CUBATURE ON WIENER SPACE: PATHWISE CONVERGENCE

CHRISTIAN BAYER AND PETER K. FRIZ

Abstract. Cubature on Wiener space [Lyons, T.; Victoir, N.; Proc. R. Soc. Lond. A 8 January 2004 vol. 460 no. 2041 169-198] provides a powerful alternative to Monte Carlo simulation for the integration of certain functionals on Wiener space. More specifically, and in the language of mathematical finance, cubature allows for fast computation of European option prices in generic diffusion models.

We give a random walk interpretation of cubature and similar (e.g. the Ninomiya–Victoir) weak approximation schemes. By using rough path analysis, we are able to establish weak convergence for general path-dependent option prices.

1. Introduction

Cubature on Wiener space (Kusuoka [6, 7], Lyons and Victoir [9], see also Litterer and Lyons [8], Ninomiya and Victoir [13], Ninomiya and Ninomiya [12]) provides a powerful alternative to Monte Carlo simulation for the integration of certain functionals on Wiener space. As of present, these functionals are of the form $f(S_T)$ where S_T is the image of a d-dimensional Brownian motion under the Itô-map (the solution map to a stochastic differential equation); the aim of cubature on Wiener space is then to provide a fast numerical algorithm to compute $E[f(S_T)]$, where the expectation is taken over the d-dimensional Wiener measure.

In the language of mathematical finance, cubature deals with European option prices in generic diffusion models. Although some exotic options can be handled in this framework (e.g. Asian options, by enhancing the state-space) general path-dependent options are not included in the presently available analysis on cubature methods. It must be admitted that cubature has been designed for fast evaluation of payoffs of the type $f(S_T)$; but even so, it may maintain its benefits in mildly path-dependent situation and, in any case, convergence to the correct value will be considered a minimal requirement by most users.

The answer to "How can it fail to converge to the correct value?" is not trivial: cubature methods are essentially derived from replacing Wiener-measure but a path-space measure supported on smooth paths $\{\omega_i\}$, subject to certain technical conditions relating to the iterated integrals of these paths. Stochastic differential equations, however, are far from stable under perturbations in the iterated integrals: recall the well-known examples of McShane [11] which give uniform approximations to Brownian motion where the limiting differential equation exhibits bias in the form of additional drift terms. (The explanation is that these approximations do *not* correctly approximate the iterated integrals of Brownian motion known as Lévy's stochastic area.) At the risk of confusing the reader, even if is guaranteed that a sample path *and its stochastic area* are uniformly correctly approximated, the limiting differential equation may still exhibit additional drift terms[1]. The point is that topology matters: uniform convergence needs to be replaced by a stronger notion

2000 *Mathematics Subject Classification.* 60F17,60H35,91G60.

Key words and phrases. Cubature, rough paths, weak approximation schemes for SDEs, pricing of path-dependent options.

P. K. Friz has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nr. 258237.

[1] ... now involving two or more iterated Lie brackets of the diffusion vector fields; see [4] for more on such subtleties.

of Hölder (or p-variation) rough path topology in order to use the stability results of rough path theory.

Our key idea is to view the iterations of cubature steps, (Lyons and Victoir [9, Theorem 3.3] for instance), via an underlying random walk of the driving signal, Brownian motion plus Lévy's area. The iterated cubature scheme corresponds precisely to stochastic differential equations in which the driving Brownian motion is replaced by k properly rescaled concatenations of the ω_i (say, chosen independently with probability λ_i at each step). Thanks to the smoothness of the ω_i, such a path has canonically defined iterated integrals; the "only" thing left to do is to establish weak convergence of this random walk to Brownian motion and Lévy's stochastic area, in the correct rough path topology. It is then an immediate consequence of the continuity of the Itô map in rough path sense (i.e. as a deterministic function of path and area in rough path topology) to see that this entails the desired weak convergence result for path dependent functionals of the type $f(S_t : 0 \le t \le T)$.

Weak convergence questions of this type were first discussed in E. Breuillard, P. Friz, M. Huesmann [2]. Unfortunately, the "Rough path Donsker" theorem obtained therein does not lend itself immediately to the present applications: a moment of reflection reveals that it would cover cubature with (1) equidistant steps and (2) in which the ω_i are straight lines (Wong-Zakai!). Our strategy is thus to develop refined arguments that allow to cover the generic cubature setting as well as its recent variations (like Ninomya–Victoir). This leads, en passant, to a more flexible version of the Donsker theorem for Brownian motion on Lie groups in topologies considerably finer than the uniform one.

The mathematical content – weak convergence of discrete structures to (Stratonovich) SDE solutions – should also be compared to the (typically Itô) diffusion limits of Markov chains (cf. Stroock and Varadhan [16, Section 11.2]), although we shall not pursue this point further here.

The current paper uses many ideas and results from rough path theory, see Lyons [10] and Friz and Victoir [5], which we primarily use for reference in this paper. For cubature on Wiener space, the authoritative reference remains Lyons and Victoir [9].

2. Cubature on Wiener space and the associated random walks

Let $B = (B_t)_{t\in[0,1]}$ denote a standard d-dimensional Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, P)$ and $\mathbf{B}_\cdot = S_2(B)_{0,\cdot}$, i.e., $\mathbf{B}$ is the Brownian motion enhanced by its Lévy area. The geometrical setting of $\mathbf{B}$ is the Lie group $G^2(\mathbb{R}^d)$, which can be defined as follows: let $e_1, \dots, e_d$ denote the canonical basis of $\mathbb{R}^d$. Then $e_i \otimes e_j$, $1 \le i, j \le d$, forms a basis for the tensor product $\mathbb{R}^d \otimes \mathbb{R}^d$. Consider the algebra $T^2(\mathbb{R}^d) \coloneqq \mathbb{R} \oplus \mathbb{R}^d \oplus \mathbb{R}^d \otimes \mathbb{R}^d$, which is understood as a step-2 nilpotent non-commutative algebra, i.e., for $\mathbf{x}_1 = z_1 + x_1 + a_1, \mathbf{x}_2 = z_2 + x_2 + a_2 \in T^2(\mathbb{R}^d)$ the product is given by

$$\mathbf{x}_1 \otimes \mathbf{x}_2 = z_1 z_2 + (z_2 x_1 + z_1 x_2) + (z_2 a_1 + z_1 a_2 + x_1 \otimes x_2).$$

Consider $\mathfrak{g}^2(\mathbb{R}^d) \subset T^2(\mathbb{R}^d)$, the Lie-algebra generated by $e_1, \dots, e_d$ together with $[e_i, e_j]$, $1 \le i < j \le d$, with the commutator defined by $[\mathbf{x}, \mathbf{y}] \coloneqq \mathbf{x} \otimes \mathbf{y} - \mathbf{y} \otimes \mathbf{x}$, $\mathbf{x}, \mathbf{y} \in T^2(\mathbb{R}^d)$. The exponential map $\exp : T^2(\mathbb{R}^d) \to T^2(\mathbb{R}^d)$ defined by

$$\exp(\mathbf{x}) \coloneqq 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbf{x}^{\otimes k} = 1 + \mathbf{x} + \frac{1}{2} \mathbf{x} \otimes \mathbf{x}$$

in the step-2 nilpotent setting, maps $\mathfrak{g}^2(\mathbb{R}^d)$ in a bijective way to the Lie group $G^2(\mathbb{R}^d) \coloneqq \exp(\mathfrak{g}^2(\mathbb{R}^d)) \subset T^2(\mathbb{R}^d)$.

This Lie group is highly relevant for rough path analysis, since it is the geometric setting of the enhanced Brownian motion mentioned before. Indeed, the $T^2(\mathbb{R}^d)$-valued process $\mathbf{B}$

defined by

$$\mathbf{B}_t := 1 + \sum_{i=1}^d B_t^i e_i + \sum_{i,j=1}^d \int_0^t B_s^i \circ dB_s^j \, e_i \otimes e_j =: S_2(B)_{0,t}, \quad 0 \le t \le 1, \tag{1}$$

lives in the Lie group $G^2(\mathbb{R}^d)$, i.e., $P\left(\mathbf{B}_t \in G^2(\mathbb{R}^d),\ t \in [0,1]\right) = 1$. In a similar way, we will consider the (step-m truncated) *signature*, see Friz and Victoir [5],

$$S_m(B)_{0,t} := 1 + \sum_{k=1}^m \sum_{i_1,\dots,i_k \in \{1,\dots,d\}} \int_{0 \le t_1 \le \cdots \le t_k \le t} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k} \, e_{i_1} \otimes \cdots \otimes e_{i_k}, \tag{2}$$

which takes values in the step-m nilpotent Lie-group $G^m(\mathbb{R}^d)$ defined analogously to $G^2(\mathbb{R}^d)$.[2]

Consider the stochastic differential equation (in Stratonovich form)

$$dX_t = V_0(X_t)dt + \sum_{i=1}^d V_i(X_t) \circ dB_t^i, \tag{3}$$

$X_0 = x_0 \in \mathbb{R}^N$. Here, $V_0, V_1, \dots, V_d : \mathbb{R}^N \to \mathbb{R}^N$ is a collection of smooth vector fields. A *cubature formula on Wiener space* is a random variable W taking values in the space $C^{1\text{-var}}([0,1], \mathbb{R}^d)$ of continuous paths of bounded variation with values in $\mathbb{R}^d$ such that we have

$$E\left[\int_{0 \le t_1 \le \cdots \le t_k \le 1} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k}\right] = E\left[\int_{0 \le t_1 \le \cdots \le t_k \le 1} dW_{t_1}^{i_1} \cdots dW_{t_k}^{i_k}\right]. \tag{4}$$

Equation (4) is supposed to hold for all multi-indices $I = (i_1, \dots, i_k) \in \{1, \dots, d\}^k$ with $k \le m$ and all $1 \le k \le m$, where m is a fixed positive integer, the *order* of the cubature formula. Moreover, we note that the paths of the process W are of bounded variation, therefore the integrals on the right hand side of (4) can be classically defined in a pathwise sense. Notice that we do not use cross-integrals between time dt and the Brownian motion dB_t. Therefore, a cubature formula in this sense can only be used to approximate SDEs with drift $V_0 \equiv 0$. We will cover the general case later in Section 3.

Rephrased in terms of the (truncated) signature, equation (4) means that

$$E\left[S_m(B)_{0,1}\right] = E\left[S_m(W)_{0,1}\right], \tag{5}$$

where the expectation takes values in the algebra $T^m(\mathbb{R}^d)$. Obviously, any cubature formula on Wiener space can be rescaled to a cubature formula on the interval $[0, \Delta t]$, $\Delta t > 0$, by replacing W with the bounded variation path

$$\delta_{\sqrt{\Delta t}}(W) : [0, \Delta t] \to \mathbb{R}^d, \quad s \mapsto \sqrt{\Delta t} W(s/\Delta t). \tag{6}$$

On the level of signatures, this corresponds to applying the *dilatation* operator $\delta_{\sqrt{\Delta t}} : G^m(\mathbb{R}^d) \to G^m(\mathbb{R}^d)$, i.e.,

$$S_m\left(\delta_{\sqrt{\Delta t}}(W)\right)_{0,\Delta t} = \delta_{\sqrt{\Delta t}}(S_m(W)_{0,1}).$$

Remark 2.1. *Note that the symbol $\delta_{\sqrt{\Delta t}}$ has different meanings on both sides of the equation: on the left hand side, it is a function from $C([0,1], \mathbb{R}^d)$ to $C([0, \Delta t], \mathbb{R}^d)$, whereas on the right hand side it is the restriction to $G^m(\mathbb{R}^d)$ of a linear map defined on the algebra $T^m(\mathbb{R}^d)$ by*

$$\delta_{\sqrt{\Delta t}}(e_{i_1} \otimes \cdots \otimes e_{i_k}) := \Delta t^{k/2} e_{i_1} \otimes \cdots \otimes e_{i_k}, \quad 0 \le k \le m, \quad i_1, \dots, i_k \in \{1, \dots, d\}.$$

[2]Note that $S_m(B)_{0,\cdot} = S_m(\mathbf{B})_{0,\cdot}$, the *Lyons lift* of the enhanced Brownian motion, reflecting the fact that $S_m(B)$ depends uniquely and continuously on $\mathbf{B}$ – whereas $\mathbf{B}$ itself is not uniquely and certainly not continuously given by B. For instance, we could have chosen the Ito-integral instead of the Stratonovich integral.

Given a mesh $\mathcal{D} = \{0 = t_0 < t_1 < \cdots < t_n = 1\}$, set $\Delta t_k := t_k - t_{k-1}$, $k = 1, \dots, n$, and $|\mathcal{D}| = \max_k \Delta t_k$. Moreover, let $W_{(1)}, \dots, W_{(n)}$ be independent copies of the cubature formula W. We define a random variable $W^{\mathcal{D}} : [0,1] \to \mathbb{R}^d$ taking values in the space of continuous paths of bounded variation by concatenation of the paths $\delta_{\sqrt{\Delta t_k}}(W_{(k)}) : [0, \Delta t_k] \to \mathbb{R}^d$, $k = 1, \dots, n$. Again, by well known properties of the signature (the Chen theorem, see for instance [2, Theorem 7.11, Exercise 7.14]), this translates to the relation

$$\begin{aligned} S_m\left(W^{\mathcal{D}}\right)_{0,1} &= S_m\left(\delta_{\sqrt{\Delta t_1}}(W_{(1)})\right)_{0,\Delta t_1} \otimes \cdots \otimes S_m\left(\delta_{\sqrt{\Delta t_n}}(W_{(n)})\right)_{0,\Delta t_n} \\ &= \delta_{\sqrt{\Delta t_1}}\left(S_m(W_{(1)})_{0,1}\right) \otimes \cdots \otimes \delta_{\sqrt{\Delta t_n}}\left(S_m(W_{(n)})_{0,1}\right), \end{aligned}$$

where $\otimes$ denotes the multiplication in the Lie group $G^m(\mathbb{R}^d)$. Finally, let $X^{\mathcal{D}}$ denote the (pathwise ODE) solution of the equation

$$dX_t^{\mathcal{D}} = \sum_{i=1}^d V_i(X_t^{\mathcal{D}}) dW_t^{\mathcal{D},i}, \tag{7}$$

$X_0^{\mathcal{D}} = x_0$. For a given function $f : \mathbb{R}^N \to \mathbb{R}$ of interest, the method of cubature on Wiener space now consists in the approximation

$$E[f(X_1)] = E\left[f(X_1^{\mathcal{D}})\right] + O\left(|D|^{(m-1)/2}\right), \tag{8}$$

provided that certain regularity assumptions are satisfied, see [9], [13] and [7]. In particular, the method provides an efficient numerical scheme, if W has been chosen in such a way that integration of (7) is "substantially simpler" then integration of the original (3), see [1]. If f is smooth, (8) holds even for uniform meshes. If f only is Lipschitz, however, then Kusuoka [7, Theorem 4] shows that (8) holds provided that one takes certain non-homogeneous meshes. The goal of this paper regarding cubature is to show that convergence even holds for (reasonable) functionals f depending on the whole path $(X_t)_{0 \le t \le 1}$.

Example 2.2. *The cubature formulas in* [9] *are discrete random variables W taking values in the space of continuous paths of bounded variations. That is, fix k paths of bounded variation $\omega_1, \dots, \omega_k : [0,1] \to \mathbb{R}^d$ and positive real numbers $\lambda_1, \dots, \lambda_k$ with $\lambda_1 + \cdots + \lambda_k = 1$. Then W is the random variable taking values in $\{\omega_1, \dots, \omega_k\}$ with $P(W = \omega_j) = \lambda_j$, $j = 1, \dots, k$. In all the concrete cubature formulas constructed in* [9]*, the paths $\omega_j(\cdot)$ are, in fact, piecewise linear.*

Example 2.3. *We can even interpret the Wong-Zakai approximation as a cubature formula on Wiener space (of order $m = 3$; the resulting convergence in* (8) *has then weak order $(m-1)/2 = 1$, precisely as the usual Euler scheme for Ito differential equations). Indeed, choose W as the linear path $W_t = tB_1$. We note that $W^{\mathcal{D}}$ can be realized (for any mesh $\mathcal{D}$) by choosing $\delta_{\sqrt{\Delta t_k}}(W_{(k)})(s) = s(B_{t_k} - B_{t_{k-1}})$, because $W_{(k)}(t) := \frac{t}{\sqrt{\Delta t_k}}(B_{t_k} - B_{t_{k-1}})$ has the same law as W and all the $W_{(k)}$ are independent. Concatenation of these paths precisely gives the piecewise linear approximation of B with nodes in $\mathcal{D}$.*

Example 2.4. *Ninomiya and Victoir* [13] *construct a cubature formula of order $m = 5$ in the following way. Let Λ be a Bernoulli random variable (taking values ± 1 with probability $1/2$ each) and let $Z^1, \dots, Z^d$ be independent standard normal random variables. Set $\varepsilon = 1/(d+1)$. For $\omega \in \Omega$, $W(\omega)$ is defined by the following formula. If $\Lambda(\omega) = -1$, we define $W(\omega)$ to be the piecewise linear path with*

$$\dot{W}^i(\omega)(s) = \begin{cases} 1/\varepsilon, & s \in [0, \varepsilon/2],\ i = 0, \\ Z^i(\omega)/\varepsilon, & s \in]\varepsilon/2 + (i-1)\varepsilon, \varepsilon/2 + i\varepsilon],\ i \in \{1, \dots, d\}, \\ 1/\varepsilon, & s \in]1 - \varepsilon/2, 1],\ i = 0, \\ 0, & \text{else}. \end{cases}$$

If $\Lambda(\omega) = 1$, $W(\omega)$ is similarly defined by

$$\dot{W}^i(\omega)(s) = \begin{cases} 1/\varepsilon, & s \in [0, \varepsilon/2],\ i = 0, \\ Z^i(\omega)/\varepsilon, & s \in]\varepsilon/2 + (d-i)\varepsilon, \varepsilon/2 + (d-i+1)\varepsilon],\ i \in \{1, \dots, d\}, \\ 1/\varepsilon, & s \in]1 - \varepsilon/2, 1],\ i = 0, \\ 0, & else. \end{cases}$$

This means, we subdivide the interval $[0, 1]$ *into* $d + 2$ *subintervals*

$$\left[0, \frac{\varepsilon}{2}\right] \cup \left]\frac{\varepsilon}{2}, \frac{3\varepsilon}{2}\right] \cup \cdots \cup \left]1 - \frac{3\varepsilon}{2}, 1 - \frac{\varepsilon}{2}\right] \cup \left]1 - \frac{\varepsilon}{2}, 1\right].$$

On each of these subintervals, $W(\omega)$ is constant in all components albeit one, which is linear. In particular, $W(\omega)$ is again piecewise linear.

At this stage, we would like to remark that we could replace the Gaussian random variables Z^i by discrete random variables having the same moments of order up to five. Then we would obtain a special case of Example 2.2 – albeit for the non-standard choice of W^0, see Section 3 below.

Let us now turn our attention to Donsker type results: for a fixed sequence of meshes $\mathcal{D}_n$ with $|\mathcal{D}_n| \to 0$ we wish to study the corresponding sequence of paths in $G^m(\mathbb{R}^d)$, i.e., we study

$$S_m\left(W^{\mathcal{D}_n}\right)_{0,\cdot} = \left(S_m\left(W^{\mathcal{D}_n}\right)_{0,t}\right)_{t \in [0,1]}.$$

By a *Donsker theorem in rough path topology* for the sequence of cubature formulas $W^{\mathcal{D}_n}$ we understand the statement that

$$S_m\left(W^{\mathcal{D}_n}\right)_{0,\cdot} \xrightarrow[n\to\infty]{} S_m(\mathbf{B})_{0,\cdot} \tag{9}$$

weakly with respect to α-Hölder rough path topology[3], for some $\alpha \in (1/3, 1/2)$ and $m \geq 2$. (In fact, elementary results of rough path theory imply then that it suffices to consider $m = 2$. Also, the claimed convergence will actually be established for *all* $\alpha < 1/2$). As a justification for calling the convergence stated in (9) a Donsker theorem, consider the following random walk. Let us again fix the mesh $\mathcal{D}_n = \{0 = t_0 < \cdots < t_n = 1\}$. (We only take n as the number of sub-intervals for the grid $\mathcal{D}_n$ for more convenient notation. The mathematics would, of course, work in precisely the same way, if the size of $\mathcal{D}_n$ was completely arbitrary, as long as $|\mathcal{D}_n| \to 0$.) Define

$$\xi_k^n = S_m\left(\delta_{\sqrt{\Delta t_k}}(W_{(k)})\right)_{0,\Delta t_k} = \delta_{\sqrt{\Delta t_k}}(S_m(W_{(k)})_{0,1}), \tag{10}$$

a random variable taking values in $G^m(\mathbb{R}^d)$. Note that $\xi_k^n = \delta_{\sqrt{\Delta t_k}}(\xi_{(k)})$, where $\xi_{(k)}$ is an independent copy of $S_m(W)_{0,1}$. Next define the $G^m(\mathbb{R}^d)$-valued, finite random walk Ξ_k^n, $k = 0, \dots, n$, by $\Xi_0^n = 1$ and

$$\Xi_{k+1}^n = \Xi_k^n \otimes \xi_{k+1}^n,$$

where 1 is the neutral element of $G^m(\mathbb{R}^d)$. Since

$$S_m\left(W^{\mathcal{D}_n}\right)_{0,t_k} = \Xi_k^n, \quad k = 0, \dots, n,$$

$S_m\left(W^{\mathcal{D}_n}\right)_{0,\cdot}$ is, indeed, a path in $G^m(\mathbb{R}^d)$ obtained from the random walk Ξ^n by (possibly random) interpolation. This gives the link to the classical Donsker theorem as well as to the paper of Breuillard, Friz and Huesmann [2]. Let us rephrase their Theorem 3 for the current setting.

---

[3] I.e., the α-Hölder topology for functions taking values in the metric space $(G^2(\mathbb{R}^d), \|\cdot\|)$., see [5, Definition 5.1, 9.15]

Proposition 2.5. *Let W be a cubature formula on Wiener space of order $m = 2$ with finite moments of all orders in the sense that*

$$\forall q \geq 1 : E\left[\left\|S_2(W)_{0,1}\right\|^q\right] < \infty,$$

where $\|\cdot\|$ denotes the Carnot-Caratheodory norm on $G^2(\mathbb{R}^d)$, see below. Moreover, assume that W is chosen in such a way that for every ω, $S_2(W(\omega))_{0,\cdot}$ is a geodesic connecting 1 and $S_2(W(\omega))_{0,1}$. Choose uniform meshes $\mathcal{D}_n = \left\{ \frac{k}{n} \;\middle|\; k = 0, \ldots, n \right\}$. Then the Donsker theorem holds in rough path topology, i.e., $S_2\left(W^{\mathcal{D}_n}\right)_{0,\cdot}$ converges to $\mathbf{B}$ in $C^{0,\alpha-H\ddot{o}l}([0,1], G^2(\mathbb{R}^d))$, for every $\alpha < 1/2$.

Recall that the Carnot-Caratheodory norm is defined by

$$\|\mathbf{x}\| := \inf\left\{ \int_0^1 |d\gamma| \;\middle|\; \gamma \in C^{1-var}\left([0,1], \mathbb{R}^d\right),\, S_2(\gamma)_{0,1} = \mathbf{x} \right\}.$$

The infimum is actually always attained, and can be parametrized as Lipschitz-continuous path with constant speed, i.e., $\left\|S_2(\gamma)_{0,t}\right\| = t\left\|S_2(\gamma)_{0,1}\right\|$ for $0 < t < 1$ and a minimizing path γ, see [5, Theorem 7.33]. As a homogeneous norm, the Carnot-Caratheodory norm is equivalent to the simpler norm

$$\|\mathbf{x}\|_2 := \max\left(|x|, |a|^{1/2}\right), \quad \mathbf{x} = 1 + x + a \in G^2(\mathbb{R}^d),$$

see [5, Theorem 7.45].

Remark 2.6. *In* [2, Theorem 1]*, the moment condition is relaxed, which gives weak convergence in α-Hölder norm for all $\alpha < \alpha^*$, for some $\alpha^* < 1/2$, which is related to the relaxed moment condition. In this paper, we shall always assume existence of all the moments. We note, however, that we could also relax this assumption, obtaining a similar result.*

3. The main result

Usually, a cubature formula W will not satisfy the conditions of Proposition 2.5, even if we only choose uniform meshes, because the corresponding interpolation $S\left(W^{\mathcal{D}_n}\right)_{0,\cdot}$ of the random walk Ξ^n will not be geodesic. Moreover, if we want to treat functions f which are not smooth, then we have to choose non-uniform meshes with $t_k = \frac{k^\gamma}{n^\gamma}$ for some $\gamma > m - 1$, see [7]. Therefore, we want to generalize Proposition 2.5 in two directions. We want to get rid of the condition of geodesic interpolation, and we want to generalize to non-uniform meshes. Fortunately, the first generalization is simple, at least for the cubature formulas actually suggested in the literature, see Example 2.2, 2.4 and also for the Wong-Zakai approximation given in Example 2.3. The second generalization, however, requires us to change the method of proof as compared to [2].

It is natural to impose some restriction on the behavior of $S\left(W^{\mathcal{D}_n}\right)_{0,\cdot}$ between two nodes of the random walk. Indeed, we have to rule out "loops" which approach infinity.

Assumption 3.1. *The cubature formula W takes values in the Cameron-Martin space $\mathcal{H}$ (of paths started at* 0*) and the Cameron-Martin norm has finite moments of all orders, i.e., for every $k \in \mathbb{N}$*

$$E\left[\|W\|_{\mathcal{H}}^k\right] = E\left[\left(\int_0^1 \left|\dot{W}(s)\right|^2 ds\right)^{k/2}\right] < \infty.$$

This assumption is both natural (a general continuous path of finite 1-variation would not be in the (1/2-ϵ)-Hölder support of the Wiener measure!) and satisfied by all (piecewise linear!) cubature formulas used in practice. We look at this in some detail in

Example 3.2. *Assume that the cubature formula W is piecewise linear, i.e., there is a positive integer ℓ and there are d-dimensional random variables $F_1, \ldots, F_\ell$ with finite moments of all orders and a mesh $0 = s_0 < \cdots < s_\ell = 1$ such that*

$$\dot{W}_s = F_l, \quad s_{l-1} < s \le s_l,\ 1 \le l \le \ell.$$

This immediately implies Assumption 3.1.

Our main theorem is (for conclusions to cubature see Corollary 3.5 below):

Theorem 3.3. *Given a cubature formula W of order $m \ge 2$ such that W_1 and the corresponding area A_1 have finite moments of all orders and Assumption 3.1 is satisfied. Then Donsker's theorem holds in rough path topology for any sequence $\mathcal{D}_n$ of meshes with $|\mathcal{D}_n| \to 0$, i.e.,*

$$S_2\left(W^{\mathcal{D}_n}\right)_{0,\cdot} \xrightarrow[n\to\infty]{} S_2(\mathbf{B})_{0,\cdot}$$

in $C^{0,\alpha-Höl}([0,1], G^2(\mathbb{R}^d))$, for every $\alpha < 1/2$.

The natural conclusion from Theorem 3.3 would be a weak convergence result for the cubature-approximation of the SDE (3) to its true solution on path-space. A little care is necessary, however, because we have ignored the drift V_0 in the SDE, i.e., our driving signal is a pure Brownian motion and does not include time. The classical approach is to add another component to both the Brownian motion and the approximating cubature paths by setting $B^0_t := t$, $W^0_t := t$ and then require the moment matching condition (4) to hold for all iterated integrals, where the multi-index $(i_1, \ldots, i_k)$ now varies over $\{0, 1, \ldots, d\}^k$, i.e., where we also consider mixed iterated integrals of Brownian motion and time t. Due to the scaling of Brownian motion "$dB_t \approx \sqrt{dt}$", it is only necessary to impose the moment matching condition for multi-indices $(i_1, \ldots, i_k)$ with $k + \#\{j | i_j = 0\} \le m$ to get weak convergence with rate $\frac{m-1}{2}$.

However, the Ninomiya-Victoir scheme does not fall into this class, because we have seen in Example 2.4 that they do not choose $W^0_t \equiv t$. Therefore, we want to generalize the above considerations slightly. Let $h : [0,1] \to \mathbb{R}$ be a deterministic, uniformly Lipschitz path with $h(0) = 0$ and $h(1) = 1$. This setting obviously includes the drift-component of the Ninomiya-Victoir scheme. We define the path ${}^hW$ by ${}^hW^i_t := W^i_t$ for $i = 1, \ldots, d$, and ${}^hW^0_t := h(t)$. As usual, we set $B^0_t := t$. We assume the usual moment matching condition to hold, i.e.,

$$E\left[S_m({}^hW)_{0,1}\right] = E\left[S_m(B)_{0,1}\right],$$

where $S_m({}^hW)$ is the step-m signature of the path ${}^hW$, more precisely

$$S_m\left({}^hW\right)_{0,1} = \sum_{k=0}^{m} \sum_{\substack{(i_1,\ldots,i_k)\in\{0,1,\ldots,d\}^k \\ k+\#\{j\,:\,i_j=0\}\le m}} \int_{0\le t_1\le\cdots\le t_k\le 1} d^hW^{i_1}_{t_1} \cdots d^hW^{i_k}_{t_k}\, e_{i_1} \otimes \cdots \otimes e_{i_k}.$$

Analogously, the signature of the Brownian motion above is understood as the signature of the now $\mathbb{R}^{d+1}$-valued process B. (This notation is ambiguous. In the following, the symbol B will usually denote the $\mathbb{R}^d$-valued Brownian motion. We only mean the extended $\mathbb{R}^{d+1}$-valued process if specifically indicated.) We note that the signatures of the $(d+1)$-dimensional processes take their values in a stratified Lie group denoted by $G^m_1(\mathbb{R}^d)$. In a similar fashion as above, we obtain – by rescaling and concatenation – a stochastic process ${}^hW^{\mathcal{D}}$ along a grid $\mathcal{D}$. Of course, we have to use a different rescaling for the component ${}^hW^0$. Indeed, following the construction in Section 2, we define $\delta_{\sqrt{\Delta t}}({}^hW) : [0, \Delta t] \to \mathbb{R}^{d+1}$ by $\delta_{\sqrt{\Delta t}}({}^hW)^i_s = \sqrt{\Delta t} W^i_{s/\Delta t}$ for $i = 1, \ldots, d$, as before, but $\delta_{\sqrt{\Delta t}}({}^hW)^0_s = \Delta t\, h(s/\Delta t)$. We continue to construct ${}^hW^{\mathcal{D}}$ by concatenation.

By rough path theory (continuity of Young pairing, e.g. [5, Section 9.4.4]) we also obtain weak convergence in path-space for the extended process ${}^hW$ to the extended, $\mathbb{R}^{d+1}$-valued Brownian motion, which even holds for any truncated signature – not only for the step-2 signature.

Corollary 3.4. *Let W be cubature formula on Wiener space of order $m \geq 2$ satisfying Assumption 3.1 and such that W_1 and the corresponding area A_1 have finite moments of all orders. Then Donsker's theorem holds in rough path topology for any sequence $\mathcal{D}_n$ of meshes with $|\mathcal{D}_n| \to 0$, i.e.,*

$$S_N\left({}^hW^{\mathcal{D}_n}\right)_{0,\cdot} \xrightarrow[n\to\infty]{} S_N(B)_{0,\cdot}$$

in $C^{0,\alpha-H\ddot{o}l}([0,1], G_1^N(\mathbb{R}^d))$, for every $\alpha < 1/2$ and any $N \geq 1$.

Moreover, we define ${}^hX^{\mathcal{D}}$ as the solution to the (random) ODE

$$d^hX_t^{\mathcal{D}} = V_0\left({}^hX_t^{\mathcal{D}}\right)dh^{\mathcal{D}_n}(t) + \sum_{i=1}^{d} V_i\left({}^hX_t^{\mathcal{D}}\right)dW_t^{\mathcal{D},i}, \tag{11}$$

where $h^{\mathcal{D}_n} := {}^hW^{\mathcal{D}_n,0}$. Then we have weak convergence of ${}^hX^{\mathcal{D}_n}$ to X on path-space.

Corollary 3.5. *Given a bounded, continuous functional $f : C^{0,\alpha-H\ddot{o}l}([0,1], \mathbb{R}^N) \to \mathbb{R}$, and assume that W, h and $\mathcal{D}_n$ satisfy the assumptions of Corollary 3.4. Then we have*

$$E\left[f\left({}^hX^{\mathcal{D}_n}\right)\right] \xrightarrow[n\to\infty]{} E[f(X)],$$

where X denotes the path $(X_t)_{t\in[0,1]}$ of the true solution of the SDE (3) and ${}^hX^{\mathcal{D}_n}$ denotes the pathwise solution of the ODE (11).

Proof. We interpret (7) as a rough differential equation, i.e., for a given (rough) path $\boldsymbol{\omega} \in C^{0,\alpha-\text{Höl}}([0,1]; G_1^2(\mathbb{R}^d))$ with $\omega = \pi_1(\boldsymbol{\omega})$ we define $\pi(\boldsymbol{\omega})_t := y_t$ by

$$dy_t = V_0(y_t)d\omega_t^0 + \sum_{i=1}^{d} V_i(y_t)d\omega_t^i.$$

In particular, we have ${}^hX^{\mathcal{D}_n} = \pi\left(S_2({}^hW^{\mathcal{D}_n})_{0,\cdot}\right)$ and $X = \pi(S_2(B)_{0,\cdot})$. By [5, Theorem 10.26], the map $\boldsymbol{\omega} \mapsto \pi(\boldsymbol{\omega})_\cdot$ is a continuous map from $C^{0,\alpha-\text{Höl}}([0,1]; G_1^2(\mathbb{R}^d))$ to $C^{0,\alpha-\text{Höl}}([0,1]; \mathbb{R}^N)$. Thus, Corollary 3.4 implies weak convergence of ${}^hX^{\mathcal{D}_n} = \pi\left(S_2({}^hW^{\mathcal{D}_n})_{0,\cdot}\right)$ to $X = \pi(S_2(B)_{0,\cdot})$ in $C^{0,\alpha-\text{Höl}}([0,1]; \mathbb{R}^N)$. □

Remark 3.6. *Since the α-Hölder topology is stronger than the usual uniform topology given by the supremum norm, Corollary 3.5 in particular holds for all bounded functionals f which are continuous in the uniform topology on path space. In the case of unbounded continuous functionals, convergence can still be guaranteed provided that some uniform integrability property holds. (Of course, in the case of call-option type derivatives, one could also try a relevant put-call-parity.) Finally, in the case of barrier options, the payoff functional is often continuous apart from a set of measure zero on path space. Naturally, non-continuities on null-sets do not hinder weak convergence of the cubature method.*

4. Random walks with independent, non-identically distributed increments

In this section, we prepare the main ingredients of a proof of Donsker's theorem for random walks with independent, but not identically distributed increments on the Lie group $G := G^2(\mathbb{R}^d)$. More precisely, let ξ be a random variable with values in G with finite moments of all orders. We shall denote the components of ξ in the basis of $\mathfrak{g} := \mathfrak{g}^2(\mathbb{R}^d)$ given by e_i, $1 \leq i \leq d$, together with $[e_i, e_j]$, $1 \leq i < j \leq d$, by X^i and $A^{i,j}$, respectively, i.e.,

$$\xi = \exp\left(\sum_{i=1}^{d} X^i e_i + \sum_{i<j} A^{i,j}[e_i, e_j]\right).$$

Thus, the condition that ξ has finite moments of all orders simply means that all the real random variables X^i, $A^{i,j}$ have finite moments of all orders $q \geq 1$. Moreover, we assume that ξ is *centered*, i.e.,

$$E[X^i] = 0, \quad i = 1, \dots, d.$$

Let us fix a mesh $\mathcal{D}_n = \{0 = t_0 < \cdots < t_n = 1\}$. For n independent copies $\xi_{(1)}, \dots, \xi_{(n)}$ of ξ, define $\xi^n_k = \delta_{\sqrt{\Delta t_k}}(\xi_{(k)})$ and the corresponding random walk

$$\Xi^n_0 = 1, \quad \Xi^n_k = \Xi^n_{k-1} \otimes \xi^n_k, \ k = 1, \dots, n.$$

For use in the next lemma, let us define the coordinate mappings x^i (mapping $x \in G$ to the component of $\log(x)$ with respect to the basis element e_i) and $x^{i,j}$ (mapping $x \in G$ to the component of $\log(x)$ with respect to the basis element $[e_i, e_j]$), $1 \leq i \leq d$, $i < j \leq d$. As usual, the corresponding vector-fields (i.e., basis of the tangent space) are denoted by $\frac{\partial}{\partial x^i}$ and $\frac{\partial}{\partial x^{i,j}}$, respectively.

Lemma 4.1. *The above random walk satisfies the central limit theorem, i.e., Ξ^n_n converges weakly to the Gaussian measure with infinitesimal generator*

$$\sum_{i<j} a^{i,j} \frac{\partial}{\partial x^{i,j}} + \frac{1}{2} \sum_{i \leq j} b^{i,j} \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j},$$

where $a^{i,j} := E[A^{i,j}]$ and $b^{i,j} := \mathrm{Cov}(X^i, X^j)$.[4]

Proof. The result is well-known in probability theory on Lie groups, see, e.g., [14]. We verify that the system of probability measures $\mu_{n,k} = (\xi^n_k)_* P$, i.e., $\mu_{n,k}$ is the law of ξ^n_k, satisfies the conditions given in [14, Theorem 3.2], namely:

(i) $\sup_n \sum_{k=1}^n \int_G \|x\|^2 \mu_{n,k}(dx) < \infty$;
(ii) $\mu_{n,k}$ is centered in the above sense;
(iii) for every $1 \leq i < j \leq d$, there is a number $a^{i,j} \in \mathbb{R}$ such that

$$a^{i,j} = \lim_{n\to\infty} \sum_{k=1}^n \int_G x^{i,j}(x) \mu_{n,k}(dx);$$

(iv) for every $1 \leq i, j \leq d$, there is a number $b^{i,j} \in \mathbb{R}$ such that

$$b^{i,j} = \lim_{n\to\infty} \sum_{k=1}^n \int_G x^i(x) x^j(x) \mu_{n,k}(dx);$$

(v) $\lim_{n\to\infty} \sum_{k=1}^n \int_{\|x\|\geq\epsilon} \|x\|^2 \mu_{n,k}(dx) = 0$ for all $\epsilon > 0$.

By homogeneity of the Carnot-Caratheodory norm, we have

$$\int_G \|x\|^2 \mu_{n,k}(dx) = E\left[\left\|\delta_{\sqrt{\Delta t_k}}(\xi)\right\|^2\right] = \Delta t_k E\left[\|\xi\|^2\right].$$

Thus,

$$\sum_{k=1}^n \int_G \|x\|^2 \mu_{n,k}(dx) = \sum_{k=1}^n \Delta t_k E\left[\|\xi\|^2\right] = E\left[\|\xi\|^2\right],$$

and the supremum over n is obviously finite, settling (i).

(ii) is satisfied by assumption on ξ. Regarding (iii), note that $x^{i,j}(\xi^n_k) = x^{i,j}(\delta_{\sqrt{\Delta t_k}}(\xi)) = \Delta t_k A^{i,j}$, where equality is understood as equality in law. Therefore, (iii) is satisfied with $a^{i,j} = E(A^{i,j}) < \infty$. A similar argument shows that (iv) holds with $b^{i,j} = \mathrm{Cov}(X^i, X^j)$.

For the proof of (v), we again use homogeneity of the Carnot-Caratheodory norm. Indeed, we have

$$\int_{\|x\|\geq\epsilon} \|x\|^2 \mu_{n,k}(dx) = \Delta t_k E\left[\mathbf{1}_{]\epsilon,\infty[}(\sqrt{\Delta t_k}\,\|\xi\|)\,\|\xi\|^2\right],$$

[4] The statement means that there is a semi-group $(\mu_t)_{t\geq 0}$ of probability measures on G having the above infinitesimal generator and μ_1 is the limiting distribution of Ξ^n_n. Moreover, this semi-group is Gaussian in the sense that $\lim_{t\searrow 0} \frac{1}{t}\mu_t(G \setminus U) = 0$ for every neighborhood U of the neutral element of the group G.

implying that

$$\sum_{k=1}^{n} \int_{\|x\| \geq \epsilon} \|x\|^2 \mu_{n,k}(dx) \leq E\left[\mathbf{1}_{\left]\frac{\epsilon}{\sqrt{|\mathcal{D}_n|}}, \infty\right[} \|\xi\|^2 \right] \leq \sqrt{P\left(\|\xi\| > \frac{\epsilon}{\sqrt{|\mathcal{D}_n|}} \right)} \cdot \sqrt{E\left[\|\xi\|^4 \right]},$$

by the Cauchy-Schwarz inequality. Now, the right hand side converges to zero by integrability of $\|\xi\|$ and $|\mathcal{D}_n| \to 0$, for every fixed $\epsilon > 0$. □

Remark 4.2. *If ξ is the step-2 signature of a cubature formula of degree $m \geq 2$, then $a^{i,j} = E[A^{i,j}] = 0$, $1 \leq i < j \leq d$, and, moreover, $b^{i,j} = \mathrm{Cov}(X^i, X^j) = \delta_{ij}$. Thus, the generator of the limiting Gaussian measure in Lemma 4.1 coincides with the generator of the Brownian motion on G, i.e., with the generator of* $\mathbf{B}$.

Next we state a moment estimate, which will enable us to prove tightness of the family of interpolated random walks in rough path topology.

Proposition 4.3. *For every $p \in \mathbb{N}$, $p \geq 1$ we can find a constant C independent of k and n such that*

$$E\left[\left\| \Xi_k^n \right\|^{4p} \right] \leq C t_k^{2p}.$$

Proof. The proof heavily relies on *Burkholder's inequality*, see [3]. Recall that the discrete time Burkholder inequality establishes the existence of constants c_p, C_p for $1 < p < \infty$ such that for every p-integrable real martingale Y_n and any $n \in \mathbb{N}$ we have

$$c_p \sup_n \|S_n\|_{L^p} \leq \sup_n \|Y_n\| \leq C_p \sup_n \|S_n\|_{L^p}$$

where, setting $Y_0 := 0$, $S_n := \sqrt{\sum_{k=1}^n (Y_k - Y_{k-1})^2}$ is the square root of the quadratic variation of Y. By choosing $Y_{n+l} \equiv Y_n$ for $l > 0$, this immediately implies the corresponding finite version

$$c_p \|S_n\|_{L^p} \leq \|Y_n\|_{L^p} \leq C_p \|S_n\|_{L^p}. \tag{12}$$

By equivalence of homogeneous norms, see, for instance, [5, Theorem 7.44], we can replace the Carnot-Caratheodory norm $\|\cdot\|$ on $G^2(\mathbb{R}^d)$ by the homogeneous norm

$$\|\mathbf{x}\|_2 := \max\left(\left|\pi_1(\log(\mathbf{x}))\right|, \sqrt{\left|\pi_2(\log(\mathbf{x}))\right|} \right) \leq \left|\pi_1(\log(\mathbf{x}))\right| + \sqrt{\left|\pi_2(\log(\mathbf{x}))\right|}, \quad \mathbf{x} \in G^2(\mathbb{R}^d),$$

where π_1 and π_2 denote the projection to the first and second level components of $\mathbf{x}$, i.e., when $\mathbf{x} = 1 + x + a \in G^2(\mathbb{R}^d)$, then $\pi_1(\mathbf{x}) = x \in \mathbb{R}^d$ and $\pi_2(\mathbf{x}) = a \in \mathbb{R}^d \otimes \mathbb{R}^d$. Thus, the assertion of the proposition is equivalent to the existence of a constant C (only depending on p) such that

$$E\left[\left|\pi_1(\log(\Xi_k^n))\right|^{4p} \right] \leq C t_k^{2p}, \tag{13}$$

$$E\left[\left|\pi_2(\log(\Xi_k^n))\right|^{2p} \right] \leq C t_k^{2p}. \tag{14}$$

We start by proving (13). By the Campbell-Baker-Hausdorff formula, we have

$$\left|\pi_1(\log(\Xi_k^n))\right| = \left| \sum_{i=1}^d \left(\sum_{l=1}^k X_l^i \right) e_i \right| \leq C \sum_{i=1}^d \left| \sum_{l=1}^k X_l^i \right|.$$

Here, $X_l^i = \sqrt{\Delta t_l} X_{(l)}^i$ and C is a constant, which does neither depend on the partition $\mathcal{D}_n$ nor on k. For the remainder of the proof, we will use this symbol for constants that may vary from line to line, but do not depend on $\mathcal{D}_n$ or on k. This implies that

$$E\left[\left|\pi_1(\log(\Xi_k^n))\right|^{4p} \right] \leq C \sum_{i=1}^d E\left[\left| \sum_{l=1}^k X_l^i \right|^{4p} \right].$$

Now we apply Burkholder's inequality to the martingale $Y_k = \sum_{l=1}^k X_l^i$ with the exponent $4p$ to get

$$E\left[\left|\pi_1(\log(\Xi_k^n))\right|^{4p}\right] \le C\sum_{i=1}^d E\left[\left|\sum_{l=1}^k (X_l^i)^2\right|^{2p}\right] = Ct_k^{2p}\sum_{i=1}^d E\left[\left|\sum_{l=1}^k \frac{\Delta t_l}{t_k}(X_{(l)}^i)^2\right|^{2p}\right].$$

Noting that the sum inside the expectation is a convex combination, we apply Jensen's inequality for the convex function x^{2p} and get

$$E\left[\left|\pi_1(\log(\Xi_k^n))\right|^{4p}\right] \le Ct_k^{2p}\sum_{i=1}^d\sum_{l=1}^k \frac{\Delta t_l}{t_k}E\left[(X_{(l)}^i)^{4p}\right] = C\left(\sum_{i=1}^d E\left[(X_{(l)}^i)^{4p}\right]\right)t_k^{2p},$$

which is of the form required in (13).

For (14), we again start with the Campbell-Baker-Hausdorff formula and get

$$\begin{aligned}
E\left[\left|\pi_2(\log(\Xi_k^n))\right|^{2p}\right] &= E\left[\left|\sum_{1\le i<j\le d}\left[\sum_{l=1}^k A_l^{i,j} + \frac{1}{2}\sum_{1\le l_1<l_2\le k}\left(X_{l_1}^i X_{l_2}^j - X_{l_1}^j X_{l_2}^i\right)\right][e_i,e_j]\right|^{2p}\right]\\
&\le C\sum_{1\le i<j\le d} E\left[\left|\sum_{l=1}^k A_l^{i,j} + \frac{1}{2}\sum_{1\le l_1<l_2\le k}\left(X_{l_1}^i X_{l_2}^j - X_{l_1}^j X_{l_2}^i\right)\right|^{2p}\right]\\
&\le C\sum_{1\le i<j\le d}\left(E\left[\left|\sum_{l=1}^k A_l^{i,j}\right|^{2p}\right] + E\left[\left|\sum_{1\le l_1<l_2\le k} X_{l_1}^i X_{l_2}^j\right|^{2p}\right] + E\left[\left|\sum_{1\le l_1<l_2\le k} X_{l_1}^j X_{l_2}^i\right|^{2p}\right]\right). \qquad (15)
\end{aligned}$$

Now fix some $1 \le i < j \le d$. Again by Jensen's inequality, we have

$$E\left[\left|\sum_{l=1}^k A_l^{i,j}\right|^{2p}\right] = t_k^{2p}E\left[\left|\sum_{l=1}^k \frac{\Delta t_l}{t_k}A_{(l)}^{i,j}\right|^{2p}\right] \le t_k^{2p}\sum_{l=1}^k\frac{\Delta t_l}{t_k}E\left[\left|A_{(l)}^{i,j}\right|^{2p}\right] = E\left[\left|A_{(1)}^{i,j}\right|^{2p}\right]t_k^{2p}. \qquad (16)$$

On the other hand, note that

$$\sum_{1\le l_1<l_2\le k} X_{l_1}^i X_{l_2}^j = \sum_{l_2=1}^k X_{l_2}^j\left(\sum_{l_1=1}^{l_2-1} X_{l_1}^i\right)$$

is a martingale (indexed by k). Thus, Burkholder's inequality (12) for the exponent $2p$ gives

$$E\left[\left|\sum_{1\le l_1<l_2\le k} X_{l_1}^i X_{l_2}^j\right|^{2p}\right] \le CE\left[\left|\sum_{l_2=1}^k (X_{l_2}^j)^2\left(\sum_{l_1=1}^{l_2-1}X_{l_1}^i\right)^2\right|^p\right] = Ct_k^pE\left[\left|\sum_{l_2=1}^k\frac{\Delta t_{l_2}}{t_k}(X_{(l_2)}^j)^2\left(\sum_{l_1=1}^{l_2-1}X_{l_1}^i\right)^2\right|^p\right].$$

Now we again apply Jensen's inequality and then the Cauchy-Schwarz inequality, and obtain

$$\begin{aligned}
E\left[\left|\sum_{1\le l_1<l_2\le k} X_{l_1}^i X_{l_2}^j\right|^{2p}\right] &\le Ct_k^p\sum_{l_2=1}^k\frac{\Delta t_{l_2}}{t_k}E\left[\left|(X_{(l_2)}^j)^2\left(\sum_{l_1=1}^{l_2-1}X_{l_1}^i\right)^2\right|^p\right]\\
&\le Ct_k^p\sum_{l_2=1}^k\frac{\Delta t_{l_2}}{t_k}\left(E\left[\left|X_{(l_2)}^j\right|^{4p}\right]\right)^{1/2}\left(E\left[\left|\sum_{l_1=1}^{l_2-1}X_{l_1}^i\right|^{4p}\right]\right)^{1/2}. \qquad (17)
\end{aligned}$$

By applying Burkholder's and Jensen's inequalities for a final time, we get for the left-most term in the above inequality

$$E\left[\left|\sum_{l_1=1}^{l_2-1} X^i_{l_1}\right|^{4p}\right] \le C t_{l_2-1}^{2p} E\left[\left|\sum_{l_1=1}^{l_2-1} \frac{\Delta t_{l_1}}{t_{l_2-1}} (X^i_{(l_1)})^2\right|^{2p}\right]$$
$$\le C t_{l_2-1}^{2p} \sum_{l_1=1}^{l_2-1} \frac{\Delta t_{l_1}}{t_{l_2-1}} E\left[\left|X^i_{(l_1)}\right|^{4p}\right] \le C t_{l_2-1}^{2p} E\left[\left|X^i_{(1)}\right|^{4p}\right].$$

Inserting the last inequality into (17), we obtain

$$E\left[\left|\sum_{1\le l_1<l_2\le k} X^i_{l_1} X^j_{l_2}\right|^{2p}\right] \le C t_k^p \sum_{l_2=1}^{k} \frac{\Delta t_{l_2}}{t_k} t_{l_2-1}^p \left(E\left[\left|X^j_{(1)}\right|^{4p}\right]\right)^{1/2} \left(E\left[\left|X^i_{(1)}\right|^{4p}\right]\right)^{1/2}$$
$$\le C \left(E\left[\left|X^j_{(1)}\right|^{4p}\right]\right)^{1/2} \left(E\left[\left|X^i_{(1)}\right|^{4p}\right]\right)^{1/2} t_k^{2p}.$$

Together with (15) and (16) this shows (14), and the proposition follows. □

5. Proof of the main results

Analogously to [2, Theorem 1] we can now state our

Theorem 5.1. *Let $\mathcal{D}_n = \{0 = t_0 < \cdots < t_n = 1\}$ be a sequence of meshes with $|\mathcal{D}_n| \to 0$ and let $\Xi^n = \left(\Xi^n_k\right)_{k=0}^n$ be a centered random walk in $G^2(\mathbb{R}^d)$ along the mesh (as defined in Section 4), whose increments have moments of all orders. Additionally, we impose $E(\pi_2(\xi^n_k)) = 0$. Define a sequence $\overline{\Xi}^n$ of stochastic processes with values in $G^2(\mathbb{R}^d)$ by $\overline{\Xi}^n_{t_k} = \Xi^n_k$ for $k = 0, \dots, n$ and by* geodesic *interpolation for $t \in [t_k, t_{k+1}]$. Then*

$$\overline{\Xi}^n \xrightarrow[n\to\infty]{} \mathbf{B}$$

in $C^{0,\alpha-H\ddot{o}l}\left([0,1], G^2(\mathbb{R}^d)\right)$ for all $\alpha < 1/2$.

Proof. Kolmogorov's tightness criterion, see, for instance, [15], implies that $\overline{\Xi}^n$ is tight (in $C^{0,\alpha-\text{Höl}}$) provided that for any $u, v \in [0,1]$

$$\sup_n E\left[d\left(\overline{\Xi}^n_v, \overline{\Xi}^n_u\right)^a\right] \le c\,|v-u|^{1+b}, \tag{18}$$

where a, b, c are positive constants with $\alpha < b/a$ and d denotes the Carnot-Caratheodory distance defined by $d(x,y) = \left\|x^{-1}y\right\|$. We choose $a = 4p$ and $b = 2p-1$, then Proposition 4.3 implies that (18) holds for $u, v \in \mathcal{D}_n$. For arbitrary $v < u$, assume that $t_i \le v < t_{i+1}$ and $t_j \le u < t_{j+1}$ (for $t_i, t_{i+1}, t_j, t_{j+1} \in \mathcal{D}_n$). Using (by the geodesic interpolation)

$$d\left(\overline{\Xi}^n_v, \overline{\Xi}^n_{t_{i+1}}\right) = \frac{t_{i+1}-v}{\Delta t_{i+1}} d\left(\overline{\Xi}^n_{t_i}, \overline{\Xi}^n_{t_{i+1}}\right), \quad d\left(\overline{\Xi}^n_{t_j}, \overline{\Xi}^n_u\right) = \frac{u-t_j}{\Delta t_{j+1}} d\left(\overline{\Xi}^n_{t_j}, \overline{\Xi}^n_{t_{j+1}}\right) \tag{19}$$

and the triangle inequality, we obtain

$$E\left[d\left(\overline{\Xi}^n_v, \overline{\Xi}^n_u\right)^a\right] \le \tilde{c}\left[\left(\frac{t_{i+1}-v}{\Delta t_{i+1}}\right)^{2p} (t_{i+1}-v)^{2p} + (t_j - t_{i+1})^{2p} + \left(\frac{u-t_j}{\Delta t_{j+1}}\right)^{2p} (u-t_j)^{2p}\right]$$
$$\le c\,|u-v|^{2p},$$

for some constant c only depending on p.

This shows that the sequence of stochastic processes $\overline{\Xi}^n$ is tight in rough-path-topology. Moreover, Lemma 4.1 shows that the finite-dimensional marginal distributions of $\overline{\Xi}^n$ converge to those of $\mathbf{B}$. Thus, we obtain the theorem for $\alpha < \frac{4p}{2p-1}$ and, with $p \to \infty$, for any $\alpha < 1/2$. □

Proof of Theorem 3.3. In Theorem 5.1 above, we have already proved our main result for $N = 2$ and for the special case of geodesic interpolation, i.e., for the case that $S_2\left(W^{\mathcal{D}_n}\right)_{0,t}$, $t \in [0,1]$, provides a geodesic interpolation between the grid points, i.e., between $S_2\left(W^{\mathcal{D}_n}\right)_{0,t_k}$, $t_k \in \mathcal{D}_n$. We note that the extension of the result to $N > 2$ is immediate since the Lyons lift $S_N : G^2(\mathbb{R}^d) \to G^N(\mathbb{R}^d)$ is continuous in rough path topology, see, for instance, [5, Corollary 9.11].

We need to give a proof for the case $N = 2$ but without geodesic interpolation. As before, assume that we are given $t_i \le v < t_{i+1}$, $t_j \le u < t_{j+1}$, with $t_i, t_j, t_{i+1}, t_{j+1} \in \mathcal{D}_n$. For any path $\omega \in \mathcal{H}$, the Cameron-Martin space, we have, see [5, Proposition 15.7],

$$|\omega|_{1-\text{var};[s,t]} \le \sqrt{|t-s|}\, \|\omega\|_{\mathcal{H};[s,t]},$$

where $|\cdot|_{1-\text{var};[s,t]}$ denotes the first variation of a path restricted to $[s,t]$ and $\|\cdot\|_{\mathcal{H};[s,t]}$ denotes the Cameron-Martin norm, likewise restricted to $[s,t]$, i.e.,

$$\|\omega\|^2_{\mathcal{H};[s,t]} = \int_s^t |\dot{\omega}(u)|^2\, du.$$

Notice that

$$\left\|W^{\mathcal{D}_n}\right\|^2_{\mathcal{H};[t_j,u]} = \int_{t_j}^u \left|\dot{W}^{\mathcal{D}_n}_t\right|^2 dt = \frac{1}{\Delta t_{j+1}} \int_{t_j}^u \left|\dot{W}\left(\frac{t-t_j}{\Delta t_{j+1}}\right)\right|^2 dt \le \int_0^1 \left|\dot{W}_t\right|^2 dt = \|W\|^2_{\mathcal{H}}.$$

Therefore, we can bound

$$\begin{aligned} d\left(S_2\left(W^{\mathcal{D}_n}\right)_{0,t_j}, S_2\left(W^{\mathcal{D}_n}\right)_{0,u}\right) &= \left\|S_2\left(W^{\mathcal{D}_n}\right)_{t_j,u}\right\| \\ &\le \left|W^{\mathcal{D}_n}\right|_{1-\text{var};[t_j,u]} \\ &\le \sqrt{u-t_j}\, \|W\|_{\mathcal{H}}. \end{aligned}$$

By Assumption 3.1, we get

$$E\left[d\left(S_2\left(W^{\mathcal{D}_n}\right)_{0,t_j}, S_2\left(W^{\mathcal{D}_n}\right)_{0,u}\right)^{4p}\right] \le (u-t_j)^{2p} E\left[\|W\|^{4p}_{\mathcal{H}}\right] \le C(u-t_j)^{2p}, \tag{20}$$

and similarly

$$E\left[d\left(S_2\left(W^{\mathcal{D}_n}\right)_{0,v}, S_2\left(W^{\mathcal{D}_n}\right)_{0,t_{i+1}}\right)^{4p}\right] \le C(t_{i+1}-v)^{2p}. \tag{21}$$

Therefore, we can show tightness of the sequence of processes $S_2\left(W^{\mathcal{D}_n}\right)_{0,\cdot}$ as in the proof of Theorem 5.1 above. □

Now we finally return to the Donsker theorem for cubature paths with an adjourned, Lipschitz component h.

Proof of Corollary 3.4. Let $h^{\mathcal{D}_n} := {}^hW^{\mathcal{D}_n,0}$ denote the 0-component of ${}^hW^{\mathcal{D}_n}$. Moreover, let $\mathrm{id} : [0,1] \to [0,1]$ denote the identity, $\mathrm{id}(t) = t$. We note that $h^{\mathcal{D}_n}$ converges to id in $C^{0,\beta-\text{Höl}}([0,1];\mathbb{R})$ for any $\beta < 1$. Indeed, let $0 < t < 1$ and let $t_i \le t < t_{i+1}$ denote the grid points closest to t. Then apparently

$$h^{\mathcal{D}_n}(t) = t_i + \Delta t_{i+1} h\left(\frac{t-t_i}{\Delta t_{i+1}}\right).$$

For $h(0) = 0$, we get $|h(t)| \le L|t|$, where L is the Lipschitz constant of h. From this one can easily conclude that $\left\|h^{\mathcal{D}_n} - \mathrm{id}\right\|_\infty \le (1+L)|\mathcal{D}_n|$ and $h^{\mathcal{D}_n}$ converges to id uniformly on $[0,1]$ with uniform Lipschitz bounds for $|\mathcal{D}_n| \to 0$. This implies the convergence in β-Hölder topology for any $\beta < 1$.

By this result together with Theorem 3.3 for the convergence of $S_2(W)$, we can immediately conclude that $S_2\left({}^hW^{\mathcal{D}_n}\right)_{0,\cdot}$ converges to $S_2(B)_{0,\cdot}$ in $C^{0,\alpha-\text{Höl}}\left([0,1]; G^2_1(\mathbb{R}^d)\right)$. By continuity of the Lyons lift $S_2(x) \mapsto S_N(x)$ in rough path topology, the statement of the corollary follows. □

References

[1] Christian Bayer, Peter K. Friz, and Ronnie Loeffen. Semi-closed form cubature and applications to financial diffusion models. *Quantitative Finance*, 2013. to appear.

[2] Emmanuel Breuillard, Peter Friz, and Martin Huesmann. From random walks to rough paths. *Proc. Amer. Math. Soc.*, 137(10):3487–3496, 2009.

[3] D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probability*, 1:19–42, 1973.

[4] Peter Friz and Harald Oberhauser. Rough path limits of the Wong-Zakai type with a modified drift term. *J. Funct. Anal.*, 256(10):3236–3256, 2009.

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

[6] Shigeo Kusuoka. Approximation of expectation of diffusion process and mathematical finance. In *Taniguchi Conference on Mathematics Nara '98*, volume 31 of *Adv. Stud. Pure Math.*, pages 147–165. Math. Soc. Japan, Tokyo, 2001.

[7] Shigeo Kusuoka. Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. In *Advances in mathematical economics. Vol. 6*, volume 6 of *Adv. Math. Econ.*, pages 69–83. Springer, Tokyo, 2004.

[8] Christian Litterer and Terry Lyons. Cubature on Wiener space continued. In Jiro Akahori, Shigeyoshi Ogawa, and Shinzo Watanabe, editors, *Stochastic Processes and Applications to Mathemaical Finance*, Singapore, pages 197–218. World Scientific Publishing, 2007.

[9] Terry Lyons and Nicolas Victoir. Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 460(2041):169–198, 2004. Stochastic analysis with applications to mathematical finance.

[10] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[11] E. J. McShane. Stochastic differential equations and models of random processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory*, pages 263–294, Berkeley, Calif., 1972. Univ. California Press.

[12] Mariko Ninomiya and Syoiti Ninomiya. A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method. *Finance Stoch.*, 13(3):415–443, 2009.

[13] Syoiti Ninomiya and Nicolas Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance*, 15(1-2):107–121, 2008.

[14] Gyula Pap. Central limit theorems on nilpotent Lie groups. *Probab. Math. Statist.*, 14(2):287–312, 1993.

[15] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[16] Daniel W. Stroock and S. R. S. Varadhan. Limit theorems for random walks on Lie groups. *Sankhyā Ser. A*, 35(3):277–294, 1973.

WIAS Berlin, Mohrenstrasse 39, 10117 Berlin, Germany
E-mail address: christian.bayer@wias-berlin.de

TU Berlin, Institute of Mathematics, MA 7-2, Strasse des 17. Juni 136, 10623 Berlin, Germany

WIAS Berlin, Mohrenstrasse 39, 10117 Berlin, Germany
E-mail address: friz@math.tu-berlin.de, peter.friz@wias-berlin.de