Universiteit Leiden
The Netherlands

Scaling limits in algebra, geometry, and probability

Arzhakova, E.

Citation
Arzhakova, E. (2022, February 23). *Scaling limits in algebra, geometry, and probability*. Retrieved from https://hdl.handle.net/1887/3276037

Version:	Publisher's Version
License:	[Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden]
Downloaded from:	https://hdl.handle.net/1887/3276037

Note: To cite this publication please use the final published version (if applicable).

Chapter 5

Rates of convergence in Central Limit Theorem for ergodic toral automorphisms[1]

5.1 Introduction and main result

Let $\mathbb{T}^d$ be a d-dimensional torus. Consider the standard projection $\pi\colon \mathbb{R}^d \to \mathbb{T}^d$ given by $\pi(x_1, \dots, x_d) = (x_1 \mod 1, \dots, x_d \mod 1)$ and a matrix $S \in \mathrm{GL}(d, \mathbb{Z})$ such that $\det S = \pm 1$. The *toral automorphism* $T_S\colon \mathbb{T}^d \to \mathbb{T}^d$ associated to the matrix S is given by $\pi \circ S = T_S \circ \pi$. Alternatively, one can simply write $T_S(x) = Sx \mod 1$. The toral automorphism T_S is ergodic if and only if the associated matrix S has no eigenvalues which are roots of unity. The transformation T_S preserves the normalised Lebesgue measure m on $\mathbb{T}^d$.

The eigendirections of a matrix S described above induce a decomposition of $\mathbb{R}^d = E^s_S \oplus E^n_S \oplus E^u_S$ where E^s_S is the eigenspace of S corresponding to the eigenvalues with modulus smaller than 1 (stable directions), E^n_S is the eigenspace of S corresponding to the eigenvalues with modulus 1 (neutral directions), and E^u_S is the eigenspace of S corresponding to the eigenvalues with modulus larger than 1 (unstable directions). An important subclass of ergodic toral automorphisms is formed by the *hyperbolic toral automorphisms* for which $E^n_S = \{0\}$ (see Figure 5.1). In other words, the matrix S associated to a hyperbolic toral automorphism has no eigenvalues of unit absolute value. In summary, we consider the following classes of toral auto-

[1]This chapter is based on: E. Arzhakova, D. Terhesiu, Rates of convergence in Central Limit Theorem for ergodic toral automorphisms, in progress

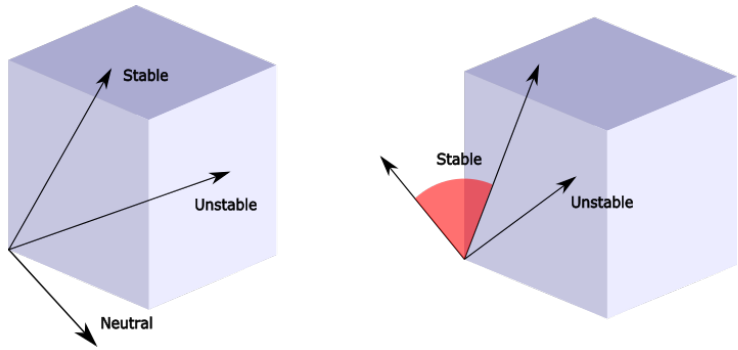

Figure 5.1: Ergodic toral automorphism (left) and a hyperbolic toral automorphism (right). The difference is in the fact that the hyperbolic toral automorphism does not have neutral eigendirections.

morphisms (TA): Toral automorphisms $\supset$ ergodic TA $\supset$ hyperbolic TA.

Example 5.1.1 (Ergodic non-hyperbolic toral automorphism.)**.** The following matrix induces an example of an ergodic non-hyperbolic toral automorphism:

$$S = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

The characteristic polynomial $\chi_S(\lambda) = \lambda^4 - 2\lambda^3 - 2\lambda + 1$ has two real eigenvalues:

$$\lambda_+ = \frac{1}{2}(1 + 3^{1/2} + 12^{1/4}) > 1, \quad \lambda_- = \frac{1}{2}(1 + 3^{1/2} - 12^{1/4}) < 1;$$

and two complex eigenvalues:

$$\lambda_0^{(1)} = \frac{1}{2}(1 - 3^{1/2} + 12^{1/4}i), \quad \lambda_0^{(2)} = \frac{1}{2}(1 - 3^{1/2} - 12^{1/4}i),$$

whose absolute values are equal to 1. Neither of the eigenvalues is a root of unity, therefore, the induced toral automorphism is ergodic. However, it is not hyperbolic due to the presence of two eigenvalues of unit absolute value.

Up to now, several probabilistic aspects of ergodic toral automorphisms have been studied (with respect to the invariant measure m) and we start by recalling some of the landmark results. We remark that hyperbolic toral automorphisms are much easier to study because the dynamical system $(\mathbb{T}^d, T_S)$ has a Markov partition (roughly, every element of the partition gets mapped to a union of partition elements) and as a consequence, the study of $(\mathbb{T}^d, T_S, m)$ can be reduced to that of two-sided finite Markov shifts for which a well developed theory exists (we refer to Section 5.2 for further details).

Due to the presence of neutral directions the study of non-hyperbolic toral automorphisms is much harder; in particular, up to trivial examples, ergodic non-hyperbolic toral automorphisms do not have Markov partitions (see, for instance, [10, 62, 63] and references therein). Resorting to the construction of some clever measurable partition and building on a previous result of Katznelson [54], Lind [63] proved exponential decay of correlation for the general class of θ-Hölder functions $v, w \in C^\theta$ on $\mathbb{T}^d$, that is, $|\int_{\mathbb{T}^d} v\, w \circ T_S^n\, dm - \int_{\mathbb{T}^d} v\, dm \int_{\mathbb{T}^d} w\, dm| \leqslant C\rho^n \|v\|_\theta |w\|_\theta$ for some uniform constant C and some $\rho \in (0,1)$. Hereafter, we denote the class of θ-Hölder functions by C^θ.

Exploiting the partitions introduced in [63], Le Borgne [10] constructed appropriate filtrations to show that under mild assumptions on the Fourier coefficients on functions v on $\mathbb{T}^d$, the Gordin method [40] of martingale differences can be applied to obtain the Central Limit Theorem (CLT) along with its refinements: Weak Invariance Principle (WIP), that is, convergence to Brownian motion, and Strong Invariance Principle (SIP), which is a strong version of the law of the iterated logarithm. For a rough overview of the martingale difference for dynamical systems we refer to Section 5.2. Below we recall the above mentioned terminology along with the result in [10].

Denote the n-th ergodic sum of $v : \mathbb{T}^d \to \mathbb{R}, v \in L^2(m)$ by $S_n v = \sum_{k=0}^{n-1} v \circ T_S^k$. Given a centered function on $\mathbb{T}^d$ (that is, $\int_{\mathbb{T}^d} v\, dm = 0$), (v, T_S) satisfies the CLT with non-zero variance if there exists $\sigma > 0$ such that $\frac{1}{\sqrt{n}} S_n v$ converges in distribution to a Gaussian random variable $Z \sim \mathcal{N}(0, \sigma^2)$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with mean zero and variance σ^2. This means that as $n \to \infty$,

$$\sup_{\alpha \in \mathbb{R}} \left| m\left(\frac{S_n f}{\sqrt{n}} < \alpha\right) - \mathbb{P}\left(Z < \alpha\right) \right| \to 0. \tag{5.1}$$

Recall that given $v : \mathbb{T}^d \to \mathbb{R}$, (v, T_S) satisfies WIP if $W_n(t) = \{\frac{1}{\sqrt{n}} \sum_{k=0}^{[nt]} v \circ T_S^k,\ t \in [0,1]\}$ converges in the space $(D[0,1], \mathbb{R})$ (the space of functions which have left-hand limits and are continuous from the right on $(0,1)$) to a Brownian motion with variance σ^2. Further, (v, T_S) satisfies the SIP if (enlarging $\mathbb{T}^d$ if necessary) there exists a sequence of independent identically distributed (iid) Gaussian random variables Y_k on $(\mathbb{T}^d, T_S, m)$ with mean zero and variance σ^2 such that

$$\sup_{1\leqslant M\leqslant n}\Big|\sum_{k=0}^{M-1} v\circ T_S^k - \sum_{k=0}^{M-1} Y_k\Big| = o(n^{1/2}(\log\log n)^{1/2}) \text{ almost surely as } n\to\infty. \tag{5.2}$$

With these specified, we recall

Theorem 5.1.2. [10] Let v be a centered function on $\mathbb{T}^d$ and $v \in L^2(m)$. Assume that for every $b > 0$ the Fourier coefficients $\widehat{v}(n)$ of v satisfy

$$\sum_{|n|>b} |\widehat{v}(n)|^2 \leqslant R\log^{-\theta}(b)$$

for some $R > 0, \theta > 2$. Assume that v is not a coboundary, i.e., there exists no $h \in L^2(m)$ such that $v = h - h\circ T_S$.

Then (v, T_S) satisfies: i) the CLT with non-zero variance

$$\sigma^2 = \sum_{k\in\mathbb{Z}} \mathbb{E}_m[v\cdot v\circ T_S^{|k|}];$$

ii) WIP and iii) SIP with rate as in (5.2).

The role of the assumption that v is not coboundary is to ensure that $\sigma > 0$.

Since the works of [10, 63] the results have been improved in two directions. Exponential mixing of all orders of Hölder functions was proved via different methods by Dolgopyat [26] and Pène [90]:

Theorem 5.1.3. [26, 90] Let $v_i \in C^\theta(\mathbb{T}^d)$ for some $\theta \in (0,1)$. Then, there exists $\rho \in (0,1)$ such that for any $n_0, \dots n_s \in \mathbb{N}$,

$$\Big|\int_{\mathbb{T}^d}\prod_{i=1}^s v_i\circ T_S^{n_i}\,dm - \prod_{i=1}^s\Big(\int_{\mathbb{T}^d} v_i\,dm\Big)\Big| \leqslant C\rho^{\min_{i\neq j}|n_i-n_j|}\prod_{i=1}^s \|v_i\|_{C^\theta},$$

for some uniform constant C.

Combining Theorem 5.1.3 with Theorem 5.1.2, Gorodnik and Spatzier [44] show that Theorem 5.1.2 holds for the whole class of Hölder functions, with no restriction on the Fourier coefficients. In fact, [44, Theorem 6.2] is phrased for the much larger class of ergodic automorphisms on compact nilmanifolds (not just on $\mathbb{T}^d$), of which particularities do not constitute the subject of this work.

In a different direction, a few works obtain rates of convergence in the CLT and the SIP of Theorem 5.1.2. First, we mention that the work of Le Borgne and Pène [15] gives optimal Berry-Essen error rates in CLT for ergodic toral automorphisms on $\mathbb{T}^3$. With the notation used in (5.1), we recall

Theorem 5.1.4. [15, A consequence of Theorem 2.] Consider an ergodic toral automorphism $(\mathbb{T}^3, T_S, m)$. Suppose that $v : \mathbb{T}^3 \to \mathbb{R}$ satisfies the assumptions of Theorem 5.1.2, in particular the same conditions on the decay of Fourier coefficients of v. Then

$$\sup_{\alpha\in\mathbb{R}} \left| m\left(\frac{S_n v}{\sqrt{n}} < \alpha\right) - \mathbb{P}\left(Z < \alpha\right) \right| = O\left(\frac{1}{\sqrt{n}}\right).$$

We also mention the work of Dedecker, Merlevède and Pène [23] who build on the technical part of the proof of Theorem 5.1.2 to enlarge the class of functions (increasing the range of θ from $\theta > 2$ to $\theta > 1$) and also improve the rate in (5.2) from $o(n^{1/2}(\log\log n)^{1/2})$ to $O(n^{1/4}(\log n))$. This estimate is not implied by and does not imply error rates in the CLT.

It is not clear to us how to generalise Theorem 5.1.4 to general $(\mathbb{T}^d, T_S, m)$, $d > 3$ and also to the entire class of Hölder functions. To our knowledge, error rates in CLT for the general class of ergodic toral automorphisms seem to be absent from the up to date literature. Our main focus is to provide promising results in this direction.

To state our main result, we introduce further terminology. Let

$$\Phi_{\sigma^2}(h) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{\sigma^2}{2}u} h(u)\, du$$

be the expectation of the function $h : \mathbb{R} \to \mathbb{R}$ with respect to the one-dimensional centered distribution $\mathcal{N}(0,\sigma^2)$. Let $\mathcal{W}$ be the class of Lipschitz functions on $\mathbb{R}$. Consider a system $(\mathbb{T}^d, T_S, m)$, a function $v : \mathbb{T}^d \to \mathbb{R}$, and let $S_n v$ be its ergodic sum. Given $Z \sim \mathcal{N}(0,\sigma^2)$ on $(\Omega, \mathscr{F}, \mathbb{P})$, the Wasserstein distance $d_{\mathcal{W}}$ between $\frac{S_n v}{\sqrt{n}}$ and Z is given by

$$d_{\mathcal{W}}\left(\frac{S_n v}{\sqrt{n}}, Z\right) = \sup_{h\in\mathcal{W}} \left| m\left(h\left(\frac{S_n v}{\sqrt{n}}\right)\right) - \Phi_{\sigma^2}(h) \right|. \tag{5.3}$$

Replacing the class $\mathcal{W}$ with the the class of the step functions $\mathbb{K} = \{1_{[-\infty,x]} : x \in \mathbb{R}\}$ gives the Kolmogorov distance, which is the one used in the Berry-Esseen result of Theorem 5.1.4. We recall that results in the Wasserstein distance are weaker since

$$d_{\mathcal{W}}\left(\frac{S_n v}{\sqrt{n}}, Z\right) \leqslant C\left(d_{\mathbb{K}}\left(\frac{S_n v}{\sqrt{n}}), Z\right)\right)^{1/2}, \tag{5.4}$$

for some uniform C.

With these specified, we state the main result of this work

Theorem 5.1.5. Consider $(\mathbb{T}^d, T_S, m)$ and suppose the stable and unstable eigenspaces of S are such that $\dim(E_S^s) = \dim(E_S^u) = 1$. Let $v\colon \mathbb{T}^d \to \mathbb{R}$ be a centered Hölder function such that v is not an L^2-coboundary. Then, (v, T_S) satisfies CLT with non-zero variance $\sigma^2 = \sum_{k\in\mathbb{Z}} \mathbb{E}_m[v \cdot v \circ T_S^{|k|}]$. Moreover,

$$d_{\mathcal{W}}\left(\frac{S_n v}{\sqrt{n}}, Z\right) = O\left(\frac{\log n}{\sqrt{n}}\right).$$

We remark that a similar method of proof works in the situation where the stable and unstable eigendirections of the associated matrix S have different dimensions and also for the multivariate Hölder observable $v : \mathbb{T}^d \to \mathbb{R}^q$, $q \geqslant 1$. For simplicity, in this chapter we omit these generalisations. Given the relation between the Wasserstein and Kolmogorov distances in (5.4), it seems that the current result is far from an optimal Berry-Esseen bound. However, this is not the case for multidimensional observables: in this case the best one can do is to use [91, Theorem 1.1] to improve the result in the Wasserstein distance in Theorem 5.1.5 to $O(n^{-1/2})$, therefore getting rid of $\log n$. In this sense, the present results are very promising.

We emphasise that in the present chapter we provide a new proof of the CLT and that the result of Theorem 5.1.5 is new for $(\mathbb{T}^d, T_S, m)$, $d > 3$ and the entire class of Hölder functions on $\mathbb{T}^d$. Much more importantly, we believe that our proof extends to the cases of random ergodic toral automorphisms, and, eventually, non-linear ergodic toral automorphisms, where previous methods simply break down. This is the subject of work in progress. At this stage we mention that our method of proof relies on the use of the CLT results for dynamical systems via the Stein method obtained by Hella et al. [49] and a careful check of their assumptions using that the eigenfunctions of T_S have components that are Diophantine irrationals.

We conclude this introduction remarking that due to the exponential mixing result for Hölder functions [63], the WIP in Theorem 5.1.2 is a direct consequence of the CLT. This is because the other required condition for WIP, namely, the tightness, can be checked as in the proof of [82, Theorem 1.4]. The SIP in Theorem 5.1.2 is more delicate and we do not address this here.

5.2 A brief survey of the methods of proof of CLT for dynamical systems

In this Section we discuss several methods to prove CLT in the framework of dynamical systems and, in particular, in the framework of hyperbolic toral automorphisms. Where appropriate, we explain why these methods break down for non-hyperbolic ergodic toral automorphisms. The application of some methods (the analogue of the characteristic function for independent random variables and the martingale difference) are illustrated using the simple example of the doubling map.

5.2.1 Gordin's homoclinic points method [41].

Suppose (X, d) is a metric space and $T : X \to X$ is a homeomorphism. Two points $x, y \in X$ are called *homoclinic* if $d(T^n x, T^n y) \to 0$ as $n \to \pm\infty$. The notion goes back to the works of Poincare, and the homoclinic equivalence relation plays an important role in various areas of the theory of dynamical systems.

One of the most convenient settings to study homoclinic structures is that of group automorphisms $T \in \mathrm{Aut}(X)$ of a compact abelian group X. It is sufficient to identify points x that are homoclinic to 0, i.e., $d(T^n x, 0) \to 0$ as $n \to \pm\infty$. Such points form a group of homoclinic points $\Delta(T, X)$.

Gordin also introduced a notion of the homoclinic transformation: an invertible map $R : X \to X$ is called homoclinic to T, if the operators $U_T f(x) = f(Tx)$, $U_R f(x) = f(Rx)$ (called the *Koopman operators* of T and R, respectively), satisfy $U_T^{-n} U_R U_T^n \to \mathrm{Id}$, where Id is the identity operator. Clearly, if $x_0 \in \Delta(X, T)$, then $Rx = x + x_0$ is homoclinic to T. The so-called Gordin group $\mathsf{Gor}(X, T)$ is formed by all invertible non-singular transformations R which are homoclinic to T. For hyperbolic toral automorphisms, the groups $\Delta(\mathbb{T}^d, T_S)$ and $\mathsf{Gor}(\mathbb{T}^d, T_S)$ are isomorphic, i.e., any homoclinic transformation arises from a homoclinic point.

Using Stein's method, Gordin established CLT for functions which are coboundaries with respect to the homoclinic transformations. Here we recall a simplified version of Gordin's CLT for homoclinic points [41]:

Theorem 5.2.1. Suppose T is a group automorphism of a compact abelian group X preserving the Haar measure λ, and $\bar{x} \in \Delta(X,T)$ is a homoclinic point. Suppose a function $f \in L^\infty(X,\lambda)$ satisfies

$$f(x) = F(x+\bar{x}) - F(x)$$

for some $F \in L^2(X,\lambda)$, and moreover,

$$\sum_{n\in\mathbb{Z}} ||f(\cdot + T^n\bar{x}) - f(\cdot)||_{L^\infty} < \infty.$$

Then the sequence

$$Z_n(x) = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(T^k x)$$

converges to the Gaussian distribution $\mathcal{N}(0,\sigma^2)$, and $\sigma^2 \geqslant 0$ is given by the following absolutely converging series

$$\sigma^2 = \sum_{k=-\infty}^{\infty} \langle F(x), f(T^k x + \bar{x}) - f(T^k x)\rangle_{L^2}.$$

In case of hyperbolic toral automorphisms, the above result immediately applies to a large class of sufficiently regular functions f on $\mathbb{T}^d$.

Unfortunately, non-hyperbolic ergodic toral automorphisms have trivial groups of homoclinic points: $\Delta(\mathbb{T}^d, T_S) = \{0\}$ (see [66] and Theorem 4.1 of [67]), and hence, the above result cannot be applied.

In [42] Gordin extended the homoclinic point approach to non-hyperbolic ergodic toral automorphims using the martingale difference method.

An interesting observation is that the homoclinic point method is also applicable to $\mathbb{Z}^d$-actions [43], and in fact, in striking contrast with $\mathbb{Z}$-actions discussed in this chapter, the method works for some non-expansive algebraic dynamical systems as well, e.g., those which arise naturally in connection to spanning trees.

5.2.2 Characteristic functions method for dynamical systems

We first recall the method in the i.i.d. set up. Consider a sequence of i.i.d. random variables $\{Y_j\}_{j\geqslant 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ with mean 0 and positive variance. One easy proof of the CLT goes via the Levy's continuity theorem. Let $\chi(t) = \mathbb{E}_{\mathbb{P}}[e^{itY_j}]$ be the characteristic function of Y_j, set $S_n = \sum_{j=0}^{n-1} Y_j$ and note that $\mathbb{E}_{\mathbb{P}}[e^{it\frac{S_n}{\sqrt{n}}}] = \chi(t/\sqrt{n})^n$. Recall that the characteristic function of $Z \sim \mathcal{N}(0,\sigma^2)$ is $\exp(-\sigma^2t^2/2)$. Also, since Y_j, $j \geqslant 1$, has finite second moment and zero mean, one has that $1-\chi(t) = \sigma^2t^2/2(1+o(1))$ as $t\to 0$. Thus, as $n \to \infty$,

$$\chi(t/\sqrt{n})^n = \Big(1 - \frac{\sigma^2t^2}{2n} + o(\sigma^2t^2/2)\Big)^n \to \exp(-\sigma^2t^2/2), \qquad t \in \mathbb{R}.$$

By the Levy's continuity theorem we conclude that the sequence $\frac{1}{\sqrt{n}}S_n$ converges in distribution to $\mathcal{N}(0,\sigma^2)$.

In the framework of measure preserving dynamical systems (X,T,μ), given $f\colon X \to \mathbb{R}$, we are interested in the convergence in distribution of the normalised ergodic sums $\frac{1}{\sqrt{n}}S_nf = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} f\circ T^k$ to $Z \sim \mathcal{N}(0,\sigma^2)$, for some $\sigma > 0$. For simplicity, we assume that $\int f\,d\mu = 0$. Using the Levy's continuity theorem, one can rephrase this problem in terms of convergence of the corresponding characteristic functions $\chi_{\frac{1}{\sqrt{n}}S_nf}(t)$ to the characteristic function of a limiting random variable. Originating from the work of Nagaev [84], a standard method to prove this convergence is by using spectral properties of transfer operators. Starting with the work of Aaronson and Denker [2], this method became classic for establishing various probabilistic results, including stable laws and local limit theorems. We briefly recall the main elements of the method.

The transfer operator $\mathcal{L} : L^1(\mu) \to L^1(\mu)$ for (X,T,μ), is defined by the equality $\int_X \mathcal{L}u\cdot v\,d\mu = \int_X u\cdot v\circ T\,d\mu$, where $u,v \in L^1(\mu)$ and $v \in L^\infty(\mu)$. The basic idea of the Nagaev method is that since

$$\int_X e^{itS_nf}ud\mu = \int_X \mathcal{L}(t)^n ud\mu, \qquad \text{where } \mathcal{L}(t)u = \mathcal{L}(e^{itf}u), u \in L^1.$$

the study of characteristic functions for dynamical systems (possibly with heavy dependencies) can be reduced to the study of properties of the perturbed transfer operator $\mathcal{L}(t)$. Note that $\mathcal{L}(0) = \mathcal{L}$. The necessary requirement for this method is to find a good function space on which the family of operators $\{\mathcal{L}(t)\}_{t\geqslant 0}$ satisfies good spectral properties: see, for instance, the survey of Gouëzel [45]. In short, this comes down to finding a Banach space $\mathcal{B}$ with norm $\|\cdot\|$ on which

i) $\mathcal{L}$ has a decomposition of the form

$$\mathcal{L}(0)^n u = \int u \, d\mu + Q(0)^n u,$$

where $Q(0)$ is an operator on $\mathcal{B}$ such that $\|Q(0)^n\| \leqslant \theta^n$, for some $\theta \in (0,1)$.

ii) the family $\{\mathcal{L}(t)\}_{t \geqslant 0}$ satisfies 'good' continuity properties. For the CLT a sufficient (but not necessary, see [45] and references therein) condition is that for $t \in B_\delta(0)$,

$$\|\mathcal{L}(t) - \mathcal{L}(0)\| \leqslant Ct^2$$

for some uniform constant C.

Items (i) and (ii) can be easily established for the simple example of the doubling map $T\colon [0,1] \to [0,1]$ given by the formula $Tx = 2x \mod 1$ (see Figure 5.2). However, apart from simple examples of unit interval maps, establishing (i) and (ii) is highly non-trivial. We refer to the list of references in [45] for some non-trivial examples.

For $Tx = 2x \mod 1$, and letting T^{-1} denote the left inverse branch, the existence of the Markov partition $\mathcal{P} = \{(T^{-(j+1)}1, T^{-j}1]\}_{j \geqslant 0}$ together with the expansion of the map is the key. Using the pointwise formula for the transfer operator as in [45], one establishes (i) and (ii) in the Banach space of piecewise C^2 functions; piecewise C^2 means C^2 on the elements of $\mathcal{P}$. Item (i) holds with $\theta = 1/T' = 1/2$ and item (ii) holds for any observable

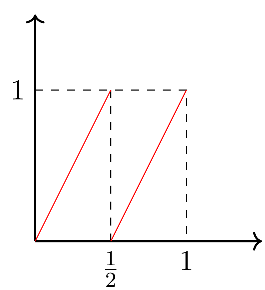

Figure 5.2: The doubling map $Tx = 2x \mod 1$.

with finite second moment.

Nagaev method for hyperbolic toral automorphisms. Nowadays, the hyperbolic toral automorphisms are known to have the spectral gap property (item i) above) in several anisotropic Banach spaces of distributions (i.e., generalised functions): see the survey of Liverani [68]. As invertible transformations, they cannot have spectral gaps in usual Banach spaces embedded in L^∞ [68]. The role of the Banach spaces in [68] is to allow a different treatment of the expanding and contracting directions. In such Banach spaces the existence of the Markov partition for hyperbolic toral automorphisms is not required, though the absence of the neutral direction is crucial.

A more traditional treatment of hyperbolic toral automorphisms exploits the existence of Markov partitions and the isomorphism with the two-sided Markov shift, where classical methods apply. We recall that as in [14] a standard way of treating two-sided Markov shifts is to collapse the stable (contracting) or unstable (expanding) directions. For the one-sided Markov shift there are several Banach spaces known to provide spectral gap: see for instance the work [2] for a brief overview. The lift of limit theorems from one-sided shifts to two-sided shifts is also classic since the work of Bowen [14].

Good Banach spaces for to ergodic toral automorphisms do not exist due to the presence of neutral direction (which in turn, does not not allow one to establish the existence of Markov partitions).

5.2.3 Martingale difference approach

Unlike the homoclinic method and the characteristic functions method, the martingale difference method can be applied in the context of non-hyperbolic ergodic toral automorphisms. The main current stochastic results on the ergodic toral automorphisms, including CLT, are obtained using the martingale difference method. In the following subsection we recall the main ingredients of the method, illustrate them with the simple example of the doubling map, and state the known limit results on the ergodic toral automorphisms.

Consider a probability space $(X, \mathcal{B}, \mu)$. A sequence $Y_n\colon X \to \mathbb{R}$ of random variables is a *martingale difference sequence* if

1. there exists a non-decreasing sequence of σ-algebras (filtration)

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{B}$$

   such that Y_n is measurable with respect to $\mathcal{F}_n$;

2. the conditional expectations satisfy $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = 0$ for $n > 0$.

A sequence $Y_n \colon X \to \mathbb{R}$ of random variables is a *reverse martingale difference sequence* if

1. there exists a non-increasing sequence of σ-algebras

$$\mathcal{B} \supset \mathcal{F}_1 \supset \mathcal{F}_2 \supset \ldots$$

   such that Y_n is measurable with respect to $\mathcal{F}_n$;

2. the conditional expectations satisfy $\mathbb{E}[Y_n|\mathcal{F}_{n+1}] = 0$ for $n > 0$.

Theorem 5.2.2 ([47,69] The (reverse) martingale difference theorem.)**.** Suppose that $\{Y_n\}$ is a martingale difference with respect to $\{\mathcal{F}_i\}$. If the following two conditions hold:

- $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i^2|\mathcal{F}_{i-1}] \xrightarrow{P} \sigma^2 < \infty$;
- $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i^2 \mathbf{1}_{|Y_i|>\epsilon\sqrt{n}}|\mathcal{F}_{i-1}] \xrightarrow{P} 0$ for every $\epsilon > 0$,

then the sequence $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$ converges in distribution to $\mathcal{N}(0,\sigma^2)$ [40, 47]. Suppose that $\{Y_n\}$ is a reverse martingale difference with respect to $\{\mathcal{F}_i\}$. If the following two conditions hold:

- $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i^2|\mathcal{F}_{i+1}] \xrightarrow{P} \sigma^2 < \infty$;
- $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i^2 \mathbf{1}_{|Y_i|>\epsilon\sqrt{n}}|\mathcal{F}_{i+1}] \xrightarrow{P} 0$ for every $\epsilon > 0$,

then the sequence $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$ converges in distribution to $\mathcal{N}(0,\sigma^2)$.

The (reverse) martingale difference method can be applied to prove CLT in the framework of a dynamical system $(X, \mathcal{F}, \mu, T)$ and functions $f \in L^2$.

The idea is to find a non-decreasing (or non-increasing) filtration $\{\mathcal{F}_i\}$ and functions $h, g \in L^2$ such that

$$f = h + g - g \circ T,$$

where $\{h \circ T^i\}$ is a (reverse) martingale difference with respect to $\{\mathcal{F}_i\}$ which satisfies the conditions of Theorem 5.2.2 with $\sigma^2 > 0$. Since $\frac{1}{\sqrt{n}} \sum_1^n f \circ T^i = \frac{1}{\sqrt{n}} \sum_1^n h \circ T^i + \frac{1}{\sqrt{n}}(g - g \circ T^{n+1})$ the CLT for f follows from CLT for h.

Let us illustrate the reverse martingale difference method with a simple example of the doubling map. Consider a system $([0,1], \mathcal{B}, \mu, T)$ where $\mathcal{B}$ is the Borel σ-algebra, μ is the Lebesgue measure, and T is the doubling map given by $Tx = 2x \mod 1$ (see Figure 5.2). Denote by U the Koopman operator of T (i.e., $Uf = f \circ T$) and by $\mathcal{L}$ the transfer operator of T.

Consider f to be Lipschitz, not a coboundary, and with zero mean. Denote the space of Lipschitz functions by $\mathcal{W}$ with norm $\|\cdot\|_{\mathcal{W}}$; since $||\mathcal{L}f||_{\mathcal{W}} \leqslant \frac{1}{2}||f||_{\mathcal{W}}$, the function $g = \sum_{i=1}^{\infty} \mathcal{L}^i f$ is well-defined and Lipschitz. Introduce a function $h = f + g - g \circ T$; we claim that $h \circ T^n$ is a reverse martingale difference with respect to a non-increasing filtration $\mathcal{B} \supset T^{-1}\mathcal{B} \supset T^{-2}\mathcal{B} \ldots$. In order to show this, it suffices to check that $\mathbb{E}[h|T^{-1}\mathcal{B}] = \int_{T^{-1}B} h \, d\mu = 0$ for every $B \in \mathcal{B}$. It is easy to verify that

$$\int_{T^{-1}B} h \, d\mu = \int_B h \cdot (\mathbf{1} \circ T) \, d\mu = \int_B \mathcal{L}h \, d\mu.$$

Note that $\mathcal{L}h = \mathcal{L}f + \mathcal{L} \sum_{i=1}^{\infty} \mathcal{L}^i f - \mathcal{L}U \sum_{i=1}^{\infty} \mathcal{L}^i f = 0$ because $\mathcal{L}U = I$. This ensures that $h \circ T^n$ is a reverse martingale difference. It is stationary and ergodic, therefore, the CLT with positive variance holds, see, for instance, [47].

In [10], the properties of distributions of stable leaves [63] were used to construct a filtration that leads to a proof of the CLT and more refined limit properties using the martingale difference method for ergodic toral automorphisms. We recall Theorem 5.1.2 stated in the introduction.

5.3 Stein's method for establishing CLT with rates of convergence

We remind the reader that the aim of the present work is to study the rates of convergence in CLT for the class of ergodic toral automorphisms, namely, to prove Theorem 5.1.5. So far, we have discussed several methods to prove CLT in the context of dynamical systems. However, neither of the methods mentioned above is suitable to obtain optimal rates of convergence in CLT for ergodic toral automorphisms and a wide class of functions. The characteristic functions method requires the presence of the spectral gap of the transfer operator; the spectral gap is present in some Banach spaces for the hyperbolic toral automorphisms but not for non-hyperbolic ergodic toral automorphisms. The homoclinic method relies on the existence of homoclinic points of the automorphism and, therefore, it only works for the family of hyperbolic toral automorphisms but it is not applicable to the whole family of ergodic toral automorphisms. The martingale difference method allows one to prove CLT for the class of ergodic toral automorphisms. As recalled in Section 5.1, variations of the martingale difference method give error rates in CLT for ergodic toral automorphisms $(\mathbb{T}^3, T_S, m)$ and a large class of observables, see Theorem 5.1.4.

In this Section we discuss the Stein method as in [49] and motivate the application of this particular variation of the Stein method to study the rate of convergence in CLT for ergodic toral automorphisms.

5.3.1 Description of the Stein method with rates of convergence

The Stein method as in [49] studies the Wasserstein distance between $W^N = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} f \circ T^k$ and $Z \sim \mathcal{N}(0, \sigma^2)$ under certain conditions on f and T. Therefore, the method allows one not only to prove CLT but also to establish the associated rates of convergence in the Wasserstein distance which provides a smooth metric on the space of distributions. The main difference between the Stein method and the characteristic functions or martingale difference methods is that the Stein method relies on the decay of correlations and it does not use spectral properties of the transfer operator such as the spectral gap. The fact that ergodic toral automorphisms enjoy exponential mixing of all orders [26, 90] ensures that some assumptions of the Stein method are trivial whereas other assumptions require serious effort.

We briefly explain the main idea of the Stein method as in [49]. Suppose that $Z \sim \mathcal{N}(0, \sigma^2)$; for each $h \in \mathcal{W}$ (the space of Lipschitz functions with Lipschitz constant 1) there exists a measurable solution A_h of the Stein equation:

$$\sigma^2 A_h'(x) - x A_h(x) = h(x) - \mathbb{E}[h(Z)].$$

The form of the equation can intuitively be explained by the fact that a random variable $W \sim \mathcal{N}(0, \sigma^2)$ if and only if $\sigma^2\mathbb{E}[A'(W)] - \mathbb{E}[WA(W)] = 0$ for all absolutely continuous[2] functions A for which these expectations exist. The right hand side of the Stein equation allows us to estimate the distance between the distributions of W and Z. We select the Wasserstein distance (see (5.3) for definition) and obtain

$$d_{\mathcal{W}}(W, Z) = \sup_{h \in \mathcal{W}} \Big|\mathbb{E}[h(W)] - \mathbb{E}[h(Z)]\Big| = \sup_{A_h \colon h \in \mathcal{W}} \Big|\sigma^2\mathbb{E}[A_h'(W)] - \mathbb{E}[WA_h(W)]\Big|.$$

For each $h \in \mathcal{W}$ such that $||h'||_\infty < \infty$ the solution A_h is a bounded function with bounded first and second derivatives. The upper bound on the Wasserstein distance follows from [49, Lemma 3.2]:

$$d_{\mathcal{W}}(W, Z) \leqslant \sup_{A \in \mathcal{A}_{\mathcal{W}}} \Big|\sigma^2\mathbb{E}[A'(W)] - \mathbb{E}[WA(W)]\Big|, \tag{5.5}$$

where $\mathcal{A}_{\mathcal{W}} = \{A \colon ||A||_\infty \leqslant 2, ||A'||_\infty \leqslant \sqrt{2/\pi}\sigma^{-1}, ||A''||_\infty \leqslant 2\sigma^{-2}\}$.

Fix $N > 0$ and suppose that $W^N = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1} X_i$ where $\{X_i\}_{i=0}^{N-1}$ is a set of random variables. Our goal is to estimate $d_{\mathcal{W}}(W^N, Z)$ and by equation (5.5) it suffices to bound the quantity $\Big|\sigma^2\mathbb{E}[A'(W^N)] - \mathbb{E}[W^N A(W^N)]\Big|$ for $A \in \mathcal{A}_{\mathcal{W}}$. To do this, fix $n, K \leqslant N$ and consider $W^n = W^N - \frac{1}{\sqrt{N}}\sum_{i=\max\{0,n-K\}}^{\min\{n-K,N\}} X_i$. Then,

$$\mathbb{E}[W^N A(W^N)] = \frac{1}{\sqrt{N}}\mathbb{E}\Big[\sum_{i=0}^{N-1} X_i\left(A(W^N) - A(W^i)\right)\Big] + \frac{1}{\sqrt{N}}\mathbb{E}\Big[\sum_{i=0}^{N-1} X_i A(W^i)\Big].$$

Denote the first summand of the right-hand side by S_1 and the second summand by S_2. Note that if the random variables X_i are independent then $S_2 = 0$; it seems plausible that if the correlations of X_i decay fast enough and K is big enough then S_2 is close to zero. However, an upper bound on S_2 is an assumption of the method (see Assumption 2 of Theorem 2.4 in [49]) and has to be checked separately for every system.

Since A, A' are absolutely continuous and $||A''||_\infty < 2\sigma^{-2}$ we have that

$$S_1 \approx \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=\max\{0,n-K\}}^{\min\{n+k,N-1\}} \mathbb{E}[X_i A'(W^N) X_j].$$

[2]A function $A \colon \mathbb{R} \to \mathbb{R}$ is called absolutely continuous if it has a derivative A' almost everywhere, the derivative is locally integrable, and $A(y) = A(x) + \int_x^y A'(t)dt$.

For large K one can approximate $\frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=\max\{0,n-K\}}^{\min\{n+K,N-1\}}\mathbb{E}[X_iX_j] \approx \sigma^2$ and we conclude that if the correlations of random variables X_i decay fast enough (see Assumption 1 of Theorem 2.4 in [49]) then $\mathbb{E}[W^N A(W^N)] \approx \sigma^2\mathbb{E}[A'(W^N)]$ and, hence, $d_{\mathcal{W}}(W^N, Z) \approx 0$.

Let us now formulate the precise statement of Theorem 2.4 of [49]. Consider a probability space (X,μ), a measure-preserving transformation $T\colon X \to X$, and a bounded measurable centered function $f\colon X \to \mathbb{R}$. For brevity we write $\int f = \int_X f\,d\mu$. Fix two integers $0 \leqslant K < N$ and write $W^N = \sum_{j=0}^{N-1}\frac{1}{\sqrt{N}} f\circ T^j$, $W^n = W^N - \frac{1}{\sqrt{N}}\sum_{j=n-K}^{n+K} f\circ T^j$.

Theorem 5.3.1. [49] Suppose that the following conditions are satisfied for a bounded measurable centered function $f\colon X \to \mathbb{R}$:

1. There exist constants $C_2, C_4 > 0$ and a non-increasing function $\rho\colon \mathbb{N}_0 \to \mathbb{R}_+$ with $\rho(0) = 1$ and $\sum_{i=1}^{\infty} i\rho(i) < \infty$ such that for any $k \geqslant 0$ and $0 \leqslant l \leqslant m \leqslant n < N$,

   (a) $\left|\int f\cdot(f\circ T^k)\right| \leqslant C_2\rho(k)$;

   (b) $\left|\int f\cdot(f\circ T^l)\cdot(f\circ T^m)\cdot(f\circ T^n)\right| \leqslant C_4\min\{\rho(l),\rho(n-m)\}$;

   (c) $\left|\int\Big(f\cdot(f\circ T^l)\cdot(f\circ T^m)\cdot(f\circ T^n)\Big) - \int\Big(f\cdot(f\circ T^l)\Big)\int\Big((f\circ T^m)\cdot(f\circ T^n)\Big)\right| \leqslant C_4\rho(m-l)$.

2. There exists a function $\tilde\rho_N\colon \mathbb{N}_0 \to \mathbb{R}_+$ such that for any differentiable $A\colon \mathbb{R}\to\mathbb{R}$ with A' absolutely continuous and $\max_{0\leqslant k\leqslant 2}||A^{(k)}||_\infty < 1$, and for any $0 \leqslant n < N$,
$$\left|\int(A(W^n)\cdot f\circ T^n)\right| \leqslant \tilde\rho_N(K);$$

3. f is not L^2 - coboundary.

Then, $0 < \sigma^2 = \int(f^2) + 2\sum_{n=1}^{\infty}\int(f\cdot f\circ T^n) < \infty$ and if $Z \sim \mathcal{N}(0,\sigma^2)$ then
$$d_{\mathcal{W}}(W^N, Z) \leqslant C_{\#}\left(\frac{K+1}{\sqrt{N}} + \sum_{i=K+1}^{\infty}\rho(i)\right) + C'_{\#}\sqrt{N}\tilde\rho_N(K)$$
where $0 < C_{\#}, C'_{\#} < \infty$ do not depend on N, K.

Assumption 1 follows immediately from good enough decay on multiple mixing. We recall this is the case for the doubling map $Tx = 2x \mod 1$. Assumption 2 states conditions on the decay of correlations of $f \circ T^n$ and $A(W^n)$ where A is a bounded function with bounded and absolutely continuous first derivative and bounded second derivative. As we discussed above, Assumption 2 is a natural consequence of the idea of the Stein method to estimate the quantity $d_{\mathcal{W}}(W^N, Z)$ by $\left|\sigma^2\mathbb{E}[A'(W^N)] - \mathbb{E}[W^N A(W^N)]\right|$. This Assumption can be difficult to verify as its proof normally requires several technical steps (for the steps we refer the reader to Section 7.1 of [49]). In the following example we show how to verify Assumption 2 for the doubling map.

Example 5.3.2 (Assumption 2 for the doubling map)**.** Following Section 7.2 of [49] we reason that the Assumption 2 holds in a simple case of the doubling map and a Lipschitz function f with zero mean such that $|f(x) - f(y)| \leqslant L|x-y|$. Fix $0 \leqslant n < N$ and recall that $W^n = W^N - \sum_{j=n-K}^{n+K} f \circ T^j$. It is natural to present W^n as the sum of two parts:

$$W^n = W^n_- + W^n_+ = \sum_{j=0}^{n-K-1} f \circ T^j + \sum_{j=n+K+1}^{N-1} f \circ T^j.$$

To simplify the summand W^n_- we introduce a partition of the unit interval $\{\xi_q = ((q-1)2^{-n}, q2^{-n})\}_{q\in\{1,\ldots 2^n\}}$ such that W^n_- is almost equal to a constant c_q on each atom ξ_q of the partition. This allows us to bound the desired quantity $\left|\int (f \circ T^n) \cdot A(W^n_- + W^n_+)\right|$ by a simpler quantity $\left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right|$:

$$\left|\int (f \circ T^n) \cdot A(W^n)\right| \leqslant \left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right| + \frac{L||A'||_\infty ||f||_\infty}{\sqrt{N}2^K}.$$

To estimate the quantity $\left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right|$ one notices that

$$\left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right| = \left|\sum_q \int_{\xi_q} \left(f A(c_q + \tilde{W}^n_+ \circ T^{K+1})\right) \circ T^n\right|$$

$$= \left|2^{-n} \sum_q \int_{\mathbb{T}^d} f A(c_q + \tilde{W}^n_+) \circ T^{K+1}\right| \leqslant \frac{L||A||_\infty}{2^K}$$

where $\tilde{W}^n_+ = W^n_+ \circ T^{n+K+1}$ and the last equality follows from the fact that $||\mathcal{L}^K f||_\theta \leqslant 2^{-K}||f||_\theta$. Since $\max_{0\leqslant k \leqslant 2} ||A^{(k)}||_\infty < 1$ we have

$$\left|\int (f \circ T^n) \cdot A(W^n)\right| \leqslant \frac{L||A||_\infty}{2^K} + \frac{L||A'||_\infty ||f||_\infty}{\sqrt{N}2^K} \leqslant \frac{L}{2^K} + \frac{L||v||_\infty}{\sqrt{N}2^K} = \tilde{\rho}_N(K).$$

We recall that assumption 3 on f not being coboundary is a typical assumption to ensure $\sigma > 0$. The significance of 5.3.1 for proving rates of convergence as in 5.1.5 is emphasised in the following corollary:

Corollary 5.3.3. [49] Suppose that for any $N > 2$ the assumptions of Theorem 5.3.1 hold for f with $\rho(i) = \lambda^i, \tilde{\rho}_N(i) = C'\lambda^i N^a$, and $K = [\log N^b / \log \lambda]$ with fixed $0 < \lambda < 1, 1 < a + 1 \leqslant b$. Then, f satisfies CLT and there exists a constant C that does not depend on N such that

$$d_{\mathcal{W}}(W^N, Z) \leqslant C \frac{\log N}{\sqrt{N}}.$$

5.4 Proof of CLT with rates of convergence for ergodic toral automorphisms

In this Section we prove Theorem 5.1.5. The method of proof is to apply Theorem 5.3.1 in the setting of Theorem 5.1.5. We start with a discussion of the three Assumptions of Theorem 5.3.1. Throughout, the integration is with respect to the normalised Lebesgue measure m. The proof below is written making one more simplifying assumption (not appearing as such in 5.1.5) namely that v is Lipschitz. We stress that this is just for the ease of the computation, as to not keep track of Hölder exponents. Else, the proof below can be easily adapted to work for Hölder functions.

Assumption 1: There exist constants $C_2, C_4 > 0$ and a non-increasing function $\rho\colon \mathbb{N}_0 \to \mathbb{R}_+$ with $\rho(0) = 1$ and $\sum_{i=1}^{\infty} i\rho(i) < \infty$ such that for any $k \geqslant 0$ and $0 \leqslant l \leqslant m \leqslant n < N$,

1. $\left| \int v \cdot (v \circ T_S^k) \right| \leqslant C_2 \rho(k);$
2. $\left| \int v \cdot (v \circ T_S^l) \cdot (v \circ T_S^m) \cdot (v \circ T_S^n) \right| \leqslant C_4 \min\{\rho(l), \rho(n-m)\};$
3. $\left| \int \Big(v \cdot (v \circ T_S^l) \cdot (v \circ T_S^m) \cdot (v \circ T_S^n) \Big) - \int \Big(v \cdot (v \circ T_S^l) \Big) \int \Big((v \circ T_S^m) \cdot (v \circ T_S^n) \Big) \right| \leqslant C_4 \rho(m-l).$

This assumption states conditions on decay of correlations of variables of the form $v \circ T_S^k$ of orders 2 and 4. The condition $\sum_{i=1}^{\infty} i\rho(i) < \infty$ implies that the speed of decay is at least $\rho(k) = k^{-2-\epsilon}$. In particular, it is easy to verify that Assumption 1 holds for systems with exponential mixing of all orders (hence, for Hölder functions and ergodic toral automorphisms [26, 90]) since the latter condition implies much stronger mixing properties than Assumption 1 requires.

Recall that the map T_S is exponentially mixing for Hölder functions: see Theorem 5.1.3. Thus, assumptions 1a and 1b of Assumption 1 are immediate. Assumption 1c follows from the exponential mixing of the second order.

Assumption 2: There exists a function $\tilde{\rho}_N : \mathbb{N}_0 \to \mathbb{R}_+$ such that for any differentiable $A\colon \mathbb{R} \to \mathbb{R}$ with A' absolutely continuous and $\max_{0\leqslant k\leqslant 2} ||A^{(k)}||_\infty < 1$, and for any $0 \leqslant n < N$,

$$\Big| \int (A(W^n) \cdot v \circ T_A^n) \Big| \leqslant \tilde{\rho}_N(K).$$

Assumption 3: This is also an assumption in v 5.1.5.

In short, Assumption 1 holds in the setting of Theorem 5.1.5 and Assumption 3 is an assumption of Theorem 5.1.5. Therefore, in order to apply Theorem 5.3.1 we are left to verify that Assumption 2 holds for (v, T_S); in other words, we are left to prove the following proposition:

Proposition 5.4.1. Let $(\mathbb{T}^d, T_S, m)$ and $v\colon \mathbb{T}^d \to \mathbb{R}$ satisfy the assumptions of 5.1.5. Fix $0 \leqslant K < N$ and for $0 \leqslant n < N$ denote

$$W^n = \frac{1}{\sqrt{N}} \left(\sum_{i=0}^{n-K-1} v \circ T_S^i + \sum_{i=n+K+1}^{N-1} v \circ T_S^i \right).$$

Then, for any $A\colon \mathbb{R} \to \mathbb{R}$ with A', A'' defined almost everywhere such that $\max_{k=0,1,2} ||A^{(k)}||_\infty \leqslant 1$ there exist constants $|\theta| < 1$ and C which do not depend on N, K, n such that

$$\Big| \int (v \circ T_S^n) \cdot A(W^n) \Big| \leqslant C \cdot \theta^K \sqrt{N} \tag{5.6}$$

holds for all $0 \leqslant n < N$.

Proof. The matrix S associated to the ergodic toral automorphism $T_S : \mathbb{T}^d \to \mathbb{T}^d$ has a characteristic polynomial p_S which is irreducible and of Salem type.

In other words, it has a single root $\lambda \in \mathbb{R}$ outside the unit disc, a single root λ^{-1} inside the unit disc, and other $d-2$ roots lie on the unit circle. Because of the ergodicity of T_S no root of p_S is a root of unity. Since the automorphism T_S is ergodic with respect to the normalised Lebesgue measure m, it is also exponentially mixing for pairs of Hölder observables [44, 63].

For every $x \in \mathbb{T}^d$ the tangent space T_x is isomorphic to $\mathbb{R}^d$ and admits eigenspace decomposition: $T_x = E_x^u \oplus E_x^n \oplus E_x^s$ where E_x^u is a 1-dimensional unstable eigenspace corresponding to the largest eigenvalue λ, E_x^s is a 1-dimensional stable eigenspace corresponding to the eigenvalue $\pm\lambda^{-1}$, and E_x^n is a $(d-2)$-dimensional neutral eigenspace corresponding to eigenvalues having unit absolute values. Since p_S is irreducible, it follows that all its roots are distinct, and the roots on the unit circle come in pairs of complex conjugates. Hence, the action induced by T_S on E_x^n is an isometry. The action induced by T_S on E_x^u is expanding and the action induced by T_S on E_x^s is contracting. Therefore, even though v is Lipschitz, the function W^n is not Lipschitz continuous uniformly in N; for instance, it grows rapidly in the unstable direction. Thus, we cannot directly apply the results of [44, 63] on decay of correlations for Lipschitz functions to prove Proposition 5.4.1.

Let us present W^n as a sum of two terms, $W^n = W_-^n + W_+^n$, where

$$W_-^n = \frac{1}{\sqrt{N}} \sum_{i=0}^{n-K-1} v \circ T_S^i \quad \text{and} \quad W_+^n = \frac{1}{\sqrt{N}} \sum_{i=n+K+1}^{N-1} v \circ T_S^i.$$

Lemma 5.4.2. There exist a finite partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$ and a set of numbers $\{c_q\}_{q\in Q}$ such that

$$\int_{\mathbb{T}^d} A(W^n)\cdot(v\circ T_S^n)\, dm \leqslant \Big| \sum_q \mu(\xi_q) \int_{\xi_q} A(c_q+W_+^n)\cdot(v\circ T_S^n)\, d\nu_q \Big| + C_1\lambda^{-K/3(d-2)}\sqrt{N},$$

where $\nu_q(\cdot) = m(\xi_d)^{-1} m(\cdot \cap \xi_q)$ and C_1 does not depend on N, K, n.

Proof. Introduce a partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$ into approximately $\lambda^{n+\frac{d-1}{3(d-2)}K}$ parallelepipeds of equal size with sides parallel to eigendirections. The length of the side parallel to the unstable direction is λ^{-n}, the length of the side parallel to the stable direction is $\lambda^{-K/3(d-2)}$, and the lengths of all sides parallel to neutral directions is $\lambda^{-K/3(d-2)}$. Note that the action of T_S is expanding in the unstable direction with coefficient λ, contracting in the stable direction with coefficient λ^{-1}, and is an isometry in neutral eigenspace; the action of T_S on the atoms of the partition and on the partition described above is shown in Figure 5.3.

On each atom ξ_q we introduce an induced measure $\nu_q(\cdot) = m(\xi_d)^{-1}m(\cdot \cap \xi_q)$ which is the normalised Lebesgue measure conditioned to ξ_q. Then, W_-^n is nearly constant on each ξ_q; in other words, the variation $W_-^n|_{\xi_q}$ is proportional to $N^{\frac{1}{2}}\lambda^{-K/3(d-2)}$. Indeed, take

$$c_q = \int W_-^n \, d\nu_q = m(\xi_q)^{-1} \int_{\xi_q} W_-^n \, dm.$$

Then

$$\begin{aligned}
\sup_{x\in\xi_q} |W_-^n(x) - c_q| &\leq \sup_{x,y\in\xi_q} |W_-^n(x) - W_-^n(y)| \\
&\leq \frac{1}{\sqrt{N}} \sum_{j=0}^{n-K-1} \sup_{x,y\in\xi_q} |v^j(x) - v^j(y)| \\
&\leq \frac{L}{\sqrt{N}} \sum_{j=0}^{n-K-1} \operatorname{diam}(T_S^j \xi_q),
\end{aligned}$$

where $\operatorname{diam}(T_S^j\xi_q)$ stands for $\sup_{x,y\in T_S^j\xi_q} |x-y|$. We can bound the diameter of the parallelepiped ξ_q by the sum of its sides: $\operatorname{diam}(\xi_q) \leqslant \lambda^{-n} + \lambda^{-K/3(d-2)} + \lambda^{-K/3(d-2)}(d-2)$. The automorphism T_S acts on the distances in the unstable direction by multiplying it by λ, in the stable direction – by multiplying it by λ^{-1}, and T_S is an isometry in the neutral directions. We conclude that

$$\begin{aligned}
\sup_{x\in\xi_q} |W_-^n(x) - c_q| &\leq \frac{L}{\sqrt{N}} \sum_{j=0}^{n-K-1} \left(\lambda^j\lambda^{-n} + \lambda^{-j}\lambda^{-K/3(d-2)} + (d-2)\lambda^{-K/3(d-2)}\right) \\
&\leq \frac{L}{\sqrt{N}} \left(\frac{\lambda^{n-K}-1}{\lambda-1}\lambda^{-n} + \frac{\lambda}{\lambda-1}\lambda^{-K/3(d-2)}\right) \\
&+ \frac{L}{\sqrt{N}} \left((n-K)(d-2)\lambda^{-K/3(d-2)}\right) \\
&\leq \frac{L}{\sqrt{N}}\lambda^{-K/3(d-2)} \left(\frac{1}{\lambda-1} + N(d-2)\right).
\end{aligned}$$

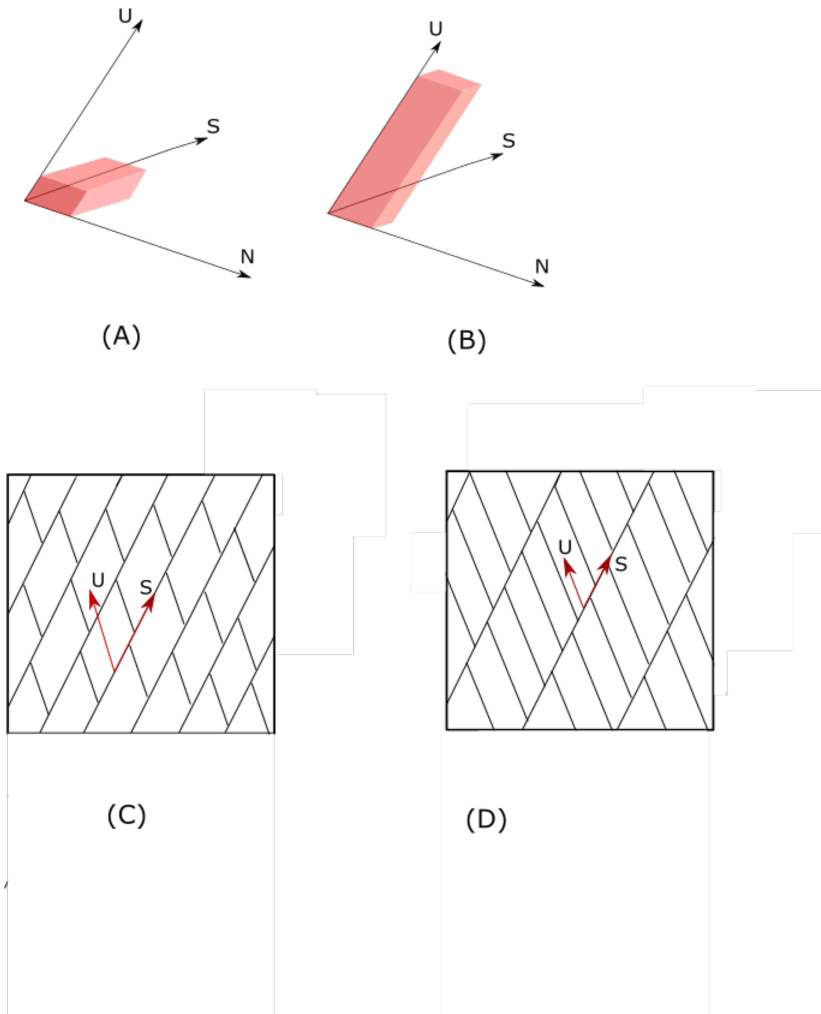

Figure 5.3: (A): an atom of the partition $\{\xi_q\}_{q\in Q}$ is a parallelepiped ξ_q whose sides are parallel to eigendirections. (B): the image of ξ_q under the action of T_S. (C): a projection on stable and unstable directions of the partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$. (D): the image of the projection on stable and unstable directions of the partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$ under the action of T_S.

By the mean value theorem we conclude that

$$\begin{aligned}\max_{q\in Q}\sup_{x\in\xi_q}|A(W^n(x)) \quad &- \quad A(c_q+W_+^n(x))| \\ &\leq \quad \|A'\|_\infty \sup_{q\in Q, x\in\xi_q}|W^n(x)-(c_q+W_n^+(x))| \\ &\leq \quad \frac{L\|A'\|_\infty\lambda^{-K/3(d-2)}}{\sqrt{N}}\left(\frac{1}{\lambda-1}+N(d-2)\right)\end{aligned}$$

and

$$\begin{aligned}\textstyle\int_{\mathbb{T}^d} \quad & A(W^n)\cdot v\circ T_S^n\,dm \leq \left|\sum_q m(\xi_q)\int_{\xi_q}A(W^n)\cdot v\circ T_S^n\,d\nu_q\right| \\ \leq \quad & \Big|\sum_q m(\xi_q)\int_{\xi_q}A(c_1+W_+^n)\cdot v\circ T_S^n\,d\nu_q + \\ + \quad & \sum_q m(\xi_q)\int_{\xi_q}\frac{L\|A'\|_\infty\lambda^{-K/3(d-2)}}{\sqrt{N}}\left(\frac{1}{\lambda-1}+N(d-2)\right)\cdot v\circ T_S^n\,d\nu_q\Big| \\ \leq \quad & \Big|\sum_q m(\xi_q)\int_{\xi_q}A(c_q+W_+^n)\cdot v\circ T_S^n\,d\nu_q\Big| \\ + \quad & \frac{L\|A'\|_\infty\|f\|_\infty\lambda^{-K/3(d-2)}}{\sqrt{N}}\left(\frac{1}{\lambda-1}+N(d-2)\right).\end{aligned}$$

The proof is complete. □

Let us rewrite $W_+^n = \widetilde{W}_+^n\circ T_S^{n+K+1}$ where $\widetilde{W}_+^n = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-n-K-2} v\circ T_S^i$. Since $T_S^n : \xi_q\to T_S^n(\xi_q)$ is a diffeomorphism on $\{\xi_q\}$ and, therefore, $T_S^n(\xi_q)$ has no self-intersections, $m_q := (T_S^n)_*\nu_q = m(\xi_q)m(\cdot\cap T_S^n(\xi_q))$ is Lebesgue measure conditioned to $T_S^n(\xi_q)$. The sets $T_S^n(\xi_q)$ form a partition of $\mathbb{T}^d$ consisting of skinny parallelepipeds of approximate unstable side length 1, stable side length $\lambda^{-n-K/3(d-2)}$, and neutral sides length $\lambda^{-K/3(d-2)}$. Then,

$$\begin{aligned}\sum_{q\in Q} \quad m(\xi_q) \quad & \int_{\xi_q}A(c_q+W_+^n)\cdot v\circ T_S^n\,d\nu_q \\ = \quad & \sum_{q\in Q}m(\xi_q)\int_{\xi_q}\left(v\cdot A(c_q+\widetilde{W}_+^n)\circ T_S^{K+1}\right)\circ T_S^n\,d\nu_q \\ = \quad & \sum_{q\in Q}m(\xi_q)\int_{T_S^n(\xi_q)}v\cdot A(c_q+\widetilde{W}_+^n)\circ T_S^{K+1}\,dm_q \\ = \quad & \int_{\mathbb{T}^d}v\cdot A(c(x)+\widetilde{W}_+^n)\circ T_S^{K+1}\,dm.\end{aligned}$$

where $c(x) = c_q$ if $x \in T_S^n(\xi_q)$. The function $A(c_q + \widetilde{W}_+^n)$ is discontinuous at $\partial T_S^n(\xi_q)$ and not uniformly Lipschitz inside each $T_S^n(\xi_q)$, so we cannot directly apply the exponential decay of correlations result to finish the proof. Instead, our next step is to prove the correlation bounds using measure disintegration and the Koksma inequality.

Lemma 5.4.3. There exists a constant C_2 which does not depend on N, K, n such that

$$\int_{\mathbb{T}^d} v \cdot A(c(x) + \widetilde{W}_+^n) \circ T_S^{K+1}\, dm \leqslant C_2 \lambda^{-K/3(d-2)}.$$

Proof. Divide each atom $T_S^n(\xi_q)$ into approximately $\lambda^{K/3(d-2)}$ parallelepipeds $\{\zeta_{q'}\}_{q'\in Q'}$ of unstable side length $\lambda^{-K/3(d-2)}$. Then, each $\zeta_{q'}$ has unstable side length $\lambda^{-K/3(d-2)}$, stable side length $\lambda^{-n-K/3(d-2)}$, and neutral sides length $\lambda^{-K/3(d-2)}$; the set $\{\zeta_{q'}\}_{q'\in Q'}$ is a partition of $\mathbb{T}^d$. On each $\zeta_{q'}$, v varies no more than $Ld\lambda^{-K/3(d-2)}$, so for $m_{q'} = \frac{m(\cdot\cap\zeta_{q'})}{m(\zeta_{q'})}$ and $v_{q'} := \int v\, dm_{q'}$,

$$\begin{aligned}\Big| \int_{\mathbb{T}^d} & v \cdot A(c(x) + \widetilde{W}_+^n(x)) \circ T_S^{K+1}\, dm - \\ & - \sum_{q'\in Q'} m(\zeta_{q'}) v_{q'} \int A(c(x) + \widetilde{W}_+^n(x)) \circ T_S^{K+1}\, dm_{q'} \Big| \\ & \leq Ld\|A\|_\infty \lambda^{-K/3(d-2)}.\end{aligned}$$

Let P be a $(d-1)$-dimensional section in the direction $E^c \oplus E^u$ passing through 0 in $\mathbb{T}^d$. Let $\{\ell_y\}_{y\in P}$ be a partition of $\mathbb{T}^d$ into arcs in the stable direction such that each ℓ_y intersects P in its endpoints (one of which is y) and doesn't intersect P in its interior. Each ℓ_y is an interval on $\mathbb{T}^d$ whose length depends on y; however, locally the lengths are the same and they change discretely in a finite number of discontinuity points. The number of discontinuity points is bounded from above and the bound depends only on d. It follows that the leaves ℓ_y of the same length form a finite number of parallelepipeds that partition the torus (see Figure 5.4) and the lengths of ℓ_y are bounded from above and from below.

The stable foliation induces a disintegration of the Lebesgue measure m: for each measurable $v : \mathbb{T}^d \to \mathbb{R}$,

$$\int_{\mathbb{T}^d} v\, dm = \int_P \int_{\ell_y} v\, d\nu_{\ell_y}\, dm_P.$$

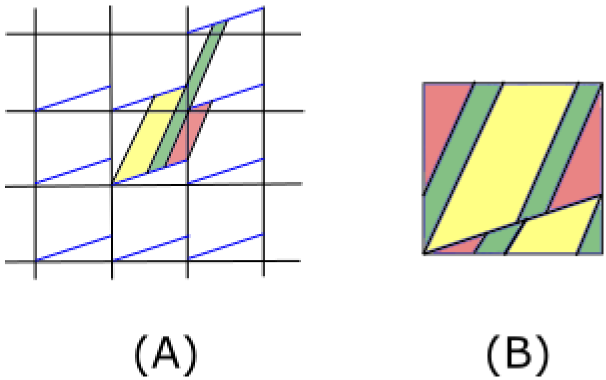

Figure 5.4: (A): Two-dimensional projection of $\mathbb{T}^d$ is a factor of $\mathbb{R}^2$ by $\mathbb{Z}^2$. The projection of P is shown by blue lines. The fibers of the stable foliation $\{\ell_y\}_{y\in P}$ are shown in yellow, green, and red; the boxes of different colours depict the regions of $\mathbb{T}^d$ with different fiber length. (B): The boxes where the fibers $\{\ell_y\}_{y\in P}$ have different length induce a partition of $\mathbb{T}^d$. The number of boxes is bounded from above and the bound depends only on d.

Note that since ℓ_y are intervals, the measure ν_{ℓ_y} is a scaled copy of the 1-dimensional Lebesgue measure for each y.

For $q' \in Q'$ consider $Z_{q'} := T_S^{K+1}(\zeta_{q'})$: it is a long skinny parallelepiped that wraps about $\lambda^{K+1-K/3(d-2)}$ around the torus in the unstable direction, and has width $\lambda^{-K/3(d-2)}$ in the neutral directions, and has width equal to $\lambda^{-n-K-1-K/3(d-2)}$ in the stable direction (see Figure 5.5). The parallelepiped $Z_{q'}$ intersects ℓ_y in intervals $\{I_j^y\}_{j=1,\dots,j_y}$ of length $\lambda^{-n-1-K(1+1/3(d-2))}$. Since the intervals I_j^y have length $\lambda^{-n-K-1-K/3(d-2)}$, the function v restricted to I_j^y varies very little. Hence, if we replace $\nu_y|_{I_j^y}$ by a single Dirac mass $\nu_y(I_j^y)\delta_{x_j^y}$ for some $x_j^y \in I_j^y$, we make only an error of order $\lambda^{-n-1-K(1+1/3(d-2))}$. Such errors are naturally absorbed in the other exponential errors we identified above.

Since $Z_{q'}$ consists of roughly $\lambda^{K+1-K/3(d-2)}$ pieces of unit (unstable) length, (stable) width $\lambda^{-n-K-1-K/3(d-2)}$, and projecting along the stable direction to $(d-1)$-dimensional parallelepipeds of $(d-1)$-dimensional area equal to $\left(\lambda^{-K/3(d-2)}\right)^{d-2}$ in P (which are roughly distributed uniformly over P), $Z_{q'}$ intersect ℓ_y roughly

$$j_y \sim \operatorname{len}(\ell_y)\lambda^{K+1-K/3(d-2)}\left(\lambda^{-K/3(d-2)}\right)^{d-2} = \operatorname{len}(\ell_y)\lambda^{\frac{K}{3}(2-\frac{1}{d-2})}$$

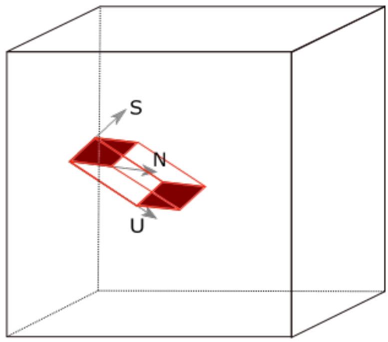

Figure 5.5: (A): an atom $\zeta_{q'}$. S, N, U stand for stable, neutral, and unstable eigendirections, respectively. (B): the parallelepiped $Z_{q'} := T_S^{K+1}(\zeta_{q'})$ is a skinny parallelepiped which wraps around $\mathbb{T}^d$ multiple times in the unstable direction.

times. This is the number of intervals I_j^y.

Since all eigenspaces are in irrational algebraic directions, consecutive intersections of parallel translations of such eigenspaces are obtained by a rotation over an irrational algebraic number. The components $\{v_i^u\}_{i=1,\dots,d}$ of the unit vector $\vec{v^u} \in E^u$ are rationally independent because T_S is ergodic. Also $\{v_i^u\}$ are algebraic numbers, and hence, by the Siegel-Roth theorem [96, 101], they are Diophantine of order ϵ for any $\epsilon > 0$. [3] It follows that the points x_j^y are roughly uniformly distributed on ℓ_y in the sense that the *discrepancy* (see [28, 60])

$$\mathfrak{D}_R^*(\{x_n\}_n) = \sup_k \sup_{[a,b)\subset \ell} \left| \frac{1}{R} \#\{k+1 \le n \le k+R : x_n \in [a,b)\} - (b-a) \right|$$

behaves as that of the consecutive points in the orbit of a rotation over a Diophantine number:

$$\mathfrak{D}_{j_y}^*(\{\alpha n\}) \le C_\epsilon j_y^{-\frac{1}{1+\nu}+\epsilon} \le C\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})},$$

see [60, Theorem 3.2-3.4]. Recall that f and therefore $\widetilde{W}_+^n$ and also $A(c+\widetilde{W}_-^n)$ is Lipschitz (uniformly in N, n and K) in the stable direction. Now we can

[3] A number α is called Diophantine if for every $\epsilon > 0$ there is a constant C such that $|\alpha - \frac{p}{q}| \ge Cq^{-2-\epsilon}$ for all $p, q \in \mathbb{Z}$, $q \neq 0$.

estimate $\int_{\ell_y} g\, d\nu_{\ell_y}$ using the Koksma inequality:

$$\left|\sum_{j=1}^{N} g(x_j^y) - \int_{\ell_y} g(x)\, dx\right| \leq \mathrm{Var}(g)\, \mathfrak{D}^*_{j_y}((x_j^y)).$$

Taking $g(x) = A(c(x) + \widetilde{W}^n_+(x))$, we obtain

$$\begin{aligned} \Big|\frac{1}{j_y}\sum_{j=1}^{j_y} A(c(x_j^y) + \widetilde{W}^n_+(x_j^y)) \quad &- \quad \int_{\ell_y} A(c(x) + \widetilde{W}^n_+)\, d\nu_\ell\Big| \\ &\leq \quad \mathrm{Var}(A(c(\cdot) + \widetilde{W}^n_+(\cdot))|_{\ell_y})\mathfrak{D}^*_{j_y} \\ &\ll \quad L\mathrm{Var}(A)\lambda^{-\frac{K}{3}(2-1/(d-2))}. \end{aligned}$$

Using the estimate above and the measure disintegration, we can estimate the integral over $Z_{q'}$:

$$\begin{aligned} &\int \quad A(c(x) + \widetilde{W}^n_+(x)) \circ T_S^{K+1}\, dm_{q'} \\ &\approx \quad \int_P \int_{\ell_y} A(c + \widetilde{W}^n_-(x)) d\nu_\ell\, dm_S + O(L\|A'\|_\infty \lambda^{-\frac{K}{3}(2-\frac{1}{d-2})}) \\ &= \quad \int_{\mathbb{T}^d} A(c + \widetilde{W}^n_-) dm + O(L\|A'\|_\infty \lambda^{-\frac{K}{3}(2-\frac{1}{d-2})}). \end{aligned}$$

Finally, summing over all $\zeta_{q'}$ with weights $v_{q'}$, we find

$$\begin{aligned} &\int_{\mathbb{T}^2} v \cdot A(c(x) + \widetilde{W}^n_+) \circ T_S^{K+1}\, dm \\ &\quad = \sum_{q'\in Q'} m(\zeta_{q'})\Big(v_{q'} \int A(c(x) + \widetilde{W}^n_+) \circ T_S^{K+1}\, dm_{q'} + O(Ld\|A\|_\infty \lambda^{-K/3(d-2)})\Big) \\ &\quad \leq \sum_{q'\in Q'} m(\zeta_{q'}) \left(v_{q'} \int_{\mathbb{T}^d} A(c + \widetilde{W}^n_-)\, dm + O(L\|A'\|_\infty \lambda^{-\frac{K}{3}(2-\frac{1}{d-2})})\right) \\ &\quad + O(Ld\|A\|_\infty \lambda^{-K/3(d-2)}) \\ &\quad \leq \left(\sum_{q'\in Q'} m(\zeta_{q'}) v_{q'} \int_{\mathbb{T}^d} A(c + \widetilde{W}^n_-)\, dm\right) \\ &\quad + O\left(L\|A'\|_\infty \lambda^{-\frac{K}{3}(2-\frac{1}{d-2})} + Ld\|A\|_\infty \lambda^{-K/3(d-2)}\right) \\ &\quad = \int v\, dm \int A(c + \widetilde{W}^n_-)\, dm + O\left(L\|A'\|_\infty \lambda^{-\frac{K}{3}(2-\frac{1}{d-2})} + Ld\|A\|_\infty \lambda^{-K/3(d-2)}\right) \\ &\quad = O\left(L\|A'\|_\infty \lambda^{-\frac{K}{3}(2-\frac{1}{d-2})} + Ld\|A\|_\infty \lambda^{-K/3(d-2)}\right) \end{aligned}$$

where the last equality follows from the assumption that $\int v\, dm = 0$. □

Lemma 5.4.2 and Lemma 5.4.3 prove the Assumption 2 of Theorem 5.3.1 with $\tilde{\rho}_N(k) = \tilde{C}\theta^k\sqrt{N}$. Theorem 5.3.1 implies that

$$d_{\mathcal{W}}(W^N, Z) \leqslant C_\# \left(\frac{K+1}{\sqrt{N}} + \sum_{i=K+1}^{\infty} \rho(i) \right) + C'_\# \sqrt{N} \tilde{\rho}_N(K) \qquad (5.7)$$

where $\rho(k) = \theta^k$ and $\tilde{\rho}_N(k) = \tilde{C}\theta^k\sqrt{N}$ for some $\theta < 1$. Plugging ρ and $\tilde{\rho}_N$ into 5.7 and choosing $K(N) = \frac{3\log N}{2\log\theta}$ (see Corollary 5.3.3) yields

$$d_{\mathcal{W}}(W^N, Z) \leqslant C_\# \left(\frac{3\log N + 1}{2\log\theta\sqrt{N}} + \frac{\theta}{N^{3/2}(1-\theta)} \right) + \frac{C'_\#}{\sqrt{N}} \leqslant C\frac{\log N}{\sqrt{N}}$$

which concludes the proof of Theorem 5.1.5.

□