Electronic Journal of Probability

Electron. J. Probab. **29** (2024), article no. 111, 1–56.
ISSN: 1083-6489 https://doi.org/10.1214/24-EJP1174

Almost sure diffusion approximation in averaging via rough paths theory*

Peter K. Friz[†] Yuri Kifer[‡]

Abstract

The paper deals with the fast-slow motions setups in the continuous time $\frac{dX^\varepsilon(t)}{dt} = \frac{1}{\varepsilon}\sigma(X^\varepsilon(t))\xi(t/\varepsilon^2) + b(X^\varepsilon(t)),\ t \in [0,T]$ and the discrete time $X_N((n+1)/N) = X_N(n/N) + N^{-1/2}\sigma(X_N(n/N))\xi(n) + N^{-1}b(X_N(n/N)),\ n = 0,1,...,[TN]$ where σ and b are smooth matrix and vector functions, respectively, ξ is a centered stationary vector stochastic process and $\varepsilon, 1/N$ are small parameters. We derive, first, estimates in the strong invariance principles for sums $S_N(t) = N^{-1/2}\sum_{0\le k<[Nt]}\xi(k)$ and iterated sums $\mathbb{S}_N^{ij}(t) = N^{-1}\sum_{0\le k<l<[Nt]}\xi_i(k)\xi_j(l)$ together with the corresponding results for integrals in the continuous time case which, in fact, yields almost sure invariance principles for iterated sums and integrals of any order and, moreover, implies laws of iterated logarithm for these objects. Then, relying on the rough paths theory, we obtain strong almost sure approximations of processes X^ε and X_N by corresponding diffusion processes Ξ^ε and Ξ_N, respectively. Previous results for the above setup dealt either with weak or moment diffusion approximations and not with almost sure approximation which is the new and natural generalization of well known works on strong invariance principles for sums of weakly dependent random variables.

Keywords: averaging; diffusion approximation; ϕ-mixing; stationary process; shifts; dynamical systems.
MSC2020 subject classifications: 34C29; 60F15; 60L20.
Submitted to EJP on August 30, 2023, final version accepted on July 16, 2024.

1 Introduction

The study of the asymptotic behavior as $\varepsilon \to 0$ of solutions X^ε of systems of ordinary differential equations having the form

$$\frac{dX^\varepsilon(t)}{dt} = \frac{1}{\varepsilon}B(X^\varepsilon(t), \xi(t/\varepsilon^2)) + b(X^\varepsilon(t),\ \xi(t/\varepsilon^2)),\ t \in [0,T] \tag{1.1}$$

*PKF acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689), and also seed support for DFG CRC/TRR 388 "Rough Analysis, Stochastic Dynamics and Related Topics".
[†]Institut für Mathematik, TU Berlin, and WIAS Berlin, Germany. E-mail: friz@math.tu-berlin.de
[‡]Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel. E-mail: kifer@math.huji.ac.il

has already more than a half century history. Here $B(\cdot,\xi(s))$ and $b(\cdot,\xi(s))$ are (random) smooth vector fields on $\mathbb{R}^d$ and ξ is a stationary process which is viewed as a fast motion while X^ε is considered as a slow motion. Assuming that

$$EB(x,\xi(0))\equiv 0 \tag{1.2}$$

for all x it was shown in a series of papers [30], [43] and [6] that X^ε converges weakly as $\varepsilon\to 0$ to a diffusion process provided ξ is sufficiently fast mixing with respect to σ-algebras generated by ξ itself. The latter condition is quite restrictive when ξ is generated by a dynamical system, i.e. when $\xi(t)=g\circ F^t$ where g is a vector function and F^t is a flow (continuous time dynamical system) preserving certain measure which makes ξ a stationary process. In order to derive weak convergence of X^ε to a diffusion for ξ built by a sufficiently large class of observables g and dynamical systems other approaches were developed recently based mainly on the rough paths theory (see e.g. [33, 34, 1, 14, 13, 20]). All above mentioned results can be obtained both in the continuous time setup (1.1) and in the discrete time setup given by the following recurrence relation

$$X^\varepsilon((n+1)\varepsilon^2)=X^\varepsilon(n\varepsilon^2)+\varepsilon B(X^\varepsilon(n\varepsilon^2),\,\xi(n))+\varepsilon^2 b(X^\varepsilon(n\varepsilon^2),\,\xi(n)) \tag{1.3}$$

where $0\le n<[T/\varepsilon^2]$ and $\xi(n),\,n\ge 0$ is a stationary sequence of random vectors. We will study (1.3) for $X_N=X^{1/\sqrt{N}}$ as $N\to\infty$. The above results can be viewed as a substantial generalization of the functional central limit theorem since when $B(x,\zeta)$ does not depend on x and $b\equiv 0$ the process X^ε weakly converges to the Brownian motion (with a covariance matrix).

Motivated by strong invariance principle results for sums (see, for instance, [9], [36], [17] and more recent [41] and [27]) the second author obtained in [31] and [32] certain results on strong L^p diffusion approximations of solutions of (1.1) and (1.3) under the condition (1.2). In this paper we are interested in strong diffusion approximation results in a more traditional form saying that X^ε (or X_N) can be redefined on a richer probability space where there exists a diffusion process Ξ^ε (or Ξ_N) such that $\varepsilon^{-\delta}\sup_{0\le t\le T}|X^\varepsilon(t)-\Xi^\varepsilon(t)|$ (or $N^\delta\max_{0\le n\le TN}|X_N(n/N)-\Xi_N(n/N)|$) remains bounded almost surely (a.s.) for all $\varepsilon>0$ (or $N\ge 1$) where $\delta>0$ does not depend on ε or N. All papers on almost sure approximations (strong invariance principles) dealt before with sums (or integrals in time) of random variables (or vectors), and so the limiting process was always a Brownian motion while almost sure diffusion approximation results do not seem to appear before in the literature. These results do not follow from [31] and [32] and, on the other hand, our methods do not yield moment estimates from these papers.

Our methods rely on the rough paths theory [39], as exposed in [24, 22], and in order to adapt the equations (1.1) and (1.3) to this setup we consider (as in [33, 14], but see Remark 2.7) a more restricted situation assuming $B(x,\xi(\cdot))=\sigma(x)\xi(\cdot)$ where $\xi(k),\,k\ge 0$ is a ϕ-mixing sequence of random vectors, $\sigma(x)$ is a smooth matrix function and $b(x,\xi(\cdot))=b(x)$ does not depend on the second variable. We obtain, first, strong invariance principles for sums $S_N(t)=N^{-1/2}\sum_{0\le k<[Nt]}\xi(k)$ and iterated sums $\mathbb{S}_N^{ij}(t)=N^{-1}\sum_{0\le k<l<[Nt]}\xi_i(k)\xi_j(l)$ in p-variation rough path sense, which is a stronger result than the standard strong invariance principle, and then, relying on a quantitative form of the local Lipschitz property of the Itô-Lyons map for càdlàg rough differential equations [25], in conjunction with a law of iterated logarithm type growth control for Brownian rough paths, we obtain our estimates for strong diffusion approximations of processes X^ε and X_N. This is in contrast to [33, 14] and subsequent works which "only" rely on continuity of the Itô-Lyons map. Similar results will be derived here in the continuous time setup (1.1) with $B(x,\xi(\cdot))=\sigma(x)\xi(\cdot)$ where $\xi(t),\,t\ge 0$ is a vector stochastic process constructed by a suspension procedure over a ϕ-mixing discrete time process. Again, we obtain first strong invariance principles for integrals $\varepsilon\int_0^{t\varepsilon^{-2}}\xi(s)ds$ and iterated integrals $\varepsilon^2\int_0^{t\varepsilon^{-2}}\xi_j(s)ds\int_0^s\xi_i(u)du$ and then rely on the rough paths machinery.

In fact, rough paths theory yields an extension to almost sure invariance principles for iterated sums and integrals from second to any order and as a byproduct of these results for weakly dependent random vectors we obtain for these objects corresponding laws of iterated logarithm. Almost sure diffusion approximations, strong invariance principles for iterated sums and laws of iterated logarithms for them never appeared before in the literature. We also note the current interest of iterated integrals and their discrete counterparts from a data science / time-series perspective [5, 21, 16].

The structure of this paper is as follows. In the next section we describe our precise setup and main results. In Sections 3 and 4 we prove the strong invariance principles in appropriate variational norms for sums and iterated sums, respectively. In Section 5 we obtain the strong invariance principles in the continuous time setup for integrals and iterated integrals. In Section 6 we provide a brief introduction to rough paths theory and show how together with the above strong invariance principles it yields our strong diffusion approximations results. In Section 7 we obtain extensions to strong invariance principles for iterated sums and integrals of any order which yields also laws of iterated logarithm for these objects.

2 Preliminaries and main results

2.1 Discrete time case

We start with the discrete time setup which consists of a complete probability space $(\Omega,\mathcal{F},P)$, a stationary sequence of e-dimensional random vectors $\xi(n)$, $-\infty<n<\infty$ and a two parameter family of countably generated σ-algebras $\mathcal{F}_{m,n}\subset\mathcal{F}$, $-\infty\le m\le n\le\infty$ such that $\mathcal{F}_{mn}\subset\mathcal{F}_{m'n'}\subset\mathcal{F}$ if $m'\le m\le n\le n'$ where $\mathcal{F}_{m\infty}=\cup_{n:\,n\ge m}\mathcal{F}_{mn}$ and $\mathcal{F}_{-\infty n}=\cup_{m:\,m\le n}\mathcal{F}_{mn}$. We will measure the dependence between σ-algebras $\mathcal{G}$ and $\mathcal{H}$ by the ϕ-coefficient defined by

$$\begin{aligned}\phi(\mathcal{G},\mathcal{H})&=\sup\{|\tfrac{P(\Gamma\cap\Delta)}{P(\Gamma)}-P(\Delta)|:\,P(\Gamma)\ne0,\,\Gamma\in\mathcal{G},\,\Delta\in\mathcal{H}\}\\&=\tfrac12\sup\{\|E(g|\mathcal{G})-Eg\|_\infty:\,g\text{ is }\mathcal{H}\text{-measurable and }\|g\|_\infty=1\}\end{aligned}\tag{2.1}$$

(see [8]) where $\|\cdot\|_\infty$ is the L^∞-norm. For each $n\ge0$ we set also

$$\phi(n)=\sup_m\phi(\mathcal{F}_{-\infty,m},\mathcal{F}_{m+n,\infty}).\tag{2.2}$$

If $\phi(n)\to0$ as $n\to\infty$ then the probability measure P is called ϕ-mixing with respect to the family $\{\mathcal{F}_{mn}\}$. Unlike [31], in order to ensure more applicability of our results to dynamical systems, we do not assume that $\xi(n)$ is $\mathcal{F}_{nn}$-measurable and instead we will work with the approximation coefficient

$$\rho(n)=\sup_m\|\xi(m)-E(\xi(m)|\mathcal{F}_{m-n,m+n})\|_\infty.\tag{2.3}$$

To save notations we will still write $\mathcal{F}_{mn}$, $\phi(n)$ and $\rho(n)$ for $\mathcal{F}_{[m][n]}$, $\phi([n])$ and $\rho([n])$, respectively, if m and n are not integers, where $[\cdot]$ denotes the integral part.

We will deal with the recurrence relation

$$X_N(n+1/N)=X_N(n/N)+\frac{1}{\sqrt N}\sigma(X_N(n/N))\xi(n)+\frac1N b(X_N(n/N))\tag{2.4}$$

and this definition is extended to all $t\in[0,T]$ by setting $X_N(t)=X_N(n/N)$ whenever $n/N\le t<(n+1)/N$. We will assume that

$$E\xi(0)=0\tag{2.5}$$

and that σ is a $d \times e$ matrix function and b is a d-dimensional vector function both defined on $\mathbb{R}^d$. To avoid excessive technicalities these coefficients and the process ξ are supposed to satisfy the following uniform bounds

$$\|\sigma\|_{C^3},\, \|b\|_{C^3} \le L \tag{2.6}$$

and

$$\|\xi(0)\|_\infty \le L \tag{2.7}$$

for some $L \ge 1$, where $\|\cdot\|_{C^3}$ is a matrix or a vector function C^3 norm and $\|\cdot\|_\infty$ is the L^∞ norm.

Introduce the $e \times e$ matrix $\varsigma = (\varsigma_{ij})$ where

$$\varsigma_{ij} = \lim_{k\to\infty} \frac{1}{k} \sum_{m=0}^{k} \sum_{n=0}^{k} \varsigma_{ij}(n-m), \text{ and } \varsigma_{il}(n-m) = E(\xi_i(m)\xi_l(n)) \tag{2.8}$$

taking into account that the limit above will exist under conditions of our theorem below (see (4.6)–(4.7) in Section 4 below). Define also

$$c_i(x) = \sum_{j,l=1}^{e} \sum_{k=1}^{d} \frac{\partial \sigma_{ij}(x)}{\partial x_k} \hat{\varsigma}_{lj} \sigma_{kl}(x),\ i = 1, ..., d \tag{2.9}$$

where

$$\hat{\varsigma}_{jl} = \lim_{k\to\infty} \frac{1}{k} \sum_{n=0}^{k} \sum_{m=-k}^{n-1} \varsigma_{jl}(n-m) = \sum_{n=1}^{\infty} E(\xi_j(0)\xi_l(n)) \tag{2.10}$$

and the latter limit exists under our conditions (which follows from (4.6)–(4.7) below). Let Ξ be the unique solution of the stochastic differential equation

$$d\Xi(t) = \sigma(\Xi(t))dW(t) + (b(\Xi(t)) + c(\Xi(t)))dt \tag{2.11}$$

where W is the e-dimensional Brownian motion with the covariance matrix ς (at the time 1). In what follows we will write $A(u) = O(\alpha(u))$ a.s. for a family of random variables $A(u)$, $u \in U \subset \mathbb{R}$ and a nonrandom sequence $\alpha(u)$, $u \in U \subset \mathbb{R}$ if $A(u)/\alpha(u)$ is bounded for all $u \in U$ by an almost surely finite random variable.

Theorem 2.1. *Suppose that X_N is defined by (2.4), (2.5)–(2.7) hold true and assume that*

$$\sup_{n\ge 0} n^4(\phi(n) + \rho(n)) < \infty. \tag{2.12}$$

Then the stationary sequence of random vectors $\xi(n)$, $-\infty < n < \infty$ can be redefined preserving its distributions on a sufficiently rich probability space which contains also a e-dimensional Brownian motion $\mathcal{W}$ with the covariance matrix ς (at the time 1) so that for each $T > 0$ and Ξ_N solving (2.11) with $W(t) = W_N(t) = N^{-1/2}\mathcal{W}(Nt)$, $0 \le t \le T$,

$$\sup_{0\le t\le T} |X_N(t) - \Xi_N(t)| = O(N^{-\delta}) \quad \textit{a.s. for all} \quad N \ge 1 \tag{2.13}$$

where $X_N(0) = \Xi_N(0)$ and $\delta > 0$ does not depend on $N \ge 1$.

We stress that a sufficiently rich probability space has here a quite precise meaning that one can define on it a sequence of independent uniformly didtributed on $[0.1]$ random variables which are independent of the sequence $\xi(n)$, $n \in \mathbb{Z}$ (see Theorem 3.8 below).

Under certain additional conditions, which are always satisfied when $d = 1$ and σ is bounded away from zero, the problem can be reduced to the strong invariance principle for sums (cf. [32]) which is well known and this would imply then Theorem 2.1.

Nevertheless, in the multidimensional case such reduction requires substantial additional assumptions on σ (see [32]). For the proof of Theorem 2.1 we will follow, first, the strategy appeared more than forty years ago in [9] leading to the strong (almost sure) invariance principle in the supremum norm. Moreover, we extend this to estimates of the almost sure invariance principles in a p-variation norm both for sums S_N and iterated sums $\mathbb{S}_N$ given by

$$S_N(t) = N^{-1/2} \sum_{0\le k<[Nt]} \xi(k) \quad \text{and} \quad \mathbb{S}_N^{ij}(t) = N^{-1} \sum_{0\le k<l<[Nt]} \xi_i(k)\xi_j(l), \tag{2.14}$$

where $S_N(0) = \mathbb{S}_N^{ij}(0) = 0$, and then rely on the local Lipschitz property of the Itô-Lyons map for rough differential equations (see, for instance, [22]) together with the growth estimates of the corresponding Lipschitz constant (see Theorem 6.1 in Section 6). Observe also that by (2.10),

$$\lim_{N\to\infty} E\mathbb{S}_N^{ij}(t) = t\hat{\varsigma}_{ij}.$$

As it is customary in the rough paths theory we use slightly different definitions for processes denoted by usual letters and the ones denoted by blackboard letters. Namely, for a process $Q(t)$, $t\ge 0$ we write $Q(s,t) = Q(t) - Q(s)$ when $t \ge s$. On the other hand, if $\mathbb{Q}(t) = \sum_{0\le k<l<[tN]} \eta(k)\zeta(l)$ or $\mathbb{Q}(t) = \int_0^t \eta(u)d\zeta(u)$ (where η and ζ are one-dimensional processes and when the latter integral makes sense) then

$$\mathbb{Q}(s,t) = \sum_{[sN]\le k<l<[tN]} \eta(k)\zeta(l) \text{ and } \mathbb{Q}(s,t) = \int_s^t (\eta(u) - \eta(s))d\zeta(u),$$

respectively (see Section 6.1.1 for some recalls). These notations affect the following definition of p-variation norms. For any path $\gamma(t)$, $t \in [0,T]$ in a Euclidean space having left and right limits and $p \ge 1$ the p-variation norm of γ on an interval $[U,V]$, $U < V$ is given by

$$\|\gamma\|_{p,[U,V]} = \big(\sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |\gamma(s,t)|^p\big)^{1/p} \tag{2.15}$$

where the supremum is taken over all partitions $\mathcal{P} = \{U = t_0 < t_1 < ... < t_n = V\}$ of $[U,V]$ and the sum is taken over the corresponding subintervals $[t_i, t_{i+1}]$, $i = 0,1,...,n-1$ of the partition while $\gamma(s,t)$ is taken according to the definitions above depending on the process under consideration. The main step in the proof of Theorem 2.1 is the following result.

Theorem 2.2. *Suppose that (2.12) holds true. Then the stationary sequence of random vectors $\xi(n)$, $-\infty < n < \infty$ can be redefined preserving its distributions on a sufficiently rich probability space which contains also a e-dimensional Brownian motion $\mathcal{W}$ with the covariance matrix ς (at the time 1) so that for each $T > 0$ and $W_N(t) = N^{-1/2}\mathcal{W}(Nt)$, $0 \le t \le T$,*

$$\|S_N - W_N\|_{p,[0,T]} = O(N^{-\delta}) \quad \text{a.s.} \tag{2.16}$$

and

$$\max_{1\le i,j\le d} \|\mathbb{S}_N^{ij} - \mathbb{W}_N^{ij}\|_{\frac{p}{2},[0,T]} = O(N^{-\delta}) \quad \text{a.s.} \tag{2.17}$$

where $p \in (2,3)$, $\delta > 0$ does not depend on $N \ge 1$ (and may be different from δ in (2.13)), S_N, $\mathbb{S}_N$ are given by (2.14) and

$$\mathbb{W}_N^{ij}(t) = \int_0^t W_N^i(s)dW_N^j(s) + t\sum_{l=1}^{\infty} E(\xi_i(0)\xi_j(l)). \tag{2.18}$$

Observe that the estimate (2.16) obtained in the p-variation norm is a stronger result than the standard strong invariance principle considered in other papers such as [9], [36], [17], [41] and [27]. The strong invariance principle for S_N, i.e. an a.s. eventual estimate of $|S_N - W_N|$ in the supremum norm, is essentially well known but in [9] and [36] it is proved assuming that the σ-algebras $\mathcal{F}_{mn}$ are generated by the random vectors $\xi(m), ..., \xi(n)$ themselves (which restricts applications) while in [41] and [27] this is proved also under different from ours conditions. Thus, we cannot rely here on a direct reference and will provide the proof of this part, as well. In fact, we will need not only this result itself but also the specific construction of the Brownian motions W_N emerging there which is used in the proof of (2.17). The strong invariance principle for iterated sums in the form (2.17) does not seem to appear in the literature before. Moreover, we will see in Section 7 that the strong invariance principle for second order iterated sums obtained in Theorem 2.2 implies strong invariance principles for multiple iterated sums of any order which, in turn, yields laws of iterated logarithm for them relying on the laws of iterated logarithm for multiple stochastic integrals $\int_0^t d\mathcal{W}^{i_\ell}(t_\ell)\int_0^{t_\ell} ... \int_0^{t_2} d\mathcal{W}^{i_1}(t_1)$ (see [2]).

Corollary 2.3. *Under the conditions of Theorem 2.2 for all* $\ell \geq 1$ *and* $0 \leq i_1, ..., i_\ell \leq d$*,*

$$P\Big\{\limsup_{t\to\infty} \frac{\sum_{0\leq k_1<k_2<...<k_\ell<t} \xi_{i_1}(k_1)\xi_{i_2}(k_2)\cdots\xi_{i_\ell}(k_\ell)}{(2t\log\log t)^{\ell/2}} = d_{i_1,i_2,...,i_\ell}\Big\} = 1$$

where $1 \leq i_1 \leq i_2 \leq ... \leq i_\ell \leq e$ *and the constant* $d_{i_1,i_2,...,i_\ell}$ *can be computed from the formula given in Corollary 7.10.*

Important classes of processes satisfying our conditions come from dynamical systems. Let F be a C^2 Axiom A diffeomorphism (in particular, Anosov) in a neighborhood of an attractor or let F be an expanding C^2 endomorphism of a Riemannian manifold Ω (see [7]), g be either a Hölder continuous vector function or a vector function which is constant on elements of a Markov partition and let $\xi(n) = \xi(n, \omega) = g(F^n\omega)$. Here the probability space is $(\Omega, \mathcal{B}, P)$ where P is a Gibbs invariant measure corresponding to some Hölder continuous function and $\mathcal{B}$ is the Borel σ-field. In this setup the assumption (2.6) on boundedness of ξ turns out to be quite natural. Let ζ be a finite Markov partition for F then we can take $\mathcal{F}_{kl}$ to be the finite σ-algebra generated by the partition $\cap_{i=k}^{l} F^i\zeta$. In fact, we can take here not only Hölder continuous g's but also indicators of sets from $\mathcal{F}_{kl}$. The conditions of Theorems 2.1 and 2.2 allow all such functions since the dependence of Hölder continuous functions on m-tails, i.e. on events measurable with respect to $\mathcal{F}_{-\infty,-m}$ or $\mathcal{F}_{m,\infty}$, decays exponentially fast in m and the condition (2.12) is even weaker than that. A related class of dynamical systems corresponds to F being a topologically mixing subshift of finite type which means that F is the left shift on a subspace Ω of the space of one (or two) sided sequences $\omega = (\omega_i, i \geq 0), \omega_i = 1, ..., l_0$ such that $\omega \in \Omega$ if $\pi_{\omega_i\omega_{i+1}} = 1$ for all $i \geq 0$ where $\Pi = (\pi_{ij})$ is an $l_0 \times l_0$ matrix with 0 and 1 entries and such that Π^n for some n is a matrix with positive entries. Again, we have to take in this case g to be Hölder continuous bounded functions on the sequence space above, P to be a Gibbs invariant measure corresponding to some Hölder continuous function and to define $\mathcal{F}_{kl}$ as the finite σ-algebra generated by cylinder sets with fixed coordinates having numbers from k to l. The exponentially fast ψ-mixing, which is stronger than ϕ-mixing required here, is well known in the above cases (see [7]). Among other dynamical systems with exponentially fast ψ-mixing we can mention also the Gauss map $Fx = \{1/x\}$ (where $\{\cdot\}$ denotes the fractional part) of the unit interval with respect to the Gauss measure G and more general transformations generated by f-expansions (see [28]). Gibbs-Markov maps which are known to be exponentially fast ϕ-mixing (see, for instance, [41]) can be also taken as F in Theorem 2.1 with $\xi(n) = g \circ F^n$ as above.

2.2 Continuous time case

Here we start with a complete probability space $(\Omega, \mathcal{F}, P)$, a P-preserving invertible transformation $\vartheta : \Omega \to \Omega$ and a two parameter family of countably generated σ-algebras $\mathcal{F}_{m,n} \subset \mathcal{F}$, $-\infty \leq m \leq n \leq \infty$ such that $\mathcal{F}_{mn} \subset \mathcal{F}_{m'n'} \subset \mathcal{F}$ if $m' \leq m \leq n \leq n'$ where $\mathcal{F}_{m\infty} = \cup_{n:\, n \geq m}\mathcal{F}_{mn}$ and $\mathcal{F}_{-\infty n} = \cup_{m:\, m \leq n}\mathcal{F}_{mn}$. The setup includes also a (roof or ceiling) function $\tau : \Omega \to (0, \infty)$ such that for some $\hat{L} > 0$,

$$\hat{L}^{-1} \leq \tau \leq \hat{L}. \tag{2.19}$$

Next, we consider the probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ such that $\hat{\Omega} = \{\hat{\omega} = (\omega, t) : \omega \in \Omega,\, 0 \leq t \leq \tau(\omega)\}$, $(\omega, \tau(\omega)) = (\vartheta\omega, 0)\}$, $\hat{\mathcal{F}}$ is the restriction to $\hat{\Omega}$ of $\mathcal{F} \times \mathcal{B}_{[0,\hat{L}]}$, where $\mathcal{B}_{[0,\hat{L}]}$ is the Borel σ-algebra on $[0, \hat{L}]$ completed by the Lebesgue zero sets, and for any $\Gamma \in \hat{\mathcal{F}}$,

$$\hat{P}(\Gamma) = \bar{\tau}^{-1}\int \mathbb{I}_\Gamma(\omega, t)dtdP(\omega) \text{ where } \bar{\tau} = \int \tau dP = E\tau$$

and E denotes the expectation on the space $(\Omega, \mathcal{F}, P)$.

Finally, we introduce a vector valued stochastic process $\xi(t) = \xi(t, (\omega, s))$, $-\infty < t < \infty$, $0 \leq s \leq \tau(\omega)$ on $\hat{\Omega}$ satisfying

$$\xi(t, (\omega, s)) = \xi(t + s, (\omega, 0)) = \xi(0, (\omega, t + s)) \text{ if } 0 \leq t + s < \tau(\omega) \text{ and}$$
$$\xi(t, (\omega, s)) = \xi(0, (\vartheta^k\omega, u)) \text{ if } t + s = u + \sum_{j=0}^{k}\tau(\vartheta^j\omega) \text{ and } 0 \leq u < \tau(\vartheta^k\omega).$$

This construction is called in dynamical systems a suspension and it is a standard fact that ξ is a stationary process on the probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ and in what follows we will write also $\xi(t, \omega)$ for $\xi(t, (\omega, 0))$.

We will assume that $X^\varepsilon(t) = X^\varepsilon(t, \omega)$ from (1.1) considered as a process on $(\Omega, \mathcal{F}, P)$ solves the equation

$$\frac{dX^\varepsilon(t)}{dt} = \frac{1}{\varepsilon}\sigma(X^\varepsilon(t))\xi(t/\varepsilon^2) + b(X^\varepsilon(t)),\ t \in [0, T] \tag{2.20}$$

where the matrix function σ and the process ξ satisfy (2.6). Set $\eta(\omega) = \int_0^{\tau(\omega)}\xi(s, \omega)ds$ and

$$\rho(n) = \sup_m \max\big(\|\tau \circ \vartheta^m - E(\tau \circ \vartheta^m|\mathcal{F}_{m-n,m+n})\|_\infty, \tag{2.21}$$
$$\|ess\sup_{0 \leq s \leq \tau(\theta^m\omega)}|\xi(0, (\vartheta^m\omega, s)) - E(\xi(0, (\vartheta^m\omega, s))|\mathcal{F}_{m-n,m+n})|\|_\infty\big).$$

Observe also that $\eta(k) = \eta \circ \vartheta^k$ is a stationary sequence of random vectors.

Next, we consider a diffusion process Ξ solving the stochastic differential equation

$$d\Xi(t) = \sigma(\Xi(t))dW(t) + (\bar{\tau}b(\Xi(t)) + c(\Xi(t)))dt, \tag{2.22}$$

with d-dimensional Brownian motion W having the covariance matrix $\varsigma = (\varsigma_{ij})$ at the time 1 given by

$$\varsigma_{ij} = \lim_{n\to\infty}\frac{1}{n}\sum_{k,l=0}^{n}E(\eta_i(k)\eta_j(l)) \tag{2.23}$$

and with

$$c_i(x) = \sum_{j,l=1}^{e}\sum_{k=1}^{d}\frac{\partial\sigma_{ij}(x)}{\partial x_k}\Big(\hat{\varsigma}_{lj} + \int_0^{\tau(\omega)}\xi_j(s, \omega)ds\int_0^s\xi_l(u, \omega)du\Big)\sigma_{kl}(x), \tag{2.24}$$
$$\hat{\varsigma}_{ij} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}\sum_{l=-n}^{k-1}E(\eta_i(l)\eta_j(k)) = \sum_{n=1}^{\infty}E(\eta_i(0)\eta_j(n)).$$

The limits in (2.23) and (2.24) exist in view of estimates similar to (4.6)–(4.7) below. Notice the difference in the definitions of $c(x)$ in (2.9) and in (2.24) which is due to the fact that $c(x)$ is defined here in terms of the process η and not ξ. The following result is a continuous time version of Theorem 2.1 which can be viewed also as a substantial extension of [17].

Theorem 2.4. *Assume that $E\eta = E\int_0^\tau \xi(t)dt = 0$ and (2.6), (2.7) and (2.12) (for ρ defined in (2.21)) hold true, as well. Then the stationary vector process $\xi(t)$, $-\infty < t < \infty$ can be redefined preserving its distributions on a sufficiently rich probability space which contains also a e-dimensional Brownian motion $\mathcal{W}$ with the covariance matrix ς (at the time 1) so that for each $T > 0$ and Ξ^ε solving (2.22) with $W(t) = W^\varepsilon(t) = \varepsilon\mathcal{W}(\varepsilon^{-2}t)$, $0 \leq t \leq T$,*

$$\sup_{0\leq t\leq T} |X^\varepsilon(t) - \Xi^\varepsilon(t/\bar\tau)| = O(\varepsilon^\delta) \;\; \text{a.s.} \tag{2.25}$$

where $X^\varepsilon(0) = \Xi^\varepsilon(0)$ and $\delta > 0$ does not depend on $\varepsilon > 0$.

We observe that if the stationary process ξ on the probability space $(\hat\Omega, \hat{\mathcal{F}}, \hat P)$ would be sufficiently fast mixing then the proof of Theorem 2.4 could proceed essentially in the same way as in Theorem 2.1 and, in fact, the former could be derived from the latter by discretizing time. But, in general, this is not the case in applications to dynamical systems no matter what speed of decay of the coefficients ϕ and ρ on the base space $(\Omega, \mathcal{F}, P)$ is assumed, and so substantial additional work is required here. The key step in the proof of Theorem 2.4 is to obtain a continuous time version of Theorem 2.2 which is the following result proved in Section 5. After that we will rely on rough paths theory arguments of Section 6. Set

$$V^\varepsilon(t) = \varepsilon\int_0^{t\bar\tau\varepsilon^{-2}} \xi(s)ds,\;\; \mathbb{V}^\varepsilon_{ij}(t) = \varepsilon^2\int_0^{t\bar\tau\varepsilon^{-2}} \xi_j(s)ds\int_0^s \xi_i(u)du,\; i,j = 1,...,d$$

and observe that in view of estimates in Section 5.3, more precisely (5.16),

$$\lim_{\varepsilon\to 0} E\mathbb{V}^\varepsilon_{ij}(t) = t\sum_{l=1}^\infty E(\eta_i(0)\eta_j(l)) + tE\Big(\int_0^{\tau(\omega)} \xi_j(s,\omega)ds\int_0^s \xi_i(u,\omega)du\Big).$$

Theorem 2.5. *Suppose that (2.12) and (2.19) hold true with ρ defined in (2.21). Then the stationary vector process $\xi(t)$, $-\infty < t < \infty$ can be redefined preserving its distributions on a sufficiently rich probability space which contains also a e-dimensional Brownian motion $\mathcal{W}$ with the covariance matrix ς at the time 1 given by (2.23) so that for each $T > 0$ and $W^\varepsilon(t) = \varepsilon\mathcal{W}(\varepsilon^{-2}t)$, $0 \leq t \leq T$,*

$$\|V^\varepsilon - W^\varepsilon\|_{p,[0,T]} = O(\varepsilon^\delta) \quad \text{a.s.} \tag{2.26}$$

and

$$\|\mathbb{V}^\varepsilon_{ij} - \mathbb{W}^\varepsilon_{ij}\|_{p/2,[0,T]} = O(\varepsilon^\delta) \quad \text{a.s.} \tag{2.27}$$

where $p \in (2,3)$, $\delta > 0$ does not depend on ε, a.s. is taken simultaneously over $\varepsilon \in (0,1)$ and

$$\mathbb{W}^\varepsilon_{ij}(t) = \int_0^t W^\varepsilon_i(s)dW^\varepsilon_j(s) + t\sum_{l=1}^\infty E(\eta_i(0)\eta_j(l)) + tE\Big(\int_0^{\tau(\omega)} \xi_j(s,\omega)ds\int_0^s \xi_i(u,\omega)du\Big).$$

Again, as explained in Section 7 the strong invariance principle (2.27) for second order iterated integrals implies strong invariance principles for iterated integrals of all orders. Taking this into account we can use Theorem 2.5 not only for the proof of Theorem 2.4 but also can apply (2.27) and its extension to multiple iterated integrals in order to obtain laws of iterated logarithm for iterated integrals of any order relying on the laws of iterated logarithm for multiple stochastic integrals (see, for instance, [2]).

Corollary 2.6. *Under the conditions of Theorem 2.5, for all $\ell \geq 1$,*

$$P\Big\{\limsup_{t\to\infty} \frac{\int_0^t \xi_{i_\ell}(t_\ell)dt_\ell\int_0^{t_\ell}\xi_{i_{\ell-1}}(t_{\ell-1})dt_{\ell-1}\cdots\int_0^{t_2}\xi_{i_1}(t_1)dt_1}{(2t\log\log t)^{\ell/2}} = d_{i_1,i_2,...,i_\ell}\Big\} = 1$$

where $1 \le i_1 \le i_2 \le ... \le i_\ell \le e$ *and the constant* $d_{i_1,i_2,...,i_\ell}$ *can be computed from the formula given in Corollary 7.13.*

The main application to dynamical systems we have here in mind is a C^2 Axiom A flow F^t near an attractor which using Markov partitions can be represented as a suspension over an exponentially fast ψ-mixing transformation so that we can take $\xi(t) = g \circ F^t$ for a Hölder continuous function g and the probability P being a Gibbs invariant measure constructed by a Hölder continuous potential on the base of the Markov partition (see, for instance, [10]).

Remark 2.7. It is not clear how to extend our methods to derive Theorems 2.1 and 2.4 for the processes X_N and X^ε given by the general equations (1.3) and (1.1), respectively. This would require, say in the discrete time case, to obtain estimates for strong (almost sure) approximations for the sums $S_N(x,t) = N^{-1/2} \sum_{0 \le k < [Nt]} B(x, \xi(k))$ and iterated sums $\mathbb{S}_N^{ij}(x,t) = N^{-1} \sum_{0 \le k < l < [Nt]} B_i(x,\xi(k)) B_j(x, \xi(l))$ in the supremum (even a Hölder) in x norm. This, essentially, amounts to a strong invariance principle in a Banach space which was proved for sums in [36] but only with logarithmic estimates which does not seem to allow to extend it to iterated sums and to p-variation norms. If this were possible then we could proceed relying on local Lipschitz continuity of the Banach space version of the Itô-Lyons map for rough differential equations (see, for instance, Theorem 3.6 in [13]; also [34, 1]).

Remark 2.8. Our results can be obtained assuming moment rather than uniform bounds, namely, in place of $\|\xi\|_\infty < \infty$ we can assume that $\|\xi\|_m < \infty$ for some m big enough. To do this it is helpful to replace the ϕ-mixing coefficient by more general dependence coefficients between pairs of σ-algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ defined by

$$\varpi_{q,p}(\mathcal{G}, \mathcal{H}) = \sup\{\|E(g|\mathcal{G}) - Eg\|_p : \, g \text{ is } \mathcal{H} - \text{measurable and } \|g\|_q \le 1\}.$$

The proofs proceed then essentially in the same way estimating moments of conditional expectations not by Lemma 3.1 below, as we do here, but by Corollary 3.6 from [35] and relying on some related estimates from Section 3 there.

Remark 2.9. Our method in the proof of Theorem 2.5 can be slightly modified to extend Theorem 2.2 to random vectors built by nonuniformly hyperbolic dynamical systems modelled by Young towers which are discrete time suspensions assuming boundedness of appropriate moments of the return time function.

3 Strong approximations for sums

3.1 General lemmas

First, we will formulate three general results which will be used throughout this paper. The following lemma is well known (see, for instance, Corollary to Lemma 2.1 in [30] or Lemma 1.3.10 in [29]).

Lemma 3.1. *Let* $H(x,\omega)$ *be a bounded measurable function on the space* $(\mathbb{R}^d \times \Omega,\, \mathcal{B} \times \mathcal{F})$*, where* $\mathcal{B}$ *is the Borel* σ*-algebra, such that for each* $x \in \mathbb{R}^d$ *the function* $H(x,\cdot)$ *is measurable with respect to a* σ*-algebra* $\mathcal{G} \subset \mathcal{F}$*. Let* V *be an* $\mathbb{R}^d$*-valued random vector measurable with respect to another* σ*-algebra* $\mathcal{H} \subset \mathcal{F}$*. Then with probability one,*

$$|E(H(V,\omega)|\mathcal{H}) - h(V)| \le 2\phi(\mathcal{G},\mathcal{H})\|H\|_\infty \tag{3.1}$$

where $h(x) = EH(x,\cdot)$ *and the* ϕ*-dependence coefficient was defined in (2.1). In particular (which is essentially an equivalent statement), let* $H(x_1,x_2)$*,* $x_i \in \mathbb{R}^{d_i}$*,* $i = 1,2$ *be a bounded Borel function and* V_i *be* $\mathbb{R}^{d_i}$*-valued* $\mathcal{G}_i$*-measurable random vectors,* $i = 1,2$ *where* $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ *are sub* σ*-algebras. Then with probability one,*

$$|E(H(V_1,V_2)|\mathcal{G}_1) - h(V_1)| \le 2\phi(\mathcal{G}_1,\mathcal{G}_2)\|H\|_\infty.$$

We will employ several times the following general moment estimate which appeared as Lemma 3.2.5 in [29] for random variables and was extended to random vectors in Lemma 3.4 from [31].

Lemma 3.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration of σ-algebras $\mathcal{G}_j$, $j \geq 1$ and a sequence of random d-dimensional vectors η_j, $j \geq 1$ such that η_j is $\mathcal{G}_j$-measurable, $j = 1, 2, ...$ Suppose that for some integer $M \geq 1$,*

$$A_{2M} = \sup_{i \geq 1} \sum_{j \geq i} \|E(\eta_j | \mathcal{G}_i)\|_{2M} < \infty$$

where $\|\eta\|_p = (E|\eta|^p)^{1/p}$ and $|\eta|$ is the Euclidean norm of a (random) vector η. Then for any integer $n \geq 1$,

$$E|\sum_{j=1}^{n} \eta_j|^{2M} \leq 3(2M)! d^M A_{2M}^{2M} n^M. \tag{3.2}$$

In order to obtain uniform moment estimates required by Theorem 2.1 we will need the following general estimate which appeared as Lemma 3.7 in [31].

Lemma 3.3. *Let $\eta_1, \eta_2, ..., \eta_N$ be random d-dimensional vectors and $\mathcal{H}_1 \subset \mathcal{H}_2 \subset ... \subset \mathcal{H}_N$ be a filtration of σ-algebras such that η_m is $\mathcal{H}_m$-measurable for each $m = 1, 2, ..., N$. Assume also that $E|\eta_m|^{2M} < \infty$ for some $M \geq 1$ and each $m = 1, ..., n$. Set $S_m = \sum_{j=1}^{m} \eta_j$. Then*

$$\begin{aligned} &E \max_{1 \leq m \leq n} |S_m|^{2M} \leq 2^{2M-1} \big((\tfrac{2M}{2M-1})^{2M} E|S_n|^{2M} \\ &+ E \max_{1 \leq m \leq n-1} |\sum_{j=m+1}^{n} E(\eta_j | \mathcal{H}_m)|^{2M} \big) \leq 2^{2M-1} A_{2M}^{2M} (3(2M)! d^M n^M + n) \end{aligned} \tag{3.3}$$

where A_{2M} is the same as in Lemma 3.2.

The following result will be used for moment estimates of sums and iterated sums of random variables, the latter part seems to be completely new.

Lemma 3.4. *Let $\zeta(k)$, $\mu(k)$, $k = 0, 1, ...$ be two sequences of random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$\begin{gathered} E\zeta(k) = E\mu(k) = 0 \text{ for all } k \geq 0, \ \sup_{k \geq 0} E(|\zeta(k)|^{2M} + |\mu(k)|^{2M}) < \infty \text{ and} \\ |\zeta(k) - E(\zeta(k) | \mathcal{F}_{k-n,k+n})|, \ |\mu(k) - E(\mu(k) | \mathcal{F}_{k-n,k+n})| \leq \rho(n) \end{gathered}$$

where the probability P is ϕ-mixing with respect to the family of σ-algebras $\mathcal{F}_{kl}$ (as described in Section 2) with ϕ and $\rho \geq 0$ satisfying $\sum_{n=0}^{\infty} n(\rho(n) + \phi(n)) < \infty$. Set

$$\mathbb{R}(\ell, m, n) = \sum_{k=m}^{n} \sum_{j=\ell(k)}^{k-1} (\zeta(j)\mu(k) - E(\zeta(j)\mu(k)))$$

where $0 \leq m \leq n$ are integers and $0 \leq \ell(k) < k$ is an integer valued function (maybe constant). Then

$$E \max_{m \leq n < N} |\mathbb{R}(\ell, m, n)|^{2M} \leq C_1^{\zeta,\mu}(M)(N - m)^M \max_{m \leq k \leq N} (k - \ell(k))^M \tag{3.4}$$

where $C^{\zeta,\mu}(M) > 0$ does not depend on N, m or ℓ. In fact, $C^{\zeta,\mu}(M)$ depends only on M, ρ and ϕ while it does not depend on the sequences $\zeta(k)$ and $\mu(k)$ themselves. In particular,

$$E \max_{m \leq n < N} |\sum_{k=m}^{n} \mu(k)|^{2M} \leq C_1^{1,\mu}(M)(N - m)^M$$

which is obtained simplifying the proof below just by disregarding the sequence $\zeta(j)$*. If* $\xi(k),\ k=0,1,...$ *is a sequence of random vectors with the components* $\xi_i(k),\ i=1,...,d$ *satisfying the conditions above, then*

$$E\max_{m\le n<N}|\sum_{k=m}^n\xi(k)|^{2M}\le C_1^\xi(M)(N-m)^M$$

for $C_1^\xi(M)=d^{2M-1}\sum_{i=1}^d C_1^{1,\xi_i}$ *where* $|\cdot|$ *denotes the Euclidean norm in* $\mathbb{R}^d$*.*

Proof. Set

$$\zeta_r(k)=E(\zeta(k)|\mathcal{F}_{k-r,k+r})\text{ and }\mu_r(k)=E(\mu(k)|\mathcal{F}_{k-r,k+r}).$$

Then

$$\zeta(k)=\lim_{i\to\infty}\zeta_{2^i}(k)=\zeta_1(k)+\sum_{r=1}^\infty(\zeta_{2^r}(k)-\zeta_{2^{r-1}}(k))$$

and

$$\mu(k)=\lim_{i\to\infty}\mu_{2^i}(k)=\mu_1(k)+\sum_{r=1}^\infty(\mu_{2^r}(k)-\mu_{2^{r-1}}(k))$$

where convergence is in the L^∞ sense since

$$\|\zeta_{2^r}(k)-\zeta_{2^{r-1}}(k)\|_\infty,\ \|\mu_{2^r}(k)-\mu_{2^{r-1}}(k)\|_\infty\le 2(\rho(2^r)+\rho(2^{r-1})).$$

For $q,r=0,1,...$ denote

$$\varrho_{q,r}(j,k)=(\zeta_{2^q}(j)-\zeta_{2^{q-1}}(j))(\mu_{2^r}(k)-\mu_{2^{r-1}}(k))$$
$$-E\big((\zeta_{2^q}(j)-\zeta_{2^{q-1}}(j))(\mu_{2^r}(k)-\mu_{2^{r-1}}(k))\big)$$

and $Q_{q,r}(k)=\sum_{j=\ell(k)}^{k-1}\varrho_{q,r}(j,k)$ where we set for convenience $\rho(2^{-1})=\sup_{k\ge0}(\|\zeta(k)\|_{2M}+\|\mu(k)\|_{2M})$ and $\zeta_{2^{-1}}(j)=\mu_{2^{-1}}(k)=0$ for all $j,k\ge0$. Then

$$\mathbb{R}(\ell,m,n)=\sum_{q,r=0}^\infty\sum_{k=m}^n Q_{q,r}(k).$$

Next, introduce $\mathcal{G}_m=\mathcal{G}_m^{q,r}=\mathcal{F}_{-\infty,m+\max(2^q,2^r)}$ and observe that $Q_{q,r}(k)$ is $\mathcal{G}_k$-measurable. We will apply Lemmas 3.2 and 3.3 to the sums $\mathbb{R}(\ell,m,n)$.

First, write

$$Q_{q,r}(k)=Q^{(1)}_{q,r,n}(k)+Q^{(2)}_{q,r,n}(k)+Q^{(3)}_{q,r,n}(k)$$

where

$$Q^{(1)}_{q,r,n}(k)=\sum_{\ell(k)\le j<\frac{k+n}{2}-2\max(2^q,2^r),\,j<k}\varrho_{q,r}(j,k),$$
$$Q^{(2)}_{q,r,n}(k)=\sum_{\frac{k+n}{2}-2\max(2^q,2^r)\le j<\frac{k+n}{2}+2\max(2^q,2^r),\,j<k}\varrho_{q,r}(j,k),$$
$$\text{and }\ Q^{(3)}_{q,r,n}(k)=\sum_{\frac{k+n}{2}+2\max(2^q,2^r)\le j<k}\varrho_{q,r}(j,k).$$

If $k-n\ge 4\max(2^q,2^r)$ then $\frac{k+n}{2}-\max(2^q,2^r)-n-\max(2^q,2^r)\ge0$ and we can write

$$\|E(Q^{(1)}_{q,r}(k)|\mathcal{G}_n)\|_{2M}\le 2\|E\big(E(\mu_{2^r}(k)-\mu_{2^{r-1}}(k)|\mathcal{F}_{-\infty,\frac{k+n}{2}-\max(2^q,2^r)})$$
$$\times\textstyle\sum_{\ell(k)\le j<\frac{k+n}{2}-2\max(2^q,2^r),\,j<k}(\zeta_{2^q}(j)-\zeta_{2^{q-1}}(j))|\mathcal{G}_n\big)\|_{2M}$$
$$\le 2\|E(\mu_{2^r}(k)-\mu_{2^{r-1}}(k)|\mathcal{F}_{-\infty,\frac{k+n}{2}-\max(2^q,2^r)})$$
$$\times\textstyle\sum_{\ell(k)\le j<\frac{k+n}{2}-2\max(2^q,2^r),\,j<k}(\zeta_{2^q}(j)-\zeta_{2^{q-1}}(j))\|_{2M}$$
$$\le 8\phi(\tfrac{k-n}{2})(\rho(2^r)+\rho(2^{r-1}))\|\textstyle\sum_{\ell(k)\le j<\frac{k+n}{2}-2\max(2^q,2^r),\,j<k}(\zeta_{2^q}(j)-\zeta_{2^{q-1}}(j))\|_{2M}$$

where we use (2.1), (2.2) and that $k - 2^r - \frac{k+n}{2} + \max(2^q, 2^r) \geq \frac{k-n}{2}$. If $0 \leq k - n < 4\max(2^q, 2^r)$ then we estimate

$$\|E(Q^{(1)}_{q,r}(k)|\mathcal{G}_n)\|_{2M}$$
$$\leq 4(\rho(2^r) + \rho(2^{r-1}))\| \sum_{\ell(k)\leq j<\frac{k+n}{2}-2\max(2^q,2^r),\, j<k}(\zeta_{2^q}(j) - \zeta_{2^{q-1}}(j))\|_{2M}.$$

In order to bound the $2M$-moment norm of the last sum we will use Lemma 3.2 setting $\mathcal{H}_i = \mathcal{F}_{-\infty,i+2^q}$ and relying on (2.1)-(2.2) we estimate for $j \geq i + 2^{q+1}$,

$$\|E(\zeta_{2^q}(j) - \zeta_{2^{q-1}}(j)|\mathcal{H}_i)\|_{2M} \leq 4(\rho(2^q) + \rho(2^{q-1}))\phi(j - i - 2^{q+1}).$$

For $i \leq j < i + 2^{q+1}$ we use just the obvious estimate

$$\|E(\zeta_{2^q}(j) - \zeta_{2^{q-1}}(j)|\mathcal{H}_i)\|_{2M} \leq 2(\rho(2^q) + \rho(2^{q-1})).$$

Hence,

$$A^{\zeta}_{2M} = \sup_{i\geq 0}\sum_{j\geq i}\|E(\zeta_{2^q}(j) - \zeta_{2^{q-1}}(j)|\mathcal{H}_i)\|_{2M} \leq 2(\rho(2^q) + \rho(2^{q-1}))(2^{q+1} + 2\sum_{i=0}^{\infty}\phi(i)),$$

and so by Lemma 3.2,

$$\| \sum_{\ell(k)\leq j<\frac{k+n}{2}-2\max(2^q,2^r),\, j<k}(\zeta_{2^q}(j) - \zeta_{2^{q-1}}(j))\|_{2M}$$
$$\leq (3(2M)!)^{1/2M}\sqrt{d}A^{\zeta}_{2M}\max_{m\leq k\leq n}\sqrt{k-\ell(k)}.$$

Next, by (2.1) and (2.2),

$$|E(Q^{(2)}_{q,r,n}(k)|\mathcal{G}_n)| \leq \sum_{\frac{k+n}{2}-2\max(2^q,2^r)\leq j<\frac{k+n}{2}+2\max(2^q,2^r),\, j<k}|E(\varrho_{q,r}(j,k)|\mathcal{G}_n)|$$
$$\leq 64\max(2^q, 2^r)\phi(\tfrac{k-n}{2} - 4\max(2^q, 2^r))(\rho(2^r) + \rho(2^{r-1}))(\rho(2^q) + \rho(2^{q-1}))$$

since each $\varrho_{q,r}(j,k)$ here is $\mathcal{F}_{\frac{k+n}{2}-3\max(2^q,2^r),\infty}$-measurable, $\frac{k+n}{2} - 3\max(2^q, 2^r) - n - \max(2^q, 2^r) = \frac{k-n}{2} - 4\max(2^q, 2^r)$ and we take here $\phi(x) = 1$ for any $x \leq 0$. Using again (2.1) and (2.2) we have also

$$|E(Q^{(3)}_{q,r,n}(k)|\mathcal{G}_n)| \leq \sum_{\frac{k+n}{2}+2\max(2^q,2^r)\leq j<k}|E(\varrho_{q,r}(j,k)|\mathcal{G}_n)|$$
$$\leq 8(\rho(2^r) + \rho(2^{r-1}))(\rho(2^q) + \rho(2^{q-1}))(k - n)\phi(\tfrac{k-n}{2})$$

since each $\varrho_{q,r}(j,k)$ here is $\mathcal{F}_{\frac{k+n}{2}+\max(2^q,2^r),\infty}$-measurable and $\frac{k+n}{2} + \max(2^q, 2^r) - n - \max(2^q, 2^r) = \frac{k-n}{2}$.

Hence, by Lemmas 3.2 and 3.3,

$$\|\max_{m\leq n\leq N}|\sum_{k=m}^{n}Q_{q,r}(k)|\|_{2M} \leq 2A^{q,r}_{2M}\big(3(2M)!d^M(N-m)^M + N - m\big)^{1/2M}$$

where by the above

$$A^{q,r}_{2M} = \sup_{0\leq n\leq N}\sum_{k\geq n}\|E(Q_{q,r}(k)|\mathcal{G}_n)\|_{2M}$$
$$\leq 64(\rho(2^r) + \rho(2^{r-1}))(\rho(2^q) + \rho(2^{q-1}))$$
$$\times\big((3(2M)!)^{1/2M}\sqrt{d}(\sum_{i=0}^{\infty}\phi(i) + \max(2^q, 2^r))^2\max_{m\leq j\leq N}\sqrt{j-\ell(j)}$$
$$+ \max(2^q, 2^r)(4\max(2^q, 2^r) + \sum_{i=0}^{\infty}\phi(i)) + \sum_{i=1}^{\infty}i\phi(i)\big).$$

This together with

$$\|\max_{m\leq n\leq N}|\mathbb{R}(\ell, m, n)|\|_{2M} \leq \sum_{q,r=0}^{\infty}\|\max_{m\leq n\leq N}|\sum_{k=m}^{n}Q_{q,r}(k)|\|_{2M}$$

yields (3.4) completing the proof of the lemma. □

3.2 Characteristic functions estimates

Next, we will follow the same path as in [9] (see also [17] and references there) which leads to strong approximation (almost sure invariance principle) theorem for sums of weakly dependent random vectors. For each $n \geq 1$ introduce the characteristic function

$$f_n(w) = E\exp(i\langle w,\, n^{-1/2}\sum_{k=0}^{n-1}\xi(k)\rangle),\; w \in \mathbb{R}^e$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. In the same way as in Lemma 3.10 from [32] we obtain

Lemma 3.5. *For any $n \geq 1$,*

$$|f_n(w) - \exp(-\frac{1}{2}\langle \varsigma w,\, w\rangle)| \leq C_2 e^3 n^{-\wp} \tag{3.5}$$

for all $w \in \mathbb{R}^e$ with $|w| \leq n^{\wp/2}$ where we can take $\wp \leq \frac{1}{20}$ and a constant $C_2 > 0$ independent of n and e.

Next, we follow [9] and [36] introducing blocks of high polynomial power length with gaps between them. Set $m_0 = 0$ and recursively $n_k = m_{k-1} + [k^\beta]$, $m_k = n_k + [k^{\beta/4}]$, $k = 1, 2, ...$ where $\beta > 0$ is big and will be chosen later on. Now we define sums

$$Q_k = \sum_{m_{k-1}\leq j<n_k} E\big(\xi(j)|\mathcal{F}_{j-\frac{1}{3}[(k-1)^{\beta/4}],\,j+\frac{1}{3}[(k-1)^{\beta/4}]}\big)$$
$$\text{and}\quad R_k = \sum_{n_k\leq j<m_k}\xi(j),\; k = 1, 2, ...$$

where the first sums play the role of blocks while the second ones are gaps whose total contribution turns out to be negligible for our purposes. Set also $\ell_N(t) = \max\{k : m_k \leq Nt\}$ and $\ell_N = \ell_N(T)$.

Lemma 3.6. *With probability one for all $N \geq 1$,*

$$\sup_{0\leq t\leq T}|S_N(t) - N^{-1/2}\sum_{1\leq k\leq \ell_N(t)} Q_k| = O(N^{-\delta}) \tag{3.6}$$

provided $\delta > 0$ is small and $\beta > 0$ is large enough.

Proof. Denote the left hand side of (3.6) by I, then

$$I \leq \sup_{0\leq t\leq T} I_1(t) + \sup_{0\leq t\leq T} I_2(t) + \sup_{0\leq t\leq T} I_3(t).$$

Here,

$$\begin{aligned} I_1(t) = N^{-1/2}\big|\sum_{1\leq k\leq \ell_N(t)}\big(\sum_{m_{k-1}\leq j<n_k}(\xi(j)\\ -E(\xi(j)|\mathcal{F}_{j-\frac{1}{3}[(k-1)^{\beta/4}],\,j+\frac{1}{3}[(k-1)^{\beta/4}]})\big)\big|\\ \leq N^{-1/2}\sum_{1\leq k\leq \ell_N(t)} k^\beta\rho(\tfrac{1}{3}[(k-1)^{\beta/4}]) \leq C_3 N^{-(\frac{1}{2}-\frac{1}{\beta+1})} \end{aligned} \tag{3.7}$$

since

$$(\frac{TN}{2})^{\frac{1}{\beta+1}} - 2 \leq \ell_N \leq (TN)^{\frac{1}{\beta+1}} \tag{3.8}$$

and by (2.12),

$$C_3 = \max_{k\geq 1}(k^\beta\rho(\frac{1}{3}[(k-1)^{\beta/4}])) < \infty.$$

It remains to estimate

$$I_2(t) = N^{-1/2}\big|\sum_{1\leq k\leq \ell_N(t)}\sum_{n_k\leq j<m_k}\xi(j)\big|$$
$$\text{and}\;\; I_3(t) = N^{-1/2}\big|\sum_{m_{\ell_N(t)}\leq j<[Nt]}\xi(j)\big|.$$

By (3.8) and Lemma 3.4 with $C_1(M)=C_1^\xi(M)$,

$$E\sup_{0\le t\le T}I_2^{2M}(t)\le N^{-M}\sum_{0\le l\le \ell_N}E\big|\sum_{1\le k\le l}\sum_{n_k\le j<m_k}\xi(j)\big|^{2M}$$
$$\le C_1(M)N^{-M}\sum_{0\le l\le\ell_N}(\sum_{1\le k\le l}k^{\beta/4})^M\le C_4(M)N^{-\frac{1}{4}(3-\frac{7}{\beta+1})M+1}$$

where $C_4(M)=C_1(M)3^{(\frac{\beta}{4}+2)M+1}$. By the Chebyshev inequality

$$P\{\sup_{0\le t\le T}I_2(t)>N^{-\delta}\}\le N^{2M\delta}E\sup_{0\le t\le T}I_2^{2M}(t). \tag{3.9}$$

Choosing $\beta>0$ big and $\delta>0$ small enough so that $\frac{7}{\beta+1}+8\delta\le 1$ and taking $M\ge 6$ we obtain that the right hand side of (3.9) is bounded by N^{-2}, and so by the Borel-Cantelli lemma

$$\sup_{0\le t\le T}I_2(t)=O(N^{-\delta})\quad\text{a.s.} \tag{3.10}$$

Next,

$$\sup_{0\le t\le T}I_3^{2M}(t)=N^{-M}\max_{1\le k\le\ell_N}\max_{m_k\le j<m_{k+1}\wedge Nt}|\sum_{m_k\le i\le j}\xi(i)|^{2M}$$
$$\le N^{-M}\sum_{1\le k\le\ell_N}\sum_{m_k\le j<m_{k+1}}|\sum_{m_k\le i\le j}\xi(i)|^{2M}.$$

By Lemma 3.4 with $C_1(M)=C_1^\xi(M)$,

$$E|\sum_{m_k\le i\le j}\xi(i)|^{2M}\le C_1(M)(j-m_k+1)^M,$$

and so by (3.8),

$$E\sup_{0\le t\le T}I_3^{2M}(t)\le C_1(M)N^{-M}\sum_{k=1}^{\ell_N}(m_{k+1}-m_k+1)^{M+1}\le C_5(M)N^{-\frac{M}{\beta+1}+1}$$

where $C_5(M)=C_1(M)3^{M+1}$. By the Chebyshev inequality

$$P\{\sup_{0\le t\le T}I_3(t)>N^{-\delta}\}\le N^{2M\delta}E\sup_{0\le t\le T}I_3^{2M}(t). \tag{3.11}$$

Choosing $\delta\le\frac{1}{4(\beta+1)}$ and $M\ge 12(\beta+1)$ we bound the right hand side of (3.11) by N^{-2} which together with the Borel-Cantelli lemma yields that

$$\sup_{0\le t\le T}I_3(t)=O(N^{-\delta})\quad\text{a.s.} \tag{3.12}$$

Finally, (3.6) follows from (3.7), (3.10) and (3.12). □

Next, set

$$\mathcal{G}_k=\mathcal{F}_{-\infty,n_k+\frac{1}{3}[k^{\beta/4}]}, \tag{3.13}$$

and so Q_k is $\mathcal{G}_k$-measurable. The following result is a corollary of Lemmas 3.1 and 3.5.

Lemma 3.7. *For any $k\ge 1$,*

$$|E(\exp(i\langle w,(n_k-m_{k-1})^{-1/2}Q_k\rangle|\mathcal{G}_{k-1})-\exp(-\tfrac{1}{2}\langle\varsigma w,w\rangle)|$$
$$\le 2\phi(\tfrac{1}{3}[(k-1)^{\beta/4}])+\rho(\tfrac{1}{3}[(k-1)^{\beta/4}])+C_2d^3[k^\beta]^{-\wp} \tag{3.14}$$

for all $w\in\mathbb{R}^e$ with $|w|\le(n_k-m_{k-1})^{\wp/2}$.

Proof. Set

$$F_k = \exp(i\langle w, (n_k - m_{k-1})^{-1/2} Q_k\rangle).$$

Then by Lemma 3.1,

$$|E(F_k|\mathcal{G}_{k-1}) - EF_k| \le 2\phi(\frac{1}{3}[(k-1)^{\beta/4}]).$$

Since $|e^{i(a+b)} - e^{ib}| \le |a|$, we obtain from (2.3) taking into account the stationarity of $\xi(k)$'s that,

$$|EF_k - f_{n_k - m_{k-1}}(w)| \le |w| k^{\beta/2} \rho(\frac{1}{3}[(k-1)^{\beta/4}]),$$

and (3.14) follows from (3.5). □

3.3 Strong approximations

We will rely on the following result which is a version of Theorem 1 in [9] with some features taken from Theorem 4.6 in [17] (see also Theorem 3 in [42]).

Theorem 3.8. *Let $\{V_k, k \ge 1\}$ be a sequence of random vectors with values in $\mathbb{R}^e$ defined on some probability space $(\Omega, \mathcal{F}, P)$ and such that V_k is measurable with respect to $\mathcal{G}_k$, $k = 1, 2, ...$ where $\mathcal{G}_k$, $k \ge 1$ is a filtration of countably generated sub-σ-algebras of $\mathcal{F}$. Assume that the probability space is rich enough so that there exists on it a sequence of uniformly distributed on $[0, 1]$ independent random variables U_k, $k \ge 1$ independent of $\vee_{k\ge 1}\mathcal{G}_k$. Let G be a probability distribution on $\mathbb{R}^e$ with the characteristic function g. Suppose that for some nonnegative numbers ν_m, δ_m and $K_m \ge 10^8 d$,*

$$E\big|E(\exp(i\langle w, V_k\rangle)|\mathcal{G}_{k-1}) - g(w)\big| \le \nu_k \tag{3.15}$$

for all $w \in \mathbb{R}^e$ with $|w| \le K_k$ and

$$G\{x : |x| \ge \frac{1}{4} K_k\} < \delta_k. \tag{3.16}$$

Then there exists a sequence $\{W_k, k \ge 1\}$ of $\mathbb{R}^e$-valued random vectors defined on $(\Omega, \mathcal{F}, P)$ with the properties

(i) W_k is $\mathcal{G}_k \vee \sigma\{U_k\}$-measurable for each $k \ge 1$;

(ii) each W_k, $k \ge 1$ has the distribution G and W_k is independent of $\sigma\{U_1, ..., U_{k-1}\} \vee \mathcal{G}_{k-1}$, and so also of $W_1, ..., W_{k-1}$;

(iii) Let $\varrho_k = 16 K_k^{-1} \log K_k + 2\nu_k^{1/2} K_k^d + 2\delta_k^{1/2}$. Then

$$P\{|V_k - W_k| \ge \varrho_k\} \le \varrho_k. \tag{3.17}$$

In order to apply Theorem 3.8 we take $V_k = (n_k - m_{k-1})^{-1/2} Q_k$, $\mathcal{G}_k$ given by (3.13) and

$$g(w) = \exp(-\frac{1}{2}\langle \varsigma w, w\rangle)$$

so that G is the mean zero d-dimensional Gaussian distribution with the covariance matrix ς. Relying on Lemma 3.7 we take $\wp = \frac{1}{20}$,

$$K_k = (n_k - m_{k-1})^{\wp/4d} \le (n_k - m_{k-1})^{\wp/2} \text{ and } \nu_k = C_6 k^{-\beta\wp}$$

for some $C_6 > 0$ independent of $k \ge 1$. By the Chebyshev inequality we have also

$$\begin{aligned} G\{x : |x| \ge \tfrac{K_k}{4}\} &= P\{|\Psi| \ge \tfrac{1}{4}(n_k - m_{k-1})^{\wp/4d}\} \\ &\le 4d\|\varsigma\|(n_k - m_{k-1})^{-\wp/2d} \le C_7 k^{-\wp\beta/2d} \end{aligned}$$

for some $C_7 > 0$ which does not depend on k.

Now Theorem 3.8 provides us with random vectors $W_k,\ k \geq 1$ satisfying the properties (i)–(iii), in particular, the random vector W_k has the mean zero Gaussian distribution with the covariance matrix ς, it is independent of $W_1, ..., W_{k-1}$ and the property (iii) holds true with

$$\varrho_k = 4\frac{\wp}{d}(n_k - m_{k-1})^{-\wp/4d}\log(n_k - m_{k-1}) + 2C_6^{1/2}k^{-\beta\wp/4} + 2C_7^{1/2}k^{-\beta\wp/4d}.$$

Next, we choose $\beta > 160d$ which gives

$$\varrho_k \leq C_8 k^{-2} \tag{3.18}$$

for all $k \geq 1$ where $C_8 > 0$ does not depend on k.

Next, let $W(t),\ t \geq 0$ be a e-dimensional Brownian motion with the covariance matrix ς at the time 1. Then the sequence of random vectors $\tilde{W}_k = (n_k - m_{k-1})^{-1/2}(W(n_k) - W(m_{k-1})),\ k = 1, 2, ...$ and $W_k,\ k \geq 1$ have the same distributions. Moreover, we can redefine the process $\xi(n),\ -\infty < n < \infty$ and choose a Brownian motion $W(s),\ s \geq 0$ (with the covariance matrix ς at the time 1) preserving their distributions so that the joint distribution of the sequences of pairs (V_k, W_k) and of $(\tilde{V}_k, \tilde{W}_k)$, where $\tilde{V}_k$ is constructed as V_k but using the redefined process $\xi(n)$, will be the same. Indeed, by the Kolmogorov extension theorem (see, for instance, [4]) such pair of processes exists if we impose consistent restrictions on their joint finite dimensional distributions. But since the pair of processes $\xi(n),\ n \geq 0$ and $W_k,\ k \geq 1$ satisfying the conditions of Theorem 3.8 exist as asserted there, these conditions are consistent and the required pair of processes indeed exists. A more precise justification of this relies on Lemma A1 from [9]. Namely, let R be the joint distribution of the process $\xi(n),\ -\infty < n < \infty$ and of the sequence $W_k,\ k \geq 1$ and let $\tilde{R}$ be the joint distribution of the sequence $\tilde{W}_k,\ k \geq 1$ and a e-dimensional Brownian motion $W(t),\ t \geq 0$ with the covariance matrix ς at the time 1. Since the second marginal of R coincides with the first marginal of $\tilde{R}$, it follows by Lemma A1 of [9] that the process ξ and the sequence $W_k,\ k \geq 1$ can be redefined on a richer probability space where there exists a Brownian motion $W(t),\ t \geq 0$ with the covariance matrix ς (at the time 1) such that $W_k = (n_k - m_{k-1})^{-1/2}(W(n_k) - W(m_{k-1}))$, and so from now on we will rely on this equality and assume that these W_k's satisfy (3.17) with ϱ_k satisfying (3.18). It follows by the Borel-Cantelli lemma that there exists a random variable $D = D(\omega) < \infty$ a.s. such that

$$|V_k - W_k| \leq Dk^{-2} \quad \text{a.s.} \tag{3.19}$$

Now we can obtain the following result.

Lemma 3.9. *With probability one,*

$$\sup_{0 \leq t \leq T} |\sum_{1 \leq k \leq \ell_N(t)} Q_k - W(tN)| = O(N^{\frac{1}{2} - \delta}) \tag{3.20}$$

for some $\delta > 0$ which does not depend on N.

Proof. We have

$$J_N(t) = |\sum_{1 \leq k \leq \ell_N(t)} Q_k - W(tN)| \leq J_N^{(1)}(t) + |J_N^{(2)}(t)|$$

where by (3.8) and (3.19),

$$J_N^{(1)}(t) = \sum_{1 \leq k \leq \ell_N(t)} (n_k - m_{k-1})^{1/2}|V_k - W_k| \leq D \sum_{1 \leq k \leq \ell_N} [k^\beta]^{1/2}k^{-2} \leq D2^{\beta/2}N^{\frac{1}{2}(1 - \frac{3}{\beta+1})} \tag{3.21}$$

and

$$J_N^{(2)}(t)=W(tN)-W(m_{\ell_N(t)})+\sum_{1\le k\le \ell_N(t)}(W(m_k)-W(n_k)).$$

Clearly, $J_N^{(2)}(t)$, $t\ge 0$ is a martingale in t, and so

$$E\sup_{0\le t\le T}|J_N^{(2)}(t)|^{2M}\le (\tfrac{2M}{2M-1})^{2M}E|J_N^{(2)}(T)|^{2M} \\ \le (\tfrac{4M}{2M-1})^{2M}(E|W(TN)-W(m_{\ell_N})|^{2M}+E|J_N^{(3)}(T)|^{2M}) \tag{3.22}$$

where $J_N^{(3)}(t)=\sum_{1\le k\le \ell_N}(W(m_k)-W(n_k))$. Now, by (3.8),

$$E|W(TN)-W(m_{\ell_N})|^{2M}\le \|\varsigma^{1/2}\|^{2M}(\prod_{j=1}^M(2j-1))(2\ell_N^\beta)^M \\ \le \|\varsigma^{1/2}\|^{2M}\sqrt{2M!}2^MN^{(1-\frac{1}{\beta+1})M}. \tag{3.23}$$

Next, $J_N^{(3)}(T)$ is a sum of independent random vectors and we will estimate last term in the right hand side of (3.22) relying on Lemma 3.2 with $\eta_j=W(m_j)-W(n_j)$ and $\mathcal{G}_j=\sigma\{W(t),\,t\le m_j\}$ for $j=1,...,\ell_N$. Then

$$A_{2M}=\sup_{i\ge 1}\sum_{j\ge i}\|E(\eta_j|\mathcal{G}_i)\|_{2M}=\sup_{i\ge 1}\|\eta_i\|_{2M}\le\sqrt{2M}(\ell_N+1)^{\beta/8},$$

and so by Lemma 3.2,

$$E|J_N^{(3)}(T)|^{2M}\le 3(2M)!(2M)^M(\ell_N+1)^{M(1+\frac{\beta}{4})} \\ \le 3(2M)!(2M)^M2^{M(1+\frac{\beta}{4})}N^{\frac{M}{4}(1+\frac{3}{\beta+1})}. \tag{3.24}$$

By (3.22)–(3.24) and the Chebyshev inequality

$$P\{\sup_{0\le t\le T}|J_N^{(2)}(t)|\ge N^{\frac12-\delta}\}\le N^{-M(1-2\delta)}E\sup_{0\le t\le T}|J_N^{(2)}(t)|^{2M} \\ \le C_9(M)(N^{-M(\frac{1}{\beta+1}-2\delta)}+N^{-M(\frac34-\frac{1}{\beta+1}-2\delta)}) \tag{3.25}$$

where $C_9(M)>0$ does not depend on N. Now choose $\beta\ge 2$, $\delta<\frac{1}{4(\beta+1)}$ and $M\ge 2(\beta+1)$ then the right hand side of (3.25) forms a converging sequence, and so by the Borel-Cantelli lemma

$$\sup_{0\le t\le T}|J_N^{(2)}(t)|=O(N^{\frac12-\delta})\quad\text{a.s.}$$

which together with (3.21) completes the proof of Lemma 3.9. □

Now combining Lemmas 3.6 and 3.9 we obtain that for some $\delta>0$ and all $N\ge 1$,

$$\sup_{0\le t\le T}|S_N(t)-W_N(t)|=O(N^{-\delta})\quad\text{a.s.}, \tag{3.26}$$

where $W_N(t)=N^{-1/2}W(tN)$ is another Brownian motion.

3.4 p-variation norm estimates

Thus, (2.16) is proved in the supremum norm and we will extend it next for the p-variation norm. First, for any $\alpha\in(0,1)$ and $\beta\in(0,1/2)$ (which has nothing to do with β in the previous subsection) we estimate by Lemma 3.4 and the Chebyshev inequality

$$P\big\{\max_{[TN]\wedge(k+N^\alpha)\ge l>k\ge 0}\tfrac{|S_N(l/N)-S_N(k/N)|}{|(l-k)/N|^{\frac12-\beta}}>1\big\} \tag{3.27}$$
$$\le\sum_{[TN]\wedge(k+N^\alpha)\ge l>k\ge 0}P\big\{|\sum_{k\le j<l}\xi(j)|>N^\beta(l-k)^{\frac12-\beta}\big\} \\ \le N^{-2\beta M}\sum_{[TN]\wedge(k+N^\alpha)\ge l>k\ge 0}(l-k)^{-M+2\beta M}E|\sum_{k\le j<l}\xi(j)|^{2M} \\ \le C_1^\xi(M)TN^{1-2\beta M}\sum_{N^\alpha\ge j\ge 1}j^{2\beta M}\le 2^{2\beta M+1}C_1^\xi(M)TN^{-2\beta M(1-\alpha)+1+\alpha}.$$

Next, we choose $M \geq \frac{3+\alpha}{2\beta(1-\alpha)}$ which makes the right hand side of (3.27) a term in a converging series. Hence, by the Borel–Cantelli lemma for any $\alpha \in (0,1)$ and $\beta \in (0,1/2)$ there exists a finite a.s. random variable $C^S_{\alpha,\beta} = C^S_{\alpha,\beta}(\omega)$ such that for all $N \geq 1$,

$$|S_N(l/N) - S_N(k/N)| \leq C^S_{\alpha,\beta}\Big|\frac{l}{N} - \frac{k}{N}\Big|^{\frac{1}{2}-\beta} \text{ if } k + N^\alpha \geq l > k \geq 0,\, l \leq [TN]. \tag{3.28}$$

Next, define

$$\hat{W}_N(t) = W_N(\frac{[Nt]}{N}) = \sum_{0\leq k<[Nt]} (W_N(\frac{k+1}{N}) - W_N(\frac{k}{N})).$$

Let $0 = t_0 < t_1 < ... < t_m = T$ and observe that if $[t_i N] = [t_{i+1}N]$ then

$$S_N(t_{i+1}) = S_N(t_i) \quad \text{and} \quad \hat{W}_N(t_{i+1}) = \hat{W}_N(t_i).$$

Hence,

$$\begin{aligned} &\textstyle\sum_{0\leq i<m} |S_N(t_{i+1}) - \hat{W}_N(t_{i+1}) - S_N(t_i) + \hat{W}_N(t_i)|^p \\ &= \textstyle\sum_{0\leq j<n} |S_N(\frac{k_{j+1}}{N}) - W_N(\frac{k_{j+1}}{N}) - S_N(\frac{k_j}{N}) + W_N(\frac{k_j}{N})|^p \\ &\leq J^{(1)}_N + 2^{p-1}(J^{(2)}_N + J^{(3)}_N) \end{aligned} \tag{3.29}$$

where

$$J^{(1)}_N = \sum_{0\leq j<n,\, k_{j+1}-k_j>N^\alpha} |S_N(\frac{k_{j+1}}{N}) - W_N(\frac{k_{j+1}}{N}) - S_N(\frac{k_j}{N}) + W_N(\frac{k_j}{N})|^p,$$

$$J^{(2)}_N = \sum_{0\leq j<n,\, k_{j+1}-k_j\leq N^\alpha} |S_N(\frac{k_{j+1}}{N}) - S_N(\frac{k_j}{N})|^p,$$

$$J^{(3)}_N = \sum_{0\leq j<n,\, k_{j+1}-k_j\leq N^\alpha} |W_N(\frac{k_{j+1}}{N}) - W_N(\frac{k_j}{N})|^p,$$

$k_j = [t_{i_j} N]$ and $0 = t_{i_0} < t_{i_1} < ... < t_{i_n} = T$ is the maximal subsequence of $t_0, t_1, ..., t_m$ such that $[t_{i_j} N] < [t_{i_{j+1}} N]$, $j = 0, 1, ..., n-1$.

In order to estimate $J^{(1)}_N$ we use (3.26) and observe that there exist no more than $[TN^{1-\alpha}]$ intervals $[k_j, k_{j+1}]$ such that $k_{j+1} - k_j > N^\alpha$ which gives that

$$J^{(1)}_N \leq C^{(1)} N^{1-\alpha-p\delta} \tag{3.30}$$

for some a.s. finite random variable $C^{(1)} = C^{(1)}(\omega) > 0$ which does not depend on N, n or on the choice of $t_1, ..., t_m$ and we choose α so close to 1 that $1 - \alpha - p\delta < 0$. In order to estimate $J^{(2)}_N$ we use (3.28) and observe that $\sum_{0\leq j<n} |k_{j+1} - k_j| \leq TN$ which gives

$$\begin{aligned} J^{(2)}_N &\leq (C^S_{\alpha,\beta})^p N^{-p(\frac{1}{2}-\beta)} \textstyle\sum_{0\leq j<n,\, k_{j+1}-k_j\leq N^\alpha} |k_{j+1} - k_j|^{p(\frac{1}{2}-\beta)} \\ &\leq (C^S_{\alpha,\beta})^p T N^{-(1-\alpha)(\frac{p}{2}-1-p\beta)} \end{aligned} \tag{3.31}$$

where we use that

$$|k_{j+1} - k_j|^{p(\frac{1}{2}-\beta)} \leq N^{\alpha(p(\frac{1}{2}-\beta)-1)} |k_{j+1} - k_j|$$

when $|k_{j+1} - k_j| \leq N^\alpha$ and assume that

$$0 < \beta < \frac{1}{2} - \frac{1}{p} \tag{3.32}$$

which is consistent since $p > 2$.

It remains to estimate $J_N^{(3)}$ (which essentially is contained in the proof of Hölder continuity of the Brownian motion). We start with the estimate similar to (3.27) relying on the Chebyshev inequality and the standard moment estimates of Gaussian random vectors,

$$P\Big\{ \max_{[TN]\wedge(k+N^\alpha)\geq l>k\geq 0} \frac{|W_N(l/N)-W_N(k/N)|}{|(l-k)/N|^{\frac{1}{2}-\beta}} > 1\Big\} \tag{3.33}$$
$$\leq \sum_{[TN]\wedge(k+N^\alpha)\geq l>k\geq 0} P\Big\{\frac{|W_N(l/N)-W_N(k/N)|}{|(l-k)/N|^{\frac{1}{2}-\beta}} > 1\Big\}$$
$$\leq C_{10}(M)N^{-2\beta M(1-\alpha)+1+\alpha}$$

where $C_{10}(M) > 0$ does not depend on $N \geq 1$. Choose again $M \geq \frac{3+\alpha}{2\beta(1-\alpha)}$ which makes the right hand side of (3.33) a term in a converging series. Again, by the Borel–Cantelli lemma for any $\alpha \in (0,1)$ and $\beta \in (0,1/2)$ there exists a finite a.s. random variable $C_{\alpha,\beta} = C_{\alpha,\beta}(\omega)$ such that for all $N \geq 1$,

$$|W_N(l/N) - W_N(k/N)| \leq C_{\alpha,\beta}\Big|\frac{l}{N} - \frac{k}{N}\Big|^{\frac{1}{2}-\beta} \text{ if } k + N^\alpha \geq l > k \geq 0,\, l \leq [TN]. \tag{3.34}$$

Proceeding as in (3.31) we obtain that

$$J_N^{(3)} \leq C^p_{\alpha,\beta} T N^{-(1-\alpha)(\frac{p}{2}-1-p\beta)} \tag{3.35}$$

where we assume again (3.32). Finally, we conclude from (3.29)–(3.31) and (3.35) that

$$\|S_N - \hat{W}_N\|_{p,[0,T]} = O(N^{-\tilde{\delta}}) \quad \text{a.s.}, \tag{3.36}$$

where $\tilde{\delta} = \min(\alpha+p\delta-1, (1-\alpha)(\frac{p}{2}-1-p\beta)) > 0$, assuming (3.32) and choosing $\alpha > 1-p\delta$.

Next, set

$$\bar{\mathbb{W}}_N^{ij}(t) = \int_0^t W_N^i(u)dW_N^j(u)$$

and

$$\hat{\mathbb{W}}_N^{ij}(t) = \sum_{0\leq l<[Nt]} (W_N^j(\frac{l+1}{N}) - W_N^j(\frac{l}{N}))W_N^i(\frac{l}{N}).$$

The relation (2.16) from Theorem 2.2 follows from (3.36) and the following result which will be used also in Sections 4 and 5.

Lemma 3.10. *For all $T > 0$ and $p \in (2,3)$,*

$$\|W_N - \hat{W}_N\|_{p,[0,T]} = O(N^{-\delta}) \quad \textit{a.s. as} \quad N \geq 1 \tag{3.37}$$

and

$$\max_{1\leq i,j\leq d} \|\bar{\mathbb{W}}_N^{ij} - \hat{\mathbb{W}}_N^{ij}\|_{\frac{p}{2},[0,T]} = O(N^{-\delta}) \quad \textit{a.s. as} \quad N \geq 1 \tag{3.38}$$

for some $\delta > 0$ which does not depend on $N \geq 1$.

We postpone the proof of this result till the end of Section 6. Observe that Lemma 3.10 gives estimates for a version of the Euler–Maruyama rough paths approximation in p-variation norms for a family of processes depending on a parameter which by itself controls the step of approximations.

4 Strong approximations for iterated sums

4.1 Estimates in the supremum norm

As in the previous section we will prove first the estimate (2.17) in the supremum norm and then will extend it to the $p/2$-variation norm. Set $m_N = [N^{1-\kappa}]$ with a small $\kappa > 0$ which will be chosen later on, $\nu_N(l) = \max\{jm_N : \, jm_N < l\}$ if $l > m_N$ and

$$R_i(k) = R_i(k,N) = \sum_{l=(k-1)m_N}^{km_N-1} \xi_i(l) \text{ for } k = 1,2,...,\iota_N(T)$$

where $\iota_N(t) = [[Nt]m_N^{-1}]$. For $1 \le i,j \le e$ define

$$\mathbb{U}_N^{ij}(t) = N^{-1} \sum_{l=m_N}^{\iota_N(t)m_N-1} \xi_j(l) \sum_{k=0}^{\nu_N(l)} \xi_i(k) = N^{-1} \sum_{1<l\le\iota_N(t)} R_j(l) \sum_{k=0}^{(l-1)m_N-1} \xi_i(k). \tag{4.1}$$

Set also

$$\bar{\mathbb{S}}_N^{ij}(t) = \mathbb{S}_N^{ij}(t) - t\sum_{l=1}^{\infty} E(\xi_i(0)\xi_j(l)).$$

We will need the following result.

Lemma 4.1. *For all* $i,j = 1,...,e$ *and* $N \ge 1$*,*

$$\sup_{0\le t\le T} |\bar{\mathbb{S}}_N^{ij}(t) - \mathbb{U}_N^{ij}(t)| = O(N^{-\delta_1}) \text{ a.s.} \tag{4.2}$$

for some $\delta_1 > 0$ *which does not depend on* N*.*

Proof. First, we write

$$|\bar{\mathbb{S}}_N^{ij}(t) - \mathbb{U}_N^{ij}(t)| \le |\mathcal{I}_N^{(1)}(t)| + |\mathcal{I}_N^{(2)}(t)| + |\mathcal{I}_N^{(3)}(t)| + |\mathcal{I}_N^{(4)}(t)| \tag{4.3}$$

where

$$\mathcal{I}_N^{(1)}(t) = \mathcal{I}_N^{1;ij}(t) = N^{-1} \sum_{l=m_N}^{\iota_N(t)m_N-1} \sum_{k=\nu_N(l)+1}^{l-1} \big(\xi_j(l)\xi_i(k) - E(\xi_j(l)\xi_i(k))\big),$$

$$\mathcal{I}_N^{(2)}(t) = \mathcal{I}_N^{2;ij}(t) = N^{-1} \sum_{l=\iota_N(t)m_N}^{[Nt]-1} \sum_{k=0}^{l-1} \big(\xi_j(l)\xi_i(k) - E(\xi_j(l)\xi_i(k))\big),$$

$$\mathcal{I}_N^{(3)}(t) = \mathcal{I}_N^{3;ij}(t) = N^{-1} \sum_{l=1}^{m_N-1} \sum_{k=0}^{l-1} \big(\xi_j(l)\xi_i(k) - E(\xi_j(l)\xi_i(k))\big)$$

and

$$\mathcal{I}_N^{(4)}(t) = \mathcal{I}_N^{4;ij}(t) = \mathcal{I}_N^{4,1;ij}(t) - \mathcal{I}_N^{4,2;ij}(t)$$

with

$$\mathcal{I}_N^{4,1;ij}(t) = N^{-1} \sum_{l=1}^{[Nt]-1} \sum_{k=0}^{l-1} E(\xi_j(l)\xi_i(k)) - t\sum_{l=1}^{\infty} E(\xi_i(0)\xi_j(l))$$

and

$$\mathcal{I}_N^{4,2;ij}(t) = N^{-1} \sum_{l=m_N}^{\iota_N(t)m_N-1} \sum_{k=0}^{\nu_N(l)} E(\xi_j(l)\xi_i(k)).$$

By Lemma 3.4,

$$E\sup_{0\le t\le T}|\mathcal{I}_N^{(1;i,j)}(t)|^{2M}=E\max_{0\le k<[TN]}|\mathcal{I}_N^{(1;i,j)}(k/N)|^{2M} \tag{4.4}$$
$$\le C_{11}(M)T^MN^{-M\kappa},\ E\sup_{0\le t\le T}|\mathcal{I}_N^{(2;i,j)}(t)|^{2M}\le C_{11}(M)T^MN^{-M\kappa}$$
$$\text{and }\ E\sup_{0\le t\le T}|\mathcal{I}_N^{(3;i,j)}(t)|^{2M}\le C_{11}(M)N^{-2M\kappa}$$

where $C_{11}(M)>0$ does not depend on N. Choosing δ_1 and M such that $\delta_1<\frac{1}{2}\kappa$ and $M\ge 2(\kappa-2\delta_1)^{-1}$ we use, first, the Chebyshev inequality and then the Borel–Cantelli lemma to obtain that with probability one,

$$\sup_{0\le t\le T}|\mathcal{I}_N^{(1;i,j)}(t)|=O(N^{-\delta_1}),\ \sup_{0\le t\le T}|\mathcal{I}_N^{(2;i,j)}(t)|=O(N^{-\delta_1}) \tag{4.5}$$
$$\text{and }\ \sup_{0\le t\le T}|\mathcal{I}_N^{(3;i,j)}(t)|=O(N^{-2\delta_1}).$$

It remains to estimate $\mathcal{I}_N^{(4;i,j)}(t)$. By (2.3), (2.6) and Lemma 3.1,

$$|E(\xi_j(l)\xi_i(k))|\le 2L\rho(|k-l|/3)+|E\big(E(\xi_j(l)|\mathcal{F}_{l-[\frac{1}{3}|k-l|],l+[\frac{1}{3}|k-l|]}) \tag{4.6}$$
$$\times E(\xi_i(k)|\mathcal{F}_{k-[\frac{1}{3}|k-l|],k+[\frac{1}{3}|k-l|]})\big)|\le 2L(L\phi(|k-l|/3)+\rho(|k-l|/3)).$$

By the stationarity

$$\sum_{k=0}^{l-1}E(\xi_j(l)\xi_i(k))=\sum_{n=1}^{l}E(\xi_j(n)\xi_i(0))$$

and by (2.12) and (4.6) the limit

$$\mathcal{L}_{ij}=\lim_{l\to\infty}\sum_{n=1}^{l}E(\xi_j(n)\xi_i(0))=\sum_{n=1}^{\infty}E(\xi_j(n)\xi_i(0))$$

exists and, moreover,

$$|\sum_{k=0}^{l-1}E(\xi_j(l)\xi_i(k))-\mathcal{L}_{ij}|\le 2L\sum_{n=l+1}^{\infty}(L\phi(n/3)+\rho(n/3)). \tag{4.7}$$

It follows from (2.12) and (2.23) that

$$\sup_{0\le t\le T}|\mathcal{I}_N^{(4,1;i,j)}(t)|\le 2LN^{-1}\sum_{l=0}^{\infty}\sum_{n=l+1}^{\infty}(L\phi(n/3)+\rho(n/3))+N^{-1}|\mathcal{L}_{ij}|\le C_{12}N^{-1} \tag{4.8}$$

for some $C_{12}>0$ which does not depend on N.

Finally, by (4.6),

$$|\sum_{k=0}^{\nu_N(l)}E(\xi_j(l)\xi_i(k))|\le 2L\sum_{n=l-\nu_N(l)}^{\infty}(L\phi(n/3)+\rho(n/3)),$$

and so by (2.12),

$$\sup_{0\le t\le T}|\mathcal{I}_N^{(4,2;i,j)}(t)|\le 2LTN^{-(1-\kappa)}\sum_{l=1}^{m_N}\sum_{n=l}^{\infty}(L\phi(n/3)+\rho(n/3))\le C_{13}N^{-(1-\kappa)} \tag{4.9}$$

for some $C_{13}>0$ which does not depend on N. The lemma now follows from (4.5), (4.8) and (4.9) together with the Chebyshev inequality and the Borel–Cantelli lemma. □

Now set

$$S_N^i(t)=N^{-1/2}\sum_{0\le k<[Nt]}\xi_i(k),\ i=1,...,e$$

and observe that

$$\mathbb{U}_N^{ij}(t)=\sum_{2\le l\le \iota_N(t)}\big(S_N^j\big(\frac{lm_N}{N}\big)-S_N^j\big(\frac{(l-1)m_N}{N}\big)\big)S_N^i\big(\frac{(l-1)m_N}{N}\big).$$

Define

$$\mathbb{V}_N^{ij}(t)=\sum_{2\le l\le \iota_N(t)}\big(W_N^j\big(\frac{lm_N}{N}\big)-W_N^j\big(\frac{(l-1)m_N}{N}\big)\big)W_N^i\big(\frac{(l-1)m_N}{N}\big)$$

where $W_N=(W_N^1,...,W_N^e)$ is the e-dimensional Brownian motion with the covariance matrix ς (at the time 1) appearing in (2.16) which was constructed in Section 3. Then

$$\begin{aligned}\sup_{0\le t\le T}|\mathbb{U}_N^{ij}(t)-\mathbb{V}_N^{ij}(t)|\le\textstyle\sum_{2\le l\le\iota_N(T)}\Big(\big(\big|S_N^j\big(\tfrac{lm_N}{N}\big)-W_N^j\big(\tfrac{lm_N}{N}\big)\big|\\ +\big|S_N^j\big(\tfrac{(l-1)m_N}{N}\big)-W_N^j\big(\tfrac{(l-1)m_N}{N}\big)\big|\big)\big|S_N^i\big(\tfrac{(l-1)m_N}{N}\big)\big|\\ +\big|W_N^j\big(\tfrac{lm_N}{N}\big)-W_N^j\big(\tfrac{(l-1)m_N}{N}\big)\big|\big|S_N^i\big(\tfrac{(l-1)m_N}{N}\big)-W_N^i\big(\tfrac{(l-1)m_N}{N}\big)\big|\Big).\end{aligned}\tag{4.10}$$

By Lemma 3.4,

$$E\max_{2\le l\le\iota_N(T)}\big|S_N^i\big(\frac{(l-1)m_N}{N}\big)\big|^{2M}\le C_{14}(M)T^M\tag{4.11}$$

for some $C_{14}(M)>0$ which does not depend on N. By the Chebyshev inequality for any $\gamma>0$,

$$P\{\max_{2\le l\le\iota_N(T)}\big|S_N^i\big(\frac{(l-1)m_N}{N}\big)\big|>N^\gamma\}\le C_{14}(M)T^MN^{-2M\gamma}.\tag{4.12}$$

Taking $M\ge\gamma^{-1}$ the right hand side of (4.12) becomes a term of a converging series and by the Borel–Cantelli lemma we obtain that for any $\gamma>0$,

$$\max_{2\le l\le\iota_N(T)}\big|S_N^i\big(\frac{(l-1)m_N}{N}\big)\big|=O(N^\gamma)\quad\text{a.s.}\tag{4.13}$$

Next, write

$$\begin{aligned}&E\max_{2\le l\le\iota_N(T)}\big|W_N^j\big(\tfrac{lm_N}{N}\big)-W_N^j\big(\tfrac{(l-1)m_N}{N}\big)\big|^{2M}\\ &\le\textstyle\sum_{2\le l\le\iota_N(T)}E\big|W_N^j\big(\tfrac{lm_N}{N}\big)-W_N^j\big(\tfrac{(l-1)m_N}{N}\big)\big|^{2M}.\end{aligned}$$

Using the standard moment estimates for the Brownian motion and relying on the Chebyshev inequality and the Borel–Cantelli lemma we obtain similarly to (4.12) and (4.13) that for $\gamma<\kappa/2$,

$$\max_{2\le l\le\iota_N(T)}\big|W_N^j\big(\frac{lm_N}{N}\big)-W_N^j\big(\frac{(l-1)m_N}{N}\big)\big|=O(N^{-\gamma})\quad\text{a.s.}\tag{4.14}$$

Now, combining (2.16), (4.10), (4.13) and (4.14) we obtain that

$$\sup_{0\le t\le T}|\mathbb{U}_N^{ij}(t)-\mathbb{V}_N^{ij}(t)|=O(N^{-\delta_2})\quad\text{a.s.}\tag{4.15}$$

where $\delta_2=\delta-\kappa-\gamma$ and we choose κ and γ so small that $\delta_2>0$.

Next, observe that

$$\begin{aligned}&\sup_{0\le t\le T}\big|\textstyle\int_0^tW_N^i(s)dW_N^j(s)-\mathbb{V}_N^{ij}(t)\big|\\ &\le\sup_{0\le t\le T}|J_N^{(1;i,j)}(t)|+\sup_{0\le t\le m_NN^{-1}}|J_N^{(2;i,j)}(t)|+\sup_{0\le t\le T}|J_N^{(3;i,j)}(t)|\end{aligned}\tag{4.16}$$

where

$$J_N^{(1;i,j)}(t)=\sum_{2\le l\le\iota_N(t)}\int_{(l-1)m_NN^{-1}}^{lm_NN^{-1}}\big(W_N^i(s)-W_N^i\big(\frac{(l-1)m_N}{N}\big)\big)dW^j(s),$$
$$J_N^{(2;i,j)}(t)=\int_0^t W_N^i(s)dW_N^j(s)\text{ and }J_N^{(3;i,j)}(t)=\int_{(\iota_N(t)\vee 1)m_NN^{-1}}^t W_N^i(s)dW_N^j(s).$$

By the standard (martingale) moment estimates for stochastic integrals (see, for instance, [40], Section 1.7),

$$E\sup_{0\le t\le T}|J_N^{(1;i,j)}(t)|^{2M}\le C_{15}(M,T)N^{-M\kappa},$$
$$E\sup_{0\le t\le m_NN^{-1}}|J_N^{(2;i,j)}(t)|^{2M}\le C_{15}(M,T)N^{-M\kappa}\quad\text{and}$$

$$E\sup\nolimits_{0\le t\le T}|J_N^{(3;i,j)}(t)|^{2M}$$
$$\le\sum\nolimits_{2\le l\le\iota_N(T)}E\sup\nolimits_{0\le u\le m_N}|\int_{(l-1)m_NN^{-1}}^{((l-1)m_N+u)N^{-1}}W_N^i(s)dW_N^j(s)|^{2M}$$
$$\le C_{15}(M,T)N^{-(M-1)\kappa}$$

where $C_{15}(M,T)>0$ does not depend on N. Taking $\delta_3<\frac{1}{2}\kappa$ and $M\ge(2\kappa+1)(\kappa-2\delta_3)^{-1}$ and employing the Chebyshev inequality together with the Borel–Cantelli lemma in the same way as above, we conclude that

$$\sup_{0\le t\le T}|J_N^{(1;i,j)}(t)|+\sup_{0\le t\le T}|J_N^{(2;i,j)}(t)|+\sup_{0\le t\le T}|J_N^{(3;i,j)}(t)|=O(N^{-\delta_3})\text{ a.s.}\tag{4.17}$$

This together with (4.2), (4.15) and (4.16) completes the proof of (2.17) in the supremum norm.

4.2 $p/2$-variation norm estimates

Next, we extend the supremum norm estimate of Section 4.1 to the $p/2$-variation norm estimate which will yield (2.17) of Theorem 2.2. First, we will derive certain Hölder continuity type estimates for our sums. For $0\le s<t\le T$ and $i,j=1,...,e$ set

$$\bar{\mathbb{S}}_N^{ij}(s,t)=N^{-1}\sum_{[sN]\le k<l<[Nt]}\xi_i(k)\xi_j(l)-(t-s)\sum_{l=1}^\infty E(\xi_i(0)\xi_j(l)),$$

and so

$$\mathbb{S}_N^{ij}(s,t)=\bar{\mathbb{S}}_N^{ij}(s,t)+(t-s)\sum_{l=1}^\infty E(\xi_i(0)\xi_j(l)),\tag{4.18}$$

recalling that by (2.12) and (4.6) the series in the right hand side of (4.18) converges absolutely.

Lemma 4.2. *There exists a finite a.s. random variable $C^{\$}_{\alpha,\beta}>0$ which does not depend on n,m or N such that*

$$\big|\bar{\mathbb{S}}_N^{ij}(\frac{m}{N},\frac{n}{N}))\big|\le C^{\$}_{\alpha,\beta}|\frac{n}{N}-\frac{m}{N}|^{1-\beta}\text{ provided }m+N^\alpha\ge n>m\ge 0,\ [TN]>n.\tag{4.19}$$

Proof. We will estimate $\bar{\mathbb{S}}_N^{ij}(s,t)$ relying on Lemma 3.4. Set

$$\mu_{kl}^{ij}=\xi_j(l)\sum_{r=k}^{l-1}\xi_i(r)\text{ and }\Sigma_N^{ij}(s,t)=N^{-1}\sum_{l=[sN]}^{[tN]-1}(\mu^{ij}_{[sN]l}-E\mu^{ij}_{[sN]l}).$$

By (2.3), (2.6) and (4.6) for $0 \le m < n < [TN]$,

$$|\bar{\mathbb{S}}_N^{ij}(\frac{m}{N}, \frac{n}{N}) - \Sigma_N^{ij}(\frac{m}{N}, \frac{n}{N})| \le (\frac{n}{N} - \frac{m}{N}) \sum_{l=1}^{\infty} |E(\xi_i(0)\xi_j(l))|. \tag{4.20}$$

Next, for any $\alpha, \beta \in (0,1)$ by the Chebyshev inequality

$$\begin{aligned} \mathcal{P} = P\big\{ \max_{m+N^\alpha \ge n > m \ge 0,\, [TN]>n} \frac{|\Sigma_N^{ij}(\frac{m}{N}, \frac{n}{N})|}{|(n-m)/N|^{1-\beta}} > 1 \big\} \\ \le \sum_{m+N^\alpha \ge n > m \ge 0,\, [TN]>n} P\big\{ |\sum_{m\le r<n} (\mu_{mr}^{ij} - E\mu_{mr}^{ij})| > N^\beta (n-m)^{1-\beta} \big\} \\ \le N^{-2\beta M} \sum_{m+N^\alpha \ge n > m \ge 0,\, [TN]>n} (n-m)^{-2M(1-\beta)} \\ \times E|\sum_{m\le r<n} (\mu_{mr}^{ij} - E\mu_{mr}^{ij})|^{2M}. \end{aligned} \tag{4.21}$$

By Lemma 3.4,

$$E|\sum_{m\le r<n} (\mu_{mr}^{ij} - E\mu_{mr}^{ij})|^{2M} \le C_{16}(M)(n-m)^{2M} \tag{4.22}$$

for some $C_{16}(M) > 0$ which does not depend on m, n or N. This together with (4.21) yields that

$$\mathcal{P} \le C_{17}(M) N^{-2M\beta(1-\alpha)+1+\alpha} \tag{4.23}$$

where $C_{17}(M) = C_{16}(M) 2^{2M\beta+1}(2M\beta+1)^{-1}$. For any $\alpha, \beta \in (0,1)$ we pick up $M \ge 1$ such that $2M\beta(1-\alpha) - 1 - \alpha \ge 2$ which makes the right hand side of (4.23) a term of a converging sequence. Thus, by the Borel–Cantelli lemma we conclude that there exists a finite a.s. random variable $C_{\alpha,\beta}^\Sigma = C_{\alpha,\beta}^\Sigma(\omega)$ such that for all $N \ge 1$,

$$|\Sigma_N^{ij}(\frac{m}{N}, \frac{n}{N}))| \le C_{\alpha,\beta}^\Sigma |\frac{n}{N} - \frac{m}{N}|^{1-\beta} \text{ provided } m + N^\alpha \ge n > m \ge 0,\, [TN] > n,$$

which together with (4.18) and (4.20) yields (4.19) for all $N \ge 1$. □

Next, we proceed similarly to Section 3.4. For $0 \le s < t \le T$ define

$$\hat{\mathbb{W}}_N^{ij}(s,t) = \sum_{[Ns]\le l<[Nt]} (W_N^j(\frac{l+1}{N}) - W_N^j(\frac{l}{N}))(W_N^i(\frac{l}{N}) - W_N^i(\frac{[Ns]}{N}))$$

and

$$\bar{\mathbb{W}}_N^{ij}(s,t) = \int_s^t (W_N^i(u) - W_N^i(s)) dW_N^j(u)$$

while $\hat{\mathbb{W}}_N^{ij}(0,t)$ and $\bar{\mathbb{W}}_N^{ij}(0,t)$ will be denoted, as before, just by $\hat{\mathbb{W}}_N^{ij}(t)$ and $\bar{\mathbb{W}}_N^{ij}(t)$, respectively. Let $0 \le t_0 < t_1 < ... < t_m = T$ and observe that if $[t_q N] = [t_{q+1} N]$ then

$$\hat{\mathbb{W}}_N^{ij}(t_q, t_{q+1}) = 0 \text{ and } |\bar{\mathbb{S}}_N^{ij}(t_q, t_{q+1})| = D_{ij}(t_{q+1} - t_q) \le D_{ij} N^{-1}$$

where $D_{ij} = |\sum_{l=1}^\infty E(\xi_i(0)\xi_j(l))| < \infty$. Hence,

$$\begin{aligned} \sum_{0\le q<m} |\bar{\mathbb{S}}_N^{ij}(t_q, t_{q+1}) - \hat{\mathbb{W}}_N^{ij}(t_q, t_{q+1})|^{p/2} \\ \le \sum_{0\le r<n} |\bar{\mathbb{S}}_N^{ij}(\tfrac{k_r}{N}, \tfrac{k_{r+1}}{N}) - \hat{\mathbb{W}}_N^{ij}(\tfrac{k_r}{N}, \tfrac{k_{r+1}}{N})|^{p/2} \\ + J_N^{(0)} \le J_N^{(0)} + J_N^{(1)} + 2^{p/2-1}(J_N^{(2)} + J_N^{(3)}), \end{aligned} \tag{4.24}$$

where

$$J_N^{(0)} = D_{ij}^{p/2} \sum_{q:|t_{q+1}-t_q|<N^{-1}} |t_{q+1}-t_q|^{p/2} \le D_{ij}^{p/2} T N^{-(\frac{p}{2}-1)}, \tag{4.25}$$

$$J_N^{(1)} = \sum_{0\le r<n, k_{r+1}-k_r>N^\alpha} \big|\bar{\mathbb{S}}_N^{ij}(\frac{k_r}{N}, \frac{k_{r+1}}{N}) - \hat{\mathbb{W}}_N^{ij}(\frac{k_r}{N}, \frac{k_{r+1}}{N})\big|^{p/2},$$

$$J_N^{(2)} = \sum_{0\le r<n, k_{r+1}-k_r\le N^\alpha} \big|\bar{\mathbb{S}}_N^{ij}(\frac{k_r}{N}, \frac{k_{r+1}}{N})\big|^{p/2} \text{ and,}$$

$$J_N^{(3)} = \sum_{0\le r<n, k_{r+1}-k_r\le N^\alpha} \big|\hat{\mathbb{W}}_N^{ij}(\frac{k_r}{N}, \frac{k_{r+1}}{N})|^{p/2},$$

$k_r = [t_{q_r}N]$ and $0 = t_{q_0} < t_{q_1} < ... < t_{q_m} = T$ is the maximal subsequence of $t_0, t_1, ..., t_m$ such that $[t_{q_r}N] < [t_{q_{r+1}}N]$, $r = 0, 1, ..., m-1$.

Observe that

$$\bar{\mathbb{S}}_N^{ij}(s,t) = \bar{\mathbb{S}}_N^{ij}(t) - \bar{\mathbb{S}}_N^{ij}(s) - S_N^i(s)(S_N^j(t) - S_N^j(s)),$$
$$\bar{\mathbb{W}}_N^{ij}(s,t) = \bar{\mathbb{W}}_N^{ij}(t) - \bar{\mathbb{W}}_N^{ij}(s) - W_N^i(s)(W_N^j(t) - W_N^j(s))$$

and

$$\hat{\mathbb{W}}_N^{ij}(s,t) = \hat{\mathbb{W}}_N^{ij}(t) - \hat{\mathbb{W}}_N^{ij}(s) - W_N^i(\frac{[Ns]}{N})(W_N^j(\frac{[Nt]}{N}) - W_N^j(\frac{[Ns]}{N})).$$

In order to estimate $J_N^{(1)}$ we use (2.16) and (2.17) for the supremum norm proved in Sections 3.3 and 4.1 together with Lemma 3.10 and observe that there exists no more than $[TN^{1-\alpha}]$ disjoint intervals (k_r, k_{r+1}) in $[0, [TN]]$ with the length exceeding N^α which gives as in (3.30) that

$$J_N^{(1)} \le C_{18} N^{1-\alpha-p\delta/2}(1 + \sup_{0\le t\le T} |S_N^j(t)|^{p/2} + \sup_{0\le t\le T} |W_N^i(t)|^{p/2} + \sup_{0\le t\le T} |W_N^j(t)|^{p/2}) \tag{4.26}$$

where $C_{18} > 0$ is an a.s. finite random variable which does not depend on N or on the choice of $t_1, ..., t_m$. Using (2.16) we have

$$\sup_{0\le t\le T} |S_N^j(t)|^{p/2} \le 2^{p/2-1}(C_{19}N^{-p\delta/2} + \sup_{0\le t\le T} |W_N^j(t)|^{p/2}) \text{ whenever } 1 \le j \le e \tag{4.27}$$

where $C_{19} > 0$ is an a.s. finite random variable which does not depend on N. By the standard martingale moment estimates for the Brownian motion for any $M \ge 1$,

$$E \sup_{0\le t\le T} |W_N^j(t)|^{pM} \le C_{20}(M,T) < \infty \text{ whenever } 1 \le j \le e,$$

where $C_{20}(M,T) > 0$ does not depend on N. Applying as above the Chebyshev inequality and then the Borel–Cantelli lemma we see that for any $\gamma > 0$,

$$\sup_{0\le t\le T} |W_N^j(t)| = O(N^\gamma) \quad \text{a.s.}$$

This together with (4.26) and (4.27) gives

$$J_N^{(1)} = O(N^{1-\alpha-p\delta/2-p\gamma/2}) \tag{4.28}$$

where we choose α so close to 1 that $1 - \alpha - p\delta/2 < 0$.

By (4.19) we obtain similarly to (3.31) that

$$J_N^{(2)} \le C_{\alpha,\beta}^{\$} N^{-p(1-\beta)/2} \textstyle\sum_{0\le j<n,\, k_{j+1}-k_j\le N^\alpha} |k_{j+1}-k_j|^{p(1-\beta)/2} \le C_{\alpha,\beta}^{\$} T N^{-(1-\alpha)(\frac{p}{2}-1-p\beta/2)} \tag{4.29}$$

where we assume that

$$0 < \beta < 1 - \frac{2}{p} \tag{4.30}$$

which is consistent since $p > 2$.

In order to estimate $J_N^{(3)}$ we proceed in the same way as in (3.33)–(3.35) and (4.21) writing

$$\begin{aligned} &P\big\{ \max_{m+N^\alpha \geq n > m \geq 0,\, [TN]>n} \frac{|\sum_{m\leq l<n}(W_N^j(\frac{l+1}{N}) - W_N^j(\frac{l}{N}))(W_N^i(\frac{l}{N}) - W_N^i(\frac{m}{N}))|}{|(n-m)/N|^{1-\beta}} \\ &> 1\big\} \leq N^{2M(1-\beta)} \sum_{m+N^\alpha \geq n > m \geq 0,\, [TN]>n} (n-m)^{-2M(1-\beta)} \\ &\times E\big| \sum_{m\leq l<n}(W_N^j(\tfrac{l+1}{N}) - W_N^j(\tfrac{l}{N}))(W_N^i(\tfrac{l}{N}) - W_N^i(\tfrac{m}{N}))\big|^{2M} \\ &\leq C_{21}(M) N^{-2M\beta} \sum_{m+N^\alpha \geq n > m \geq 0,\, [TN]>n} (n-m)^{2M\beta} \\ &\leq C_{21}(M) T N^{-2M\beta(1-\alpha)+1+\alpha} \end{aligned} \tag{4.31}$$

where $C_{21}(M) > 0$ does not depend on $N \geq 1$ and we rely on the Chebyshev inequality and the standard moment estimates for the martingale $M_n = \sum_{m\leq l<n}(W_N^j(\frac{l+1}{N}) - W_N^j(\frac{l}{N}))(W_N^i(\frac{l}{N}) - W_N^i(\frac{m}{N}))$ (see, for instance, [40]) or, alternatively, use Lemma 3.4. Choosing $M \geq (3+\alpha)(2\beta(1-\alpha))^{-1}$ we obtain in the right hand side of (4.31) a term of a converging sequence and application of the Borel–Cantelli lemma provides us with a finite a.s. random variable $C_{\alpha,\beta}^W > 0$ such that for all $N \geq 1$,

$$\big| \int_{m/N}^{n/N} (W_N^i(u) - W_N^i(m/N)) dW_N^j(u) \big| \leq C_{\alpha,\beta}^W \big| \frac{n}{N} - \frac{m}{N} \big|^{1-\beta} \tag{4.32}$$

whenever $N^\alpha \geq n - m > 0$, $n < [TN]$, $m \geq 0$. In the same way as in (4.29) we have now

$$J_N^{(3)} \leq C_{\alpha,\beta}^W T N^{-(1-\alpha)(p/2-1-p\beta/2)}. \tag{4.33}$$

Collecting (4.18), (4.24), (4.28), (4.29) and (4.30) we obtain (2.17) for some $\delta > 0$ which together with Lemma 3.10 completes the proof of Theorem 2.2. □

5 Continuous time case: proof of Theorem 2.5

5.1 Basic estimates

First observe that by (2.6), (2.19) and (2.21),

$$\begin{aligned} &|(\eta \circ \vartheta^m(\omega) - E(\eta \circ \vartheta^m | \mathcal{F}_{m-n,m+n})(\omega)| \\ &= \big| \int_0^{\tau\circ\vartheta^m(\omega)} \xi(s, \vartheta^m\omega) ds - E(\int_0^{\tau\circ\vartheta^m(\omega)} \xi(s, \vartheta^m\omega) ds | \mathcal{F}_{m-n,m+n})(\omega) \big| \\ &\leq 2L|\tau \circ \vartheta^m(\omega) - E(\tau \circ \vartheta^m | \mathcal{F}_{m-n,m+n})(\omega)| + \big| \int_0^{E(\tau\circ\vartheta^m(\omega)|\mathcal{F}_{m-n,m+n})} (\xi(s, \vartheta^m\omega) \\ &- E(\xi(s, \vartheta^m\omega) | \mathcal{F}_{m-n,m+n})(\omega)) ds \big| \leq (2L + \hat{L})\rho(n), \end{aligned} \tag{5.1}$$

and so the sequence $\eta(k)$, $k \in \mathbb{Z}$ satisfies the conditions of Theorem 2.1.

With a slight abuse of notations we set now for each $t \in [0, T]$,

$$S_\varepsilon(t) = \varepsilon \sum_{0\leq k<[\varepsilon^{-2}t]} \eta(k) \text{ and } \mathbb{S}_\varepsilon^{ij}(t) = \varepsilon^2 \sum_{0\leq k<l<[\varepsilon^{-2}t]} \eta_i(k)\eta_j(l).$$

In view of the assumptions of Theorem 2.4 we can apply Theorem 2.2 to $S_{1/\sqrt{N}}$ and $\mathbb{S}_{1/\sqrt{N}}$ to obtain

$$\|S_{1/\sqrt{N}} - W_N\|_{p,[0,T]} = O(N^{-\delta}) \quad \text{a.s.} \tag{5.2}$$

and

$$\max_{0\le i,j\le d}\|\mathbb{S}^{ij}_{1/\sqrt{N}}-\mathbb{W}^{ij}_N\|_{p/2,[0,T]}=O(N^{-\delta})\quad\text{a.s.}\tag{5.3}$$

where $W_N=N^{-1/2}\mathcal{W}(Nt)$ and $\mathcal{W}$ is the universal Brownian motion constructed via Theorem 3.8 for the sequence $\eta(k),\, k\in\mathbb{Z}$ in place of the sequence $\xi(k),\, k\in\mathbb{Z}$ considered there. Here

$$\mathbb{W}^{ij}_N(t)=\int_0^t W^i_N(s)dW^j_N(s)+t\sum_{l=1}^\infty E(\eta_i(0)\eta_j(l)).$$

Next, set $N_\varepsilon=[\varepsilon^{-2}]$. Then, by (2.6) and (2.19),

$$|S_\varepsilon(t)-S_{1/N_\varepsilon}(t)|\le|\varepsilon\sqrt{N_\varepsilon}-1|N_\varepsilon^{-1}|\textstyle\sum_{0\le k<[N_\varepsilon t]}\eta(k)|\tag{5.4}$$
$$+\varepsilon|\textstyle\sum_{[N_\varepsilon t]\le k<[\varepsilon^{-2}t]}\eta(k)|\le\varepsilon^2|S_{1/\sqrt{N_\varepsilon}}(t)|+\varepsilon(T+1)L\hat{L}.$$

This together with (5.2) and the arguments similar to Section 3.4 yields easily that

$$\|S_\varepsilon-W_{[\varepsilon^{-2}]}\|_{p,[0,T]}=O(\varepsilon^\delta)\quad\text{a.s.}\tag{5.5}$$

for some $\delta>0$ where a.s. is simultaneously over $\varepsilon\in(0,1)$.

Next,

$$|\mathbb{S}^{ij}_\varepsilon(t)-\mathbb{S}^{ij}_{1/N_\varepsilon}(t)|\le|\varepsilon^2\sqrt{N_\varepsilon}-1|N_\varepsilon^{-1}|\textstyle\sum_{0\le k<l<[N_\varepsilon t]}\eta_i(k)\eta_j(l)|\tag{5.6}$$
$$+\varepsilon^2|\textstyle\sum_{[N_\varepsilon t]\le l<[\varepsilon^{-2}t]}\eta_j(k)\sum_{k=0}^{l-1}\eta_i(k)|$$
$$\le\varepsilon^2|\mathbb{S}^{ij}_{1/\sqrt{N_\varepsilon}}(t)|+\varepsilon L\hat{L}|\textstyle\sum_{[N_\varepsilon t]\le l<[\varepsilon^{-2}t]}S^i_{1/\sqrt{N_\varepsilon}}((l-1)\varepsilon^2)|.$$

Relying on (5.2), (5.3) and the arguments similar to Section 4.2 we derive easily that

$$\|\mathbb{S}^{ij}_\varepsilon-\mathbb{W}^{ij}_{[\varepsilon^{-2}]}\|_{p/2,[0,T]}=O(\varepsilon^\delta)\quad\text{a.s.}\tag{5.7}$$

for some $\delta>0$ where a.s. is simultaneously over $\varepsilon\in(0,1)$. In fact, employing the standard moment estimates for the Brownian motion together with its Hölder continuity and the Borel–Cantelli lemma we can replace $W_{[\varepsilon^{-2}]}(t)=[\varepsilon^{-2}]^{-1/2}\mathcal{W}([\varepsilon^{-2}]t)$ in (5.5) by $W^\varepsilon(t)=\varepsilon\mathcal{W}(\varepsilon^{-2}t)$ and $\mathbb{W}^{ij}_{[\varepsilon^{-2}]}$ in (5.7) by

$$\tilde{\mathbb{W}}^\varepsilon_{ij}(t)=\int_0^t W^\varepsilon_i(s)dW^\varepsilon_j(s)+t\sum_{l=1}^\infty E(\eta_i(0)\eta_j(l)).$$

Hence, in addition to (5.5) and (5.7) we have also

$$\|S_\varepsilon-W^\varepsilon\|_{p,[0,T]}=O(\varepsilon^\delta)\quad\text{a.s.}\tag{5.8}$$

and

$$\|\mathbb{S}^{ij}_\varepsilon-\tilde{\mathbb{W}}^\varepsilon_{ij}\|_{p/2,[0,T]}=O(\varepsilon^\delta)\quad\text{a.s.}\tag{5.9}$$

5.2 A renewal type lemma

The following result will be used several times in this section and though it is essentially known we will provide its self-contained proof for completeness.

Lemma 5.1. *Let $n(s)=n(s,\omega)=0$ if $\tau(\omega)>s$ and*

$$n(s)=n(s,\omega)=\max\{k:\sum_{j=0}^{k-1}\tau\circ\vartheta^j(\omega)\le s\}.$$

Then for any $M \geq 1$ *and* $s, t \geq 0$*,*

$$E|n(s\bar{\tau}) - s|^{2M} \leq K(M)s^M \text{ and } E \sup_{0\leq s\leq t} |n(s\bar{\tau}) - s|^{2M} \leq K(M)t^{M+1}, \tag{5.10}$$

where $K(M) > 0$ *does not depend on* s *and* t*, and for any* $\gamma > 0$*,*

$$|n(s\bar{\tau}) - s| = O(s^{\frac{1}{2}+\gamma}) \quad a.s. \tag{5.11}$$

Proof. Observe first that without loss of generality it suffices to prove (5.10) and (5.11) for $s = k,\ k = 1, 2, ...$ since $n(s\bar{\tau}) - n([s]\bar{\tau}) \leq \hat{L}^2$ in view of (2.19). Next,

$$m^{-M} E|n(m\bar{\tau}) - m|^{2M} \leq 1 + \sum_{k=1}^{\infty} P\{|n(m\bar{\tau}) - m| > k^{1/2M}\sqrt{m}\}. \tag{5.12}$$

Now we have the following events inclusions

$$\{|n(m\bar{\tau}) - m| > k^{1/2M}\sqrt{m}\} \subset \{\textstyle\sum_{0\leq j\leq m+k^{1/2M}\sqrt{m}} \tau\circ\vartheta^j < m\bar{\tau}\}$$
$$\cup\{\textstyle\sum_{0\leq j\leq m-k^{1/2M}\sqrt{m}} \tau\circ\vartheta^j > m\bar{\tau}\} \subset \Gamma_k \cup \Delta_k$$

where

$$\Gamma_k = \{\sum_{0\leq j\leq m+k^{1/2M}\sqrt{m}} (\tau\circ\vartheta^j - \bar{\tau}) < -[k^{1/2M}\sqrt{m}]\bar{\tau}\}$$

and

$$\Delta_k = \{\sum_{0\leq j\leq m-k^{1/2M}\sqrt{m}} (\tau\circ\vartheta^j - \bar{\tau}) > [k^{1/2M}\sqrt{m}]\bar{\tau}\}.$$

Next, by Lemma 3.4,

$$E(\sum_{0\leq j\leq n-1} (\tau\circ\vartheta^j - \bar{\tau}))^{6M} \leq C_1^{\tau}(3M)n^{3M}$$

where the index τ in C_1^{τ} means that we apply Lemma 3.4 for sums of τ's in place of ξ's there. Hence, by the Chebyshev inequality for $k, m \geq 1$,

$$\max(P(\Gamma_k),\ P(\Delta_k)) \leq C_1^{\tau}(3M)(m + k^{1/2M}\sqrt{m})^{3M}[k^{1/2M}\sqrt{m}]^{-6M}\bar{\tau}^{-6M}$$
$$\leq \tilde{K}(M)k^{-3/2}$$

for some $\tilde{K}(M) > 0$ which does not depend on k and m. Summing in $k \geq 1$ we obtain the first estimate in (5.10) from (5.12). The second estimate in (5.10) follows by

$$E \max_{0\leq m\leq [t]+1} |n(m\bar{\tau}) - m|^{2M} \leq \sum_{0\leq m\leq [t]+1} E|n(m\bar{\tau}) - m|^{2M} \leq K(M) \sum_{0\leq m\leq [t]+1} m^M.$$

Finally, by (5.10) and the Chebyshev inequality,

$$P\{|n(m\bar{\tau}) - m| > m^{\frac{1}{2}+\gamma}\} \leq K(M)m^{-M(1+2\gamma)}m^M = K(M)m^{-2M\gamma}.$$

Choosing $M \geq \gamma^{-1}$ we obtain in the right hand side here a term of a converging sequence and by the Borel–Cantelli lemma (5.11) follows. □

5.3 Proof of (2.26) and (2.27) in the supremum norm

Set

$$U^\varepsilon(t)=\varepsilon\sum_{0\le k<n(t\bar\tau\varepsilon^{-2})}\eta(k)\text{ and }\mathbb{U}^\varepsilon_{ij}(t)=\varepsilon^2\sum_{0\le k<l<n(t\bar\tau\varepsilon^{-2})}\eta_i(k)\eta_j(l).$$

Then for $M\ge 1$,

$$\begin{gathered}\sup_{0\le t\le T}|S_\varepsilon(t)-U^\varepsilon(t)|^{2M}\\ =\varepsilon^{2M}\sup_{0\le t\le T}|\sum_{\min([\varepsilon^{-2}t],n(t\bar\tau\varepsilon^{-2}))\le k<\max([\varepsilon^{-2}t],n(t\bar\tau\varepsilon^{-2}))}\eta(k)|^{2M}\\ \le\varepsilon^{2M}\sum_{0\le m\le\varepsilon^{-2}T}\sup_{0\le t\le T}\max_{0\le k\le|n(t\bar\tau\varepsilon^{-2})-\varepsilon^{-2}t|}|\sum_{l=0}^k\eta(l+m)|^{2M}\\ \le\varepsilon^{2M}\sum_{0\le m\le\varepsilon^{-2}T}\max_{0\le k\le\varepsilon^{-3/2}}|\sum_{l=0}^k\eta(l+m)|^{2M}\\ +\varepsilon^{2M}\sum_{0\le m\le\varepsilon^{-2}T}\max_{0\le k\le T\hat L^2\varepsilon^{-2}}|\sum_{l=0}^k\eta(l+m)|^{2M}\mathbb{I}_{\sup_{0\le t\le T}|n(t\bar\tau\varepsilon^{-2})-\varepsilon^{-2}t|>\varepsilon^{-3/2}}\end{gathered}$$

since $|n(t\bar\tau\varepsilon^{-2})-\varepsilon^{-2}t|\le T\hat L^2\varepsilon^{-2}$. Applying Lemmas 3.3, 3.4 or Theorem B from [44] to sums $\sum_{l=0}^k\eta(l)$ and taking into account stationarity of the sequence $\eta(l)$, $l\ge 0$ we obtain that

$$E\max_{0\le k\le\varepsilon^{-3/2}}|\sum_{l=0}^k\eta(l+m)|^{2M}\le C_{22}(M)\varepsilon^{-3M/2}$$

for some $C_{22}(M)>0$ which does not depend on $\varepsilon>0$. Using, in addition, Lemma 5.1 together with the Cauchy-Schwarz and Chebyshev inequalities we derive that

$$\begin{gathered}E(\max_{0\le k\le T\hat L^2\varepsilon^{-2}}|\sum_{l=0}^k\eta(l+m)|^{2M}\mathbb{I}_{\sup_{0\le t\le T}|n(t\bar\tau\varepsilon^{-2})-\varepsilon^{-2}t|>\varepsilon^{-3/2}})\\ \le\big(E\max_{0\le k\le T\hat L\varepsilon^{-2}}|\sum_{l=0}^k\eta(l+m)|^{2(M+1)}\big)^{\frac{M}{M+1}}\\ \times\big(P\{\sup_{0\le t\le T}|n(t\bar\tau\varepsilon^{-2})-\varepsilon^{-2}t|>\varepsilon^{-3/2}\}\big)^{\frac{1}{M+1}}\\ \le\tilde C(M)(T\hat L^2\varepsilon^{-2}+1)^M\varepsilon^{3(M+1)}(E\sup_{0\le t\le T}|n(t\bar\tau\varepsilon^{-2})-\varepsilon^{-2}t|^{2(M+1)^2})^{\frac{1}{M+1}}\\ \le C_{23}(M)\varepsilon^{-M}\end{gathered}$$

for some $\tilde C(M)$, $C_{23}(M)>0$ which do not depend on $\varepsilon\in(0,1)$.

Hence,

$$\sup_{0\le t\le T}|S_\varepsilon(t)-U^\varepsilon(t)|^{2M}\le C_{22}(M)T\varepsilon^{\frac{M}{2}-2}+C_{23}(M)\varepsilon^{M-2},$$

and so by the Chebyshev inequality,

$$P\{\sup_{0\le t\le T}|S_\varepsilon(t)-U^\varepsilon(t)|\ge\varepsilon^{1/8}\}\le(C_{22}(M)+C_{23}(M))T\varepsilon^{\frac{M}{4}-2}.$$

Taking $M\ge 24$ and $\varepsilon=\varepsilon_n=n^{-1/2}$ we obtain by the Borel–Cantelli lemma that

$$\sup_{0\le t\le T}|S_{\varepsilon_n}(t)-U^{\varepsilon_n}(t)|=O(n^{-1/16})\text{ a.s.}$$

When $(n+1)^{-1/2}\le\varepsilon\le n^{-1/2}$ then

$$|S_\varepsilon(t)-S_{\varepsilon_n}(t)|\le(2T+1)Ln^{-1/2}\text{ and }|U^\varepsilon(t)-U^{\varepsilon_n}(t)|\le(2T+1)L\hat Ln^{-1/2},$$

and so

$$\sup_{0\le t\le T}|S_\varepsilon(t)-U^\varepsilon(t)|=O(\varepsilon^{1/8})\text{ a.s.}\tag{5.13}$$

Next we estimate

$$\sup_{0\le t\le T}|\mathbb{S}^{ij}_\varepsilon(t)-\mathbb{U}^\varepsilon_{ij}(t)|^{2M}\le 2^{2M-1}\varepsilon^{4M}(I^{2M}_{\varepsilon,1}+I^{2M}_{\varepsilon,2}I^{2M}_{\varepsilon,3})$$

where

$$I_{\varepsilon,1} = \sup_{0\le t\le T} |\sum_{\min([\varepsilon^{-2}t],n(t\bar{\tau}\varepsilon^{-2}))\le k<l<\max([\varepsilon^{-2}t],n(t\bar{\tau}\varepsilon^{-2}))} \eta_i(k)\eta_j(l)|,$$
$$I_{\varepsilon,2} = |\sum_{0\le k<\min([\varepsilon^{-2}T],n(T\bar{\tau}\varepsilon^{-2}))} \eta_i(k)| \text{ and}$$
$$I_{\varepsilon,3} = \sup_{0\le t\le T} |\sum_{\min([\varepsilon^{-2}t],n(t\bar{\tau}\varepsilon^{-2}))\le l<\max([\varepsilon^{-2}t],n(t\bar{\tau}\varepsilon^{-2}))} \eta_j(l)|.$$

Similarly to the above

$$I^{2M}_{\varepsilon,1} \le \sum_{0\le m\le\varepsilon^{-2}T}\ \sum_{0\le n\le\varepsilon^{-3/2}} J^{2M}(m,n) + \sum_{0\le m\le\varepsilon^{-2}T}\ \sum_{0\le n\le L\hat{L}T\varepsilon^{-2}} J^{2M}_\varepsilon(m,n)$$

where

$$J(m,n) = \sum_{0\le k<l<n} \eta_i(k+m)\eta_j(l+m) \text{ and}$$
$$J_\varepsilon(m,n) = J(m,n)\mathbb{I}_{\sup_{0\le t\le T}|n(t\bar{\tau}\varepsilon^{-2})-\varepsilon^{-2}t|>\varepsilon^{-3/2}}.$$

By the stationarity of the sequence $\eta(r)$, $r = 0, 1, ...,$

$$EJ^{2M}(m,n) = EJ^{2M}(0,n) \le 2^{2M-1}(EJ^{2M}_1(n) + EJ^{2M}_2(n))$$

where

$$J_1(n) = J(0,n) - J_2(n) \text{ and } J_2(n) = \sum_{0\le k<l<n} E(\eta_i(k)\eta_j(l)).$$

By Lemma 3.4,

$$EJ^{2M}_1(n) \le C_{24}(M)n^{2M}$$

for some $C_{24}(M) > 0$ which does not depend on n.

In view of the assumptions of Theorem 2.4 on the coefficients ϕ and ρ the estimates (4.6) and (4.7) considered for the sequence $\eta(r)$, $r \ge 0$ (in place of $\xi(r)$, $r \ge 0$ there) hold true, as well, which implies that

$$|J_2(n)| \le C_{25}n$$

for some $C_{25} > 0$ which does not depend on n. Now, by the Cauchy–Schwarz and Chebyshev inequalities, similarly to the above,

$$EJ^{2M}_\varepsilon(m,n) = E(J_1(n) + J_2(n))^{2M}\mathbb{I}_{\sup_{0\le t\le T}|n(t\bar{\tau}\varepsilon^{-2})-\varepsilon^{-2}t|>\varepsilon^{-3/2}}$$
$$\le \left(E(J_1(n) + J_2(n))^{2(M+1)}\right)^{\frac{M}{M+1}} \left(P\{\sup_{0\le t\le T}|n(t\bar{\tau}\varepsilon^{-2}) - \varepsilon^{-2}t| > \varepsilon^{-3/2}\}\right)^{\frac{1}{M+1}}$$
$$\le C_{26}(M)n^{2M}\varepsilon^M$$

for some $C_{26}(M) > 0$ which does not depend on n and ε. Collecting the above inequalities we obtain that,

$$\varepsilon^{4M}EI^{2M}_{\varepsilon,1} \le C_{27}(M)\varepsilon^{M-4}$$

for some $C_{27}(M) > 0$ which does not depend on ε.

Next, relying on Lemma 3.4 considered for the sequence $\eta(r)$, $r \ge 0$ and taking into account that $n(T\bar{\tau}\varepsilon^{-2}) \le T\hat{L}^2\varepsilon^{-2}$ we obtain that

$$EI^{2M}_{\varepsilon,2} \le C_{28}\varepsilon^{-2M}$$

for some $C_{28} > 0$ which does not depend on ε. Since,

$$I_{\varepsilon,3} = \varepsilon^{-1}\sup_{0\le t\le T}|S_\varepsilon(t) - U^\varepsilon(t)|$$

we can use the estimates at the beginning of this subsection to obtain that

$$\varepsilon^{4M}E(I^{2M}_{\varepsilon,2}I^{2M}_{\varepsilon,3}) \le C_{29}\varepsilon^{\frac{M}{2}-2}$$

for some $C_{29} > 0$ which does not depend on $\varepsilon > 0$. Proceeding similarly to the above with the Chebyshev inequality and the Borel–Cantelli lemma we arrive finally at

$$\max_{0\le i,j\le d}\sup_{0\le t\le T}|\mathbb{S}_\varepsilon(t)-\mathbb{U}^\varepsilon_{ij}(t)|=O(\varepsilon^{1/8})\quad \text{a.s.} \tag{5.14}$$

Next, we compare V^ε with U^ε and $\mathbb{V}^\varepsilon$ with $\mathbb{U}^\varepsilon$ with V^ε and $\mathbb{V}^\varepsilon$ defined before Theorem 2.5. Clearly, by (2.6) and (2.19),

$$\sup_{0\le t\le T}|V^\varepsilon(t)-U^\varepsilon(t)|\le \varepsilon L\hat{L}. \tag{5.15}$$

Next,

$$\sup_{0\le t\le T}|\mathbb{V}^\varepsilon_{ij}(t)-tEF_{ij}-\mathbb{U}^\varepsilon_{ij}(t)|\le \sup_{0\le t\le T}|J_1(t)|+\sup_{0\le t\le T}|J_2(t)|+\sup_{0\le t\le T}|J_3(t)| \tag{5.16}$$

where $F_{ij}(\omega)=\int_0^{\tau(\omega)}\xi_j(s,\omega)ds\int_0^s\xi_i(u,\omega)du$,

$$J_1(t)=\varepsilon^2\int_{\sigma(t\bar{\tau}\varepsilon^{-2})}^{t\bar{\tau}\varepsilon^{-2}}\xi_j(s)\int_0^s\xi_i(u)duds,\ \sigma(s)=\sigma(s,\omega)=\sum_{j=0}^{n(s)-1}\tau\circ\vartheta^j(\omega),$$

$$J_2(t)=\varepsilon^2(\int_0^{\sigma(t\bar{\tau}\varepsilon^{-2})}\xi_j(s)\int_{\sigma(s)}^s\xi_i(u)duds-n(t\bar{\tau}\varepsilon^{-2})EF_{ij})$$

$$\text{and}\quad J_3(t)=(tEF_{ij}-\varepsilon^2 n(t\bar{\tau}\varepsilon^{-2})EF_{ij}).$$

Now, by (2.6) and (2.19),

$$|J_1(t)|\le L\varepsilon^2\textstyle\int_{\sigma(t\bar{\tau}\varepsilon^{-2})}^{t\bar{\tau}\varepsilon^{-2}}ds|\int_0^s\xi_i(u)du| \tag{5.17}$$

$$\le L^2\hat{L}\varepsilon^2+L\hat{L}\varepsilon^2|\textstyle\int_0^{t\bar{\tau}\varepsilon^{-2}}\xi_i(u)du|\le L^2\hat{L}\varepsilon^2(1+\hat{L})+L\hat{L}\varepsilon^2|\sum_{k=0}^{n(t\bar{\tau}\varepsilon^{-2})}\eta_i(k)|.$$

By (2.6), (2.19) and (5.11) for any $\gamma > 0$,

$$\varepsilon^2\Big|\sum_{k=0}^{n(t\bar{\tau}\varepsilon^{-2})}\eta_i(k)-\sum_{k=0}^{[t\varepsilon^{-2}]-1}\eta_i(k)\Big|=O(\varepsilon^{1-\gamma})\ \text{a.s.}$$

It follows that

$$\sup_{0\le t\le T}|J_1(t)|\le C_{30}\varepsilon^{1-\gamma}+L\hat{L}\varepsilon^2\max_{1\le k\le[T\varepsilon^{-2}]}|\sum_{l=0}^{k-1}\eta_i(l)| \tag{5.18}$$

for some a.s. finite random variable $C_{30}=C_{30}(\omega)>0$ which does not depend on ε.

Next, applying Lemma 3.4 to the sequence $\eta(k)$, $k\ge 0$ which is possible in view of (5.1), we obtain

$$E\max_{1\le k\le[T\varepsilon^{-2}]}|\sum_{l=0}^{k-1}\eta_i(l)|^{2M}\le\sum_{1\le k\le[T\varepsilon^{-2}]}E|\sum_{l=0}^{k-1}\eta_i(l)|^{2M}\le C_1^\eta(M)T^M\varepsilon^{-2M}$$

where $C_1^\eta>0$ does not depend on ε. Hence, by the Chebyshev inequality

$$P\{\varepsilon^2\max_{1\le k\le[T\varepsilon^{-2}]}|\sum_{l=0}^{k-1}\eta_i(l)|>\varepsilon^{1-\gamma}\}\le C_1^\eta(M)T^M\varepsilon^{M\gamma}. \tag{5.19}$$

Taking $\varepsilon=\varepsilon_n=\frac{1}{\sqrt{n}}$, $\gamma\in(0,1)$, $M\ge 2/\gamma$ and applying the Borel-Cantelli lemma we obtain that

$$n^{-1}\max_{1\le k\le[Tn]}|\sum_{l=0}^{k-1}\eta_i(l)|=O(n^{-\frac{1}{2}(1-\gamma)})\ \text{a.s.}$$

If $\frac{1}{\sqrt{n}} \leq \varepsilon \leq \frac{1}{\sqrt{n-1}}$ then

$$\varepsilon^2 \max_{1\leq k\leq [T\varepsilon^{-2}]} |\sum_{l=0}^{k-1} \eta_i(l)|$$
$$\leq \frac{1}{n-1} \max_{1\leq k\leq [Tn]} |\sum_{l=0}^{k-1} \eta_i(l)| = O(n^{-\frac{1}{2}(1-\gamma)}) = O(\varepsilon^{1-\gamma}) \text{ a.s.},$$

and so

$$\sup_{0\leq t\leq T} |J_1(t)| = O(\varepsilon^{1-\gamma}) \quad \text{a.s.} \tag{5.20}$$

Next,

$$|J_2(t)| = \varepsilon^2 |\sum_{k=0}^{n(t\bar{\tau}\varepsilon^{-2})} (F_{ij} \circ \vartheta^k - EF_{ij})| \leq |J_4(t)| + 2(L\hat{L})^2 \varepsilon^2 |n(t\bar{\tau}\varepsilon^{-2}) - [t\varepsilon^{-2}]| \tag{5.21}$$

where

$$J_4(t) = \varepsilon^2 \sum_{k=0}^{[t\varepsilon^{-2}]} (F_{ij} \circ \vartheta^k - EF_{ij}).$$

Now observe that by (2.6), (2.19) and (2.21),

$$|F_{ij} \circ \vartheta^k - E(F_{ij} \circ \vartheta^k | \mathcal{F}_{k-n,k+n})| \tag{5.22}$$
$$= | \int_0^{\tau\circ\vartheta^k(\omega)} \xi_j(s, \vartheta\omega) ds \int_0^s \xi_i(u, \vartheta^k\omega) du$$
$$- E(\int_0^{\tau\circ\vartheta^k(\omega)} \xi_j(s, \vartheta\omega) ds \int_0^s \xi_i(u, \vartheta^k\omega) du | \mathcal{F}_{k-n,k+n})|$$
$$\leq 2L\hat{L} |\tau \circ \vartheta^k(\omega) - E(\tau \circ \vartheta^k | \mathcal{F}_{k-n,k+n})(\omega)| + G_{ij}(E(\tau \circ \vartheta^k | \mathcal{F}_{k-n,k+n})(\omega), \omega)$$

where

$$G_{ij}(r, \omega) = | \int_0^r \int_0^s \big(\xi_j(s, \vartheta^k\omega)\xi_i(u, \vartheta^k\omega) - E(\xi_j(s, \vartheta^k\omega)\xi_i(u, \vartheta^k\omega) | \mathcal{F}_{k-n,k+n})\big) ds du|$$
$$\leq 2Lr^2 \rho(n).$$

Hence, the left hand side of (5.22) does not exceed $2L\hat{L}\rho(n)(1 + \hat{L})$, and so we can apply Lemma 3.4 with $F_{ij} \circ \vartheta^k$'s in place of $\xi(k)$'s to obtain that

$$E \sup_{0\leq t\leq T} J_4^{2M}(t) \leq \varepsilon^{4M} E \max_{1\leq n<[T\varepsilon^{-2}]} |\sum_{k=0}^{n-1} (F_{ij} \circ \vartheta^k - EF_{ij})|^{2M} \tag{5.23}$$
$$\leq \varepsilon^{2M} C_1^F(M) T^M$$

where $C_1^F > 0$ does not depend on ε. Arguing as for (5.20) we see that for any $\gamma > 0$,

$$\sup_{0\leq t\leq T} |J_4(t)| = O(\varepsilon^{1-\gamma}) \quad \text{a.s.}$$

which together with (5.11) and (5.21) gives that for any $\gamma > 0$,

$$\sup_{0\leq t\leq T} |J_2(t)| = O(\varepsilon^{1-\gamma}) \quad \text{a.s.} \tag{5.24}$$

Estimating J_3 by (5.11) we obtain that for any $\gamma > 0$,

$$\sup_{0\leq t\leq T} |J_3(t)| \leq |EF_{ij}| \sup_{0\leq t\leq T} |t - \varepsilon^2 n(t\bar{\tau}\varepsilon^{-2})| = O(\varepsilon^{1-\gamma}) \quad \text{a.s.} \tag{5.25}$$

Finally, collecting (5.8), (5.9), (5.13), (5.14), (5.15), (5.18), (5.24) and (5.25) we obtain (2.26) and (2.27) but only for the supremum norm.

5.4 Completing the proof of Theorem 2.5

For $0 \le s < t$ set

$$V^\varepsilon(s,t) = V^\varepsilon(t) - V^\varepsilon(s) = \varepsilon \int_{s\bar{\tau}\varepsilon^{-2}}^{t\bar{\tau}\varepsilon^{-2}} \xi(u)du,$$

$$U^\varepsilon(s,t) = U^\varepsilon(t) - U^\varepsilon(s) = \varepsilon \sum\nolimits_{n(s\bar{\tau}\varepsilon^{-2}) \le k < n(t\bar{\tau}\varepsilon^{-2})} \eta(k)$$
$$\text{and } \hat{U}(s,t) = \varepsilon \sum\nolimits_{[s\varepsilon^{-2}] \le k < [t\varepsilon^{-2}]} \eta(k).$$

First, observe that by (2.6) for any $\beta \in (0, 1/2)$,

$$|V^\varepsilon(s,t)| \le \varepsilon^{-1} L\bar{\tau}(t-s) \le L\bar{\tau}(t-s)^{\frac{1}{2}-\beta}\varepsilon^{-1}(t-s)^{\frac{1}{2}+\beta} = O((t-s)^{\frac{1}{2}-\beta}) \tag{5.26}$$

provided

$$(t-s) = O(\varepsilon^{\frac{2}{1+2\beta}}). \tag{5.27}$$

Next, we are going to obtain Hölder type uniform estimates of $|V^\varepsilon(s,t)|(t-s)^{(\frac{1}{2}-\beta)}$ similar to (5.26) for small $t-s$ satisfying

$$s + \varepsilon^{1-\alpha} \ge t \ge s + \varepsilon^2 > s \ge 0 \tag{5.28}$$

where $\alpha \in (0,1)$ is close to 1 and it is chosen similarly to Section 3.4. Observe that by (2.6) and (2.19),

$$|V^\varepsilon(s,t) - V^\varepsilon([s\varepsilon^{-2}]\varepsilon^2,\, [t\varepsilon^{-2}]\varepsilon^2)| \le 2\varepsilon L\hat{L},$$

and so under (5.28),

$$\frac{|V^\varepsilon(s,t)|}{(t-s)^{\frac{1}{2}-\beta}} \le \sqrt{2}\frac{|V^\varepsilon([s\varepsilon^{-2}]\varepsilon^2,\, [t\varepsilon^{-2}]\varepsilon^2)|}{([t\varepsilon^{-2}] - [s\varepsilon^{-2}])^{\frac{1}{2}-\beta}\varepsilon^{1-2\beta}} + 2\varepsilon^{2\beta}L\hat{L}.$$

Hence,

$$\sup\nolimits_{0\le s<s+\varepsilon^2\le t\le s+\varepsilon^{1-\alpha},\, t\le T} \frac{|V^\varepsilon(s,t)|}{(t-s)^{\frac{1}{2}-\beta}} \tag{5.29}$$
$$\le \sqrt{2}\max\nolimits_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}} \frac{|V^\varepsilon(k\varepsilon^2, l\varepsilon^2)|}{(l-k)^{\frac{1}{2}-\beta}\varepsilon^{1-2\beta}} + 2\varepsilon^{2\beta}L\hat{L}.$$

In order to estimate the moments of the right hand side of (5.29) we introduce

$$\hat{V}^\varepsilon_{kl}(\omega,u) = \varepsilon\int_{k\bar{\tau}}^{l\bar{\tau}} \xi(v,(\omega,u))dv = \varepsilon\int_{k\bar{\tau}+u}^{l\bar{\tau}+u} \xi(v,(\omega,0))dv$$

and observe that by (2.6) and (2.19),

$$|V^\varepsilon(k\varepsilon^2, l\varepsilon^2) - \hat{V}^\varepsilon_{kl}(\omega,u)| \le 2uL\varepsilon \le 2L\hat{L}\varepsilon. \tag{5.30}$$

Since ξ is a stationary process on the probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ we see that

$$\int |\hat{V}^\varepsilon_{kl}(\omega,u)|^{2M} d\hat{P}(\omega,u) = \int |\hat{V}^\varepsilon_{0l-k}(\omega,u)|^{2M} d\hat{P}(\omega,u),$$

and so by (5.30),

$$E|V^\varepsilon(k\varepsilon^2, l\varepsilon^2)|^{2M} \le 2^{4M-2}\hat{L}^2 E|V^\varepsilon(0,(l-k)\varepsilon^2)|^{2M} \tag{5.31}$$
$$+2^{4M-1}(2^{2M}+1)\hat{L}(L\hat{L})^{2M}\varepsilon^{2M}.$$

By (5.15),

$$|V^\varepsilon(0,(l-k)\varepsilon^2) - U^\varepsilon(0,(l-k)\varepsilon^2)| \le 2\varepsilon L\hat{L},$$

and so

$$E|V^\varepsilon(0,(l-k)\varepsilon^2)|^{2M} \le 2^{2M-1}E|U^\varepsilon(0,(l-k)\varepsilon^2)|^{2M} + 2^{4M-1}(L\hat{L})^{2M}\varepsilon^{2M}. \tag{5.32}$$

In order to estimate the right hand side of (5.32) we observe first that

$$|U^\varepsilon(0,(l-k)\varepsilon^2) - \hat{U}^\varepsilon(0,(l-k)\varepsilon^2)| \le \varepsilon L\hat{L}|n((l-k)\bar{\tau}) - (l-k)|,$$

and so

$$E|U^\varepsilon(0,(l-k)\varepsilon^2)|^{2M} \le 2^{2M-1}E|\hat{U}^\varepsilon(0,(l-k)\varepsilon^2)|^{2M} \tag{5.33}$$
$$+2^{2M-1}(L\hat{L})^{2M}\varepsilon^{2M}E|n((l-k)\bar{\tau}) - (l-k)|^{2M}.$$

By Lemma 3.4 applied to sums of $\eta(k)$'s we obtain that

$$E|\hat{U}^\varepsilon(0,(l-k)\varepsilon^2)|^{2M} \le C_1^\eta(M)(l-k)^M\varepsilon^{2M} \tag{5.34}$$

where $C_1^\eta > 0$ does not depend on ε, l and k. By (5.10) and (5.31)–(5.34),

$$E|V^\varepsilon(k\varepsilon^2, l\varepsilon^2)|^{2M} \le C_{31}\varepsilon^{2M}(l-k)^M \tag{5.35}$$

for some $C_{31} > 0$ which does not depend on ε, k and l. Hence, by the Chebyshev inequality

$$P\big\{\max_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}} \frac{|V^\varepsilon(k\varepsilon^2, l\varepsilon^2)|}{(l-k)^{\frac{1}{2}-\beta}\varepsilon^{1-2\beta}} > 1\big\} \tag{5.36}$$
$$\le \sum_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}} P\big\{\frac{|V^\varepsilon(k\varepsilon^2, l\varepsilon^2)|}{(l-k)^{\frac{1}{2}-\beta}\varepsilon^{1-2\beta}} > 1\big\}$$
$$\le \sum_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}} \varepsilon^{-2M(1-2\beta)}(l-k)^{-M(1-2\beta)}E|V^\varepsilon(k\varepsilon^2, l\varepsilon^2)|^{2M}$$
$$\le C_{31}\varepsilon^{4\beta M}\sum_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}}(l-k)^{2\beta M} \le C_{31}T\varepsilon^{2\beta M(1-\alpha)-4}.$$

Now, take $\varepsilon = \varepsilon_N = \frac{1}{\sqrt{N}}$ and $M \ge 4\beta^{-1}(1-\alpha)^{-1}$, then the right hand side of (5.36) is a term of a converging sequence, and so by the Borel–Cantelli lemma there exists an a.s. finite random variable $C_{32} > 0$ which does not depend on ε, k and l and such that

$$|V^\varepsilon(k\varepsilon^2, l\varepsilon^2)| \le C_{32}(l\varepsilon^2 - k\varepsilon^2)^{\frac{1}{2}-\beta}$$

whenever $\varepsilon^{-(1+\alpha)} \ge l-k \ge 1$. Since for any $\frac{1}{\sqrt{N+1}} \le \varepsilon \le \frac{1}{\sqrt{N}}$, we have $N+1 \ge \varepsilon^{-2} \ge N$, and so

$$\sup_{0\le s<t\le T}|V^\varepsilon(s,t) - V^{\varepsilon_N}(s,t)| \le 2T\hat{L}\varepsilon,$$

we conclude from here and from (5.29) that

$$|V^\varepsilon(s,t)| \le C_{\alpha,\beta}|t-s|^{\frac{1}{2}-\beta} \tag{5.37}$$

whenever (5.28) holds true, where $C_{\alpha,\beta} > 0$ is a a.s. finite random variable which does not depend on s, t and ε. Taking into account (5.26) and (5.27) we conclude that (5.37) holds true just under the condition $|t-s| \le \varepsilon^{1-\alpha}$.

Now we can complete the proof of (2.26) from Theorem 2.5. Set $N_\varepsilon = [\varepsilon^{-2}]$ and define $\hat{W}^\varepsilon(t) = W^\varepsilon(N_\varepsilon^{-1}[N_\varepsilon t])$. Let $0 = t_0 < t_1 < ... < t_m = T$. Next, we proceed similarly to (3.29)–(3.35) writing

$$\sum_{0\le i<m}|V^\varepsilon(t_i, t_{i+1}) - \hat{W}^\varepsilon(t_i, t_{i+1})|^p \le J_1^\varepsilon + 2^{p-1}(J_2^\varepsilon + J_3^\varepsilon)$$

where for $\alpha \in (0,1)$,

$$J_1^\varepsilon = \sum_{0\le i<m,\, t_{i+1}-t_i>N_\varepsilon^{-(1-\alpha)}} |V^\varepsilon(t_i,t_{i+1}) - \hat{W}^\varepsilon(t_i,t_{i+1})|^p,$$
$$J_2^\varepsilon = \sum_{0\le i<m,\, t_{i+1}-t_i\le N_\varepsilon^{-(1-\alpha)}} |V^\varepsilon(t_i,t_{i+1})|^p \quad \text{and}$$
$$J_3^\varepsilon = \sum_{0\le i<m,\, t_{i+1}-t_i\le N_\varepsilon^{-(1-\alpha)}} |\hat{W}^\varepsilon(t_i,t_{i+1})|^p.$$

Taking into account that

$$|V^\varepsilon(t_i,t_{i+1}) - V^\varepsilon(N_\varepsilon^{-1}[N_\varepsilon t_i], N_\varepsilon^{-1}[N_\varepsilon t_{i+1}])| \le 2\varepsilon(1-\varepsilon^2)^{-1}L\hat{L},$$

that there exists no more than $[TN_\varepsilon^{1-\alpha}]$ intervals $[t_i,t_{i+1}]$ with $t_{i+1}-t_i > N_\varepsilon^{-(1-\alpha)}$ and relying on the supremum norm estimate in (2.26) we obtain that

$$J_1^\varepsilon \le 2^{2p-1}T(L\hat{L})^p\varepsilon^{p-2(1-\alpha)}(1-\varepsilon^2)^{-p}$$
$$+2^{p-1}\sum_{0\le j<n,k_{j+1}-k_j>N_\varepsilon^\alpha}\sum_{0\le j<n,k_{j+1}-k_j>N_\varepsilon^\alpha}|V^\varepsilon(\tfrac{k_j}{N_\varepsilon},\tfrac{k_{j+1}}{N_\varepsilon}) - \hat{W}^\varepsilon(\tfrac{k_j}{N_\varepsilon},\tfrac{k_{j+1}}{N_\varepsilon})|^p$$
$$\le 2^{2p-1}\varepsilon^{-2(1-\alpha)}((L\hat{L})^p\varepsilon^p(1-\varepsilon^2)^{-p} + O(\varepsilon^{p\delta}))$$

where $k_j = [t_{i_j}N_\varepsilon]$ and $0 = t_{i_0} < t_{i_1} < ... < t_{i_n} = T$ is the maximal subsequence of $t_0,t_1,...,t_m$ with $[t_{i_{j+1}}N_\varepsilon] > [t_{i_j}N_\varepsilon]$.

Next, using (5.37) we obtain similarly to (3.31) that

$$J_2^\varepsilon \le C_{\alpha,\beta}^p \sum_{0\le i<m,\, t_{i+1}-t_i\le N_\varepsilon^{-(1-\alpha)}} |t_{i+1}-t_i|^{p(\frac{1}{2}-\beta)}$$
$$\le C_{\alpha,\beta}^p N_\varepsilon^{-(1-\alpha)(p(\frac{1}{2}-\beta)-1)}\sum_{0\le i<m}|t_{i+1}-t_i| \le C_{\alpha,\beta}^p T\varepsilon^{2(1-\alpha)(\frac{p}{2}-1-p\beta)}. \tag{5.38}$$

Proceeding similarly to (3.33)–(3.35) for $\hat{W}^\varepsilon$ we obtain also

$$J_3^\varepsilon \le \tilde{C}_{\alpha,\beta}^p T\varepsilon^{2(1-\alpha)(\frac{p}{2}-1-p\beta)}$$

where $C_{\alpha,\beta},\tilde{C}_{\alpha,\beta} > 0$ are a.s. finite random variables which do not depend on $\varepsilon > 0$ or the choice of $t_1,...,t_m$. Taking α so close to 1 that $p\delta > 2(1-\alpha)$ and choosing β satisfying (3.32) we obtain that

$$\|V^\varepsilon - \hat{W}^\varepsilon\|_{p,[0,T]} = O(\varepsilon^{\tilde{\delta}}) \quad \text{a.s.}$$

for some $\tilde{\delta} > 0$ which does not depend on $\varepsilon > 0$. This together with Lemma 3.10 completes the proof of (2.26) from Theorem 2.5.

It remains to complete the proof of (2.27). For $0 \le s < t$ set

$$\mathbb{V}_{ij}^\varepsilon(s,t) = \varepsilon^2\int_{s\bar{\tau}\varepsilon^{-2}}^{t\bar{\tau}\varepsilon^{-2}}\xi_j(v)dv\int_{s\bar{\tau}\varepsilon^{-2}}^{v}\xi_i(u)du,$$
$$\mathbb{U}_{ij}^\varepsilon(s,t) = \varepsilon^2\sum_{n(s\bar{\tau}\varepsilon^{-2})\le k<l<n(t\bar{\tau}\varepsilon^{-2})}\eta_i(k)\eta_j(l),$$
$$\hat{\mathbb{U}}_{ij}^\varepsilon(s,t) = \varepsilon^2\sum_{[s\varepsilon^{-2}]\le k<l<[t\varepsilon^{-2}]}\eta_i(k)\eta_j(l) \quad \text{and}$$
$$\hat{\mathbb{W}}_{ij}^\varepsilon(s,t) = \sum_{[N_\varepsilon s]\le l<[N_\varepsilon t]}(W_j^\varepsilon(\tfrac{l+1}{N_\varepsilon}) - W_j^\varepsilon(\tfrac{l}{N_\varepsilon}))(W_i^\varepsilon(\tfrac{l}{N_\varepsilon}) - W_i^\varepsilon(\tfrac{[N_\varepsilon s]}{N_\varepsilon}))$$

while $\hat{\mathbb{W}}_{ij}^\varepsilon(0,t)$ equals $\hat{\mathbb{W}}_{ij}^\varepsilon(t)$ defined before Lemma 3.10. By (2.6) for any $\beta \in (0,1)$,

$$|\mathbb{V}_{ij}^\varepsilon(s,t)| \le \varepsilon^{-2}L^2\bar{\tau}^2(t-s)^2 \le L^2\bar{\tau}^2(t-s)^{1-\beta}\varepsilon^{-2}(t-s)^{1+\beta} = O((t-s)^{1-\beta}) \tag{5.39}$$

provided

$$(t-s) = O(\varepsilon^{\frac{2}{1+\beta}}). \tag{5.40}$$

Next, we are going to obtain Hölder type estimates of $|\mathbb{V}^\varepsilon_{ij}(s,t)|(t-s)^{-(1-\beta)}$ similar to (5.39) for $t-s$ satisfying (5.28). Observe that by (2.6) and (2.19),

$$|\mathbb{V}^\varepsilon_{ij}(s,t)-\mathbb{V}^\varepsilon_{ij}([s\varepsilon^{-2}]\varepsilon^2,\,[t\varepsilon^{-2}]\varepsilon^2)|\le\varepsilon^2(L\hat{L})^2+\varepsilon^2 L\int_{[t\varepsilon^{-2}]\bar{\tau}}^{t\varepsilon^{-2}\bar{\tau}}dv|\int_{s\varepsilon^{-2}\bar{\tau}}^{v}\xi_i(u)du|$$
$$+\varepsilon^2|\int_{[s\varepsilon^{-2}]\bar{\tau}}^{[t\varepsilon^{-2}]\bar{\tau}}\xi_j(v)dv\int_{[s\varepsilon^{-2}]\bar{\tau}}^{s\varepsilon^{-2}\bar{\tau}}\xi_i(u)du|$$
$$\le 4\varepsilon^2(L\hat{L})^2+\varepsilon L\hat{L}(|V^\varepsilon_i(s,t)|+|V^\varepsilon_j(s,t)|),$$

and so under (5.28),

$$\frac{|\mathbb{V}^\varepsilon_{ij}(s,t)|}{(t-s)^{1-\beta}}\le 2\frac{|\mathbb{V}^\varepsilon_{ij}([s\varepsilon^{-2}]\varepsilon^2,\,[t\varepsilon^{-2}]\varepsilon^2)|}{([t\varepsilon^{-2}]-[s\varepsilon^{-2}])^{1-\beta}\varepsilon^{2(1-\beta)}}+4(L\hat{L})^2\varepsilon^{2\beta}+L\hat{L}\varepsilon^\beta\frac{|V^\varepsilon_i(s,t)|+|V^\varepsilon_j(s,t)|}{(t-s)^{\frac12(1-\beta)}}.$$

Hence,

$$\sup_{0\le s<s+\varepsilon^2\le t\le s+\varepsilon^{1-\alpha},\,t\le T}\frac{|\mathbb{V}^\varepsilon_{ij}(s,t)|}{(t-s)^{1-\beta}} \tag{5.41}$$
$$\le 2\max_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\,l\le T\varepsilon^{-2}}\frac{|\mathbb{V}^\varepsilon_{ij}(k\varepsilon^2,l\varepsilon^2)|}{(l-k)^{1-\beta}\varepsilon^{2(1-\beta)}}+4\varepsilon^{2\beta}L\hat{L}$$
$$+L\hat{L}\varepsilon^\beta\sup_{0\le s<s+\varepsilon^2\le t\le s+\varepsilon^{1-\alpha},\,t\le T}\frac{|V^\varepsilon_i(s,t)|+|V^\varepsilon_j(s,t)|}{(t-s)^{\frac12(1-\beta)}}.$$

In order to estimate the moments of the right hand side of (5.41) we introduce

$$\hat{\mathbb{V}}^\varepsilon_{ij}(k,l)(\omega,u)=\varepsilon^2\int_{k\bar{\tau}}^{l\bar{\tau}}\xi(v,(\omega,u))dv\int_{k\bar{\tau}}^{v}\xi(w,(\omega,u))dw$$

and observe that by (2.6) and (2.19) similarly to the above,

$$|\mathbb{V}^\varepsilon_{ij}(k\varepsilon^2,l\varepsilon^2)-\hat{\mathbb{V}}^\varepsilon_{ij}(k,l)(\omega,u)|\le 4\varepsilon^2(L\hat{L})^2+\varepsilon L\hat{L}(|V^\varepsilon_i(k\varepsilon^2,l\varepsilon^2)|+|V^\varepsilon_j(k\varepsilon^2,l\varepsilon^2)|). \tag{5.42}$$

Again, we use the stationarity of the process ξ on the probability space $(\hat{\Omega},\hat{\mathcal{F}},\hat{P})$ to obtain that

$$\int|\hat{\mathbb{V}}^\varepsilon_{ij}(k,l)(\omega,u)|^{2M}d\hat{P}(\omega,u)=\int|\hat{\mathbb{V}}^\varepsilon_{ij}(0,l-k)(\omega,u)|^{2M}d\hat{P}(\omega,u),$$

and so by (5.42),

$$E|\mathbb{V}^\varepsilon_{ij}(k\varepsilon^2,l\varepsilon^2)|^{2M}\le 2^{4M-2}\hat{L}^2E|\mathbb{V}^\varepsilon_{ij}(0,(l-k)\varepsilon^2)|^{2M} \tag{5.43}$$
$$+2^{6M-1}(2^{2M}+1)\hat{L}(L\hat{L})^{2M}\varepsilon^{4M}$$
$$+2^{4M-2}\varepsilon^{2M}(L\hat{L})^{2M}(E|V^\varepsilon_i(k\varepsilon^2,l\varepsilon^2)|^{2M}+|V^\varepsilon_j(k\varepsilon^2,l\varepsilon^2)|^{2M}).$$

Next, by (5.16), (5.17), (5.21) and (5.25),

$$|\mathbb{V}^\varepsilon_{ij}(0,(l-k)\varepsilon^2)-\varepsilon^2(l-k)EF_{ij}-\mathbb{U}^\varepsilon_{ij}(0,(l-k)\varepsilon^2)| \tag{5.44}$$
$$\le(L\hat{L})^2\varepsilon^2(l-k)+L^2\hat{L}\varepsilon^2(2+\hat{L})$$
$$+4(L\hat{L})^2\varepsilon^2|n((l-k)\bar{\tau})-(l-k)|+\varepsilon^2|\sum_{m=0}^{l-k-1}(F_{ij}\circ\vartheta^m-EF_{ij})|$$

where we took into account that $|F_{ij}|\le(L\hat{L})^2$. Also, we obtain easily by (2.6) that

$$|\mathbb{U}^\varepsilon_{ij}(0,(l-k)\varepsilon^2)-\hat{\mathbb{U}}^\varepsilon_{ij}(0,(l-k)\varepsilon^2)| \tag{5.45}$$
$$\le\varepsilon^2L^2|n((l-k)\bar{\tau})-(l-k)|^2+\varepsilon^2L|n((l-k)\bar{\tau})-(l-k)||\hat{U}^\varepsilon(0,(l-k)\varepsilon^2)|.$$

By Lemma 3.4,

$$E|\hat{\mathbb{U}}^\varepsilon_{ij}(0,(l-k)\varepsilon^2)|^{2M}\le C_{33}\varepsilon^{4M}(l-k)^{2M}$$

where $C_{33} > 0$ does not depend on l, k or ε. Combining this with (5.10), (5.23), (5.34), (5.36) and (5.43)–(5.45) we obtain that

$$E|\mathbb{V}^{\varepsilon}_{ij}(k\varepsilon^2, l\varepsilon^2)|^{2M} \le C_{34}\varepsilon^{4M}(l-k)^{2M} \tag{5.46}$$

for some $C_{34} > 0$ which does not depend on ε, k or l. Hence, by the Chebyshev inequality

$$\begin{aligned} &P\Big\{ \max_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}} \frac{|\mathbb{V}^{\varepsilon}_{ij}(k\varepsilon^2, l\varepsilon^2)|}{(l-k)^{1-\beta}\varepsilon^{2(1-\beta)}} > 1\Big\} \\ &\le \sum_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}} P\Big\{ \frac{|\mathbb{V}^{\varepsilon}_{ij}(k\varepsilon^2, l\varepsilon^2)|}{(l-k)^{1-\beta}\varepsilon^{2(1-\beta)}} > 1\Big\} \\ &\le C_{34}\varepsilon^{4M} \sum_{0\le k<l\le k+\varepsilon^{-(1+\alpha)},\, l\le T\varepsilon^{-2}} (l-k)^{2\beta M} \le C_{34}\varepsilon^{2M\beta(1-\alpha)-4} \end{aligned} \tag{5.47}$$

where $\alpha \in (0,1)$ is close to 1 and it is chosen similarly to Section 4.2. Taking $\varepsilon = \varepsilon_N = \frac{1}{\sqrt{N}}$ and $M \ge 3\beta^{-1}(1-\alpha)^{-1}$, using (5.39) when (5.40) is satisfied, relying on the Borel–Cantelli lemma and arguing as for (5.37) we obtain that

$$|\mathbb{V}^{\varepsilon}_{ij}(s,t)| \le C_{\alpha,\beta}|t-s|^{1-\beta} \tag{5.48}$$

whenever $|t-s| \le \varepsilon^{1-\alpha}$ holds true, where $C_{\alpha,\beta} > 0$ is another a.s. finite random variable which does not depend on s, t and ε.

Let $0 = t_0 < t_1 < ... < t_m = T$ and write

$$\sum_{0\le q<m} |\mathbb{V}^{\varepsilon}_{ij}(t_q, t_{q+1}) - \hat{\mathbb{W}}^{\varepsilon}_{ij}(t_q, t_{q+1})|^{p/2} \le \mathcal{J}^{\varepsilon}_1 + 2^{\frac{p}{2}-1}(\mathcal{J}^{\varepsilon}_2 + \mathcal{J}^{\varepsilon}_3)$$

where for $\alpha \in (0,1)$,

$$\begin{aligned} \mathcal{J}^{\varepsilon}_1 &= \sum_{0\le q<m,\, t_{i+1}-t_i > N_\varepsilon^{-(1-\alpha)}} |\mathbb{V}^{\varepsilon}_{ij}(t_q, t_{q+1}) - \hat{\mathbb{W}}^{\varepsilon}_{ij}(t_q, t_{q+1})|^{p/2}, \\ \mathcal{J}^{\varepsilon}_2 &= \sum_{0\le q<m,\, t_{q+1}-t_q \le N_\varepsilon^{-(1-\alpha)}} |\mathbb{V}^{\varepsilon}_{ij}(t_q, t_{q+1})|^{p/2} \quad \text{and} \\ J^{\varepsilon}_3 &= \sum_{0\le q<m,\, t_{q+1}-t_q \le N_\varepsilon^{-(1-\alpha)}} |\hat{\mathbb{W}}^{\varepsilon}_{ij}(t_q, t_{q+1})|^{p/2}. \end{aligned}$$

Observe that

$$\mathbb{V}^{\varepsilon}_{ij}(s,t) = \mathbb{V}^{\varepsilon}_{ij}(t) - \mathbb{V}^{\varepsilon}_{ij}(s) - V^{\varepsilon}_i(s)(V^{\varepsilon}_j(t) - V^{\varepsilon}_j(s))$$

and

$$\hat{\mathbb{W}}^{\varepsilon}_{ij}(s,t) = \hat{\mathbb{W}}_j(t) - \hat{\mathbb{W}}^{\varepsilon}_{ij}(s) - \hat{W}^{\varepsilon}_i(s)(\hat{W}^{\varepsilon}_j(t) - \hat{W}^{\varepsilon}_j(s)).$$

We estimate $\mathcal{J}^{\varepsilon}_1$ taking into account Lemma 3.10, the supremum norm estimates in (2.26) and (2.27), the fact that there exist no more than $[TN_\varepsilon^{1-\alpha}]$ intervals $[t_q, t_{q+1}]$ with length exceeding $N_\varepsilon^{-(1-\alpha)}$ and then proceed similarly to (4.26)–(4.28) obtaining that

$$\mathcal{J}^{\varepsilon}_1 = O(\varepsilon^{p\delta+2(\alpha-1)})$$

where $1-\alpha > 0$ can be taken arbitrarily small and $\delta > 0$ comes from Lemma 3.10 and the supremum norm estimates of Theorem 2.5. Next, $\mathcal{J}^{\varepsilon}_2$ is estimated using (5.48) taking into account that $\sum_{0\le q<m} |t_{q+1} - t_q| = T$ and arguing similarly to (3.31) and (5.38). Finally, $\mathcal{J}^{\varepsilon}_3$ is estimated similarly to (3.33)–(3.35) and (4.31)–(4.33). This completes the proof of (2.27). Now, Theorem 2.4 follows from Theorem 2.5 and the rough paths theory arguments given in Section 6.4.2 below.

6 Rough paths and diffusion approximation

We start this section with a review of some elements of rough path theory, pointing whenever possible to [22]. Most results therein are formulated in Hölder spaces, the extension to càdlàg p-variation spaces is found in [23, 25].

6.1 Review and local Lipschitz continuity of Itô-Lyons map

6.1.1 Rough paths

Consider a càdlàg path $U : [0,T] \to \mathbb{R}^e$ of finite p-variation on $[0,T]$, so that

$$\|U\|_{p,[0,T]} := \left(\sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |U(s,t)|^p \right)^{\frac{1}{p}} < \infty \tag{6.1}$$

with path increments $U(s,t) := U(t) - U(s) \in \mathbb{R}^e$, additive in the sense that $U(s,t) + U(t,u) = U(s,u)$. For $p \in [1,\infty)$ this defines a seminorm (it does not separate constants). For $p = 1$, iterated Riemann–Stieltjes (RS) integration defines second order increments $\mathbb{U}(s,t) \in \mathbb{R}^e \otimes \mathbb{R}^e$, càdlàg in both variables,

$$\mathbb{U}^{ij}(s,t) := \int_{(s,t]} (U^i(r-) - U^i(s))dU^j(r), \qquad 1 \le i,j \le e. \tag{6.2}$$

Such increments are non-additive; elementary (additivity) properties of the integral gives *Chen's relation* (cf. [22, Ch. 2])

$$\mathbb{U}(s,t) + U(s,t) \otimes U(t,u) + \mathbb{U}(t,u) = \mathbb{U}(s,u), \quad 0 \le s \le t \le u \le T, \tag{6.3}$$

with tensor notation that relieves us from spelling out coordinates. Thanks to classical works of Young (cf. [22, Ch. 4]) this extends to $p \in [1,2)$, such that[1]

$$\|\mathbb{U}\|_{(p/2),[0,T]} = \left(\sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |\mathbb{U}(s,t)|^{\frac{p}{2}} \right)^{\frac{2}{p}} < \infty. \tag{6.4}$$

Let now $p \in [2,3)$. There is no more (Riemann–Stieltjes or Young) meaning to (6.2), instead we consider $\mathbb{U}$ as part of what we mean by a path: by definition, a (level-2, càdlàg) p-*rough path* (over $\mathbb{R}^e$, on $[0,T]$) is a pair $\mathbf{U} = (U,\mathbb{U})$, càdlàg, where one imposes the algebraic Chen relation (6.3) and the analytic regularity conditions (6.1), (6.4), so that

$$\|\mathbf{U}\|_{p,[0,T]} := \|U\|_{p,[0,T]} + \|\mathbb{U}\|_{(p/2),[0,T]} < \infty\ . \tag{6.5}$$

If $U(0)$ is fixed, or upon identifying paths with identical increments, one can equivalently regard $\mathbf{U}$ as 2-parameter (càdlàg) function $(s,t) \mapsto \mathbf{U}(s,t) = (U(s,t),\mathbb{U}(s,t)) \in \mathbb{R}^e \oplus (\mathbb{R}^e \otimes \mathbb{R}^e) =: G$, a (Lie)group equipped with multiplication $(a,M)\star(b,N) = (a+b, M+a\otimes b+N)$, inverse $(a,M)^{-1} := (-a, -M+a\otimes a)$, and identity $(0,0)$. Addivity of U and Chen's relation then take the appealing form

$$\mathbf{U}(s,t) \star \mathbf{U}(t,u) = \mathbf{U}(s,u). \tag{6.6}$$

From $\mathbf{U}(s,t) = \mathbf{U}(0,s)^{-1} \star \mathbf{U}(0,t)$ we see that $t \mapsto \mathbf{U}(0,t)$ contains all information which suggests an essentially equivalent definition of rough path as genuine G-valued càdlàg

[1] The spaces $\mathbb{R}^e, \mathbb{R}^e \otimes \mathbb{R}^e$ are equipped with compatible norms, all denoted by $|\cdot|$, compatible in the sense that $|v \otimes w| \le |v||w|$ for all $v,w \in \mathbb{R}^e$.

path $t \mapsto \mathbf{U}(t)$, with induced group increments $\mathbf{U}(s,t) = (U(s,t), \mathbb{U}(s,t)) = \mathbf{U}(s)^{-1} \star \mathbf{U}(t)$, subject only to the regularity condition (6.5).

In many cases $\mathbb{U}$ arises from some sort of (possibly stochastic) integration of some path (or process) U against itself, and hence scales like λ^2 upon replacing U by λU. This suggests a purely analytic *dilation* of rough paths, with pointwise definition

$$\delta_\lambda \mathbf{U}(s,t) := (\lambda U(s,t), \lambda^2 \mathbb{U}(s,t)). \tag{6.7}$$

The *homogenous rough path norm*

$$|||\mathbf{U}|||_{p,[0,T]} := \|U\|_{p,[0,T]} + \|\mathbb{U}\|^{1/2}_{(p/2),[0,T]} < \infty\ . \tag{6.8}$$

then has the desirable property $|||\delta_\lambda \mathbf{U}|||_{p,[0,T]} = \lambda |||\mathbf{U}|||_{p,[0,T]}, \lambda \geq 0$, and is often preferable to its non-homogenous counterpart (6.5). The latter however gives rise to the *(inhomogenous) p-rough path distance*[2]

$$\|\mathbf{U}; \tilde{\mathbf{U}}\|_{p,[0,T]} := \|\mathbf{U} - \tilde{\mathbf{U}}\|_{p,[0,T]} := \|U - \tilde{U}\|_{p,[0,T]} + \|\mathbb{U} - \tilde{\mathbb{U}}\|_{(p/2),[0,T]}, \tag{6.9}$$

with respect to which the Itô-Lyons map turns out locally Lipschitz continuous. (At occasions, its homogenous counterpart can also be useful.)

6.1.2 Semimartingales as rough paths

The main motivation for this construction comes from stochastic analysis. Indeed, if $U = U(t,\omega)$ is càdlàg semimartingale, on $\mathbb{R}^e$, then a.s. its Itô lift $\mathbf{U}^{\text{Itô}}(t;\omega) = (U(t;\omega), \mathbb{U}^{\text{Itô}}(0,t;\omega))$, with increments $U(s,t;\omega) = U(t;\omega) - U(s;\omega) \in \mathbb{R}^e$ and

$$\mathbb{U}^{ij;\text{Itô}}(s,t;\omega) = \left(\int_{(s,t]} (U^i(r-) - U^i(s)) dU^j(r) \right)(\omega), \qquad 1 \leq i,j \leq e, \tag{6.10}$$

has the correct p and $p/2$ variation regularity, any $p > 2$, and hence constitutes a p-rough path, for any $p \in (2,3)$, over $\mathbb{R}^e$ and on compact time horizon $[0,T]$. The afore-mentioned p-variation regularity of a semimartingale is classical (see [38, Thm. 1]; the argument relies on representing a càdàg martingale as time-changed Brownian motion). In case $e = 1$, this already gives (via Itô's formula) the $p/2$ variation regularity of $\mathbb{U}^{\text{Itô}}$; for the general case of multidimensional càdlàg semimartingales see [12, 26], for the continuous case see [24, Ch.14] and references therein.

6.2 Rough differential equations

Let b and $\sigma_1, ..., \sigma_e$ be vector fields on $\mathbb{R}^d$, sufficiently smooth for all derivatives below to exist. As is common in this context, we regard $(\sigma_1, ..., \sigma_e)$ as $(d \times e)$-matrix valued function $\sigma : \mathbb{R}^d \to L(\mathbb{R}^e, \mathbb{R}^d)$. By one of several equivalent definitions, e.g. [22, Ch.8.7], it is said that Y solves the (càdlàg) *rough differential equation (RDE)*

$$dY = b(Y^-)dt + \sigma(Y^-)d\mathbf{U}$$

iff, for all $0 \leq s < t \leq T$, and $i = 1, ..., d$ one has[3]

$$Y_t - Y_s = b(Y_s)(t-s) + \sigma(Y_s)U_{s,t} + D\sigma(Y_s)\sigma(Y_s)\mathbb{U}_{s,t} + R_{s,t}\ , \tag{6.11}$$

[2] The notation $\|\mathbf{U}; \tilde{\mathbf{U}}\|_{p,[0,T]}$ is a gentle reminder of the non-linear nature of rough path spaces, as is evident from Chen's relation.

[3] In coordinates, $(D\sigma(Y_s)\sigma(Y_s)\mathbb{U}_{s,t})_i = \sum_{1 \leq l,j \leq e} \sum_{1 \leq k \leq d} \partial_k \sigma_{ij}(Y_s) \sigma_{kl}(Y_s) \mathbb{U}^{lj}_{s,t}$.

with small remainder R, in the sense that

$$\sup_{\mathcal{P}(\varepsilon)} \sum_{[s,t]\in\mathcal{P}(\varepsilon)} |R_{s,t}| \to 0 \quad \text{as } \varepsilon \to 0,$$

with supremum taken over all partitions $\mathcal{P}(\varepsilon)$ of $[0,T]$, with mesh-size less than ε. This definition first encodes that $(Y,Y') = (Y,\sigma(Y)) \in \mathscr{D}_U^{p/2}$, is a controlled (càdlàg) rough path, cf. [22, Ch.4]), [25], and "$p/2$"-remainder $Y^{\#}_{s,t} = D\sigma(Y_s)\sigma(Y_s)\mathbb{U}_{s,t} + R_{s,t}$ for which $\|Y^{\#}\|_{(p/2)} < \infty$. As a consequence (of càdlàg rough integration theory [23]), we can sum over $[s,t] \in \mathcal{P}$, which is a partition of $[0,u]$, and see that Y satisfies a bona fide rough integral equation (RDE), for all $u \in (0,T]$,

$$Y_u = Y_0 + \int_{(0,u]} b(Y_s^-)ds + \int_{(0,u]} \sigma(Y_s^-)d\mathbf{U}. \tag{6.12}$$

Conversely, (6.11) is satisfied by every solution of this integral equation. We are interested in discrete-time approximations. Concerning the drift term this amounts to replace ds by dA^N with the step function $A^N_s := [sN]/N$. We note that $\Delta^N_s := s - A^N(s) \to 0$, uniformly, with the rate 1, i.e. $\|\Delta^N\|_\infty = O(N^{-1})$. We also have uniform 1-variation bounds on compacts. By an easy interpolation argument (see e.g [24, Sec. 8.5]), $\|\Delta^N\|_{q,[0,T]} \le \|\Delta^N\|_\infty^{1-1/q} \|\Delta^N\|_{1,[0,T]}^{1/q}$, we see that we have q-variation convergence with the rate $(1-1/q) > 0$, whenever $q > 1$. This motivates to consider extensions to more general drift terms of the form

$$dY = b(Y^-)dA + \sigma(Y^-)d\mathbf{U}. \tag{6.13}$$

with càdlàg $t \mapsto A(t)$ of finite q-variation. For $q \le p/2$ and bounded b, which suffices for our purposes, the contribution of this drift term can be absorbed in $Y^{\#}$, hence can be treated as a perturbation of the drift-free case. The following theorem gives (well-known) conditions for well-posedness and a quantitative, local Lipschitz estimate for the solution (a.k.a. Itô-Lyons) map, comparing Y to the solution of another RDE,

$$d\tilde{Y} = b(\tilde{Y}^-)d\tilde{A} + \sigma(\tilde{Y}^-)d\tilde{\mathbf{U}}. \tag{6.14}$$

6.1 Theorem. *Let $p \in [2,3), q \in [1,p/2]$ and $b,\sigma \in C^3_b$. Then there exist unique càdlàg solutions to* (6.13) *and* (6.14) *with given initial data Y_0 and $\tilde{Y}_0$. Moreover, the solution map is locally Lipschitz in the precise sense*

$$\|Y - \tilde{Y}\|_{p,[0,T]} \le Ce^{C\ell^3}\big\{\|A - \tilde{A}\|_{q,[0,T]} + \|\mathbf{U} - \tilde{\mathbf{U}}\|_{p,[0,T]} + |Y_0 - \tilde{Y}_0|\big\}$$

for some $C = C(p;b,\sigma)$, whenever

$$\max\{||A||_{q,[0,T]}, ||\tilde{A}||_{q,[0,T]}, |||\mathbf{U}|||_{p,[0,T]}, |||\tilde{\mathbf{U}}|||_{p,[0,T]}\} \le \ell$$

Remark 6.1. (i) In our application $p > 2$ so we can take $q > 1$, as fits our needs. (ii) In a setting of continuous geometric rough paths this estimate is found in [24, Thm 10.26]. The càdlàg extension is found in [25, Thm 3.9] but only written in the drift free case $b \equiv 0$. The Lipschitz estimate for càdlàg RDEs with drift appears in [13], but without explicit dependence on ℓ. (ii) We have not pushed for optimal assumptions on b, σ. In the present form, this gives us the convenience, used in the proof below, to reduce everything to the drift-free case.

Proof. As noted before (6.11) we write $(\sigma_1, ..., \sigma_e) \leftrightarrow \sigma$, i.e. identify the noise vector fields with the map $\mathbb{R}^d \ni y \mapsto ((\xi^i) \mapsto \sum_i \sigma_i(y)\xi^i) \in L(\mathbb{R}^e, \mathbb{R}^d)$. We can treat (6.13) and (6.14) as (drift-free) RDEs with vector fields

$$(b, \sigma_1, ..., \sigma_e) \leftrightarrow \sigma^{\text{ext}},$$

with $\sigma^{\text{ext}}(y) \in L(\mathbb{R}^{1+e}, \mathbb{R}^d)$, driven by the $\mathbf{U}^{\text{ext}}$, the canonically defined rough path associated to $(A, U, \mathbb{U})$, with all "missing" iterated integrals (between A and components of U) canonically defined in the Young sense (see Section 4.1 in [22]). Moreover, standard estimates for Young integrals imply

$$|||\mathbf{U}^{\text{ext}}|||_{p,[0,T]} \leq c\Big(||A||_{q,[0,T]} + |||\mathbf{U}|||_{p,[0,T]}\Big),$$

for some constant $c = c(p, q)$, as well as,

$$\|\mathbf{U}^{\text{ext}} - \tilde{\mathbf{U}}^{\text{ext}}\|_{p,[0,T]} \leq c\,\ell\Big(\|A - \tilde{A}\|_{q,[0,T]} + \|\mathbf{U} - \tilde{\mathbf{U}}\|_{p,[0,T]}\Big). \tag{6.15}$$

The claimed estimates then follow by applying [25, Thm 3.9]. □

We state a corollary for families of RDEs, indexed by $N \in \mathbb{N}$, of the form

$$dY_N = b(Y_N^-)dA_N + \sigma(Y_N^-)d\mathbf{U}_N, \quad d\tilde{Y}_N = b(\tilde{Y}_N^-)d\tilde{A}_N + \sigma(\tilde{Y}_N^-)d\tilde{\mathbf{U}}_N$$

with initial data $Y_N(0) = Y(0)$ and $\tilde{Y}_N(0) = \tilde{Y}(0)$, respectively.

Corollary 6.2. *Let p, q, b, σ be as in the previous theorem. Assume $Y(0) = \tilde{Y}(0)$ and let there exist constants C and $\delta > 0$ such that for all $N \in \mathbb{N}$ we have*

$$\|A_N - \tilde{A}_N\|_{q;[0,T]} + \|\mathbf{U}_N - \tilde{\mathbf{U}}_N\|_{p;[0,T]} \leq CN^{-\delta}, \tag{6.16}$$

as well as

$$\|A_N\|_{q;[0,T]} + \|\tilde{A}_N\|_{q;[0,T]} \leq C$$

and

$$|||\tilde{\mathbf{U}}_N|||_{p;[0,T]} \leq C \log\log(N \vee 3). \tag{6.17}$$

Then, for any $\delta' \in (0, \delta)$, and some constant C' not dependent on N,

$$\sup_{0 \leq t \leq T} |Y_N(t) - \tilde{Y}_N(t)| \leq C' N^{-\delta'}.$$

Proof. Without loss of generality $T = 1$ and $C \geq 1$. Apply Theorem 6.1 with $\mathbf{U} = \mathbf{U}_N, \tilde{\mathbf{U}} = \tilde{\mathbf{U}}_N$ and

$$||\tilde{\mathbf{U}}_N|||_p + |||\mathbf{U}_N - \tilde{\mathbf{U}}_N|||_p \leq C \log\log(N \vee 3) + C =: \ell_N$$

It is easy to see that, for every $\eta > 0$, and as $N \to \infty$,

$$C \exp(C\ell_N^3) = O(N^\eta).$$

In combination with $\|\mathbf{U}_N - \tilde{\mathbf{U}}_N\|_{p;[0,T]} = O(N^{-\delta})$ the results follows. □

A word on the assumptions of the previous corollary. With $A_N(s) := [sN]/N$ and $\tilde{A}_N(s) \equiv s$, we already pointed out, as part of the motivation that led us to (6.13), an easy interpolation argument that gives $\|A_N - \tilde{A}_N\|_{q;[0,T]} = O(N^{-(1-1/q)})$. More interestingly, the iterated-logarithmic bound (6.17) precisely holds for families of rescaled Brownian rough paths, as we will now see.

6.3 Brownian rough paths with parameters (Σ, Γ)

6.3.1 Brownian rough paths

Consider a e-dimensional Brownian motion $\mathcal{W} = \mathcal{W}(\omega)$ with a given covariance $\Sigma = E[\mathcal{W}(1) \otimes \mathcal{W}(1)] \in \mathbb{R}^e \otimes \mathbb{R}^e$. Specializing (6.10) to the present situation, we have the *Itô Brownian rough path* $\mathbf{W}^{\text{Itô}} = (\mathcal{W}, \mathbb{W}^{\text{Itô}})$. For any $\Gamma \in \mathbb{R}^e \otimes \mathbb{R}^e$, we may then consider the second level perturbation $\mathbf{W} = (\mathcal{W}, \mathbb{W})$, with

$$\mathbb{W}(s,t) = \mathbb{W}^{\text{Itô}}(s,t) + \Gamma(t-s) = \int_s^t (\mathcal{W}(r) - \mathcal{W}(s)) \otimes d\mathcal{W}(r) + \Gamma(t-s). \tag{6.18}$$

This yields a class of *Brownian rough paths*, with law determined by the parameters (Σ, Γ). Such a Brownian rough paths is really a (Lie) group-valued Brownian motions in the sense that $t \mapsto \mathbf{W}(t) \in G$ has stationary and independent (group) increments, with Brownian scaling valid in the sense that $t \mapsto \delta_\lambda \mathbf{W}(t/\lambda^2), \lambda > 0$, is again a Brownian rough path, equal in law to $\mathbf{W}$. (Dilation δ was introduced in (6.7).) A familiar situation is $\Gamma = \frac{1}{2}\Sigma$, the resulting Brownian rough path is then precisely the *Stratonovich Brownian rough path*, with

$$\mathbb{W}^{\text{Strato};ij}(s,t) = \int_s^t (\mathcal{W}^i(r) - \mathcal{W}^i(s)) \circ d\mathcal{W}^j(r) = \mathbb{W}^{\text{Itô};ij}(s,t) + \frac{1}{2}\Sigma^{ij}(t-s).$$

If we furthermore specialize to $\Sigma = \text{Id}$, so that $\mathcal{W}$ is a *standard* Brownian motion, the Brownian rough paths with parameters $(\text{Id}, 0)$ (resp. $(\text{Id}, \frac{1}{2}\text{Id})$) will be referred to as *standard* Itô- (resp. Stratonovich) Brownian rough paths. See Chapter 4 in [22] for a detailed discussion.

6.3.2 Differential equations driven by Brownian rough paths

Call $Y = Y(\omega)$ the solution to the rough differential equation driven by a typical realization of the Brownian rough paths, that is

$$dY = b(Y)dt + \sigma(Y)d\mathbf{W} = b(Y)dt + \sigma(Y)d(\mathcal{W}, \mathbb{W}).$$

It is well-known [22, Theorem 9.1] that this yields a solution to the Itô, resp. Stratonovich, stochastic differential equation whenever $\mathbb{W} = \mathbb{W}^{\text{Itô}}$, resp. $\mathbb{W}^{\text{Strato}}$. (This extends further to semimartingales, [24, Ch.14], [12].) For a general Brownian rough path, with $\mathbb{W}(s,t) = \mathbb{W}^{\text{Itô}}(s,t) + \Gamma(t-s)$ as given in (6.18), the very definition of a RDE solution (6.11), with $(U_{s,t}, \mathbb{U}_{s,t})$ replaced by $(\mathcal{W}_{s,t}, \mathbb{W}_{s,t})$ immediately shows that

$$dY = b(Y)dt + \sigma(Y)d(\mathcal{W}, \mathbb{W}) = \tilde{b}(Y)dt + \sigma(Y)d(\mathcal{W}, \mathbb{W}^{\text{Itô}}) \tag{6.19}$$

with the drift vector field $\tilde{b} = b + c$ determined for $i = 1, .., d$ by

$$c_i(y) = (D\sigma(y)\sigma(y)\Gamma)_i = \sum_{1 \le l,j \le e} \sum_{1 \le k \le d} \partial_k \sigma_{ij}(Y_s)\sigma_{kl}(Y_s)\Gamma^{lj}. \tag{6.20}$$

The reason is simply that the defining second order term, part of the very definition (6.11), expands as

$$\sum_{1 \le l,j \le e} \sum_{1 \le k \le d} \partial_k \sigma_{ij}(Y_s)\sigma_{kl}(Y_s)\mathbb{W}^{lj}_{s,t} = \sum_{1 \le l,j \le e} \sum_{1 \le k \le d} \partial_k \sigma_{ij}(Y_s)\sigma_{kl}(Y_s)\mathbb{W}^{\text{Itô},lj}_{s,t} + c_i(t-s).$$

Appealing again to [22, Theorem 9.1], we see that the (random) RDE solution to (6.19), with deterministic initial data is a strong solution to the Itô SDE $dY = \tilde{b}(Y)dt + \sigma(Y)dB$.

6.3.3 Rescaling Brownian rough paths and LIL type estimates

Consider now a Brownian rough path $\mathbf{W} = (\mathcal{W}, \mathbb{W})$ with parameters (Σ, Γ) and introduce

$$\mathbf{W}_N(t) = \delta_{N^{-1/2}} \mathbf{W}(Nt)$$

for $N \in \mathbb{N}$. With $\mathbf{W}_N = (W_N, \mathbb{W}_N)$ this means $W_N(t) = N^{-1/2}\mathcal{W}(Nt)$ and

$$\mathbb{W}_N(s,t) = N^{-1}\mathbb{W}(Ns, Nt) = \int_s^t (W_N(r) - W_N(s)) \otimes dW_N(r) + \Gamma(t-s)$$

so that each $(W_N, \mathbb{W}_N)$ is a Brownian rough path with the same parameters (Γ, Σ).

Proposition 6.3. *For every $T > 0$, and $p > 2$,*

$$|||\mathbf{W}_N|||_{p;[0,T]} \le C_T(\omega)\sqrt{\log\log(N \vee 3)}.$$

for some a.s. finite random variable C_T, and all $N \in \mathbb{N}$.

Proof. By assumption, $\mathbf{W}_N$ is obtained by scaling from $\mathbf{W} = (\mathcal{W}, \mathbb{W})$, a Brownian rough path with parameters (Σ, Γ). It suffices to treat the case of standard Stratonovich Brownian rough path, i.e. $(\Sigma, \Gamma) = (\mathrm{Id}, \frac{1}{2}\mathrm{Id}))$, as introduced in Section 6.3.1. Indeed, this reduction is easily obtained from writing $\mathcal{W} = \sqrt{\Sigma}B$ in terms of a standard e-dimensional Brownian motion B and

$$\mathbb{W}(s,t) = \int_s^t \mathcal{W}(s,r) \otimes \circ d\mathcal{W}(r) + (\Gamma - \tfrac{1}{2}\mathrm{Id})(t-s)$$

so that, for some constant $c = c(\Sigma, \Gamma, T)$,

$$\|\mathbb{W}\|_{(p/2),[0,T]} \le c(\|\mathbb{B}\|_{(p/2),[0,T]} + 1)$$

where

$$\mathbb{B}(s,t) = \int_s^t (B(r) - B(s)) \otimes \circ dB(r).$$

We thus consider the case of the standard Stratonovich Brownian rough path from here on. This allows to use directly the Strassen law established in p-variation rough path topology [37] which states that, with iterated logarithm $\log_2 = \log \circ \log$,

$$\mathbf{Z}_N := \delta_{(2N\log_2 N)^{-1/2}} \mathbf{W}(N\cdot)$$

is a.s. relatively compact, in the space of geometric p-rough paths. The set of limit points given the canonical lift of the *Cameron–Martin unit ball*, that is

$$\left\{ (H, \mathbb{H}) : H : [0,T] \to \mathbb{R}^e, \text{absolutely continuous: } H(0) = 0, \int_0^T |\dot{H}(t)|^2 dt \le 1, \right\},$$

where it is understood in the above that $\mathbb{H}(s,t) = \int_s^t H(s,r) \otimes \dot{H}(r)dr$. Using in particular Cauchy-Schwarz,

$$\|H\|_{p,[0,T]} \le \|H\|_{1,[0,T]} = \int_0^T |\dot{H}(t)|dt \le \sqrt{T}\Big(\int_0^T |\dot{H}(t)|^2 dt\Big)^{1/2};$$

also, the right-hand side of $|\mathbb{H}(s,t)|^{1/2} \le \|H\|_{1,[s,t]} \le \sqrt{|t-s|}\sqrt{\int_s^t |\dot{H}(r)|^2 dr}$ telescopes to $\sqrt{T}\sqrt{\int_0^T |\dot{H}(t)|^2 dt}$ upon summation over any partition of $[0,T]$, hence

$$\|\mathbb{H}\|_{(p/2),[0,T]} \le \|\mathbb{H}\|_{(1/2),[0,T]} \le T\int_0^T |\dot{H}(t)|^2 dt.$$

Restricting to H in the Cameron–Martin unit ball,

$$|||\mathbf{H}|||_{p,[0,T]} = \|H\|_{p,[0,T]} + \|\mathbb{H}\|^{1/2}_{(p/2),[0,T]} \le 2\sqrt{T}\|\dot{H}\|_{L^2} \le 2\sqrt{T}.$$

By Strassen's law for the Brownian rough path [37], a.s. with $N \to \infty$,

$$\inf_{\|\dot{H}\|_{L^2}\le 1} |||\mathbf{Z}_N;\mathbf{H}|||_{p,[0,T]} \to 0,$$

so we can pick (H_N) in the Cameron-Martin unit ball so that, a.s. $|||\mathbf{Z}_N;\mathbf{H}_N|||_p \to 0$. But then

$$|||\mathbf{Z}_N|||_{p,[0,T]} \le |||\mathbf{Z}_N;\mathbf{H}_N|||_{p,[0,T]} + |||\mathbf{H}_N|||_{p,[0,T]} = O(1)$$

so that $|||\mathbf{Z}_N|||_p \le C(\omega)$, for some a.s. finite random variable $C(\omega) = C_T(\omega)$. Now

$$|||\mathbf{Z}_N|||_{p,[0,T]} = (2\log_2 N)^{-1/2}|||\delta_{N^{-1/2}}\mathbf{W}(N.)|||_{p,[0,T]} = (2\log_2 N)^{-1/2}|||\mathbf{W}_N|||_{p,[0,T]}$$

hence, absorbing $2^{1/2}$ in the constant $C(\omega)$, we have

$$|||\mathbf{W}_N|||_{p,[0,T]} \le C(\omega)(\log_2 N)^{1/2}. \qquad \square$$

Remark 6.4. Proposition 6.3 holds with integer N replaced by $1/\varepsilon$, as a.s. estimate, uniform over all $\varepsilon \in (0,1]$. A suitable "continuous" formulation of the Strassen law for the Brownian rough path is found in [24, Ex. 13.46], always in conjunction with Brownian scaling and the remark that Hölder – refines p-variation (rough path) topology.

6.4 Diffusion approximations

6.4.1 Discrete dynamics and proof of Theorem 2.1

We rewrite (2.4) as

$$\begin{aligned} X_N((n+1)/N) = X_N(n/N) + \tfrac{1}{N}b(X_N(n/N)) \\ +\sigma(X_N(n/N))(S_N((n+1)/N) - S_N(n/N)) \end{aligned} \tag{6.21}$$

and further as a càdlàg differential equation. Specifically, we regard the rescaled partial sum process S_N as piecewise constant (càdlàg) process and also write $t_N := [tN]/N$ so that

$$dX_N = b\Big(X_N^-\Big)dt_N + \sigma(X_N^-)dS_N.$$

This equation makes sense (equivalently) as Riemann-Stieltjes integral equations and as (càdlàg) rough integral equation, written as

$$dX_N = b\Big(X_N^-\Big)dt_N + \sigma(X_N^-)d\mathbf{S}_N,$$

where $\mathbf{S}_N = (S_N, \mathbb{S}_N)$ is the (pathwise) canonical lift of S_N, a piecewise constant càdlàg process. The assumptions of Theorem 2.1 guarantee, by Theorem 2.2, that

$$\|\mathbf{S}_N;\mathbf{W}_N\|_{p,[0,T]} = O(N^{-\delta}) \quad \text{a.s.}$$

where, in the terminology of Section 6.3.1, we have that $\mathbf{W}_N = (W_N, \mathbb{W}_N)$ is a Brownian rough path, obtained by rescaling a universal Brownian rough path, with parameters (Σ, Γ) identified in Theorem 2.2, with covariance $\Sigma = \varsigma$ from (2.8), and $\Gamma = \hat{\varsigma}$ from (2.8), (2.9), with components given by,

$$\Gamma_{ij} = \hat{\varsigma}_{ij} = \lim_{k\to\infty}\frac{1}{k}\sum_{n=0}^{k}\sum_{m=-k}^{n-1} E(\xi_i(m)\xi_j(n)) = E\Big(\sum_{l=1}^{\infty}\xi_i(0)\xi_j(l)\Big),\ i,j = 1,...,e.$$

With Section 6.3.2, we see that Proposition 6.3 applies to the family $\mathbf{W}_N$ and we can conclude with Corollary 6.2 that

$$\sup_{0\le t\le T} |X_N(t) - \Xi_N(t)| = O(N^{-\delta}) \quad \text{a.s. as} \quad N \ge 1$$

where Ξ_N is the unique solution of the rough differential equation

$$d\Xi_N(t) = \sigma(\Xi_N(t))d\mathbf{W}_N(t) + b(\Xi_N(t))dt = \sigma(\Xi_N(t))d\mathbf{W}_N^{\text{Itô}}(t) + \tilde{b}(\Xi_N(t))dt \tag{6.22}$$

with $\tilde{b} = b + c$ where thanks to (6.20), $c = c(x)$ is given by

$$c_i(x) = \sum_{j,l=1}^{e} \sum_{k=1}^{d} \frac{\partial \sigma_{ij}(x)}{\partial x_k} \hat{\varsigma}_{lj} \sigma_{kl}(x),\ i = 1, ..., d.$$

By basic consistency results of stochastic and rough integration ([12], also [22, Ch. 5]) the (random) RDE solution Ξ_N is also the (unique) solution to the classical Itô stochastic differential equations

$$d\Xi_N(t) = \sigma(\Xi_N(t))dW_N(t) + \tilde{b}(\Xi_N(t))dt.$$

6.4.2 Continuous dynamics and proof of Theorem 2.4

We recall from Theorem 2.5 the definition

$$V^\varepsilon(t) = \varepsilon \int_0^{t\bar{\tau}\varepsilon^{-2}} \xi(s)ds$$

where $\bar{\tau} \in (0, \infty)$. Performing a deterministic time-change $t \to t/\bar{\tau}$ if necessary, we can assume $\bar{\tau} = 1$ and write (2.20) as

$$dX^\varepsilon(t) = b(X^\varepsilon(t))dt + \sigma(X^\varepsilon(t))dV^\varepsilon,$$

and further (equivalently) as rough integral equation

$$dX^\varepsilon(t) = b(X^\varepsilon(t))dt + \sigma(X^\varepsilon(t))d\mathbf{V}^\varepsilon,\ t \in [0, T],$$

where $\mathbf{V}^\varepsilon = (V^\varepsilon, \mathbb{V}^\varepsilon)$ is the (pathwise) canonical lift of V^ε, i.e. $\mathbb{V}^\varepsilon(s,t) = \int_s^t (\delta V^\varepsilon)(s,r) \otimes dV^\varepsilon(r)$. Theorem 2.5 tells us precisely that

$$\|\mathbf{V}^\varepsilon; \mathbf{W}^\varepsilon\|_{p,[0,T]} = O(\varepsilon^\delta) \quad \text{a.s.}$$

for some $\delta > 0$ a.s. taken simultaneously over $\varepsilon \in (0, 1)$. By construction, preceding equation (5.8), the family of Brownian rough paths $\{\mathbf{W}^\varepsilon : \varepsilon \in (0, 1)\}$ is obtained by rescaling a universal Brownian rough path. In the terminology of Section 6.3.1, we have that $\mathbf{W}^\varepsilon = (W^\varepsilon, \mathbb{W}^\varepsilon)$ is a Brownian rough path (by construction, with parameters (Σ, Γ) where the covariance $\Sigma = \varsigma$ is given by (2.23) and Γ comes from Theorem 2.5, i.e.

$$\Gamma_{ij} = \hat{\varsigma}_{ij} + E \int_0^{\tau(\omega)} \xi_j(s, \omega)ds \int_0^s \xi_i(u, \omega)du, \quad i, j = 1, ..., d.$$

To describe the limiting dynamics consider, for each $\varepsilon \in (0, 1)$, the unique solution to the (random) rough differential equation

$$d\Xi^\varepsilon(t) = \sigma(\Xi^\varepsilon(t))d\mathbf{W}^\varepsilon(t) + b(\Xi^\varepsilon(t))\bar{\tau}dt = \sigma(\Xi^\varepsilon(t))d\mathbf{W}^{\text{Itô},\varepsilon}(t) + \tilde{b}(\Xi^\varepsilon(t))dt \tag{6.23}$$

with $\tilde{b} = b\bar{\tau} + c$ where, thanks to (6.20),

$$c_i(x) = \sum_{j,l=1}^{e} \sum_{k=1}^{d} \frac{\partial \sigma_{ij}(x)}{\partial x_k} \big(\Gamma_{lj}\big) \sigma_{kl}(x).$$

The basic continuity result for RDEs, Theorem 6.1, applies a fortiori to continuous p-variation rough paths, so that the arguments given in the càdàg setting in the previous section, adapt immediately (cf. Remark 6.4) to the continuous setting. In particular, we see

$$\sup_{0 \le t \le T} |X^\varepsilon(t) - \Xi^\varepsilon(t/\bar{\tau})| = O(\varepsilon^\delta) \text{ a.s.} \tag{6.24}$$

We then remark that, by basic consistency results of stochastic and rough integration ([22, Ch. 5]), each process Ξ^ε is also the (unique) solution to the classical Itô stochastic differential equation

$$d\Xi^\varepsilon(t) = \sigma(\Xi^\varepsilon(t)) dW^\varepsilon(t) + \tilde{b}(\Xi^\varepsilon(t)) dt$$

and we obtain (2.25). □

6.5 Euler–Maruyama approximation of the Itô Brownian rough paths

We recall the setup. Let $\{W_N : N \in \mathbb{N}\}$ be a family of Brownian motions defined on the same probability space and $\mathbf{W}_N = (W_N, \mathbb{W}_N)$ be the corresponding Itô Brownian rough paths. Set

$$\hat{W}_N(t) = W_N([tN]/N), \quad \hat{\mathbb{W}}_N(s,t) = \int_{(s,t]} (\hat{W}_N(r-) - \hat{W}_N(s)) \otimes d\hat{W}_N(r)$$

so that $(\hat{W}_N, \hat{\mathbb{W}}_N)$ is the canonical (càdlàg) rough path lift of piecewise constant approximations to W_N. We now prove Lemma 3.10, restated here for the reader's convenience.

Lemma 6.5. *For any $T > 0$ and $p > 2$, there exists $\delta > 0$ such that, almost surely,*

$$\|W_N - \hat{W}_N\|_{p,[0,T]} = O(N^{-\delta}), \qquad \|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{\frac{p}{2},[0,T]} = O(N^{-\delta}).$$

Proof. We proceed in four steps. (1) We first get an L^q-version, any $q < \infty$, of these estimates in case $p = \infty$ where we recall $\|X\|_{\infty,[0,T]} = \sup |X(t) - X(s)|$ with sup taken over all $(s,t) \in \Delta_T := \{(s,t) : 0 \le s \le t \le T\}$. (2) Uniform (in N) p-variation estimates, any $p > 2$. (3) An interpolation argument gives us L^q-estimates in p'-variation, any $p' > 2$. (4) At last, the Borel-Cantelli lemma allows us to switch to a.s. convergence. (We insist that all $(W_N, \mathbb{W}_N)$ had identical law, so that in the steps (1)-(3) we could have written $(W, \mathbb{W})$. In the step (4) however, this notation is fully justified.)

Step(1) Write $t^- := [Nt]/N, t^+ = t^- + 1/N$ so that $\hat{W}_N(t) = W_N(t^-)$. Trivially, $\hat{W}_N = W_N$ at times $t \in D_N := \{t_i \equiv i/N : 0 \le i \le NT\}$. For arbitrary times $(s,t) \in \Delta_T$ we use the Hölder modulus, with exponent $\alpha < 1/2$, to see

$$\begin{aligned} \sup_{(s,t)\in\Delta_T} |(W_N(t) - W_N(s)) - (\hat{W}_N(t) - \hat{W}_N(s))| &\le 2 \sup_{t\in[0,T]} |W_N(t) - W_N(t^-)| \\ &\le 2\|W_N\|_{\alpha;[0,T]} (1/N)^\alpha. \end{aligned}$$

By a classical result of Fernique (cf. below for a more general result with precise reference) the law of $\|W_N\|_{\alpha;[0,T]}$ (independent of N) enjoys Gaussian concentration, in the sense that $\mathbb{E}(e^{c\|W_N\|_\alpha^2}) < \infty$, for some $c = c(\alpha, T) > 0$. By expanding $\exp(.)$ we see that the L^q-norm of any r.v. with Gaussian concentration is finite, and in fact, for a constant $C = C(\alpha, T)$,[4]

$$\| \|W_N - \hat{W}_N\|_{\infty,[0,T]} \|_{L^q(\Omega)} \le C\sqrt{q}(1/N)^\alpha.$$

[4] Here and below we dependencies of constants w.r.t. q are made explicit when easy to do so, although this is not required for this proof.

For second level estimates, we first consider the case of partition points, i.e. $(s,t) \in \Delta_{T,N} := \Delta_T \cap D_N^2$. In this case,

$$\mathbb{W}_N(s,t) - \hat{\mathbb{W}}_N(s,t) = \sum_{i=m}^{n-1} (\mathbb{W}_N)(t_i, t_{i+1}) =: S(n) - S(m)$$

noting that $S(n) \equiv \sum_{i=0}^{n-1} (\mathbb{W}_N)(t_i, t_{i+1})$ defines a random walk with centred independent increments, hence a discrete martingale. We then have

$$\sup_{(s,t)\in\Delta_{T,N}} |\mathbb{W}_N(s,t) - \hat{\mathbb{W}}_N(s,t)| \leq \max_{1\leq n<[NT]} |S(n)| =: S^\star([NT]).$$

To deal with non-partition points we focus on $s^- \leq s < s^+ \leq t = t^-$. (The general case, with $t^- \leq t < t^+$, is treated in the same fashion.) With Chen's relation, we have

$$(i) := \mathbb{W}_N(s,t) - \mathbb{W}_N(s^+,t) = \mathbb{W}_N(s,s^+) + W_N(s,s^+) \otimes W_N(s^+,t) +$$

and hence (cf. footnote in Section 6.1.1 on compatibility of norms on $\mathbb{R}^d \otimes \mathbb{R}^d$ and $\mathbb{R}^d$),

$$|(i)| \leq \|\mathbb{W}_N\|_{2\alpha,[0,T]} (1/N)^{2\alpha} + \|W_N\|^2_{\alpha,[0,T]} (1/N)^\alpha |t-s|^\alpha$$

Similarly, using the very defininition of $(\hat{W}_N, \hat{\mathbb{W}}_N)$, we have

$$\begin{aligned} |(ii)| := |\hat{\mathbb{W}}_N(s,t) - \hat{\mathbb{W}}_N(s^+,t)| &= |\hat{W}_N(s,s^+) \otimes \hat{W}_N(s^+,t) + \hat{\mathbb{W}}_N(s,s^+)| \\ &= |W_N(s^-,s^+) \otimes \hat{W}_N(s^+,t)| \\ &\leq \|W_N\|^2_{\alpha,[0,T]} (1/N)^\alpha |t-s|^\alpha \end{aligned}$$

The terms $(i), (ii)$ account for the difference between s and $s^+ \in D_N$. Similarly, one accounts for the difference between $t^- \in D_N$ and t with terms $\widetilde{(i)}, \widetilde{(ii)}$, with identical estimate, so that

$$\begin{aligned} |\mathbb{W}_N(s,t) - \hat{\mathbb{W}}_N(s,t)| &\leq |\mathbb{W}_N(s^+,t) - \hat{\mathbb{W}}_N(s^+,t)| + |(i)| + |(ii)| + |\widetilde{(i)}| + |\widetilde{(ii)}|. \\ &\leq S^\star([NT]) + 2(\|\mathbb{W}_N\|_{2\alpha,[0,T]} \frac{1}{N^{2\alpha}} + 2\|W_N\|^2_{\alpha,[0,T]} \frac{T^\alpha}{N^\alpha}). \end{aligned}$$

Using Doob's inequality, the fact that $S([NT])$ is an element in the second Wiener-Ito chaos (and integrability properties thereof, see e.g. Theorem D.8 in [24], and finally independence of the $\mathbb{W}_N(t_i, t_{i+1})$, we can bound

$$\begin{aligned} \|S^\star([NT])\|_{L^q} &\leq \frac{q}{q-1} \|S([NT])\|_{L^q} \lesssim q \|S([NT])\|_{L^2} \\ &= q \left\| \sum_{i=m}^{[NT]-1} \mathbb{W}_N(t_i, t_{i+1}) \right\|_{L^2} = q \sqrt{\sum_{i=0}^{[NT]-1} (t_i, t_{i+1})^2} \leq q \sqrt{\frac{T+1}{N}} \end{aligned}$$

On the other hand, Fernique estimate for Brownian rough paths, Corollary 13.14 in [24], gives Gaussian concentration of $|||\mathbf{W}_N|||_{\alpha,[0,t]} = \|W_N\|_{\alpha,[0,T]} + \|\mathbb{W}_N\|^{1/2}_{2\alpha,[0,T]}$ which implies a $O(q)$-bound for the L^q-norm of both $\|\mathbb{W}_N\|_{2\alpha,[0,T]}$ and $\|W\|^2_{\alpha,[0,T]}$. Putting it all together, we see for some constant C which does not depend on N or q,

$$\| \sup_{(s,t)\in\Delta_T} |\mathbb{W}_N(s,t) - \hat{\mathbb{W}}_N(s,t)| \|_{L^q} \leq C \frac{q}{N^\alpha}.$$

Step(2) From Proposition 6.17 in [25] we can see that for some constant $C = C(p,T)$ but not dependent on q,

$$\sup_N \| \|\hat{\mathbb{W}}_N\|_{p/2,[0,T]} \|_{L^q} \leq Cq < \infty.$$

(The corresponding first level estimate is trivial in view of the ω-wise estimate $\sup_N \|\hat{W}_N\|_{p,[0,T]} \leq \|W\|_{p,[0,T]}$ and Gaussian concentration of the right-hand side.)

Step(3) We proceed by interpolation, as e.g. in Lemma 5.2 in [25]. Let $2 < p < p'$. In what follows, we spell out the second level estimates (the first level estimates are similar but easier),

$$\begin{aligned} \|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p'/2,[0,T]} &\leq \|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{\infty,[0,T]}^{1-p/p'} \|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p/2,[0,T]}^{p/p'} \\ &\leq \|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{\infty,[0,T]}^{1-p/p'} \times K_N^{p/p'}(\omega) \end{aligned}$$

with $K_N^{p/p'}(\omega)$ can be taken as $2^{p/p'}$ times $\|\mathbb{W}_N\|_{p/2,[0,T]}^{p/p'} + \|\hat{\mathbb{W}}_N\|_{p/2,[0,T]}^{p/p'}$. Let $1/q = 1/q' + 1/q''$ and apply Hölder's inequality to see, also setting $r' = q'(1 - p/p'), r'' = q''p/p'$

$$\begin{aligned} \|\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p'/2,[0,T]}\|_{L^q(\Omega)} &\leq \|\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{\infty,[0,T]}^{1-p/p'}\|_{L^{q'}} \|K^{p/p'}\|_{L^{q''}} \\ &= \|\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{\infty,[0,T]}\|_{L^{r'}}^{1-p/p'} \|K_N\|_{L^{r''}}^{p/p'}. \end{aligned}$$

Thanks to step (2), we have bounds on $\|K_N\|_{L^{r''}}$ which are uniform in N and in fact such that $\sup_N \|K_N\|_{L^{r''}} \leq Cr''$ for some constant $C > 0$ which does not depend on N and r''. But then

$$\|\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p'/2,[0,T]}\|_{L^q(\Omega)} \leq \tilde{C}(r'(1/N)^\eta)^{1-p/p'}(r'')^{p/p'}$$

for another constant $\tilde{C} > 0$ which does not depend on N, r' and r''. The precise choices are not that important but we can take $p = (p'+2)/2$, $q' = q'' = q/2$. In the end,

$$\|\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p'/2,[0,T]}\|_{L^q(\Omega)} \leq C(p')q(1/N)^{\eta'}$$

with $\eta' = \eta(1 - p/p') > 0$.

Step(4) A Borel-Cantelli argument then leads to the a.s. estimates. Indeed, for any $\varepsilon \in (0, \eta')$ pick $q > 1/\varepsilon$ so that, from Chebyshev-Markov's inequality,

$$P\{N^{\eta'-\varepsilon}\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p'/2,[0,T]} > 1\} \leq N^{q(\eta'-\varepsilon)}E(\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p'/2,[0,T]}^q) = O(N^{-q\varepsilon}).$$

Since $q\varepsilon > 1$, this bound is summable in N and the Borel-Cantelli lemma tells us that $\|\mathbb{W}_N - \hat{\mathbb{W}}_N\|_{p'/2,[0,T]} \leq 1/N^{\eta'-\varepsilon}$ for all $N \geqslant N_0$ for some $N_0 = N_0(\omega)$. The almost convergence thus holds with rate $\delta = \eta' - \varepsilon > 0$. (The first level estimates are similar.) □

7 Rough paths and law of iterated logarithm for iterated sums and integrals

The present section is devoted to higher order extensions of Theorem 2.2. In conjunction with a higher-order Strassen law for Brownian rough paths, also shown below, we arrive at a (functional) law of iterated logarithm for iterated sums. We rely here on some higher-order concepts of rough paths. (Detailed references are given but without a systematic review.)

7.1 Lyons' extension for càdlàg rough paths

Let $\mathbf{X} = (X, \mathbb{X})$ be càdlàg p-rough path, $p \in [2, 3)$. One defines inductively iterated rough integrals

$$\bar{\mathbf{X}}^\ell(s,t) = \int_{(s,t]} \bar{\mathbf{X}}^{\ell-1}(s, r-) \otimes d\mathbf{X}(r-) \quad \in (\mathbb{R}^e)^{\otimes \ell}.$$

The entire stack $\mathrm{Ext}(\mathbf{X})(s,t) := \bar{\mathbf{X}}(s,t) = (1, \bar{\mathbf{X}}^1(s,t), \ldots, \bar{\mathbf{X}}^\ell(s,t), \ldots)$, with values in the tensor series over $\mathbb{R}^e$, is known as *Lyons' extension* of $\mathbf{X}$. It is equivalently given as $\bar{\mathbf{X}}(s,t) = \bar{\mathbf{X}}^{-1}(s) \otimes \bar{\mathbf{X}}(t)$, in terms of a linear rough differential equation

$$d\bar{\mathbf{X}}(t) = \bar{\mathbf{X}}(t-) \otimes d\mathbf{X}(t), \;\; \bar{\mathbf{X}}(0) = 1.$$

Example 7.1. If $\mathbb{X}(s,t) = \int_s^t (X(r-) - X(s)) \otimes dX(r)$ for some càdlàg bounded variation path X, then, for all levels $\ell \geqslant 1$,

$$\bar{\mathbf{X}}^\ell(s,t) = \int_{\{s \le r_1 \le \cdots \le r_\ell \le t\}} dX(r_1-) \otimes \cdots \otimes dX(r_\ell -) =: \int_{\Delta^\ell_{s,t}} dX \otimes . \cdots \otimes dX$$

(In case of a piecewise constant càdlàg path X, this becomes an iterated sum.)

Example 7.2. In case of (Itô, resp. Stratonovich) Brownian rough path, we have

$$\bar{\mathbf{W}}^{\text{Itô};\ell}(s,t) = \int_{\Delta^\ell_{s,t}} dW \otimes \cdots \otimes dW, \quad \bar{\mathbf{W}}^{\text{Strato};\ell}(s,t) = \int_{\Delta^\ell_{s,t}} \circ dW \otimes \cdots \otimes \circ dW$$

in the usual Itô- resp. Stratonovich sense.

Example 7.3. In case of a general Brownian rough path $\mathbf{W} = (W, \mathbb{W})$ with parameters (Σ, Γ), we can understand its Lyons extension $\bar{\mathbf{W}} = \mathrm{Ext}(\mathbf{W})$ elegantly as solution to the Itô linear rough differential equation

$$d\bar{\mathbf{W}}(t) = \bar{\mathbf{W}}(t) \otimes dW(t) + \bar{\mathbf{W}}(t) \otimes \Gamma dt, \quad \bar{\mathbf{X}}(0) = 1,$$

followed by setting $\bar{\mathbf{W}}(s,t) = \bar{\mathbf{W}}(s)^{-1} \otimes \bar{\mathbf{W}}(t)$. Given any word $w = (i_1 \cdots i_\ell) \in \{1, ..., e\}^\ell$, of length $|w| = \ell$, and writing $e_w = e_{i_1 \cdots i_\ell} = e_{i_1} \otimes \cdots \otimes e_{i_\ell}$, the components of $\bar{\mathbf{W}}^\ell \in (\mathbb{R}^e)^{\otimes \ell}$ also admit explicit combinatorial expressions, namely

$$\langle \bar{\mathbf{W}}(s,t), e_w \rangle = \langle \bar{\mathbf{W}}^{\text{Itô}}(s,t), e_w \rangle + \sum_v c_v \langle \bar{\mathbf{W}}^{\text{Itô}}(s,t), e_v \rangle$$

with summation over all words v obtained from w by contracting one or more neighbouring pairs $(i_j, i_{j+1}) \in \{1, ..., e\}^2$ to a single letter 0, with the additional convention that $W_0(t) = t$. (That is, $\bar{\mathbf{W}}^{\text{Itô}}$ here should really be understood as the stack of iterated Itô integrals of $(1+e)$-dimensional time-space Brownian motion (W_0, W).) The constants c_v are multiplicative functions of Γ. For instance, if v is obtained by contracting, say, two pairs, $(i_j, i_{j+1}), (i_k, i_{k+1})$, with $1 < j+1 < k < \ell$, then $c_v = \Gamma_{i_j, i_j+1} \Gamma_{i_k, i_k+1}$. This follows in exactly the same way as [11, Prop. 22] and can be seen as algebraic renormalization procedure for rough paths.

Theorem 7.4. *Let $\mathbf{X} = (X, \mathbb{X}), \tilde{\mathbf{X}} = (\tilde{X}, \tilde{\mathbb{X}})$ be càdlàg p-rough paths, $p \in [2,3)$, with*

$$|||\mathbf{X}|||_{p,[0,T]} \vee |||\tilde{\mathbf{X}}|||_{p,[0,T]} \le R \in [1, \infty).$$

Then, for every $\ell \in \mathbb{N}$ there exists $c = O(R^\ell)$, as $R \to \infty$, such that

$$\|\mathrm{Ext}(\mathbf{X})^\ell - \mathrm{Ext}(\tilde{\mathbf{X}})^\ell\|_{p/\ell,[0,T]} \le c(\|X - \tilde{X}\|_{p,[0,T]} + \|\mathbb{X} - \tilde{\mathbb{X}}\|_{p/2,[0,T]}).$$

Proof. This is a variation of [39, Thm 2.2.2], see also [22, Ex 4.6], though what we need is not a direct consequence of these statements (which are given in terms of continuous control functions $\omega(s,t)$ resp. in a Hölder setting with $\omega(s,t) = t - s$). We only illustrate the case $\ell = 3$, the general case being similar, giving a new argument based on local Lipschitz of higher-oder rough integration. (The case $\ell > 3$ goes along the same lines, cf. [22, Sec 4.5].) Recall that the space of (first order) controlled rough paths, $\mathcal{V} = (V, V') \in \mathscr{D}_X^{p/2}$, is Banach with norm[5] $\|\mathcal{V}\|_{X;p/2} \equiv \|\delta V - V' \delta X\|_{p/2} + \|V'\|_p$. For a p-rough path $\mathbf{X} = (X, \mathbb{X})$ we have

$$\mathcal{V} \times \mathbf{X} \mapsto \left(\int (V, V')^- d(X, \mathbb{X}), V, V' \right) =: (Z, Z', Z'') =: \mathcal{Z} \in D^{p/3}_{\mathbf{X}}$$

[5] Here and below, $(\delta X)(s,t) = X(t) - X(s)$ denotes the increments of paths in a linear space. We also write, accordingly, $(\delta V - V'\delta X)(s,t) = V(t) - V(s) - V'(s)(X(t) - X(s))$, $(Z''\mathbb{X})(s,t) = Z''(s)\mathbb{X}(s,t)$ and so on.

where $\mathscr{D}_{\mathbf{X}}^{p/3}$, the space of second order controlled rough paths, is Banach with norm

$$\|\mathcal{Z}\|_{\mathbf{X};p/3} := \|\delta Z - Z'\delta X - Z''\mathbb{X}\|_{p/3} + \|\delta Z' - Z''\delta X\|_{p/2} + \|Z''\|_p.$$

Given another rough path $\widetilde{\mathbf{X}}$, the generic (local) Lipschitz estimate for rough integration gives

$$\begin{aligned}\|\mathcal{Z};\tilde{\mathcal{Z}}\|_{\mathbf{X},\tilde{\mathbf{X}};p/3} &\equiv \|\delta Z - Z'\delta X - Z''\mathbb{X} - (\tilde{Z} - \tilde{Z}'\delta\tilde{X} - \tilde{Z}''\tilde{\mathbb{X}})\|_{p/3} \\ &+ \|\delta Z' - Z''\delta X - (\tilde{Z}' - \tilde{Z}''\delta\tilde{X})\|_{p/2} \\ &+ \|Z'' - \tilde{Z}''\|_p \le c(\|X - \tilde{X}\|_p + \|\mathbb{X} - \tilde{\mathbb{X}}\|_{p/2} + \|\mathcal{V};\tilde{\mathcal{V}}\|_{X,\tilde{X};p/2})\end{aligned}$$

where a constant c can be taken uniformly provided $\mathbf{X}, \widetilde{\mathbf{X}}$ and $\mathcal{V}, \tilde{\mathcal{V}}$ remain bounded in their appropriate (rough resp. controlled rough path) spaces, and $\|\mathcal{V};\tilde{\mathcal{V}}\|_{X,\bar{X};p/2} \equiv \|V' - \tilde{V}'\|_p + \|\delta V - V'\delta X - (\tilde{V} - \tilde{V}'\delta\tilde{X})\|_{p/2}$. We can now apply this with $(V, V') = (\mathbb{X}, X)$, and $(\tilde{V}, \tilde{V}') = (\tilde{\mathbb{X}}, \tilde{X})$, the crucial remark being that in this case

$$\|\mathcal{V};\tilde{\mathcal{V}}\|_{X,\tilde{X};p/2} = \|X - \tilde{X}\|_p + \|\mathbb{X} - \tilde{\mathbb{X}}\|_{p/2}.$$

A moment of reflection (and Chen's relation) reveals that the $p/3$-variation component of $\|\mathcal{Z};\tilde{\mathcal{Z}}\|_{\mathbf{X},\tilde{\mathbf{X}};p/3}$ is then nothing but the $p/3$-variation of the map

$$(s,t) \mapsto \int_s^t \mathbb{X}(s, r-) \otimes dX(r) - \int_s^t \tilde{\mathbb{X}}(s, r-) \otimes d\tilde{X}(r);$$

we thus see that, for $\ell = 3$,

$$\|\mathrm{Ext}(\mathbf{X})^\ell - \mathrm{Ext}(\tilde{\mathbf{X}})^\ell\|_{p/\ell,[0,T]} \le c(\|X - \tilde{X}\|_p + \|\mathbb{X} - \tilde{\mathbb{X}}\|_{p/2}).$$

The argument shows that c can be taken uniformly as $\mathbf{X}, \tilde{\mathbf{X}}$ remain in a bounded set, such as a ball of radius R. To make the dependence on R explicit, we use scaling. Note that $\mathbf{Y} := \delta_{1/R}\mathbf{X}$, and similar for $\tilde{\mathbf{Y}}$, are of (at most) unit size in the norm $|||\cdot|||_{p,[0,T]}$. Application of the above estimate, noting $\mathrm{Ext}(\mathbf{Y})^\ell = (1/R)^\ell \mathrm{Ext}(\mathbf{X})^\ell$, similarly for $\tilde{\mathbf{Y}}$, and

$$\|Y - \tilde{Y}\|_p + \|\mathbb{Y} - \tilde{\mathbb{Y}}\|_{p/2} = R^{-1}\|X - \tilde{X}\|_p + R^{-2}\|\mathbb{X} - \tilde{\mathbb{X}}\|_{p/2} \le \|X - \tilde{X}\|_p + \|\mathbb{X} - \tilde{\mathbb{X}}\|_{p/2},$$

using $R \ge 1$, shows that c can be taken as $O(R^\ell)$. □

7.2 The law of iterated logarithm for iterated sums

7.2.1 Almost sure invariance principle in rough paths metrics beyond level-2

Theorem 7.5. *The conclusion of Theorem 2.2, with fixed $p \in (2,3)$, can be extended to any level $\ell \in \mathbb{N}$. That is,*

$$\|\overline{\mathbf{S}}_N^\ell - \overline{\mathbf{W}}_N^\ell\|_{p/\ell,[0,T]} = O(N^{-\delta}) \quad a.s.$$

where $\overline{\mathbf{S}}_N^\ell(s,t)$ is given by the rescaled ℓ-fold iterated summation

$$\overline{\mathbf{S}}_N^\ell(s,t) = N^{-\ell/2} \sum_{[Ns] \le k_1 < \cdots < k_\ell < [Nt]} \xi(k_1) \otimes \cdots \otimes \xi(k_\ell) \quad \in (\mathbb{R}^e)^{\otimes \ell}$$

and $\overline{\mathbf{W}}_N = (1, \overline{\mathbf{W}}_N^1, \overline{\mathbf{W}}_N^2, \ldots)$ can be given as the solution of a "drift-corrected" Itô stochastic differential equation

$$d\overline{\mathbf{W}}_N = \overline{\mathbf{W}}_N \otimes dW_N + \overline{\mathbf{W}}_N \otimes \Gamma dt, \quad \overline{\mathbf{W}}_N(0) = 1,$$

with associated increments $\overline{\mathbf{W}}_N(s,t) = \overline{\mathbf{W}}_N(s)^{-1} \otimes \overline{\mathbf{W}}_N(t)$, and driving Brownian motion $W_N(t) = N^{-1/2}\mathcal{W}(Nt)$.

Remark 7.6. Decomposing $\bar{\mathbf{W}}^\ell(s,t) = \sum \langle \bar{\mathbf{W}}(s,t), e_w \rangle e_w$ with sum over all words of length $|w| = \ell$, we note that an explicit combinatorial expression of these coefficients, as linear combinations of iterated Itô-integrals of time-space Brownian motion, was given in Example 7.3.

Proof. By Theorem 2.2 the claimed estimate holds true for $\ell = 1, 2$ so that a.s.

$$\|S_N - W_N\|_{p,[0,T]} + \|\mathbb{S}_N - \mathbb{W}_N\|_{p/2,[0,T]} = O(N^{-\delta}).$$

Thanks to $p \in (2,3)$, we can appeal to Theorem 7.4, noting that the polynomial growth $c = c(R)$ therein is more than enough to allow us to proceed similarly to Section 6.4 (Proposition 6.3 and Corollary 6.2). It then suffices to recall that the Lyons extension $\mathrm{Ext}(S_N, \mathbb{S}_N)$ is precisely given by $\overline{\mathbf{S}}_N$, as was pointed out in Example 7.1, and that the description of $\overline{\mathbf{W}}_N = \mathrm{Ext}(W_N, \mathbb{W}_N)$ is given in Example 7.3. ☐

7.2.2 Strassen's functional LIL for Brownian rough paths

A (possibly degenerate) covariance matrix gives rise to a (possibly degenerate) inner product structure, $\langle \tilde{v}, \tilde{v} \rangle_{\Sigma^{-1}} := \langle v, v \rangle = \sum_{i=1}^e v_i^2$ when $\tilde{v} = \sqrt{\Sigma} v$ and $+\infty$ else. Note that $\langle v, v \rangle_{\Sigma^{-1}} = \langle \Sigma^{-1} v, v \rangle$ in the non-degenerate case. An absolutely continuous path $H : [0,T] \to \mathbb{R}^e$, with $H(0) = 0$, is called *Cameron–Martin path* if $\|H\|^2_{\mathcal{H};[0,T]} := \int_0^T \langle \dot{H}(t), \dot{H}(t) \rangle_{\Sigma^{-1}} dt < \infty$. Every such path lifts canonically to a rough path $(H, \mathbb{H})$ with $\mathbb{H} = \int \delta H \otimes dH$, i.e. $\mathbb{H}(s,t) = \int_s^t (H(r) - H(s)) \otimes \dot{H}(r) dr$. We take $T = 1$ in what follows and also need the Cameron–Martin unit ball,

$$\mathcal{K} := \{H \in \mathcal{H} : \|H\|_{\mathcal{H};[0,1]} < \infty\}.$$

Let now $\mathbf{W} = (W, \mathbb{W})$ be a Brownian rough path with parameters (Σ, Γ), with increments of its Lyons extension of the form

$$\bar{\mathbf{W}}(s,t) = (1, \bar{\mathbf{W}}^1(s,t), \ldots, \bar{\mathbf{W}}^\ell(s,t), \ldots).$$

Set also

$$\mathcal{K}^\ell := \left\{\mathbf{H}^\ell : H \in \mathcal{K}\right\}, \quad \mathbf{H}^\ell(s,t) = \int_{\Delta^\ell_{s,t}} dH \otimes \ldots \otimes dH. \tag{7.1}$$

Proposition 7.7. *For any $\ell \in \mathbb{N}$ and $p > 2$, as $n \to \infty$, a.s.*

$$\inf_{H \in \mathcal{K}} \left\| (2n \log\log n)^{-\frac{\ell}{2}} \bar{\mathbf{W}}^\ell(n\cdot, n\cdot) - \mathbf{H}^\ell \right\|_{p/\ell;[0,1]} \to 0,$$

and the set of limit points of the above sequence equals $\mathcal{K}^\ell$.

Remark 7.8. The same proof yields the same statement in stronger $\alpha\ell$-Hölder sense, $\alpha = 1/p \in (1/3, 1/2)$.

Proof. Strassen's functional LIL for Brownian motion is a well-known consequence of Schilder's theorem. Typically formulated in ∞-topology (e.g. [19]), extensions to iterated stochastic integrals [2] and to α-Hölder topology [3] have appeared in the literature. Similarly, a Schilder theorem for the Brownian rough path gives Strassen's law in rough path topology, as was first seen in the p-variation, then α-Hölder rough path topology, see [37, 24] and references therein.

Since the afore-mentioned results only deal with standard Brownian motion B with $\mathbb{B} = \int \delta B \otimes \circ dB$ we quickly treat the case of a general $\mathbf{W} = (W, \mathbb{W})$, a Brownian rough path with parameters (Σ, Γ). To this end, let $\sigma = \sqrt{\Sigma}$ and note that

$$W = \langle \sigma, B(t) \rangle,\ \mathbb{W}(s,t) = \left\langle \sigma \otimes \sigma, \int \delta B \otimes dB \right\rangle + \Gamma(t-s) = \langle \sigma \otimes \sigma, \mathbb{B} \rangle + \tilde{\Gamma}(t-s),$$

with Itô-Stratonovich corrected $\tilde{\Gamma} = \Gamma - \Sigma/2$. The map

$$(B, \mathbb{B}) \mapsto (W, \mathbb{W}) = \mathbf{W} \mapsto (\bar{\mathbf{W}}^1, \cdots, \bar{\mathbf{W}}^\ell) \equiv \bar{\mathbf{W}}^{\le \ell}$$

is continuous between the appropriate rough path spaces, the contraction principle then shows that

$$\{(\varepsilon \bar{\mathbf{W}}^1, \ldots, \varepsilon^\ell \bar{\mathbf{W}}^\ell) : \varepsilon > 0\}$$

satisfies a LDP with speed ε^2 in p-variation (or $1/p$-Hölder) rough path topology, with good rate function

$$I(\mathbf{H}) = \frac{1}{2} \|H\|^2_{\mathcal{H};[0,1]}$$

whenever $\mathbf{H} = \left(\mathbf{H}^1, \ldots, \mathbf{H}^\ell\right)$ is the canonical lift of H, and $+\infty$ else. Note that this rate function depends on Σ but not on Γ. As in [37, 24], it then follows that Strassen's functional LIL holds in the stated higher order generality,

$$\left\| (2n \log \log n)^{-\frac{\ell}{2}} \bar{\mathbf{W}}^\ell (n\cdot, n\cdot); \boldsymbol{\mathcal{K}}^\ell \right\|_{p/\ell;[0,1]} \to_{n\to\infty} 0,$$

together with the stated characterization of the limit points of this sequence. (We used notation $\|x; A\| = \inf_{y \in A} \|x - y\|$.) □

7.2.3 Strassen's theorem for iterated sums

Consider

$$\bar{\boldsymbol{\xi}}^\ell (m, n) := \sum_{m \le k_1 < \cdots < k_\ell < n} \xi(k_1) \otimes \cdots \otimes \xi(k_\ell)$$

and recall that $\mathcal{K} \subset \mathcal{H}$ defines the unit Cameron–Martin ball.

Theorem 7.9. *For any $\ell \in \mathbb{N}$ and $p > 2$, as $N \to \infty$, a.s.*

$$\inf_{H \in \mathcal{K}} \left\| (2N \log \log N)^{-\frac{\ell}{2}} \bar{\boldsymbol{\xi}}^\ell ([N\cdot], [N\cdot]) - \mathbf{H}^\ell \right\|_{p/\ell;[0,1]} \to 0 \tag{7.2}$$

and the set of limit points given by (7.1). *In particular,*

$$\inf_{H \in \mathcal{K}} \left| (2N \log \log N)^{-\frac{\ell}{2}} \sum_{0 \le k_1 < \cdots < k_\ell < N} \xi(k_1) \otimes \cdots \otimes \xi(k_\ell) - \mathbf{H}^\ell(0,1) \right| \to 0, \tag{7.3}$$

with set of limit points given by $\left\{\mathbf{H}^\ell(0,1) : H \in \mathcal{K}\right\}$.

Proof. (i) Recall $\overline{\mathbf{S}}^\ell_N(s,t) = N^{-\ell/2} \bar{\boldsymbol{\xi}}^\ell([Ns],[Nt])$ is precisely the Lyons lift of $(S_N, \mathbb{S}_N)$. Recall also $\overline{\mathbf{W}}^\ell_N(s,t) = N^{-\ell/2} \overline{\mathbf{W}}^\ell(Ns, Nt)$. Then $\overline{\mathbf{S}}^\ell_N(\cdot,\cdot) - \overline{\mathbf{W}}^\ell_N(\cdot,\cdot) = O(N^{-\delta})$ from Theorem 7.5 above shows

$$\bar{\boldsymbol{\xi}}^\ell([N\cdot],[N\cdot]) - \overline{\mathbf{W}}^\ell(N\cdot, N\cdot) = O(N^{\ell/2 - \delta}),$$

always in p/ℓ-variation sense on $[0,1]$, hence

$$(2N \log \log N)^{-\frac{\ell}{2}} \bar{\boldsymbol{\xi}}^\ell([N\cdot],[N\cdot]) - (2N \log \log N)^{-\frac{\ell}{2}} \overline{\mathbf{W}}^\ell(N\cdot, N\cdot) = o(1)$$

and so the functional LIL for $\bar{\boldsymbol{\xi}}^\ell$, as stated in (7.2), follows directly from the one for $\overline{\mathbf{W}}^\ell$ in Proposition 7.7 above. As for (7.3) it suffices to note that any variation norm on $[0,1]$ dominates the increment over the unit interval. □

The following corollary can be seen as generalization of [2, Cor. 3.2] which dealt with iterated Brownian integrals.

Corollary 7.10. *Let* $A \in (\mathbb{R}^e)^{\otimes \ell}$ *and define the tensor contraction* $\mathcal{X}_N^\ell := \left\langle A, \sum_{0 \le k_1 < \cdots < k_\ell < N} \xi(k_1) \otimes \cdots \otimes \xi(k_\ell) \right\rangle$ *with values in the reals. Then*

$$P\left(\limsup_{N\to\infty} \frac{\mathcal{X}_N^\ell}{(2N \log\log N)^{\frac{\ell}{2}}} = M \right) = 1$$

with

$$M = \sup\left\{ \left\langle A, \int_{\Delta_{0,1}^\ell} dH \otimes \ldots \otimes dH \right\rangle : \|H\|_{\mathcal{H};[0,T]} \le 1 \right\}. \tag{7.4}$$

Proof. Immediate from (7.3) and accompanying description of the limit set. □

7.3 The law of iterated logarithm for iterated integrals

The arguments of the last section immediately extend to the case of iterated integrals, and in particular lead to a proof of Corollary 2.6. The analogue of Theorem 7.5 reads

Theorem 7.11. *The conclusion of Theorem 2.5, with fixed* $p \in (2,3)$*, can be extended to any level* $\ell \in \mathbb{N}$*. That is,*

$$\|\overline{\mathbf{V}}_\ell^\varepsilon - \overline{\mathbf{W}}_\ell^\varepsilon\|_{p/\ell,[0,T]} = O(N^{-\delta}) \quad \text{a.s.}$$

where $\overline{\mathbf{V}}_\ell^\varepsilon$ *is given by the rescaled* ℓ*-fold iterated integrals,*

$$\overline{\mathbf{V}}_\ell^\varepsilon(s,t) = \varepsilon^\ell \int_{\{s\bar\tau\varepsilon^{-2} \le r_1 \le \cdots \le r_\ell \le t\bar\tau\varepsilon^{-2}\}} \xi(r_1) \otimes \cdots \otimes \xi(r_\ell) dr_1 \cdots dr_\ell$$

and $\overline{\mathbf{W}}_N = (1, \overline{\mathbf{W}}_N^1, \overline{\mathbf{W}}_N^2, \ldots)$ *exactly as in Theorem 7.5, just with updated covariance for the Brownian motion* $\overline{\mathbf{W}}_N^1 = W_N$*, namely the covariance given in Theorem 2.5, and* $\Gamma = \{\Gamma_{ij} : 1 \le i,j \le e\}$ *given by*

$$\Gamma_{ij} = \sum_{l=1}^\infty E(\eta_i(0)\eta_j(l)) + E\Big(\int_0^{\tau(\omega)} \xi_j(s,\omega) ds \int_0^s \xi_i(u,\omega) du \Big).$$

Proof. Similar to Theorem 7.5: by Theorem 2.5 the claimed estimate holds true for $\ell = 1,2$ and thanks to $p \in (2,3)$, we can appeal to Theorem 7.4, noting that the Lyons extension of V^ε is precisely given by $\overline{\mathbf{V}}^\varepsilon = (V^\varepsilon, \mathbb{V}^\varepsilon)$. □

We can now deduce, as in the discrete case, a functional LIL for iterated integrals from the corresponding statement for Brownian rough paths; Proposition 7.7. We set

$$\bar{\boldsymbol{\xi}}^\ell(s,t) := \int_{\{s \le r_1 \le \cdots \le r_\ell \le t\}} \xi(r_1) \otimes \cdots \otimes \xi(r_\ell) dr_1 \cdots dr_\ell \quad \in (\mathbb{R}^e)^{\otimes \ell}.$$

Theorem 7.12. *For any* $\ell \in \mathbb{N}$ *and* $p > 2$*, as* $N \to \infty$*, a.s.*

$$\inf_{H \in \mathcal{K}} \left\| (2N \log\log N)^{-\frac{\ell}{2}} \bar{\boldsymbol{\xi}}^\ell(\bar\tau N \cdot, \bar\tau N \cdot) - \mathbf{H}^\ell \right\|_{p/\ell;[0,1]} \to 0$$

and the set of limit points given by (7.1)*. In particular,*

$$\inf_{H \in \mathcal{K}} \left| (2N \log\log N)^{-\frac{\ell}{2}} \bar{\boldsymbol{\xi}}^\ell(0, \bar\tau N) - \mathbf{H}^\ell(0,1) \right| \to 0.$$

with set of limit points given by $\{\mathbf{H}^\ell(0,1) : H \in \mathcal{K}\}$*.*

We have, as before,

Corollary 7.13. *Let* $A \in (\mathbb{R}^e)^{\otimes \ell}$ *and define the real-valued tensor contraction* $\mathcal{X}_N^\ell := \left\langle A, \bar{\boldsymbol{\xi}}^\ell(0,N) \right\rangle$*. Then, with* M *given in* (7.4)*,*

$$P\left(\limsup_{N\to\infty} \frac{\mathcal{X}_N^\ell}{(2N\log\log N)^{\frac{\ell}{2}}} = M/\bar{\tau} \right) = 1.$$

7.4 Remarks on iterated sums and integrals

Iterated integrals and sums of the type considered in Corollaries 7.10 and 7.13 are of interest in data science. Specifically, iterated integrals have given rise to a popular feature set of machine learning applications, the lectures notes [15] constitute an excellent source of information. Iterated sums, a.k.a. *iterated-sums signatures* are a natural variant, specifically for feature extraction of time-series, see e.g. [5, 21, 16]. Strictly speaking, they allow for additional integer powers of the ξ's. An extension of Corollary 7.10 in this direction is not difficult, e.g. using results of [25], but this would require an algebraic setup in terms of quasi-shuffle or Grossmann–Larson Hopf algebras that would lead us too far astray.

References

[1] I. Bailleul, R. Catellier, *Rough flows and homogenization in stochastic turbulence*, Journal of Differential Equations 263(8) (2017), 4894–4928. MR3680942

[2] P. Baldi, *Large deviations and functional iterated logarithm law for diffusion processes*, Prob. Theory Rel. Fields 71 (1986), 435–453. MR0824713

[3] P. Baldi, G. Ben Arous, G. Kerkyacharian, *Large deviations and the Strassen theorem in Hölder norm*, Stoch. Proc. Appl. 42(1) (1992), 171–180. MR1172514

[4] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., J. Willey, New York, 1999. MR1700749

[5] P. Bonnier, P. Kidger, I. Perez Arribas, C. Salvi, T. Lyons, *Deep signature transforms*, in: Adv. Neur. Info. Proc. Syst. 32 (NIPS 2019), 2019.

[6] A.N. Borodin, *A limit theorem for solutions of differential equations with random right-hand side*, Theory Probab. Appl. 22 (1977), 482–497. MR0517995

[7] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math., vol. 470, Springer-Verlag, Berlin, 1975. MR0442989

[8] R.C. Bradley, *Introduction to Strong Mixing Conditions*, Kendrick Press, Heber City, 2007. MR2325295

[9] I. Berkes, W. Philipp, *Approximation theorems for independent and weakly dependent random vectors*, Annals Probab. 7 (1979), 29–54. MR0515811

[10] R. Bowen, D. Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. 29 (1975), 181–202. MR0380889

[11] Y. Bruned, I. Chevyrev, P.K. Friz, R. Preiß, *A rough path perspective on renormalization*, J. Funct. Anal. 277(11) (2019), 108283. MR4013830

[12] I. Chevyrev, P.K. Friz, *Canonical RDEs and general semimartingales as rough paths*, Ann. Probab. 47 (2019), 420–463. MR3909973

[13] I. Chevyrev, P.K. Friz, A. Korepanov, I. Melbourne, H. Zhang, *Deterministic homogenization under optimal moment assumptions for fast-slow systems. Part 2*, Ann. l'Inst. H. Poincaré, Prob. Stat. 58 (2022), 1328–1350. MR4452636

[14] I. Chevyrev, P.K. Friz, A. Korepanov, I. Melbourne, H. Zhang, *Multiscale systems, homogenization, and rough paths*, in: Friz P., König W., Mukherjee C., Olla S. (eds.) Probability and Analysis in Interacting Physical Systems. VAR75 2016. Springer Proceedings in Mathematics & Statistics, vol. 283, Springer, 2019. MR3968507

[15] I. Chevyrev, A. Kormilitzin, *A Primer on the Signature Method in Machine Learning*, arXiv:1603.03788.

[16] C. Cuchiero, L. Gonon, L. Grigoryeva, J.P. Ortega, J. Teichmann, *Discrete-time signatures and randomness in reservoir computing*, IEEE Transactions on Neural Networks and Learning Systems, 2021. MR4506280

[17] H. Dehling, W. Philipp, *Empirical process technique for dependent data*, in: H.G. Dehling, T. Mikosch, M. Sorenson (eds.), *Empirical Process Technique for Dependent Data*, pp. 3–113, Birkhäuser, Boston, 2002. MR1958777

[18] M. Denker, W. Philipp, *Approximation by Brownian motion for Gibbs measures and flows under a function*, Ergod. Th. & Dynam. Sys. 4 (1984), 541–552. MR0779712

[19] J.-D. Deuschel, D. Stroock, *Large Deviations*, vol. 342. American Mathematical Soc., 2001. MR0997938

[20] J.-D. Deuschel, T. Orenshtein, N. Perkowski, *Additive functionals as rough paths*, Ann. Probab. 49(3) (2021), 1450–1479. MR4255150

[21] J. Diehl, K. Ebrahimi-Fard, N. Tapia, *Time-warping invariants of multidimensional time series*, Acta Appl. Math. 170 (2020), 265–290. MR4163237

[22] P.K. Friz, M. Hairer, *A Course on Rough Paths: With an Introduction to Regularity Structures*, 2nd edition, Universitext Springer, 2020. MR4174393

[23] P.K. Friz, A. Shekhar, *General rough integration, Lévy rough paths and a Lévy–Kintchine-type formula*, Ann. Probab. 45(4) (2017), 12707–2765. MR3693973

[24] P.K. Friz, N.B. Victoir, *Multidimensional Stochastic Processes as Rough Paths*, Cambridge Studies in Advanced Mathematics, vol. 120, Cambridge Univ. Press, 2010. MR2604669

[25] P.K. Friz, H. Zhang, *Differential equations driven by rough paths with jumps*, J. Diff. Equat. 264 (2018), 6226–6301. MR3770049

[26] P.K. Friz, P. Zorin-Kranich, *Rough semimartingales and p-variation estimates for martingale transforms*, Ann. Probab. 51(2) (2023), 397–441. MR4546622

[27] S. Gouëzel, *Almost sure invariance principle for dynamical systems by spectral methods*, Ann. Probab. 38 (2010), 1619–1671.

[28] L. Heinrich, *Mixing properties and central limit theorem for a class of non-identical piecewise monotonic C^2-transformations*, Mathematische Nachricht. 181 (1996), 185–214. MR1409076

[29] Ye. Hafouta, Yu. Kifer, *Nonconventional Limit Theorems and Random Dynamics*, World Scientific, Singapore, 2018. MR3793181

[30] R.Z. Khasminskii, *A limit theorem for the solution of differential equations with random right-hand sides*, Theory Probab. Appl. 11 (1966), 390–406. MR0203789

[31] Yu. Kifer, *Strong diffusion approximation in averaging and value computation in Dynkin's games*, Ann. Appl. Probab. 34(1A) (2024), 103–147. arXiv:2011.07907. MR4696274

[32] Yu. Kifer, *Strong diffusion approximation in averaging with dynamical systems fast motions*, Israel J. Math. 251 (2022), 595–634. MR4527552

[33] D. Kelly, I. Melbourne, *Smooth approximation of stochastic differential equations*, Ann. Probab. 44 (2016), 479–520. MR3456344

[34] D. Kelly, I. Melbourne, *Deterministic homogenization for fast–slow systems with chaotic noise*, Journal of Functional Analysis 272(10) (2017), 4063–4102. MR3626033

[35] Yu. Kifer, S.R.S. Varadhan, *Nonconventional limit theorems in discrete and continuous time via martingales*, Ann. Probab. 42 (2014), 649–688. MR3178470

[36] J. Kuelbs, W. Philipp, *Almost sure invariance principles for partial sums of mixing B-valued random variables*, Annals Probab. 8 (1980), 1003–1036. MR0602377

[37] M. Ledoux, Z. Qian, T. Zhang, *Large deviations and support theorem for diffusion processes via rough paths*, Stoch. Proc. Appl. 102(2) (2002) 265–283. MR1935127

[38] D. Lepingle, *La variation d'ordre p des semi-martingales*, Z. Wahrsch. Verw. Gebiete 36(4) (1976), 295–316. MR0420837

[39] T. Lyons, *Differential equations driven by rough signals*, Revista Mat. Iberoamericana 14(2) (1998), 215–310. MR1654527

[40] X. Mao, *Stochastic Differential Equations and Applications*, 2nd edition, Woodhead, Oxford, 2010.

[41] I. Melbourne, M. Nicol, *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*, Ann. Probab. 37 (2009), 478–505. MR2510014

[42] D. Monrad, W. Philipp, *Nearby variables with nearby laws and a strong approximation theorem for Hilbert space valued martingales*, Probab. Th. Rel. Fields 88 (1991), 381–404. MR1100898

[43] G.C. Papanicolaou, W. Kohler, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math. 27 (1974), 641–668. MR0368142

[44] R.J. Serfling, *Moment inequalities for the maximum cumulative sum*, Ann. Math. Stat. 41 (1970), 1227–1234. MR0268938