MULTISCALE SYSTEMS, HOMOGENIZATION, AND ROUGH PATHS

ILYA CHEVYREV, PETER K. FRIZ, ALEXEY KOREPANOV, IAN MELBOURNE, AND HUILIN ZHANG

Dedicated to Professor S.R.S Varadhan on the occasion of his 75th birthday

Abstract. In recent years, substantial progress was made towards understanding convergence of fast-slow deterministic systems to stochastic differential equations. In contrast to more classical approaches, the assumptions on the fast flow are very mild. We survey the origins of this theory and then revisit and improve the analysis of Kelly-Melbourne [Ann. Probab. Volume 44, Number 1 (2016), 479-520], taking into account recent progress in p-variation and càdlàg rough path theory.

Contents

Key words and phrases. Fast-slow systems, homogenization, rough paths.

1. Introduction

The purpose of this article is to survey and improve several recent developments in the theories of homogenization and rough paths, and the interaction between them. From the side of homogenization, we are interested in the programme initiated by [57] and continued in [33] of studying fast-slow systems without mixing assumptions on the fast flow. From the side of rough paths, we are interested in surveying recent extensions of the theory to the discontinuous setting [28, 18, 31] (see also [17, 21, 40, 74] for related results); the continuous theory, for the purposes of this survey, is well-understood [30]. The connection between the two sides first arose in [41, 42] in which the authors were able to employ rough path techniques (in the continuous and discontinuous setting) to study systems widely generalising those considered in [57, 33].

In this article we address both continuous and discrete systems. The continuous fast-slow systems take the form of the ODEs

$$\frac{d}{dt}x_\varepsilon = a(x_\varepsilon, y_\varepsilon) + \varepsilon^{-1}b(x_\varepsilon, y_\varepsilon)\ , \qquad \frac{d}{dt}y_\varepsilon = \varepsilon^{-2}g(y_\varepsilon)\ . \tag{1}$$

The equations are posed on $\mathbb{R}^d \times M$ for some compact Riemannian manifold M, and $g : M \to TM$ is a suitable vector field. We assume a fixed initial condition $x_\varepsilon(0) = \xi$ for some fixed $\xi \in \mathbb{R}^d$, while the initial condition for y_ε is drawn randomly from (M, λ), where λ is a Borel probability measure on M.

For the discrete systems, we are interested in dynamics of the form

$$X^{(n)}_{j+1} = X^{(n)}_j + n^{-1}a(X^{(n)}_j, Y_j) + n^{-1/2}b(X^{(n)}_j, Y_j)\ , \quad Y_{j+1} = TY_j\ . \tag{2}$$

The equations are again posed on $\mathbb{R}^d \times M$, and $T : M \to M$ is an appropriate transformation. As before, $X^{(n)}_0 = \xi \in \mathbb{R}^d$ is fixed and Y_0 is drawn randomly from a probability measure λ on M.

Let $x_\varepsilon : [0, 1] \to \mathbb{R}^d$ denote either the solution to (1), or the piecewise constant path $x_\varepsilon(t) = X^{(\lfloor 1/\varepsilon^2 \rfloor)}_{\lceil t/\varepsilon^2 \rceil - 1}$, where $X^{(n)}_j$ is the solution to (2). The primary goal of this article is to show convergence in law $x_\varepsilon \to X$ in the uniform (or stronger) topology as $\varepsilon \to 0$. Here X is a stochastic process, which in our situation will be the solution to an SDE.

Throughout this note we shall focus on the case where $a(x, y) \equiv a(x)$ depends only on x and $b(x, y) \equiv b(x)v(y)$, where $v : M \to \mathbb{R}^m$ is an observable of y and $b : \mathbb{R}^d \to L(\mathbb{R}^m, \mathbb{R}^d)$ This is precisely the situation considered in [41]. One restriction of the method in [41] is the use of Hölder rough path topology which necessitates moment conditions on the fast dynamics which are suboptimal from the point of view of homogenization. Our main insight is that switching from α-Hölder to p-variation rough path topology allows for optimal moment assumptions on the fast dynamics. The non-product case was previously handled, also with suboptimal moment assumptions, in [42] and also [7], using infinite-dimensional and flow-based rough paths respectively. We briefly discuss this and some other extensions in Section 5, leaving a full analysis of the general (non-product) case, under equally optimal moment assumptions, to a forthcoming artice [19].

An example of fast dynamics. There are many examples to which the results presented here apply, however we feel it is important to have a concrete (and simple to state) example in mind from the very beginning. In this regard, Pomeau & Manneville [64] introduced a class of maps that exhibit *intermittency* as part of their study of turbulent bursts. The most-studied example [49] is the one-dimensional map $T : M \to M$, $M = [0, 1]$, given by

$$Ty = \begin{cases} y(1 + 2^\gamma y^\gamma) & y < \frac{1}{2} \\ 2y - 1 & y \geq \frac{1}{2} \end{cases} . \tag{3}$$

Here $\gamma \geq 0$ is a parameter. When $\gamma = 0$ this is the doubling map $Ty = 2y \bmod 1$ which is uniformly expanding (see Section 2.3). For $\gamma > 0$, there is a neutral fixed point at 0 ($T'(0) = 1$) which has more and more influence as γ increases. For each value of $\gamma \in [0, 1)$, there is a unique absolutely continuous invariant probability measure μ. This measure is ergodic and equivalent (in fact equal when $\gamma = 0$) to the Lebesgue measure.

Suppose that $v\colon M \to \mathbb{R}^m$ is Hölder continuous and $\int v \, d\mu = 0$. Let

$$v_n = \sum_{0 \leq j < n} v \circ T^j.$$

By [49, 76], for $\gamma \in [0, \frac{1}{2})$, the random variable $n^{-1/2} v_n$, defined on the probability space (M, μ), converges in law to a normal distribution. (Convergence in law also holds on (M, Leb).) In other words, the *central limit theorem* (CLT) holds. However, by [35], the CLT fails for $\gamma > \frac{1}{2}$ (instead there is convergence to a stable law of index γ^{-1}); for $\gamma = \frac{1}{2}$ the CLT holds but with non-standard normalization $(n \log n)^{-1/2}$. Hence from now on we restrict to $\gamma \in [0, \frac{1}{2})$.

Define

$$S_n = \sum_{0 \leq i \leq j < n} (v \circ T^i) \otimes (v \circ T^j) .$$

The approach to homogenization of fast-slow systems in [41] requires convergence of the pair of stochastic processes $\left(n^{-1/2} v_{\lfloor nt \rfloor}, n^{-1} S_{\lfloor nt \rfloor}\right)$ to an enhanced Brownian motion, which is established for all $\gamma \in [0, \frac{1}{2})$. (See Sections 2.3 and 4.1.) Further, the approach based on Hölder rough path theory requires that $\|v_n\|_{2q} = O(n^{1/2})$ and $\|S_n\|_q = O(n)$ for some $q > 3$. These estimates are established in [41] for $\gamma \in [0, \frac{2}{11})$. An improvement in [45] covers $\gamma \in [0, \frac{1}{4})$ and this is known to be sharp [56, 53]. Hence the parameter regime $\gamma \in [\frac{1}{4}, \frac{1}{2})$ is beyond the Hölder rough path theory. In contrast, the p-variation rough path theory described here requires the moment estimates only for some $q > 1$ and [45] applies for all $\gamma \in [0, \frac{1}{2})$. Hence we are able to prove homogenization theorems in the full range $\gamma \in [0, \frac{1}{2})$.

The remainder of this article is organized as follows. In Section 2, we discuss the WIP and chaotic dynamics, and several situations of homogenization where rough path theory is not required. In Section 3, we introduce the parts of rough path theory required in the Brownian motion setting of this paper. This is applied to fast-slow systems in Section 4. In Section 5, we mention extensions and related work.

Acknowledgements: I.C. is funded by a Junior Research Fellowship of St John's College, Oxford. P.K.F. acknowledges partial support from the ERC, CoG-683164, and DFG research unit FOR2402. A.K. and I.M. acknowledge partial support from the European Advanced Grant StochExtHomog (ERC AdG 320977). H.Z. thanks the Institut für Mathematik, TU Berlin, for its hospitality.

2. Emergence of randomness in deterministic dynamical systems

In this section, we review a simplified situation where the ordinary weak invariance principle (see below) suffices, and rough path theory is not required.

2.1. **The weak invariance principle.**

Consider a family of stochastic processes indexed by $\varepsilon \in (0,1)$, say $W_\varepsilon = W_\varepsilon(t,\omega)$ with values in $\mathbb{R}^m$. We are interested in convergence of the respective laws. In the case of continuous sample paths (including smooth or piecewise linear) we say that the *weak invariance principle* (WIP) holds if

$$W_\varepsilon \to_w W \text{ in } C([0,1],\mathbb{R}^m) \text{ as } \varepsilon \to 0 \ ,$$

where W is an m-dimension Brownian motion with covariance matrix Σ; in the case of càdlàg sample paths (including piecewise constant) we mean

$$W_\varepsilon \to_w W \text{ in } D([0,1],\mathbb{R}^m) \text{ as } \varepsilon \to 0 \ ,$$

where C resp. D denotes the space of continuous resp. càdlàg paths, equipped with the uniform topology.[1] For notational simplicity only, assume (W_ε) are defined on a common probability space (Ω, F, λ); we then write $W_\varepsilon \to_\lambda W$ to indicate convergence in law, i.e. $\mathbb{E}_\lambda[f(W_\varepsilon)] \to E_\lambda[f(W)]$ for all bounded continuous functionals.

In many cases, one has convergence of second moments. This allows to compute the covariance of the limiting Brownian motion,

$$\Sigma = \mathbb{E}(W(1)\otimes W(1)) = \lim_{\varepsilon\to 0} \mathbb{E}_\lambda(W_\varepsilon(1)\otimes W_\varepsilon(1)) \ . \tag{4}$$

The WIP is also known as the *functional central limit theorem*, with the CLT for finite-dimensional distributions as a trivial consequence. Conversely, the CLT for f.d.d. together with tightness gives the WIP.

Donsker's invariance principle [23] is the prototype of a WIP: consider a centered m-dimensional random walk $Z_n := \xi_1 + \cdots + \xi_n$, with $\mathbb{R}^m$-valued IID increments of zero mean and finite covariance Σ. Extend to a piecewise linear resp. (càdlàg) piecewise constant process $(Z_t : t \geq 0)$. Then the WIP holds for the rescaled random walk

$$Z^\varepsilon_t := \varepsilon Z_{t/\varepsilon^2} \ ,$$

and the limiting Brownian motion has covariance Σ. This result has an important generalization to a *(functional) martingale CLT*: using similar notation, assume (Z_n) is a zero mean, L^2-martingale, with stationary and ergodic increments (ξ_i). Then, with the identical rescaling, the WIP holds true, with convergence of second moments [14] (or e.g. [10, Thm. 18.3]).

Another interesting example is given by *physical Brownian motion* with positive mass $\varepsilon^2 > 0$ and friction matrix M, where the trajectory is given by

$$X^\varepsilon_t := \varepsilon \int_0^{t/\varepsilon^2} Y_s\, ds$$

and Y follows an m-dimensional OU process, $dY = -MY dt + dB$, $Y_0 = y_0$. Here, M is an $m \times m$-matrix whose spectrum has positive real part, and B is an m-dimensional standard Brownian motion. One checks without difficulties [63, 26] that a WIP holds, even in the sense of weak convergence in the Hölder space $C^\alpha([0,1],\mathbb{R}^m)$ with any $\alpha < 1/2$. The covariance matrix of the limiting Brownian

[1]Since our limit processes here - a Brownian motion - is continuous, there is no need to work with the Skorokhod topology on D.

is given by $\Sigma = M^{-1}(M^{-1})^T$, as can be seen from the Newton dynamics $\varepsilon^2 \ddot{X}^\varepsilon = -M\dot{X}^\varepsilon + \dot{B}$ with white noise $\dot{B}$.

Finally, *sufficiently chaotic deterministic dynamical systems* are a rich source of WIPs. To fix ideas, consider a compact Riemannian manifold M with a Lipschitz vector field g and corresponding flow g_t, for which there is an ergodic, invariant Borel probability measure μ on M. We regard (g_t) as an M-valued stochastic process, given by $g_t(y_0)$ with initial condition y_0 distributed according to λ, another Borel probability measure on M. (It is possible but not necessary to have $\lambda = \mu$.) Consider further a suitable observable $v : M \to \mathbb{R}^m$ with $\mathbb{E}_\mu v = 0$. A family of C^1-processes $(W_\varepsilon)_{\varepsilon>0}$, with values in $\mathbb{R}^m$, is then given by

$$W_\varepsilon(t) = \varepsilon \int_0^{t\varepsilon^{-2}} v \circ g_s \, ds \ .$$

As will be reviewed in Section 2.3 below, also in a discrete time setting, in many situations a WIP holds. That is,

$$W_\varepsilon \to_\lambda W \text{ in } C([0,1], \mathbb{R}^m) \text{ as } \varepsilon \to 0 \ .$$

Typically one also has convergence of second moments, so that W is a Brownian motion with covariance Σ given by (4). Under (somewhat restrictive) assumptions on the decay of correlations, this can be simplified to a Green-Kubo type formula

$$\Sigma = \int_0^\infty \mathbb{E}_\mu \{ v \otimes (v \circ g_s) + (v \circ g_s) \otimes v \} ds \ .$$

2.2. **First applications to fast-slow systems.** In the setting of deterministic, sufficiently chaotic dynamical systems discussed in the previous paragraph, Melbourne–Stuart [57] consider the fast-slow system posed on $\mathbb{R}^d \times M$ (with $m = d$),

$$\dot{x}_\varepsilon = a(x_\varepsilon, y_\varepsilon) + \varepsilon^{-1} v(y_\varepsilon) \ , \qquad \dot{y}_\varepsilon = \varepsilon^{-2} g(y_\varepsilon) \ ,$$

with deterministic initial data $x_\varepsilon(0) = x_0$ and $y_\varepsilon(0)$ sampled randomly with probability λ. We wish to study the limiting dynamics of the slow variable x_ε. Assuming for simplicity $a(x, y) = a(x)$, the basic observation is to rewrite

$$\dot{x}_\varepsilon = a(x_\varepsilon) + \dot{W}_\varepsilon \ .$$

We see that the noise W_ε enters the equation in an additive fashion and one checks without difficulty that the "Itô-map" $W_\varepsilon \mapsto x_\varepsilon$ extends continuously (w.r.t. uniform convergence) to any continuous noise path. Now assume validity of a WIP, i.e. $W_\varepsilon \to_\lambda W$. Then, together with continuity of the Itô-map, one obtains the desired limiting SDE dynamics of the slow variable as

$$dX = dW + a(X) dt \ .$$

In the general case when a depends on x_ε and y_ε, the drift term is given by $\bar{a}(x) = \int_M a(x, y) \, d\mu(y)$

In subsequent work, Gottwald–Melbourne [33] consider the one-dimensional case $d = m = 1$ with

$$\dot{x}_\varepsilon = a(x_\varepsilon, y_\varepsilon) + \varepsilon^{-1} b(x_\varepsilon) v(y_\varepsilon) \ , \qquad \dot{y}_\varepsilon = \varepsilon^{-2} g(y_\varepsilon) \ .$$

Again, taking $a(x, y) = a(x)$ for simplicity, the limiting SDE turns out to be of Stratonovich form

$$dX = a(X) dt + b(X) \circ dW \ .$$

The essence of the proof is a robust representation of such SDEs. Indeed, taking $a \equiv 0$ for notational simplicity, an application of the (first order) Stratonovich chain rule exhibits the explicit solution as $X_t = e^{W_t b}(X_0)$, where $e^{W_t b}$ denotes the flow at "time" $W_t \in \mathbb{R}$ along the vector field b; this clearly depends continuously on X_0 and W w.r.t. uniform convergence. Hence, as in the additive case, the problem is reduced to having a WIP. This line of reasoning can be pushed a little further, namely to the case $\dot{x}_\varepsilon = a(x_\varepsilon, y_\varepsilon) + \varepsilon^{-1} V(x_\varepsilon) v(y_\varepsilon)$ with commuting vector fields $V = (V_1, ..., V_m)$, a.k.a. the Doss–Sussmann method, but fails for general vector fields, not to mention the non-product case when $V(x)v(y)$ is replaced by $b(x, y)$. This is a fundamental problem which is addressed by Lyons' theory of rough paths.

Gottwald-Melbourne [33] consider also discrete time fast-slow systems posed on $\mathbb{R}^d \times M$,

$$X_{j+1}^{(n)} = X_j^{(n)} + n^{-1} a(X_j^{(n)}, Y_j) + n^{-1/2} b(X_j^{(n)}) v(Y_j) \ , \quad Y_{j+1} = TY_j \ ,$$

Again we suppose for notational simplicity that $a(x, y) = a(x)$. We continue to suppose that μ is an ergodic T-invariant probability measure on M and that $\mathbb{E}_\mu v = 0$. Also, λ is another probability measure on M. Recall that $x_\varepsilon(t) = X_{[t/\varepsilon]-1}^{[1/\varepsilon]}$ and assume validity of a WIP, i.e. $W_\varepsilon \to_\lambda W$. When $b \equiv 1$, it is shown in [33] that $x_\varepsilon \to_\lambda X$ where $dX = a(X)\,dt + dW$. For $d = m = 1$, under additional mixing assumptions it is shown that $x_\varepsilon \to_\lambda X$ where $dX = \tilde{a}(X)\,dt + b(X)\,dW$ with

$$\tilde{a}(x) = a(x) + b(x)b'(x) \sum_{n=1}^{\infty} \mathbb{E}_\mu(v \; v \circ T^n).$$

2.3. **Chaotic dynamics: CLT and the WIP.** In this subsection, we describe various classes of dynamical systems with good statistical properties, focusing attention on the CLT and WIP.

Somewhat in contrast to rough path theory, the ergodic theory of smooth dynamical systems is much simpler for discrete time than for continuous time – indeed the continuous time theory proceeds by reducing to the discrete time case. Also, the simplest examples are noninvertible. The reasons behind this are roughly as follows. Since the papers of Anosov [4] and Smale [72], it is has been understood that the way to study dynamical systems is to exploit expansion and contraction properties. The simplest systems are uniformly expanding; these are necessarily discrete time and noninvertible. Anosov and Axiom A (uniformly hyperbolic) diffeomorphisms have uniformly contracting and expanding directions. Anosov and Axiom A (uniformly hyperbolic) flows have a neutral time direction and are uniformly contracting and expanding in the remaining directions. The neutral direction makes flows much harder to study. The mixing properties of uniformly hyperbolic flows are still poorly understood (see for example the review in [54]); fortunately the CLT and WIP do not rely on mixing.

Accordingly, we consider in turn expanding maps, hyperbolic diffeomorphisms, and hyperbolic flows, in Subsections 2.3.1, 2.3.2 and 2.3.3 respectively. This includes the uniform cases mentioned in the previous paragraph, but also dynamical systems that are nonuniformly expanding/hyperbolic, which is crucial for incorporating large classes of examples.

2.3.1. *Uniformly and nonuniformly expanding maps.* The CLT and WIP are proved in [37, 39] for large classes of dynamical systems (in fact, they prove a stronger statistical property, known as the almost sure invariance principle). For recent developments in this direction, see [20, 44] and references therein. In particular, the CLT and WIP hold for smooth uniformly expanding maps and for systems modelled by Young towers with summable decay of correlations [76], which provide a rich source of examples including the intermittent maps (3).

Here we review the results in various situations, focusing on various issues that are of importance for fast-slow systems: CLT, WIP, covariance matrices, nondegeneracy. Also, we mention the notions of spectral decomposition, mixing up to a finite cycle, basins of attraction, and strong distributional convergence, which are necessary for understanding how the theory is applied.

Smooth uniformly expanding maps. The simplest chaotic dynamical system is the doubling map $T : M \to M$, $M = [0, 1]$, given by $Ty = 2y \bmod 1$. More generally, let $T : M \to M$ be a C^2 map on a compact Riemannian manifold M and let $\mathcal{B}$ be the σ-algebra of Borel sets. The map is *uniformly expanding* if there are constants $C > 0$, $L > 1$ such that $\|DT^n|_y z\| \geq CL^n\|z\|$ for all $y \in M$, $z \in T_yM$. By [46], there is a unique ergodic T-invariant Borel probability measure μ on M equivalent to the volume measure. (Recall that μ is T-invariant if $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$, and is ergodic if $\mu(B) = 0$ or $\mu(B) = 1$ for all $B \in \mathcal{B}$ with $TB \subset B$.) By Birkhoff's ergodic theorem [11] (an extension of the strong law of large numbers) the sum $v_n = \sum_{0\leq j<n} v \circ T^j$ satisfies $n^{-1}v_n \to \int_M v\, d\mu$ a.e. for all $v \in L^1$.

To make further progress it is necessary to impose some regularity on the observable v; the CLT fails in general for continuous observables. Hence, we suppose that v is Hölder. Specifically, fix $\kappa \in (0, 1)$ and let $C_0^\kappa(M)$ be the space of C^κ observables $v : M \to \mathbb{R}$ with $\int_M v\, d\mu = 0$. It is well known [12, 67, 71] that there are constants $\gamma \in (0, 1)$ and $C > 0$ depending only on T and κ such that

$$(5) \qquad \Big| \int_M v\, w \circ T^n \, d\mu \Big| \leq C\gamma^n \|v\|_{C^\kappa} \|w\|_1 \quad \text{for all } v \in C_0^\kappa(M),\ w \in L^1,\ n \geq 1.$$

An immediate consequence is that the limit

$$\sigma^2 := \lim_{n\to\infty} n^{-1} \int_M v_n^2 \, d\mu ,$$

exists and that

$$\sigma^2 = \int_M v^2 \, d\mu + 2 \sum_{n=1}^{\infty} \int_M v\, v \circ T^n \, d\mu .$$

For $1 \leq p \leq \infty$, we recall that the Koopman operator $U : L^p \to L^p$ is given by $Uv = v \circ T$ and that the transfer operator $P : L^q \to L^q$ is given by $\int_M Pv\; w\, d\mu = \int_M v\; w \circ T\, d\mu$ for $v \in L^q$, $w \in L^p$, where $p^{-1} + q^{-1} = 1$. These operators satisfy $\|U\|_p = 1$ and $\|P\|_p \leq 1$ for $1 \leq p \leq \infty$. In addition, $PU = I$ and $UP = \mathbb{E}(\,\cdot\,|T^{-1}\mathcal{B})$. Note that property (5) is equivalent to

$$(6) \qquad \|P^n v\|_\infty \leq C\gamma^n \|v\|_{C^\kappa} \quad \text{for all } v \in C_0^\kappa(M),\ n \geq 1.$$

Following the classical approach of Gordin [32], we define $\chi = \sum_{j=1}^{\infty} P^j v$ and $m = v - \chi \circ T + \chi$. By (6), $m, \chi \in L^\infty$. It follows from the definitions that $m \in \ker P$ and hence that $n^{-1} \int_M m_n^2 \, d\mu = \int_M m^2 \, d\mu$ for all n. Since

$$v_n - m_n = \chi \circ T^n - \chi \in L^\infty ,$$

it follows that $\int_M m^2 \, d\mu = \sigma^2$.

Moreover, $\mathbb{E}(m|T^{-1}\mathcal{B}) = UPm = 0$, so $\{m \circ T^n,\, n \ge 0\}$ is an L^∞ stationary ergodic reverse martingale difference sequence. Hence, standard martingale limit theorems apply. In particular, by [9, 52] we obtain the CLT for m and thereby v:

$$n^{-1/2}v_n \to_\mu N(0,\sigma)^2 \quad \text{as } n \to \infty.$$

We refer to the decomposition $v = m + \chi \circ T - \chi$ as an L^∞ *martingale-coboundary decomposition*, since m is a reverse martingale increment. The coboundary term $\chi \circ T - \chi \in L^\infty$ telescopes under iteration and therefore is often negligible. Next, define the process $W_n \in C[0,1]$ by setting $V_n(t) = n^{-1/2}v_{nt}$ for $t = 0, 1/n, 2/n \dots$ and linearly interpolating. By [14, 52] we obtain the WIP:

$$W_n \to_\mu W \quad \text{in } C[0,1] \text{ as } n \to \infty,$$

where W is a Brownian motion with variance σ^2.

The CLT and WIP are said to be *degenerate* if $\sigma^2 = 0$. We now show that this is extremely rare. Since $v = m + \chi \circ T - \chi$ and $\sigma^2 = \int_M m^2\, d\mu$, we obtain that $\sigma^2 = 0$ if and only if $v = \chi \circ T - \chi$ where $\chi \in L^\infty$. Moreover, the series $\chi = \sum_{n=1}^\infty P^n v$ converges in C^κ (see for example [68]), so in particular χ is continuous. Let $C^\kappa_{\rm deg}$ consist of observables $v \in C^\kappa_0(M)$ with $\sigma^2 = 0$.

Proposition 2.1. *$C^\kappa_{\rm deg}$ is a closed, linear subspace of infinite codimension in $C^\kappa_0(M)$.*

Proof. Suppose that $v \in C^\kappa_{\rm deg}$ so $v = \chi \circ T - \chi$ where χ is continuous. Iterating, we obtain $v_k = \chi \circ T^k - \chi$. Hence if $y \in M$ is a period k point, i.e. $T^k y = y$, then $v_k(y) = 0$. Since periodic points are dense in M [72] we obtain infinitely many linear constraints on v. □

Analogous results hold for vector-valued observables $v : M \to \mathbb{R}^m$. Let $v \in C^\kappa_0(M, \mathbb{R}^m)$. Then

$$\lim_{n\to\infty} n^{-1} \int_M v_n \otimes v_n \, d\mu = \Sigma, \tag{7}$$

where $\Sigma \in \mathbb{R}^m \otimes \mathbb{R}^m$ is symmetric and positive semidefinite, and

$$\Sigma = \int_M v \otimes v\, d\mu + \sum_{n=1}^\infty \int_M \{v \otimes (v \circ T^n) + (v \circ T^n) \otimes v\}\, d\mu. \tag{8}$$

Define $v_n \in \mathbb{R}^m$ and $W_n \in C([0,1], \mathbb{R}^m)$ as before. By the above results, $n^{-1/2}c^T v_n$ converges in distribution to a normal distribution with variance $c^T \Sigma c$ for each $c \in \mathbb{R}^m$, and hence by Cramer-Wold we obtain the multi-dimensional CLT $n^{-1/2}v_n \to_d N(0, \Sigma)$. Similarly, $W_n \to_w W$ in $C([0,1], \mathbb{R}^m)$ where W is m-dimensional Brownian motion with covariance Σ. Finally $c^T\Sigma c = 0$ for $c \in \mathbb{R}^m$ if and only if $c^T v \in C^\kappa_{\rm deg}$. Hence the degenerate case $\det \Sigma = 0$ occurs only on a closed subspace of infinite codimension.

Since the CLT is a consequence of the WIP, generally we only mention the WIP in the remainder of this subsection.

Returning to the ergodic theorem, if $v \in L^1$, then $n^{-1}v_n(y_0) \to \int_M v\, d\mu$ for μ almost every initial condition $y_0 \in M$. Since μ is equivalent to volume, we could equally choose the initial condition y_0 randomly with respect to volume, which is perhaps more natural since volume is the intrinsic measure on M. Similar considerations apply to the WIP. Based on ideas of [24], it follows from [79, Cor. 2] that if $v \in C_0^\kappa(M,\mathbb{R}^m)$ then $W_n \to_\lambda W$ in $C([0,1],\mathbb{R}^m)$ for every absolutely continuous Borel probability measure λ (including μ and volume as special cases). This property is often called *strong distributional convergence* [79]. Of course $C_0^\kappa(M,\mathbb{R}^m)$ is defined using μ regardless of the choice of λ.

Piecewise expanding maps. There are numerous extensions of the above arguments in various directions. For example, Keller [39, Thm. 3.5] considers piecewise $C^{1+\varepsilon}$ transformations $T : M \to M$, $M = [0,1]$, with finitely many monotone branches and $|T'| \ge L$ for some $L > 1$. There exists an ergodic T-invariant absolutely continuous probability measure (acip) μ. Let $\Lambda = \operatorname{supp}\mu$. Recall that Λ is *mixing* if $\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$ for all measurable sets $A, B \subset \Lambda$. In this case, by [39, Thm. 3.3], condition (5) holds (with M replaced by Λ). Hence we obtain the WIP for all $v \in C_0^\kappa(\Lambda,\mathbb{R}^m)$ with Σ given as in (7) and (8). Also $\det\Sigma = 0$ if and only if there exists $c \in \mathbb{R}^m$ such that $c^T v = \chi \circ T - \chi$ for some $\chi : \Lambda \to \mathbb{R}$ in L^∞.

If Λ is not mixing, then condition (5) fails. Nevertheless, by [39, Thm. 3.3] Λ is *mixing up to a finite cycle*: we can write Λ as a disjoint union $\Lambda = A_1 \cup \cdots \cup A_k$ for some $k \ge 2$ such that T permutes the A_j cyclically and $T^k : A_j \to A_j$ is mixing with respect to $\mu|A_j$ for each j. Moreover, condition (5) holds for the map $T^k : A_j \to A_j$. It is easily verified that the WIP goes through for $T : \Lambda \to \Lambda$ and that the limit formula (7) for Σ remains valid. (Of course in the nonmixing case, (8) no longer makes sense.)

The *basin of attraction* of the ergodic probability measure μ is defined as

$$B_\mu = \{y \in M : \lim_{n\to\infty} n^{-1}v_n(y) = \textstyle\int_M v\,d\mu \quad \text{for all } v : M \to \mathbb{R} \text{ continuous}\}.$$

(The ergodic theorem guarantees that, modulo a zero measure set, $\operatorname{supp}\mu \subset B_\mu$, but in general B_μ can be much larger.) The acip μ need not be unique but by [39, Thm. 3.3] there is *a spectral decomposition*: there exist finitely many absolutely continuous ergodic invariant probability measures $\mu_1, \dots, \mu_k$ such that $\mathrm{Leb}(B_{\mu_1} \cup \cdots \cup B_{\mu_k}) = 1$ and the results described above for μ hold separately for each of $\mu_1, \dots, \mu_k$.

For related results on C^2 one-dimensional maps with infinitely many branches, we refer to [69]. For higher-dimensional piecewise smooth maps, see for example [15, 70]. Again there is a spectral decomposition into finitely many attractors which are mixing up to a finite cycle. After restricting to an appropriate subset and considering a suitable iterate of T, condition (5) holds and we obtain the WIP etc as described above.

In general, extra work is required to deduce that degeneracy is infinite codimension as in Proposition 2.1. We note that the approach in [15] fits within the Young tower approach of [75, 76] where it is possible to recover Proposition 2.1 as described below.

Nonuniformly expanding maps. An important method for studying nonuniformly expanding maps $T : M \to M$ is to construct a *Young tower* as in [76]. This incorporates the maps (3) discussed in the introduction.

Let M be a bounded metric space with finite Borel measure ρ and let $T : M \to M$ be a nonsingular transformation ($\rho(T^{-1}B) = 0$ if and only if $\rho(B) = 0$ for $B \in \mathcal{B}$). Let $Y \subset M$ be a subset of positive measure, and let α be an at most countable measurable partition of Y with $\rho(a) > 0$ for all $a \in \alpha$. We suppose that there is an integrable *return time* function $\tau : Y \to \mathbb{Z}^+$, constant on each a with value $\tau(a) \ge 1$, and constants $L > 1$, $\kappa \in (0, 1)$, $C_0 > 0$, such that for each $a \in \alpha$,

(1) $F = T^\tau$ restricts to a (measure-theoretic) bijection from a onto Y.
(2) $d(Fx, Fy) \ge Ld(x, y)$ for all $x, y \in a$.
(3) $d(T^\ell x, T^\ell y) \le C_0 d(Fx, Fy)$ for all $x, y \in a$, $0 \le \ell < \tau(a)$.
(4) $\zeta_0 = \frac{d\rho|_Y}{d\rho|_Y \circ F}$ satisfies $|\log \zeta_0(x) - \log \zeta_0(y)| \le C_0 d(Fx, Fy)^\kappa$ for all $x, y \in a$.

The *induced map* $F = T^\tau : Y \to Y$ has a unique acip μ_Y.

Remark 2.2. *For the intermittent maps* (3)*, we can take* $Y = [\frac{1}{2}, 1]$ *and we can choose* τ *to be the first return to* Y*. In general, it is not required that* τ *is the first return time to* Y*.*

Define the Young tower [76], $\Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \le \ell \le \tau(y) - 1\}$, and the tower map

$$f : \Delta \to \Delta, \qquad f(y, \ell) = \begin{cases} (y, \ell + 1), & \ell \le \tau(y) - 2 \\ (Fy, 0), & \ell = \tau(y) - 1 \end{cases}. \tag{9}$$

The projection $\pi_\Delta : \Delta \to \Lambda$, $\pi_\Delta(y, \ell) = T^\ell y$, defines a semiconjugacy from f to T. Define the ergodic acip $\mu_\Delta = \mu_Y \times \{\text{counting}\} / \int_Y \tau \, d\mu_Y$ for $f : \Delta \to \Delta$. Then $\mu = (\pi_\Delta)_* \mu_\Delta$ is an ergodic acip for $T : M \to M$ and μ is mixing up to a finite cycle.

Young [76] proved that if μ is mixing and $\mu_Y(y \in Y : \tau(y) > n) = O(n^{-(\beta+1)})$ for some $\beta > 0$, then

$$\Big| \int_M v \, w \circ T^n \, d\mu \Big| \le Cn^{-\beta} \|v\|_{C^\kappa} \|w\|_\infty \quad \text{for all } v \in C_0^\kappa(M),\ w \in L^\infty,\ n \ge 1, \tag{10}$$

In particular, $\beta > 1$ corresponds to *summable decay of correlations.* (For the maps (3), $\beta = \gamma^{-1} - 1$, so $\beta > 1$ corresponds to $\gamma < \frac{1}{2}$.) Equivalently, $\|P^n v\|_1 \le Cn^{-\beta} \|v\|_{C^\kappa}$ and by interpolation $\|P^n v\|_p \le C^{1/p} n^{-\beta/p} \|v\|_{C^\kappa}$ for all $p \ge 1$.

For $\beta > 1$, we have that $\|P^n v\|_1$ is summable for $v \in C_0^\kappa(M, \mathbb{R}^m)$, and a standard calculation shows that formulas (7) and (8) for Σ hold. Also, the series $\chi = \sum_{n=1}^\infty P^n v$ converges in L^p for all $p < \beta$ and we obtain an L^p martingale-coboundary decomposition $v = m + \chi \circ T - \chi$. For $\beta > 2$, we have $m, \chi \in L^2$ and the WIP follows. With extra work it can be shown that the WIP holds for all $\beta > 1$. We refer to [43, 48, 51, 73] for further details. See also [44, 55]. By [55, Rem. 2.11], the degenerate case $\det \Sigma = 0$ is infinite codimension in the sense of Proposition 2.1.

In the case where μ is mixing only up to a finite cycle, the WIP etc go through unchanged, except that formula (8) does not make sense.

2.3.2. *Hyperbolic diffeomorphisms.* A WIP for Axiom A diffeomorphisms can be found in [22]. The WIP is also well-known to hold for systems modelled by Young towers with exponential tails [75] as well as those with summable decay of correlations (for an explicit and completely general argument, see [60]). This is a very flexible setting that covers large classes of nonuniformly hyperbolic diffeomorphisms (with singularities).

The results for hyperbolic diffeomorphisms $T : M \to M$ are similar to those in Subsection 2.3.1, subject to two complications. The first complication affects the proofs. Since T is invertible, the transfer operator P is an isometry on L^q for all

q. In particular, $\ker P = \{0\}$. Hence the approach in Subsection 2.3.1 cannot be applied directly. The method for getting around this is rather convoluted and is described at the end of this subsection.

The second complication affects the statement of the results. Typically, the invariant measures of interest are supported on zero volume sets and hence there are no acips. We say that μ is a *physical measure* if the basin of attraction B_μ has positive volume. (This is automatic for acips but is an extra assumption now.) Let Vol denote the normalized volume on B_μ.

There is an important class of physical measures μ, known as Sinai-Ruelle-Bowen (SRB) measures [77], for which the WIP with respect to Vol (and hence, by strong distributional convergence, every absolutely continuous probability measure λ on B_μ) follows from the WIP with respect to μ. Hence it is natural to consider observables v with $\int_\Lambda v\,d\mu = 0$ and to ask that $W_n \to_\lambda W$ for absolutely continuous probability measures λ on B_μ.

Axiom A diffeomorphisms. Let M be a compact Riemannian manifold. A C^2 diffeomorphism $T : M \to M$ is said to be *Anosov* [4] if there is a continuous DT-invariant splitting $TM = E^s \oplus E^u$ (into stable and unstable directions) where $\|DT^n|E^s\| \le Ca^n$ and $\|DT^{-n}|E^u\| \le Ca^n$ for $n \ge 1$. Here $C > 0$ and $a \in (0,1)$ are constants.

Smale [72] introduced the notion of *Axiom A* diffeomorphism extending the definition in [4]. Since we are interested in SRB measures, we restrict attention to attracting sets, bypassing the full definitions in [72]. Recall that a closed T-invariant set $\Lambda \subset M$ is *attracting* if there is a neighbourhood U of Λ such that $\lim_{n\to\infty} \operatorname{dist}(T^n y, \Lambda) = 0$ for all $y \in U$. An attracting set is called *Axiom A* if there is a continuous DT-invariant splitting $T_\Lambda M = E^s \oplus E^u$ over Λ, again with the properties $\|DT^n|E^s\| \le Ca^n$ and $\|DT^{-n}|E^u\| \le Ca^n$ for $n \ge 1$. To avoid trivialities, we suppose that $\dim E^u_y \ge 1$ for all $y \in \Lambda$. (We allow $\dim E^s_y = 0$ though this is just the uniformly expanding case.)

By [72], there is a spectral decomposition of Λ into finitely many attracting sets, called *Axiom A attractors* with the property that none of them can be decomposed further. Moreover, periodic points are dense in Λ.

If Λ is an Axiom A attractor, then by [12, 67, 71] there is a unique ergodic invariant probability measure μ on Λ such that $\operatorname{Leb}(B_\mu) > 0$. Moreover, μ is mixing up to a finite cycle.

All the results described in Subsection 2.3.1 for uniformly expanding maps hold for Axiom A attractors. Specifically, let $C^\kappa_0(\Lambda, \mathbb{R}^m)$ denote the space of C^κ observables $v : \Lambda \to \mathbb{R}^m$ with $\int_\Lambda v\,d\mu = 0$. Then the WIP holds on (Λ, μ) and (B_μ, λ) with Σ satisfying formula (7). Moreover $C^\kappa_{\rm deg} = \{v \in C^\kappa_0(\Lambda, \mathbb{R}^m) : \det \Sigma = 0\}$ is a closed subspace of infinite codimension in $C^\kappa_0(\Lambda, \mathbb{R}^m)$. If in addition μ is mixing, then formula (8) holds.

Nonuniformly hyperbolic diffeomorphisms and Young towers. A large class of attractors Λ for nonuniformly hyperbolic diffeomorphisms (with singularities) $T : M \to M$ can be modelled by two-sided Young towers with exponential tails [75] and subexponential tails [76]. The Young tower set up covers numerous classes of examples as surveyed in [16, 75, 77, 78] including Axiom A attractors, Lorentz gases, Hénon-like attractors [8], and intermittent solenoids [60]. See also [1, 2, 3].

We end this subsection with a very rough sketch of the method of proof of the WIP for Young towers. This includes the Axiom A attractors as a special case for which standard references are [12, 62]. The idea is again to induce to a map $F = T^\tau : Y \to Y$ that is a uniformly hyperbolic transformation with countable partition and full branches, as described in Young [75], with an integrable inducing time $\tau : Y \to \mathbb{Z}^+$ that is constant on partition elements. (Again τ is not necessarily the first return time.) The construction in [75] ensures that there exists an SRB measure μ_Y for F. Starting from F and τ, we construct a "two-sided" Young tower $f : \Delta \to \Delta$ as in (9) with ergodic invariant probability $\mu_\Delta = \mu_Y \times \{\text{counting}\}/\int_Y \tau\, d\mu_Y$. The projection $\pi_\Delta : \Delta \to \Lambda$, $\pi_\Delta(y, \ell) = T^\ell y$, defines a semiconjugacy from f to T, and $\mu = (\pi_\Delta)_*\mu_\Delta$ is the desired SRB measure for $T : M \to M$. Moreover, μ is mixing up to a finite cycle.

Given $v \in C_0^\kappa(M, \mathbb{R}^m)$, we define the lifted observable $\hat{v} = v \circ \pi_\Delta : \Delta \to \mathbb{R}^m$. It suffices to work from now on with $\hat{v}$.

Next, there is a quotienting procedure which projects out the stable directions reducing to an expanding map. Formally, this consists of a "uniformly expanding" map $\bar{F} : \bar{Y} \to \bar{Y}$ and a projection $\pi : Y \to \bar{Y}$ such that $\bar{F} \circ \pi = \pi \circ F$ and such that $\tau(y) = \tau(y')$ whenever $\pi y = \pi y'$. In particular, τ projects to a well-defined return time $\tau : \bar{Y} \to \mathbb{Z}^+$. Using $\bar{F}$ and τ we construct a "one-sided" Young tower $\bar{f} : \bar{\Delta} \to \bar{\Delta}$. The projection π extends to $\pi : \Delta \to \bar{\Delta}$ with $\pi(y, \ell) = (\pi y, \ell)$ and we define $\bar{\mu}_\Delta = \pi_*\mu_\Delta$. The map $\bar{f}$ plays the role of a "nonuniformly expanding map".

As in Subsection 2.3.1, we consider the tails $\mu_Y(\tau > n)$. In the exponential tail setting of [75], $\mu_Y(\tau > n) = O(\gamma^n)$ for some $\gamma \in (0, 1)$ and a version of the "Sinai trick" (see for example [55, Lem. 3.2]) shows that $v \circ \pi_\Delta = \hat{v} + \chi_1 \circ f - \chi_1$ where $\chi_1 \in L^\infty$ and $\hat{v}(y) = \hat{v}(y')$ whenever $\pi y = \pi y'$. In particular, $\hat{v}$ projects to a well-defined observable $\bar{v} : \bar{\Delta} \to \mathbb{R}^m$.

This construction can be carried out so that $\bar{v}$ is sufficiently regular that the analogue of condition (6) holds, where P is the transfer operator on $\bar{\Delta}$. Hence we obtain an L^∞ martingale-coboundary decomposition $\bar{v} = m + \chi_2 \circ \bar{f} - \chi_2$ on $\bar{\Delta}$. This gives the associated decomposition

$$v \circ \pi_\Delta = \hat{v} = m \circ \pi + \chi \circ f - \chi,$$

on Δ where $\chi = \chi_1 + \chi_2 \circ \pi$.

Now the argument is finished, since we can apply the methods from Subsection 2.3.1 to obtain the WIP, etc, for m on $(\bar{\Delta}, \bar{\mu}_\Delta)$, and hence $m \circ \pi$ on (Δ, μ_Δ), $\hat{v}$ on (Δ, μ_Δ), and v on (Λ, μ).

Finally, we consider the case $\mu(\tau > n) = O(n^{-(\beta+1)})$ with $\beta > 1$. In certain situations (nonuniform expansion but uniform contraction) the Sinai trick works as above and reduces to the situation in (10). The general case is more complicated but is covered by [60, Cor. 2.2]. Again, this is optimal since there are many examples with $\beta = 1$ where the CLT with standard scaling does not hold.

2.3.3. *Hyperbolic flows.* Let $\dot{y} = g(y)$ be an ODE defined by a C^2 vector field $g : M \to TM$ on a compact Riemannian manifold M. Let $g_t : M \to M$ denote the corresponding flow. Let $X \subset M$ be a codimension one cross-section transverse to the flow and let $\varphi : X \to \mathbb{R}^+$ be a return time function, namely a function such that $g_{\varphi(x)}(x) \in X$ for $x \in X$. The map $T = f_\varphi : X \to X$ is called the *Poincaré map*. We assume (possibly after shrinking X) that $\inf \varphi > 0$. Given an ergodic invariant probability measure μ_X on X and $\varphi \in L^1(X)$, we construct an ergodic invariant probability measure μ on M as follows. Define the suspension

$$X^\varphi = \{(x, u) \in X \times \mathbb{R} : 0 \le u \le \varphi\}/\sim, \qquad (x, \varphi(x)) \sim (Tx, 0).$$

The suspension flow $T_t : X^\varphi \to X^\varphi$ is given by $T_t(x,u) = (x, u+t)$ modulo identifications. The probability measure $\mu^\varphi = (\mu_X \times \text{Leb}\,)/\int_X \varphi\, d\mu_X$ is ergodic and T_t-invariant. Moreover, $\pi : X^\varphi \to M$ given by $\pi(x,u) = T_u x$ is a semiconjugacy from T_t to g_t and $\mu = \pi_*\mu^\varphi$ is the desired ergodic invariant probability measure on M.

Now suppose that $v \in C_0^\kappa(M, \mathbb{R}^m)$ and define the induced observable

$$V : X \to \mathbb{R}^m, \qquad V(x) = \int_0^{\varphi(x)} v(g_u x)\, du.$$

By a purely probabilistic argument [34] (based on [65, 58, 36, 61]), the WIP for $V : X \to \mathbb{R}^m$ with the map T implies a WIP for $v : M \to \mathbb{R}^m$. That is, setting $v_t = \int_0^t v \circ g_s\, ds$ and $W_n(t) = n^{-1/2} v_{nt}$, we obtain $W_n \to_\mu W$ where W is Brownian motion with covariance $\Sigma = \Sigma_X / \int_X \varphi\, d\mu_X$ and Σ_X is the covariance in the WIP for V.

By Bowen [12], Axiom A flows can be realized as suspension flows over uniformly hyperbolic diffeomorphisms, and the above considerations yield the WIP for attractors for Axiom A flows [22]. The same is true for large classes of nonuniformly hyperbolic flows modelled as suspensions over Young towers with summable decay of correlations, including Lorentz gases, Lorenz attractors [6] and singular hyperbolic attractors [5]. In these situations, μ_X and μ are SRB measures on X and M respectively. Also the nondegeneracy property in Proposition 2.1 applies to Σ_X and thereby $\Sigma = \Sigma_X / \int_X \varphi\, d\mu_X$. Moreover,

$$\Sigma = \lim_{t\to\infty} t^{-1} \int_M v_t \otimes v_t\, d\mu,$$

and under extra (rather restrictive) mixing assumptions

$$\Sigma = \int_0^\infty \int_M \{v \otimes (v \circ g_t) + (v \circ g_t) \otimes v\}\, d\mu\, dt.$$

3. General rough path theory

3.1. **Limit theorems from rough path analysis.** Consider a (for simplicity only: finite-dimensional) Banach space $(\mathcal{B}, \|\cdot\|)$ and fixed $p \in [2,3)$. Define the group $G := \mathcal{B} \oplus (\mathcal{B} \otimes \mathcal{B})$ 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 $(0,0)$. A (level-2) *p-rough path* (over $\mathcal{B}$, on $[0,1]$) is a path $\mathbf{X} = (\mathbf{X}_t : 0 \le t \le 1)$ with values and increments $\mathbf{X}_{s,t} := \mathbf{X}_s^{-1} \star \mathbf{X}_t := (X_{s,t}, \mathbb{X}_{s,t}) \in G$ of finite p-variation condition, $p \in [2,3)$, either in the sense ("*homogeneous rough path norm*")

$$\|\!|\mathbf{X}|\!\|_{p\text{-var}} := \|X\|_{p\text{-var}} + \|\mathbb{X}\|^{1/2}_{(p/2)\text{-var}} < \infty\ , \tag{11}$$

or, equivalently, in terms of the *inhomogeneous rough path norm*

$$\|\mathbf{X}\|_{p\text{-var}} := \|X\|_{p\text{-var}} + \|\mathbb{X}\|_{(p/2)\text{-var}} < \infty\ ; \tag{12}$$

we used the notation, applicable to any Ξ from $\{0 \le s \le t \le 1\}$ into a normed space, any $q > 0$,

$$\|\Xi\|_{q\text{-var}} := \left(\sup_{\mathcal{P}} \sum_{[s,t]\in\mathcal{P}} |\Xi_{s,t}|^q \right)^{\frac{1}{q}} < \infty\ . \tag{13}$$

Write $\mathscr{C}^{p\text{-var}}([0,1],\mathcal{B})$ resp. $\mathscr{D}^{p\text{-var}}([0,1],\mathcal{B})$ for the space of such (continuous resp. càdlàg) p-rough paths; and also $\mathscr{C}$ resp. $\mathscr{D}$ for the space of continuous resp. càdlàg paths with values in $\mathcal{B}\oplus(\mathcal{B}\otimes\mathcal{B})$. The space of α-Hölder rough paths, $\mathscr{C}^{\alpha\text{-Höl}}$ with $\alpha=1/p\in(1/3,1/2]$ forms a popular subclass of $\mathscr{C}^{p\text{-var}}$. A weakly geometric p-rough path, in symbols $\mathbf{X}\in\mathscr{C}_g^{p\text{-var}}$, satisfies a "product rule" of the type $\mathrm{Sym}(\mathbb{X}_t)=(1/2)X_t\otimes X_t$; effectively $\mathbf{X}$ takes values in a sub-group $H\subset G$. (We remark that, when $\mathcal{B}=\mathbb{R}^m$, the (Lie) groups H,G, can be identified with, respectively, the step-2 truncated free nilpotent group with m generators and the step-2 truncated Butcher group with m decorations of its nodes.) Every continuous BV path lifts canonically via

$$\mathbb{X}_t=\int_0^t(X_s-X_0)\otimes dX_s\ ,$$

and gives rise to a (continuous) weakly geometric p-rough path. Conversely, every $\mathbf{X}\in\mathscr{C}_g^{p\text{-var}}$ is the uniform limit of smooth paths, with uniform p-variation bounds. Similarly, every càdlàg BV path X lifts canonically via

$$\mathbb{X}_t=\int_{(0,t]}(X_s^- -X_0)\otimes dX_s$$

to a (càdlàg) p-rough path in $\mathscr{D}^{p\text{-var}}$. We introduce, on $\mathscr{D}^{p\text{-var}}$ (and then by restriction on $\mathscr{C}^{p\text{-var}}$ and $\mathscr{C}_g^{p\text{-var}}$) the *(inhomogeneous) p-rough path distance*[2]

$$\|\mathbf{X};\tilde{\mathbf{X}}\|_{p\text{-var}}:=\|X-\tilde{X}\|_{p\text{-var}}+\|\mathbb{X}-\tilde{\mathbb{X}}\|_{(p/2)\text{-var}}\ . \tag{14}$$

(A similar Hölder rough path distance can be defined on $\mathscr{C}^{\alpha\text{-Höl}}$ and $\mathscr{C}_g^{\alpha\text{-Höl}}$).

Consider sufficiently regular vector fields $V_0:\mathbb{R}^d\to\mathbb{R}^d$ and $V:\mathbb{R}^d\to L(\mathcal{B},\mathbb{R}^d)$. By definition, Y solves the *rough differential equation (RDE)*

$$dY=V_0(Y^-)dt+V(Y^-)d\mathbf{X}$$

if, for all $0\le s<t\le 1$, writing DV for the derivative,[3]

$$Y_t-Y_s=V_0(Y_s)(t-s)+V(Y_s)X_{s,t}+DV(Y_s)V(Y_s)\mathbb{X}_{s,t}+R_{s,t}\ , \tag{15}$$

where, for the "remainder term" R, we require, writing $\mathcal{P}(\varepsilon)$ for a partition of $[0,1]$ with mesh-size less than ε,

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

This definition first encodes that Y is controlled (cf. [28]) by $X\in D^{p\text{-var}}$, with derivative $Y'=W(Y_s)\in D^{p\text{-var}}$ and remainder $Y^{\#}_{s,t}=W(Y_s)\mathbb{X}_{s,t}+R_{s,t}$ with $\|Y^{\#}\|_{(p/2)\text{-var}}<\infty$. As a consequence, Y satisfies a bona fide rough integral equation, for all $t\in(0,1]$,

$$Y_t=y_0+\int_{(0,t]}V_0(Y_s^-)ds+\int_{(0,t]}V(Y_s^-)d\mathbf{X}_s\ .$$

Conversely, (15) is satisfied by every solution to this integral equation. See e.g. [27] for more details on this construction in the Hölder rough path case, and [28, 31] for the càdlàg p-variation case; this contains the discrete Hölder setting of [40]. The

[2] In view of the genuine non-linearity of rough path spaces, we refrain from writing $\|\mathbf{X}-\tilde{\mathbf{X}}\|_{p\text{-var},[0,1]}$.

[3] In coordinates, when $\mathcal{B}=\mathbb{R}^m$, we have $DV(Y_s)V(Y_s)\mathbb{X}_{s,t}=\partial_\alpha V_\gamma(Y_s)V_\beta^\alpha(Y_s)\mathbb{X}_{s,t}^{\beta,\gamma}$ with summation over $\alpha=1,\dots,d$ and $\beta,\gamma=1,\dots,m$.

following theorem, in the case of continuous p-rough paths, is due to Lyons [50], the recent extension to càdlàg rough paths is taken from [31]. We write C^{p+} to indicate $C^{p+\varepsilon}$, for some $\varepsilon > 0$.

Theorem 3.1 (Continuity of RDE solution map)**.** *Let $p \in [2, 3)$. Consider a càdlàg rough path $\mathbf{X} \in \mathscr{D}^{p\text{-}var}([0, 1], \mathcal{B})$, assume $V_0 \in C^{1+}$ and $V \in C^{p+}$. Then there exists a unique càdlàg solution $Y \in D([0, 1], \mathbb{R}^d)$ to the rough differential equation*

$$dY_t = V_0(Y_t^-)dt + V(Y_t^-)d\mathbf{X}_t, \quad Y_0 = y_0 \in \mathbb{R}^d \ ,$$

and the solution is locally Lipschitz in the sense that

$$\|Y - \tilde{Y}\|_{p\text{-}var} \lesssim \|\mathbf{X}; \tilde{\mathbf{X}}\|_{p\text{-}var} + |y_0 - \tilde{y}_0|$$

with proportionality constant uniform over bounded classes of driving p-rough paths.

The p-variation rough path distance can be replaced by a p-variation Skorokhod type rough path metric, which adds more flexibility when the limiting (rough) path has jumps, but we won't need this generality here. Checking p-variation rough path convergence can be done by interpolation: uniform convergence plus uniform p'-variation bounds, for some $p' < p$.

It is known that (càdlàg) semimartingales give rise to càdlàg p-rough paths for any $p \in (2, 3)$ [18]. Solving the resulting random RDE provides exactly a (robust) solution theory for the corresponding SDE. As a consequence, we have the following limit theorems of Stratonovich and Itô type, which cannot be obtained by UCV/UT type argument familiar from stochastic analysis. (Assumptions on V_0, V are as above.) The following theorem applies in particular to sequences of smooth processes, in which case Stratonovich SDEs are simply random ODEs.

Theorem 3.2 (Stratonovich-type limit theorem)**.** *Consider a sequence of continuous semimartingale drivers (B^n) with Stratonovich lift $(\mathbf{B}^{\circ,n})$, such that $\mathbf{B}^{\circ,n}$ converges to $\mathbf{B} = (B, \mathbb{B}^\circ + \Gamma)$, for some continuous BV process Γ, weakly (resp. in probability, a.s.) in the uniform topology with $\{\|\mathbf{B}^{\circ,n}\|_{p\text{-}var}(\omega)\}$ tight, for $p \in (2, 3)$. (Γ is necessarily skew-symmetric.)*

(i) For any $p' > p$, it holds that $\mathbf{B}^{\circ,n} \to \mathbf{B}$ weakly (resp. in probability, a.s.) in the p'-variation rough path topology.

(ii) Assume, in the sense of Stratonovich SDEs, [4]

$$dY_t^n = V_0(Y_t^n)dt + V(Y_t^n) \circ dB_t^n$$

such that $Y_0^n \equiv y_0^n \to y_0$. Then the Stratonovich SDE solutions Y^n converge weakly (resp. in probability, a.s.) to Y in the uniform topology, where $Y_0 = y_0$ and

$$dY_t = V_0(Y_t)dt + DV(Y_t)V(Y_t)d\Gamma_t + V(Y_t) \circ dB_t \ .$$

Moreover, $\{\|Y^n\|_{p\text{-}var}(\omega) : n \geq 1\}$ is tight and one also has weak (resp. in probability, a.s.) convergence in the p'-variation uniform metric for any $p' > p$.

We now state an analogous Itô-type result. The next theorem in particular applies to sequences of piecewise constant, càdlàg processes, in which case, Itô SDEs are simply stochastic recursions.

[4]Often B^n has continuous BV sample paths. Every such process is (trivially) a semimartingale (under its own filtration); the Stratonovich SDE interpretation is the one consistent with the ODE interpretation, in the sense of a Riemann-Stieltjes integral equation.

Theorem 3.3 (Itô-type limit theorem)**.** *Consider a sequence of càdlàg semimartingale drivers B^n, with Itô lift $\mathbf{B}^n = (B^n, \mathbb{B}^n)$, such that $\mathbf{B}^n$ converges to $\mathbf{B} = (B, \mathbb{B} + \Gamma)$, for some càdlàg BV process Γ, weakly (resp. in probability, a.s.) in the uniform topology with $\{\|\mathbf{B}^n\|_{p\text{-}var}(\omega)\}$ tight, for $p \in (2,3)$.*

(i) For any $p' > p$, it holds that $\mathbf{B}^{\circ,n} \to \mathbf{B}$ weakly (resp. in probability, a.s.) in the p'-variation rough path topology.

(ii) Assume, in the sense of Itô SDEs,

$$dY_t^n = V_0(Y_t^{n,-})dt + V(Y_t^{n,-})dB_t^n$$

such that $Y_0^n \equiv y_0^n \to y$. Then Itô SDE solutions Y^n converge weakly (resp. in probability, a.s.) to Y in the uniform topology, where $Y(0) = y$ and

$$dY_t = V_0(Y_t^-)dt + DV(Y_t)V(Y_t)d\Gamma_t + W(Y_t^-)dB_t\ .$$

Moreover, $\{\|Y^n\|_{p\text{-}var}(\omega) : n \geq 1\}$ is tight and one also has weak (resp. in probability, a.s.) convergence in the p'-variation uniform metric for any $p' > p$.

Remark 3.4. *A minor generalization of Theorem 3.3, which will be convenient later on, states that the drift term $V_0(Y_t^{n,-})dt$ in the approximate problem can be replaced by $V_0(Y_t^{n,-})d\tau^n$ where $\tau^n(t) \to t$ uniformly with uniform 1-variation bounds.*

We emphasize that in both Theorem 3.2 and 3.3, the *only purpose* of the semimartingale and adaptedness assumptions is to give a familiar interpretation of what is really a rough differential equation driven by a random rough path. (Consistency with SDEs, in a general semimartingale setting, is established in [18]). The proof is essentially a corollary of interpolation, in a weak convergence setting, with the purely deterministic Theorem 3.1, see [31] for details.

3.2. **WIPs in rough path theory.** We start with some generalities. A *Brownian rough path* (over $\mathbb{R}^m$) is an $\mathbb{R}^m \oplus (\mathbb{R}^m \otimes \mathbb{R}^m)$-valued continuous process $\mathbf{B} = (B, \mathbb{B})$ with independent increments with respect to the group structure introduced in Section 3.1, such that B is centered. (In particular, B is a classical m-dimensional Brownian motion.) It is known that sample paths $\mathbf{B}(\omega)$ are, with probability one, in $\mathscr{C}^{\alpha\text{-Höl}}$ for any $\alpha < 1/2$, and hence also in $\mathscr{C}^{p\text{-var}}$ and $\mathscr{D}^{p\text{-var}}$ for any $p > 2$. We have a full characterization of Brownian rough paths: B is a classical m-dimensional Brownian motion (with some covariance $\Sigma \in \mathbb{R}^m \otimes \mathbb{R}^m$) and

$$\mathbb{B}_{s,t} = \int_s^t B_{s,r} \otimes \circ dB_r + (t-s)\Gamma = \int_s^t B_{s,r} \otimes dB_r + (t-s)(\Gamma + \tfrac{1}{2}\Sigma)$$

for some matrix $\Gamma \in \mathbb{R}^m \otimes \mathbb{R}^m$, which we name *area drift.* (Note that $\mathbf{B}$ is geometric iff Γ is skew-symmetric.) Given a sequence of random rough paths $(\mathbf{B}_\varepsilon)$, we say that the WIP holds *in α-Hölder* (resp. *p-variation*) *rough path sense* if, as $\varepsilon \to 0$,

$$\mathbf{B}_\varepsilon \to_w \mathbf{B} \text{ in } \mathscr{C}^{\alpha\text{-Höl}} \text{ (resp. } \mathscr{C}^{p\text{-var}},\ \mathscr{D}^{p\text{-var}}) \tag{16}$$

but note that only the regimes $\alpha \in (1/3, 1/2)$ (resp. $p \in (2,3)$) correspond to a WIP in a bona fide rough path topology. As is implicit in Theorems 3.2 and 3.3, this follows from checking convergence in law in the uniform topology; that is, an *enhanced weak invariance principle* in the sense that [5]

$$\mathbf{B}_\varepsilon \to_w \mathbf{B} \text{ in } \mathscr{C} \text{ (resp. } \mathscr{D}\text{) as } \varepsilon \to 0 ,$$

together with tightness of α-Hölder (resp. p-variation) rough path norms, at the expense of replacing α (resp. p) in (16) with $\alpha' < \alpha$ (resp. $p' > p$).

In simple situations, $\mathbf{B}_\varepsilon$ is given as the canonical (Stratonovich or Itô) lift of a *good* sequence of convergent semimartingales. In this case, the limiting area drift is zero and $\{|||\mathbf{B}_\varepsilon|||_{p\text{-var}} : \varepsilon \in (0,1]\}$ is automatically tight [18], for any $p > 2$. (This gives a decisive link between classical semimartingale stability theory [38, 47] with rough path analysis.) Immediate applications then include Donsker's theorem in p-variation rough path topology, under identical (finite second) moment assumptions as the classical Donsker theorem. (With piecewise linear interpolation, the limit is the Stratonovich lift $(B, \int B \otimes \circ dB)$, with piecewise constant interpolation the limit is the Itô lift $(B, \int B \otimes dB)$.) As another immediate application, the (functional) CLT for L^2-martingales with stationary, ergodic increments is valid on a rough path level [18]. If interested in the α-Hölder rough path topology, one can use a Kolmogorov-type tightness criterion [30]. Provided $q > 1$, a uniform moment estimate of the form

$$\sup_{\varepsilon \in (0,1]} \mathbb{E}\Big[|||\mathbf{B}_\varepsilon(s,t)|||^{2q}\Big]^{1/2q} \lesssim |t-s|^{1/2} , \tag{17}$$

or equivalently,

$$\sup_{\varepsilon \in (0,1]} \mathbb{E}\Big[|B_\varepsilon(s,t)|^{2q}\Big]^{1/2q} \lesssim |t-s|^{1/2} , \quad \sup_{\varepsilon \in (0,1]} \mathbb{E}\Big[|\mathbb{B}_\varepsilon(s,t)|^{q}\Big]^{1/q} \lesssim |t-s| ,$$

gives tightness in the α-Hölder topology, for every $\alpha < 1/2 - 1/(2q)$.

Remark 3.5. *In the Hölder setting, note that only $\alpha > \frac{1}{3}$ gives a bona fide (level-2) rough path metric under which the Itô map behaves continuously. This leads to the suboptimal moment assumption $q > 3$. To obtain the WIP in the Hölder rough path sense* [13]*, this necessitates increments with* 6+ *moments. (In contrast, we have the WIP in the p-variation rough path topology under the optimal assumption of* 2+ *moments.) This is also the main drawback of using the Hölder topology in* [41].

Of course, in Gaussian situations such moment assumptions are harmless and can conveniently be reduced to $q = 1$. An instructive example is given by physical Brownian motion X^ε, as introduced in Section 2.1. The tightness condition can be seen to be satisfied for all $q < \infty$, giving α-Hölder rough path tightness for any $\alpha < 1/2$. More interestingly, X^ε has a Brownian rough path limit with non-zero area drift [26]*, provided the particle feels a Lorentz force, expressed through non-symmetry of M. (See the notation in Section 2.1.) Specifically, this is seen by writing MX^ε as Brownian motion plus a "corrector" which goes uniformly to zero, but leads, in the $\varepsilon \to 0$ limit, to an area contribution.*

Remark 3.6. *We note that (17) requires no martingale assumptions whatsoever. It is an important observation for the sequel that (17) leads to tightness not only in the α-Hölder rough path topology but also in the p-variation rough path topology: to wit, it follows from the Besov-variation embedding* [29] *that (17) implies p-variation tightness for any $p > 2$. In this way, for example, one can reprove the*

[5]Again it suffices to work with the uniform topology on both $\mathscr{C}$ and $\mathscr{D}$.

WIP in p-variation rough path topology under the almost optimal assumption of $2+$ *moments. This argument becomes important when direct martingale arguments are not possible.*

4. Applications to fast-slow systems

4.1. **Chaotic dynamics: enhanced WIP and moments.** In this subsection, we resume the discussion of chaotic dynamical systems from Section 2.3 but now focusing on some finer statistical properties, namely an enhanced WIP and moment estimates, that are required for applying rough path theory.

4.1.1. *Expanding maps.* Continuing Subsection 2.3.1, we suppose that $T : M \to M$ is a C^2 uniformly expanding map, with unique ergodic absolutely continuous invariant probability measure μ, so conditions (5) and (6) hold. In particular, for any $v \in C_0^\kappa(M, \mathbb{R}^m)$, we have an L^2 martingale-coboundary decomposition decomposition $v = m + \chi \circ T - \chi$. Define the càdlàg processes $W_n \in D([0,1], \mathbb{R}^m)$, $\mathbb{W}_n \in D([0,1], \mathbb{R}^m \otimes \mathbb{R}^m)$,

$$W_n(t) = n^{-1/2} \sum_{0 \le j < n} v \circ T^j, \qquad \mathbb{W}_n(t) = n^{-1} \sum_{0 \le i < j < n} (v \circ T^i) \otimes (v \circ T^j).$$

Recall that we have the WIP $W_n \to_\mu W$ in $D([0,1], \mathbb{R}^m)$ where W is m-dimensional Brownian motion with covariance Σ given by formulas (7) and (8). By [41, Thm. 4.3], we have the *enhanced WIP* (called iterated WIP in [41])

$$(W_n, \mathbb{W}_n) \to_\mu (W, \mathbb{W}) \quad \text{in } \mathscr{D}([0,1], \mathbb{R}^m \times (\mathbb{R}^m \otimes \mathbb{R}^m)),$$

where $\mathbb{W}(t) = \int_0^t W \otimes dW + \Gamma t$. Here $\int W \otimes dW$ is the Itô integral and the area drift $\Gamma \in \mathbb{R}^m \otimes \mathbb{R}^m$ is given by

$$\Gamma = \lim_{n \to \infty} \int_M \mathbb{W}_n(1)\, d\mu, \tag{18}$$

and satisfies

$$\Gamma = \sum_{n=1}^{\infty} \int_M v \otimes (v \circ T^n)\, d\mu. \tag{19}$$

The proof of the enhanced WIP in [41] has two main steps. The first step is to apply [47, Thm. 2.2] (alternatively [38]) to the martingale component m taking into consideration that $\{m \circ T^n,\, n \ge 0\}$ is a reverse martingale difference sequence. This yields an enhanced WIP with zero area drift. The contribution from the coboundary $\chi \circ T - \chi$ is no longer negligible, but a general result [41, Thm. 3.1] for mixing dynamical systems and L^2 coboundaries yields the (typically nonzero) area drift Γ.

Again, strong distributional convergence applies by [79, Thm. 1]. The hypotheses in [79] are verified in the course of the proof of [41, Lem. 6.3]. Hence $(W_n, \mathbb{W}_n) \to_\lambda (W, \mathbb{W})$ in $\mathscr{D}([0,1], \mathbb{R}^m \times (\mathbb{R}^m \otimes \mathbb{R}^m))$ for all absolutely continuous Borel probability measures λ.

For nonuniformly expanding maps, the enhanced WIP goes through unchanged provided the martingale-coboundary decomposition holds in L^2 (with the usual caveat that formula (19) only holds when μ is mixing). This covers the situation (10) with $\beta > 2$. As before, extra work is required for the case $\beta \in (1, 2]$. By [41, Thm. 10.2], the enhanced WIP holds for all $\beta > 1$ for nonuniformly expanding maps modelled by Young towers, including the intermittent maps (3). For such maps, we obtain optimal results: the enhanced WIP holds precisely when the ordinary CLT holds.

Turning to moments, an immediate consequence of the L^p martingale-coboundary decomposition, $p \geq 2$, and Burkhölder's inequality is that $\|v_n\|_p = O(n^{1/2})$ where the implied constant depends on v and p. As noted in [56, 53, 59], in fact

$$\|v_n\|_{2p} = O(n^{1/2}) \tag{20}$$

and this holds for L^p martingale-coboundary decompositions with $p \geq 1$. This improved result uses the additional information that $v \in L^\infty$ and a maximal inequality of [66].

In the situation (10), we have the martingale-coboundary decomposition for all $1 \leq p < \beta$. As shown in [56, 53], the estimate (20) is sharp; $\|v_n\|_q = O(n^{1/2})$ for $q < 2\beta$ but there are examples where the estimate typically fails for $q > 2\beta$.

We also require estimates for the enhanced (iterated) moment $S_n = \sum_{0 \leq i \leq j < n}(v \circ T^i) \otimes (v \circ T^j)$. Assuming an L^p martingale-coboundary decomposition with $p \geq 3$ and $v \in L^\infty$, it was shown in [41, Prop. 7.1] that $\|S_n\|_{2p/3} = O(n)$. In the Young tower setting, this has been improved in [45] to $\|S_n\|_p = O(n)$ for $p \geq 1$.

The moment estimates discussed above are all in L^p spaces with respect to μ. Clearly if $\lambda \ll \mu$ and $d\lambda/d\mu \in L^\infty$ then the same moment estimates hold also with respect to λ. In particular, we can take $\lambda = \mathrm{Vol}$ for the C^2 uniformly expanding maps. For the intermittent maps (3) it is standard that $d\mu/d\,\mathrm{Leb}$ is bounded below, so we can take $\lambda = \mathrm{Leb}$.

4.1.2. *Hyperbolic diffeomorphism.* For Axiom A diffeomorphisms and Young towers with exponential tails, we saw in Subsection 2.3.2 that there is an L^p martingale-coboundary decomposition for all p. Also, for Young towers with exponential contraction and polynomial tails $\mu_Y(\tau > n) = O(n^{-(\beta+1)})$, we have an L^p martingale-coboundary decomposition for $p < \beta$. By [41, 45], we obtain the enhanced WIP provided $\beta > 1$ and optimal moment estimates $\|v_n\|_{2p} = O(n^{1/2})$ and $\|S_n\|_p = O(n)$ for $1 \leq p < \beta$. As before, the covariance Σ and drift Γ satisfy (7) and (18), and under additional mixing assumptions we have (8) and (19).

For general Young towers with polynomial tails $\mu_Y(\tau > n) = O(n^{-\beta})$ the enhanced WIP still holds for all $\beta > 1$ by [60] but currently we only have the moment estimates $\|v_n\|_{2p} = O(n^{1/2})$ and $\|S_n\|_{2p/3} = O(n)$ for $3 \leq p < \beta$ from [41]. Obtaining optimal moment estimates here is the subject of work in progress.

4.1.3. *Hyperbolic flows.* The methods mentioned in Subsection 2.3.3 for passing the WIP from (non)uniformly hyperbolic diffeomorphisms to (non)uniformly hyperbolic flows work just as well for the enhanced WIP [41, Sec. 6]. Define

$$W_n(t) = n^{-1/2} \int_0^{nt} v \circ g_s \, ds \, , \qquad \mathbb{W}_n(t) = n^{-1} \int_0^{nt} \int_0^{s} (v \circ g_r) \otimes (v \circ g_s) \, dr \, ds \, .$$

Then $(W_n, \mathbb{W}_n) \to_\lambda (W, \mathbb{W})$ where W is Brownian motion with covariance Σ and $\mathbb{W}_n(t) = \int_0^t W \otimes dW + \Gamma_I t$. Here $\Sigma = \lim_{n\to\infty} \mathbb{E}_\lambda(W_n(1) \otimes W_n(1))$ as before, and $\Gamma_I = \lim_{n\to\infty} \mathbb{E}_\lambda \mathbb{W}_n(1)$. Alternatively, $\mathbb{W}_n(t) = \int_0^t W \otimes \circ dW + \Gamma t$, where

$\Gamma = \Gamma_I - \frac{1}{2}\Sigma$ is skew-symmetric. Under extra mixing assumptions,

$$\Sigma = \int_0^\infty \int_\Lambda \{v \otimes (v \circ g_t) + (v \circ g_t) \otimes v\}\, d\mu\, dt \ , \tag{21}$$

$$\Gamma = \frac{1}{2} \int_0^\infty \int_\Lambda \{v \otimes (v \circ g_t) - (v \circ g_t) \otimes v\}\, d\mu\, dt \ . \tag{22}$$

The situation for moments extends in a straightforward way [41, Sec. 7.2]. Define

$$v_n = \int_0^n v \circ g_s\, ds \ , \qquad S_n = \int_0^n \int_0^s (v \circ g_r) \otimes (v \circ g_s)\, dr\, ds \ .$$

Then the estimates described in Subsection 2.3.1 apply equally here.

4.2. **Continuous dynamics.** We present now an application of rough path theory to fast-slow systems where the fast variable satisfies a suitable WIP. We consider first continuous dynamics (1) in the case of multiplicative noise, i.e.,

$$\dot{x}_\varepsilon = a(x_\varepsilon) + \varepsilon^{-1} b(x_\varepsilon) v(y_\varepsilon) \ , \qquad \dot{y}_\varepsilon = \varepsilon^{-2} g(y_\varepsilon) \ . \tag{23}$$

We consider the $\mathbb{R}^m$-valued path

$$W_\varepsilon(t) = \varepsilon \int_0^{t\varepsilon^{-2}} v \circ g_s\, ds$$

and rewrite the slow dynamics in the form of a controlled ODE,

$$dx_\varepsilon = a(x_\varepsilon)dt + b(x_\varepsilon)dW_\varepsilon \ .$$

Following Kelly–Melbourne [41], this formulation invites an application of finite-dimensional rough path theory; the only modification relative to [41] is our present use of p-variation rough path metrics, which leads to optimal moment assumptions and optimal regularity assumptions on the coefficients. The key is a suitable WIP on the level of rough paths, as discussed in Section 3.

To this end, we consider the following two assumptions on the fast dynamics. Following the discussion in Section 4.1, we see that a wide range of dynamics satisfy these assumptions. For every $\varepsilon > 0$, we let $\mathbb{W}_\varepsilon$ be the canonical second iterated integral of W_ε, and for $p \in (2, 3)$, we consider the geometric p-rough path $\mathbf{W}_\varepsilon := (W_\varepsilon, \mathbb{W}_\varepsilon)$.

Assumption 4.1. *It holds that* $(W_\varepsilon, \mathbb{W}_\varepsilon) \to (W, \mathbb{W})$ *as* $\varepsilon \to 0$ *in the sense of finite-dimensional distributions on* (M, λ)*, where* W *is an* m*-dimensional Brownian motion and* $\mathbb{W}(t) = \int_0^t W \otimes \circ dW + \Gamma t$ *for some* $\Gamma \in \mathbb{R}^m \otimes \mathbb{R}^m$ *deterministic.*

Assumption 4.2. *There exists* $q > 1$ *and* $K > 0$ *such that*

$$\Big\| \int_s^t v^i \circ g_r\, dr \Big\|_{L^{2q}(\lambda)} \le K|t-s|^{1/2} \ ,$$
$$\Big\| \int_s^t \int_s^r v^i \circ g_u\, v^j \circ g_r\, du\, dr \Big\|_{L^q(\lambda)} \le K|t-s| \ ,$$

for all $s, t \ge 0$ *and* $1 \le i, j \le m$.

The first assumption identifies the possible limit points of W_ε as a rough path; the second ensures that the WIP holds in a sufficiently strong rough path topology as demonstrated by the following result. As before, all path space norms (p-var, α-Höl, etc.) are relative to the fixed interval $[0, 1]$.

Proposition 4.3. *Under Assumption 4.2, it holds that for all* $p \in (2,3)$

$$\sup_{\varepsilon\in(0,1]} \mathbb{E}\|\!|\mathbf{W}_\varepsilon\|\!|_{p\text{-}var}^{2q} < \infty \;,$$

and for all $\alpha \in (0, \frac{1}{2} - \frac{1}{2q})$

$$\sup_{\varepsilon\in(0,1]} \mathbb{E}\|\!|\mathbf{W}_\varepsilon\|\!|_{\alpha\text{-}H\ddot{o}l}^{2q} < \infty \;.$$

Proof. Viewing $\mathbf{W}_\varepsilon$ as a path in $G^2(\mathbb{R}^m) \subset \mathbb{R}^m \oplus (\mathbb{R}^m \otimes \mathbb{R}^m)$, the step-2 free nilpotent group equipped with Carnot-Carathéodory metric d, it holds that

$$|d(\mathbf{W}_\varepsilon(s), \mathbf{W}_\varepsilon(t))|_{L^{2q}(\lambda)} \lesssim |W_\varepsilon(s,t)|_{L^{2q}(\lambda)} + |\mathbb{W}_\varepsilon(s,t)|_{L^q(\lambda)}^{1/2} \lesssim |t-s|^{1/2},$$

where the final bound follows from Assumption 4.2. Let $\beta \in [0,1/2)$. Then $\mathbb{E}[|\mathbf{W}_\varepsilon|_{W^{\beta,2q}}^{2q}]$ is uniformly bounded in $\varepsilon > 0$, and thus, by the Besov-Hölder and Besov-variation embeddings [29] (see also [30, Cor. A.2, A.3]), so is $\mathbb{E}\|\!|\mathbf{W}_\varepsilon\|\!|_{(\beta-1/(2q))\text{-Höl}}^{q}$ and $\mathbb{E}\|\!|\mathbf{W}_\varepsilon\|\!|_{(1/\beta)\text{-var}}^{2q}$. □

Theorem 4.4. *Suppose Assumptions 4.1 and 4.2 hold.*

(i) For every $p > 2$*, it holds that* $\mathbf{W}_\varepsilon \to_\lambda \mathbf{W}$ *in the p-variation rough path topology.*

(ii) Let $a \in C^{1+}(\mathbb{R}^d, \mathbb{R}^d)$*,* $b \in C^{2+}(\mathbb{R}^d, \mathbb{R}^{d\times m})$*, and let* x_ε *be the solution to* (23)*. Then* $x_\varepsilon \to_\lambda X$ *in* $C^{p\text{-}var}([0,1], \mathbb{R}^d)$ *for every* $p > 2$*, where* X *is the solution to the SDE*

$$dX = \Big(a(X) + \sum_{i,j=1}^m \Gamma^{i,j} \sum_{k=1}^d b^{i,k}\partial_k b^j(X)\Big)dt + b(X)\circ dW, \quad X(0) = \xi. \tag{24}$$

Remark 4.5. *If follows from Assumptions 4.1 and 4.2 that the covariance matrix* Σ *and area drift* Γ *are given by*

$$\Sigma = \lim_{\varepsilon\to 0} \mathbb{E}_\lambda(W_\varepsilon(1) \otimes W_\varepsilon(1)), \qquad \Gamma = \lim_{\varepsilon\to 0} \mathbb{E}_\lambda \mathbb{W}_\varepsilon(1) - \frac{1}{2}\Sigma.$$

Under additional mixing assumptions, formulas (21) *and* (22) *hold.*

Proof. (i) follows from part (i) of Theorem 3.2. For (ii), observe that x_ε solves the ODE

$$dx_\varepsilon = a(x_\varepsilon)dt + b(x_\varepsilon)dW_\varepsilon \;.$$

We are thus in the framework of part (ii) of Theorem 3.2, from which the conclusion follows. □

4.3. **Discrete dynamics.** We now discuss discrete dynamics (2) in the case of multiplicative noise, i.e.,

$$X_{j+1}^{(n)} = X_j^{(n)} + n^{-1}a(X_j^{(n)}) + n^{-1/2} b(X_j^{(n)}) v(Y_j) \;,$$

where, as before, $v : M \to \mathbb{R}^m$, $b : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, and $a : \mathbb{R}^d \to \mathbb{R}^d$. As usual, $X_0^{(n)} = \xi \in \mathbb{R}^d$ is fixed and Y_0 is drawn randomly from a probability measure λ on M. To consider this system as a controlled ODE, we introduced the càglàd path

$$x_n : [0,1] \to \mathbb{R}^d \;, \quad x_n(t) = X_{\lceil nt\rceil - 1}^{(n)} \tag{25}$$

(where we set $X^{(n)}_{-1} := X^{(n)}_0$), as well as the the càdlàg paths

$$W_n : [0,1] \to \mathbb{R}^m , \quad W_n(t) = n^{-1/2} \sum_{j=0}^{\lfloor nt \rfloor - 1} v(Y_j) ,$$
$$z_n : [0,1] \to \mathbb{R} , \quad z_n(t) = \lfloor tn \rfloor / n .$$

It is easy to verify that x_n defined by (25) is the unique solution of the controlled (discontinuous) ODE

$$dx_n = a(x_n^-)dz_n + b(x_n^-)dW_n , \quad x_n(0) = \xi \in \mathbb{R}^d . \tag{26}$$

Let us denote by $\mathbb{W}_n$ the canonical second iterated integral of W_n

$$\mathbb{W}^{i,j}_{v,n}(s,t) = \int_{(s,t]} (W^{i,-}_{v,n}(r) - W^i_{v,n}(s)) dW^j_{v,n}(r) , \quad 1 \le i,j \le m .$$

Consider the following analogues of Assumptions 4.1 and 4.2.

Assumption 4.6. *It holds that $(W_n, \mathbb{W}_n) \to (W, \mathbb{W})$ as $\varepsilon \to 0$ in the sense of finite-dimensional distributions on (M, λ), where W is a Brownian motion in $\mathbb{R}^m$ and $\mathbb{W}(t) = \int_0^t W \otimes dW + \Gamma t$ for some $\Gamma \in \mathbb{R}^{m\times m}$ deterministic.*

Assumption 4.7. *There exists $q > 1$ and $K > 0$ such that for all $n \ge 1$ and $0 \le k, l \le n$,*

$$\big\| W_n(l/n) - W_n(k/n) \big\|_{L^{2q}(\lambda)} \le K n^{-1/2} |l-k|^{1/2} ,$$
$$\big\| \mathbb{W}_n(k/n, l/n) \big\|_{L^q(\lambda)} \le K n^{-1} |l-k| .$$

Proposition 4.8. *Under Assumption 4.7, for all $p \in (2,3)$*

$$\sup_{n\ge 1} \mathbb{E} |\!|\!| (W_n, \mathbb{W}_n) |\!|\!|^{2q}_{p\text{-}var} < \infty .$$

Proof. This is a direct application of [31, Prop. 6.17]. □

Remark 4.9. *If follows from Assumptions 4.6 and 4.7 that the covariance matrix Σ and the area drift Γ are given by*

$$\Sigma = \lim_{n\to\infty} \mathbb{E}_\lambda (W_n(1) \otimes W_n(1)), \qquad \Gamma = \lim_{n\to\infty} \mathbb{E}_\lambda \mathbb{W}_n(1).$$

Under additional mixing assumptions, formulas (8) *and* (19) *hold.*

Combining Theorem 3.3 and Proposition 4.8, we arrive at the following convergence result which relaxes the moment conditions required in [41].

Theorem 4.10. *Suppose that Assumptions 4.6 and 4.7 hold.*

(i) For every $p > 2$, it holds that $\mathbf{W}_n \to_\lambda \mathbf{W}$ in the p-variation rough path topology.

(ii) Let $a \in C^{1+}(\mathbb{R}^d, \mathbb{R}^d)$, $b \in C^{2+}(\mathbb{R}^d, \mathbb{R}^{d\times m})$, and let x_n be the solution to (26). *Then $x_n \to_\lambda X$ in $C^{p\text{-}var}([0,1], \mathbb{R}^d)$ for all $p > 2$, where X is the solution to the SDE*

$$dX = \Big(a(X) + \sum_{i,j=1}^m \Gamma^{i,j} \sum_{k=1}^d b^{i,k} \partial_k b^j (X) \Big) dt + b(X)\, dW , \quad X(0) = \xi .$$

5. Extension to families and non-product case

Throughout this article, we restricted attention to the case of multiplicative noise given in product form. The general form (1) was addressed in [42], though with suboptimal moment assumptions. By adapting the methods of this section to an infinite-dimensional rough paths setting similar to [42], we are able to handle, with optimal moment assumptions, a generalisation of (1) of the form

$$\frac{d}{dt}x_\varepsilon = a_\varepsilon(x_\varepsilon, y_\varepsilon) + \varepsilon^{-1} b_\varepsilon(x_\varepsilon, y_\varepsilon), \qquad \frac{d}{dt}y_\varepsilon = \varepsilon^{-2} g_\varepsilon(y_\varepsilon)\;, \tag{27}$$

where $a_\varepsilon, b_\varepsilon, g_\varepsilon$ now depend on ε, and so does the probability measure λ_ε from which $y_\varepsilon(0)$ is drawn randomly. We assume we are also given a family μ_ε of ergodic $g_{\varepsilon,t}$-invariant probability measures on M, where $g_{\varepsilon,t}$ is the flow generated by g_ε; we require that $\int_M b_\varepsilon(x,y)\, d\mu_\varepsilon(y) = 0$ for all $\varepsilon \in [0,1]$ and $x \in \mathbb{R}^d$.

We note that a similar generalisation is also possible for the discrete dynamics (2), which was not addressed in [42] even in the ε-independent setting. Details are found in our forthcoming work [19].

Let $C^\eta_\varepsilon(M, \mathbb{R}^m)$ be the space of C^η functions $v : M \to \mathbb{R}^m$ with $\int_M v\, d\mu_\varepsilon = 0$. Fix $q \in (1, \infty]$, $\kappa, \bar\kappa > 0$, $\alpha > 2 + \frac{d}{q}$. Let $a_\varepsilon \in C^{1+\bar\kappa,0}(\mathbb{R}^d \times M, \mathbb{R}^d)$ and $b_\varepsilon \in C^{\alpha,\kappa}_\varepsilon(\mathbb{R}^d \times M, \mathbb{R}^d)$ satisfying

$$\sup_{\varepsilon\in[0,1]} \|a_\varepsilon\|_{C^{1+\bar\kappa,0}} < \infty\;, \qquad \sup_{\varepsilon\in[0,1]} \|b_\varepsilon\|_{C^{\alpha,\kappa}} < \infty\;, \qquad \lim_{\varepsilon\to 0} \|b_\varepsilon - b_0\|_{C^{\alpha,\kappa}} = 0\;.$$

For $v \in C^\eta_\varepsilon(M, \mathbb{R}^m)$, define

$$W_{v,\varepsilon}(t) = \varepsilon \int_0^{\varepsilon^{-2}t} v \circ g_{\varepsilon,s}\, ds, \qquad \mathbb{W}_{v,\varepsilon}(t) = \int_0^t W_{v,\varepsilon} \otimes dW_{v,\varepsilon}.$$

We require the following assumptions.

(1) Moment bounds: there exists $K > 0$ such that for all families v_ε, $w_\varepsilon \in C^\kappa_\varepsilon(M)$, it holds that for all $s,t \geq 0$ and $\varepsilon \in [0,1]$,

$$\Big\| \int_s^t v_\varepsilon \circ g_{\varepsilon,r}\, dr \Big\|_{L^{2q}(\lambda_\varepsilon)} \leq K \|v_\varepsilon\|_{C^\kappa} |t-s|^{1/2}\;,$$
$$\Big\| \int_s^t \int_s^r v_\varepsilon \circ g_{\varepsilon,u}\, w_\varepsilon \circ g_{\varepsilon,r}\, du\, dr \Big\|_{L^q(\lambda_\varepsilon)} \leq K \|v_\varepsilon\|_{C^\kappa} \|w_\varepsilon\|_{C^\kappa} |t-s|\;.$$

(2) Enhanced WIP: there exists a bilinear operator $\mathfrak{B} : C^\eta_0(M) \times C^\eta_0(M) \to \mathbb{R}$ such that for every family $v_\varepsilon \in C^\kappa_\varepsilon(M, \mathbb{R}^m)$ with $\lim_{\varepsilon\to0} |v_\varepsilon - v_0|_{C^\kappa} = 0$, there exists an m-dimensional Brownian motion W such that

$$(W_{v_\varepsilon,\varepsilon}, \mathbb{W}_{v_\varepsilon,\varepsilon}) \to_{\lambda_\varepsilon} (W, \mathbb{W}), \quad \text{as } \varepsilon \to 0,$$

in the sense of finite-dimensional distributions, where $\mathbb{W}^{i,j}(t) = \int_0^t W^i\, dW^j + \mathfrak{B}(v_0^i, v_0^j)t$.

(3) Convergence of drift: it holds that

$$\sup_{t\in[0,1]} \|V_\varepsilon(t) - \bar a t\|_{C^{1+\bar\kappa}} \to_{\lambda_\varepsilon} 0 \quad \text{as } \varepsilon \to 0,$$

where $V_\varepsilon(t) = \int_0^t a_\varepsilon(\cdot, y_\varepsilon(r))dr$ and $\bar a = \int_M a_0(\cdot, y) d\mu_0(y)$.

Consider the SDE

$$dX = \tilde{a}(X)\,dt + \sigma(X)\,dB\ , \quad X(0) = \xi\ , \tag{28}$$

where B is the standard Brownian motion in $\mathbb{R}^d$ and $\tilde{a}$ and σ are given by

$$\tilde{a}^i(x) = \bar{a}^i(x) + \sum_{k=1}^{d} \mathfrak{B}(b_0^k(x,\cdot), \partial_k b_0^i(x,\cdot))\ , \quad i = 1,\dots,d\ ,$$

$$(\sigma(x)\sigma^T(x))^{ij} = \mathfrak{B}(b_0^i(x,\cdot), b_0^j(x,\cdot)) + \mathfrak{B}(b_0^j(x,\cdot), b_0^i(x,\cdot))\ , \quad i,j = 1,\dots,d\ .$$

If the SDE (28) has a unique solution X, then under the assumptions (1–3) above one can conclude that $x_\varepsilon \to_{\lambda_\varepsilon} X$.

References

[1] José F. Alves, Jorge M. Freitas, Stefano Luzzatto, and Sandro Vaienti. From rates of mixing to recurrence times via large deviations. *Adv. Math.*, 228(2):1203–1236, 2011.

[2] José F. Alves, Stefano Luzzatto, and Vilton Pinheiro. Markov structures and decay of correlations for non-uniformly expanding dynamical systems. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 22(6):817–839, 2005.

[3] José F. Alves and Vilton Pinheiro. Gibbs-Markov structures and limit laws for partially hyperbolic attractors with mostly expanding central direction. *Adv. Math.*, 223(5):1706–1730, 2010.

[4] D. V. Anosov. Geodesic flows on closed Riemannian manifolds of negative curvature. *Trudy Mat. Inst. Steklov.*, 90:209, 1967.

[5] V. Araujo and I. Melbourne. Mixing properties and statistical limit theorems for singular hyperbolic flows without a smooth stable foliation. *ArXiv e-prints*, November 2017.

[6] V. Araújo, I. Melbourne, and P. Varandas. Rapid mixing for the Lorenz attractor and statistical limit laws for their time-1 maps. *Comm. Math. Phys.*, 340(3):901–938, 2015.

[7] I. Bailleul and R. Catellier. Rough flows and homogenization in stochastic turbulence. *J. Differential Equations*, 263(8):4894–4928, 2017.

[8] Michael Benedicks and Lai-Sang Young. Markov extensions and decay of correlations for certain Hénon maps. *Astérisque*, (261):xi, 13–56, 2000. Géométrie complexe et systèmes dynamiques (Orsay, 1995).

[9] Patrick Billingsley. The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.*, 12:788–792, 1961.

[10] Patrick Billingsley. *Convergence of probability measures*, *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons Inc., New York, 1999.

[11] George D. Birkhoff. Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. U.S.A.*, 17(12):656–660, 1931.

[12] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Lecture Notes in Mathematics, Vol. 470. Springer-Verlag, Berlin-New York, 1975.

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

[14] B. M. Brown. Martingale central limit theorems. *Ann. Math. Statist.*, 42:59–66, 1971.

[15] Jérôme Buzzi and Véronique Maume-Deschamps. Decay of correlations for piecewise invertible maps in higher dimensions. *Israel J. Math.*, 131:203–220, 2002.

[16] N. Chernov and L. S. Young. Decay of correlations for Lorentz gases and hard balls. In *Hard ball systems and the Lorentz gas*, volume 101 of *Encyclopaedia Math. Sci.*, pages 89–120. Springer, Berlin, 2000.

[17] I. Chevyrev. Random walks and Lévy processes as rough paths. *Probab. Theory Related Fields*, 170(3-4):891–932, 2018.

[18] I. Chevyrev and P. K. Friz. Canonical RDEs and general semimartingales as rough paths. *ArXiv e-prints*, April 2017. To appear in Ann. Probab.

[19] I. Chevyrev, P.K. Friz, A. Korepanov, I. Melbourne, and H. Zhang. In preparation.

[20] Christophe Cuny and Florence Merlevède. Strong invariance principles with rate for "reverse" martingale differences and applications. *J. Theoret. Probab.*, 28(1):137–183, 2015.

[21] A. M. Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Appl. Math. Res. Express. AMRX*, no. 2:Art. ID abm009, 40 pp, 2007.
[22] Manfred Denker and Walter Philipp. Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergodic Theory Dynam. Systems*, 4(4):541–552, 1984.
[23] Monroe D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, No. 6:12, 1951.
[24] G. K. Eagleson. Some simple conditions for limit theorems to be mixing. *Teor. Verojatnost. i Primenen.*, 21(3):653–660, 1976.
[25] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.
[26] P. K. Friz, P. Gassiat, and T. J. Lyons. Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.*, 367(11):7939–7955, 2015.
[27] P. K. Frizand M. Hairer. *A Course on Rough Path Analysis, with an Introduction to Regularity Structures, Springer 2014*. Universitext. Springer, 2014.
[28] P. K. Friz and A. Shekhar. General rough integration, Lévy rough paths and a Lévy–Kintchine-type formula. *Ann. Probab.*, 45(4):2707–2765, 07 2017.
[29] P. K. Friz and N. Victoir. A variation embedding theorem and applications. *Journal of Functional Analysis*, 239(2):631 – 637, 2006.
[30] P. K. Friz and N. Victoir. *Multidimensional Stochastic Processes as Rough Paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
[31] P. K. Friz and H. Zhang. Differential equations driven by rough paths with jumps. *Journal of Differential Equations*, 264(10):6226 – 6301, 2018
[32] M. I. Gordin. The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR*, 188:739–741, 1969.
[33] Georg A. Gottwald and Ian Melbourne. Homogenization for deterministic maps and multiplicative noise. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 469(2156):20130201, 16, 2013.
[34] Georg A. Gottwald and Ian Melbourne. Central limit theorems and suppression of anomalous diffusion for systems with symmetry. *Nonlinearity*, 29(10):2941–2960, 2016.
[35] Sébastien Gouëzel. Central limit theorem and stable laws for intermittent maps. *Probab. Theory Related Fields*, 128(1):82–122, 2004.
[36] Sébastien Gouëzel. Statistical properties of a skew product with a curve of neutral points. *Ergodic Theory Dynam. Systems*, 27(1):123–151, 2007.
[37] Franz Hofbauer and Gerhard Keller. Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.*, 180(1):119–140, 1982.
[38] A. Jakubowski, J. Mémin, and G. Pagès. Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbf{D}^1$ de Skorokhod. *Probab. Theory Related Fields*, 81(1):111–137, 1989.
[39] Gerhard Keller. Generalized bounded variation and applications to piecewise monotonic transformations. *Z. Wahrsch. Verw. Gebiete*, 69(3):461–478, 1985.
[40] David Kelly. Rough path recursions and diffusion approximations. *Ann. Appl. Probab.*, 26(1):424–461, 2016.
[41] David Kelly and Ian Melbourne. Smooth approximation of stochastic differential equations. *Ann. Probab.*, 44(1):479–520, 01 2016.
[42] David Kelly and Ian Melbourne. Deterministic homogenization for fast-slow systems with chaotic noise. *Journal of Functional Analysis*, 272(10):4063 – 4102, 2017.
[43] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.
[44] A. Korepanov, Z. Kosloff, and I. Melbourne. Martingale-coboundary decomposition for families of dynamical systems, *Annales l'Institut H. Poincare. Anal. Non Lineaire*, 35(1):859–885, 2018.
[45] A. Korepanov, Z. Kosloff, and I. Melbourne. Deterministic homogenization for families of fast-slow systems. In preparation.
[46] K. Krzyżewski and W. Szlenk. On invariant measures for expanding differentiable mappings. *Studia Math.*, 33:83–92, 1969.

[47] Thomas G. Kurtz and Philip Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.*, 19(3):1035–1070, 1991.

[48] Carlangelo Liverani. Central limit theorem for deterministic systems. In *International Conference on Dynamical Systems (Montevideo, 1995)*, volume 362 of *Pitman Res. Notes Math. Ser.*, pages 56–75. Longman, Harlow, 1996.

[49] Carlangelo Liverani, Benoît Saussol, and Sandro Vaienti. A probabilistic approach to intermittency. *Ergodic Theory Dynam. Systems*, 19(3):671–685, 1999.

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

[51] Michael Maxwell and Michael Woodroofe. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.*, 28(2):713–724, 2000.

[52] D. L. McLeish. Dependent central limit theorems and invariance principles. *Ann. Probability*, 2:620–628, 1974.

[53] Ian Melbourne. Large and moderate deviations for slowly mixing dynamical systems. *Proc. Amer. Math. Soc.*, 137(5):1735–1741, 2009.

[54] I. Melbourne. Superpolynomial and polynomial mixing for semiflows and flows. *ArXiv e-prints*, October 2017. To appear in Nonlinearity.

[55] Ian Melbourne and Matthew Nicol. Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys.*, 260(1):131–146, 2005.

[56] Ian Melbourne and Matthew Nicol. Large deviations for nonuniformly hyperbolic systems. *Trans. Amer. Math. Soc.*, 360(12):6661–6676, 2008.

[57] I. Melbourne and A. M. Stuart. A note on diffusion limits of chaotic skew-product flows. *Nonlinearity*, 24(4):1361–1367, 2011.

[58] Ian Melbourne and Andrei Török. Statistical limit theorems for suspension flows. *Israel J. Math.*, 144:191–209, 2004.

[59] Ian Melbourne and Andrei Török. Convergence of moments for axiom A and non-uniformly hyperbolic flows. *Ergodic Theory Dynam. Systems*, 32(3):1091–1100, 2012.

[60] Ian Melbourne and Paulo Varandas. A note on statistical properties for nonuniformly hyperbolic systems with slow contraction and expansion. *Stoch. Dyn.*, 16(3):1660012, 13, 2016.

[61] Ian Melbourne and Roland Zweimüller. Weak convergence to stable Lévy processes for nonuniformly hyperbolic dynamical systems. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):545–556, 2015.

[62] William Parry and Mark Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.

[63] Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale methods*, volume 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.

[64] Yves Pomeau and Paul Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys.*, 74(2):189–197, 1980.

[65] M. Ratner. The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature. *Israel J. Math.*, 16:181–197, 1973.

[66] Emmanuel Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2000.

[67] David Ruelle. *Thermodynamic formalism*, volume 5 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co., Reading, Mass., 1978. The mathematical structures of classical equilibrium statistical mechanics, With a foreword by Giovanni Gallavotti and Gian-Carlo Rota.

[68] David Ruelle. The thermodynamic formalism for expanding maps. *Comm. Math. Phys.*, 125(2):239–262, 1989.

[69] Marek Rychlik. Bounded variation and invariant measures. *Studia Math.*, 76(1):69–80, 1983.

[70] Benoît Saussol. Absolutely continuous invariant measures for multidimensional expanding maps. *Israel J. Math.*, 116:223–248, 2000.

[71] Ja. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.

[72] S. Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73:747–817, 1967.

[73] Marta Tyran-Kamińska. An invariance principle for maps with polynomial decay of correlations. *Comm. Math. Phys.*, 260(1):1–15, 2005.

[74] David R. E. Williams. Path-wise solutions of stochastic differential equations driven by Lévy processes. *Rev. Mat. Iberoamericana*, 17(2):295–329, 2001.

[75] Lai-Sang Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)*, 147(3):585–650, 1998.

[76] Lai-Sang Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999.

[77] Lai-Sang Young. What are SRB measures, and which dynamical systems have them? *J. Statist. Phys.*, 108(5-6):733–754, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[78] Lai-Sang Young. Generalizations of SRB measures to nonautonomous, random, and infinite dimensional systems. *J. Stat. Phys.*, 166(3-4):494–515, 2017.

[79] Roland Zweimüller. Mixing limit theorems for ergodic transformations. *J. Theoret. Probab.*, 20(4):1059–1071, 2007.

I. Chevyrev, Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom

E-mail address: chevyrev@maths.ox.ac.uk

P.K. Friz, Institut für Mathematik, Technische Universität Berlin, and Weierstrass–Institut für Angewandte Analysis und Stochastik, Berlin, Germany

E-mail address: friz@math.tu-berlin.de

A. Korepanov, Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom

E-mail address: a.korepanov@warwick.ac.uk

I. Melbourne, Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom

E-mail address: i.melbourne@warwick.ac.uk

H. Zhang, Institut für Mathematik, Technische Universität Berlin, and Institute of Mathematics, Shandong University, Jinan, China

E-mail address: huilinzhang2014@gmail.com