

Wasserstein convergence rates in the invariance principle for deterministic dynamical systems

ZHENXIN LIU and ZHE WANG

*School of Mathematical Sciences, Dalian University of Technology,
Dalian 116024, P. R. China*

(e-mail: zxliu@dlut.edu.cn, zwangmath@hotmail.com, wangz1126@mail.dlut.edu.cn)

(Received 1 April 2022 and accepted in revised form 25 April 2023)

Abstract. In this paper, we consider the convergence rate with respect to Wasserstein distance in the invariance principle for deterministic non-uniformly hyperbolic systems. Our results apply to uniformly hyperbolic systems and large classes of non-uniformly hyperbolic systems including intermittent maps, Viana maps, unimodal maps and others. Furthermore, as a non-trivial application to the homogenization problem, we investigate the Wasserstein convergence rate of a fast–slow discrete deterministic system to a stochastic differential equation.

Key words: invariance principle, rate of convergence, Wasserstein distance, non-uniformly hyperbolic system, homogenization

2020 Mathematics Subject Classification: 37A50, 60F17 (Primary); 37D20, 60B10 (Secondary)

1. Introduction

It is well known that deterministic dynamical systems can exhibit some statistical limit properties if the system is chaotic enough and the observable satisfies some regularity conditions. In recent years, we have seen growing research interests in statistical limit properties for deterministic systems such as the law of large numbers (or Birkhoff's ergodic theorem), central limit theorem (CLT), weak invariance principle (WIP), almost sure invariance principle (ASIP), large deviations and so on.

The WIP (also known as the functional CLT) states that a stochastic process constructed by the sums of random variables with suitable scale converges weakly to Brownian motion, which is a far-reaching generalization of the CLT. Donsker's theorem [13] is the prototypical invariance principle, which deals with independent and identically distributed random variables. Later, different versions were extensively studied. In particular, many authors studied the WIP and ASIP for dynamical systems with some hyperbolicity. Denker and Philipp [11] proved the ASIP for uniformly hyperbolic diffeomorphisms and

flows. The results are stronger, as the ASIP implies the WIP and CLT. Melbourne and Nicol [24] investigated the ASIP for non-uniformly expanding maps and non-uniformly hyperbolic diffeomorphisms that can be modelled by a Young tower [33, 34], and they also obtained corresponding results for flows. After that, there are many works on the WIP for non-uniformly hyperbolic systems, which we will not mention here.

To the best of our knowledge, there are only two works on rates of convergence in the WIP for deterministic dynamical systems in spite of the fact that there are many results on the convergence itself. In his PhD thesis [1], Antoniou obtained the rate of convergence in the Lévy–Prokhorov distance for uniformly expanding maps using the martingale approximation method and applying an estimate for martingale difference arrays [22] by Kubilyus. Then, following the same method, together with Melbourne, he [2] further generalized the convergence rates to non-uniformly expanding/hyperbolic systems. Specifically, they apply a new version of the martingale-coboundary decomposition [21] by Korepanov et al.

The Wasserstein distance has been used extensively in recent years to metrize weak convergence. It is stronger and contains more information than the Lévy–Prokhorov distance since it involves the metric of the underlying space. This distance finds important applications in the fields of optimal transport, geometry, partial differential equations and so on; see, e.g., Villani [32] for details. There are some results on Wasserstein convergence rates for the CLT in the community of probability and statistics; see, e.g., [10, 26, 29]. However, to our knowledge, there are no related results on the invariance principle for dynamical systems. Motivated by [1, 2], we aim to estimate the Wasserstein convergence rate in the WIP for non-uniformly hyperbolic systems.

Following the procedure of [2], we first consider a martingale as an intermediary process. In [2], the authors apply a result of Kubilyus [22] and the key is to estimate the distance between W_n , defined in (3.1) below, and the intermediary process. In the present paper, we use the ideas in [2] to estimate the distance between W_n and the intermediary process. Hence, most of our efforts are to deal with the Wasserstein distance between the intermediary process and Brownian motion, which is handled by a martingale version of the Skorokhod embedding theorem. In this way, we obtain the rate of convergence $O(n^{-1/4+\delta})$ in the Wasserstein distance, where δ depends on the degree of non-uniformity. When the system that can be modelled by a Young tower has a superpolynomial tail, δ can be arbitrarily small.

Our results are applicable to uniformly hyperbolic systems and large classes of non-uniformly hyperbolic systems modelled by a Young tower with superpolynomial and polynomial tails. In comparison with [2], when the dynamical system has a superpolynomial tail, we can obtain the same convergence rate $O(n^{-1/4+\delta})$ for δ arbitrarily small. However, in our case, the price to pay is that the dynamical system needs to have stronger mixing properties. For example, we consider the Pomeau–Manneville intermittent map (3.2), which has a polynomial tail. By [2], the convergence rate is $O(n^{-1/4+\gamma/2+\delta})$ in the Lévy–Prokhorov distance for $\gamma \in (0, \frac{1}{2})$, but we obtain the Wasserstein convergence rate $O(n^{-1/4+\gamma/(4(1-\gamma))+\delta})$ only for $\gamma \in (0, \frac{1}{4})$. See Example 3.6 for details.

As a non-trivial application, we consider the deterministic homogenization in fast–slow dynamical systems. In [15], Gottwald and Melbourne proved that the slow variable with

suitable scales converges weakly to the solution of a stochastic differential equation. Then Antoniou and Melbourne [2] studied the weak convergence rate of the above problem based on the convergence rate in the WIP with respect to the Lévy–Prokhorov distance. In this paper, we obtain the Wasserstein convergence rate for the homogenization problem based on our results. In comparison with [2], for uniformly hyperbolic fast systems, we obtain the same convergence rate $O(\epsilon^{1/3-\delta})$, where δ can be arbitrarily small and ϵ is identified with $n^{-1/2}$. However, for non-uniformly hyperbolic fast systems, we need to request stronger mixing properties than in [2]. See Remark 5.2 for details.

The remainder of this paper is organized as follows. In §2, we give the definition and basic properties of Wasserstein distances. In §3, we review the definitions of non-uniformly expanding maps and non-uniformly hyperbolic diffeomorphisms and we state the main results in this paper. In §4, we first introduce the method of martingale approximation and summarize some required properties and then we prove the main results. In the last section, we give an application to fast–slow systems.

Throughout the paper, we use 1_A to denote the indicator function of measurable set A . As usual, $a_n = o(b_n)$ means that $\lim_{n \rightarrow \infty} a_n/b_n = 0$, $a_n = O(b_n)$ means that there exists a constant $C > 0$ such that $|a_n| \leq C|b_n|$ for all $n \geq 1$ and $\|\cdot\|_{L^p}$ means the L^p -norm. For simplicity, we write C to denote constants independent of n and C may change from line to line. We use \rightarrow_w to denote the weak convergence in the sense of probability measures [5]. We denote by $C[0, 1]$ the space of all continuous functions on $[0, 1]$ equipped with the supremum distance d_C , that is,

$$d_C(x, y) := \sup_{t \in [0,1]} |x(t) - y(t)|, \quad x, y \in C[0, 1].$$

We use \mathbb{P}_X to denote the law/distribution of random variable X and use $X =_d Y$ to mean X, Y sharing the same distribution.

2. Preliminaries

In this section, we review the definition of Wasserstein distances and some important properties about the distance. See, e.g., [8, 28, 32] for details.

Let (\mathcal{X}, d) be a Polish space, that is, a complete separable metric space, equipped with the Borel σ -algebra \mathcal{B} . Given two probability measures μ and ν on \mathcal{X} , take two random variables X and Y such that $\text{law}(X) = \mu, \text{law}(Y) = \nu$. Then the pair (X, Y) is called a *coupling* of μ and ν ; the joint distribution of (X, Y) is also called a *coupling* of μ and ν .

Definition 2.1. Let $q \in [1, \infty)$. Then, for any two probability measures μ and ν on \mathcal{X} , the *Wasserstein distance of order q* between them is defined by

$$\begin{aligned} \mathcal{W}_q(\mu, \nu) &:= \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}} d(x, y)^q \, d\pi(x, y) \right)^{1/q} \\ &= \inf\{[\mathbb{E}d(X, Y)^q]^{1/q}; \text{law}(X) = \mu, \text{law}(Y) = \nu\}, \end{aligned}$$

where $\Pi(\mu, \nu)$ is the set of all couplings of μ and ν .

PROPOSITION 2.2. (See [8, Lemma 5.2]) *Given two probability measures μ and ν on \mathcal{X} , the infimum in Definition 2.1 can be attained for some coupling (X, Y) of μ and ν .*

Those couplings achieving the infimum in Proposition 2.2 are called *optimal couplings* of μ and ν . Note also that the distance $\mathcal{W}_q(\mu, \nu)$ can be bounded above by the L^q distance of any coupling (X, Y) of μ and ν .

PROPOSITION 2.3. (See [8, Theorem 5.6] or [32, Definition 6.8]) $\mathcal{W}_q(\mu_n, \mu) \rightarrow 0$ if and only if the following two conditions hold:

- (1) $\mu_n \rightarrow_w \mu$; and
- (2) $\int_{\mathcal{X}} d(x, x_0)^q d\mu_n(x) \rightarrow \int_{\mathcal{X}} d(x, x_0)^q d\mu(x)$ for some (thus any) $x_0 \in \mathcal{X}$.

In particular, if d is bounded, then the convergence with respect to \mathcal{W}_q is equivalent to the weak convergence.

PROPOSITION 2.4. Suppose that $\mathcal{G} : \mathcal{X} \rightarrow \mathcal{X}$ is Lipschitz continuous with constant K . Then, for any two probability measures μ and ν on \mathcal{X} and $q \in [1, \infty)$,

$$\mathcal{W}_q(\mu \circ \mathcal{G}^{-1}, \nu \circ \mathcal{G}^{-1}) \leq K \mathcal{W}_q(\mu, \nu).$$

Proof. By Proposition 2.2, we can choose an optimal coupling (X, Y) of μ and ν such that

$$[\mathbf{E}d(X, Y)^q]^{1/q} = \mathcal{W}_q(\mu, \nu).$$

Then

$$\begin{aligned} \mathcal{W}_q(\mu \circ \mathcal{G}^{-1}, \nu \circ \mathcal{G}^{-1}) &\leq [\mathbf{E}d(\mathcal{G}(X), \mathcal{G}(Y))^q]^{1/q} \\ &\leq K [\mathbf{E}d(X, Y)^q]^{1/q} = K \mathcal{W}_q(\mu, \nu). \end{aligned} \quad \square$$

Remark 2.5. In the following, for simplicity, we use the notation $\mathcal{W}_p(X, Y)$ to mean $\mathcal{W}_p(\mathbb{P}_X, \mathbb{P}_Y)$. However, we should keep in mind that (X, Y) need not be an optimal coupling of $(\mathbb{P}_X, \mathbb{P}_Y)$.

The following result is known; see, e.g., [8, Lemma 5.3] or [28, Corollary 8.3.1] for details. However, the forms or proofs in these references are different from the following, which is more appropriate for our purpose. For the convenience of the reader, we also give a proof.

PROPOSITION 2.6. For any given probability measures μ and ν on \mathcal{X} and $p \in [1, \infty)$,

$$\pi(\mu, \nu) \leq \mathcal{W}_p(\mu, \nu)^{p/(p+1)},$$

where π is the Lévy–Prokhorov distance defined by

$$\pi(\mu, \nu) := \inf\{\epsilon > 0 : \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for all closed sets } A \in \mathcal{B}\}.$$

Here A^ϵ denotes the ϵ -neighbourhood of A .

Proof. Let A be a closed set. Then, for any coupling (X, Y) of μ and ν ,

$$\begin{aligned} \mathbb{P}(X \in A) &\leq \mathbb{P}(Y \in A^\epsilon) + \mathbb{P}(d(X, Y) \geq \epsilon) \\ &\leq \mathbb{P}(Y \in A^\epsilon) + \frac{\mathbf{E}d(X, Y)^p}{\epsilon^p}. \end{aligned}$$

Note that $\mathbb{P}(X \in A)$ and $\mathbb{P}(Y \in A^\epsilon)$ depend on X and Y only through their distributions. So by the arbitrariness of the coupling (X, Y) of μ and ν ,

$$\mathbb{P}(X \in A) \leq \mathbb{P}(Y \in A^\epsilon) + \epsilon^{-p} \mathcal{W}_p(\mu, \nu)^p.$$

Choosing $\epsilon = \mathcal{W}_p(\mu, \nu)^{p/(p+1)}$, we deduce that $\mathbb{P}(X \in A) \leq \mathbb{P}(Y \in A^\epsilon) + \epsilon$. Hence,

$$\pi(\mu, \nu) \leq \mathcal{W}_p(\mu, \nu)^{p/(p+1)}. \quad \square$$

3. Non-uniformly expanding/hyperbolic maps

3.1. *Non-uniformly expanding map.* Let (M, d) be a bounded metric space with Borel probability measure ρ . Let $T : M \rightarrow M$ be a non-singular (that is, $\rho(T^{-1}E) = 0$ if and only if $\rho(E) = 0$ for all Borel measurable sets E) ergodic transformation. Suppose that Y is a subset of M with positive measure and that $\{Y_j\}$ is an at most countable measurable partition of Y with $\rho(Y_j) > 0$. Let $R : Y \rightarrow \mathbb{Z}^+$ be an integrable function that is constant on each Y_j and let $T^{R(y)}(y) \in Y$ for all $y \in Y$. We call R the *return time* and $F = T^R : Y \rightarrow Y$ is the corresponding *induced map*. We do not require that R is the first return time to Y .

Let $\nu = (d\rho|_Y)/(d\rho|_Y \circ F)$ be the inverse Jacobian of F with respect to ρ . We assume that there are constants $\lambda > 1, K, C > 0$ and $\eta \in (0, 1]$ such that, for any x, y in a same partition element Y_j :

- (1) $F|_{Y_j} = T^{R(Y_j)} : Y_j \rightarrow Y$ is a (measure-theoretic) bijection for each j ;
- (2) $d(Fx, Fy) \geq \lambda d(x, y)$;
- (3) $d(T^l x, T^l y) \leq Cd(Fx, Fy)$ for all $0 \leq l < R(Y_j)$; and
- (4) $|\log \nu(x) - \log \nu(y)| \leq Kd(Fx, Fy)^\eta$.

Then, such a dynamical system $T : M \rightarrow M$ is a *non-uniformly expanding map*. If $R \in L^p(Y)$ for some $p \geq 1$, then we call $T : M \rightarrow M$ a non-uniformly expanding map of *order* p . It is standard that there is a unique absolutely continuous F -invariant probability measure μ_Y on Y with respect to the measure ρ .

We define the *Young tower* as in [33, 34]. Let $\Delta := \{(x, l) : x \in Y, l = 0, 1, \dots, R(x) - 1\}$, and define an *extension map* $f : \Delta \rightarrow \Delta$ by

$$f(x, l) := \begin{cases} (x, l + 1) & \text{if } l + 1 < R(x), \\ (Fx, 0) & \text{if } l + 1 = R(x). \end{cases}$$

We have a *projection map* $\pi_\Delta : \Delta \rightarrow M$ given by $\pi_\Delta(x, l) := T^l x$ and it is a semiconjugacy satisfying $T \circ \pi_\Delta = \pi_\Delta \circ f$. Then we obtain an ergodic f -invariant probability measure μ_Δ on Δ given by $\mu_\Delta := \mu_Y \times m / \int_Y R d\mu_Y$, where m denotes the counting measure on \mathbb{N} . Hence, there exists an *extension space* $(\Delta, \mathcal{M}, \mu_\Delta)$, where \mathcal{M} is the underlying σ -algebra on (Δ, μ_Δ) . Further, the push-forward measure $\mu = (\pi_\Delta)_* \mu_\Delta$ is an absolutely continuous T -invariant probability measure.

Given a Hölder observable $v : M \rightarrow \mathbb{R}$ with exponent $\eta \in (0, 1]$, define

$$|v|_\infty := \sup_{x \in M} |v(x)|, \quad |v|_\eta := \sup_{x \neq y} \frac{|v(x) - v(y)|}{d(x, y)^\eta}.$$

Let $C^\eta(M)$ denote the Banach space of Hölder observables with norm $\|v\|_\eta = |v|_\infty + |v|_\eta < \infty$. Consider the continuous processes W_n defined by

$$W_n(t) := \frac{1}{\sqrt{n}} \left[\sum_{j=0}^{[nt]-1} v \circ T^j + (nt - [nt])v \circ T^{[nt]} \right] \quad \text{for all } t \in [0, 1], \tag{3.1}$$

where $v \in C^\eta(M)$ with $\int_M v \, d\mu = 0$. Let $v_n := \sum_{i=0}^{n-1} v \circ T^i$ denote the Birkhoff sum.

The following lemma is a summary of known results; see [16, 21, 24, 25] for details.

LEMMA 3.1. *Suppose that $T : M \rightarrow M$ is a non-uniformly expanding map of order $p \geq 2$. Let $v : M \rightarrow \mathbb{R}$ be a Hölder observable with $\int_M v \, d\mu = 0$. Then the following statements hold.*

- (a) *The limit $\sigma^2 = \lim_{n \rightarrow \infty} \int_M (n^{-1/2} v_n)^2 \, d\mu$ exists.*
- (b) *$n^{-1/2} v_n \rightarrow_w G$ as $n \rightarrow \infty$, where G is normal with mean zero and variance σ^2 .*
- (c) *$W_n \rightarrow_w W$ in $C[0, 1]$ as $n \rightarrow \infty$, where W is a Brownian motion with mean zero and variance σ^2 .*
- (d) *If $\mu_Y(R > n) = O(n^{-(\beta+1)})$, $\beta > 1$, then*

$$\lim_{n \rightarrow \infty} \int_M |n^{-1/2} v_n|^q \, d\mu = \mathbf{E}|G|^q \quad \text{for all } q \in [0, 2\beta).$$

- (e) *$\|\max_{k \leq n} |\sum_{i=0}^{k-1} v \circ T^i|\|_{L^{2(p-1)}} \leq C \|v\|_\eta n^{1/2}$ for all $n \geq 1$.*

Proof. Items (a)–(c) are well known; see, e.g., [16, 21, 24]. Item (d) can be found in [25, Theorem 3.5]. For item (e), see [21, Corollary 2.10] for details. □

Remark 3.2. In the case of (d), Melbourne and Török [25] gave examples to illustrate that the q th moments diverge for $q > 2\beta$. Hence, the result on the order of convergent moments is essentially optimal.

THEOREM 3.3. *Let $T : M \rightarrow M$ be a non-uniformly expanding map of order $p > 2$. Suppose that $v : M \rightarrow \mathbb{R}$ is a Hölder observable with $\int_M v \, d\mu = 0$. Then $\mathcal{W}_q(W_n, W) \rightarrow 0$ in $C[0, 1]$ for all $1 \leq q < 2(p - 1)$.*

Proof. It follows from Lemma 3.1(e) that W_n has a finite moment of order $2(p - 1)$. This, together with the fact that $W_n \rightarrow_w W$ as $n \rightarrow \infty$ in Lemma 3.1(c), implies that, for each $q < 2(p - 1)$,

$$\lim_{n \rightarrow \infty} \mathbf{E} \sup_{t \in [0,1]} |W_n(t)|^q = \mathbf{E} \sup_{t \in [0,1]} |W(t)|^q$$

by [9, Theorem 4.5.2]. On the other hand, by the fact that $W_n : M \rightarrow C[0, 1]$ and the definition of push-forward measures,

$$\int_{C[0,1]} d_C(x, 0)^q \, d\mu \circ W_n^{-1}(x) = \int_M \sup_{t \in [0,1]} |W_n(t, \omega)|^q \, d\mu(\omega) = \mathbf{E} \sup_{t \in [0,1]} |W_n(t)|^q.$$

Similarly,

$$\int_{C[0,1]} d_C(x, 0)^q \, d\mu \circ W^{-1}(x) = \mathbf{E} \sup_{t \in [0,1]} |W(t)|^q.$$

Hence,

$$\lim_{n \rightarrow \infty} \int_{C[0,1]} d_C(x, 0)^q \, d\mu \circ W_n^{-1}(x) = \int_{C[0,1]} d_C(x, 0)^q \, d\mu \circ W^{-1}(x).$$

By taking $\mu_n = \mu \circ W_n^{-1}$, $\mu = \mu \circ W^{-1}$ and $x_0 = 0$ in Proposition 2.3 and the fact that $W_n \rightarrow_w W$ in Lemma 3.1(c), the result follows. □

THEOREM 3.4. *Let $T : M \rightarrow M$ be a non-uniformly expanding map of order $p \geq 4$ and suppose that $v : M \rightarrow \mathbb{R}$ is a Hölder observable with $\int_M v \, d\mu = 0$. Then there exists a constant $C > 0$ such that $\mathcal{W}_{p/2}(W_n, W) \leq Cn^{-1/4+1/(4(p-1))}$ for all $n \geq 1$.*

We postpone the proof of Theorem 3.4 to §4.

Remark 3.5.

- (1) Since $\mathcal{W}_q \leq \mathcal{W}_p$ for $q \leq p$, Theorem 3.4 provides an estimate for $\mathcal{W}_q(W_n, W)$ for all $1 \leq q \leq p/2$, $p \geq 4$.
- (2) Our result implies a convergence rate $O(n^{-1/4+\delta'})$ with respect to the Lévy–Prokhorov distance, where δ' depends only on p and δ' can be arbitrarily small as $p \rightarrow \infty$. Indeed, for two given probability measures μ and ν , we have $\pi(\mu, \nu) \leq \mathcal{W}_p(\mu, \nu)^{p/(p+1)}$; see Proposition 2.6.
- (3) The convergence rate in Theorem 3.4 may not be optimal. However, it is well known that one cannot get a better result than $O(n^{-1/4})$ by means of the Skorokhod embedding theorem; see [6, 30] for details.

Example 3.6. (Pomeau–Manneville intermittent maps) A typical example of non-uniformly expanding systems with polynomial tails is the Pomeau–Manneville intermittent map [23, 27]. Consider the map $T : [0, 1] \rightarrow [0, 1]$ given by

$$T(x) = \begin{cases} x(1 + 2^\gamma x^\gamma) & \text{if } x \in [0, \frac{1}{2}), \\ 2x - 1 & \text{if } x \in [\frac{1}{2}, 1], \end{cases} \tag{3.2}$$

where $\gamma \geq 0$ is a parameter. When $\gamma = 0$, this is $Tx = 2x \bmod 1$, which is a uniformly expanding system. It is well known that, for each $0 \leq \gamma < 1$, there is a unique absolutely continuous invariant probability measure μ . By [34], for $0 < \gamma < 1$, the map can be modelled by a Young tower with tails $O(n^{-1/\gamma})$. Further for $\gamma \in [0, \frac{1}{2})$, the CLT and WIP hold for Hölder continuous observables. We restrict the parameter $\gamma \in (0, \frac{1}{2})$; then the map is a non-uniformly expanding system of order p for any $p < 1/\gamma$. By Theorem 3.4, we obtain $\mathcal{W}_{p/2}(W_n, W) \leq Cn^{-1/4+\gamma/(4(1-\gamma))+\delta}$ for all $\gamma \in (0, \frac{1}{4})$.

Example 3.7. (Viana maps) Consider the Viana maps [31] $T_\alpha : S^1 \times \mathbb{R} \rightarrow S^1 \times \mathbb{R}$

$$T_\alpha(\omega, x) = (l\omega \bmod 1, a_0 + \alpha \sin 2\pi\omega - x^2).$$

Here, $a_0 \in (1, 2)$ is chosen in such a way that $x = 0$ is a preperiodic point for the map $g(x) = a_0 - x^2$, α is fixed to be sufficiently small and $l \in \mathbb{N}$ with $l \geq 16$. The results in [17] show that any T close to the map T_α in the C^3 topology can be modelled by a Young tower with stretched exponential tails, which is a non-uniformly expanding map of order p for all $p \geq 1$. Hence by Theorem 3.4, for all $p \geq 4$, $\mathcal{W}_{p/2}(W_n, W) \leq Cn^{-1/4+1/(4(p-1))}$.

3.2. Non-uniformly hyperbolic diffeomorphism. In this subsection, we introduce the main results for non-uniformly hyperbolic systems in the sense of Young [33, 34]. In this case, we follow the argument in [21, 24].

Let $T : M \rightarrow M$ be a diffeomorphism (possibly with singularities[†]) defined on a Riemannian manifold (M, d) . As in [24], consider a subset $Y \subset M$ which has a hyperbolic product structure: that is, there exist a continuous family of unstable disks $\{W^u\}$ and a continuous family of stable disks $\{W^s\}$ such that:

- (1) $\dim W^s + \dim W^u = \dim M$;
- (2) each W^u -disk is transversal to each W^s -disk in a single point; and
- (3) $Y = (\cup W^u) \cap (\cup W^s)$.

For $x \in Y$, $W^s(x)$ denotes the element in $\{W^s\}$ containing x .

Furthermore, there is a measurable partition $\{Y_j\}$ of Y such that each Y_j is a union of elements in $\{W^s\}$ and a W^u such that each element of $\{W^s\}$ intersects W^u in one point. Defining an integrable return time $R : Y \rightarrow \mathbb{Z}^+$ that is constant on each partition Y_j , we can get the corresponding induced map $F = T^R : Y \rightarrow Y$. The *separation time* $s(x, y)$ is the greatest integer $n \geq 0$ such that $F^n x, F^n y$ lie in the same partition element of Y .

We assume that there exist $C > 0$ and $\gamma \in (0, 1)$ such that:

- (1) $F(W^s(x)) \subset W^s(Fx)$ for all $x \in Y$;
- (2) $d(T^n(x), T^n(y)) \leq C\gamma^n$ for all $x \in Y, y \in W^s(x)$ and $n \geq 0$; and
- (3) $d(T^n(x), T^n(y)) \leq C\gamma^{s(x,y)}$ for $x, y \in W^u$ and $0 \leq n < R$.

As for the non-uniformly expanding map, we can define a Young tower. Let $\Delta := \{(x, l) : x \in Y, l = 0, 1, \dots, R(x) - 1\}$ and define an extension map $f : \Delta \rightarrow \Delta$,

$$f(x, l) := \begin{cases} (x, l + 1) & \text{if } l + 1 < R(x), \\ (Fx, 0) & \text{if } l + 1 = R(x). \end{cases}$$

We have a projection map $\pi_\Delta : \Delta \rightarrow M$ given by $\pi_\Delta(x, l) := T^l x$ and it is a semiconjugacy satisfying $T \circ \pi_\Delta = \pi_\Delta \circ f$.

Let $\bar{Y} = Y / \sim$, where $y \sim y'$ if $y' \in W^s(y)$; denote by $\bar{\pi} : Y \rightarrow \bar{Y}$ the natural projection. We can also obtain a partition $\{\bar{Y}_j\}$ of \bar{Y} , a well-defined return time $\bar{R} : \bar{Y} \rightarrow \mathbb{Z}^+$

[†]The meaning of singularity here is in the sense of Young [33], which is different from that of non-uniformly expanding maps at the beginning of §3.1.

and a corresponding induced map $\bar{F} : \bar{Y} \rightarrow \bar{Y}$, as in the case of Y . In addition, we assume that:

- (1) $\bar{F}|_{\bar{Y}_j} = \bar{T}^{\bar{R}(\bar{Y}_j)} : \bar{Y}_j \rightarrow \bar{Y}$ is a bijection for each j ; and
- (2) $v_0 = d\bar{\rho}/(d\bar{\rho} \circ \bar{F})$ satisfies $|\log v_0(y) - \log v_0(y')| \leq K\gamma^s(y,y')$, for all $y, y' \in \bar{Y}_j$, where $\bar{\rho} = \bar{\pi}_*\rho$ with ρ being the Riemannian measure.

Let $\bar{f} : \bar{\Delta} \rightarrow \bar{\Delta}$ denote the corresponding extension map. The projection $\bar{\pi} : Y \rightarrow \bar{Y}$ extends to the projection $\bar{\pi} : \Delta \rightarrow \bar{\Delta}$; here we use the same notation $\bar{\pi}$, which should not cause confusion. There exist an \bar{f} -invariant probability measure $\bar{\mu}$ on $\bar{\Delta}$ and an f -invariant probability measure μ_Δ on Δ , such that $\bar{\pi} : \Delta \rightarrow \bar{\Delta}$ and $\pi_\Delta : \Delta \rightarrow M$ are measure preserving.

THEOREM 3.8. *Let $T : M \rightarrow M$ be a non-uniformly hyperbolic transformation of order $p > 2$. Suppose that $v : M \rightarrow \mathbb{R}$ is a Hölder observable with $\int_M v \, d\mu = 0$. Then $\mathcal{W}_q(W_n, W) \rightarrow 0$ in $C[0, 1]$ for all $1 \leq q < 2(p - 1)$.*

Proof. By [21, Corollary 5.5], W_n has a finite moment of order $2(p - 1)$. The remaining proof is similar to that of Theorem 3.3. □

THEOREM 3.9. *Let $T : M \rightarrow M$ be a non-uniformly hyperbolic transformation of order $p \geq 4$ and suppose that $v : M \rightarrow \mathbb{R}$ is a Hölder observable with $\int_M v \, d\mu = 0$. Then there exists a constant $C > 0$ such that $\mathcal{W}_{p/2}(W_n, W) \leq Cn^{-1/4+1/(4(p-1))}$ for all $n \geq 1$.*

We postpone the proof of Theorem 3.9 to the next section.

Example 3.10. (Non-uniformly expanding/hyperbolic systems with exponential tails) In this case, the return time $R \in L^p$ for all p . Hence, for all $p \geq 4$, $\mathcal{W}_{p/2}(W_n, W) \leq Cn^{-1/4+1/(4(p-1))}$. Specific examples are:

- some partially hyperbolic systems with a mostly contracting direction [7, 12];
- unimodal maps and multimodal maps as in [20] for a fixed system; and
- Hénon-type attractors [19]. Let $T_{a,b} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be defined by $T_{a,b}(x, y) = (1 - ax^2 + y, bx)$ for $a < 2$, $b > 0$, where b is small enough depending on a . It follows from [3, 4] that T admits an Sinai–Ruelle–Bowen (SRB) measure and T can be modelled by a Young tower with exponential tails.

4. Proof of Theorems 3.4 and 3.9

4.1. *Martingale approximation.* The martingale approximation method [14] is one of the main methods for studying statistical limit properties. In [21], Korepanov et al obtained a new version of martingale-coboundary decomposition, which is applicable to non-uniformly hyperbolic systems. In this subsection, we recall some required properties in [21].

PROPOSITION 4.1. *Let $T : M \rightarrow M$ be a non-uniformly expanding map of order $p \geq 1$ and suppose that $v : M \rightarrow \mathbb{R}$ is a Hölder observable with $\int_M v \, d\mu = 0$. Then there is an extension $f : \Delta \rightarrow \Delta$ of T such that, for any $v \in C^\eta(M)$, there exist $m \in L^p(\Delta)$ and $\chi \in L^{p-1}(\Delta)$ with*

$$v \circ \pi_\Delta = m + \chi \circ f - \chi, \quad \mathbf{E}(m|f^{-1}\mathcal{M}) = 0.$$

Moreover, there is a constant $C > 0$ such that, for all $v \in C^\eta(M)$,

$$\|m\|_{L^p} \leq C\|v\|_\eta, \quad \|\chi\|_{L^{p-1}} \leq C\|v\|_\eta$$

and, for $n \geq 1$,

$$\left\| \max_{0 \leq j \leq n} |\chi \circ f^j - \chi| \right\|_{L^p} \leq C\|v\|_\eta n^{1/p}.$$

Proof. The proposition is a summary of Propositions 2.4, 2.5 and 2.7 in [21]. □

PROPOSITION 4.2. Fix $n \geq 1$. Then $\{m \circ f^{n-i}, f^{-(n-i)}\mathcal{M}; 1 \leq i \leq n\}$ is a martingale difference sequence.

Proof. See, for example [21, Proposition 2.9]. □

PROPOSITION 4.3. If $p \geq 2$, then $\|\max_{k \leq n} |\sum_{i=1}^k m \circ f^{n-i}|\|_{L^p} \leq C\|m\|_{L^p} n^{1/2}$ for all $n \geq 1$.

Proof. See the proof in [21, Corollary 2.10]. □

4.2. *Proof of Theorem 3.4.* Define

$$\zeta_{n,j} := \frac{1}{\sqrt{n\sigma}} m \circ f^{n-j}, \quad \mathcal{F}_{n,j} := f^{-(n-j)}\mathcal{M} \quad \text{for } 1 \leq j \leq n.$$

For $1 \leq l \leq n$, define the conditional variance

$$V_{n,l} := \sum_{j=1}^l \mathbf{E}(\zeta_{n,j}^2 | \mathcal{F}_{n,j-1}).$$

We set $V_{n,0} = 0$.

Define the stochastic process X_n with sample paths in $C[0, 1]$ by

$$X_n(t) := \sum_{j=1}^k \zeta_{n,j} + \frac{tV_{n,n} - V_{n,k}}{V_{n,k+1} - V_{n,k}} \zeta_{n,k+1} \quad \text{if } V_{n,k} \leq tV_{n,n} < V_{n,k+1}. \quad (4.1)$$

Step 1. Estimate of the Wasserstein distance between X_n and B . Let B be a standard Brownian motion, that is, $B =_d 1/\sigma W$.

LEMMA 4.4. Let $p \geq 4$. Then, for any $\delta > 0$, there exists a constant $C > 0$ such that $\mathcal{W}_{p/2}(X_n, B) \leq Cn^{-(1/4-\delta)}$ for all $n \geq 1$.

Proof. (1) Fix $n > 0$. It suffices to deal with a single row of the array $\{\zeta_{n,j}, \mathcal{F}_{n,j}, 1 \leq j \leq n\}$. By the Skorokhod embedding theorem (see Theorem A.1), there exists a probability space (depending on n) supporting a standard Brownian motion, still denoted by B , which should not cause confusion, and a sequence of non-negative random variables τ_1, \dots, τ_n such that, for $T_i = \sum_{j=1}^i \tau_j$, we have $\sum_{j=1}^i \zeta_{n,j} = B(T_i)$ with $1 \leq i \leq n$.

In particular, we set $T_0 = 0$. Then, on this probability space and for this Brownian motion, we aim to show that, for any $\delta > 0$, there exists a constant $C > 0$ such that

$$\left\| \sup_{t \in [0,1]} |X_n(t) - B(t)| \right\|_{L^{p/2}} \leq Cn^{-(1/4-\delta)} \quad \text{for all } n \geq 1.$$

Thus, the result follows from Definition 2.1.

For ease of exposition when there is no ambiguity, we will write ζ_j and V_k instead of $\zeta_{n,j}$ and $V_{n,k}$, respectively. Then, by (4.1),

$$X_n(t) = B(T_k) + \left(\frac{tV_n - V_k}{V_{k+1} - V_k} \right) (B(T_{k+1}) - B(T_k)) \quad \text{if } V_k \leq tV_n < V_{k+1}. \quad (4.2)$$

(2) Note that Theorem A.1(3) implies that

$$T_k - V_k = \sum_{i=1}^k (\tau_i - \mathbf{E}(\tau_i | \mathcal{B}_{i-1})) \quad \text{if } 1 \leq k \leq n,$$

where \mathcal{B}_i is the σ -field generated by all events up to T_i for $1 \leq i \leq n$. Therefore, $\{T_k - V_k, \mathcal{B}_k, 1 \leq k \leq n\}$ is a martingale. By the Burkholder inequality and the conditional Jensen inequality, for all $p \geq 4$,

$$\begin{aligned} \left\| \max_{1 \leq k \leq n} |T_k - V_k| \right\|_{L^{p/2}} &\leq Cn^{1/2} \max_{1 \leq k \leq n} \|\tau_k - \mathbf{E}(\tau_k | \mathcal{B}_{k-1})\|_{L^{p/2}} \\ &\leq Cn^{1/2} \max_{1 \leq k \leq n} \|\tau_k\|_{L^{p/2}}. \end{aligned}$$

It follows from Theorem A.1(4) that $\mathbf{E}(\tau_k^{p/2}) \leq 2\Gamma(p/2 + 1)\mathbf{E}(\zeta_k^p)$ for each k . So

$$\left\| \max_{1 \leq k \leq n} |T_k - V_k| \right\|_{L^{p/2}} \leq Cn^{1/2} \max_{1 \leq k \leq n} \|\zeta_k\|_{L^p}^2 = Cn^{-1/2} \|m\|_{L^p}^2. \quad (4.3)$$

On the other hand, it follows from [2, Proposition 4.1] that

$$\|V_n - 1\|_{L^{p/2}} \leq Cn^{-1/2} \|v\|_{\eta}^2. \quad (4.4)$$

(3) Based on the above estimates, by Chebyshev’s inequality,

$$\begin{aligned} \mu(|T_n - 1| > 1) &\leq \mathbf{E}|T_n - 1|^{p/2} \leq 2^{p/2-1} \{\mathbf{E}|T_n - V_n|^{p/2} + \mathbf{E}|V_n - 1|^{p/2}\} \\ &\leq Cn^{-p/4} (\|m\|_{L^p}^p + \|v\|_{\eta}^p). \end{aligned} \quad (4.5)$$

According to the Hölder inequality, (4.5) and Proposition 4.3, we deduce that

$$\begin{aligned} I &:= \left\| \mathbf{1}_{\{|T_n - 1| > 1\}} \sup_{t \in [0,1]} |X_n(t) - B(t)| \right\|_{L^{p/2}} \\ &\leq (\mu(|T_n - 1| > 1))^{1/p} \left\| \sup_{t \in [0,1]} |X_n(t) - B(t)| \right\|_{L^p} \\ &\leq (\mu(|T_n - 1| > 1))^{1/p} \left(\left\| \sup_{t \in [0,1]} |X_n(t)| \right\|_{L^p} + \left\| \sup_{t \in [0,1]} |B(t)| \right\|_{L^p} \right) \\ &\leq Cn^{-1/4}. \end{aligned}$$

(4) We now estimate $|X_n - B|$ on the set $\{|T_n - 1| \leq 1\}$: that is,

$$\begin{aligned} & \left\| \mathbf{1}_{\{|T_n-1|\leq 1\}} \sup_{t \in [0,1]} |X_n(t) - B(t)| \right\|_{L^{p/2}} \\ & \leq \left\| \mathbf{1}_{\{|T_n-1|\leq 1\}} \sup_{t \in [0,1]} |X_n(t) - B(T_k)| \right\|_{L^{p/2}} + \left\| \mathbf{1}_{\{|T_n-1|\leq 1\}} \sup_{t \in [0,1]} |B(T_k) - B(t)| \right\|_{L^{p/2}} \\ & =: I_1 + I_2. \end{aligned}$$

For I_1 , it follows from (4.2) that

$$\sup_{t \in [0,1]} |X_n(t) - B(T_k)| \leq \max_{0 \leq k \leq n-1} |B(T_{k+1}) - B(T_k)| = \max_{0 \leq k \leq n-1} |\zeta_{k+1}|.$$

By Proposition A.2,

$$\begin{aligned} I_1 &= \left\| \mathbf{1}_{\{|T_n-1|\leq 1\}} \sup_{t \in [0,1]} |X_n(t) - B(T_k)| \right\|_{L^p} \\ &\leq \left\| \mathbf{1}_{\{|T_n-1|\leq 1\}} \max_{0 \leq k \leq n-1} |\zeta_{k+1}| \right\|_{L^p} \\ &\leq \left\| \max_{0 \leq k \leq n-1} |\zeta_{k+1}| \right\|_{L^p} \\ &\leq Cn^{-(1/2-1/p)}. \end{aligned}$$

(5) We now consider I_2 on the set $\{|T_n - 1| \leq 1\}$. Take $p_1 > p$. Then it is well known that

$$\mathbf{E}|B(t) - B(s)|^{p_1} \leq c|t - s|^{p_1/2} \quad \text{for all } s, t \in [0, 2]. \tag{4.6}$$

So, it follows from Kolmogorov’s continuity theorem that, for each $0 < \gamma < 1/2 - 1/(p_1)$, the process $B(\cdot)$ admits a version, still denoted by B , such that, for almost all ω , the sample path $t \mapsto B(t, \omega)$ is Hölder continuous with exponent γ and

$$\left\| \sup_{\substack{s,t \in [0,2] \\ s \neq t}} \frac{|B(s) - B(t)|}{|s - t|^\gamma} \right\|_{L^{p_1}} < \infty.$$

In particular,

$$\left\| \sup_{\substack{s,t \in [0,2] \\ s \neq t}} \frac{|B(s) - B(t)|}{|s - t|^\gamma} \right\|_{L^p} < \infty. \tag{4.7}$$

As for $|T_k - t|$,

$$\begin{aligned} \sup_{t \in [0,1]} |T_k - t| &\leq \max_{0 \leq k \leq n-1} \sup_{t \in [V_k/V_n, V_{k+1}/V_n]} |T_k - t| \\ &\leq \max_{0 \leq k \leq n-1} \left| T_k - \frac{V_k}{V_n} \right| + \max_{0 \leq k \leq n-1} \sup_{t \in [V_k/V_n, V_{k+1}/V_n]} \left| \frac{V_k}{V_n} - t \right| \\ &\leq \max_{0 \leq k \leq n} \left| T_k - \frac{V_k}{V_n} \right| + \max_{0 \leq k \leq n-1} \left| \frac{V_{k+1}}{V_n} - \frac{V_k}{V_n} \right| \\ &\leq \max_{0 \leq k \leq n} |T_k - V_k| + \max_{0 \leq k \leq n} \left| V_k - \frac{V_k}{V_n} \right| + \max_{0 \leq k \leq n-1} \left| \frac{V_{k+1}}{V_n} - V_{k+1} \right| \end{aligned}$$

$$\begin{aligned}
 & + \max_{0 \leq k \leq n-1} |V_{k+1} - V_k| + \max_{0 \leq k \leq n-1} \left| V_k - \frac{V_k}{V_n} \right| \\
 & \leq \max_{0 \leq k \leq n} |T_k - V_k| + 3 \max_{0 \leq k \leq n} \left| V_k - \frac{V_k}{V_n} \right| + \max_{0 \leq k \leq n-1} |V_{k+1} - V_k|.
 \end{aligned}$$

Note that $T_0 = V_0 = 0$ and $\gamma \leq 1$, so

$$\sup_{t \in [0,1]} |T_k - t|^\gamma \leq \max_{1 \leq k \leq n} |T_k - V_k|^\gamma + 3^\gamma \max_{1 \leq k \leq n} \left| V_k - \frac{V_k}{V_n} \right|^\gamma + \max_{0 \leq k \leq n-1} |V_{k+1} - V_k|^\gamma.$$

Hence,

$$\begin{aligned}
 & \left\| \sup_{t \in [0,1]} |T_k - t|^\gamma \right\|_{L^p} \\
 & \leq \left\| \max_{1 \leq k \leq n} |T_k - V_k| \right\|_{L^{\gamma p}}^\gamma + 3^\gamma \left\| \max_{1 \leq k \leq n} \left| V_k - \frac{V_k}{V_n} \right| \right\|_{L^{\gamma p}}^\gamma + \left\| \max_{0 \leq k \leq n-1} |V_{k+1} - V_k| \right\|_{L^{\gamma p}}^\gamma.
 \end{aligned} \tag{4.8}$$

For the first term, since $\gamma < \frac{1}{2}$, it follows from (4.3) that

$$\left\| \max_{1 \leq k \leq n} |T_k - V_k| \right\|_{L^{\gamma p}}^\gamma \leq Cn^{-\gamma/2}. \tag{4.9}$$

For the second term, since $|V_k - V_k/V_n| = V_k|1 - 1/V_n|$,

$$\max_{1 \leq k \leq n} \left| V_k - \frac{V_k}{V_n} \right| = V_n \left| 1 - \frac{1}{V_n} \right| = |V_n - 1|.$$

Hence, by (4.4),

$$\left\| \max_{1 \leq k \leq n} \left| V_k - \frac{V_k}{V_n} \right| \right\|_{L^{\gamma p}}^\gamma = \|V_n - 1\|_{L^{\gamma p}}^\gamma \leq Cn^{-\gamma/2}. \tag{4.10}$$

As for the last term, note that $|V_k - V_{k-1}| = \mathbf{E}(\zeta_k^2 | \mathcal{F}_{k-1}) = \mathbf{E}((1/n\sigma^2)m^2 | f^{-1}\mathcal{M}) \circ f^{n-k}$ for all $1 \leq k \leq n$. So,

$$\left\| \max_{0 \leq k \leq n-1} |V_{k+1} - V_k| \right\|_{L^{\gamma p}}^\gamma = \left\| \max_{1 \leq k \leq n} \left| \mathbf{E} \left(\frac{m^2}{n\sigma^2} | f^{-1}\mathcal{M} \right) \circ f^{n-k} \right| \right\|_{L^{\gamma p}}^\gamma \leq Cn^{-(\gamma-2\gamma/p)}, \tag{4.11}$$

where the inequality follows from Proposition A.2.

Based on the above estimates (4.9)–(4.11),

$$\left\| \sup_{t \in [0,1]} |T_k - t|^\gamma \right\|_{L^p} \leq C(n^{-\gamma/2} + n^{-(\gamma-2\gamma/p)}) \leq Cn^{-\gamma/2}, \tag{4.12}$$

where the last inequality holds since $\gamma < \frac{1}{2}$, $1 - 2/p \geq \frac{1}{2}$.

On the set $\{|T_n - 1| \leq 1\}$, note that

$$\sup_{t \in [0,1]} |B(T_k) - B(t)| \leq \left[\sup_{\substack{s,t \in [0,2] \\ s \neq t}} \frac{|B(s) - B(t)|}{|s - t|^\gamma} \right] \left[\sup_{t \in [0,1]} |T_k - t|^\gamma \right].$$

Since $0 < \gamma < \frac{1}{2} - 1/(p_1)$, by the Hölder inequality, (4.7) and (4.12),

$$\begin{aligned}
 I_2 &= \left\| \mathbb{1}_{\{|T_n-1|\leq 1\}} \sup_{t \in [0,1]} |B(T_k) - B(t)| \right\|_{L^{p/2}} \\
 &\leq \left\| \left[\sup_{\substack{s,t \in [0,2] \\ s \neq t}} \frac{|B(s) - B(t)|}{|s - t|^\gamma} \right] \left[\sup_{t \in [0,1]} |T_k - t|^\gamma \right] \right\|_{L^{p/2}} \\
 &\leq \left\| \sup_{\substack{s,t \in [0,2] \\ s \neq t}} \frac{|B(s) - B(t)|}{|s - t|^\gamma} \right\|_{L^p} \left\| \sup_{t \in [0,1]} |T_k - t|^\gamma \right\|_{L^p} \\
 &\leq Cn^{-\gamma/2}.
 \end{aligned}$$

Note that p_1 can be taken arbitrarily large in (4.6), which implies that γ can be chosen sufficiently close to $\frac{1}{2}$. So, for any $\delta > 0$, we can choose p_1 large enough such that $I_2 \leq Cn^{-1/4+\delta}$. The result now follows from the above estimates for I, I_1 and I_2 . \square

Step 2. Estimate of the convergence rate between W_n and X_n . The proof is almost identical to that in [2, §4.1], so we only sketch it here.

PROPOSITION 4.5. [2, Proposition 4.6] *For $n \geq 1$, define*

$$Z_n := \max_{0 \leq i, l \leq \sqrt{n}} \left| \sum_{j=i\sqrt{n}}^{i\sqrt{n}+l-1} v \circ T^j \right|.$$

Then:

- (a) $|\sum_{j=a}^{b-1} v \circ T^j| \leq Z_n((b - a)(n^{1/2} - 1)^{-1} + 3)$ for all $0 \leq a < b \leq n$; and
- (b) $\|Z_n\|_{L^{2(p-1)}} \leq C\|v\|_n^{1/4+1/(4(p-1))}$ for all $n \geq 1$.

Define a continuous transformation $g : C[0, 1] \rightarrow C[0, 1]$ by $g(u)(t) := u(1) - u(1 - t)$.

LEMMA 4.6. *Let $p > 2$. Then there exists a constant $C > 0$ such that $\mathcal{W}_{p-1}(g \circ W_n \circ \pi_\Delta, \sigma X_n) \leq Cn^{-1/4+1/(4(p-1))}$ for all $n \geq 1$, recalling that $\pi_\Delta : \Delta \rightarrow M$ is the projection map.*

Proof. Since $\mathcal{W}_{p-1}(g \circ W_n \circ \pi_\Delta, \sigma X_n) \leq \|\sup_{t \in [0,1]} |g \circ W_n(t) \circ \pi_\Delta - \sigma X_n(t)|\|_{L^{p-1}}$, following the proof of [2, Lemma 4.7], we can obtain the conclusion. \square

Proof of Theorem 3.4. Note that $g \circ g = \text{Id}$ and g is Lipschitz with $\text{Lip}g \leq 2$. It follows from Proposition 2.4 that

$$\mathcal{W}_{p/2}(W_n, W) = \mathcal{W}_{p/2}(g(g \circ W_n), g(g \circ W)) \leq 2\mathcal{W}_{p/2}(g \circ W_n, g \circ W).$$

Since π_Δ is a semiconjugacy, $W_n \circ \pi_\Delta =_d W_n$. Also, $g(W) =_d W =_d \sigma B$. By Lemmas 4.4 and 4.6, for $p \geq 4$,

$$\begin{aligned} \mathcal{W}_{p/2}(g \circ W_n, g \circ W) &= \mathcal{W}_{p/2}(g \circ W_n \circ \pi_\Delta, W) \\ &\leq \mathcal{W}_{p/2}(g \circ W_n \circ \pi_\Delta, \sigma X_n) + \mathcal{W}_{p/2}(\sigma X_n, \sigma B) \\ &\leq Cn^{-1/4+1/(4(p-1))} + Cn^{-1/4+\delta} \leq Cn^{-1/4+1/(4(p-1))}, \end{aligned}$$

where the last inequality holds because $\delta > 0$ can be taken arbitrarily small. □

4.3. *Proof of Theorem 3.9.* The proof is based on the following Lemma 4.7 which is presented in detail in [21, §5].

LEMMA 4.7. *Let $p \geq 1, \eta \in (0, 1]$. Suppose that $T : M \rightarrow M$ is a non-uniformly hyperbolic transformation with the return time $R \in L^p$ and that $v : M \rightarrow \mathbb{R}$ is a Hölder observable. Then:*

- (1) $\bar{f} : \bar{\Delta} \rightarrow \bar{\Delta}$ is a non-uniformly expanding map of order p ; and
- (2) there exists $\theta \in (0, 1)$ such that, for all $v \in C^\eta(M)$, there exist $\phi \in C^\theta(\bar{\Delta})$ and $\psi \in L^\infty(\Delta)$ such that $v \circ \pi_\Delta = \phi \circ \bar{\pi} + \psi - \psi \circ f$. Moreover, $|\psi|_\infty \leq C\|v\|_\eta, \|\phi\|_\theta \leq C\|v\|_\eta$.

Proof of Theorem 3.9. As the definition of W_n in (3.1), define $\bar{W}_n(t) := (1/\sqrt{n}) \sum_{j=0}^{nt-1} \phi \circ \bar{f}^j$ for $t = j/n, 1 \leq j \leq n$, and linearly interpolate to obtain the process $\bar{W}_n \in C[0, 1]$, where ϕ is from Lemma 4.7. By Lemma 4.7, we have $|\bar{W}_n(t) \circ \bar{\pi} - W_n \circ \pi_\Delta|_\infty \leq Cn^{-1/2}|\psi|_\infty$ by simple computations. Since $\bar{\pi}, \pi_\Delta$ are semiconjugacies, $\mathcal{W}_{p/2}(\bar{W}_n, W_n) = \mathcal{W}_{p/2}(\bar{W}_n \circ \bar{\pi}, W_n \circ \pi_\Delta) \leq Cn^{-1/2}$. It follows from Lemma 4.7 that \bar{f} is a non-uniformly expanding map of order p and that ϕ is a Hölder continuous observable with $\int_{\bar{\Delta}} \phi d\bar{\mu} = 0$. By Theorem 3.4, for $p \geq 4, \mathcal{W}_{p/2}(\bar{W}_n, W) \leq Cn^{-1/4+1/(4(p-1))}$. Hence, $\mathcal{W}_{p/2}(W_n, W) \leq \mathcal{W}_{p/2}(W_n, \bar{W}_n) + \mathcal{W}_{p/2}(\bar{W}_n, W) \leq Cn^{-1/4+1/(4(p-1))}$. □

5. Application to homogenization problem

We consider fast–slow systems of the discrete form

$$x_\epsilon(n + 1) = x_\epsilon(n) + \epsilon^2 g(x_\epsilon(n), y(n), \epsilon) + \epsilon h(x_\epsilon(n))v(y(n)), \quad x_\epsilon(0) = \xi, \quad (5.1)$$

where $g : \mathbb{R} \times M \times \mathbb{R}^+ \rightarrow \mathbb{R}, h : \mathbb{R} \rightarrow \mathbb{R}$ satisfy some regularity conditions and $v \in C^\eta(M)$ with $\int_M v d\mu = 0$. The fast variables $y(n) \in M$ are generated by iterating a non-uniformly expanding map: that is $y(n + 1) = Ty(n), y(0) = y_0$. Here, $T : M \rightarrow M$ satisfies the setting in §3.1 The initial condition $\xi \in \mathbb{R}$ is fixed and $y_0 \in M$ is chosen randomly, which is the reason for the emergence of randomness from deterministic dynamical systems.

We have the following regularity conditions.

- (1) $g : \mathbb{R} \times M \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is bounded.
- (2) $g(x, y, 0)$ is Lipschitz in x uniformly in y with Lipschitz constant L ; that is, $|g(x_1, y, 0) - g(x_2, y, 0)| \leq L|x_1 - x_2|$ for all $x_1, x_2 \in \mathbb{R}, y \in M$.
- (3) $\sup_{x \in \mathbb{R}} \sup_{y \in M} |g(x, y, \epsilon) - g(x, y, 0)| \leq C\epsilon^{1/3}$.
- (4) $g(x, y, 0)$ is Hölder continuous in y uniformly in x : that is, $\sup_{x \in \mathbb{R}} |g(x, \cdot, 0)|_\eta < \infty$.
- (5) h is exact; that is, $h = 1/\psi'$, where ψ is a monotone differentiable function and ψ' denotes the derived function. Moreover, $h, h', h'', 1/h$ are bounded.

Let $\hat{x}_\epsilon(t) = x_\epsilon(t\epsilon^{-2})$ for $t = 0, \epsilon^2, 2\epsilon^2, \dots$, and linearly interpolate to obtain $\hat{x}_\epsilon \in C[0, 1]$. Then it follows from [15, Theorem 1.3] that $\hat{x}_\epsilon \rightarrow_w X$ in $C[0, 1]$, where X is the solution to the Stratonovich stochastic differential equation (SDE)

$$dX = \left\{ \bar{g}(X) - \frac{1}{2}h(X)h'(X) \int_M v^2 d\mu \right\} dt + h(X) \circ dW, \quad X(0) = \xi. \tag{5.2}$$

Here, W is a Brownian motion with mean zero and variance σ^2 and $\bar{g}(x) = \int_M g(x, y, 0) d\mu(y)$.

Define $W_\epsilon(t) = \epsilon \sum_{j=0}^{t\epsilon^{-2}-1} v(y(j))$ for $t = 0, \epsilon^2, 2\epsilon^2, \dots$, and linearly interpolate to obtain $W_\epsilon \in C[0, 1]$. Comparing W_ϵ with W_n , we can see that ϵ is identified with $n^{-1/2}$. Hence, it follows from Theorem 3.4 that $\mathcal{W}_{p/2}(W_\epsilon, W) = O(\epsilon^{(p-2)/2(p-1)})$.

THEOREM 5.1. *Let $T : M \rightarrow M$ be a non-uniformly expanding map of order $p \geq 4$. Suppose that the regularity conditions hold. Then there exists a constant $C > 0$ such that*

$$\mathcal{W}_{p/2}(\hat{x}_\epsilon, X) \leq \begin{cases} C\epsilon^{(p-2)/2p} & \text{if } 4 \leq p \leq 6, \\ C\epsilon^{1/3}(-\log \epsilon)^{1/4} & \text{if } p > 6. \end{cases}$$

Proof. The proof follows from the argument in [2]. First, suppose that $h(x) \equiv 1$ and let $N = \lceil \epsilon^{-4/3} \rceil$. By [2, Proposition 5.4], we can write

$$\hat{x}_\epsilon(t) = \xi + W_\epsilon(t) + D_\epsilon(t) + E_\epsilon(t) + \int_0^t \bar{g}(\hat{x}_\epsilon(s)) ds,$$

where

$$D_\epsilon(t) = \epsilon^{2/3} \sum_{n=0}^{\lceil t\epsilon^{-2/3} \rceil - 1} J_\epsilon(n), \quad J_\epsilon(n) = \epsilon^{4/3} \sum_{j=nN}^{(n+1)N-1} \tilde{g}(x_\epsilon(nN), y(j)),$$

$$\tilde{g}(x, y) = g(x, y, 0) - \bar{g}(x)$$

and

$$\left\| \sup_{t \in [0, 1]} |E_\epsilon(t)| \right\|_{L^{2(p-1)}} \leq C\epsilon^{1/3}. \tag{5.3}$$

Let $B_\epsilon(R_\epsilon) = \{\sup_{t \in [0, 1]} |\hat{x}_\epsilon(t)| \leq R_\epsilon\}$, where $R_\epsilon = (-32\sigma^2 \log \epsilon)^{1/2}$. By [2, Lemma 5.5],

$$\mu \left(\sup_{t \in [0, 1]} |\hat{x}_\epsilon(t)| \geq R_\epsilon \right) \leq C\epsilon^{(p-2)/2p}.$$

Since

$$\mu \left(\sup_{t \in [0, 1]} |D_\epsilon 1_{B_\epsilon^c(R_\epsilon)}(t)| > 0 \right) \leq \mu \left(\sup_{t \in [0, 1]} |\hat{x}_\epsilon(t)| \geq R_\epsilon \right) \leq C\epsilon^{(p-2)/2p}$$

and D_ϵ is bounded, we get

$$\left\| \sup_{t \in [0, 1]} |D_\epsilon 1_{B_\epsilon^c(R_\epsilon)}(t)| \right\|_{L^\infty} \leq C\epsilon^{(p-2)/2p}.$$

Moreover, it follows from [2, Lemma 5.6] that $\|\sup_{t \in [0,1]} |D_\epsilon 1_{B_\epsilon(R_\epsilon)}(t)|\|_{L^{2(p-1)}} \leq C\epsilon^{1/3}(-\log \epsilon)^{1/4}$. Hence,

$$\left\| \sup_{t \in [0,1]} |D_\epsilon(t)| \right\|_{L^{2(p-1)}} = O(\epsilon^{(p-2)/2p} + \epsilon^{1/3}(-\log \epsilon)^{1/4}). \tag{5.4}$$

Next, define a continuous map $G : C[0, 1] \rightarrow C[0, 1]$ as $G(u) = v$, where v is the unique solution to $v(t) = \xi + u(t) + \int_0^t \bar{g}(v(s)) ds$. Since \bar{g} is Lipschitz, according to the existence and uniqueness of solutions to ordinary differential equations, G is well defined. By Gronwall’s inequality, G is Lipschitz with $\text{Lip}G \leq e^{\text{Lip}\bar{g}}$.

Since $X = G(W)$ and $\hat{x}_\epsilon = G(W_\epsilon + D_\epsilon + E_\epsilon)$,

$$\mathcal{W}_{p/2}(\hat{x}_\epsilon, X) = \mathcal{W}_{p/2}(G(W_\epsilon + D_\epsilon + E_\epsilon), G(W)) \leq e^{\text{Lip}\bar{g}} \mathcal{W}_{p/2}(W_\epsilon + D_\epsilon + E_\epsilon, W).$$

Following $\mathcal{W}_{p/2}(W_\epsilon, W) = O(\epsilon^{(p-2)/(2(p-1))})$ and the above estimates (5.3)–(5.4), we obtain that $\mathcal{W}_{p/2}(\hat{x}_\epsilon, X) \leq C(\epsilon^{(p-2)/2p} + \epsilon^{1/3}(-\log \epsilon)^{1/4})$.

When $h \neq 1$, by a change of variables, $z_\epsilon(n) = \psi(x_\epsilon(n))$, $\hat{z}_\epsilon(t) = \psi(\hat{x}_\epsilon(t))$, we can reduce the case of multiplicative noise to the case of additive noise: that is,

$$z_\epsilon(n + 1) - z_\epsilon(n) = \epsilon v(y(n)) + \epsilon^2 G(z_\epsilon(n), y(n), \epsilon), \quad z_\epsilon(0) = \psi(\xi),$$

where $G(z, y, \epsilon) := \psi'(\psi^{-1}z)g(\psi^{-1}z, y, \epsilon) + \frac{1}{2}\psi''(\psi^{-1}z)(\psi'(\psi^{-1}z))^{-2}v^2(y) + O(\epsilon)$; see [15, 21] for the calculations. Moreover, we can verify that $G(z, y, \epsilon)$ satisfies the regularity conditions (1)–(4).

Let

$$\bar{G}(z) := \psi'(\psi^{-1}z)\bar{g}(\psi^{-1}(z)) + \frac{1}{2}\psi''(\psi^{-1}(z))(\psi'(\psi^{-1}(z)))^{-2} \int_M v^2 d\mu.$$

Consider the SDE

$$dZ = dW + \bar{G}(Z)dt, \quad Z(0) = \psi(\xi). \tag{5.5}$$

Then $\hat{z}_\epsilon \rightarrow_w Z$, where Z is the solution to (5.5) and $\mathcal{W}_{p/2}(\hat{z}_\epsilon, Z) = O(\epsilon^{(p-2)/2p} + \epsilon^{1/3}(-\log \epsilon)^{1/4})$. Because the Stratonovich integral satisfies the usual chain rule, we can see that $Z = \psi(X)$ satisfies the SDE (5.5) as in [15]. Hence,

$$\begin{aligned} \mathcal{W}_{p/2}(\hat{x}_\epsilon, X) &= \mathcal{W}_{p/2}(\psi^{-1}(\hat{z}_\epsilon), \psi^{-1}(Z)) \leq \text{Lip}(\psi^{-1})\mathcal{W}_{p/2}(\hat{z}_\epsilon, Z) \\ &= O(\epsilon^{(p-2)/2p} + \epsilon^{1/3}(-\log \epsilon)^{1/4}). \end{aligned}$$

The proof is complete. □

Remark 5.2.

- (1) Our result is also applicable to the case where the fast variables are generated by iterating a non-uniformly hyperbolic transformation.
- (2) In [2], the authors obtained the convergence rate $O(\epsilon^{1/3-\delta})$ with respect to the Lévy–Prokhorov distance, where δ depends only on p and δ can be arbitrarily small as $p \rightarrow \infty$. Compared with [2], our result implies the convergence rate in [2] by Proposition 2.6. However, the price to pay is that the non-uniformly hyperbolic fast systems need to have stronger mixing properties than in [2]. To be more specific, in [2], the fast systems are non-uniformly hyperbolic with return time $R \in L^p (p > 2)$, while, in our case, the return time $R \in L^p (p \geq 4)$.

Acknowledgements. The authors sincerely thank Prof. Ian Melbourne for his valuable suggestions. The authors are deeply grateful to the referee for his/her great patience and very careful reading of the paper and for many valuable suggestions which lead to significant improvements of the paper. This work is supported by NSFC Grants 11871132 and 11925102 and by Dalian High-level Talent Innovation Project (Grant 2020RD09).

A. Appendix

THEOREM A.1. (Skorokhod embedding theorem [18]) *Let $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1\}$ be a zero-mean, square-integrable martingale. Then there exist a probability space supporting a (standard) Brownian motion W and a sequence of non-negative variables τ_1, τ_2, \dots with the following properties. If $T_n = \sum_{i=1}^n \tau_i, S'_n = W(T_n), X'_1 = S'_1, X'_n = S'_n - S'_{n-1}$ for $n \geq 2$, and if \mathcal{B}_n is the σ -field generated by S'_1, \dots, S'_n and $W(t)$ for $0 \leq t \leq T_n$, then:*

- (1) $\{S_n, n \geq 1\} =_d \{S'_n, n \geq 1\}$;
- (2) T_n is a stopping time with respect to \mathcal{B}_n ;
- (3) $E(\tau_n | \mathcal{B}_{n-1}) = E(|X'_n|^2 | \mathcal{B}_{n-1})$ almost surely (a.s.); and
- (4) for any $p > 1$, there exists a constant $C_p < \infty$ depending only on p such that

$$E(\tau_n^p | \mathcal{B}_{n-1}) \leq C_p E(|X'_n|^{2p} | \mathcal{B}_{n-1}) = C_p E(|X'_n|^{2p} | X'_1, \dots, X'_{n-1}) \quad \text{a.s.,}$$

where $C_p = 2(8/\pi^2)^{p-1} \Gamma(p + 1)$, with Γ being the usual Gamma function.

PROPOSITION A.2.[†] *Let $\{X_k\}_{k \geq 1}$ be a sequence of identically distributed random variables defined on a common probability space with $\|X_k\|_{L^p} < \infty$ for each $k \geq 1$. Then $\|\max_{1 \leq k \leq n} |X_k|\|_{L^p} = o(n^{1/p})$ as $n \rightarrow \infty$.*

Proof. For $\epsilon > 0$,

$$|X_k|^p \leq n\epsilon + |X_k|^p 1_{\{|X_k|^p > n\epsilon\}}.$$

So,

$$\max_{1 \leq k \leq n} |X_k|^p \leq n\epsilon + \sum_{k=1}^n |X_k|^p 1_{\{|X_k|^p > n\epsilon\}}.$$

Since $\{X_k\}$ is identically distributed,

$$\mathbf{E} \max_{1 \leq k \leq n} |X_k|^p \leq n\epsilon + n\mathbf{E}[|X_k|^p 1_{\{|X_k|^p > n\epsilon\}}].$$

It follows that

$$\frac{1}{n} \mathbf{E} \max_{1 \leq k \leq n} |X_k|^p \leq \epsilon + \mathbf{E}[|X_k|^p 1_{\{|X_k|^p > n\epsilon\}}] \rightarrow \epsilon$$

as $n \rightarrow \infty$. Hence the result follows because ϵ can be taken arbitrarily small. □

[†]This estimate was suggested to us by Prof. Ian Melbourne.

REFERENCES

- [1] M. Antoniou. Rates of convergence for statistical limit laws in deterministic dynamical systems. *PhD Thesis*, University of Warwick, 2018.
- [2] M. Antoniou and I. Melbourne. Rate of convergence in the weak invariance principle for deterministic systems. *Comm. Math. Phys.* **369** (2019), 1147–1165.
- [3] M. Benedicks and L.-S. Young. Sinai–Bowen–Ruelle measures for certain Hénon maps. *Invent. Math.* **112** (1993), 541–576.
- [4] M. Benedicks and L.-S. Young. Markov extensions and decay of correlations for certain Hénon maps. *Astérisque* **261** (2000), 13–56.
- [5] P. Billingsley. *Convergence of Probability Measures (Wiley Series in Probability and Statistics: Probability and Statistics)*, 2nd edn. Wiley-Interscience, New York, 1999.
- [6] A. A. Borovkov. The rate of convergence in the invariance principle. *Teor. Veroyatnost. i Primenen.* **18** (1973), 217–234.
- [7] A. Castro. Backward inducing and exponential decay of correlations for partially hyperbolic attractors. *Israel J. Math.* **130** (2002), 29–75.
- [8] M.-F. Chen, *From Markov Chains to Non-equilibrium Particle Systems*, 2nd edn. World Scientific Publishing Co., Inc., River Edge, NJ, 2004.
- [9] K. L. Chung, *A Course in Probability Theory*, 3rd edn. Academic Press, Inc., San Diego, CA, 2001.
- [10] J. Dedecker and E. Rio. On mean central limit theorems for stationary sequences. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** (2008), 693–726.
- [11] M. Denker and W. Philipp. Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergod. Th. & Dynam. Sys.* **4** (1984), 541–552.
- [12] D. Dolgopyat. On dynamics of mostly contracting diffeomorphisms. *Comm. Math. Phys.* **213** (2000), 181–201.
- [13] M. D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6** (1951), 12pp.
- [14] M. I. Gordin. The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188** (1969), 739–741.
- [15] G. A. Gottwald and I. Melbourne. Homogenization for deterministic maps and multiplicative noise. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **469** (2013), 20130201, 16pp.
- [16] S. Gouëzel. Central limit theorem and stable laws for intermittent maps. *Probab. Theory Related Fields* **128** (2004), 82–122.
- [17] S. Gouëzel. Decay of correlations for nonuniformly expanding systems. *Bull. Soc. Math. France* **134** (2006), 1–31.
- [18] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application (Probability and Mathematical Statistics)*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York–London, 1980.
- [19] M. Hénon. A two-dimensional mapping with a strange attractor. *Comm. Math. Phys.* **50** (1976), 69–77.
- [20] A. Korepanov, Z. Kosloff and I. Melbourne. Averaging and rates of averaging for uniform families of deterministic fast-slow skew product systems. *Studia Math.* **238** (2017), 59–89.
- [21] A. Korepanov, Z. Kosloff and I. Melbourne. Martingale-coboundary decomposition for families of dynamical systems. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **35** (2018), 859–885.
- [22] K. Kubilyus. The rate of convergence in the invariance principle for martingale differences. *Liet. Mat. Rink.* **34** (1994), 482–494.
- [23] C. Liverani, B. Saussol and S. Vaienti. A probabilistic approach to intermittency. *Ergod. Th. & Dynam. Sys.* **19** (1999), 671–685.
- [24] I. Melbourne and M. Nicol. Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys.* **260** (2005), 131–146.
- [25] I. Melbourne and A. Török. Convergence of moments for Axiom A and non-uniformly hyperbolic flows. *Ergod. Th. & Dynam. Sys.* **32** (2012), 1091–1100.
- [26] F. Merlevède, J. Dedecker and E. Rio. Rates of convergence for minimal distances in the central limit theorem under projective criteria. *Electron. J. Probab.* **14** (2009), 978–1011.
- [27] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys.* **74** (1980), 189–197.
- [28] S. T. Rachev, L. B. Klebanov, S. V. Stoyanov and F. J. Fabozzi. *The Methods of Distances in the Theory of Probability and Statistics*. Springer, New York, 2013.
- [29] E. Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009), 802–817.
- [30] S. Sawyer. Rates of convergence for some functionals in probability. *Ann. Math. Stat.* **43** (1972), 273–284.

- [31] M. Viana. Multidimensional nonhyperbolic attractors. *Publ. Math. Inst. Hautes Etudes Sci.* **85** (1997), 63–96.
- [32] C. Villani. *Optimal Transport: Old and New (Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338)*. Springer-Verlag, Berlin, 2009.
- [33] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)* **147** (1998), 585–650.
- [34] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.* **110** (1999), 153–188.