0,1}} f(u)\, \mathrm{d}g(u)\, \mathrm{d}g(v) = \int_{\Delta^3_{0,1}} \mathrm{d}u_1\, \mathrm{d}u_2\, \mathrm{d}u_3 = \frac{1}{6}.$$

One the other hand, if g is smooth, we can use Fubini to see that

$$\begin{aligned}\int_{\Delta^2_{0,1}} f(u)\, \mathrm{d}g(u)\, \mathrm{d}g(v) &= \int_{[0,1]^2} f(u)g'(u)g'(v)1_{\{u<v\}}\, \mathrm{d}u\, \mathrm{d}v \\ &= \frac{1}{2}\int_{[0,1]^2} f(u)g'(u)g'(v)1_{\{u<v\}}\, \mathrm{d}u\, \mathrm{d}v \\ &\quad + \frac{1}{2}\int_{[0,1]^2} f(v)g'(v)g'(u)1_{\{v<u\}}\, \mathrm{d}u\, \mathrm{d}v \\ &= \frac{1}{2}\int_{[0,1]^2} \big(f(u)1_{\{u<v\}} + f(v)1_{\{v<u\}}\big)g'(u)g'(v)\, \mathrm{d}u\, \mathrm{d}v\end{aligned}$$

$$= \frac{1}{2} \int_{[0,1]^2} f(u \wedge v) g'(u) g'(v) \, \mathrm{d}u \, \mathrm{d}v$$
$$= \frac{1}{2} \int_{[0,1]^2} f(u \wedge v) \, \mathrm{d}\big(g(u)g(v)\big),$$

where the last integral is a $2D$ Young integral. Hence we have seen that an iterated $1D$-integral can be transformed into a usual $2D$-integral. We will use this trick for the remaining estimates.

Lemma 11. *Let $f : [0,1]^2 \to \mathbb{R}$ be a continuous function. Set*

$$\bar{f}(u_1, u_2, v_1, v_2) = f(u_1 \wedge u_2, v_1 \wedge v_2).$$

(1) *Let $u_1 < \tilde{u}_1, u_2 < \tilde{u}_2, v_1 < \tilde{v}_1, v_2 < \tilde{v}_2$ be all in $[0,1]$. Then*

$$\bar{f}\begin{pmatrix} u_1, \tilde{u}_1 \\ u_2, \tilde{u}_2 \\ v_1, \tilde{v}_1 \\ v_2, \tilde{v}_2 \end{pmatrix} = f\begin{pmatrix} u, \tilde{u} \\ v, \tilde{v} \end{pmatrix},$$

where we set

$$[u, \tilde{u}] = \begin{cases} [u_1, \tilde{u}_1] \cap [u_2, \tilde{u}_2], & \textit{if } [u_1, \tilde{u}_1] \cap [u_2, \tilde{u}_2] \neq \varnothing, \\ [0, 0], & \textit{if } [u_1, \tilde{u}_1] \cap [u_2, \tilde{u}_2] = \varnothing, \end{cases}$$

$$[v, \tilde{v}] = \begin{cases} [v_1, \tilde{v}_1] \cap [v_2, \tilde{v}_2], & \textit{if } [v_1, \tilde{v}_1] \cap [v_2, \tilde{v}_2] \neq \varnothing, \\ [0, 0], & \textit{if } [v_1, \tilde{v}_1] \cap [v_2, \tilde{v}_2] = \varnothing. \end{cases}$$

(2) *For $s < t$, $\sigma < t$ and $p \geq 1$ we have*

$$V_p\big(f, [s,t] \times [\sigma, \tau]\big) = V_p\big(\bar{f}, [s,t]^2 \times [\sigma, \tau]^2\big).$$

Proof.

(1) By definition of the higher dimensional increments,

$$\bar{f}\begin{pmatrix} u_1, \tilde{u}_1 \\ u_2, \tilde{u}_2 \\ v_1 \\ v_2 \end{pmatrix} = \bar{f}\begin{pmatrix} \tilde{u}_1 \\ \tilde{u}_2 \\ v_1 \\ v_2 \end{pmatrix} - \bar{f}\begin{pmatrix} \tilde{u}_1 \\ u_2 \\ v_1 \\ v_2 \end{pmatrix} - \bar{f}\begin{pmatrix} u_1 \\ \tilde{u}_2 \\ v_1 \\ v_2 \end{pmatrix} + \bar{f}\begin{pmatrix} u_1 \\ u_2 \\ v_1 \\ v_2 \end{pmatrix}$$
$$= f(\tilde{u}_1 \wedge \tilde{u}_2, v_1 \wedge v_2) - f(\tilde{u}_1 \wedge u_2, v_1 \wedge v_2)$$
$$- f(u_1 \wedge \tilde{u}_2, v_1 \wedge v_2) + f(u_1 \wedge u_2, v_1 \wedge v_2).$$

By a case distinction, one sees that this is equal to $f(\tilde{u}, v_1 \wedge v_2) - f(u, v_1 \wedge v_2)$. One goes on with

$$\bar{f}\begin{pmatrix} u_1, \tilde{u}_1 \\ u_2, \tilde{u}_2 \\ v_1, \tilde{v}_1 \\ v_2, \tilde{v}_2 \end{pmatrix} = \bar{f}\begin{pmatrix} u_1, \tilde{u}_1 \\ u_2, \tilde{u}_2 \\ \tilde{v}_1 \\ \tilde{v}_2 \end{pmatrix} - \bar{f}\begin{pmatrix} u_1, \tilde{u}_1 \\ u_2, \tilde{u}_2 \\ \tilde{v}_1 \\ v_2 \end{pmatrix} - \bar{f}\begin{pmatrix} u_1, \tilde{u}_1 \\ u_2, \tilde{u}_2 \\ v_1 \\ \tilde{v}_2 \end{pmatrix} + \bar{f}\begin{pmatrix} u_1, \tilde{u}_1 \\ u_2, \tilde{u}_2 \\ v_1 \\ v_2 \end{pmatrix}$$
$$= h(\tilde{v}_1 \wedge \tilde{v}_2) - h(\tilde{v}_1 \wedge v_2) - h(v_1 \wedge \tilde{v}_2) + h(v_1 \wedge v_2)$$
$$= h(\tilde{v}) - h(v),$$

where $h(\cdot) = f(\tilde{u}, \cdot) - f(u, \cdot)$. Hence

$$h(\tilde{v}) - h(v) = f(\tilde{u}, \tilde{v}) - f(u, \tilde{v}) - f(\tilde{u}, v) + f(u, v) = f\begin{pmatrix} u, \tilde{u} \\ v, \tilde{v} \end{pmatrix}.$$

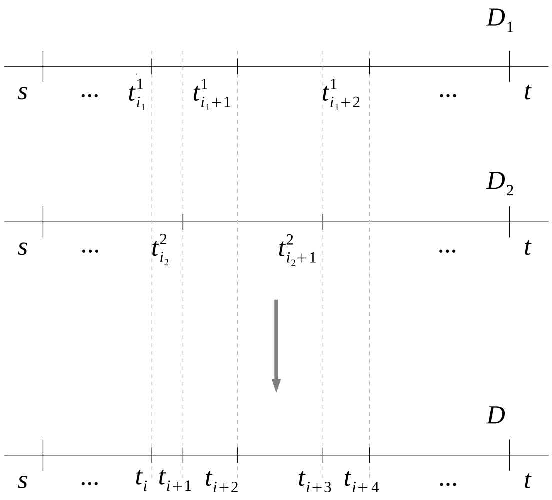

Fig. 1. The union of 2 partitions

(2) Let D be a partition of $[s,t]$ and $\tilde{D}$ a partition of $[\sigma,\tau]$. Then by 1,

$$\sum_{t_i\in D,\tilde{t}\in\tilde{D}}\left|f\begin{pmatrix}t_i,t_{i+1}\\ \tilde{t}_j,\tilde{t}_{j+1}\end{pmatrix}\right|^p=\sum_{t_i\in D,\tilde{t}\in\tilde{D}}\left|\bar{f}\begin{pmatrix}t_i,t_{i+1}\\ t_i,t_{i+1}\\ \tilde{t}_j,\tilde{t}_{j+1}\\ \tilde{t}_j,\tilde{t}_{j+1}\end{pmatrix}\right|^p\le\left(V_p\left(\bar{f},[s,t]^2\times[\sigma,\tau]^2\right)\right)^p,$$

hence $V_p(f,[s,t]\times[\sigma,\tau])\le V_p(\bar{f},[s,t]^2\times[\sigma,\tau]^2)$. Now let D_1, D_2 be partitions of $[s,t]$ and $\tilde{D}_1$, $\tilde{D}_2$ be partitions of $[\sigma,\tau]$. Set $D=D_1\cup D_2$, $\tilde{D}=\tilde{D}_1\cup\tilde{D}_2$. Then D is a partition of $[s,t]$ and $\tilde{D}$ a partition of $[\sigma,\tau]$ (see Fig. 1).

By (1),

$$\sum_{\substack{t^1_{i_1}\in D_1,t^2_{i_2}\in D_2\\ \tilde{t}^1_{j_1}\in\tilde{D}_1,\tilde{t}^2_{j_2}\in\tilde{D}_2}}\left|f\begin{pmatrix}t^1_{i_1},t^1_{i_1+1}\\ t^2_{i_2},t^2_{i_2+1}\\ \tilde{t}^1_{j_1},\tilde{t}^1_{j_1+1}\\ \tilde{t}^2_{j_2},\tilde{t}^2_{j_2+1}\end{pmatrix}\right|^p=\sum_{t_i\in D,\tilde{t}\in\tilde{D}}\left|f\begin{pmatrix}t_i,t_{i+1}\\ \tilde{t}_j,\tilde{t}_{j+1}\end{pmatrix}\right|^p\le\left(V_p\left(f,[s,t]\times[\sigma,\tau]\right)\right)^p$$

and we also get $V_p(\bar{f},[s,t]^2\times[\sigma,\tau]^2)\le V_p(f,[s,t]\times[\sigma,\tau])$.

□

Lemma 12. *Let $(X,Y)\colon[0,1]\to\mathbb{R}^2$ be a centred Gaussian process with continuous paths of finite variation and assume that ω is a symmetric control which controls the ρ-variation of $R_{(X,Y)}$ where $\rho\ge1$. Take $(s,t)\in\Delta$, $\gamma>\rho$ and set $\varepsilon^2=V_\infty(R_{X-Y},[s,t]^2)^{1-\rho/\gamma}$.*

(1) *Set $f(u_1,u_2,v_1,v_2)=E[X_{u_1}X_{u_2}X_{v_1}X_{v_2}]$. Then there is a constant $C_1=C_1(\rho)$ and a symmetric 4D grid-control $\tilde{\omega}_1$ which controls the ρ-variation of f and*

$$V_\rho\left(f,[s,t]^4\right)\le\tilde{\omega}_1\left([s,t]^4\right)^{1/\rho}=C_1\omega\left([s,t]^2\right)^{2/\rho}.$$

(2) *Set $\tilde{f}(u_1,u_2,v_1,v_2)=E[\mathbf{X}^{(2)}_{s,u_1\wedge u_2}\mathbf{X}^{(2)}_{s,v_1\wedge v_2}]$. Then there is a constant $C_2=C_2(\rho)$ such that*

$$V_\rho\left(\tilde{f},[s,t]^4\right)\le C_2\omega\left([s,t]^2\right)^{2/\rho}.$$

(3) *Set*

$$g(u_1,u_2,v_1,v_2)=E\big[(X_{u_1}X_{u_2}-Y_{u_1}Y_{u_2})(X_{v_1}X_{v_2}-Y_{v_1}Y_{v_2})\big].$$

Then there is a constant $C_3=C_3(\rho,\gamma)$ *and a symmetric* $4D$ *grid-control* $\tilde{\omega}_2$ *which controls the* γ*-variation of* g *and*

$$V_\gamma\big(g,[s,t]^4\big)\le\tilde{\omega}_2\big([s,t]^4\big)^{1/\gamma}=C_3\varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma+1/\rho}.$$

(4) *Set*

$$\tilde{g}(u_1,u_2,v_1,v_2)=E\big[\big(\mathbf{X}^{(2)}-\mathbf{Y}^{(2)}\big)_{s,u_1\wedge u_2}\big(\mathbf{X}^{(2)}-\mathbf{Y}^{(2)}\big)_{s,v_1\wedge v_2}\big].$$

Then there is a constant $C_4=C_4(\rho,\gamma)$ *such that*

$$V_\gamma\big(\tilde{g},[s,t]^4\big)\le C_4\varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma+1/\rho}.$$

Proof.

(1) Let $u_1<\tilde{u}_1$, $u_2<\tilde{u}_2$, $v_1<\tilde{v}_1$, $v_2<\tilde{v}_2$. By the Wick-formula,

$$\begin{aligned}
&\big|E[X_{u_1,\tilde{u}_1}X_{u_2,\tilde{u}_2}X_{v_1,\tilde{v}_1}X_{v_2,\tilde{v}_2}]\big|^\rho\\
&\quad\le 3^{\rho-1}\big|E[X_{u_1,\tilde{u}_1}X_{u_2,\tilde{u}_2}]E[X_{v_1,\tilde{v}_1}X_{v_2,\tilde{v}_2}]\big|^\rho+3^{\rho-1}\big|E[X_{u_1,\tilde{u}_1}X_{v_1,\tilde{v}_1}]E[X_{u_2,\tilde{u}_2}X_{v_2,\tilde{v}_2}]\big|^\rho\\
&\qquad+3^{\rho-1}\big|E[X_{u_1,\tilde{u}_1}X_{v_2,\tilde{v}_2}]E[X_{u_2,\tilde{u}_2}X_{v_1,\tilde{v}_1}]\big|^\rho\\
&\quad\le 3^{\rho-1}\omega\big([u_1,\tilde{u}_1]\times[u_2,\tilde{u}_2]\big)\omega\big([v_1,\tilde{v}_1]\times[v_2,\tilde{v}_2]\big)\\
&\qquad+3^{\rho-1}\omega\big([u_1,\tilde{u}_1]\times[v_1,\tilde{v}_1]\big)\omega\big([u_2,\tilde{u}_2]\times[v_2,\tilde{v}_2]\big)\\
&\qquad+3^{\rho-1}\omega\big([u_1,\tilde{u}_1]\times[v_2,\tilde{v}_2]\big)\omega\big([u_2,\tilde{u}_2]\times[v_1,\tilde{v}_1]\big)\\
&\quad=:\tilde{\omega}_1\big([u_1,\tilde{u}_1]\times[u_2,\tilde{u}_2]\times[v_1,\tilde{v}_1]\times[v_2,\tilde{v}_2]\big).
\end{aligned}$$

It is easy to see that $\tilde{\omega}_1$ is a symmetric grid-control and that it fulfils the stated property.

(2) A direct consequence of Lemma 4 and Lemma 11.

(3) We have

$$X_{u_1}X_{u_2}-Y_{u_1}Y_{u_2}=(X_{u_1}-Y_{u_1})X_{u_2}+Y_{u_1}(X_{u_2}-Y_{u_2}).$$

Hence for $u_1<\tilde{u}_1,u_2<\tilde{u}_2,v_1<\tilde{v}_1,v_2<\tilde{v}_2$,

$$\begin{aligned}
\tilde{f}\begin{pmatrix}u_1,\tilde{u}_1\\u_2,\tilde{u}_2\\v_1,\tilde{v}_1\\v_2,\tilde{v}_2\end{pmatrix}&=E\big[(X-Y)_{u_1,\tilde{u}_1}X_{u_2,\tilde{u}_2}(X-Y)_{v_1,\tilde{v}_1}X_{v_2,\tilde{v}_2}\big]\\
&\quad+E\big[Y_{u_1,\tilde{u}_1}(X-Y)_{u_2,\tilde{u}_2}(X-Y)_{v_1,\tilde{v}_1}X_{v_2,\tilde{v}_2}\big]\\
&\quad+E\big[(X-Y)_{u_1,\tilde{u}_1}X_{u_2,\tilde{u}_2}Y_{v_1,\tilde{v}_1}(X-Y)_{v_2,\tilde{v}_2}\big]\\
&\quad+E\big[Y_{u_1,\tilde{u}_1}(X-Y)_{u_2,\tilde{u}_2}Y_{v_1,\tilde{v}_1}(X-Y)_{v_2,\tilde{v}_2}\big].
\end{aligned}$$

For the first term we have, using Lemma 7,

$$\begin{aligned}
&\big|E\big[(X-Y)_{u_1,\tilde{u}_1}X_{u_2,\tilde{u}_2}(X-Y)_{v_1,\tilde{v}_1}X_{v_2,\tilde{v}_2}\big]\big|^\gamma\\
&\quad\le 3^{\gamma-1}\big|E\big[(X-Y)_{u_1,\tilde{u}_1}X_{u_2,\tilde{u}_2}\big]\big|^\gamma\big|E\big[(X-Y)_{v_1,\tilde{v}_1}X_{v_2,\tilde{v}_2}\big]\big|^\gamma
\end{aligned}$$

$$
\begin{aligned}
&+3^{\gamma-1}\big|E\big[(X-Y)_{u_1,\tilde u_1}(X-Y)_{v_1,\tilde v_1}\big]\big|^{\gamma}\big|E[X_{u_2,\tilde u_2}X_{v_2,\tilde v_2}]\big|^{\gamma}\\
&+3^{\gamma-1}\big|E\big[(X-Y)_{u_1,\tilde u_1}X_{v_2,\tilde v_2}\big]\big|^{\gamma}\big|E\big[X_{u_2,\tilde u_2}(X-Y)_{v_1,\tilde v_1}\big]\big|^{\gamma}\\
\le{}& 3^{\gamma-1}\varepsilon^{2\gamma}\omega\big([s,t]^2\big)^{\gamma/\rho-1}\omega\big([u_1,\tilde u_1]\times[u_2,\tilde u_2]\big)\omega\big([v_1,\tilde v_1]\times[v_2,\tilde v_2]\big)\\
&+3^{\gamma-1}\varepsilon^{2\gamma}\omega\big([u_1,\tilde u_1]\times[v_1,\tilde v_1]\big)\omega\big([u_2,\tilde u_2]\times[v_2,\tilde v_2]\big)^{\gamma/\rho}\\
&+3^{\gamma-1}\varepsilon^{2\gamma}\omega\big([s,t]^2\big)^{\gamma/\rho-1}\omega\big([u_1,\tilde u_1]\times[v_2,\tilde v_2]\big)\omega\big([u_2,\tilde u_2]\times[v_1,\tilde v_1]\big)\\
\le{}& 3^{\gamma-1}\varepsilon^{2\gamma}\omega\big([s,t]^2\big)^{\gamma/\rho-1}\big(\omega\big([u_1,\tilde u_1]\times[u_2,\tilde u_2]\big)\omega\big([v_1,\tilde v_1]\times[v_2,\tilde v_2]\big)\\
&+\omega\big([u_1,\tilde u_1]\times[v_1,\tilde v_1]\big)\omega\big([u_2,\tilde u_2]\times[v_2,\tilde v_2]\big)\\
&+\omega\big([u_1,\tilde u_1]\times[v_2,\tilde v_2]\big)\omega\big([u_2,\tilde u_2]\times[v_1,\tilde v_1]\big)\big)\\
=:{}& \tilde\omega\big([u_1,\tilde u_1]\times[u_2,\tilde u_2]\times[v_1,\tilde v_1]\times[v_2,\tilde v_2]\big).
\end{aligned}
$$

$\tilde\omega$ is a symmetric grid-control and fulfils the stated property. The other terms are treated in the same way.

(4) Follows from Lemma 8 and Lemma 11. □

Corollary 4. *Let (X,Y), ω, ρ and γ as in Lemma 6. Then there is a constant $C=C(\rho,\gamma)$ such that*

$$\big|\mathbf{X}^{i,i,j,j}_{s,t}-\mathbf{Y}^{i,i,j,j}_{s,t}\big|_{L^2}\le C\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{3/(2\rho)}$$

holds for every $(s,t)\in\Delta$ and $i\neq j$ where $\varepsilon^2=V_\infty(R_{X-Y},[s,t]^2)^{1-\rho/\gamma}$.

Proof. As seen before, we can use Fubini to obtain

$$\mathbf{X}^{i,i,j,j}_{s,t}=\int_{\Delta^2_{s,t}}\mathbf{X}^{i,i}_{s,u_1}\,\mathrm{d}X^j_{u_1}\,\mathrm{d}X^j_{u_2}=\frac{1}{2}\int_{[s,t]^2}\mathbf{X}^{i,i}_{s,u_1\wedge u_2}\,\mathrm{d}\big(X^j_{u_1}X^j_{u_2}\big)$$

and hence

$$
\begin{aligned}
\big|\mathbf{X}^{i,i,j,j}_{s,t}-\mathbf{Y}^{i,i,j,j}_{s,t}\big|_{L^2}\le{}&\frac{1}{2}\bigg|\int_{[s,t]^2}\big(\mathbf{X}^{i,i}_{s,u_1\wedge u_2}-\mathbf{Y}^{i,i}_{s,u_1\wedge u_2}\big)\,\mathrm{d}\big(X^j_{u_1}X^j_{u_2}\big)\bigg|_{L^2}\\
&+\frac{1}{2}\bigg|\int_{[s,t]^2}\mathbf{Y}^{i,i}_{s,u_1\wedge u_2}\,\mathrm{d}\big(X^j_{u_1}X^j_{u_2}-Y^j_{u_1}Y^j_{u_2}\big)\bigg|_{L^2}.
\end{aligned}
$$

We use a Young $4D$-estimate and the estimates of Lemma 12 to see that

$$
\begin{aligned}
&\bigg|\int_{[s,t]^2}\big(\mathbf{X}^{i,i}_{s,u_1\wedge u_2}-\mathbf{Y}^{i,i}_{s,u_1\wedge u_2}\big)\,\mathrm{d}\big(X^j_{u_1}X^j_{u_2}\big)\bigg|^2_{L^2}\\
&\quad=\int_{[s,t]^4}E\big[\big(\mathbf{X}^{i,i}_{s,u_1\wedge u_2}-\mathbf{Y}^{i,i}_{s,u_1\wedge u_2}\big)\big(\mathbf{X}^{i,i}_{s,v_1\wedge v_2}-\mathbf{Y}^{i,i}_{s,v_1\wedge v_2}\big)\big]\,\mathrm{d}E\big[X^j_{u_1}X^j_{u_2}X^j_{v_1}X^j_{v_2}\big]\\
&\quad\le c_1\varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma}\omega\big([s,t]^2\big)^{3/\rho}.
\end{aligned}
$$

The second term is estimated in the same way using again Lemma 12. □

Lemma 13. *Let $f:[0,1]^2\to\mathbb{R}$ and $g:[0,1]^2\times[0,1]^2\to\mathbb{R}$ be continuous where g is symmetric in the first and the last two variables. Let $(s,t)\in\Delta$ and assume that $f(s,\cdot)=f(\cdot,s)=0$. Assume also that f has finite p-variation and that the q-variation of g is controlled by a symmetric $4D$ grid-control $\tilde\omega$ where $\frac{1}{p}+\frac{1}{q}>1$. Define*

$$\Psi(u,v)=\int_{[s,u]^2\times[s,v]^2}f(u_1\wedge u_2,v_1\wedge v_2)\,\mathrm{d}g(u_1,u_2;v_1,v_2).$$

Then there is a constant $C = C(p,q)$ such that

$$V_q\big(\Psi;[s,t]^2\big) \le C V_p\big(f;[s,t]^2\big)\tilde{\omega}\big([s,t]^4\big)^{1/q}.$$

Proof. Set

$$\tilde{f}(u_1,u_2,v_1,v_2) = f(u_1 \wedge u_2, v_1 \wedge v_2).$$

Let $u < v$ and $u' < v'$. Note that

$$\begin{aligned}
&1_{[s,v]^2\times[s,v']^2} - 1_{[s,u]^2\times[s,v']^2} - 1_{[s,v]^2\times[s,u']^2} + 1_{[s,u]^2\times[s,u']^2} \\
&\quad = 1_{([s,v]^2\setminus[s,u]^2)\times[s,v']^2} - 1_{([s,v]^2\setminus[s,u]^2)\times[s,u']^2} \\
&\quad = 1_{([s,v]^2\setminus[s,u]^2)\times([s,v']^2\setminus[s,u']^2)}.
\end{aligned}$$

If we take out the square $[s,u]^2$ of the larger square $[s,v]^2$, what is left is the union of three essentially disjoint squares. More precisely,

$$\overline{[s,v]^2 \setminus [s,u]^2} = [u,v]^2 \cup \big([s,u]\times[u,v]\big) \cup \big([u,v]\times[s,u]\big).$$

The same holds for u' and v'. Hence,

$$\begin{aligned}
&\big(\overline{[s,v]^2 \setminus [s,u]^2}\big) \times \big(\overline{[s,v']^2 \setminus [s,u']^2}\big) \\
&\quad = \big([u,v]^2 \cup \big([s,u]\times[u,v]\big) \cup \big([u,v]\times[s,u]\big)\big) \\
&\qquad \times \big([u',v']^2 \cup \big([s,u']\times[u',v']\big) \cup \big([u',v']\times[s,u']\big)\big) \\
&\quad = \big([u,v]^2\times[u',v']^2\big) \cup \big([u,v]^2\times[s,u']\times[u',v']\big) \cup \big([u,v]^2\times[u',v']\times[s,u']\big) \\
&\qquad \cup \big([s,u]\times[u,v]\times[u',v']^2\big) \cup \big([s,u]\times[u,v]\times[s,u']\times[u',v']\big) \\
&\qquad \cup \big([s,u]\times[u,v]\times[u',v']\times[s,u']\big) \\
&\qquad \cup \big([u,v]\times[s,u]\times[u',v']^2\big) \cup \big([u,v]\times[s,u]\times[s,u']\times[u',v']\big) \\
&\qquad \cup \big([u,v]\times[s,u]\times[u',v']\times[s,u']\big)
\end{aligned}$$

and all these are unions of essentially disjoint sets. Using continuity and the symmetry of $\tilde{f}$ and g we have then

$$\begin{aligned}
\Psi\begin{pmatrix} u,v \\ u',v' \end{pmatrix} &= \int_{([s,v]^2\setminus[s,u]^2)\times([s,v']^2\setminus[s,u']^2)} \tilde{f}\,\mathrm{d}g \\
&= \int_{[u,v]^2\times[u',v']^2} \tilde{f}\,\mathrm{d}g + 2\int_{[u,v]^2\times[s,u']\times[u',v']} \tilde{f}\,\mathrm{d}g \\
&\quad + 2\int_{[s,u]\times[u,v]\times[u',v']^2} \tilde{f}\,\mathrm{d}g + 4\int_{[s,u]\times[u,v]\times[s,u']\times[u',v']} \tilde{f}\,\mathrm{d}g.
\end{aligned}$$

For the first integral we use Young $4D$-estimates. Since $\tilde{f}(s,\cdot,\cdot,\cdot) = \cdots = \tilde{f}(\cdot,\cdot,\cdot,s) = 0$, we can proceed as in the proof of Lemma 2 and use Lemma 11 to see that

$$\begin{aligned}
\left|\int_{[u,v]^2\times[u',v']^2} \tilde{f}\,\mathrm{d}g\right| &\le c_1 V_p\big(f,[s,t]^2\big)V_q\big(g,[u,v]^2\times[u',v']^2\big) \\
&\le c_1 V_p\big(f,[s,t]^2\big)\tilde{\omega}\big([u,v]^2\times[u',v']^2\big)^{1/q}.
\end{aligned}$$

For the second integral, we have

$$\begin{aligned}&\int_{[u,v]^2\times[s,u']\times[u',v']} \tilde{f}\,\mathrm{d}g\\&\quad=\int_{[u,v]^2\times[s,u']\times[u',v']} f(u_1\wedge u_2, v_1\wedge v_2)\,\mathrm{d}g(u_1,u_2;v_1,v_2)\\&\quad=\int_{[u,v]^2\times[s,u']} f(u_1\wedge u_2, v_1)\,\mathrm{d}\big[g\big(u_1,u_2;v_1,v'\big)-g\big(u_1,u_2;v_1,u'\big)\big].\end{aligned}$$

We now use a Young $3D$-estimate to see that

$$\begin{aligned}\left|\int_{[u,v]^2\times[s,u']\times[u',v']} \tilde{f}\,\mathrm{d}g\right| &\le c_2 V_p\big(f(\cdot\wedge\cdot,\cdot),[s,t]^3\big)\\&\quad\times V_q\big(g\big(\cdot,\cdot;\cdot,v'\big)-g\big(\cdot,\cdot;\cdot,u'\big),[u,v]^2\times\big[s,u'\big]\big).\end{aligned}$$

As in Lemma 11, one can show that $V_p(f(\cdot\wedge\cdot,\cdot),[s,t]^3)=V_p(f,[s,t]^2)$. For g, we have

$$\begin{aligned}V_q\big(g\big(\cdot,\cdot;\cdot,v'\big)-g\big(\cdot,\cdot;\cdot,u'\big),[u,v]^2\times\big[s,u'\big]\big) &\le V_q\big(g,[u,v]^2\times\big[s,u'\big]\times\big[u',v'\big]\big)\\&\le\tilde{\omega}\big([u,v]^2\times[s,t]\times\big[u',v'\big]\big)^{1/q}.\end{aligned}$$

Hence

$$\left|\int_{[u,v]^2\times[s,u']\times[u',v']} \tilde{f}\,\mathrm{d}g\right| \le c_2 V_p\big(f,[s,t]^2\big)\tilde{\omega}\big([u,v]^2\times[s,t]\times\big[u',v'\big]\big)^{1/q}.$$

Similarly, using Young $3D$ and $2D$ estimates, we get

$$\left|\int_{[s,u]\times[u,v]\times[u',v']^2} \tilde{f}\,\mathrm{d}g\right| \le c_3 V_p\big(f,[s,t]^2\big)\tilde{\omega}\big([s,t]\times[u,v]\times\big[u',v'\big]^2\big)^{1/q}$$

and

$$\left|\int_{[s,u]\times[u,v]\times[s,u']\times[u',v']} \tilde{f}\,\mathrm{d}g\right| \le c_4 V_p\big(f,[s,t]^2\big)\tilde{\omega}\big([s,t]\times[u,v]\times[s,t]\times\big[u',v'\big]\big)^{1/q}.$$

Putting all together, using the symmetry of $\tilde{\omega}$ we have shown that

$$\left|\Psi\begin{pmatrix}u,v\\u',v'\end{pmatrix}\right|^q \le c_5 V_p\big(f,[s,t]^2\big)^q\tilde{\omega}\big([u,v]\times\big[u',v'\big]\times[s,t]^2\big).$$

Since $\tilde{\omega}_2([u,v]\times[u',v']):=\tilde{\omega}([u,v]\times[u',v']\times[s,t]^2)$ is a $2D$ grid-control this shows the claim. □

We are now able to prove the remaining estimate.

Corollary 5. *Let (X,Y), ω, ρ and γ as in Lemma 6. Then there is a constant $C=C(\rho,\gamma)$ such that*

$$\big|\mathbf{X}^{j,i,i,k}_{s,t}-\mathbf{Y}^{j,i,i,k}_{s,t}\big|_{L^2}\le C\varepsilon\omega\big([s,t]^2\big)^{1/(2\gamma)}\omega\big([s,t]^2\big)^{3/(2\rho)}$$

holds for every $(s,t)\in\Delta$ and i,j,k pairwise distinct where $\varepsilon^2=V_\infty(R_{X-Y},[s,t]^2)^{1-\rho/\gamma}$.

Proof. From

$$\int_{\Delta^2_{s,w}} X^j_{s,u_1}\,\mathrm{d}X^i_{u_1}\,\mathrm{d}X^i_{u_2}=\frac{1}{2}\int_{[s,w]^2} X^j_{s,u_1\wedge u_2}\,\mathrm{d}\big(X^i_{u_1}X^i_{u_2}\big)$$

we see that

$$\mathbf{X}_{s,t}^{j,i,i,k} = \frac{1}{2} \int_s^t \left(\int_{[s,w]^2} X_{s,u_1 \wedge u_2}^j \, \mathrm{d}\big(X_{u_1}^i X_{u_2}^i\big) \right) \mathrm{d}X_w^k.$$

Hence

$$\begin{aligned} &\big|\mathbf{X}_{s,t}^{j,i,i,k} - \mathbf{Y}_{s,t}^{j,i,i,k}\big|_{L^2} \\ &\quad \le \frac{1}{2} \left| \int_s^t \Psi_1(w) \, \mathrm{d}X_w^k \right|_{L^2} + \frac{1}{2} \left| \int_s^t \Psi_2(w) \, \mathrm{d}X_w^k \right|_{L^2} + \frac{1}{2} \left| \int_s^t \Psi_3(w) \, \mathrm{d}\big(X^k - Y^k\big)_w \right|_{L^2}, \end{aligned}$$

where

$$\begin{aligned} \Psi_1(w) &= \int_{[s,w]^2} \big(X_{s,u_1 \wedge u_2}^j - Y_{s,u_1 \wedge u_2}^j\big) \, \mathrm{d}\big(X_{u_1}^i X_{u_2}^i\big), \\ \Psi_2(w) &= \int_{[s,w]^2} Y_{s,u_1 \wedge u_2}^j \, \mathrm{d}\big(X_{u_1}^i X_{u_2}^i - Y_{u_1}^i Y_{u_2}^i\big), \\ \Psi_3(w) &= \int_{[s,w]^2} Y_{s,u_1 \wedge u_2}^j \, \mathrm{d}\big(Y_{u_1}^i Y_{u_2}^i\big). \end{aligned}$$

We start with the first integral. From independence and Young $2D$-estimates,

$$\begin{aligned} \left| \int_s^t \Psi_1(w) \, \mathrm{d}X_w^k \right|_{L^2}^2 &= \int_{[s,t]^2} E\big[\Psi_1(w_1)\Psi_1(w_2)\big] \, \mathrm{d}E\big[X_{w_1}^k X_{w_2}^k\big] \\ &\le c_1 V_\rho\big(E\big[\Psi_1(\cdot)\Psi_1(\cdot)\big], [s,t]^2\big) V_\rho\big(R_{X^k}[s,t]^2\big). \end{aligned}$$

Now,

$$\begin{aligned} &E\big[\Psi_1(w_1)\Psi_1(w_2)\big] \\ &\quad = \int_{[s,w_1]^2 \times [s,w_2]^2} E\big[\big(X_{s,u_1 \wedge u_2}^j - Y_{s,u_1 \wedge u_2}^j\big)\big(X_{s,v_1 \wedge v_2}^j - Y_{s,v_1 \wedge v_2}^j\big)\big] \, \mathrm{d}E\big[X_{u_1}^i X_{u_2}^i X_{v_1}^i X_{v_2}^i\big]. \end{aligned}$$

In Lemma 12 we have seen that the ρ-variation of $E[X_\cdot^i X_\cdot^i X_\cdot^i X_\cdot^i]$ is controlled by a symmetric grid-control $\tilde{\omega}_1$. Hence we can apply Lemma 13 to conclude that

$$\begin{aligned} V_\rho\big(E\big[\Psi_1(\cdot)\Psi_1(\cdot)\big], [s,t]^2\big) &\le c_2 V_\gamma\big(R_{X-Y}; [s,t]^2\big) \tilde{\omega}_1\big([s,t]^4\big)^{1/\rho} \\ &\le c_3 \varepsilon^2 \omega\big([s,t]^2\big)^{1/\gamma} \omega\big([s,t]^2\big)^{2/\rho}. \end{aligned}$$

Clearly, $V_\rho(R_{X^k}[s,t]^2) \le \omega([s,t]^2)^{1/\rho}$ and therefore

$$\left| \int_s^t \Psi_1(w) \, \mathrm{d}X_w^k \right|_{L^2}^2 \le c_4 \varepsilon^2 \omega\big([s,t]^2\big)^{1/\gamma} \omega\big([s,t]^2\big)^{3/\rho}.$$

Now we come to the second integral. From independence,

$$\begin{aligned} \left| \int_s^t \Psi_2(w) \, \mathrm{d}X_w^k \right|_{L^2}^2 &= \int_{[s,t]^2} E\big[\Psi_2(w_1)\Psi_2(w_2)\big] \, \mathrm{d}E\big[X_{w_1}^k X_{w_2}^k\big] \\ &\le c_5 V_\gamma\big(E\big[\Psi_2(\cdot)\Psi_2(\cdot)\big], [s,t]^2\big) V_\rho\big(R_{X^k}[s,t]^2\big). \end{aligned}$$

Now

$$
\begin{aligned}
&E\big[\Psi_2(w_1)\Psi_2(w_2)\big] \\
&\quad = \int_{[s,w_1]^2\times[s,w_2]^2} E\big[Y^j_{s,u_1\wedge u_2}Y^j_{s,v_1\wedge v_2}\big]\,\mathrm{d}E\big[\big(X^i_{u_1}X^i_{u_2}-Y^i_{u_1}Y^i_{u_2}\big)\big(X^i_{v_1}X^i_{v_2}-Y^i_{v_1}Y^i_{v_2}\big)\big] \\
&\quad =: \int_{[s,w_1]^2\times[s,w_2]^2} E\big[Y^j_{s,u_1\wedge u_2}Y^j_{s,v_1\wedge v_2}\big]\,\mathrm{d}g(u_1,u_2,v_1,v_2).
\end{aligned}
$$

In Lemma 12 we have seen that the $4D$ γ-variation of g is controlled by a symmetric $4D$ grid-control $\tilde{\omega}_2$ where

$$
\tilde{\omega}_2\big([s,t]^4\big)^{1/\gamma} = c_6\varepsilon^2\omega\big([s,t]^2\big)^{1/\rho+1/\gamma}.
$$

Hence

$$
V_\gamma\big(E\big[\Psi_2(\cdot)\Psi_2(\cdot)\big],[s,t]^2\big) \le c_7 V_\rho\big(R_{Y^j};[s,t]^2\big)\tilde{\omega}_2\big([s,t]^4\big)^{1/\gamma} \le c_8\varepsilon^2\omega\big([s,t]^2\big)^{2/\rho+1/\gamma}.
$$

This gives us

$$
\left|\int_s^t \Psi_2(w)\,\mathrm{d}X^k_w\right|^2_{L^2} \le c_9\varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma}\omega\big([s,t]^2\big)^{3/\rho}.
$$

For the third integral we see again that

$$
\begin{aligned}
\left|\int_s^t \Psi_3(w)\,\mathrm{d}\big(X^k-Y^k\big)_w\right|^2_{L^2} &= \int_{[s,t]^2} E\big[\Psi_3(w_1)\Psi_3(w_2)\big]\,\mathrm{d}E\big[\big(X^k-Y^k\big)_{w_1}\big(X^k-Y^k\big)_{w_2}\big] \\
&\le c_{10}V_\rho\big(E\big[\Psi_3(\cdot)\Psi_3(\cdot)\big],[s,t]^2\big)V_\gamma\big(R_{X-Y},[s,t]^2\big).
\end{aligned}
$$

From

$$
E\big[\Psi_3(w_1)\Psi_3(w_2)\big] = \int_{[s,w_1]^2\times[s,w_2]^2} E\big[Y^j_{s,u_1\wedge u_2}Y^j_{s,v_1\wedge v_2}\big]\,\mathrm{d}E\big[Y^i_{u_1}Y^i_{u_2}Y^i_{v_1}Y^i_{v_2}\big]
$$

we see that we can apply Lemma 13 to obtain

$$
V_\rho\big(E\big[\Psi_3(\cdot)\Psi_3(\cdot)\big],[s,t]^2\big) \le c_{11}V_\rho\big(R_{Y^j};[s,t]^2\big)\omega\big([s,t]^2\big)^{2/\rho} \le c_{11}\omega\big([s,t]^2\big)^{3/\rho}.
$$

Clearly, $V_\gamma(R_{X-Y},[s,t]^2) \le \varepsilon^2\omega([s,t]^2)^{1/\gamma}$ and hence

$$
\left|\int_s^t \Psi_3(w)\,\mathrm{d}\big(X^k-Y^k\big)_w\right|^2_{L^2} \le c_{12}\varepsilon^2\omega\big([s,t]^2\big)^{1/\gamma}\omega\big([s,t]^2\big)^{3/\rho}
$$

which gives the claim. □

Remark 2. *Even though Propositions 5, 6 and 7 are only formulated for Gaussian processes with sample paths of finite variation, the estimate (5.1) is valid also for general Gaussian rough paths for $n=1,2,3,4$. Indeed, this follows from the fact that Gaussian rough paths are just defined as L^2 limits of smooth paths, cf.* [6].

5.3. *Higher levels*

Once we have shown our desired estimates for the first four levels, we can use induction to obtain also the higher levels. This is done in the next proposition.

Proposition 8. *Let X and Y be Gaussian processes as in Theorem 1. Let ρ, γ be fixed and ω be a control. Assume that there are constants $\tilde{C} = \tilde{C}(n)$ such that*

$$|\mathbf{X}^n_{s,t}|_{L^2}, |\mathbf{Y}^n_{s,t}|_{L^2} \le \tilde{C}(n) \frac{\omega(s,t)^{n/(2\rho)}}{\beta(n/(2\rho))!}$$

holds for $n = 1, \dots, [2\rho]$ and constants $C = C(n)$ such that

$$|\mathbf{X}^n_{s,t} - \mathbf{Y}^n_{s,t}|_{L^2} \le C(n)\varepsilon\omega(s,t)^{1/(2\gamma)} \frac{\omega(s,t)^{(n-1)/(2\rho)}}{\beta((n-1)/(2\rho))!}$$

holds for $n = 1, \dots, [2\rho]+1$ and every $(s,t) \in \Delta$. Here, $\varepsilon > 0$ and β is a positive constant such that

$$\beta \ge 4\rho\left(1 + 2^{([2\rho]+1)/2\rho}\left(\zeta\left(\frac{[2\rho]+1}{2\rho}\right) - 1\right)\right),$$

where ζ is just the usual Riemann zeta function. Then for every $n \in \mathbb{N}$ there is a constant $C = C(n)$ such that

$$|\mathbf{X}^n_{s,t} - \mathbf{Y}^n_{s,t}|_{L^2} \le C\varepsilon\omega(s,t)^{1/(2\gamma)} \frac{\omega(s,t)^{(n-1)/(2\rho)}}{\beta((n-1)/(2\rho))!}$$

holds for every $(s,t) \in \Delta$.

Proof. From Proposition 1 we know that for every $n \in \mathbb{N}$ there are constants $\tilde{C}(n)$ such that

$$|\mathbf{X}^n_{s,t}|_{L^2}, |\mathbf{Y}^n_{s,t}|_{L^2} \le \tilde{C} \frac{\omega(s,t)^{n/(2\rho)}}{\beta(n/(2\rho))!}$$

holds for all $s < t$. We will proof the assertion by induction over n. The induction basis is fulfiled by assumption. Suppose that the statement is true for $k = 1, \dots, n$ where $n \ge [2\rho]+1$. We will show the statement for $n+1$. Let $D = \{s = t_0 < t_1 < \cdots < t_j = t\}$ be any partition of $[s,t]$. Set

$$\begin{aligned} \bar{\mathbf{X}}_{s,t} &:= (1, \mathbf{X}^1_{s,t}, \dots, \mathbf{X}^n_{s,t}, 0) \in T^{n+1}(\mathbb{R}^d), \\ \bar{\mathbf{X}}^D_{s,t} &:= \bar{\mathbf{X}}_{s,t_1} \otimes \cdots \otimes \bar{\mathbf{X}}_{t_{j-1},t} \end{aligned}$$

and the same for $\mathbf{Y}$. We know that $\lim_{|D|\to 0} \bar{\mathbf{X}}^D_{s,t} = S_{n+1}(\mathbf{X})_{s,t}$ a.s. and the same holds for $\mathbf{Y}$ (indeed, this is just the definition of the Lyons lift, cf. [14], Theorem 2.2.1). By multiplicativity, $\pi_k(\bar{\mathbf{X}}^D_{s,t}) = \mathbf{X}^k_{s,t}$ for $k \le n$. We will show that for any dissection D we have

$$|\pi_{n+1}(\bar{\mathbf{X}}^D_{s,t} - \bar{\mathbf{Y}}^D_{s,t})|_{L^2} \le C(n+1)\varepsilon\omega(s,t)^{1/(2\gamma)} \frac{\omega(s,t)^{n/(2\rho)}}{\beta(n/(2\rho))!}.$$

We use the notation $(\mathbf{X}^D)^k := \pi_k(\bar{\mathbf{X}}^D)$. Assume that $j \ge 2$. Let D' be the partition of $[s,t]$ obtained by removing a point t_i of the dissection D for which

$$\omega(t_{i-1}, t_{i+1}) \le \begin{cases} \frac{2\omega(s,t)}{j-1} & \text{for } j \ge 3, \\ \omega(s,t) & \text{for } j = 2 \end{cases}$$

holds (Lemma 2.2.1 in [14] shows that there is indeed such a point). By the triangle inequality,

$$|(\mathbf{X}^D - \mathbf{Y}^D)^{n+1}|_{L^2} \le |(\mathbf{X}^D - \mathbf{X}^{D'})^{n+1} - (\mathbf{Y}^D - \mathbf{Y}^{D'})^{n+1}|_{L^2} + |(\mathbf{X}^{D'} - \mathbf{Y}^{D'})^{n+1}|_{L^2}.$$

We estimate the first term on the right hand side. As seen in the proof of [14, Theorem 2.2.1], $(\mathbf{X}^D_{s,t} - \mathbf{X}^{D'}_{s,t})^{n+1} = \sum_{l=1}^n \mathbf{X}^l_{t_{i-1},t_i}\mathbf{X}^{n+1-l}_{t_i,t_{i+1}}$. Set $\mathbf{R}^l = \mathbf{Y}^l - \mathbf{X}^l$. Then

$$
\begin{aligned}
&\big(\mathbf{X}^D_{s,t} - \mathbf{X}^{D'}_{s,t}\big)^{n+1} - \big(\mathbf{Y}^D_{s,t} - \mathbf{Y}^{D'}_{s,t}\big)^{n+1} \\
&\quad = \sum_{l=1}^n \mathbf{X}^l_{t_{i-1},t_i}\mathbf{X}^{n+1-l}_{t_i,t_{i+1}} - \big(\mathbf{X}^l_{t_{i-1},t_i} + \mathbf{R}^l_{t_{i-1},t_i}\big)\big(\mathbf{X}^{n+1-l}_{t_i,t_{i+1}} + \mathbf{R}^{n+1-l}_{t_i,t_{i+1}}\big) \\
&\quad = \sum_{l=1}^n -\mathbf{X}^l_{t_{i-1},t_i}\mathbf{R}^{n+1-l}_{t_i,t_{i+1}} - \mathbf{R}^l_{t_{i-1},t_i}\mathbf{Y}^{n+1-l}_{t_i,t_{i+1}}.
\end{aligned}
$$

By the triangle inequality, equivalence of L^q-norms in the Wiener Chaos, our moment estimate for $\mathbf{X}^k$ and $\mathbf{Y}^k$ and the induction hypothesis,

$$
\begin{aligned}
&\big|\big(\mathbf{X}^D_{s,t} - \mathbf{X}^{D'}_{s,t}\big)^{n+1} - \big(\mathbf{Y}^D_{s,t} - \mathbf{Y}^{D'}_{s,t}\big)^{n+1}\big|_{L^2} \\
&\quad \le c_1(n+1)\sum_{l=1}^n \big|\mathbf{X}^l_{t_{i-1},t_i}\big|_{L^2}\big|\mathbf{R}^{n+1-l}_{t_i,t_{i+1}}\big|_{L^2} + \big|\mathbf{R}^l_{t_{i-1},t_i}\big|_{L^2}\big|\mathbf{Y}^{n+1-l}_{t_i,t_{i+1}}\big|_{L^2} \\
&\quad \le c_2(n+1)\sum_{l=1}^n \varepsilon\omega(t_i,t_{i+1})^{1/(2\gamma)}\frac{\omega(t_{i-1},t_i)^{l/(2\rho)}}{\beta(l/(2\rho))!}\frac{\omega(t_i,t_{i+1})^{(n-l)/(2\rho)}}{\beta((n-l)/(2\rho))!} \\
&\qquad + \varepsilon\omega(t_{i-1},t_i)^{1/(2\gamma)}\frac{\omega(t_{i-1},t_i)^{(l-1)/(2\rho)}}{\beta((l-1)/(2\rho))!}\frac{\omega(t_i,t_{i+1})^{(n+1-l)/(2\rho)}}{\beta((n+1-l)/(2\rho))!} \\
&\quad \le 2c_2\varepsilon\omega(s,t)^{1/(2\gamma)}\sum_{l=0}^n \frac{\omega(t_{i-1},t_i)^{l/(2\rho)}}{\beta(l/(2\rho))!}\frac{\omega(t_i,t_{i+1})^{(n-l)/(2\rho)}}{\beta(n-l/(2\rho))!} \\
&\quad = \frac{4\rho}{\beta^2}c_2\varepsilon\omega(s,t)^{1/(2\gamma)}\frac{1}{2\rho}\sum_{l=0}^n \frac{\omega(t_{i-1},t_i)^{l/(2\rho)}}{(l/(2\rho))!}\frac{\omega(t_i,t_{i+1})^{(n-l)/(2\rho)}}{((n-l)/(2\rho))!} \\
&\quad \le 4\rho c_2\varepsilon\omega(s,t)^{1/(2\gamma)}\frac{\omega(t_{i-1},t_{i+1})^{n/(2\rho)}}{\beta^2(n/(2\rho))!},
\end{aligned}
$$

where we used the neo-classical inequality (cf. [11]) and superadditivity of the control function. Hence for $j \ge 3$,

$$
\begin{aligned}
\big|\big(\mathbf{X}^D_{s,t} - \mathbf{X}^{D'}_{s,t}\big)^{n+1} - \big(\mathbf{Y}^D_{s,t} - \mathbf{Y}^{D'}_{s,t}\big)^{n+1}\big|_{L^2} &\le 4\rho c_2\varepsilon\omega(s,t)^{1/(2\gamma)}\frac{\omega(t_{i-1},t_{i+1})^{n/(2\rho)}}{\beta^2(n/(2\rho))!} \\
&\le \left(\frac{2}{j-1}\right)^{n/(2\rho)} 4\rho c_2\varepsilon\omega(s,t)^{1/(2\gamma)}\frac{\omega(s,t)^{n/(2\rho)}}{\beta^2(n/(2\rho))!}.
\end{aligned}
$$

For $j = 2$ we get

$$
\big|\big(\mathbf{X}^D_{s,t} - \mathbf{X}^{D'}_{s,t}\big)^{n+1} - \big(\mathbf{Y}^D_{s,t} - \mathbf{Y}^{D'}_{s,t}\big)^{n+1}\big|_{L^2} \le 4\rho c_2\varepsilon\omega(s,t)^{1/(2\gamma)}\frac{\omega(s,t)^{n/(2\rho)}}{\beta^2(n/(2\rho))!}
$$

but then $D' = \{s,t\}$ and therefore $|(\mathbf{X}^{D'}_{s,t} - \mathbf{Y}^{D'}_{s,t})^{n+1}|_{L^2} = 0$. Hence by successively dropping points we see that

$$
\big|\big(\mathbf{X}^D_{s,t} - \mathbf{Y}^D_{s,t}\big)^{n+1}\big|_{L^2} \le \left(1 + \sum_{j=3}^\infty \left(\frac{2}{j-1}\right)^{n/(2\rho)}\right) 4\rho c_2\varepsilon\omega(s,t)^{1/(2\gamma)}\frac{\omega(s,t)^{n/(2\rho)}}{\beta^2(n/(2\rho))!}
$$

holds for all partitions D. Since $n \geq [2\rho] + 1$,

$$\sum_{j=3}^{\infty} \left(\frac{2}{j-1} \right)^{n/(2\rho)} \leq \sum_{j=3}^{\infty} \left(\frac{2}{j-1} \right)^{([2\rho]+1)/(2\rho)} \leq 2^{([2\rho]+1)/(2\rho)} \left(\zeta \left(\frac{[2\rho]+1}{2\rho} \right) - 1 \right)$$

and thus

$$\left| \left(\mathbf{X}^D_{s,t} - \mathbf{Y}^D_{s,t} \right)^{n+1} \right|_{L^2} \leq \frac{4\rho(1 + 2^{([2\rho]+1)/(2\rho)}(\zeta(([2\rho]+1)/(2\rho)) - 1))}{\beta} c_2 \varepsilon \omega(s,t)^{1/(2\gamma)} \frac{\omega(s,t)^{n/(2\rho)}}{\beta(n/(2\rho))!}.$$

By the choice of β, we get the uniform bound

$$\left| \left(\mathbf{X}^D_{s,t} - \mathbf{Y}^D_{s,t} \right)^{n+1} \right|_{L^2} \leq c_2 \varepsilon \omega(s,t)^{1/(2\gamma)} \frac{\omega(s,t)^{n/(2\rho)}}{\beta(n/(2\rho))!}$$

which holds for all partitions D. Noting that a.s. convergence implies convergence in L^2 in the Wiener chaos, we obtain our claim by sending $|D| \to 0$. □

Corollary 6. *Let (X, Y), ω, ρ and γ as in Lemma 6. Then for all $n \in \mathbb{N}$ there are constants $C = C(\rho, \gamma, n)$ such that*

$$\left| \mathbf{X}^n_{s,t} - \mathbf{Y}^n_{s,t} \right|_{L^2} \leq C \varepsilon \omega\left([s,t]^2\right)^{1/(2\gamma)} \omega\left([s,t]^2\right)^{(n-1)/(2\rho)}$$

holds for every $(s,t) \in \Delta$ where $\varepsilon^2 = V_\infty(R_{X-Y}, [0,1]^2)^{1-\rho/\gamma}$.

Proof. For $n = 1, 2, 3, 4$ this is the content of Propositions 5, 6 and 7. By making the constants larger if necessary, we also get

$$\left| \mathbf{X}^n_{s,t} - \mathbf{Y}^n_{s,t} \right|_{L^2} \leq c(n) \varepsilon \omega\left([s,t]^2\right)^{1/(2\gamma)} \frac{\omega([s,t]^2)^{(n-1)/(2\rho)}}{\beta((n-1)/(2\rho))!}$$

with β chosen as in Proposition 8. We have already seen that

$$\left| \mathbf{X}^n_{s,t} \right|_{L^2}, \left| \mathbf{Y}^n_{s,t} \right|_{L^2} \leq \tilde{c}(n) \frac{\omega([s,t]^2)^{n/(2\rho)}}{\beta(n/(2\rho))!}$$

holds for constants $\tilde{c}(n)$ where $n = 1, 2, 3$. Since $\rho < 2$, we have $[2\rho] + 1 \leq 4$. From Proposition 8 we can conclude that

$$\left| \mathbf{X}^n_{s,t} - \mathbf{Y}^n_{s,t} \right|_{L^2} \leq c(n) \varepsilon \omega\left([s,t]^2\right)^{1/(2\gamma)} \frac{\omega([s,t]^2)^{(n-1)/(2\rho)}}{\beta((n-1)/(2\rho))!}$$

holds for every $n \in \mathbb{N}$ and constants $c(n)$. Setting $C(n) = \frac{c(n)}{\beta((n-1)/(2\rho))!}$ gives our claim. □

6. Main result

Assume that X is a Gaussian process as in Theorem 1 with paths of finite p-variation. Consider a sequence $(\Lambda_k)_{k \in \mathbb{N}}$ of continuous operators

$$\Lambda_k : C^{p\text{-}var}\left([0,1], \mathbb{R}\right) \to C^{1\text{-}var}\left([0,1], \mathbb{R}\right).$$

If $x = (x^1, \ldots, x^d) \in C^{p\text{-}var}([0,1], \mathbb{R}^d)$, we will write $\Lambda_k(x) = (\Lambda_k(x^1), \ldots, \Lambda_k(x^d))$. Assume that Λ_k fulfils the following conditions:

(1) $\Lambda_k(x) \to x$ in the $|\cdot|_\infty$-norm if $k \to \infty$ for every $x \in C^{p\text{-}var}([0,1], \mathbb{R}^d)$.

(2) If R_X has finite controlled ρ-variation, then, for some $C = C(\rho)$,

$$\sup_{k,l\in\mathbb{N}} |R_{(\Lambda_k(X),\Lambda_l(X))}|_{\rho\text{-}var;[0,1]^2} \le C|R_X|_{\rho\text{-}var;[0,1]^2}.$$

Our main result is the following:

Theorem 5. *Let X be a Gaussian process as in Theorem 1 for $\rho < 2$ and $K \ge V_\rho(R_X, [0,1]^2)$. Then there is an enhanced Gaussian process $\mathbf{X}$ with sample paths in $C^{0,p\text{-}var}([0,1], G^{[p]}(\mathbb{R}^d))$ w.r.t. $(\Lambda_k)_{k\in\mathbb{N}}$ where $p \in (2\rho, 4)$, i.e.*

$$\left|\rho_{p\text{-}var}\big(S_{[p]}\big(\Lambda_k(X)\big), \mathbf{X}\big)\right|_{L^r} \to 0$$

for $k \to \infty$ and every $r \ge 1$. Moreover, choose γ such that $\gamma > \rho$ and $\frac{1}{\gamma} + \frac{1}{\rho} > 1$. Then for $q > 2\gamma$ and every $N \in \mathbb{N}$ there is a constant $C = C(q, \rho, \gamma, K, N)$ such that

$$\left|\rho_{q\text{-}var}\big(S_N\big(\Lambda_k(X)\big), S_N(\mathbf{X})\big)\right|_{L^r} \le C r^{N/2} \sup_{0\le t\le 1} \left|\Lambda_k(X)_t - X_t\right|_{L^2(\mathbb{R}^d)}^{1-\rho/\gamma}$$

holds for every $k \in \mathbb{N}$.

Proof. The first statement is a fundamental result about Gaussian rough paths, see [7], Theorem 15.33. For the second, take $\delta > 0$ and set

$$\gamma' = (1+\delta)\gamma \quad \text{and} \quad \rho' = (1+\delta)\rho.$$

By choosing δ smaller if necessary we can assume that $\frac{1}{\rho'} + \frac{1}{\gamma'} > 1$ and $q > 2\gamma'$. Set

$$\omega_{k,l}(A) = |R_{(\Lambda_k(X),\Lambda_l(X))}|_{\rho'\text{-}var;A}^{\rho'}$$

for a rectangle $A \subset [0,1]^2$ and

$$\varepsilon_{k,l} = V_\infty\big(R_{(\Lambda_k(X)-\Lambda_l(X))}, [0,1]^2\big)^{1/2-\rho'/(2\gamma')} = V_\infty\big(R_{(\Lambda_k(X)-\Lambda_l(X))}, [0,1]^2\big)^{1/2-\rho/(2\gamma)}.$$

From Theorem 2 we know that $\omega_{k,l}$ is a $2D$ control function which controls the ρ'-variation of $R_{(\Lambda_k(X),\Lambda_l(X))}$. From Corollary 6 we can conclude that there is a constant c_1 such that

$$\left|\pi_n\big(S_N\big(\Lambda_k(X)\big)_{s,t} - S_N\big(\Lambda_l(X)\big)_{s,t}\big)\right|_{L^2} \le c_1 \varepsilon_{k,l} \omega_{k,l}\big([s,t]^2\big)^{1/(2\gamma')} \omega_{k,l}\big([s,t]^2\big)^{(n-1)/(2\rho')}$$

holds for every $n = 1, \dots, N$, $(s,t) \in \Delta$ and $k, l \in \mathbb{N}$. Now,

$$\begin{aligned}\omega_{k,l}\big([s,t]^2\big)^{(n-1)/(2\rho')} &= \left(\frac{\omega_{k,l}([s,t]^2)}{\omega_{k,l}([0,1]^2)}\right)^{(n-1)/(2\rho')} \omega_{k,l}\big([0,1]^2\big)^{(n-1)/(2\rho')} \\ &\le \omega_{k,l}\big([s,t]^2\big)^{(n-1)/(2\gamma')} \omega_{k,l}\big([0,1]^2\big)^{(n-1)/(2\rho')-(n-1)/(2\gamma')}.\end{aligned}$$

From Theorem 2 and our assumptions on the Λ_k we know that

$$\omega_{k,l}\big([0,1]^2\big)^{1/\rho'} \le c_2 |R_X|_{\rho'\text{-}var;[0,1]^2} \le c_3 V_\rho\big(R_X, [0,1]^2\big) \le c_4\big(\rho, \rho', K\big)$$

holds uniformly over all k, l. Hence

$$\left|\pi_n\big(S_N\big(\Lambda_k(X)\big)_{s,t} - S_N\big(\Lambda_l(X)\big)_{s,t}\big)\right|_{L^2} \le c_5 \varepsilon_{k,l} \omega_{k,l}\big([s,t]^2\big)^{n/(2\gamma')}.$$

Proposition 1 shows with the same argument that

$$\left|\pi_n\big(S_N\big(\Lambda_k(X)\big)_{s,t}\big)\right|_{L^2} \le c_6 \omega_{k,l}\big([s,t]^2\big)^{n/(2\rho')} \le c_7 \omega_{k,l}\big([s,t]^2\big)^{n/(2\gamma')}$$

for every $k \in \mathbb{N}$ and the same holds for $S_N(\Lambda_l(X))_{s,t}$. From [7], Proposition 15.24 we can conclude that there is a constant c_8 such that

$$\left|\rho_{q\text{-}var}\left(S_N\left(\Lambda_k(X)\right), S_N\left(\Lambda_l(X)\right)\right)\right|_{L^r} \le c_8 r^{N/2} \varepsilon_{k,l}$$

holds for all $k, l \in \mathbb{N}$. In particular, we have shown that $(S_N(\Lambda_k(X)))_{k\in\mathbb{N}}$ is a Cauchy sequence in L^r and it is clear that the limit is given by the Lyons lift $S_N(\mathbf{X})$ of the enhanced Gaussian process $\mathbf{X}$. Now fix $k \in \mathbb{N}$. For every $l \in \mathbb{N}$,

$$\begin{aligned}\left|\rho_{q\text{-}var}\left(S_N\left(\Lambda_k(X)\right), S_N(\mathbf{X})\right)\right|_{L^r} &\le \left|\rho_{q\text{-}var}\left(S_N\left(\Lambda_k(X)\right), S_N\left(\Lambda_l(X)\right)\right)\right|_{L^r} \\ &\quad + \left|\rho_{q\text{-}var}\left(S_N\left(\Lambda_l(X)\right), S_N(\mathbf{X})\right)\right|_{L^r} \\ &\le c_8 r^{N/2} \varepsilon_{k,l} + \left|\rho_{q\text{-}var}\left(S_N\left(\Lambda_l(X)\right), S_N(\mathbf{X})\right)\right|_{L^r}.\end{aligned}$$

It is easy to see that

$$\varepsilon_{k,l} \to V_\infty\left(R_{(\Lambda_k(X)-X)}, [0,1]^2\right)^{1/2-\rho/(2\gamma)} \quad \text{for } l \to \infty$$

and since

$$\left|\rho_{q\text{-}var}\left(S_N\left(\Lambda_l(X)\right), S_N(\mathbf{X})\right)\right|_{L^r} \to 0 \quad \text{for } l \to \infty$$

we can conclude that

$$\left|\rho_{q\text{-}var}\left(S_N\left(\Lambda_k(X)\right), S_N(\mathbf{X})\right)\right|_{L^r} \le c_8 r^{N/2} V_\infty\left(R_{(\Lambda_k(X)-X)}, [0,1]^2\right)^{1/2-\rho/(2\gamma)}$$

holds for every $k \in \mathbb{N}$. Finally, we have for $[\sigma,\tau] \times [\sigma',\tau'] \subset [0,1]^2$

$$\left|R_{(\Lambda_k(X)-X)}\begin{pmatrix}\sigma, \tau \\ \sigma', \tau'\end{pmatrix}\right|_{\mathbb{R}^{d\times d}} \le 4 \sup_{0\le s<t\le 1} \left|R_{(\Lambda_k(X)-X)}(s,t)\right|_{\mathbb{R}^{d\times d}}$$

and hence

$$V_\infty\left(R_{(\Lambda_k(X)-X)}, [0,1]^2\right) \le 4 \sup_{0\le s<t\le 1} \left|R_{(\Lambda_k(X)-X)}(s,t)\right|_{\mathbb{R}^{d\times d}}.$$

Furthermore, for any $s < t$,

$$\left|R_{(\Lambda_k(X)-X)}(s,t)\right|_{\mathbb{R}^{d\times d}} \le \left|\Lambda_k(X)_s - X_s\right|_{L^2(\mathbb{R}^d)} \left|\Lambda_k(X)_t - X_t\right|_{L^2(\mathbb{R}^d)} \le \sup_{0\le t\le 1} \left|\Lambda_k(X)_t - X_t\right|^2_{L^2(\mathbb{R}^d)}$$

and therefore

$$V_\infty\left(R_{(\Lambda_k(X)-X)}, [0,1]^2\right)^{1/2-\rho/(2\gamma)} \le c_9 \sup_{0\le t\le 1} \left|\Lambda_k(X)_t - X_t\right|^{1-\rho/\gamma}_{L^2(\mathbb{R}^d)}$$

which shows the result. □

The next Theorem gives pathwise convergence rates for the Wong–Zakai error for suitable approximations of the driving signal.

Theorem 6. *Let X be as in Theorem 1 for $\rho < 2$, $K \ge V_\rho(R_X, [0,1]^2)$ and $X^{(k)} = \Lambda_k(X)$. Consider the SDEs*

$$\mathrm{d}Y_t = V(Y_t)\,\mathrm{d}\mathbf{X}_t, \quad Y_0 \in \mathbb{R}^n, \tag{6.1}$$

$$\mathrm{d}Y_t^{(k)} = V\left(Y_t^{(k)}\right)\mathrm{d}X_t^{(k)}, \quad Y_0^{(k)} = Y_0 \in \mathbb{R}^n, \tag{6.2}$$

where $|V|_{Lip^\theta} \le \nu < \infty$ *for a* $\theta > 2\rho$. *Assume that there is a constant* C_1 *and a sequence* $(\varepsilon_k)_{k\in\mathbb{N}} \subset \bigcup_{r\ge 1} l^r$ *such that*

$$\sup_{0\le t\le 1} \left|X_t^{(k)} - X_t\right|_{L^2}^2 \le C_1 \varepsilon_k^{1/\rho} \quad \textit{for all } k \in \mathbb{N}.$$

Choose η, q *such that*

$$0 \le \eta < \min\left\{\frac{1}{\rho} - \frac{1}{2}, \frac{1}{2\rho} - \frac{1}{\theta}\right\} \quad \textit{and} \quad q \in \left(\frac{2\rho}{1-2\rho\eta}, \theta\right).$$

Then both SDEs (6.1) *and* (6.2) *have unique solutions* Y *and* $Y^{(k)}$ *and there is a finite random variable* C *and a null set* M *such that*

$$\left|Y^{(k)}(\omega) - Y(\omega)\right|_{\infty;[0,1]} \le \left|Y^{(k)}(\omega) - Y(\omega)\right|_{q\text{-}var;[0,1]} \le C(\omega)\varepsilon_k^\eta \tag{6.3}$$

holds for all $k \in \mathbb{N}$ *and* $\omega \in \Omega \setminus M$. *The random variable* C *depends on* $\rho, q, \eta, \nu, \theta, K, C_1$, *the sequence* $(\varepsilon_k)_{k\in\mathbb{N}}$ *and the driving process* X *but not on the equation itself. The same holds for the set* M.

Remark 3. *Note that this means that we have universal rates*, *i.e. the set* M *and the random variable* C *are valid for all starting points* (*and also vector fields subject to a uniform* Lip^θ*-bound*). *In particular, our convergence rates apply to solutions viewed as* C^l*-diffeomorphisms where* $l = [\theta - q]$, *cf.* [7], *Theorem* 11.12 *and* [5].

Proof of Theorem 6. Note that $\gamma > \rho$ and $\frac{1}{\rho} + \frac{1}{\gamma} > 1$ is equivalent to $0 < \frac{1}{2\rho} - \frac{1}{2\gamma} < \frac{1}{\rho} - \frac{1}{2}$. Hence there is a $\gamma_0 > \rho$ such that $\eta = \frac{1}{2\rho} - \frac{1}{2\gamma_0}$ and $\frac{1}{\rho} + \frac{1}{\gamma_0} > 1$. Furthermore, $2\gamma_0 = \frac{2\rho}{1-2\rho\eta} < q$. Choose $\gamma_1 > \gamma_0$ such that still $2\gamma_1 < q$ and $\eta < \frac{1}{2\rho} - \frac{1}{2\gamma_1} < \frac{1}{\rho} - \frac{1}{2}$, hence $\frac{1}{\rho} + \frac{1}{\gamma_1} > 1$ hold. Set $\alpha := \frac{1}{2\rho} - \frac{1}{2\gamma_1} - \eta > 0$. From Theorem 5 we know that for every $r \ge 1$ and $N \in \mathbb{N}$ there is a constant c_1 such that

$$\left|\rho_{q\text{-}var}\left(S_N\left(X^{(k)}\right), S_N(\mathbf{X})\right)\right|_{L^r} \le c_1 r^{N/2} \sup_{0\le t\le 1} \left|X_t^{(k)} - X_t\right|_{L^2}^{1-\rho/\gamma} \le c_2 r^{N/2} \varepsilon_k^{1/(2\rho)-1/(2\gamma)}$$

holds for every $k \in \mathbb{N}$. Hence

$$\left|\frac{\rho_{q\text{-}var}(S_N(X^{(k)}), S_N(\mathbf{X}))}{\varepsilon_k^\eta}\right|_{L^r} \le c_2 r^{N/2} \varepsilon_k^\alpha$$

for every $k \in \mathbb{N}$. From the Markov inequality, for any $\delta > 0$,

$$\sum_{k=1}^\infty P\left[\frac{\rho_{q\text{-}var}(S_N(X^{(k)}), S_N(\mathbf{X}))}{\varepsilon_k^\eta} \ge \delta\right] \le \frac{1}{\delta^r} \sum_{k=1}^\infty \left|\frac{\rho_{q\text{-}var}(S_N(X^{(k)}), S_N(\mathbf{X}))}{\varepsilon_k^\eta}\right|_{L^r}^r \le c_3 \sum_{k=1}^\infty \varepsilon_k^{\alpha r}.$$

By assumption, we can choose r large enough such that the series converges. With Borel–Cantelli we can conclude that

$$\frac{\rho_{q\text{-}var}(S_N(X^{(k)}), S_N(\mathbf{X}))}{\varepsilon_k^\eta} \to 0$$

outside a null set M for $k \to \infty$. We set

$$C_2 := \sup_{k\in\mathbb{N}} \frac{\rho_{q\text{-}var}(S_N(X^{(k)}), S_N(\mathbf{X}))}{\varepsilon_k^\eta} < \infty \quad \text{a.s.}$$

Since C_2 is the supremum of $\mathcal{F}$-measurable random variables it is itself $\mathcal{F}$-measurable. Now set $N = [q]$ which turns $\rho_{q\text{-}var}$ into a rough path metric. Note that since $\theta > 2\rho$, (6.1) and (6.2) have indeed unique solutions Y and $Y^{(k)}$. We substitute the driver $\mathbf{X}$ by $S_N(\mathbf{X})$ resp. $X^{(k)}$ by $S_N(X^{(k)})$ in the above equations, now considered as RDEs in the

q-rough paths space. Since $\theta > q$, both (RDE-) equations have again unique solutions and it is clear that they coincide with Y and $Y^{(k)}$. From

$$\rho_{q\text{-}var}\big(S_N\big(X^{(k)}\big),\mathbf{1}\big) \le \rho_{q\text{-}var}\big(S_N\big(X^{(k)}\big),S_N(\mathbf{X})\big) + \rho_{q\text{-}var}\big(S_N(\mathbf{X}),\mathbf{1}\big) \le C_1 + \rho_{q\text{-}var}\big(S_N(\mathbf{X}),\mathbf{1}\big)$$

we see that for every $\omega \in \Omega \setminus M$ the $S_N(X^{(k)}(\omega))$ are uniformly bounded for all k in the topology given by the metric $\rho_{q\text{-}var}$. Thus we can apply local Lipschitz-continuity of the Itō–Lyons map (see [7], Theorem 10.26) to see that there is a random variable C_3 such that

$$\big|Y^{(k)} - Y\big|_{q\text{-}var;[0,1]} \le C_3 \rho_{q\text{-}var}\big(S_N\big(X^{(k)}\big),S_N(\mathbf{X})\big) \le C_3 \cdot C_2 \varepsilon_k^{\eta}$$

holds for every $k \in \mathbb{N}$ outside M. Finally,

$$\big|Y_t^{(k)} - Y_t\big| = \big|Y_{0,t}^{(k)} - Y_{0,t}\big| \le \big|Y^{(k)} - Y\big|_{q\text{-}var;[0,t]} \le \big|Y^{(k)} - Y\big|_{q\text{-}var;[0,1]}$$

is true for all $t \in [0,1]$ and the claim follows. □

6.1. *Mollifier approximations*

Let ϕ be a mollifier function with support $[-1,1]$, i.e. $\phi \in C_0^{\infty}([-1,1])$ is positive and $|\phi|_{L^1} = 1$. If $x : [0,1] \to \mathbb{R}$ is a continuous path, we denote by $\bar{x} : \mathbb{R} \to \mathbb{R}$ its continuous extension to the whole real line, i.e.

$$\bar{x}_u = \begin{cases} x_0 & \text{for } x \in (-\infty, 0], \\ x_u & \text{for } x \in [0,1], \\ x_1 & \text{for } x \in [1, \infty). \end{cases}$$

For $\varepsilon > 0$ set

$$\phi_\varepsilon(u) := \frac{1}{\varepsilon}\phi(u/\varepsilon) \quad \text{and}$$

$$x_t^{\varepsilon} := \int_{\mathbb{R}} \phi_\varepsilon(t-u)\bar{x}_u \,\mathrm{d}u.$$

Let $(\varepsilon_k)_{k\in\mathbb{N}}$ be a sequence of real numbers such that $\varepsilon_k \to 0$ for $k \to \infty$. Define

$$\Lambda_k(x) := x^{\varepsilon_k}.$$

In [7], Chapter 15.2.3 it is shown that the sequence $(\Lambda_k)_{k\in\mathbb{N}}$ fulfils the conditions of Theorem 5.

Corollary 7. *Let X be as in Theorem 1 and assume that there is a constant C such that $V_\rho(R_X;[s,t]^2) \le C|t-s|^{1/\rho}$ holds for all $s < t$. Choose $(\varepsilon_k)_{k\in\mathbb{N}} \in \bigcup_{r\ge 1} l^r$ and set $X^{(k)} = X^{\varepsilon_k}$. Then the solutions $Y^{(k)}$ of the SDE (6.2) converge pathwise to the solution Y of (6.1) in the sense of (6.3) with rate $\mathrm{O}(\varepsilon_k^{\eta})$ where η is chosen as in Theorem 6.*

Proof. It suffices to note that for every $\varepsilon > 0$, $Z \in \{X^1, \dots, X^d\}$ and $t \in [0,1]$ we have

$$\begin{aligned} E\big[\big|Z_t^{\varepsilon} - Z_t\big|^2\big] &= E\bigg[\bigg(\int_{\mathbb{R}} \phi_\varepsilon(t-u)(\bar{Z}_u - Z_t)\,\mathrm{d}u\bigg)^2\bigg] \\ &= E\bigg[\bigg(\int_{[t-\varepsilon,t+\varepsilon]} \phi_\varepsilon(t-u)(\bar{Z}_u - Z_t)\,\mathrm{d}u\bigg)^2\bigg] \\ &= E\bigg[\int_{[t-\varepsilon,t+\varepsilon]^2} \phi_\varepsilon(t-u)\phi_\varepsilon(t-v)(\bar{Z}_u - Z_t)(\bar{Z}_v - Z_t)\,\mathrm{d}u\,\mathrm{d}v\bigg] \end{aligned}$$

$$= \int_{[t-\varepsilon,t+\varepsilon]^2} \phi_\varepsilon(t-u)\phi_\varepsilon(t-v) E\big[(\bar{Z}_u - Z_t)(\bar{Z}_v - Z_t)\big]\,\mathrm{d}u\,\mathrm{d}v$$

$$\leq \sup_{\substack{t\in[0,1]\\ |h_1|,|h_2|\leq\varepsilon}} \big|E\big[(\bar{Z}_{t+h_1} - Z_t)(\bar{Z}_{t+h_2} - Z_t)\big]\big|$$

$$\leq \sup_{\substack{t\in[0,1]\\ |h|\leq\varepsilon}} E\big[(\bar{Z}_{t+h} - Z_t)^2\big] \leq c_1 \varepsilon^{1/\rho}$$

from which follows that $\sup_{0\leq t\leq 1} |X_t^{\varepsilon_k} - X_t|_{L^2}^2 \leq c_1 \varepsilon_k^{1/\rho}$. We conclude with Theorem 6. □

6.2. *Piecewise linear approximations*

If $D = \{0 = t_0 < t_1 < \cdots < t_{\#D-1} = 1\}$ is a partition of $[0,1]$ and $x : [0,1] \to \mathbb{R}$ a continuous path, we denote by x^D the piecewise linear approximation of x at the points of D, i.e. x^D coincides with x at the points t_i and if $t_i \leq t < t_{i+1}$ we have

$$\frac{x_{t_{i+1}}^D - x_t^D}{t_{i+1} - t} = \frac{x_{t_{i+1}} - x_{t_i}}{t_{i+1} - t_i}.$$

Let $(D_k)_{k\in\mathbb{N}}$ be a sequence of partitions of $[0,1]$ such that $|D_k| := \max_{t_i \in D_k}\{|t_{i+1} - t_i|\} \to 0$ for $k \to \infty$. If $x : [0,1] \to \mathbb{R}$ is continuous, we define

$$\Lambda_k(x) := x^{D_k}.$$

In [7], Chapter 15.2.3 it is shown that $(\Lambda_k)_{k\in\mathbb{N}}$ fulfils the conditions of Theorem 5. If R_X is the covariance of a Gaussian process, we set

$$|D|_{R_X,\rho} = \Big(\max_{t_i\in D} V_\rho\big(R_X; [t_i, t_{i+1}]^2\big)\Big)^\rho.$$

Corollary 8. *Let X be as in Theorem 1. Choose a sequence of partitions $(D_k)_{k\in\mathbb{N}}$ of the interval $[0,1]$ such that $(|D_k|_{R_X,\rho})_{k\in\mathbb{N}} \in \bigcup_{r\geq 1} l^r$ and set $X^{(k)} = X^{D_k}$. Then the solutions $Y^{(k)}$ of the SDE (6.2) converge pathwise to the solution Y of (6.1) in the sense of (6.3) with rate $\mathrm{O}(\varepsilon_k^\eta)$ where $(\varepsilon_k)_{k\in\mathbb{N}} = (|D_k|_{R_X,\rho})_{k\in\mathbb{N}}$ and η is chosen as in Theorem 6.*

Proof. Let D be any partition of $[0,1]$ and $t \in [t_i, t_{i+1}]$ where $t_i, t_{i+1} \in D$. Take $Z \in \{X^1, \ldots, X^d\}$. Then

$$Z_t^D - Z_t = Z_{t_i,t_{i+1}} \frac{t - t_i}{t_{i+1} - t_i} - Z_{t_i,t}.$$

Therefore

$$\big|Z_t^D - Z_t\big|_{L^2} \leq |Z_{t_i,t_{i+1}}|_{L^2} + |Z_{t_i,t}|_{L^2} \leq 2V_\rho\big(R_X; [t_i, t_{i+1}]^2\big)^{1/2} \leq 2|D|_{R_X,\rho}^{1/(2\rho)}.$$

We conclude with Theorem 6. □

Example 1. *Let $X = B^H$ be the fractional Brownian motion with Hurst parameter $H \in (1/4, 1/2]$. Set $\rho = \frac{1}{2H} < 2$. Then one can show that R_X has finite ρ-variation and $V_\rho(R_X; [s,t]^2) \leq c(H)|t-s|^{1/\rho}$ for all $(s,t) \in \Delta$ (see [8], Example 1). Assume that the vector fields in (6.1) are sufficiently smooth by which we mean that $1/\rho - 1/2 \leq 1/(2\rho) - 1/\theta$, i.e.*

$$\theta \geq \frac{2\rho}{\rho - 1} = \frac{1}{1/2 - H}.$$

Let $(D_k)_{k\in\mathbb{N}}$ *be the sequence of uniform partitions. By Corollary 8, for every* $\eta < 2H - 1/2$ *there is a random variable* C *such that*

$$\left|Y^{(k)} - Y\right|_\infty \le C\left(\frac{1}{k}\right)^\eta \quad a.s.$$

hence we have a Wong–Zakai convergence rate arbitrary close to $2H - 1/2$. *In particular, for the Brownian motion, we obtain a rate close to* $1/2$, *see also* [9] *and* [5]. *For* $H \to 1/4$, *the convergence rate tends to* 0 *which reflects the fact that the Lévy area indeed diverges for* $H = 1/4$, *see* [2].

6.3. *The simplified step-N Euler scheme*

Consider again the SDE

$$\mathrm{d}Y_t = V(Y_t)\,\mathrm{d}X_t, \quad Y_0 \in \mathbb{R}^n$$

interpreted as a pathwise RDE driven by the lift $\mathbf{X}$ of a Gaussian process X which fulfils the conditions of Theorem 1. Let D be a partition of $[0, 1]$. We recall the simplified step-N Euler scheme from the introduction:

$$\begin{aligned} Y_0^{\mathrm{sEuler}^N;D} &= Y_0, \\ Y_{t_{j+1}}^{\mathrm{sEuler}^N;D} &= Y_{t_j}^{\mathrm{sEuler}^N;D} + V_i\big(Y_{t_j}^{\mathrm{sEuler}^N;D}\big)X^i_{t_j,t_{j+1}} + \frac{1}{2}\mathcal{V}_{i_1}V_{i_2}\big(Y_{t_j}^{\mathrm{sEuler}^N;D}\big)X^{i_1}_{t_j,t_{j+1}}X^{i_2}_{t_j,t_{j+1}} \\ &\quad + \cdots + \frac{1}{N!}\mathcal{V}_{i_1}\cdots\mathcal{V}_{i_{N-1}}V_{i_N}\big(Y_{t_j}^{\mathrm{sEuler}^N;D}\big)X^{i_1}_{t_j,t_{j+1}}\cdots X^{i_N}_{t_j,t_{j+1}}, \end{aligned}$$

where $t_j \in D$. In this section, we will investigate the convergence rate of this scheme. For simplicity, we will assume that

$$V_\rho\big(R_X;[s,t]^2\big) = \mathrm{O}\big(|t-s|^{1/\rho}\big)$$

which can always be achieved at the price of a deterministic time-change based on

$$[0,1] \ni t \mapsto \frac{V_\rho(R_X;[0,t]^2)^\rho}{V_\rho(R_X;[0,1]^2)^\rho} \in [0,1].$$

Set $D_k = \{\frac{i}{k}\colon i = 0,\ldots,k\}$.

Corollary 9. *Let* $p > 2\rho$ *and assume that* $|V|_{Lip^\theta} < \infty$ *for* $\theta > p$. *Choose* η *and* N *such that*

$$\eta < \min\left\{\frac{1}{\rho} - \frac{1}{2}, \frac{1}{2\rho} - \frac{1}{\theta}\right\} \quad \text{and} \quad N \le [\theta].$$

Then there are random variables C_1 *and* C_2 *such that*

$$\max_{t_j\in D_k}\big|Y_{t_j} - Y_{t_j}^{\mathrm{sEuler}^N;D_k}\big| \le C_1\left(\frac{1}{k}\right)^\eta + C_2\left(\frac{1}{k}\right)^{(N+1)/p-1} \quad a.s.\ \text{for all } k \in \mathbb{N}.$$

Proof. Recall the step-N Euler scheme from the introduction (or cf. [7], Chapter 10). Set $X^{(k)} = X^{D_k}$ and let $Y^{(k)}$ be the solution of the SDE (6.2). Then $Y_{t_j}^{\mathrm{sEuler}^N;D_k} = (Y^{(k)})_{t_j}^{\mathrm{Euler}^N;D_k}$ for every $t_j \in D_k$ and therefore, using the triangle inequality,

$$\max_{t_j\in D_k}\big|Y_{t_j} - Y_{t_j}^{\mathrm{sEuler}^N;D_k}\big| \le \sup_{t\in[0,1]}\big|Y_t - Y_t^{(k)}\big| + \max_{t_j\in D_k}\big|Y_{t_j}^{(k)} - \big(Y^{(k)}\big)_{t_j}^{\mathrm{Euler}^N;D_k}\big|.$$

By the choice of D_k we have $|D_k|_{R_X,\rho} = \mathrm{O}(k^{-1})$. Applying Corollary 8 we obtain for the first term $|Y - Y^{(k)}|_\infty = \mathrm{O}(k^{-\eta})$. Refering to [7], Theorem 10.30 we see that the second term is of order $\mathrm{O}(k^{-((N+1)/p-1)})$. $\square$

Remark 4. *Assume that the vector fields are sufficiently smooth, i.e.* $\theta \geq \frac{2\rho}{\rho-1}$. *Then we obtain an error of* $\mathrm{O}(k^{-(2/p-1/2)}) + \mathrm{O}(k^{-((N+1)/p-1)})$, *any* $p > 2\rho$. *That means that in the case* $\rho = 1$, *the step*-2 *scheme* (*i.e. the simplified Milstein scheme*) *gives an optimal convergence rate of* (*almost*) $1/2$. *For* $\rho \in (1, 2)$, *the step*-3 *scheme gives an optimal rate of* (*almost*) $1/\rho - 1/2$. *In particular, we see that using higher order schemes does not improve the convergence rate since in that case*, *the Wong–Zakai error persists*. *In the fractional Brownian motion case*, *the simplified Milstein scheme gives an optimal convergence rate of* (*almost*) $1/2$ *for the Brownian motion and for* $H \in (1/4, 1/2)$ *the step*-3 *scheme gives an optimal rate of* (*almost*) $2H - 1/2$. *This answers a conjecture stated in* [4].

References

[1] T. Cass and P. Friz. Densities for rough differential equations under Hoermander's condition. *Ann. of Math. (2)* **171** (2010) 2115–2141. MR2680405

[2] L. Coutin and Z. Qian. Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields* **122** (2002) 108–140. MR1883719

[3] A. M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express. AMRX* (2007) Art. ID abm009, 40. MR2387018

[4] A. Deya, A. Neuenkirch and S. Tindel. A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (2012) 518–550.

[5] P. Friz and S. Riedel. Convergence rates for the full Brownian rough paths with applications to limit theorems for stochastic flows. *Bull. Sci. Math.* **135** (2011) 613–628. MR2838093

[6] P. Friz and N. Victoir. Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (2010) 369–413. MR2667703

[7] P. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge Univ. Press, Cambridge, 2010. MR2604669

[8] P. Friz and N. Victoir. A note on higher dimensional p-variation. *Electron. J. Probab.* **16** (2011) 1880–1899. MR2842090

[9] I. Gyöngy and A. Shmatkov. Rate of convergence of Wong–Zakai approximations for stochastic partial differential equations. *Appl. Math. Optim.* **54** (2006) 315–341. MR2268661

[10] M. Hairer. Rough stochastic PDEs. *Comm. Pure Appl. Math.* **64** (2011) 1547–1585. MR2832168

[11] K. Hara and M. Hino. Fractional order Taylor's series and the neo-classical inequality. *Bull. Lond. Math. Soc.* **42** (2010) 467–477. MR2651942

[12] Y. Hu and D. Nualart. Rough path analysis via fractional calculus. *Trans. Amer. Math. Soc.* **361** (2009) 2689–2718. MR2471936

[13] S. Janson. *Gaussian Hilbert Spaces*. Cambridge Univ. Press, New York, 1997. MR1474726

[14] T. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoam.* **14** (1998) 215–310. MR1654527

[15] T. Lyons and Z. Qian. *System Control and Rough Paths*. Oxford Univ. Press, New York, 2002. MR2036784

[16] A. Neuenkirch, S. Tindel and J. Unterberger. Discretizing the fractional Lévy area. *Stochastic Process. Appl.* **120** (2010) 223–254. MR2576888

[17] C. Reutenauer. *Free Lie Algebras*. Clarendon Press, New York, 1993. MR1231799

[18] N. Towghi. Multidimensional extension of L. C. Young's inequality. *JIPAM J. Inequal. Pure Appl. Math.* **3** (2002) 13 (electronic).