*The Annals of Probability*
2016, Vol. 44, No. 1, 479–520
DOI: 10.1214/14-AOP979


# SMOOTH APPROXIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

BY DAVID KELLY[1] AND IAN MELBOURNE[2]

*University of North Carolina and University of Warwick*

Consider an Itô process $X$ satisfying the stochastic differential equation $dX = a(X)\,dt + b(X)\,dW$ where $a, b$ are smooth and $W$ is a multidimensional Brownian motion. Suppose that $W_n$ has smooth sample paths and that $W_n$ converges weakly to $W$. A central question in stochastic analysis is to understand the limiting behavior of solutions $X_n$ to the ordinary differential equation $dX_n = a(X_n)\,dt + b(X_n)\,dW_n$.

The classical Wong–Zakai theorem gives sufficient conditions under which $X_n$ converges weakly to $X$ provided that the stochastic integral $\int b(X)\,dW$ is given the Stratonovich interpretation. The sufficient conditions are automatic in one dimension, but in higher dimensions the correct interpretation of $\int b(X)\,dW$ depends sensitively on how the smooth approximation $W_n$ is chosen.

In applications, a natural class of smooth approximations arise by setting $W_n(t) = n^{-1/2}\int_0^{nt} v \circ \phi_s\,ds$ where $\phi_t$ is a flow (generated, e.g., by an ordinary differential equation) and $v$ is a mean zero observable. Under mild conditions on $\phi_t$, we give a definitive answer to the interpretation question for the stochastic integral $\int b(X)\,dW$. Our theory applies to Anosov or Axiom A flows $\phi_t$, as well as to a large class of nonuniformly hyperbolic flows (including the one defined by the well-known Lorenz equations) and our main results do not require any mixing assumptions on $\phi_t$.

The methods used in this paper are a combination of rough path theory and smooth ergodic theory.

**1. Introduction.** Let $X$ be a $d$-dimensional Itô process defined by a stochastic differential equation (SDE) of the form

$$dX = a(X)\,dt + b(X)\,dW, \tag{1.1}$$

where $a : \mathbb{R}^d \to \mathbb{R}^d$ is $C^{1+}$, $b : \mathbb{R}^d \to \mathbb{R}^{d\times e}$ is $C^{2+}$, and $W$ is an $e$-dimensional Brownian motion with $e \times e$-dimensional covariance matrix $\Sigma$.

Given a sequence of $e$-dimensional processes $W_n$ with smooth sample paths, we consider the sequence of ordinary differential equations (ODEs)

$$dX_n = a(X_n)\,dt + b(X_n)\,dW_n, \tag{1.2}$$

Received March 2014; revised October 2014.

[1]Supported in part by ONR Grant N00014-12-1-0257.

[2]Supported in part by European Advanced Grant StochExtHomog (ERC AdG 320977).

*MSC2010 subject classifications.* Primary 60H10; secondary 37D20, 37D25, 37A50.

*Key words and phrases.* Interpretation of stochastic integrals, Wong–Zakai approximation, uniform and nonuniform hyperbolicity, rough paths, iterated invariance principle.

where $dW_n = \dot{W}_n\,dt$. We suppose that an initial condition $\xi \in \mathbb{R}^d$ is fixed throughout and consider solutions $X$ and $X_n$ satisfying $X(0) = X_n(0) = \xi$.

Let $T > 0$. The sequence $W_n$ is said to satisfy the *weak invariance principle (WIP)* if $W_n \to_w W$ in $C([0,T],\mathbb{R}^e)$. Assuming the WIP, a central question in stochastic analysis is to determine whether $X_n \to_w X$ in $C([0,T],\mathbb{R}^d)$ for a suitable interpretation of the stochastic integral $\int b(X)\,dW$ implicit in (1.1). The Wong–Zakai theorem [53] gives general conditions under which convergence holds with the Stratonovich interpretation for the stochastic integral. These conditions are automatically satisfied in the one-dimensional case $d = e = 1$, but may fail in higher dimensions. See also Sussmann [51]. In two dimensions, McShane [31] gave the first counterexamples, and Sussmann [52] provided numerous further counterexamples.

From now on, we replace (1.1) by the SDE

$$dX = a(X)\,dt + b(X) * dW, \tag{1.3}$$

to emphasize the issue with the interpretation of the stochastic integral. General principles suggest that the limiting stochastic integral should be Stratonovich modified by an antisymmetric drift term:

$$b(X) * dW = b(X) \circ dW + \frac{1}{2}\sum_{\alpha,\beta,\gamma} D^{\beta\gamma}\partial^\alpha b^\beta(X) b^{\alpha\gamma}(X)\,dt.$$

Here, and throughout the paper, we sum over $1 \le \alpha \le d$, $1 \le \beta,\gamma \le e$, and $b^{\alpha\gamma}$ and $b^\beta$ denote the $(\alpha,\gamma)$th entry and $\beta$th column, respectively, of $b$. Moreover, $\{D^{\beta\gamma}\}$ is an antisymmetric matrix that is to be determined. [Hence, an alternative to (1.3) would be to consider $dX = \tilde{a}(X)\,dt + b(X) \circ dW$ with the emphasis on determining the correct drift term $\tilde{a}$.]

In applications, smooth processes $W_n$ that approximate Brownian motion arise naturally from differential equations as follows [18, 21, 36, 42, 43]. Let $\phi_t : M \to M$ be a smooth flow on a finite-dimensional manifold $M$ preserving an ergodic measure $\nu$ and let $v : M \to \mathbb{R}^e$ be a smooth observable with $\int_M v\,d\nu = 0$. Define

$$W_n(t) = n^{-1/2}\int_0^{nt} v \circ \phi_s\,ds. \tag{1.4}$$

For large classes of uniformly and nonuniformly hyperbolic flows [11, 20, 33, 35], it can be shown that $W_n$ satisfies the WIP. In this paper, we consider such flows, and give a definitive answer to the question of how to correctly interpret the stochastic integral $\int b(X) * dW$ in order to ensure that $X_n \to_w X$.

*An important special case*. Let $d = e = 2$ and take $a \equiv 0$, $b(x_1,x_2) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & x_1 \end{smallmatrix}\right)$. The ODE (1.2) becomes

$$dX_n^1 = dW_n^1, \qquad dX_n^2 = X_n^1\,dW_n^2,$$

so with the initial condition $\xi = 0$ we obtain $X_n^1 \equiv W_n^1$ and $X_n^2(t) = \int_0^t W_n^1\,dW_n^2$. Weak convergence of $W_n$ to $W$ does not determine the weak limit of $\int_0^t W_n^1\,dW_n^2$.

However, according to rough path theory [29], this is the key obstruction to solving the central problem in this paper. Generally, define the family of smooth processes $\mathbb{W}_n \in C([0,\infty), \mathbb{R}^{e\times e})$,

$$\mathbb{W}_n^{\beta\gamma}(t) = \int_0^t W_n^\beta \, dW_n^\gamma, \qquad 1 \le \beta, \gamma \le e. \tag{1.5}$$

The theory of rough paths implies that under some mild moment estimates, the weak limit of $(W_n, \mathbb{W}_n)$ determines the weak limit of $X_n$ in (1.2) and the correct interpretation for the stochastic integral in (1.3).

Hence, a large part of this paper is dedicated to proving an *iterated WIP* for the pair $(W_n, \mathbb{W}_n)$.

*Anosov and Axiom A flows*. One well-known class of flows to which our results apply is given by the Axiom A (uniformly hyperbolic) flows introduced by Smale [50]. This includes Anosov flows [3]. We do not give the precise definitions, since they are not needed for understanding the paper, but a rough description is as follows. (See [6, 46, 48] for more details.)

Let $\phi_t : M \to M$ be a $C^2$ flow defined on a compact manifold $M$. A flow-invariant subset $\Omega \subset M$ is *uniformly hyperbolic* if for all $x \in \Omega$ there exists a $D\phi_t$-invariant splitting transverse to the flow into uniformly contracting and expanding directions. The flow is *Anosov* if the whole of $M$ is uniformly hyperbolic. More generally, an *Axiom A* flow is characterised by the property that the dynamics decomposes into finitely many hyperbolic equilibria and finitely many uniformly hyperbolic subsets $\Omega_1, \dots, \Omega_k$, called *hyperbolic basic sets*, such that the flow on each $\Omega_i$ is transitive (there is a dense orbit).

If $\Omega$ is a hyperbolic basic set, there is a unique $\phi_t$-invariant ergodic probability measure (called an *equilibrium measure*) associated to each Hölder function on $\Omega$. [In the special case that $\Omega$ is an attractor, there is a distinguished equilibrium measure called the physical measure or SRB measure (after Sinai, Ruelle, Bowen).]

In the remainder of the Introduction, we assume that $\Omega$ is a hyperbolic basic set with equilibrium measure $\nu$ (corresponding to a Hölder potential). We exclude the trivial case where $\Omega$ consists of a single periodic orbit.

We can now state our main results. For $u : \Omega \to \mathbb{R}^q$, we define $\mathbb{E}_\nu(u) \in \mathbb{R}^q$ and $\mathrm{Cov}_\nu(u) \in \mathbb{R}^{q\times q}$ by setting $\mathbb{E}_\nu(u) = \int_\Omega u \, d\nu$ and $\mathrm{Cov}_\nu^{\beta\gamma}(u) = \mathbb{E}_\nu(u^\beta u^\gamma) - \mathbb{E}_\nu(u^\beta)\mathbb{E}_\nu(u^\gamma)$.

THEOREM 1.1 (Iterated WIP). *Suppose that $\Omega \subset M$ is a hyperbolic basic set with equilibrium measure $\nu$ and that $v : \Omega \to \mathbb{R}^e$ is Hölder with $\int_\Omega v \, d\nu = 0$. Define $W_n$ and $\mathbb{W}_n$ as in* (1.4) *and* (1.5). *Then*:

(a) $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ *in* $C([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$ *as* $n \to \infty$, *where*:
  (i) $W$ *is an $e$-dimensional Brownian motion with covariance matrix* $\Sigma = \mathrm{Cov}(W(1)) = \lim_{n\to\infty} \mathrm{Cov}_\nu(W_n(1))$.
  (ii) $\mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta \circ dW^\gamma + \frac{1}{2} D^{\beta\gamma} t$ *where* $D = 2\lim_{n\to\infty} \mathbb{E}_\nu(\mathbb{W}_n(1)) - \Sigma$.

(b) *If in addition the integral* $\int_0^\infty \int_\Omega v^\beta v^\gamma \circ \phi_t \, dt$ *exists for all* $\beta, \gamma$, *then*

$$\Sigma^{\beta\gamma} = \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_t + v^\gamma v^\beta \circ \phi_t) \, d\nu \, dt$$

*and*

$$D^{\beta\gamma} = \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_t - v^\gamma v^\beta \circ \phi_t) \, d\nu \, dt.$$

THEOREM 1.2 (Convergence to SDE). *Suppose that* $\Omega \subset M$ *is a hyperbolic basic set with equilibrium measure* $\nu$ *and that* $v : X \to \mathbb{R}^e$ *is Hölder with* $\int_\Omega v \, d\nu = 0$. *Let* $W_n$, $W$ *and* $D$ *be as in Theorem* 1.1. *Let* $a : \mathbb{R}^d \to \mathbb{R}^d$ *be* $C^{1+}$ *and* $b : \mathbb{R}^d \to \mathbb{R}^{d\times e}$ *be* $C^{2+}$, *and define* $X_n$ *to be the solution of the ODE* (1.2) *with* $X_n(0) = \xi$.

*Then* $X_n \to_w X$ *in* $C([0,\infty), \mathbb{R}^d)$ *as* $n \to \infty$, *where* $X$ *satisfies the SDE*

$$dX = \left\{ a(X) + \frac{1}{2} \sum_{\alpha,\beta,\gamma} D^{\beta\gamma} \partial^\alpha b^\beta(X) b^{\alpha\gamma}(X) \right\} dt + b(X) \circ dW, \qquad X(0) = \xi.$$

*Mixing assumptions on the flow*. The only place where we use mixing assumptions on the flow is in Theorem 1.1(b) to obtain closed form expressions for the diffusion and drift coefficients $\Sigma$ and $D$. In general, these integrals need not converge for Axiom A flows even when $v$ is $C^\infty$.

Dolgopyat [12] proved exponential decay of correlations for Hölder observables $v$ of certain Anosov flows, including geodesic flows on compact negatively curved surfaces. This was extended by Liverani [27] to Anosov flows with a contact structure, including the case of geodesic flows in all dimensions. Theorem 1.1(b) holds for the flows considered in [12, 27]. Nevertheless, for typical Anosov flows, the extra condition in Theorem 1.1(b) is not known to hold for Hölder observables.

Dolgopyat [13] introduced the weaker notion of *rapid mixing*, namely decay of correlations at an arbitrary polynomial rate, and proved that typical Axiom A flows enjoy this property. By [16], an open and dense set of Axiom A flows are rapid mixing. However, this theory applies only to observables $v$ that are sufficiently smooth, and the degree of smoothness is not readily computable. On the positive side, Theorem 1.1(b) holds for typical Axiom A flows provided $v$ is $C^\infty$.

In the absence of a good theory of mixing for flows, we have chosen (as in [36]) to develop our theory in such a way that the dependence on mixing is minimized. Instead we rely on statistical properties of flows, which is a relatively well-understood topic.

A more complicated closed form expression for $\Sigma$ and $D$ that does not require mixing conditions on the flow can be found in Corollary 8.1.

*Beyond uniform hyperbolicity*. In this Introduction, for ease of exposition we have chosen to focus on the case of uniformly hyperbolic flows (Anosov or Axiom A). However, our results hold for large classes of nonuniformly hyperbolic flows. In particular, Young [54] introduces a class of nonuniformly hyperbolic diffeomorphisms, that includes uniformly hyperbolic (Axiom A) diffeomorphisms, as well as Hénon-like attractors [5]. For flows with a Poincaré map that is nonuniformly hyperbolic in the sense of [54], Theorems 1.1 and 1.2 go through unchanged.

The nonuniformly hyperbolic diffeomorphisms in [54] (but not necessarily the corresponding flows) have exponential decay of correlations for Hölder observables. Young [55] considers nonuniformly hyperbolic diffeomorphisms with subexponential decay of correlations. Many of our results go through for flows with a Poincaré map that is nonuniformly hyperbolic in the more general sense of [55]. In particular, our results are valid for the classical Lorenz equations.

These extensions are discussed at length in Section 10.

*Structure of the proofs.* In the smooth ergodic theory literature, there are numerous results on the WIP where $W_n \to_w W$. Usually such results are obtained first for processes $W_n$ arising from a discrete time dynamical system. Results for flows are then obtained as a corollary of the discrete time case, see for example [9, 33, 35, 37, 40, 45]. Hence, it is natural to solve the discrete time analogue of Theorem 1.1 first before extending to continuous time. This is the approach followed in this paper. We first prove the discrete time iterated WIP, Theorem 2.1 below. Then we derive the continuous time WIP, Theorem 1.1, as a consequence, before obtaining Theorem 1.2 using rough path theory. For completeness, we also state and prove the discrete time analogue of Theorem 1.2 (see Theorem 2.2 below), even though this is not required for the proof of Theorem 1.2.

For the proof of the discrete time iterated WIP, it is convenient to use the standard method of passing from invertible maps to noninvertible maps. So we prove the iterated WIP first for noninvertible maps, then for invertible maps, and finally for continuous time systems.

*Structure of the paper*. The remainder of this paper is organized as follows. Sections 2 to 5 deal with the discrete time iterated WIP. Section 2 states our main results for discrete time. In Section 3, we present a result on cohomological invariance of weak limits of iterated processes. This result seems of independent theoretical interest but in this paper it is used to significantly simplify calculations. In Sections 4 and 5, we prove the iterated WIP for discrete time systems that are noninvertible and invertible, respectively.

In Section 6, we return to the case of continuous time and prove a purely probabilistic result about lifting the iterated WIP from discrete time to continuous time. In Section 7, we state and prove some moment estimates that are required to apply rough path theory. In Section 8, we prove the iterated WIP stated in Theorem 1.1. Then in Section 9, we prove Theorem 1.2 and its discrete time analogue.

In Section 10, we discuss various generalizations of our main results that go beyond the Axiom A case. In particular, we consider large classes of systems that are nonuniformly hyperbolic in the sense of [54, 55].

We conclude this Introduction by mentioning related work of Dolgopyat [14], Theorem 5 and [15], Theorem 3(b). These results, which rely on very different techniques from those developed here, prove the analogue of Theorem 1.2 for a class of partially hyperbolic discrete time dynamical systems. The intersection with our work consists of Anosov diffeomorphisms and time-one maps of Anosov flows with better than summable decay of correlations. As discussed above, our main results do not rely on mixing for flows; only the formulas require mixing. Also, we consider the entire Axiom A setting (including Smale horseshoes and flows that possess a horseshoe in the Poincaré map) and our results apply to systems that are nonuniformly hyperbolic in the sense of Young (including Hénon and Lorenz attractors).

*Notation.* As usual, we let $\int b(X)\,dW$ and $\int b(X)\circ dW$ denote the Itô and Stratonovich integrals, respectively.

We use the "big $O$" and $\ll$ notation interchangeably, writing $a_n = O(b_n)$ or $a_n \ll b_n$ if there is a constant $C > 0$ such that $a_n \le C b_n$ for all $n \ge 1$.

## 2. Statement of the main results for discrete time.

In this section, we state the discrete time analogues of our main Theorems 1.1 and 1.2.

Let $f : M \to M$ be a $C^2$ diffeomorphism defined on a compact manifold $M$. Again we focus on the case where $\Lambda \subset M$ is a (nontrivial) hyperbolic basic set with equilibrium measure $\mu$. The definitions are identical to those for Axiom A flows, with the simplification that the direction tangent to the flow is absent. (Hyperbolic basic sets are denoted throughout by $\Omega$ in the flow case described in Section 1 and by $\Lambda$ in the current discrete time setting. The analysis of the flow case includes passing from the hyperbolic basic set $\Omega$ for the flow to a hyperbolic basic set $\Lambda$ for a suitable Poincaré map; hence the need for distinct notation.)

We assume in this section that $\Lambda$ is mixing: $\lim_{n\to\infty}\int_\Lambda w_1 w_2 \circ f^n\,d\mu = \int_\Lambda w_1\,d\mu \int_\Lambda w_2\,d\mu$ for all $w_1, w_2 \in L^2$ (this assumption is relaxed in Section 10).

Let $v : \Lambda \to \mathbb{R}^e$ be Hölder with $\int_\Lambda v\,d\mu = 0$. Define the cadlag processes $W_n \in D([0,\infty),\mathbb{R}^e)$, $\mathbb{W}_n \in D([0,\infty),\mathbb{R}^{e\times e})$,

$$
\begin{aligned}
& W_n(t) = n^{-1/2}\sum_{j=0}^{[nt]-1} v\circ f^j, \\
(2.1)\qquad & \mathbb{W}_n^{\beta\gamma}(t) = \int_0^t W_n^\beta\,dW_n^\gamma = n^{-1}\sum_{0\le i<j\le[nt]-1} v^\beta\circ f^i\, v^\gamma\circ f^j.
\end{aligned}
$$

Since our limiting processes have continuous sample paths, throughout we use the sup-norm topology on $D([0,\infty),\mathbb{R}^e)$ unless otherwise stated.

THEOREM 2.1 (Iterated WIP, discrete time). *Suppose that $\Lambda \subset M$ is a mixing hyperbolic basic set with equilibrium measure $\mu$, and that $v : \Lambda \to \mathbb{R}^e$ is Hölder with $\int_\Lambda v\,d\mu = 0$. Define $W_n$ and $\mathbb{W}_n$ as in (2.1). Then $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ in $D([0,\infty),\mathbb{R}^e \times \mathbb{R}^{e\times e})$ as $n \to \infty$, where:*

(i) *$W$ is an $e$-dimensional Brownian motion with covariance matrix* $\Sigma = \operatorname{Cov}(W(1)) = \lim_{n\to\infty} \operatorname{Cov}_\mu(W_n(1))$ *given by*

$$\Sigma^{\beta\gamma} = \int_\Lambda v^\beta v^\gamma \, d\mu + \sum_{n=1}^\infty \int_\Lambda (v^\beta v^\gamma \circ f^n + v^\gamma v^\beta \circ f^n) \, d\mu.$$

(ii) $\mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta \, dW^\gamma + E^{\beta\gamma} t$ *where* $E = \lim_{n\to\infty} \mathbb{E}_\mu(\mathbb{W}_n(1))$ *is given by*

$$E^{\beta\gamma} = \sum_{n=1}^\infty \int_\Lambda v^\beta v^\gamma \circ f^n \, d\mu.$$

Given $a : \mathbb{R}^d \to \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^{d\times e}$, we define $X_n \in D([0,\infty), \mathbb{R}^d)$, to be the solution to an appropriately discretized version of equation (1.2). Namely, we set $X_n(t) = X_{[nt],n}$ where

$$X_{j+1,n} = X_{j,n} + n^{-1} a(X_{j,n}) + b(X_{j,n}) \left( W_n\left(\frac{j+1}{n}\right) - W_n\left(\frac{j}{n}\right) \right), \qquad X_{0,n} = \xi.$$

THEOREM 2.2 (Convergence to SDE, discrete time). *Suppose that $\Lambda \subset M$ is a mixing hyperbolic basic set with equilibrium measure $\mu$, and that $v : \Lambda \to \mathbb{R}^e$ is Hölder with $\int_\Lambda v \, d\mu = 0$. Let $W_n$, $W$ and $E$ be as in Theorem 2.1. Let $a : \mathbb{R}^d \to \mathbb{R}^d$ be $C^{1+}$ and $b : \mathbb{R}^d \to \mathbb{R}^{d\times e}$ be $C^{2+}$, and define $X_n \in D([0,\infty), \mathbb{R}^d)$ as above.*

*Then $X_n \to_w X$ in $D([0,\infty), \mathbb{R}^d)$ as $n \to \infty$, where $X$ satisfies the SDE*

$$dX = \left\{ a(X) + \sum_{\alpha,\beta,\gamma} E^{\beta\gamma} \partial^\alpha b^\beta(X) b^{\alpha\gamma}(X) \right\} dt + b(X) \, dW, \qquad X(0) = \xi.$$

**3. Cohomological invariance for iterated integrals.** In this section, we present a result which is of independent theoretical interest but which in particular significantly simplifies the subsequent calculations.

Let $f : \Lambda \to \Lambda$ be an invertible or noninvertible map with invariant probability measure $\mu$. Suppose that $v, \hat{v} : \Lambda \to \mathbb{R}^e$ are mean zero observables lying in $L^2$. Define $W_n \in D([0,\infty), \mathbb{R}^e)$ and $\mathbb{W}_n \in D([0,\infty), \mathbb{R}^{e\times e})$ as in (2.1), and similarly define $\widehat{W}_n \in D([0,\infty), \mathbb{R}^e)$ and $\widehat{\mathbb{W}}_n \in D([0,\infty), \mathbb{R}^{e\times e})$ starting from $\hat{v}$ instead of $v$.

We say that $v$ and $\hat{v}$ are *$L^2$-cohomologous* if there exists $\chi : \Lambda \to \mathbb{R}^e$ lying in $L^2$ such that $v = \hat{v} + \chi \circ f - \chi$. It is then easy to see that $W_n$ satisfies the WIP if and only if $\widehat{W}_n$ satisfies the WIP and moreover the weak limits of $W_n$ and $\widehat{W}_n$ coincide. However, the weak limits of $\mathbb{W}_n$ and $\widehat{\mathbb{W}}_n$ need not coincide. The following result supplies the correction factor needed to recover identical weak limits.

THEOREM 3.1. *Suppose that $f : \Lambda \to \Lambda$ is mixing and that $v, \hat{v} \in L^2(\Lambda, \mathbb{R}^e)$ are $L^2$-cohomologous mean zero observables. Let $1 \le \beta, \gamma \le e$. Then the limit*

$\lim_{n\to\infty}\sum_{j=1}^n\int_\Lambda(v^\beta v^\gamma\circ f^j-\hat v^\beta\hat v^\gamma\circ f^j)\,d\mu$ *exists and*

$$\mathbb{W}_n^{\beta\gamma}(t)-\widehat{\mathbb{W}}_n^{\beta\gamma}(t)\to t\sum_{j=1}^\infty\int_\Lambda\bigl(v^\beta v^\gamma\circ f^j-\hat v^\beta\hat v^\gamma\circ f^j\bigr)\,d\mu\qquad a.e.,$$

*as* $n\to\infty$, *uniformly on compact subsets of* $[0,\infty)$.

*In particular, the weak limits of the processes*

$$\mathbb{W}_n^{\beta\gamma}(t)-t\sum_{j=1}^n\int_\Lambda v^\beta v^\gamma\circ f^j\,d\mu,\qquad\widehat{\mathbb{W}}_n^{\beta\gamma}(t)-t\sum_{j=1}^n\int_\Lambda\hat v^\beta\hat v^\gamma\circ f^j\,d\mu,$$

*coincide* (*in the sense that if one limit exists*, *then so does the other and they are equal*).

PROOF. Write $v=\hat v+a$, $a=\chi\circ f-\chi$, and $A_n(t)=n^{-1/2}\sum_{j=0}^{[nt]-1}a\circ f^j$. Then

$$\mathbb{W}_n^{\beta\gamma}(t)-\widehat{\mathbb{W}}_n^{\beta\gamma}(t)=\int_0^t W_n^\beta\,dW_n^\gamma-\int_0^t\widehat{W}_n^\beta\,d\widehat{W}_n^\gamma=\int_0^t A_n^\beta\,dW_n^\gamma+\int_0^t\widehat{W}_n^\beta\,dA_n^\gamma.$$

Now

$$\begin{aligned}\int_0^t A_n^\beta\,dW_n^\gamma&=n^{-1}\sum_{j=0}^{[nt]-1}\sum_{i=0}^{j-1}a^\beta\circ f^i v^\gamma\circ f^j=n^{-1}\sum_{j=0}^{[nt]-1}(\chi^\beta\circ f^j-\chi^\beta)v^\gamma\circ f^j\\&=n^{-1}\sum_{j=0}^{[nt]-1}(\chi^\beta v^\gamma)\circ f^j-n^{-1}\chi^\beta\sum_{j=0}^{[nt]-1}v^\gamma\circ f^j,\end{aligned}$$

which converges to $t\int_\Lambda\chi^\beta v^\gamma\,d\mu$ a.e. by the ergodic theorem.

A similar argument for the remaining term, after changing order of summation yields that $\int_0^t\widehat{W}_n^\beta\,dA_n^\gamma\to-t\int_\Lambda\hat v^\beta\chi^\gamma\circ f\,d\mu$ a.e.

Hence, we have shown that

$$\mathbb{W}_n^{\beta\gamma}(t)-\widehat{\mathbb{W}}_n^{\beta\gamma}(t)\to t\biggl(\int_\Lambda\chi^\beta v^\gamma\,d\mu-\int_\Lambda\hat v^\beta\chi^\gamma\circ f\,d\mu\biggr).\tag{3.1}$$

Next,

$$v^\beta v^\gamma\circ f^j-\hat v^\beta\hat v^\gamma\circ f^j=(\chi^\beta\circ f-\chi^\beta)v^\gamma\circ f^j+\hat v^\beta(\chi^\gamma\circ f-\chi^\gamma)\circ f^j,$$

and so

$$\sum_{j=1}^n\int_\Lambda v^\beta v^\gamma\circ f^j\,d\mu-\sum_{j=1}^n\int_\Lambda\hat v^\beta\hat v^\gamma\circ f^j\,d\mu$$

$$=\sum_{j=1}^n\int_\Lambda\{(\chi^\beta\circ f-\chi^\beta)v^\gamma\circ f^j+\hat v^\beta(\chi^\gamma\circ f-\chi^\gamma)\circ f^j\}\,d\mu\tag{3.2}$$

$$= \sum_{j=1}^{n} \int_\Lambda \{(\chi^\beta \circ f^{n-j+1} - \chi^\beta \circ f^{n-j}) v^\gamma \circ f^n + \hat{v}^\beta (\chi^\gamma \circ f^{j+1} - \chi^\gamma \circ f^j)\} \, d\mu$$

$$= \int_\Lambda \chi^\beta v^\gamma \, d\mu - \int_\Lambda \hat{v}^\beta \chi^\gamma \circ f \, d\mu + L_n,$$

where $L_n = \int_\Lambda (\hat{v}^\beta \chi^\gamma \circ f^{n+1} - \chi^\beta v^\gamma \circ f^n) \, d\mu \to 0$ as $n \to \infty$ by the mixing assumption. The result is immediate from (3.1) and (3.2). $\square$

COROLLARY 3.2. *Let $f : \Lambda \to \Lambda$ be mixing and let $v, \hat{v} \in L^2(\Lambda, \mathbb{R}^e)$ be $L^2$-cohomologous mean zero observables.*

*Suppose that $(\widehat{W}_n, \widehat{\mathbb{W}}_n) \to_w (\widehat{W}, \widehat{\mathbb{W}})$ in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$ as $n \to \infty$. Then $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$ as $n \to \infty$, where $W = \widehat{W}$ and*

$$\mathbb{W}^{\beta\gamma}(t) = \widehat{\mathbb{W}}^{\beta\gamma}(t) + t \sum_{j=1}^{\infty} \int_\Lambda \big(v^\beta v^\gamma \circ f^j \, d\mu - \hat{v}^\beta \hat{v}^\gamma \circ f^j\big) \, d\mu.$$

REMARK 3.3. For completeness, we describe the analogous result for semiflows. Again the result is of independent theoretical interest even though we make no use of it in this paper.

Let $\phi_t : \Omega \to \Omega$ be a mixing (semi)flow with invariant probability measure $\nu$. Suppose that $v, \hat{v} : \Omega \to \mathbb{R}^e$ are mean zero observables lying in $L^2$. Define $W_n$ and $\mathbb{W}_n$ as in (1.4) and (1.5), and similarly define $\widehat{W}_n$ and $\widehat{\mathbb{W}}_n$ starting from $\hat{v}$ instead of $v$.

We say that $v$ and $\hat{v}$ are *$L^2$-cohomologous* if there exists $\chi : \Omega \to \mathbb{R}^e$ lying in $L^2$ such that $\int_0^t v \circ \phi_s \, ds = \int_0^t \hat{v} \circ \phi_s \, ds + \chi \circ \phi_t - \chi$. Again, $W_n$ satisfies the WIP if and only if $\widehat{W}_n$ satisfies the WIP and the weak limits coincide. As in Theorem 3.1, we find that the limit $\lim_{n\to\infty} \int_0^n \int_\Omega (v^\beta v^\gamma \circ \phi_s - \hat{v}^\beta \hat{v}^\gamma \circ \phi_s) \, d\nu$ exists and

$$\mathbb{W}_n^{\beta\gamma}(t) - \widehat{\mathbb{W}}_n^{\beta\gamma}(t) \to t \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_s - \hat{v}^\beta \hat{v}^\gamma \circ \phi_s) \, d\nu \, ds \qquad \text{a.e.},$$

as $n \to \infty$, uniformly on compact subsets of $[0, \infty)$. The proof is almost identical to that of Theorem 3.1, and hence is omitted.

**4. Iterated WIP for noninvertible maps.** A sufficient condition for Theorem 2.1 is that $f : \Lambda \to \Lambda$ is a mixing uniformly expanding map. More generally, in this section we consider a class of nonuniformly expanding maps with sufficiently rapid decay of correlations. The underlying hypotheses can be satisfied only by noninvertible maps; see Section 5 for more general hypotheses appropriate for invertible maps.

In Section 4.1 we give more details on the class of maps that is considered in this section. In Section 4.2, we prove the iterated WIP for these maps.

4.1. *Noninvertible maps.* Let $f:\Lambda\to\Lambda$ be an ergodic measure-preserving map defined on a probability space $(\Lambda,\mu)$ and let $v:\Lambda\to\mathbb{R}^d$ be an integrable observable with $\int_\Lambda v\,d\mu=0$. Let $P:L^1(\Lambda)\to L^1(\Lambda)$ be the transfer operator for $f$ given by $\int_\Lambda Pw_1w_2\,d\mu=\int_\Lambda w_1Uw_2\,d\mu$ for $w_1\in L^1(\Lambda)$, $w_2\in L^\infty(\Lambda)$ where $Uw=w\circ f$.

DEFINITION 4.1. Let $p\ge 1$. We say that $v$ *admits an $L^p$ martingale-coboundary decomposition* if there exists $m,\chi\in L^p(\Lambda,\mathbb{R}^e)$ such that

$$v=m+\chi\circ f-\chi,\qquad m\in\ker P. \tag{4.1}$$

We refer to $m$ as the *martingale part* of the decomposition.

REMARK 4.2. The reason for calling $m$ a martingale will become clearer in Section 4.2. For the time being, we note that it is standard and elementary that $PU=I$ and $UP=E(\cdot|f^{-1}\mathcal{B})$ where $\mathcal{B}$ is the underlying $\sigma$-algebra. In particular $E(m|f^{-1}\mathcal{B})=0$.

Our main result in this section is the following.

THEOREM 4.3. *Suppose that $f$ is mixing and that the decomposition* (4.1) *holds with $p=2$. Then the conclusion of Theorem* 2.1 *is valid.*

PROPOSITION 4.4. *Let $p\ge 1$. A sufficient condition for* (4.1) *to hold is that $v\in L^\infty$ and there are constants $C>0$, $\tau>p$ such that*

$$\left|\int_\Lambda vw\circ f^n\,d\mu\right|\le C\|w\|_\infty n^{-\tau}\qquad \textit{for all } w\in L^\infty, n\ge 1. \tag{4.2}$$

PROOF. By duality, $\|P^nv\|_1\le Cn^{-\tau}$. Also, $\|P^nv\|_\infty\le\|v\|_\infty$ and it follows that $\|P^nv\|_p\le\|v\|_\infty^{1-1/p}(Cn^{-\tau})^{1/p}$ which is summable.

Define $\chi=\sum_{n=1}^\infty P^nv\in L^p$, and write $v=m+\chi\circ f-\chi$ where $m\in L^p$. Applying $P$ to both sides and using the fact that $PU=I$, we obtain that $m\in\ker P$. $\square$

There are large classes of noninvertible maps for which the decay condition (4.2) has been established for sufficiently regular $v$; see Section 10. In particular, for uniformly expanding maps the decay is exponential for Hölder continuous $v$, so $\tau$ and $p$ can be chosen arbitrarily large.

In the remainder of this subsection, we reduce Theorem 4.3 to the martingale part. Define the cadlag processes $M_n\in D([0,\infty),\mathbb{R}^e)$, $\mathbb{M}_n\in D([0,\infty),\mathbb{R}^{e\times e})$,

$$M_n(t)=n^{-1/2}\sum_{j=0}^{[nt]-1}m\circ f^j,$$

$$\mathbb{M}_n^{\beta\gamma}(t)=\int_0^t M_n^\beta\,dM_n^\gamma=n^{-1}\sum_{0\le i<j\le[nt]-1}m^\beta\circ f^im^\gamma\circ f^j.$$

Theorem 4.3 follows from the following lemma.

LEMMA 4.5. *Suppose that $f$ is ergodic and that $m \in L^2(\Lambda, \mathbb{R}^e)$ with $Pm = 0$. Then $(M_n, \mathbb{M}_n) \to_w (W, I)$ in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$, as $n \to \infty$, where $W$ is an $e$-dimensional Brownian motion with covariance matrix* $\operatorname{Cov}(W(1)) = \int_\Lambda mm^T\,d\mu$ *and $I^{\beta\gamma}(t) = \int_0^t W^\beta\,dW^\gamma$.*

PROOF OF THEOREM 4.3. We apply Corollary 3.2 with $\hat{v} = m$. Note that $\int_\Lambda mm^T \circ f^j\,d\mu = \int_\Lambda P^j mm^T\,d\mu = 0$ for all $j \geq 1$. By Theorem 3.1, $E = \sum_{j=1}^\infty vv^T \circ f^j\,d\mu$ is a convergent series. By Corollary 3.2, $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ where $\Sigma = \operatorname{Cov}(W(1)) = \int_\Lambda mm^T\,d\mu$ and $\mathbb{W}(t) = I(t) + Et$.

It remains to prove that $\Sigma^{\beta\gamma} = \lim_{n\to\infty} \operatorname{Cov}_\mu^{\beta\gamma}(W_n(1)) = \int_\Lambda v^\beta v^\gamma\,d\mu + \sum_{n=1}^\infty \int_\Lambda (v^\beta v^\gamma \circ f^n + v^\gamma v^\beta \circ f^n)\,d\mu$ and that $E = \lim_{n\to\infty} \mathbb{E}_\mu(\mathbb{W}_n(1))$.

Define $v_n = \sum_{j=0}^{n-1} v \circ f^j$, $m_n = \sum_{j=0}^{n-1} m \circ f^j$. Then

$$\int_\Lambda m_n m_n^T\,d\mu = \sum_{0\leq i,j\leq n-1} \int_\Lambda m \circ f^i (m \circ f^j)^T\,d\mu = n\Sigma.$$

Equivalently, $c^T\Sigma c = n^{-1}\int_\Lambda (c^T m_n)^2\,d\mu$ for all $c \in \mathbb{R}^e$, $n \geq 1$. Let $\|\cdot\|_2$ denote the $L^2$ norm on $(\Lambda, \mu)$. We have that $n^{1/2}(c^T\Sigma c)^{1/2} = \|c^T m_n\|_2$. By (4.1), $v_n - m_n = \chi \circ f^n - \chi$. Using $f$-invariance of $\mu$,

$$\begin{aligned} \big|\|c^T v_n\|_2 - n^{1/2}(c^T\Sigma c)^{1/2}\big| &= \big|\|c^T v_n\|_2 - \|c^T m_n\|_2\big| \leq \|c^T(v_n - m_n)\|_2 \\ &\leq 2\|c^T\chi\|_2, \end{aligned}$$

and hence $\lim_{n\to\infty} n^{-1/2}\|c^T v_n\|_2 = (c^T\Sigma c)^{1/2}$. Equivalently,

$$\Sigma = \lim_{n\to\infty} n^{-1}\int_\Lambda v_n v_n^T\,d\mu = \lim_{n\to\infty} \operatorname{Cov}_\mu(W_n(1)). \tag{4.3}$$

Let $a_r = \int_\Lambda v \circ f^r v^T\,d\mu$ and $s_k = \sum_{r=1}^k a_r$. Compute that

$$\begin{aligned} \sum_{0\leq j<i\leq n-1} \int_\Lambda v \circ f^{i-j} v^T\,d\mu &= \sum_{1\leq r<n} (n-r)\int_\Lambda v \circ f^r v^T\,d\mu \\ &= \sum_{1\leq r<n} (n-r)a_r = \sum_{k=1}^n s_k. \end{aligned}$$

Hence,

$$\begin{aligned} \lim_{n\to\infty} n^{-1} \sum_{0\leq j<i\leq n-1} \int_\Lambda v \circ f^{i-j} v^T\,d\mu &= \lim_{n\to\infty} n^{-1}\sum_{k=1}^n s_k = \lim_{n\to\infty} s_n \\ &= \sum_{r=1}^\infty \int_\Lambda v \circ f^r v^T\,d\mu. \end{aligned} \tag{4.4}$$

Similarly,

$$(4.5)\qquad \lim_{n\to\infty} n^{-1} \sum_{0\le i<j\le n-1} \int_\Lambda v(v\circ f^{j-i})^T\,d\mu = \sum_{r=1}^\infty \int_\Lambda v(v\circ f^r)^T\,d\mu.$$

Write

$$\begin{aligned} n^{-1}\int_\Lambda v_n v_n^T\,d\mu &= n^{-1}\sum_{0\le i,j\le n-1}\int_\Lambda v\circ f^i(v\circ f^j)^T\,d\mu \\ &= \int_\Lambda vv^T\,d\mu + n^{-1}\sum_{0\le j<i\le n-1}\int_\Lambda v\circ f^{i-j}v^T\,d\mu \\ &\quad + n^{-1}\sum_{0\le i<j\le n-1}\int_\Lambda v(v\circ f^{j-i})^T\,d\mu. \end{aligned}$$

By (4.3), (4.4), (4.5), $\Sigma = \int_\Lambda vv^T\,d\mu + \sum_{r=1}^\infty \int_\Lambda (v\circ f^r v^T + v(v\circ f^r)^T)\,d\mu$.

Finally, $\mathbb{E}_\mu(\mathbb{W}_n(1)) = n^{-1}\sum_{0\le i<j\le n-1}\int_\Lambda v(v\circ f^{j-i})^T\,d\mu$, so it follows from (4.5) that $\lim_{n\to\infty}\mathbb{E}_\mu(\mathbb{W}_n(1)) = E$. $\square$

4.2. *Proof of Lemma* 4.5.

REMARK 4.6. The $M_n \to_w W$ part of Lemma 4.5 is standard but we give the proof for completeness. The statement can be obtained from the proof of Lemma 4.5 by ignoring the $\mathbb{M}_n$ component. In particular, our use of this fact in the proof of Lemma 4.8 below is not circular.

Recall that $m$ is $\mathcal{B}$-measurable and $m\in\ker P$ so $E(m|f^{-1}\mathcal{B})=0$. Similarly, $m\circ f^j$ is $f^{-j}\mathcal{B}$-measurable and $E(m\circ f^j|f^{-(j+1)}\mathcal{B}) = E(m|f^{-1}\mathcal{B})\circ f^j = 0$. If the sequence of $\sigma$-algebras $f^{-j}\mathcal{B}$ formed a filtration, then $M_n$ would be a martingale and we could apply Kurtz and Protter [24], Theorem 2.2 (see also [22]) to obtain a limit for $(M_n,\mathbb{M}_n)$.

In fact, the $\sigma$-algebras are decreasing: $f^{-j}\mathcal{B}\supset f^{-(j+1)}\mathcal{B}$ for all $j$. To remedy this, we pass to the natural extension $\tilde f:\tilde\Lambda\to\tilde\Lambda$. This is an invertible map with ergodic invariant measure $\tilde\mu$, and there is a measurable projection $\pi:\tilde\Lambda\to\Lambda$ such that $\pi\tilde f = f\pi$ and $\pi_*\tilde\mu=\mu$. The observable $m:\Lambda\to\mathbb{R}^e$ lifts to an observable $\tilde m = m\circ\pi:\tilde\Lambda\to\mathbb{R}^e$ and the joint distributions of $\{m\circ f^j : j\ge 0\}$ are identical to those of $\{\tilde m\circ\tilde f^j : j\ge 0\}$.

Define

$$\widetilde M_n(t) = n^{-1/2}\sum_{j=0}^{[nt]-1}\tilde m\circ\tilde f^j,$$

$$\widetilde{\mathbb{M}}_n^{\beta\gamma}(t) = \int_0^t \widetilde M_n^\beta\,d\widetilde M_n^\gamma = n^{-1}\sum_{0\le i<j\le[nt]-1}\tilde m^\beta\circ\tilde f^i\,\tilde m^\gamma\circ\tilde f^j.$$

Then $(\widetilde{M}_n, \widetilde{\mathbb{M}}_n) = (M_n, \mathbb{M}_n) \circ \pi$ and $\pi$ is measure preserving, so it is equivalent to prove that

$$(\widetilde{M}_n, \widetilde{\mathbb{M}}_n) \to_w (W, I) \qquad \text{in } \mathrm{D}\big([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e}\big). \tag{4.6}$$

Let $\widetilde{\mathcal{B}} = \pi^{-1}\mathcal{B}$. Again $\tilde{f}^{-j}\widetilde{\mathcal{B}} \supset \tilde{f}^{-(j+1)}\widetilde{\mathcal{B}}$ but this means that $\{\mathcal{F}_j, j \ge 1\} = \{\tilde{f}^j\widetilde{\mathcal{B}}, j \ge 1\}$ is an *increasing* sequence of $\sigma$-algebras. Moreover, $\tilde{m} \circ \tilde{f}^{-j}$ is $\mathcal{F}_j$-measurable and $E(\tilde{m} \circ \tilde{f}^{-j} | \mathcal{F}_{j-1}) = 0$. Hence, the "backward" process

$$\widetilde{M}_n^-(t) = n^{-1/2} \sum_{j=-[nt]}^{-1} \tilde{m} \circ \tilde{f}^j$$

forms an ergodic stationary martingale. Similarly, define

$$\widetilde{\mathbb{M}}_n^{\beta\gamma,-}(t) = \int_0^t \widetilde{M}_n^{\beta,-}\, d\widetilde{M}_n^{\gamma,-} = n^{-1} \sum_{[-nt] \le j < i \le -1} \tilde{m}^\beta \circ \tilde{f}^i \tilde{m} \circ \tilde{f}^j.$$

Note that $\int_{\tilde{\Lambda}} \tilde{m}\tilde{m}^T\, d\tilde{\mu} = \int_\Lambda m m^T\, d\mu$.

PROPOSITION 4.7. *$(\widetilde{M}_n^-, \widetilde{\mathbb{M}}_n^-) \to_w (W, I)$ in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$ as $n \to \infty$.*

PROOF. We verify the hypotheses of Kurtz and Protter [24], Theorem 2.2 (with $\delta = \infty$ and $A_n \equiv 0$). We have already seen that $\widetilde{M}_n^-$ is a martingale. Also, by the calculation in the proof of Theorem 4.3, $E(\widetilde{M}_n^{\gamma,-}(t)^2) = n^{-1}\| \sum_{j=1}^{[nt]} \tilde{m}^\gamma \circ \tilde{f}^{-j}\|_2^2 = t\int_{\tilde{\Lambda}} (\tilde{m}^\gamma)^2\, d\tilde{\mu}$ independent of $n$, so condition C2.2(i) in [24], Theorem 2.2, is trivially satisfied.

The WIP for stationary ergodic $L^2$ martingales (e.g., [8, 30]) implies that $\widetilde{M}_n^- \to_w W$ in $D([0,\infty), \mathbb{R}^e)$. In particular, $(\widetilde{M}_n^{\beta,-}, \widetilde{M}_n^{\gamma,-}) \to_w (W^\beta, W^\gamma)$ in $D([0,\infty), \mathbb{R}^2)$. Hence, the result follows from [24], Theorem 2.2. $\square$

It remains to relate weak convergence of $(\widetilde{M}_n^-, \widetilde{\mathbb{M}}_n^-)$ and $(\widetilde{M}_n, \widetilde{\mathbb{M}}_n)$. It suffices to work in $D([0,T], \mathbb{R}^e \times \mathbb{R}^{e\times e})$ for each fixed integer $T \ge 1$.

LEMMA 4.8. *Let $g(u)(t) = u(T) - u(T-t)$ and $h(u,v)(t) = u(T-t)(v(T) - v(T-t))$. Let $*$ denote matrix transpose in $\mathbb{R}^{e\times e}$. Then*

$$(\widetilde{M}_n, \widetilde{\mathbb{M}}_n) \circ \tilde{f}^{-nT} = \big(g(\widetilde{M}_n^-), \big(g(\widetilde{\mathbb{M}}_n^-) - h(\widetilde{M}_n^-)\big)^*\big) + F_n,$$

*where* $\sup_{t\in[0,T]} F_n(t) \to 0$ *a.e.*

PROOF. In this proof, we suppress the tildes. First, we show that $M_n \circ f^{-nT} = g(M_n^-) + F_n^0$, where $\sup_{t\in[0,T]} F_n^0(t) \to 0$ a.e.

We have

$$M_n(t)\circ f^{-nT} = n^{-1/2}\sum_{j=0}^{[nt]-1} m\circ f^j\circ f^{-nT} = n^{-1/2}\sum_{j=-nT}^{[nt]-1-nT} m\circ f^j$$
$$= M_n^-(T) - M_n^-(T-t) + F_n^0(t).$$

Here, $F_n^0$ consists of at most one term and we can write

$$\left|F_n^0(t)\right| \le n^{-1/2}\Big|\max_{j=1,\dots,nT} m\circ f^{-j}\Big|.$$

It suffices to work componentwise, so suppose without loss that $e=1$. By the ergodic theorem, $n^{-1}\sum_{j=1}^n m^2\circ f^{-j}\to\int_\Lambda m^2\,d\mu$, and so $n^{-1}m^2\circ f^{-n}\to 0$. It follows that $n^{-1}\max_{j=0,\dots,nT} m^2\circ f^{-j}\to 0$ a.e. and so $\sup_{t\in[0,T]} F_n^0(t)\to 0$ a.e.

Next, we show that $\mathbb{M}_n\circ f^{-nT} = (g(\mathbb{M}_n^-) - h(M_n^-))^* + F_n$, where $\sup_{t\in[0,T]} F_n(t)\to 0$ a.e. We have

$$\mathbb{M}_n^{\beta\gamma}(t) = n^{-1}\sum_{j=0}^{[nt]-1}\left(\sum_{i=0}^{j-1} m^\beta\circ f^i\right)m^\gamma\circ f^j,$$

$$\mathbb{M}_n^{\beta\gamma,-}(t) = n^{-1}\sum_{j=-[nt]+1}^{-1}\left(\sum_{i=[-nt]}^{j-1} m^\gamma\circ f^i\right)m^\beta\circ f^j.$$

Hence,

$$\begin{aligned}
&\mathbb{M}_n^{\beta\gamma}(t)\circ f^{-nT}\\
&\quad= n^{-1}\sum_{j=-nT}^{[nt]-1-nT}\ \sum_{i=-nT}^{j-1} m^\beta\circ f^i m^\gamma\circ f^j\\
&\quad= n^{-1}\left(\sum_{j=-nT}^{-nT}+\sum_{j=-nT+1}^{-1}-\sum_{j=[nt]-nT+1}^{-1}-\sum_{j=[nt]-nT}^{[nt]-nT}\right)\sum_{i=-nT}^{j-1} m^\beta\circ f^i m^\gamma\circ f^j\\
&\quad= F_n^1(t)+\mathbb{M}_n^{\gamma\beta,-}(T)-E_n(t)-F_n^2(t),
\end{aligned}\tag{4.7}$$

where

$$F_n^1(t) = n^{-1}\sum_{i=-nT}^{-nT-1} m^\beta\circ f^i m^\gamma\circ f^{-nT},$$

$$F_n^2(t) = \left(n^{-1/2}\sum_{i=-nT}^{[nt]-nT-1} m^\beta\circ f^i\right)\left(n^{-1/2}m^\gamma\circ f^{[nt]-nT}\right),$$

$$E_n(t) = n^{-1}\sum_{j=[nt]-nT+1}^{-1}\ \sum_{i=-nT}^{j-1} m^\beta\circ f^i m^\gamma\circ f^j.$$

Now $F_n^1(t)$ consists of only two terms and clearly converges to 0 almost everywhere. The first factor in $F_n^2$ converges weakly to $W^\beta$ (see Remark 4.6) and the second factor converges to 0 almost everywhere by the ergodic theorem. Hence, $\sup_{t\in[0,T]} Z|F_n^r(t)| \to 0$ a.e. for $r = 1, 2$. Moreover,

$$
\begin{aligned}
E_n(t) &= n^{-1} \sum_{j=[nt]-nT+1}^{-1} \left( \sum_{i=-nT}^{-nT+[nt]-1} + \sum_{i=-nT+[nt]}^{j-1} \right) m^\beta \circ f^i m^\gamma \circ f^j \\
&= H_n(t) + \mathbb{M}_n^{\gamma\beta,-}(T-t) + F_n^3(t),
\end{aligned}
\tag{4.8}
$$

where

$$
\begin{aligned}
H_n(t) &= \left( n^{-1/2} \sum_{j=[nt]-nT}^{-1} m^\gamma \circ f^j \right) \left( n^{-1/2} \sum_{i=-nT}^{-nT+[nt]-1} m^\beta \circ f^i \right) \\
&= M_n^{\gamma,-}(T-t)\big(M_n^{\beta,-}(T) - M_n^{\beta,-}(T-t)\big),
\end{aligned}
\tag{4.9}
$$

and $F_n^3(t) = n^{-1} \sum_{i=-nT}^{-nT+[nt]-1} m^\beta \circ f^i m^\gamma \circ f^{[nt]-nT+1}$. Again, $\sup_{t\in[0,T]} |F_n^3(t)| \to 0$ a.e. by the ergodic theorem. The result follows from (4.7), (4.8), (4.9). $\square$

PROPOSITION 4.9. *Let $\widetilde{D}([0,T], \mathbb{R}^q)$ denote the space of caglad functions from $[0,T]$ to $\mathbb{R}^q$ with the standard Skorokhod $\mathcal{J}_1$ topology. Suppose that $A_n = B_n + F_n$ where $A_n \in D([0,T], \mathbb{R}^q)$, $B_n \in \widetilde{D}([0,T], \mathbb{R}^q)$, and $F_n \to 0$ uniformly in probability. If $Z$ has continuous sample paths and $B_n \to_w Z$ in $\widetilde{D}([0,T], \mathbb{R}^q)$, then $A_n \to_w Z$ in $D([0,T], \mathbb{R}^q)$.*

PROOF. It is clear that the limiting finite distributions of $A_n$ coincide with those of $B_n$, so it suffices to show that $A_n$ inherits tightness from $B_n$. One way to see this is to consider the following Arzela–Ascoli-type characterization [49], valid in both $D([0,T], \mathbb{R}^q)$ and $\widetilde{D}([0,T], \mathbb{R}^q)$.

Tightness of $B_n$ in $\widetilde{D}([0,T], \mathbb{R}^q)$ implies that for any $\varepsilon > 0$, $k \ge 1$, there exists $C > 0$, $\delta_k > 0$, $n_k \ge 1$ such that $P(|B_n|_\infty > C) < \varepsilon$ for all $n \ge 1$ and $P(\omega(B_n, \delta_k) > 1/k) < \varepsilon$ for all $n \ge n_k$, where

$$
\omega(\psi, \delta) = \sup_{t-\delta<t'<t<t''<t+\delta} \min\{|\psi(t) - \psi(t')|, |\psi(t) - \psi(t'')|\}
$$

(where $t, t', t''$ are restricted to $[0,T]$). These criteria are also satisfied by $F_n$ for trivial reasons, and hence by $A_n$ establishing tightness of $A_n$ in $D([0,T], \mathbb{R}^q)$. $\square$

COROLLARY 4.10. *$(\widetilde{M}_n, \widetilde{\mathbb{M}}_n) \to_w (g(W), (g(I) - h(W))^*)$ in $D([0,T], \mathbb{R}^e \times \mathbb{R}^{e\times e})$ as $n \to \infty$.*

PROOF. Recalling the notation from Lemma 4.8, observe that the functional $\chi : D([0,T], \mathbb{R}^e \times \mathbb{R}^{e\times e}) \to \widetilde{D}([0,T], \mathbb{R}^e \times \mathbb{R}^{e\times e})$ given by $\chi(u,v) = (g(u), (g(v) - h(u))^*)$ is continuous. Hence, it follows from Proposition 4.7 and the continuous mapping theorem that $(g(\widetilde{M}_n^-), (g(\widetilde{\mathbb{M}}_n^-) - h(\widetilde{M}_n^-))^*) \to_w (g(W), (g(I) - h(W))^*)$ in $\widetilde{D}([0,T], \mathbb{R}^e \times \mathbb{R}^{e\times e})$. The result is now immediate from Lemma 4.8 and Proposition 4.9. $\square$

LEMMA 4.11. $(g(W), (g(I) - h(W))^*) =_d (W, I)$ *in* $D([0,T], \mathbb{R}^e \times \mathbb{R}^{e\times e})$.

PROOF. *Step* 1. $g(W) =_d W$ in $D([0,T], \mathbb{R}^e)$. To see this, note that both processes are Gaussian with continuous sample paths and $g(W)(0) = W(0) = 0$. One easily verifies that $\mathrm{Cov}(g(W)(t_1), g(W)(t_2)) = t_1\Sigma$ for all $0 \le t_1 \le t_2 \le T$. Hence, $g(W) =_d W$.

*Step* 2. Introduce the process $J(t) = \int_0^t g(W)\,dg(W)$. We claim that $(g(W), J) =_d (W, I)$. To see this, let $Y_n(t) = \sum_{j=0}^{[nt]-1} W(j/n)(W((j+1)/n) - W(j/n))$ so $(W, Y_n) \to_w (W, I)$. Similarly, let $Z_n(t) = \sum_{j=0}^{[nt]-1} g(W)(j/n)(g(W)((j+1)/n) - g(W)(j/n))$ so $(g(W), Z_n) \to_w (g(W), J)$. It is clear that $(W, Y_n) =_d (g(W), Z_n)$ so the claim follows.

*Step* 3. We complete the proof by showing that $J = (g(I) - h(W))^*$. Let $1 \le \beta, \gamma \le e$. We show that $g(I)^{\beta\gamma} - h(W)^{\beta\gamma} = J^{\gamma\beta}$.

Now $J^{\gamma\beta}(t) = \int_0^t g(W)^\gamma\, dg(W)^\beta = \lim_{n\to\infty} S_n$ where the limit is in probability and

$$
\begin{aligned}
S_n &= \sum_{k=0}^{[nt]-1} g(W)^\gamma\Big(\frac{k}{n}\Big)\Big(g(W)^\beta\Big(\frac{k+1}{n}\Big) - g(W)^\beta\Big(\frac{k}{n}\Big)\Big) \\
&= \sum_{k=0}^{[nt]-1} \Big(W^\gamma(T) - W^\gamma\Big(T - \frac{k}{n}\Big)\Big) \\
&\qquad\qquad \times \Big(W^\beta\Big(T - \frac{k}{n}\Big) - W^\beta\Big(T - \frac{k+1}{n}\Big)\Big) \\
&= \sum_{k=0}^{[nt]-1}\sum_{j=0}^{k-1} \Big(W^\gamma\Big(T - \frac{j}{n}\Big) - W^\gamma\Big(T - \frac{j+1}{n}\Big)\Big) \\
&\qquad\qquad \times \Big(W^\beta\Big(T - \frac{k}{n}\Big) - W^\beta\Big(T - \frac{k+1}{n}\Big)\Big) \\
&= \sum_{j=0}^{[nt]-2}\sum_{k=j+1}^{[nt]-1} \Big(W^\beta\Big(T - \frac{k}{n}\Big) - W^\beta\Big(T - \frac{k+1}{n}\Big)\Big) \\
&\qquad\qquad \times \Big(W^\gamma\Big(T - \frac{j}{n}\Big) - W^\gamma\Big(T - \frac{j+1}{n}\Big)\Big)
\end{aligned}
$$

$$= \sum_{j=0}^{[nt]-2} \Big( W^{\beta}\Big(T - \frac{j+1}{n}\Big) - W^{\beta}\Big(T - \frac{[nt]}{n}\Big)\Big) \times \Big( W^{\gamma}\Big(T - \frac{j}{n}\Big) - W^{\gamma}\Big(T - \frac{j+1}{n}\Big)\Big).$$

On the other hand, $\{g(I) - h(W)\}^{\beta\gamma}(t) = \int_{T-t}^{T} (W^{\beta} - W^{\beta}(T-t))\, dW^{\gamma} = \lim_{n\to\infty} T_n$ where

$$\begin{aligned} T_n &= \sum_{i=[n(T-t)]}^{nT-1} \Big( W^{\beta}\Big(\frac{i}{n}\Big) - W^{\beta}(T-t)\Big)\Big( W^{\gamma}\Big(\frac{i+1}{n}\Big) - W^{\gamma}\Big(\frac{i}{n}\Big)\Big) \\ &= \sum_{j=0}^{-[-nt]-1} \Big( W^{\beta}\Big(T - \frac{j+1}{n}\Big) - W^{\beta}(T-t)\Big) \\ &\qquad \times \Big( W^{\gamma}\Big(T - \frac{j}{n}\Big) - W^{\gamma}\Big(T - \frac{j+1}{n}\Big)\Big). \end{aligned}$$

We claim that $\lim_{n\to\infty}(T_n - S_n) = 0$ a.e. from which the result follows. When $nt$ is an integer, $S_n = T_n$. Otherwise, $T_n - S_n = A_n + B_n$ where

$$\begin{aligned} A_n &= \sum_{j=0}^{[nt]-2} \Big( W^{\beta}(T-t) - W^{\beta}\Big(T - \frac{[nt]}{n}\Big)\Big)\Big( W^{\gamma}\Big(T - \frac{j}{n}\Big) - W^{\gamma}\Big(T - \frac{j+1}{n}\Big)\Big) \\ &= \Big( W^{\beta}(T-t) - W^{\beta}\Big(T - \frac{[nt]}{n}\Big)\Big)\Big( W^{\gamma}(T) - W^{\gamma}\Big(T - \Big(\frac{[nt]-1}{n}\Big)\Big)\Big) \end{aligned}$$

and

$$\begin{aligned} B_n &= \Big( W^{\beta}\Big(T - \Big(\frac{[nt]+1}{n}\Big)\Big) - W^{\beta}(T-t)\Big) \\ &\qquad \times \Big( W^{\gamma}\Big(T - \frac{[nt]}{n}\Big) - W^{\gamma}\Big(T - \Big(\frac{[nt]+1}{n}\Big)\Big)\Big). \end{aligned}$$

The claim follows since $A_n \to 0$ and $B_n \to 0$ as $n \to \infty$. $\square$

PROOF OF LEMMA 4.5. This follows from Corollary 4.10 and Lemma 4.11. $\square$

**5. Iterated WIP for invertible maps.** In this section, we prove an iterated WIP for invertible maps, and as a special case we prove Theorem 2.1.

For an invertible map $f : \Lambda \to \Lambda$, the transfer operator $P$ is an isometry on $L^p$ for all $p$, so the hypotheses used in Section 4 are not applicable. We require the following more general setting.

Suppose that in addition to the underlying probability space $(\Lambda, \mu)$ and measure-preserving map $f : \Lambda \to \Lambda$, there is an additional probability space $(\bar{\Lambda}, \bar{\mu})$ and measure-preserving map $\bar{f} : \bar{\Lambda} \to \bar{\Lambda}$, and there is a semiconjugacy

$\pi : \Lambda \to \bar{\Lambda}$ with $\pi_* \mu = \bar{\mu}$ such that $\pi \circ f = \bar{f} \circ \pi$. (The system on $\bar{\Lambda}$ is called a factor of the system on $\Lambda$.) We let $P$ denote the transfer operator for $\bar{f}$.

DEFINITION 5.1. Let $v : \Lambda \to \mathbb{R}^e$ be of mean zero and let $p \geq 1$. We say that $v$ *admits an $L^p$ martingale-coboundary decomposition* if there exists $m, \chi \in L^p(\Lambda, \mathbb{R}^e)$, $\bar{m} \in L^p(\bar{\Lambda}, \mathbb{R}^e)$, such that

$$v = m + \chi \circ f - \chi, \qquad m = \bar{m} \circ \pi, \qquad \bar{m} \in \ker P. \tag{5.1}$$

The definition is clearly more general than Definition 4.1, but the consequences are unchanged.

THEOREM 5.2. *Suppose that $f$ is mixing and that the decomposition* (5.1) *holds with $p = 2$. Then the conclusion of Theorem* 2.1 *is valid.*

PROOF. By Theorem 3.1, we again reduce to considering the martingale part $m$. Define the cadlag processes $(M_n, \mathbb{M}_n)$ and $(\overline{M}_n, \overline{\mathbb{M}}_n)$ starting from $m$ and $\bar{m}$, respectively. Then $(M_n, \mathbb{M}_n) = (\overline{M}_n, \overline{\mathbb{M}}_n) \circ \pi$. Hence, we reduce to proving the iterated WIP for $(M_n, \mathbb{M}_n) = (\overline{M}_n, \overline{\mathbb{M}}_n)$. Since $\bar{m} \in \ker P$, we are now in the situation of Section 4, and the result follows from Lemma 4.5. □

For the remainder of this paper, hypotheses about the existence of a martingale-coboundary decomposition refer only to the more general decomposition in (5.1).

5.1. *Applications of Theorem* 5.2. We consider first the case of Axiom A (uniformly hyperbolic) diffeomorphisms. By Bowen [6], any (nontrivial) hyperbolic basic set can be modeled by a two-sided subshift of finite type $f : \Lambda \to \Lambda$. The alphabet consists of $k$ symbols $\{0, 1, \ldots, k-1\}$ and there is a transition matrix $A \in \mathbb{R}^{k \times k}$ consisting of zeros and ones. The phase space $\Lambda$ consists of bi-infinite sequences $y = (y_i) \in \{0, 1, \ldots, k-1\}^{\mathbb{Z}}$ such that $A_{y_i, y_{i+1}} = 1$ for all $i \in \mathbb{Z}$, and $f$ is the shift $(fy)_i = y_{i+1}$.

For any $\theta \in (0, 1)$, we define the metric $d_\theta(x, y) = \theta^{s(x,y)}$ where the separation time $s(x, y)$ is the greatest integer $n \geq 0$ such that $x_i = y_i$ for $|i| \leq n$. Define $F_\theta(\Lambda)$ to be the space of $d_\theta$-Lipschitz functions $v : \Lambda \to \mathbb{R}^e$ with Lipschitz constant $|v|_\theta = \sup_{x \neq y} |x - y|/d_\theta(x, y)$ and norm $\|v\|_\theta = |v|_\infty + |v|_\theta$ where $|v|_\infty$ is the sup-norm. For each $\theta$, this norm makes $F_\theta(\Lambda)$ into a Banach space.

As usual, we have the corresponding one-sided shift $\bar{f} : \bar{\Lambda} \to \bar{\Lambda}$ where $\bar{\Lambda} = \{0, 1, \ldots, k-1\}^{\{0,1,2,\ldots\}}$, and the associated function space $F_\theta(\bar{\Lambda})$. There is a natural projection $\pi : \Lambda \to \bar{\Lambda}$ that is a semiconjugacy between the shifts $f$ and $\bar{f}$, and Lipschitz observables $\bar{v} \in F_\theta(\bar{\Lambda})$ lift to Lipschitz observables $v = \bar{v} \circ \pi \in F_\theta(\Lambda)$.

A $k$-cylinder in $\bar{\Lambda}$ is a set of the form $[a_0, \ldots, a_{k-1}] = \{y \in \bar{\Lambda} : y_i = a_i$ for all $i = 0, \ldots, k-1\}$, where $a_0, \ldots, a_{k-1} \in \{0, 1, \ldots, k-1\}$. The underlying $\sigma$-algebra $\overline{\mathcal{B}}$ is defined to be the $\sigma$-algebra generated by the $k$-cylinders. Note that $\bar{f} : \bar{\Lambda} \to \bar{\Lambda}$ is measurable with respect to this $\sigma$-algebra. We define $\mathcal{B}$ to be the smallest $\sigma$-algebra on $\Lambda$ such that $\pi : \Lambda \to \bar{\Lambda}$ and $f : \Lambda \to \Lambda$ are measurable.

For any potential function in $F_\theta(\bar{\Lambda})$ we obtain a unique equilibrium state $\bar{\mu}$. This is an ergodic $\bar{f}$-invariant probability measure defined on $(\bar{\Lambda}, \overline{\mathcal{B}})$. Define $\mu$ on $(\Lambda, \mathcal{B})$ to be the unique $f$-invariant measure such that $\pi_*\mu = \bar{\mu}$. Again, $\mu$ is an ergodic probability measure.

We assume that there is an integer $m \geq 1$ such that all entries of $A^m$ are nonzero. Then the shift $f$ is mixing with respect to $\mu$.

PROOF OF THEOREM 2.1. To each $y \in \bar{\Lambda}$ associate a $y^* \in \Lambda$ such that (i) $y_i^* = y_i$ for all $i \geq 0$ and (ii) $x_0 = y_0$ implies that $x_i^* = y_i^*$ for each $i \leq 0$ (e.g., for the full shift, take $y_i^* = 0$ for $i < 0$).

Given the observable $v \in F_\theta(\Lambda)$, define $\chi_1(x) = \sum_{n=0}^{\infty} v(f^n x^*) - v(f^n x)$. Then $\chi_1 \in L^\infty$ and $v = \hat{v} + \chi_1 \circ f - \chi_1$ where $\hat{v}$ "depends only on the future" and projects down to an observable $\bar{v} : \bar{\Lambda} \to \mathbb{R}$. Moreover, by Sinai [48], $\bar{v} \in F_{\theta^{1/2}}(\bar{\Lambda})$. It is standard that there exist constants $a, C > 0$ such that $|\int_\Lambda \bar{v} w \circ f^n \, d\mu| \leq C\|\bar{v}\|_{\theta^{1/2}} \|w\|_1 e^{-an}$ for all $w \in L^1$, $n \geq 1$. By Proposition 4.4, (4.1) holds for all $p$ (even $p = \infty$). That is, there exist $\bar{m}$, $\bar{\chi}_2 \in L^\infty(\bar{\Lambda})$ such that $\bar{v} = \bar{m} + \bar{\chi}_2 \circ \bar{f} - \bar{\chi}_2$ where $\bar{m} \in \ker P$. It follows that $\hat{v} = m + \chi_2$ where $m = \bar{m} \circ \pi$, $\chi_2 = \bar{\chi}_2 \circ \pi$. Setting $\chi = \chi_1 + \chi_2$, we obtain an $L^\infty$ martingale-coboundary decomposition for $v$ in the sense of (5.1). Now apply Theorem 5.2. $\square$

Our results hold for also for the class of nonuniformly hyperbolic diffeomorphisms studied by Young [54]. The maps in [54] enjoy exponential decay of correlations for Hölder observables.

More generally, it is possible to consider the situation of Young [55] where the decay of correlations is at a polynomial rate $n^{-\tau}$. Provided $\tau > 2$ and there is exponential contraction along stable manifolds, then the conclusion of Theorem 2.1 goes through unchanged. These conditions can be relaxed further; see Section 10.

**6. Iterated WIP for flows.** In this section, we prove a continuous time version of the iterated WIP by reducing from continuous time to discrete time. Theorem 6.1 below is formulated in a purely probabilistic setting, extending the approach in [19, 37, 40].

We suppose that $f : \Lambda \to \Lambda$ is a map with ergodic invariant probability measure $\mu$. Let $r : \Lambda \to \mathbb{R}^+$ be an integrable roof function with $\bar{r} = \int_\Lambda r \, d\mu$. We suppose throughout that $r$ is bounded below (away from zero). Define the suspension $\Lambda^r = \{(x, u) \in \Lambda \times \mathbb{R} : 0 \leq u \leq r(x)\}/\sim$ where $(x, r(x)) \sim (fx, 0)$. Define the suspension flow $\phi_t(x, u) = (x, u + t)$ computed modulo identifications. The measure $\mu^r = \mu \times \text{Lebesgue}/\bar{r}$ is an ergodic invariant probability measure for $\phi_t$. Using the notation of the Introduction, we write $(\Omega, \nu) = (\Lambda^r, \mu^r)$.

Now suppose that $v : \Omega \to \mathbb{R}^e$ is integrable with $\int_\Omega v \, d\nu = 0$. Define the smooth processes $W_n \in C([0, \infty), \mathbb{R}^e)$, $\mathbb{W}_n \in C([0, \infty), \mathbb{R}^e \times \mathbb{R}^{e \times e})$,

$$W_n(t) = n^{-1/2} \int_0^{nt} v \circ \phi_s \, ds,$$

$$\mathbb{W}_n^{\beta\gamma}(t) = \int_0^t W_n^\beta \, dW_n^\gamma = n^{-1} \int_0^{nt} \int_0^s v^\beta \circ \phi_r v^\gamma \circ \phi_s \, dr \, ds.$$

Define $\tilde{v} : \Lambda \to \mathbb{R}^e$ by setting $\tilde{v}(x) = \int_0^{r(x)} v(x,u)\,du$, and define the cadlag processes $\widetilde{W}_n \in D([0,\infty), \mathbb{R}^e)$, $\widetilde{\mathbb{W}}_n \in D([0,\infty), \mathbb{R}^{e\times e})$,

$$\widetilde{W}_n(t) = n^{-1/2} \sum_{j=0}^{[nt]-1} \tilde{v} \circ f^j,$$

$$\widetilde{\mathbb{W}}_n^{\beta\gamma}(t) = \int_0^t \widetilde{W}_n^\beta d\widetilde{W}_n^\gamma = n^{-1} \sum_{0 \le i < j \le [nt]-1} \tilde{v}^\beta \circ f^i \tilde{v}^\gamma \circ f^j.$$

We assume that the discrete time case is understood, so we have that

$$(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \to_w (\widetilde{W}, \widetilde{\mathbb{W}}) \qquad \text{in } D\big([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e}\big), \tag{6.1}$$

where $\widetilde{W}$ is $e$-dimensional Brownian motion and $\widetilde{\mathbb{W}}^{\beta\gamma}(t) = \int_0^t \widetilde{W}^\beta \, d\widetilde{W}^\gamma + \tilde{E}^{\beta\gamma} t$. Here, the probability space for the processes on the left-hand side is $(\Lambda, \mu)$.

Define $H : \Omega \to \mathbb{R}^e$ by setting $H(x,u) = \int_0^u v(x,s)\,ds$.

THEOREM 6.1. *Suppose that $\tilde{v} \in L^2(\Lambda)$ and $|H||v| \in L^1(\Omega)$. Assume (6.1) and that*

$$n^{-1/2} \sup_{t\in[0,T]} |H \circ \phi_{nt}| \to_w 0 \qquad \text{in } C\big([0,\infty), \mathbb{R}^e\big), \tag{6.2}$$

$$\lim_{n\to\infty} n^{-1} \Big\| \max_{1\le k \le nT} \Big| \sum_{1 \le i \le k} \tilde{v} \circ f^i \Big| \Big\|_2 = 0. \tag{6.3}$$

*Then $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ in $C([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$ where the probability space on the left-hand side is $(\Omega, \nu)$, and*

$$W = (\bar{r})^{-1/2} \widetilde{W}, \qquad \mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta \, dW^\gamma + E^{\beta\gamma} t,$$

$$E^{\beta\gamma} = (\bar{r})^{-1} \tilde{E}^{\beta\gamma} + \int_\Omega H^\beta v^\gamma \, d\nu.$$

REMARK 6.2. The regularity conditions on $\tilde{v}$ and $|H||v|$ are satisfied if $v \in L^\infty(\Omega, \mathbb{R}^e)$ and $r \in L^2(\Lambda, \mathbb{R})$, or if $v \in L^2(\Omega, \mathbb{R}^e)$ and $r \in L^\infty(\Lambda, \mathbb{R})$. Moreover, assumption (6.2) is satisfied under these conditions by Proposition 6.6(b).

If $\tilde{v}$ admits an $L^2$ martingale-coboundary decomposition (5.1), then condition (6.3) holds by Burkholder's inequality [10].

In the remainder of this section, we prove Theorem 6.1. Recall the notation $v_t = \int_0^t v \circ \phi_s \, ds$, $\tilde{v}_n = \sum_{j=0}^{n-1} \tilde{v} \circ f^j$, $r_n = \sum_{j=0}^{n-1} r \circ f^j$. For $(x,u) \in \Omega$ and $t > 0$, we define the lap number $N(t) = N(x,u,t) \in \mathbb{N}$:

$$N(t) = \max\{n \ge 0 : r_n(x) \le u + t\}.$$

Define $g_n(t) = N(nt)/n$.

LEMMA 6.3. *$(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \circ g_n \to_w ((\bar{r})^{-1/2}\widetilde{W}, (\bar{r})^{-1}\widetilde{\mathbb{W}})$ in $D(([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$.*

PROOF. By (6.1), $(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \to_w (\widetilde{W}, \widetilde{\mathbb{W}})$ on $(\Lambda, \mu)$. Extend $(\widetilde{W}_n, \widetilde{\mathbb{W}}_n)$ to $\Omega$ by setting $\widetilde{W}_n(x,u) = \widetilde{W}_n(x)$, $\widetilde{\mathbb{W}}_n(x,u) = \widetilde{\mathbb{W}}_n(x)$.

We claim that $(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \to_w (\widetilde{W}, \widetilde{\mathbb{W}})$ on $(\Omega, \nu)$. Define $\bar{g}(t) = t/\bar{r}$. By the ergodic theorem, $g_n(t) = N(nt)/n = tN(nt)/(nt) \to \bar{g}(t)$ almost everywhere on $(\Omega, \nu)$. Hence, $(\widetilde{W}_n, \widetilde{\mathbb{W}}_n, g_n) \to_w (\widetilde{W}, \widetilde{\mathbb{W}}, \bar{g})$ on $(\Omega, \nu)$. It follows from the continuous mapping theorem that

$$\begin{aligned}\{(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \circ g_n(t), t \geq 0\} \to_w \{(\widetilde{W}, \widetilde{\mathbb{W}}) \circ g(t), t \geq 0\} \\ = \{(\widetilde{W}(t/\bar{r}), \widetilde{\mathbb{W}}(t/\bar{r})), t \geq 0\} \\ = \{((\bar{r})^{-1/2}\widetilde{W}(t), (\bar{r})^{-1}\widetilde{\mathbb{W}}(t)), t \geq 0\}\end{aligned}$$

on $(\Omega, \nu)$ completing the proof.

It remains to verify the claim, Let $c = \operatorname{essinf} r$ and form the probability space $(\Omega, \mu_c)$ where $\mu_c = (\mu \times \text{Lebesgue}|_{[0,c]})/c$. Then it is immediate that $(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \to_w (\widetilde{W}, \widetilde{\mathbb{W}})$ on $(\Omega, \mu_c)$. To pass from $\mu_c$ to $\nu$, and hence to prove the claim, we apply [56], Theorem 1. Since $\mu_c$ is absolutely continuous with respect to $\nu$, it suffices to prove for all $\varepsilon, T > 0$ that

$$\lim_{n\to\infty} \mu_r\Bigl(\sup_{t\in[0,T]} |P_n(t) \circ f - P_n(t)| > \varepsilon\Bigr) = 0, \tag{6.4}$$

for $P_n = \widetilde{W}_n$ and $P_n = \widetilde{\mathbb{W}}_n$. We give the details for the latter since that is the more complicated case. Compute that $\widetilde{\mathbb{W}}_n^{\beta\gamma}(t) \circ f - \widetilde{\mathbb{W}}_n^{\beta\gamma}(t) = n^{-1}\sum_{1\leq i<[nt]} \tilde{v}^\gamma \circ f^i \tilde{v}^\beta \circ f^{[nt]} - n^{-1}\sum_{1\leq j<[nt]} \tilde{v}^\gamma \tilde{v}^\beta \circ f^j$ and so

$$\begin{aligned}\Bigl\|\sup_{[0,T]} |\widetilde{\mathbb{W}}_n^{\beta\gamma} \circ f - \widetilde{\mathbb{W}}_n^{\beta\gamma}|\Bigr\|_1 &\leq \|\tilde{v}^\beta\|_2 n^{-1}\Bigl\|\max_{1\leq k\leq nT}\Bigl|\sum_{1\leq i<k} \tilde{v}^\gamma \circ f^i\Bigr|\Bigr\|_2 \\ &\quad + \|\tilde{v}^\gamma\|_2 n^{-1}\Bigl\|\max_{1\leq k\leq nT}\Bigl|\sum_{1\leq j\leq k} \tilde{v}^\beta \circ f^j\Bigr|\Bigr\|_2 \to 0\end{aligned}$$

by (6.3). Hence, (6.4) follows from Markov's inequality. □

It follows from the definition of lap number that

$$\phi_t(x,u) = (f^{N(t)}x, u + t - r_{N(t)}(x)).$$

We have the decomposition

$$\begin{aligned}v_t(x,u) &= \int_0^{N(t)} v(\phi_s(x,0))\,ds + H \circ \phi_t(x,u) - H(x,u) \\ &= \tilde{v}_{N(t)}(x) + H \circ \phi_t(x,u) - H(x,u).\end{aligned} \tag{6.5}$$

We also require the following elementary result.

PROPOSITION 6.4. *Let $a_n$ be a real sequence and $b > 0$. If $\lim_{n\to\infty} n^{-b} a_n = 0$, then $\lim_{n\to\infty} n^{-b} \sup_{t\in[0,T]} |a_{[nt]}| = 0$.*

LEMMA 6.5. *$(W_n, \mathbb{W}_n) = (\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \circ g_n + F_n$, where $F_n \to_w F$ in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$ and $F(t) = (0, \int_\Omega H^\beta v^\gamma \, d\nu)t$.*

PROOF. Using (6.5), we can write

$$W_n(t) = n^{-1/2} v_{nt} = n^{-1/2} \tilde{v}_{N(t)} + n^{-1/2} H \circ \phi_{nt} - n^{-1/2} H.$$

By definition, $\widetilde{W}_n(N(nt)/n) = n^{-1/2} \tilde{v}_{N(t)}$. Hence, by assumption (6.2), we obtain the required decomposition for $W_n$.

Similarly,

$$\begin{aligned} \mathbb{W}_n^{\beta\gamma}(t) &= \int_0^t W_n^\beta \, dW_n^\gamma = \int_0^t v_{ns}^\beta v^\gamma \circ \phi_{ns} \, ds \\ &= \int_0^t [\tilde{v}_{N(ns)}^\beta + H^\beta \circ \phi_{ns} - H^\beta] v^\gamma \circ \phi_{ns} \, ds = A_n(t) + B_n(t), \end{aligned} \tag{6.6}$$

where

$$A_n(t) = \int_0^t \tilde{v}_{N(ns)}^\beta v^\gamma \circ \phi_{ns} \, ds, \qquad B_n(t) = n^{-1} \int_0^{nt} [H^\beta \circ \phi_s - H^\beta] v^\gamma \circ \phi_s \, ds.$$

By the ergodic theorem,

$$n^{-1} H^\beta \int_0^n v^\gamma \circ \phi_s \, ds = H^\beta (n)^{-1} \int_0^n v^\gamma \circ \phi_s \, ds \to H^\beta \int_\Omega v^\gamma \, d\nu = 0.$$

Hence, by Proposition 6.4, $n^{-1} \sup_{t\in[0,T]} |H^\beta \int_0^{nt} v^\gamma \circ \phi_s \, ds| \to 0$ a.e. Similarly,

$$n^{-1} \int_0^n H^\beta \circ \phi_s v^\gamma \circ \phi_s \, ds = n^{-1} \int_0^n (H^\beta v^\gamma) \circ \phi_s \, ds \to \int_\Omega H^\beta v^\gamma \, d\nu.$$

Applying Proposition 6.4 with $b = 1$ and $a_n = \int_0^n H^\beta \circ \phi_s v^\gamma \circ \phi_s \, ds - n \int_\Omega H^\beta v^\gamma \, d\nu$, we obtain that $n^{-1} \int_0^{nt} H^\beta \circ \phi_s v^\gamma \circ \phi_s \, ds \to \int_\Omega H^\beta v^\gamma \, d\nu$ uniformly on $[0, T]$ a.e. Hence, $B_n(t) \to t \int_\Omega H^\beta v^\gamma \, d\nu$ uniformly on $[0, T]$ a.e.

To deal with the term $A_n$, we introduce the return times $t_{n,j} = t_{n,j}(x, u)$, with $0 = t_{n,0} < t_{n,1} < t_{n,2} < \cdots$ such that $N(nt) = j$ for $t \in [t_{n,j}, t_{n,j+1})$. Note that $t_{n,j}(x, u) = (r_j(x) - u)/n$ for $j \ge 1$. Since $r$ is bounded below, we have that $\lim_{j\to\infty} t_{n,j} = \infty$ for each $n$.

Compute that

$$\begin{aligned} A_n(t) &= \sum_{j=0}^{N(nt)-1} \int_{t_{n,j}}^{t_{n,j+1}} \tilde{v}_j^\beta v^\gamma \circ \phi_{ns} \, ds + \int_{t_{n,N(nt)}}^t \tilde{v}_{N(nt)}^\beta v^\gamma \circ \phi_{ns} \, ds \\ &= \sum_{j=0}^{N(nt)-1} \tilde{v}_j^\beta \int_{t_{n,j}}^{t_{n,j+1}} v^\gamma \circ \phi_{ns} \, ds + \tilde{v}_{N(nt)}^\beta \int_{t_{n,N(nt)}}^t v^\gamma \circ \phi_{ns} \, ds. \end{aligned}$$

For $j \geq 1$,

$$\int_{t_{n,j}}^{t_{n,j+1}} v \circ \phi_{ns}\, ds = \int_{t_{n,j}}^{t_{n,j+1}} v\big(f^j x, u + ns - r_j(x)\big)\, ds$$
$$= n^{-1} \int_0^{r(f^j x)} v\big(f^j x, s\big)\, ds = n^{-1} \tilde{v} \circ f^j,$$

and similarly we can write $\int_0^{t_{n,1}} v \circ \phi_{ns}\, ds = n^{-1} \int_u^{r(x)} v(x,s)\, ds = n^{-1}\tilde{v} + O(1/n)$ a.e.

By definition, $\widetilde{\mathbb{W}}_n(N(nt)/n) = n^{-1} \sum_{j=0}^{N(nt)-1} \tilde{v}_j \tilde{v} \circ f^j$. Hence, we have shown that $A_n(t) = \widetilde{\mathbb{W}}_n \circ g_n(t) + C_n(t) + O(1/n)$ a.e., where $C_n^{\beta\gamma}(t) = \tilde{v}^{\beta}_{N(nt)} \int_{t_{n,N(nt)}}^t v^{\gamma} \circ \phi_{ns}\, ds$.

Finally, we note that

$$\int_{t_{n,N(nt)}}^{t} v \circ \phi_{ns}\, ds = \int_{t_{n,N(nt)}}^{t} v\big(f^{N(nt)} x, u + ns - r_{N(nt)}(x)\big)\, ds$$
$$= n^{-1} \int_0^{u+t-r_{N(nt)}(x)} v\big(f^{N(nt)} x, s\big)\, ds$$
$$= n^{-1} H\big(f^{N(nt)} x, u + t - r_{N(nt)}(x)\big)$$
$$= n^{-1} H \circ \phi_{nt}.$$

Hence, $C_n^{\beta\gamma} = \widetilde{W}_n^{\beta} \circ g_n(t) \cdot n^{-1/2} H^{\gamma} \circ \phi_{nt} \to_w 0$ by Lemma 6.3 and assumption (6.2). □

PROOF OF THEOREM 6.1. This is immediate from Lemmas 6.3 and 6.5. □

PROPOSITION 6.6. *Sufficient conditions for assumption* (6.2) *to hold are that* (a) $H \in L^{2+}(\Omega, \mathbb{R}^e)$, *or* (b) $\tilde{v}_* \in L^2(\Lambda)$, *where* $\tilde{v}_*(x) = \int_0^{r(x)} |v(x,u)|\, du$.

PROOF. In both cases, we prove that $n^{-1/2} H \circ \phi_n \to 0$ a.e. By Proposition 6.4, $\sup_{t\in[0,T]} H \circ \phi_{nt} \to 0$ a.e.

(a) Choose $\delta > 0$ such that $H \in L^{2+\delta}$ and $\tau < \frac{1}{2}$ such that $\tau(2+\delta) > 1$. Since $\|H \circ \phi_n\|_{2+\delta} = \|H\|_{2+\delta}$, it follows from Markov's inequality that $\nu(|H \circ \phi_n| > n^{\tau}) \leq \|H\|_{2+\delta} n^{-\tau(2+\delta)}$ which is summable. By Borel–Cantelli, there is a constant $C > 0$ such that $|H \circ \phi_n| \leq C n^{-\tau}$ a.e., and hence $n^{-1/2} H \circ \phi_n \to 0$ a.e.

(b) Since $\tilde{v}_*^2 \in L^1(\Lambda)$, it follows from the ergodic theorem that $n^{-1/2} \tilde{v}_* \circ f^n \to 0$ a.e. Moreover, $N(nt)/n \to 1/\bar{r}$ a.e. on $(\Omega, \nu)$ and hence $n^{-1/2} \tilde{v}_* \circ f^{[N(nt)]} \to 0$ a.e. The result follows since $|H(x,u)| \leq \tilde{v}_*(x)$ for all $x, u$. □

REMARK 6.7. The sufficient conditions in Proposition 6.6 imply almost sure convergence, uniformly on $[0, T]$, for the term $F_n$ in Lemma 6.5.

**7. Moment estimates.** In this section, we obtain some moment estimates that are required to apply rough path theory. (Proposition 7.5 below is also required for part of Theorem 1.1; see the proof of Corollary 8.3.)

7.1. *Discrete time moment estimates.* Let $f:\Lambda\to\Lambda$ be a map (invertible or noninvertible) with invariant probability measure $\mu$. Suppose that $v:\Lambda\to\mathbb{R}^e$ is a mean zero observable lying in $L^\infty$. Define

$$v_n=\sum_{j=0}^{n-1}v\circ f^j,\qquad S_n^{\beta\gamma}=\sum_{0\le i<j<n}v^\beta\circ f^i v^\gamma\circ f^j.$$

PROPOSITION 7.1. *Suppose that $v:\Lambda\to\mathbb{R}^e$ lies in $L^\infty$ and admits an $L^p$ martingale-coboundary decomposition* (5.1) *for some $p\ge 3$. Then there exists a constant $C>0$ such that*

$$\Big\|\max_{0\le j\le n}|v_j|\Big\|_{2p}\le Cn^{1/2},\qquad \Big\|\max_{0\le j\le n}|S_j|\Big\|_{2p/3}\le Cn\qquad \textit{for all } n\ge 1.$$

PROOF. The estimate $\|v_n\|_{2p}\ll n^{1/2}$ is proved in [34], equation (3.1). Since $v_{n+a}-v_a=_d v_n$ for all $a,n$, the result for $\max_{0\le j\le n}|v_j|$ follows by [47], Corollary B1 (cf. [38], Lemma 4.1).

To estimate $S_n$ write

$$S_n^{\beta\gamma}=\sum_{0\le i<j<n}m^\beta\circ f^i v^\gamma\circ f^j+\sum_{1\le j<n}(\chi^\beta\circ f^j-\chi^\beta)v^\gamma\circ f^j.$$

We have $\|\sum_{1\le j<n}\chi^\beta\circ f^j v^\gamma\circ f^j\|_p\le n\|\chi^\beta v^\gamma\|_p\le n\|\chi^\beta\|_p\|v^\gamma\|_\infty$ and $\|\sum_{1\le j<n}\chi^\beta v^\gamma\circ f^j\|_p\le\|\chi^\beta\|_p\|\sum_{1\le j<n}v^\gamma\circ f^j\|_\infty\le n\|\chi^\beta\|_p\|v^\gamma\|_\infty$.

Next, we estimate $I_n=\sum_{0\le i<j<n}m^\beta\circ f^i v^\gamma\circ f^j$. Passing to the natural extension $\tilde f:\tilde\Lambda\to\tilde\Lambda$ in the noninvertible case (and taking $\tilde f=f$ in the invertible case), we have

$$\tilde I_n=\sum_{0\le i<j<n}\tilde m^\beta\circ\tilde f^i\tilde v^\gamma\circ\tilde f^j=\Big(\sum_{-n\le i<j<0}\tilde m^\beta\circ\tilde f^i\tilde v^\gamma\circ\tilde f^j\Big)\circ\tilde f^n=\tilde I_n^-\circ\tilde f^n,$$

so we reduce to estimating $\tilde I_n^-=\sum_{-n\le i<j<0}\tilde v^\gamma\circ\tilde f^j\tilde m^\beta\circ\tilde f^i$.

Now,

$$\tilde I_n^-=\sum_{k=1}^n X_k\qquad\text{where } X_k=\Big(\sum_{-k<j<0}\tilde v^\gamma\circ\tilde f^j\Big)\tilde m^\beta\circ\tilde f^{-k}.$$

Recall that $E(\tilde m^\beta\circ\tilde f^i\,|\,\tilde f^{-i-1}\tilde{\mathcal{B}})=0$. Hence $E(X_k\,|\,\tilde f^{k-1}\tilde{\mathcal{B}})=0$, and so $\{X_k;k\ge1\}$ is a sequence of martingale differences. For $p'>1$, Burkholder's inequality [10] states that $\|\tilde I_n^-\|_{p'}\ll\|(\sum_{k=1}^n X_k^2)^{1/2}\|_{p'}$, and it follows for $p'\ge 2$ that

$$\|\tilde I_n^-\|_{p'}^2\ll\sum_{k=1}^n\|X_k\|_{p'}^2. \tag{7.1}$$

Taking $p' = 2p/3$, it follows from Hölder's inequality that

$$\|X_k\|_{2p/3} \le \Big\| \sum_{-k<j<0} \tilde{v}^\gamma \circ \tilde{f}^j \Big\|_{2p} \|\tilde{m}^\beta \circ \tilde{f}^{-k}\|_p = \|v_{k-1}^\gamma\|_{2p} \|m^\beta\|_p \ll k^{1/2}.$$

Hence, $\|\tilde{I}_n^-\|_{2p/3} \ll n$ and so $\|S_n\|_{2p/3} \ll n$.

This time we cannot apply the maximal inequality of [47] since we do not have a good estimate for $S_{a+n} - S_a$ uniform in $a$. However, we claim that $\|S_{a+n} - S_a\|_{2p/3} \ll n + n^{1/2}a^{1/2}$. Set $A_{a,n} = (\sum_{k=a+1}^{a+n} b_k^2)^{1/2}$ with $b_k = k^{1/2}$. By the claim, $\|S_{a+n} - S_a\|_{2p/3} \ll A_{a,n}$ and it follows from [41], Theorem A (see also references therein) that $\|\max_{0\le j\le n} |S_j|\|_{2p/3} \ll n$ as required.

For the claim, observe that

$$\begin{aligned} S_{a+n}^{\beta\gamma} - S_a^{\beta\gamma} &= \sum_{j=a}^{a+n-1} \sum_{i=0}^{j-1} v^\beta \circ f^i v^\gamma \circ f^j \\ &= \sum_{j=a}^{a+n-1} \sum_{i=0}^{a-1} v^\beta \circ f^i v^\gamma \circ f^j + \sum_{j=a}^{a+n-1} \sum_{i=a}^{j-1} v^\beta \circ f^i v^\gamma \circ f^j \\ &= v_a^\beta v_n^\gamma \circ f^a + S_n^{\beta\gamma} \circ f^a. \end{aligned}$$

Hence,

$$\|S_{a+n}^{\beta\gamma} - S_a^{\beta\gamma}\|_q \le \|v_n^\gamma\|_{2q} \|v_a^\beta\|_{2q} + \|S_n\|_q \ll n^{1/2}a^{1/2} + n,$$

for $q = 2p/3$. This proves the claim. □

REMARK 7.2. The proof of Proposition 7.1 makes essential use of the fact that $v \in L^\infty$ [26, 34, 38]. Under this assumption, the estimate for $\max_{0\le j\le n} |v_j|$ requires only that $p \ge 1$ and is optimal in the sense that there are examples where $\lim_{n\to\infty} \|n^{-1/2} v_n\|_q = \infty$ for all $q > 2p$; see [38], Remark 3.7.

We conjecture that the optimal estimate for $\max_{0\le j\le n} |S_j|$ is that $\|\max_{0\le j\le n} |S_j|\|_p \ll n$ (for $p \ge 2$). Then we would only require $p > 3$ instead of $p > 9/2$ in our main results.

Recall that $W_n(t) = n^{-1/2} \sum_{j=0}^{[nt]-1} v \circ f^j$ and $\mathbb{W}_n^{\beta\gamma}(t) = \int_0^t W_n^\beta \, dW_n^\gamma$. We define the increments

$$W_n(s,t) = W_n(t) - W_n(s) \quad \text{and} \quad \mathbb{W}_n^{\beta\gamma}(s,t) = \int_s^t W_n^\beta(s,r) \, dW_n^\gamma(r).$$

COROLLARY 7.3. *Suppose that $v : \Lambda \to \mathbb{R}^e$ lies in $L^\infty$ and admits an $L^p$ martingale-coboundary decomposition (5.1) for some $p \ge 3$. Then there exists a constant $C > 0$ such that*

$$\|W_n(j/n, k/n)\|_{2p} \le C(|k-j|/n)^{1/2} \quad \textit{and}$$

$$\|\mathbb{W}_n(j/n, k/n)\|_{2p/3} \le C|k-j|/n,$$

*for all* $j, k, n \geq 1$.

PROOF. Let $t > s > 0$. By definition,

$$\begin{aligned} W_n(s,t) &= n^{-1/2} \sum_{i=[ns]}^{[nt]-1} v \circ f^i = n^{-1/2} \Bigg( \sum_{i=0}^{[nt]-[ns]-1} v \circ f^i \Bigg) \circ f^{[ns]} \\ &=_d n^{-1/2} \sum_{i=0}^{[nt]-[ns]-1} v \circ f^i = n^{-1/2} v_{[nt]-[ns]}. \end{aligned}$$

By Proposition 7.1, assuming without loss that $j < k$,

$$\| W_n(j/n, k/n) \|_{2p} = n^{-1/2} \| v_{k-j} \|_{2p} \leq C\big((k-j)/n\big)^{1/2}.$$

Similarly,

$$\begin{aligned} \mathbb{W}_n(s,t) &= n^{-1} \sum_{[ns] \leq i < j \leq [nt]-1} v^\beta \circ f^i v^\gamma \circ f^j \\ &= n^{-1} \Bigg( \sum_{0 \leq i < j < [nt]-[ns]-1} v^\beta \circ f^i v^\gamma \circ f^j \Bigg) \circ f^{[ns]} \\ &=_d n^{-1} \sum_{0 \leq i < j < [nt]-[ns]-1} v^\beta \circ f^i v^\gamma \circ f^j = n^{-1} S^{\beta\gamma}_{[nt]-[ns]}. \end{aligned}$$

By Proposition 7.1,

$$\| \mathbb{W}_n(j/n, k/n) \|_{2p/3} = n^{-1} \| S_{k-j} \|_{2p/3} \leq C(k-j)/n,$$

as required. $\square$

7.2. *Continuous time moment estimates*. Let $\phi_t : \Omega \to \Omega$ be a suspension flow as in Section 6, with Poincaré map $f : \Lambda \to \Lambda$. As before, we write $\Omega = \Lambda^r$, $\nu = \mu^r$, where $r : \Lambda \to \mathbb{R}$ is a roof function with $\bar{r} = \int r \, d\mu$. Let $v : \Omega \to \mathbb{R}^e$ with $\int_\Omega v \, d\nu = 0$.

As before, we suppose that $r$ is bounded away from zero, but now we suppose in addition that $v$ and $r$ lie in $L^\infty$. (These assumptions can be relaxed, but then the assumption on $p$ has to be strengthened in the subsequent results.)

Define

$$v_t = \int_0^t v \circ \phi_s \, ds, \qquad S_t^{\beta\gamma} = \int_0^t \int_0^u v^\beta \circ \phi_s v^\gamma \circ \phi_u \, ds \, du.$$

Let $\tilde{v} : \Lambda \to \mathbb{R}^e$ be given by $\tilde{v}(x) = \int_0^{r(x)} v(x,u) \, du$ (so $\tilde{v}$ coincides with the function defined in Section 6). The assumptions on $v$ and $r$ imply that $\tilde{v} \in L^\infty(\Lambda, \mu)$.

PROPOSITION 7.4. $N(t) \leq [t / \operatorname{ess\,inf} r] + 1$ *for all* $(x,u) \in \Omega$, $t \geq 0$.

PROOF. Compute that

$$
\begin{aligned}
r_{[t/\operatorname{ess\,inf} r]+2}(x) &= r(x) + r_{[t/\operatorname{ess\,inf} r]+1}(fx) \\
&\geq u + \big([t/\operatorname{ess\,inf} r] + 1\big) \operatorname{ess\,inf} f > u + t.
\end{aligned}
$$

Hence, the result follows from the definition of lap number. $\square$

PROPOSITION 7.5. *Suppose that $\tilde{v} : \Lambda \to \mathbb{R}^e$ admits an $L^p$ martingale-coboundary decomposition* (5.1) *for some $p \geq 3$. Then there exists a constant $C > 0$ such that*

$$
\|v_t\|_{2p} \leq C t^{1/2}, \qquad \|S_t\|_{2p/3} \leq Ct,
$$

*for all $t \geq 0$.*

PROOF. If $t \leq 1$, then we have the almost sure estimates $|v_t| \leq \|v\|_\infty t \leq \|v\|_\infty t^{1/2}$ and $|S_t| \leq \|v\|_\infty^2 t^2 \leq \|v\|_\infty^2 t$. Hence, in the remainder of the proof, we can suppose that $t \geq 1$.

For the $v_t$ estimate, we follow the argument used in [38], Lemma 4.1. By (6.5),

$$
v_t = \tilde{v}_{N(t)} + G(t),
$$

where $G(t)(x,u) = H \circ \phi_t(x,u) - H(x,u) = \int_0^u v(\phi_t(x,s))\,ds - \int_0^u v(x,s)\,ds$. In particular, $\|G(t)\|_\infty \leq 2\|r\|_\infty \|v\|_\infty \leq 2\|r\|_\infty \|v\|_\infty t^{1/2}$. By Proposition 7.4, there is a constant $R > 0$ such that $N(t) \leq Rt$ for all $t \geq 1$. Hence,

$$
|v_t| \leq \max_{0 \leq j \leq Rt} |\tilde{v}_j| + 2\|r\|_\infty \|v\|_\infty t^{1/2}.
$$

By Proposition 7.1, $\|\max_{0 \leq j \leq Rt} |\tilde{v}_j|\|_{2p} \ll t^{1/2}$. Since $r$ is bounded above and below, this estimate for $\max_{0 \leq j \leq Rt} |\tilde{v}_j|$ holds equally in $L^{2p}(\Lambda)$ and $L^{2p}(\Omega)$. Hence $\|v_t\|_{2p} \ll t^{1/2}$.

To estimate $S_t$ we make use of decompositions similar to those in Section 6. By (6.5),

$$
S_t^{\beta\gamma} = \int_0^t v_s^\beta v^\gamma \circ \phi_s\,ds = \int_0^t \big(\tilde{v}_{N(s)}^\beta + G^\beta(s)\big) v^\gamma \circ \phi_s\,ds,
$$

where $\|\int_0^t G^\beta(s) v^\gamma \circ \phi_s\,ds\|_\infty \leq 2|r|_\infty |v|_\infty^2 t$. Moreover, in the notation from the proof of Lemma 6.5 with $n = 1$,

$$
\begin{aligned}
\int_0^t \tilde{v}_{N(s)}^\beta v^\gamma \circ \phi_s\,ds &= A_1(t) \\
&= \sum_{j=0}^{N(t)-1} \tilde{v}_j^\beta \tilde{v}^\gamma \circ f^j - \tilde{v}^\beta \int_0^u v^\gamma \circ \phi_s\,ds + \tilde{v}_{N(t)}^\beta H^\gamma \circ \phi_t \\
&= \tilde{S}_{N(t)}^{\beta\gamma} - \tilde{v}^\beta \int_0^u v^\gamma \circ \phi_s\,ds + \tilde{v}_{N(t)}^\beta H^\gamma \circ \phi_t,
\end{aligned}
$$

where $\tilde{S}_n$ is as in Proposition 7.1. Now $\|\tilde{v}^{\beta}_{N(t)}\|_\infty \le \|N(t)\|_\infty \|\tilde{v}^{\beta}\|_\infty \le Rt\|r\|_\infty\|v\|_\infty$. Hence, by Proposition 7.1,

$$\left|\int_0^t \tilde{v}^{\beta}_{N(s)} v^{\gamma} \circ \phi_s \, ds\right| \le \max_{j\le Rt} |\tilde{S}^{\beta\gamma}_j| + (1+Rt)\|r\|^2_\infty \|v\|^2_\infty \ll t,$$

completing the proof. $\square$

Again we recall that $W_n(t) = n^{-1/2}\int_0^{nt} v \circ \phi_s \, ds$ and $\mathbb{W}^{\beta\gamma}_n(t) = \int_0^t W^{\beta}_n \, dW^{\gamma}_n$, and we define the increments

$$W_n(s,t) = W_n(t) - W_n(s) \quad \text{and} \quad \mathbb{W}^{\beta\gamma}_n(s,t) = \int_s^t W^{\beta}_n(s,r)\, dW^{\gamma}_n(r).$$

COROLLARY 7.6. *Suppose that $\tilde{v}$ admits an $L^p$ martingale-coboundary decomposition* (5.1) *for some $p \ge 3$. Then there exists a constant $C > 0$ such that*

$$\|W_n(s,t)\|_{2p} \le C|t-s|^{1/2} \quad \textit{and} \quad \|\mathbb{W}_n(s,t)\|_{2p/3} \le C|t-s|,$$

*for all $s,t \ge 0$.*

PROOF. This is almost identical to the proof of Corollary 7.3. $\square$

REMARK 7.7. Any hyperbolic basic set for an Axiom A flow can be written as a suspension over a mixing hyperbolic basic set $f : \Lambda \to \Lambda$ with a Hölder roof function $r$. Since every Hölder mean zero observable $\tilde{v} : \Lambda \to \mathbb{R}^e$ admits an $L^\infty$ martingale-coboundary decomposition, it follows that Proposition 7.5 and Corollary 7.6 hold for all $p$.

**8. Applications of Theorem 6.1.** In this section, we apply Theorem 6.1 to a large class of uniformly and nonuniformly hyperbolic flows. In particular, we complete the proof of Theorem 1.1. Our main results do not require mixing assumptions on the flow, but the formulas simplify in the mixing case.

Let $\phi_t : \Omega \to \Omega$ be a suspension flow as in Section 6, with mixing Poincaré map $f : \Lambda \to \Lambda$. As before, we write $\Omega = \Lambda^r$, $\nu = \mu^r$, where $r : \Lambda \to \mathbb{R}$ is a roof function with $\bar{r} = \int r \, d\mu$.

*Nonmixing flows.* First, we consider the case where $\phi_t$ is not mixing. (As usual, we suppose that the Poincaré map $f$ is mixing.)

COROLLARY 8.1. *Suppose that $f : \Lambda \to \Lambda$ is mixing and that $r \in L^1(\Lambda)$ is bounded away from zero. Let $v \in L^1(\Omega, \mathbb{R}^e)$ with $\int_\Omega v \, d\nu = 0$. Suppose further that $|H||v|$ is integrable and that assumption* (6.2) *is satisfied.*

*Assume that $\tilde{v}$ admits a martingale-coboundary decomposition* (5.1) *with $p = 2$. Then the conclusion of Theorem* 6.1 *is valid. Moreover,*

$$\begin{aligned}\Sigma^{\beta\gamma} &= \mathrm{Cov}^{\beta\gamma}\, W(1) \\ &= (\bar{r})^{-1}\int_\Lambda \tilde{v}^{\beta}\tilde{v}^{\gamma}\, d\mu + (\bar{r})^{-1}\sum_{n=1}^{\infty}\int_\Lambda (\tilde{v}^{\beta}\tilde{v}^{\gamma}\circ f^n + \tilde{v}^{\gamma}\tilde{v}^{\beta}\circ f^n)\, d\mu,\end{aligned}$$

*and* $\mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta \circ dW^\gamma + \frac{1}{2} D^{\beta\gamma} t$ *where*

$$D^{\beta\gamma} = (\bar{r})^{-1} \sum_{n=1}^{\infty} \int_\Lambda (\tilde{v}^\beta \tilde{v}^\gamma \circ f^n - \tilde{v}^\gamma \tilde{v}^\beta \circ f^n)\, d\mu + \int_\Omega (H^\beta v^\gamma - H^\gamma v^\beta)\, d\nu.$$

PROOF. By Theorem 4.3, condition (6.1) is satisfied. Specifically, $(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \to_w (\widetilde{W}, \widetilde{\mathbb{W}})$ where $\widetilde{W}$ is a Brownian motion with $\mathrm{Cov}^{\beta\gamma}(\widetilde{W}(1)) = \int_\Lambda \tilde{v}^\beta \tilde{v}^\gamma\, d\mu + \sum_{n=1}^\infty \int_\Lambda (\tilde{v}^\beta \tilde{v}^\gamma \circ f^n + \tilde{v}^\gamma \tilde{v}^\beta \circ f^n)\, d\mu$ and $\widetilde{\mathbb{W}}_n^{\beta\gamma}(t) = \int_0^t \widetilde{W}^\beta\, d\widetilde{W}^\gamma + \tilde{E}^{\beta\gamma} t$.

By Remark 6.2, hypothesis (6.3) is satisfied. Hence, by Theorem 6.1, $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ where $W = (\bar{r})^{-1/2} \widetilde{W}$ and $\mathbb{W}^{\beta\gamma}(t) = \int_0^t W_n^\beta\, dW_n^\gamma + E^{\beta\gamma} t$. It is immediate that $\Sigma = \mathrm{Cov}(W(1))$ has the desired form. Moreover, by Theorems 4.3 and 6.1,

$$E^{\beta\gamma} = (\bar{r})^{-1} \sum_{n=1}^{\infty} \int_\Lambda \tilde{v}^\beta \tilde{v}^\gamma \circ f^n\, d\mu + \int_\Omega H^\beta v^\gamma\, d\nu.$$

The Stratonovich correction gives

$$\begin{aligned} D^{\beta\gamma} &= 2E^{\beta\gamma} - \Sigma^{\beta\gamma} \\ &= (\bar{r})^{-1} \left\{ \sum_{n=1}^{\infty} \int_\Lambda (\tilde{v}^\beta \tilde{v}^\gamma \circ f^n - \tilde{v}^\gamma \tilde{v}^\beta \circ f^n)\, d\mu - \int_\Lambda \tilde{v}^\beta \tilde{v}^\gamma\, d\mu \right\} \\ &\quad + 2 \int_\Omega H^\beta v^\gamma\, d\nu. \end{aligned}$$

To complete the proof, we show that $(\bar{r})^{-1} \int_\Lambda \tilde{v}^\beta \tilde{v}^\gamma\, d\mu = \int_\Omega H^\beta v^\gamma\, d\nu + \int_\Omega H^\gamma v^\beta\, d\nu$. Compute that

$$\begin{aligned} \int_\Lambda \tilde{v}^\beta \tilde{v}^\gamma\, d\mu &= \int_\Lambda \left\{ \int_0^{r(x)} v^\beta(x,u)\, du \int_0^{r(x)} v^\gamma(x,s)\, ds \right\} d\mu \\ &= \int_\Lambda \int_0^{r(x)} v^\beta(x,u) \left\{ \int_0^u v^\gamma(x,s)\, ds + \int_u^{r(x)} v^\gamma(x,s)\, ds \right\} du\, d\mu \\ &= \int_\Lambda \int_0^{r(x)} v^\beta(x,u) H^\gamma(x,u)\, du\, d\mu \\ &\quad + \int_\Lambda \int_0^{r(x)} v^\gamma(x,s) \left( \int_0^s v^\beta(x,u)\, du \right) ds\, d\mu \\ &= \bar{r} \int_\Omega v^\beta H^\gamma\, d\nu + \int_\Lambda \int_0^{r(x)} v^\gamma(x,s) H^\beta(x,s)\, ds \\ &= \bar{r} \int_\Omega v^\beta H^\gamma\, d\nu + \bar{r} \int_\Omega v^\gamma H^\beta\, d\nu, \end{aligned}$$

as required. □

REMARK 8.2. Corollary 8.1 applies directly to Hölder observables of semiflows that are suspensions of the uniformly and nonuniformly expanding maps considered in Section 4, and of flows that are suspensions of the uniformly and nonuniformly hyperbolic diffeomorphisms considered in Section 5. In particular, this includes Axiom A flows and nonuniformly hyperbolic flows that are suspensions over Young towers with exponential tails.

*Mixing flows.* Under additional conditions, we obtain the formulas for $\Sigma$ and $D$ promised in Theorem 1.1(b).

COROLLARY 8.3. *Assume the set up of Corollary 8.1. Suppose further that $v \in L^\infty$, and that $\tilde{v}$ admits a martingale-coboundary decomposition (5.1) with $p = 3$. If the integral $\int_0^\infty \int_\Omega v^\beta v^\gamma \circ \phi_t \, d\nu \, dt$ exists, then*

$$\Sigma^{\beta\gamma} = \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_t + v^\gamma v^\beta \circ \phi_t) \, d\nu \, dt$$

*and*

$$D^{\beta\gamma} = \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_t - v^\gamma v^\beta \circ \phi_t) \, d\nu \, dt.$$

PROOF. It follows from [10] that $\|W_n\|_p = O(1)$, and hence that $\mathbb{E}_\nu |W_n|^q \to \mathbb{E}|W|^q$ for all $q < p$. In particular, taking $q = 2$, we deduce that $\mathrm{Cov}_\nu(W_n(1)) \to \Sigma$. Moreover, the calculation in the proof of Theorem 4.3 shows that $\Sigma^{\beta\gamma} = \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_t + v^\gamma v^\beta \circ \phi_t) \, d\nu \, dt$. Similarly $\mathbb{E}_\nu(\int_0^1 W_n^\beta \, dW_n^\gamma) \to \int_0^\infty \int_\Omega v^\beta v^\gamma \circ \phi_t \, d\nu \, dt$.

Since $v \in L^\infty$ and $p = 3$, it follows from Proposition 7.5 that $\| \int_0^1 W_n^\beta \, dW_n^\gamma \|_2 = O(1)$. [In fact, we require only that $\| \int_0^1 W_n^\beta \, dW_n^\gamma \|_q = O(1)$ for some $q > 1$.] Hence, $\mathbb{E}_\nu(\int_0^1 W_n^\beta \, dW_n^\gamma) \to E^{\beta\gamma}$, and so

$$\begin{aligned} D^{\beta\gamma} &= 2E^{\beta\gamma} - \Sigma^{\beta\gamma} \\ &= 2\int_0^\infty \int_\Omega v^\beta v^\gamma \circ \phi_t \, d\nu \, dt - \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_t + v^\gamma v^\beta \circ \phi_t) \, d\nu \, dt \\ &= \int_0^\infty \int_\Omega (v^\beta v^\gamma \circ \phi_t - v^\gamma v^\beta \circ \phi_t) \, d\nu \, dt, \end{aligned}$$

as required. □

PROOF OF THEOREM 1.1. We use the fact that every hyperbolic basic set for an Axiom A flow can be written as a suspension over a mixing hyperbolic basic set $f : \Lambda \to \Lambda$ with a Hölder roof function $r$. Any Hölder mean zero observable $v : \Lambda \to \mathbb{R}^e$ admits an $L^\infty$ martingale-coboundary decomposition. Hence Theorem 1.1 follows from Corollaries 8.1 and 8.3. □

**9. Smooth approximation theorem.** In this section, we prove Theorems 1.2 and 2.2. To do so, we need a few tools from *rough path theory* that allow us to lift the iterated WIP into a convergence result for fast-slow systems. We do not need to introduce much new terminology since the tools we need are to some extent prepackaged for our purposes. For the continuous time results, we use the standard rough path theory [29], but for the discrete time results we use results of [23].

9.1. *Rough path theory in continuous time.* Let $U_n : [0,T] \to \mathbb{R}^e$ be a path of bounded variation. Then we can define the iterated integral $\mathbb{U}_n : [0,T] \to \mathbb{R}^{e\times e}$ by

$$\mathbb{U}_n(t) = \int_0^t U_n(r)\, dU_n(r), \tag{9.1}$$

where the integral is uniquely defined in the Riemann–Stieltjes sense. As usual, we define the increments

$$U_n(s,t) = U_n(t) - U_n(s) \quad \text{and} \quad \mathbb{U}_n(s,t) = \int_s^t U_n(s,r)\, dU_n(r).$$

Suppose that $a : \mathbb{R}^d \to \mathbb{R}^d$ is $C^{1+}$ and $b : \mathbb{R}^d \to \mathbb{R}^{d\times e}$ is $C^3$, and let $X_n$ be the solution to the equation

$$X_n(t) = \xi + \int_0^t a\big(X_n(s)\big)\, ds + \int_0^t b\big(X_n(s)\big)\, dU_n(s), \tag{9.2}$$

which is well-defined for each $n$, and moreover has a unique solution for every initial condition $\xi \in \mathbb{R}^d$. To characterize the limit of $X_n$, we use the following standard tool from rough path theory.

THEOREM 9.1. *Suppose that $(U_n, \mathbb{U}_n) \to_w (U, \mathbb{U})$ in $C([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$, where $U$ is Brownian motion and where $\mathbb{U}$ can be written*

$$\mathbb{U}(t) = \int_0^t U(s) \circ dU(s) + Dt, \tag{9.3}$$

*for some constant matrix $D \in \mathbb{R}^{e\times e}$. Suppose moreover that there exist $C > 0$ and $q > 3$ such that*

$$\big\|U_n(s,t)\big\|_{2q} \le C|t-s|^{1/2} \quad \textit{and} \quad \big\|\mathbb{U}_n(s,t)\big\|_q \le C|t-s|, \tag{9.4}$$

*hold for all $n \ge 1$ and $s,t \in [0,T]$. Then $X_n \to_w X$ in $C([0,\infty), \mathbb{R}^d)$, where*

$$dX = \Big(a(X) + \sum_{\alpha,\beta,\gamma} D^{\beta\gamma} \partial^\alpha b^\beta(X) b^{\alpha\gamma}(X)\Big)\, dt + b(X) \circ dW. \tag{9.5}$$

*If (9.4) holds for all $q < \infty$, then the $C^3$ condition on $b$ can be relaxed to $C^{2+}$.*

This result has been used in several contexts [7, 25], so we only sketch the proof.

PROOF OF THEOREM 9.1. First, suppose that $a \in C^{1+}$, $b \in C^3$. By [17], Theorem 12.10, we know that the map $(U_n, \mathbb{U}_n) \mapsto X_n$ is continuous with respect to the $\rho_\gamma$ topology (i.e., the rough path topology) for any $\gamma > 1/3$. In particular, the estimates (9.4), combined with the iterated invariance principle, guarantee that $(U_n, \mathbb{U}_n) \to_w (U, \mathbb{U})$ in the $\rho_\gamma$ topology for some $\gamma > 1/3$. It follows that $X_n \to_w X$ where $X$ satisfies the rough differential equation

$$X(t) = X(0) + \int_0^t a\big(X(s)\big)\,ds + \int_0^t b\big(X(s)\big)\,d(U, \mathbb{U})(s).$$

By definition of rough integrals, and the decomposition (9.3), $X$ satisfies (9.5).

Similarly, if the estimates (9.4) hold for all $q < \infty$, then we can apply [17], Theorem 12.10, under the relaxed condition $a \in C^{1+}$, $b \in C^{2+}$. $\square$

Now let $\phi_t : \Omega \to \Omega$ be a suspension flow as in Section 6, with Poincaré map $f : \Lambda \to \Lambda$. As before, we write $\Omega = \Lambda^r$, $\nu = \mu^r$, where $r : \Lambda \to \mathbb{R}$ is a roof function with $\bar{r} = \int r\,d\mu$. Let $v : \Omega \to \mathbb{R}^e$ with $\int_\Omega v\,d\nu = 0$ and define $\tilde{v} : \Lambda \to \mathbb{R}^e$, $\tilde{v}(x) = \int_0^{r(x)} v \circ \phi_t\,dt$.

COROLLARY 9.2. *Suppose that $f : \Lambda \to \Lambda$ is mixing and that $r \in L^\infty(\Lambda)$ is bounded away from zero. Suppose that $a \in C^{1+}$ and $b \in C^3$. Let $v \in L^\infty(\Omega, \mathbb{R}^e)$ with $\int_\Omega v\,d\nu = 0$. If $\tilde{v}$ admits a martingale-coboundary decomposition* (5.1) *with $p > \frac{9}{2}$, then the conclusion of Theorem* 1.2 *is valid.*

PROOF. Recall that $X_n$ satisfies (9.2) with $U_n = W_n$. By Corollary 8.1, $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ where $W$ is Brownian motion and $\mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta\,dW^\gamma + D^{\beta\gamma} t$. Moreover, the estimates (9.4) hold by Corollary 7.6. The result follows directly from Theorem 9.1. $\square$

PROOF OF THEOREM 1.2. Again we use the fact that every hyperbolic basic set for an Axiom A flow can be written as a suspension over a mixing hyperbolic basic set $f : \Lambda \to \Lambda$ with a Hölder roof function $r$. Any Hölder mean zero observable $v : \Lambda \to \mathbb{R}^e$ admits an $L^\infty$ martingale-coboundary decomposition. Hence, Theorem 1.2 follows from Corollary 9.2 together with the last statement of Theorem 9.1 (to allow for the weakened regularity assumption on $b$). $\square$

9.2. *Rough path theory in discrete time.* In this section, we introduce tools [23] that are the discrete time analogue of those introduced in the continuous rough path section. Let $U_n : [0, T] \to \mathbb{R}^e$ be a step function defined by

$$U_n(t) = \sum_{j=0}^{[nt]-1} \Delta U_{n,j}.$$

We also define the discrete iterated integral $\mathbb{U}_n : [0, T] \to \mathbb{R}^{e\times e}$ by

$$\mathbb{U}_n(t) = \int_0^t U_n(r)\, dU_n(r) = \sum_{0\le i<j<[n^{-2}t]} \Delta U_{n,i} \Delta U_{n,j}. \tag{9.6}$$

Note that, as usual, we use the left-Riemann sum convention. We define the increments

$$U_n(s,t) = \sum_{j=[ns]}^{[nt]-1} \Delta U_{n,j} \quad \text{and} \quad \mathbb{U}_n(s,t) = \sum_{[ns]\le i<j<[n^{-2}t]} \Delta U_{n,i} \Delta U_{n,j}.$$

Suppose that $a : \mathbb{R}^d \to \mathbb{R}^d$ is $C^{1+}$ and $b : \mathbb{R}^d \to \mathbb{R}^{d\times e}$ is $C^3$, and let $X_{n,j}$ be defined by the recursion

$$X_{n,j+1} = X_{n,j} + n^{-1} a(X_{n,j}) + b(X_{n,j}) \Delta U_{n,j} \tag{9.7}$$

with initial condition $X_{n,0} = \xi \in \mathbb{R}^d$. We then define the path $X_n : [0, T] \to \mathbb{R}^d$ by the rescaling $X_n(t) = X_{n,[nt]}$. The following theorem is the discrete time analogue of Theorem 9.1 and is proved in [23].

THEOREM 9.3. *Suppose that $(U_n, \mathbb{U}_n) \to_w (U, \mathbb{U})$ in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$, where $U$ is Brownian motion and where $\mathbb{U}$ can be written*

$$\mathbb{U}(t) = \int_0^t U(s)\, dU(s) + Et,$$

*for some constant matrix $E \in \mathbb{R}^{e\times e}$. Suppose moreover that there exist $C > 0$ and $q > 3$ such that*

$$\|U_n(j/n, k/n)\|_{2q} \le C \left|\frac{j-k}{n}\right|^{1/2} \quad \textit{and} \quad \|\mathbb{U}_n(j/n, k/n)\|_q \le C \left|\frac{j-k}{n}\right|, \tag{9.8}$$

*hold for all $n \ge 1$ and $j, k = 0, \dots, n$. Then $X_n \to_w X$ in $D([0,\infty), \mathbb{R}^d)$, where*

$$dX = \Big(a(X) + \sum_{\alpha,\beta,\gamma} E^{\beta\gamma} \partial^\alpha b^\beta(X) b^{\alpha\gamma}(X)\Big) dt + b(X)\, dW.$$

*If (9.8) holds for all $q < \infty$, then the $C^3$ condition on $b$ can be relaxed to $C^{2+}$.*

COROLLARY 9.4. *Suppose that $f : \Lambda \to \Lambda$ is mixing and that $a \in C^{1+}$, $b \in C^3$. Let $v \in L^\infty(\Omega, \mathbb{R}^e)$ with $\int_\Omega v\, d\nu = 0$. If $v$ admits a martingale-coboundary decomposition (5.1) with $p > \frac{9}{2}$, then the conclusion of Theorem 2.2 is valid.*

PROOF. We have that $X_n$ is defined by the recursion (9.7) with $\Delta U_{n,j} = n^{-1/2} v \circ f^j$. In particular, $U_n = W_n$ and $\mathbb{U}_n = \mathbb{W}_n$, as defined in Section 2. By Theorem 2.1, $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ where $W$ is Brownian motion and

$\mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta \, dW^\gamma + E^{\beta\gamma} t$. Moreover, the estimates (9.8) follow immediately from Corollary 7.3. Hence, the result follows from Theorem 9.3. □

PROOF OF THEOREM 2.2. Again, any Hölder mean zero observable $v : \Lambda \to \mathbb{R}^e$ admits an $L^\infty$ martingale-coboundary decomposition. Hence, Theorem 2.2 follows from Corollary 9.4 together with the last statement of Theorem 9.3. □

**10. Generalizations.** Our main results, Theorems 1.1, 1.2 for continuous time, Theorems 2.1, 2.2 for discrete time, are formulated for the well known, but restrictive, class of uniformly hyperbolic (Axiom A) diffeomorphisms and flows. In this section, we extend these results to a much larger class of systems that are nonuniformly hyperbolic in the sense of Young [54, 55]. Also, as promised, we show how to relax the mixing assumption in Theorems 2.1 and 2.2.

In Section 10.1, we consider the case of noninvertible maps modeled by Young towers. Then in Sections 10.2 and 10.3, we consider the corresponding situations for invertible maps and continuous time systems.

10.1. *Noninvertible maps modeled by Young towers.* In the noninvertible setting, a *Young tower* $f : \Delta \to \Delta$ is defined as follows. First we recall the notion of a Gibbs–Markov map $F : Y \to Y$.

Let $(Y, \mu_Y)$ be a probability space with a countable measurable partition $\alpha$, and let $F : Y \to Y$ be a Markov map. Given $x, y \in Y$, define the separation time $s(x, y)$ to be the least integer $n \ge 0$ such that $F^n x$, $F^n y$ lie in distinct partition elements of $\alpha$. It is assumed that the partition separates orbits. Given $\theta \in (0, 1)$ we define the metric $d_\theta(x, y) = \theta^{s(x,y)}$.

If $v : Y \to \mathbb{R}$ is measurable, we define $|v|_\theta = \sup_{x \ne y} |v(x) - v(y)|/d_\theta(x, y)$ and $\|v\|_\theta = \|v\|_\infty + |v|_\theta$. The space $F_\theta(Y)$ of observables $v$ with $\|v\|_\theta < \infty$ forms a Banach space with norm $\|\cdot\|_\theta$.

Let $g$ denote the inverse of the Jacobian of $F$ for the measure $\mu_Y$. We require the *good distortion* property that $|\log g|_\theta < \infty$. The map $F$ is said to be *Gibbs–Markov* if it has good distortion and *big images*: $\inf_{a \in \alpha} \mu_Y(Fa) > 0$. A special case of big images is the *full branch* condition $Fa = Y$ for all $a \in \alpha$. Gibbs–Markov maps with full branches are automatically mixing.

If $F : Y \to Y$ is a mixing Gibbs–Markov map, then observables in $F_\theta(Y)$ have exponential decay of correlations against $L^1$ observables. In particular, Theorems 2.1 and 2.2 apply in their entirety to mean zero observables $v : Y \to \mathbb{R}^e$ with components in $F_\theta(Y)$ for mixing Gibbs–Markov maps.

Given a full branch Gibbs–Markov map $F : Y \to Y$, we now introduce a *return time function* $\varphi : Y \to \mathbb{Z}^+$ assumed to be constant on partition elements. We suppose that $\varphi$ is integrable and set $\bar\varphi = \int_Y \varphi \, d\mu_Y$. Define the Young tower

$$\Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \le \ell < \varphi(y)\},$$

and define the tower map $f : \Delta \to \Delta$ by setting

$$f(y,\ell) = \begin{cases} (y,\ell+1), & \ell \le \varphi(y)-2, \\ (Fy,0), & \ell = \varphi(y)-1. \end{cases} \tag{10.1}$$

Then $\mu = \mu_Y \times \text{Lebesgue}/\bar{\varphi}$ is an ergodic $f$-invariant probability measure on $\Delta$. Note that the system $(\Delta, \mu, f)$ is uniquely determined by $(Y, \mu_Y, F)$ together with $\varphi$.

The separation time $s(x, y)$ extends to the tower by setting $s((x,\ell),(y,\ell')) = 0$ for $\ell \ne \ell'$ and $s((x,\ell),(y,\ell)) = s(x,y)$. The metric $d_\theta$ extends accordingly to $\Delta$ and we define the space $F_\theta(\Delta)$ of observables $v : \Delta \to \mathbb{R}$ that lie in $L^\infty(\Delta)$ and are Lipschitz with respect to this metric.

The tower map $f : \Delta \to \Delta$ is mixing if and only if $\gcd\{\varphi(a) : a \in \alpha\} = 1$. In the mixing case, it follows from Young [54, 55] that the rate of decay of correlations on the tower $\Delta$ is determined by the tail function

$$\mu(\varphi > n) = \mu\big(y \in Y : \varphi(y) > n\big).$$

In [54], it is shown that exponential decay of $\mu(\varphi > n)$ implies exponential decay of correlations for observables in $F_\theta(\Delta)$, and [55] shows that if $\mu(\varphi > n) = O(n^{-\beta})$ then correlations for such observables decay at a rate that is $O(n^{-(\beta-1)})$. For systems that are *modeled by a Young tower*, Hölder observables for the underlying dynamical system lift to observables in $F_\theta(\Delta)$ (for appropriately chosen $\theta$) and thereby inherit the above results on decay of correlations. Similarly, if we define $F_\theta(\Delta, \mathbb{R}^e)$ to consist of observables $v : \Delta \to \mathbb{R}^e$ with components in $F_\theta(\Delta)$, then results on weak convergence for vector-valued Hölder observables are inherited by the lifted observables in $F_\theta(\Delta, \mathbb{R}^e)$ and so it suffices to prove everything at the Young tower level.

THEOREM 10.1. *Suppose that $f : \Delta \to \Delta$ is a mixing Young tower with return time function $\varphi : Y \to \mathbb{Z}^+$ satisfying $\mu(\varphi > n) = O(n^{-\beta})$. Let $v \in F_\theta(\Delta, \mathbb{R}^e)$ with $\int_\Delta v \, d\mu = 0$. Then*:

(a) *Iterated WIP*: *If $\beta > 3$, then the conclusions of Theorem 2.1 are valid.*

(b) *Convergence to SDE*: *If $\beta > \frac{11}{2}$, then the conclusions of Theorem 2.2 are valid for all $a \in C^{1+}$, $b \in C^3$.*

*In particular, Theorems 2.1 and 2.2 are valid for systems modeled by Young towers with exponential tails for all $a \in C^{1+}$, $b \in C^{2+}$.*

PROOF. In the setting of noninvertible (one-sided) Young towers [55], given $v \in F_\theta(\Delta)$ with mean zero, there is a constant $C$ such that

$$\left| \int_\Delta v w \circ f^n \, d\mu \right| \le C \|w\|_\infty n^{-(\beta-1)} \qquad \text{for all } w \in L^\infty, n \ge 1.$$

Hence, by Proposition 4.4, there is an $L^p$ martingale-coboundary decomposition (4.1) for any $p < \beta - 1$. The desired results follow from Theorem 4.3 and Corollary 9.2, respectively. $\square$

If $\beta > 2$, or more generally $\varphi \in L^2$, the WIP is well known. In fact, $\varphi \in L^2$ suffices also for the iterated WIP and the mixing assumption on $f$ is unnecessary, as shown in the next result. These assumptions are optimal, since the ordinary CLT is generally false when $\varphi \notin L^2$.

THEOREM 10.2. *Suppose that $\Delta$ is a Young tower with return time function $\varphi \in L^2$. Let $v \in F_\theta(\Delta, \mathbb{R}^e)$ with $\int_\Delta v\, d\mu = 0$. Then $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ where $W$ is an $e$-dimensional Brownian motion with covariance matrix*

$$\begin{aligned}\Sigma^{\beta\gamma} &= \operatorname{Cov}^{\beta\gamma} W(1) \\ &= (\bar{\varphi})^{-1} \int_Y \tilde{v}^\beta \tilde{v}^\gamma \, d\mu_Y + (\bar{\varphi})^{-1} \sum_{n=1}^{\infty} \int_Y \big(\tilde{v}^\beta \tilde{v}^\gamma \circ F^n + \tilde{v}^\gamma \tilde{v}^\beta \circ F^n\big)\, d\mu_Y,\end{aligned}$$

*and $\mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta \, dW^\gamma + E^{\beta\gamma} t$ where*

$$E^{\beta\gamma} = (\bar{\varphi})^{-1} \sum_{n=1}^{\infty} \int_Y \tilde{v}^\beta \tilde{v}^\gamma \circ F^n \, d\mu_Y + \int_\Delta H^\beta v^\gamma \, d\mu, \qquad H(y, \ell) = \sum_{j=0}^{\ell-1} v(y, j).$$

*If moreover $\mu(\varphi > n) = O(n^{-\beta})$ for some $\beta > \frac{11}{2}$, then the conclusion of Theorem 2.2 (convergence to SDE) holds for all $a \in C^{1+}$, $b \in C^3$.*

PROOF. We use the discrete analogue of the inducing method used in the proof of Theorem 6.1. Define $\tilde{v} : Y \to \mathbb{R}^e$ by setting $\tilde{v}(y) = \sum_{j=0}^{\varphi(y)-1} v(f^j y)$. Then $\tilde{v}$ lies in $L^2$ and $\int_Y \tilde{v}\, d\mu_Y = 0$. Let $P$ denote the transfer operator for $F : Y \to Y$. Although $\tilde{v} \notin F_\theta(Y, \mathbb{R}^e)$ an elementary calculation [33], Lemma 2.2, shows that $P\tilde{v} \in F_\theta(Y, \mathbb{R}^e)$. In particular, $P\tilde{v}$ has exponential decay of correlations against $L^1$ observables. It follows that $\chi = \sum_{j=1}^{\infty} P^j \tilde{v}$ converges in $L^\infty$, and hence following the proof of Proposition 4.4, we obtain that $\tilde{v}$ admits an $L^2$ martingale-coboundary decomposition.

Define the cadlag processes $\widetilde{W}_n$, $\widetilde{\mathbb{W}}_n$ as in (2.1) using $\tilde{v}$ instead of $v$. It follows from Theorem 4.3 that $(\widetilde{W}_n, \widetilde{\mathbb{W}}_n) \to_w (\widetilde{W}, \widetilde{\mathbb{W}})$ where $\widetilde{W}$ is an $e$-dimensional Brownian motion and $\widetilde{\mathbb{W}}^{\beta\gamma}(t) = \int_0^t \widetilde{W}^\beta \, d\widetilde{W}^\gamma + \tilde{E}^{\beta\gamma} t$ with

$$\operatorname{Cov}^{\beta\gamma} \tilde{W}(1) = \int_Y \tilde{v}^\beta \tilde{v}^\gamma \, d\mu_Y + \sum_{n=1}^{\infty} \int_Y \big(\tilde{v}^\beta \tilde{v}^\gamma \circ F^n + \tilde{v}^\gamma \tilde{v}^\beta \circ F^n\big)\, d\mu_Y,$$

and $\tilde{E}^{\beta\gamma} = \sum_{n=1}^{\infty} \int_Y \tilde{v}^\beta \tilde{v}^\gamma \circ F^n \, d\mu_Y$.

Arguing as in the proof of Theorem 6.1, and noting Remark 6.2, we obtain that $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ where

$$W = (\bar{\varphi})^{-1/2}\widetilde{W}, \qquad \mathbb{W}^{\beta\gamma}(t) = \int_0^t W^\beta \, dW^\gamma + E^{\beta\gamma} t,$$
$$E^{\beta\gamma} = (\bar{\varphi})^{-1}\tilde{E}^{\beta\gamma} + \int_\Delta H^\beta v^\gamma \, d\mu.$$

Finally, to prove the last statement of the theorem, it suffices by Corollary 9.2 to show that $v$ admits an $L^p$ martingale-coboundary decomposition with $p > \frac{9}{2}$. We already saw that this holds for $\Delta$ mixing, equivalently $d = \gcd\{\varphi(a) : a \in \alpha\} = 1$. If $d > 1$, then $\Delta$ can be written as a disjoint union of $d$ towers $\Delta_k$ each with a Gibbs–Markov map that is a copy of $F$ and return time function $1_{\Delta_k}\varphi/d$. Each of these $d$ towers is mixing under $f^d$, and the towers are cyclically permuted by $f$. Hence,

$$\sum_{m=1}^{\infty} P^m \tilde{v} = \sum_{k,r=1}^{\infty} \sum_{m=0}^{\infty} P^{md+r}\Big(1_{\Delta_k}\tilde{v} - d\int_\Delta 1_{\Delta_k}\tilde{v}\, d\mu\Big).$$

But $\|P^{md}(1_{\Delta_k}\tilde{v} - d\int_\Delta 1_{\Delta_k}\tilde{v}\, d\mu)\|_p \ll m^{-\beta}$. Hence, we can define $\chi = \sum_{m=1}^\infty P^m \tilde{v} \in L^p$ yielding the desired decomposition $\tilde{v} = m + \chi \circ f - \chi$. $\square$

EXAMPLE 10.3. A prototypical family of nonuniformly expanding maps are intermittent maps $f : [0, 1] \to [0, 1]$ of Pomeau–Manneville type [28, 44] given by

$$fx = \begin{cases} x(1 + 2^\alpha x^\alpha), & x \in [0, \frac{1}{2}), \\ 2x - 1, & x \in [\frac{1}{2}, 1]. \end{cases}$$

For each $\alpha \in [0, 1)$, there is a unique absolutely continuous invariant probability measure $\mu$. For $\alpha \in (0, 1)$, there is a neutral fixed point at 0 and the system is modeled by a mixing Young tower with tails that are $O(n^{-\beta})$ where $\beta = \alpha^{-1}$.

Hence, the results of this paper apply in their entirety for $\alpha \in [0, \frac{2}{11})$. Further, it is well known that the WIP holds if and only if $\alpha \in [0, \frac{1}{2})$, and we recover this result, together with the iterated WIP, for $\alpha \in [0, \frac{1}{2})$.

10.2. *Invertible maps modeled by Young towers.* A large class of nonuniformly hyperbolic diffeomorphisms (possibly with singularities) can be modeled by two-sided Young towers with exponential and polynomial tails. For such towers, Theorems 10.1 and 10.2 go through essentially without change. The definitions are much more technical, but we sketch some of the details here.

Let $(M, d)$ be a Riemannian manifold. Young [54] introduced a class of nonuniformly hyperbolic maps $T : M \to M$ with the property that there is an ergodic $T$-invariant SRB measure for which exponential decay of correlations holds for Hölder observables. We refer to [54] for the precise definitions, and restrict here to providing the notions and notation required for understanding the results presented here. In particular, there is a "uniformly hyperbolic" subset $Y \subset M$ with partition $\{Y_j\}$ and return time function $\varphi : Y \to \mathbb{Z}^+$ (denoted $R$ in [54]) constant on partition elements. For each $j$, it is assumed that $T^{\varphi(j)}(Y_j) \subset Y$. We define the *induced map* $F = T^{\varphi} : Y \to Y$.

Define the *(two-sided) Young tower* $\Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \leq \ell < \varphi(y)\}$ and define the tower map $f : \Delta \to \Delta$ using the formula (10.1).

It is assumed moreover that there is an $F$-invariant foliation of $Y$ by "stable disks," and that this foliation extends up the tower $\Delta$. We obtain the quotient tower map $\bar{f} : \bar{\Delta} \to \bar{\Delta}$ and quotient induced map $\bar{F} = \bar{f}^{\varphi} : \bar{Y} \to \bar{Y}$. The hypotheses in [54] guarantee that:

PROPOSITION 10.4. *There exists an ergodic $T$-invariant probability measure $\nu$ on $M$, and ergodic invariant probability measures $\mu_\Delta$, $\mu_{\bar{\Delta}}$, $\mu_Y$, $\mu_{\bar{Y}}$ defined on $\Delta$, $\bar{\Delta}$, $Y$, $\bar{Y}$, respectively, such that*:

(a) *The projection $\pi : \Delta \to M$ given by $\pi(y, \ell) = T^{\ell} y$, and the projections $\bar{\pi} : \Delta \to \bar{\Delta}$ and $\bar{\pi} : Y \to \bar{Y}$ given by quotienting, are measure preserving.*

(b) *The return time function $\varphi : Y \to \mathbb{Z}^+$ is integrable with respect to $\mu_Y$ (and hence also with respect to $\mu_{\bar{Y}}$ when regarded as a function on $\bar{Y}$).*

(c) $\mu_\Delta = \mu_Y \times \text{counting}/\int_Y \varphi \, d\mu$ *and* $\mu_{\bar{\Delta}} = \mu_{\bar{Y}} \times \text{counting}/\int_Y \varphi \, d\mu$.

(d) *The system $(\bar{Y}, \bar{F}, \mu_{\bar{Y}})$ is a full branch Gibbs–Markov map with partition $\alpha = \{\bar{Y}_j\}$. Hence, $\bar{f} : \bar{\Delta} \to \bar{\Delta}$ is a one-sided Young tower as in Section 10.1.*

(e) $\mu_Y(\varphi > n) = O(e^{-an})$ *for some $a > 0$.*

(f) *Let $v : M \to \mathbb{R}$ be Hölder with $\int_M v \, d\nu = 0$. Then $v \circ \pi = \bar{v} \circ \bar{\pi} + \chi_1 \circ f - \chi_1$ where $\chi_1 \in L^\infty(\Delta)$ and $\bar{v} \in F_\theta(\bar{\Delta})$ for some $\theta \in (0, 1)$.*

PROOF. Parts (a)–(e) can be found in [54]. For part (f) see, for example, [32, 33]. □

COROLLARY 10.5. *Theorems 2.1 and 2.2 are valid for Hölder mean zero observables of systems modeled by (two-sided) mixing Young towers with exponential tails.*

PROOF. By Proposition 10.4(d) and the proof of Theorem 10.1, for any $p$ we can decompose $\bar{v} \in F_\theta(\Delta)$ as $\bar{v} = \bar{m} + \bar{\chi}_2 \circ \bar{f} - \bar{\chi}_2$ where $\bar{m}, \bar{\chi}_2 \in L^\infty(\bar{\Delta})$ and $\bar{m}$ lies in the kernel of the transfer operator corresponding to $\bar{F} : \bar{\Delta} \to \bar{\Delta}$. Now let $m = \bar{m} \circ \bar{\pi}$ and $\chi = \chi_1 + \bar{\chi}_2 \circ \bar{\pi}$ where $\chi_1$ is as in Proposition 10.4(f). We have shown that $v \circ \pi$ admits an $L^\infty$ martingale-coboundary decomposition (5.1). By Theorem 5.2, we obtain the required results for $v \circ \pi$, and hence for $v$. □

By [5], this includes the important example of Hénon-like attractors. Again the results hold with the appropriate modifications (in the formulas for $\Sigma$ and $E$) for nonmixing towers with exponential tails.

A similar situation holds for systems modeled by (two-sided) Young towers with polynomial tails where Proposition 10.4(a)–(d) are unchanged and part (e) is replaced by the condition that $\mu_Y(\varphi > n) = O(n^{-\beta})$. In general, part (f) needs modifying. The simplest case is where there is sufficiently fast uniform contraction along stable manifolds (exponential as assumed in [2, 32, 33], or polynomial as in [1]). Then part (f) is unchanged allowing us to reduce to the situations of Theorem 10.1 in the mixing case, $\beta > 3$, and Theorem 10.2 in the remaining cases.

In the general setting of Young towers with subexponential tails, there is contraction/expansion only on visits to $Y$ and Proposition 10.4(f) fails. In this case, an alternative construction [39] can be used to reduce from $M$ to $Y$ and then to $\bar{Y}$. Define the induced observable $\tilde{v}$ on $Y$ by setting $\tilde{v}(y) = \sum_{\ell=0}^{\varphi(y)-1} v(T^\ell y)$. If $\varphi \in L^p$ (which is the case for all $p < \beta$) then it is shown in [39] that $\tilde{v} = \bar{m} \circ \bar{\pi} + \chi \circ F - \chi$ where $\bar{m} \in L^p(\bar{Y})$ lies in the kernel of the transfer operator for $\bar{F} : \bar{Y} \to \bar{Y}$ and $\chi \in L^p(Y)$. Thus, if $\varphi \in L^2$, we obtain the iterated WIP for $\tilde{v}$, and hence for $v$.

10.3. *Semiflows and flows modeled by Young towers.* Finally, we note that the results for noninvertible and invertible maps modeled by a Young tower pass over to suspension semiflows and flows defined over such maps. Using the methods in Sections 6 and 7, we reduce from observables defined on the flow to observables defined on the Young tower, where we can apply the results from Sections 10.1 and 10.2. We refer to [33] for a description of numerous examples of flows that can be reduced to maps in this way.

We mention here the classical Lorenz attractor for which Theorems 1.1 and 1.2 follow as a consequence of such a construction. There are numerous methods to proceed with the Lorenz attractor, but probably the simplest is as follows. The Poincaré map is a Young tower with exponential tails, but the roof function for the flow has a logarithmic singularity, and hence is unbounded. An idea in [4] is to remodel the flow as a suspension with bounded roof function over a mixing Young tower $\Delta$ with slightly worse, namely stretched exponential, tails. In particular, the return time function for $\Delta$ still lies in $L^p$ for all $p$. Hölder observables for the flow can now be shown to induce to observables in $F_\theta(\Delta)$, thereby reducing to the situation of Section 10.2. Moreover, the flow for the Lorenz attractor has exponential contraction along stable manifolds, and this is inherited by each of the Young tower models described above. Hence, we can reduce to the situation in Theorem 10.1 with $\beta$ arbitrarily large.

**Acknowledgements.** We are grateful to Martin Hairer and Andrew Stuart for helpful comments.

## REFERENCES

[1] ALVES, J. F. and AZEVEDO, D. (2013). Statistical properties of diffeomorphisms with weak invariant manifolds. Preprint.

[2] ALVES, J. F. and PINHEIRO, V. (2008). Slow rates of mixing for dynamical systems with hyperbolic structures. *J. Stat. Phys.* **131** 505–534. MR2386574

[3] ANOSOV, D. V. (1967). Geodesic flows on closed Riemannian manifolds of negative curvature. *Proc. Steklov Inst.* **90** 1–209. MR0224110

[4] BÁLINT, P. and MELBOURNE, I. (2010). Decay of correlations for flows with unbounded roof function, including the infinite horizon planar periodic Lorentz gas. Preprint.

[5] BENEDICKS, M. and YOUNG, L.-S. (2000). Markov extensions and decay of correlations for certain Hénon maps. *Astérisque* **261** 13–56. MR1755436

[6] BOWEN, R. (1975). *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*. *Lecture Notes in Mathematics* **470**. Springer, Berlin. MR0442989

[7] BREUILLARD, E., FRIZ, P. and HUESMANN, M. (2009). From random walks to rough paths. *Proc. Amer. Math. Soc.* **137** 3487–3496. MR2515418

[8] BROWN, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66. MR0290428

[9] BUNIMOVICH, L. A., SINAĬ, Y. G. and CHERNOV, N. I. (1991). Statistical properties of two-dimensional hyperbolic billiards. *Uspekhi Mat. Nauk* **46** 43–92. MR1138952

[10] BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42. MR0365692

[11] DENKER, M. and PHILIPP, W. (1984). Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergodic Theory Dynam. Systems* **4** 541–552. MR0779712

[12] DOLGOPYAT, D. (1998). On decay of correlations in Anosov flows. *Ann. of Math.* (2) **147** 357–390. MR1626749

[13] DOLGOPYAT, D. (1998). Prevalence of rapid mixing in hyperbolic flows. *Ergodic Theory Dynam. Systems* **18** 1097–1114. MR1653299

[14] DOLGOPYAT, D. (2004). Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.* **356** 1637–1689 (electronic). MR2034323

[15] DOLGOPYAT, D. (2005). Averaging and invariant measures. *Mosc. Math. J.* **5** 537–576, 742. MR2241812

[16] FIELD, M., MELBOURNE, I. and TÖRÖK, A. (2007). Stability of mixing and rapid mixing for hyperbolic flows. *Ann. of Math.* (2) **166** 269–291. MR2342697

[17] FRIZ, P. K. and VICTOIR, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths*: *Theory and Applications*. *Cambridge Studies in Advanced Mathematics* **120**. Cambridge Univ. Press, Cambridge. MR2604669

[18] GIVON, D., KUPFERMAN, R. and STUART, A. (2004). Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity* **17** R55–R127. MR2097022

[19] GOUËZEL, S. (2007). Statistical properties of a skew product with a curve of neutral points. *Ergodic Theory Dynam. Systems* **27** 123–151. MR2297091

[20] GOUËZEL, S. (2010). Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probab.* **38** 1639–1671. MR2663640

[21] HUISINGA, W., SCHÜTTE, C. and STUART, A. M. (2003). Extracting macroscopic stochastic dynamics: Model problems. *Comm. Pure Appl. Math.* **56** 234–269. MR1934621

[22] JAKUBOWSKI, A., MÉMIN, J. and PAGÈS, G. (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbf{D}^1$ de Skorokhod. *Probab. Theory Related Fields* **81** 111–137. MR0981569

[23] KELLY, D. T. B. (2014). Rough path recursions and diffusion approximations. Preprint.

[24] KURTZ, T. G. and PROTTER, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070. MR1112406

[25] LEJAY, A. and LYONS, T. (2005). On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. In *Current Trends in Potential Theory*. *Theta Ser. Adv. Math.* **4** 63–84. Theta, Bucharest. MR2243956

[26] LESIGNE, E. and VOLNÝ, D. (2001). Large deviations for martingales. *Stochastic Process. Appl*. **96** 143–159. MR1856684

[27] LIVERANI, C. (2004). On contact Anosov flows. *Ann. of Math.* (2) **159** 1275–1312. MR2113022

[28] LIVERANI, C., SAUSSOL, B. and VAIENTI, S. (1999). A probabilistic approach to intermittency. *Ergodic Theory Dynam. Systems* **19** 671–685. MR1695915

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

[30] MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 620–628. MR0358933

[31] MCSHANE, E. J. (1972). Stochastic differential equations and models of random processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. III*: *Probability Theory* 263–294. Univ. California Press, Berkeley, CA. MR0402921

[32] MELBOURNE, I. (2007). Rapid decay of correlations for nonuniformly hyperbolic flows. *Trans. Amer. Math. Soc*. **359** 2421–2441 (electronic). MR2276628

[33] MELBOURNE, I. and NICOL, M. (2005). Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys*. **260** 131–146. MR2175992

[34] MELBOURNE, I. and NICOL, M. (2008). Large deviations for nonuniformly hyperbolic systems. *Trans. Amer. Math. Soc*. **360** 6661–6676. MR2434305

[35] MELBOURNE, I. and NICOL, M. (2009). A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Ann. Probab.* **37** 478–505. MR2510014

[36] MELBOURNE, I. and STUART, A. M. (2011). A note on diffusion limits of chaotic skew-product flows. *Nonlinearity* **24** 1361–1367. MR2776125

[37] MELBOURNE, I. and TÖRÖK, A. (2004). Statistical limit theorems for suspension flows. *Israel J. Math*. **144** 191–209. MR2121540

[38] MELBOURNE, I. and TÖRÖK, A. (2012). Convergence of moments for Axiom A and nonuniformly hyperbolic flows. *Ergodic Theory Dynam. Systems* **32** 1091–1100. MR2995657

[39] MELBOURNE, I. and VARANDAS, P. (2014). A note on statistical properties for nonuniformly hyperbolic systems with subexponential contraction and expansion. Preprint.

[40] MELBOURNE, I. and ZWEIMÜLLER, R. (2015). Weak convergence to stable Lévy processes for nonuniformly hyperbolic dynamical systems. *Ann. Inst. Henri Poincaré Probab. Stat*. **51** 545–556. MR3335015

[41] MÓRICZ, F. (1976). Moment inequalities and the strong laws of large numbers. *Z. Wahrsch. Verw. Gebiete* **35** 299–314. MR0407950

[42] PAPANICOLAOU, G. C. and KOHLER, W. (1974). Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math*. **27** 641–668. MR0368142

[43] PAVLIOTIS, G. A. and STUART, A. M. (2008). *Multiscale Methods*: *Averaging and Homogenization*. *Texts in Applied Mathematics* **53**. Springer, New York. MR2382139

[44] POMEAU, Y. and MANNEVILLE, P. (1980). Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys*. **74** 189–197. MR0576270

[45] RATNER, M. (1973). The central limit theorem for geodesic flows on $n$-dimensional manifolds of negative curvature. *Israel J. Math*. **16** 181–197. MR0333121

[46] RUELLE, D. (1978). *Thermodynamic Formalism*. *Encyclopedia of Mathematics and Its Applications* **5**. Addison-Wesley, Reading, MA. MR0511655

[47] SERFLING, R. J. (1970). Moment inequalities for the maximum cumulative sum. *Ann. Math. Statist*. **41** 1227–1234. MR0268938

[48] SINAĬ, J. G. (1972). Gibbs measures in ergodic theory. *Russ. Math. Surv.* **27** 21–70.

[49] SKOROHOD, A. V. (1956). Limit theorems for stochastic processes. *Teor. Veroyatn. Primen*. **1** 289–319. MR0084897

[50] SMALE, S. (1967). Differentiable dynamical systems. *Bull. Amer. Math. Soc.* (*N.S.*) **73** 747–817. MR0228014

[51] SUSSMANN, H. J. (1978). On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probab.* **6** 19–41. MR0461664

[52] SUSSMANN, H. J. (1991). Limits of the Wong–Zakai type with a modified drift term. In *Stochastic Analysis* 475–493. Academic Press, Boston, MA. MR1119845

[53] WONG, E. and ZAKAI, M. (1965). On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.* **36** 1560–1564. MR0195142

[54] YOUNG, L.-S. (1998). Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* (2) **147** 585–650. MR1637655

[55] YOUNG, L.-S. (1999). Recurrence times and rates of mixing. *Israel J. Math.* **110** 153–188. MR1750438

[56] ZWEIMÜLLER, R. (2007). Mixing limit theorems for ergodic transformations. *J. Theoret. Probab.* **20** 1059–1071. MR2359068

COURANT INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
NEW YORK, NEW YORK 10012-1185
USA
E-MAIL: dtkelly@cims.nyu.edu

MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: i.melbourne@warwick.ac.uk