Limit Theorems for Slowly Mixing Dynamical Systems

by

Matt Galton

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Mathematics Institute

September 2019
Contents

List of Figures iii
Acknowledgments iv
Declarations v
Abstract vi

Chapter 1 Introduction 1

Chapter 2 Preliminary Material 5
  2.1 Rates of Mixing .................................................. 5
  2.2 Auxiliary Results ................................................ 9

Chapter 3 Limit Theorems for Non-Invertible Maps 12
  3.1 A Martingale Coboundary Decomposition .................. 12
  3.2 Central Limit Theorem .......................................... 15
    3.2.1 Scalar-Valued Observables ............................... 15
    3.2.2 Vector-Valued Observables ............................. 20
  3.3 Weak Invariance Principle ................................. 24
  3.4 Iterated Weak Invariance Principle ....................... 35
    3.4.1 Suspensions and Examples .............................. 37
    3.4.2 Proof of the Iterated WIP ............................ 40

Chapter 4 Limit Theorems for Invertible Maps 52
  4.0.1 Sufficient Criteria for Assumption (A_{inv}) .......... 53
  4.0.2 An Example – Intermittent Baker’s map ............... 55
  4.1 A Martingale Coboundary Decomposition .................. 57
  4.2 Central Limit Theorem ........................................ 67
  4.3 Weak Invariance Principle ................................. 70
4.4 Iterated Weak Invariance Principle ................................. 72
  4.4.1 Suspensions and Examples ................................. 73
  4.4.2 Proof of the Iterated WIP ................................. 77
List of Figures

3.1 A graph of the intermittent map $T$ defined by (3.1), for $\alpha = 0.9$. . . 16
3.2 A typical sample path of the process $W_n$ for $d = 1$. . . . . . . . 25
3.3 A graphical representation of the suspension flow $T_t$. . . . . . . . . 39
Acknowledgments

Firstly, huge thanks go to my supervisor Ian Melbourne for his constant support and guidance throughout my PhD. I couldn’t have asked for a better supervisor.

Thanks must also go to my family, for their support and serving as a welcome respite throughout my PhD. To my dad, thanks for the uncountable curries and beer. To my mum, thanks for lending a kind ear. To my brother, thanks for distracting me from my PhD with rugby and comedy nights when needed, and paying for all of our lunches!

To Cerys, words can’t describe how much your optimism, love and support has helped me over the last four years. Sneaking doodles into my notes never failed to make me laugh. We’ve made Marie proud! Thanks also to my second family, Stephen, Ann, Becky, Tom and Jess, for the laughs and letting me use your home for holiday trips in the sun.

To everyone that shared a coffee or an evening playing computer games with me, pretending to listen to my problems, thanks.

Finally, thanks to Warwick’s Maths department for the wine and cheese, and the Stats department for the coffee!
Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree, apart from the material in Chapter 2 and Sections 3.1, 3.2, which was previously submitted to the University of Warwick for the degree of Masters of Science. This material has undergone significant change, and new results have been added since submission.

I plan on publishing a paper on the non-invertible and invertible iterated weak invariance principle.
Abstract

In this thesis we prove the iterated weak invariance principle for ergodic, probability-preserving dynamical systems with respect to $L^\infty$ observables under a mild mixing assumption. When the dynamics can be modelled by a Young tower the iterated weak invariance principle is already known under optimal conditions. The setting where $T$ is not necessarily modelled by a Young tower still has gaps, however. This is the setup considered in this thesis, and it is flexible enough to include time-one maps of suitable (semi-)flows. For both non-invertible and invertible maps, we improve upon the previous best results. For non-invertible maps, our mixing assumption is optimal when correlations decay at polynomial rates.
Chapter 1

Introduction

Many chaotic dynamical systems have similar properties to random processes, such as the law of large numbers (or Birkhoff’s ergodic theorem), the central limit theorem (CLT) (see for example [9, 13, 22, 28, 32, 47, 50]) and the weak invariance principle (WIP) (see [9, 13, 22, 50]).

For limit theorems in continuous time, a standard mechanism exists whereby it often suffices to show a limit theorem for a suitable Poincaré map [39]. This has been successfully applied to billiard flows [3], geodesic flows over negatively curved manifolds [13, 47] and the classical Lorenz equations [23], amongst others. Consequently this thesis only considers discrete time dynamical systems.

Many advances have focussed on particular dynamical systems, such as the collision map for various billiards [9], expanding maps of the interval with a neutral fixed point [33, 45, 53] and Hénon maps [4]. In this thesis we take an abstract approach: Let $T : \Lambda \rightarrow \Lambda$ be an ergodic, measure-preserving transformation over a probability space $(\Lambda, \mathcal{A}, \mu)$ and let $v : \Lambda \rightarrow \mathbb{R}^d$ be a mean zero, $L^\infty(\Lambda, \mathbb{R}^d)$ observable. From an abstract perspective, the literature available for limit theorems in discrete time is vast [5, 11, 14, 17, 25, 31, 34, 35] and various different techniques are available [18, 43, 49]. We focus on the martingale approach of Gordin [18]. The main idea is to relate the observable $v$ to a martingale through a coboundary and apply readily available martingale limit theorems [21].

For $t \in [0, 1]$ define $W_n(t) := n^{-1/2} \sum_{j=0}^{\lfloor nt \rfloor} v \circ T^j$, and define the $d \times d$ matrix-valued process $\mathcal{W}_n(t) := \int_0^t W_n^i \ dW_n^j$, where the integral is understood in the Stieltjes sense. We say $(W_n, \mathcal{W}_n)$ satisfies an iterated weak invariance principle (iterated WIP) if $(W_n, \mathcal{W}_n) \rightarrow_w (W, \mathcal{W})$ holds over the Skorohod space $(D([0, 1], \mathbb{R}^d \times \mathbb{R}^{d \times d}), \| \cdot \|_\infty)$, where $\rightarrow_w$ denotes weak convergence, $W$ is a suitable $d$ dimensional Brownian motion and $\mathcal{W}$ is a $d \times d$ matrix-valued process.
Motivating the Iterated WIP

Note that $W_n$ actually depends on two variables, $W_n(x,t) = n^{-1/2} \sum_{j=0}^{[nt]-1} v(T^j x)$. Roughly speaking, if we choose $x \in \Lambda$ randomly according to $\mu$, then we can consider $W_n(x,\cdot) : [0,1] \to \mathbb{R}^d$ as a random path. The WIP says that on very large time scales, $W_n$ looks like random noise. More precisely $W_n \to_w W$.

A natural question is to suppose that $W_n$ are driving some differential equations, and ask if the solutions converge in some sense. This is known as homogenization [10, 19, 29, 30, 38]. Since $W_n$ converges to a random noise, one might hope that the solutions converge to a process containing a random noise component. To be more precise, define $z_n(t) = [nt]/n$ and suppose $a \in C^1(\mathbb{R}^m, \mathbb{R}^m)$ and $b \in C^{2+}(\mathbb{R}^m, \mathbb{R}^{m \times d})$. Let $x_n$ be the unique $m$-dimensional solution of the (discontinuous) ODE

$$dx_n = a(x_n^-)dz_n + b(x_n^-)dW_n, \quad x_n(0) = \zeta \in \mathbb{R}^m,$$

where $\zeta$ is fixed. We hope that $x_n$ converges weakly to some random process $X$, and $X$ is the unique solution to an SDE of the form

$$dX = a(X)dt + b(X) \diamond dW, \quad X(0) = \zeta,$$

where $b(X) \diamond dW$ is a suitable stochastic integral (see [10, Theorem 4.10] for more details). Unfortunately, the WIP alone does not uniquely determine the weak limit of the iterated integrals $\int W_n^i dW_n^j$, so things can become problematic:

**Example.** In equation (1.1) take $m = d = 2$, $a \equiv 0$, $b(x_n^1, x_n^2) = \begin{pmatrix} 1 & 0 \\ 0 & x_n^1 \end{pmatrix}$ and $\zeta = 0$. We now have

$$dx_n^1 = dW_n^1, \quad dx_n^2 = x_n^1 dW_n^2.$$ 

Therefore $x_n^1(t) = W_n^1(t)$ and $x_n^2(t) = \int_0^t W_n^1 dW_n^2$. The mapping $\chi : D([0,1], \mathbb{R}^2) \to D([0,1], \mathbb{R}^2)$, $\chi(g)(t) = (g^1(t), \int_0^t g^1(t) dg^2)$ is not continuous and so the WIP $W_n \to_w W$ does not guarantee weak convergence of $x_n$.

The saving grace is that understanding the weak limit of the iterated integrals is nearly enough to guarantee convergence of $x_n$:

**Theorem** ([10, Theorem 4.10]). If $(W_n, \mathcal{W}_n)$ satisfies an iterated WIP and a suitable moment bound (see [10, Assumption 4.7]), then $x_n \to_w X$, where $x_n$ and $X$ are solutions to equations (1.1) and (1.2) respectively.
In this thesis we focus on the iterated WIP for both invertible and non-invertible maps. For ease of exposition, the following discussion considers correlations decaying at polynomial rates.

Non-Invertible maps

Suppose for every $n \geq 1$, $i = 1, \ldots, d$ and $w \in L^\infty(\Lambda)$ we have

$$\left| \int v^i w \circ T^n \, d\mu \right| \leq C\|w\|\infty n^{-\alpha},$$

where $C > 0$ and $\alpha > 0$.

The CLT and WIP hold when $\alpha > 1$ and are due to Liverani [32, Theorem 1.1] and Dedecker and Rio [11, Theorem 1], respectively.

If $T$ is modelled by a Young tower [52, 53], then the iterated WIP is known under the optimal condition $\alpha > 1$ [29, Theorem 10.2]. By optimal we mean that there are examples where the CLT fails for $\alpha \leq 1$ [20].

When $T$ is not necessarily modelled by a Young tower, the iterated WIP holds under the hypothesis $\alpha > 2$ by [29, Theorem 4.3]. In this thesis, we improve this result to $\alpha > 1$, so the iterated WIP holds exactly when the CLT holds. Our result is optimal for correlations decaying at polynomial rates. We also give examples where this new result applies.

Invertible maps

The mixing condition for invertible maps is slightly different to that of non-invertible maps: Assume there exists a sub-sigma algebra $A_0 \subset A$ such that $T^{-1}A_0 \subset A_0$ and assume that for every $n \geq 1$,

$$\|E_0(v \circ T^{-n})\|_1 \leq Cn^{-\alpha}, \quad \|E_0(v \circ T^n) - v \circ T^n\|_1 \leq Cn^{-\beta},$$

where $C > 0$ and $\alpha, \beta > 0$. The first condition is analogous to the mixing condition in the non-invertible case.

The CLT and WIP hold when $\alpha, \beta > 1$ and are due to Liverani [32, Theorem 1.2] and Dedecker and Rio [11, Corollary 4(b)], respectively.

If $T$ is modelled by a Young tower, take $A_0$ as the sigma algebra generated by the stable foliation. Then the first condition corresponds to the rate of decay of correlations for the quotient map, and the second condition corresponds to the average rate of contraction along stable leaves [12]. The iterated WIP is known under the condition $\alpha, \beta > 1$ [40, Corollary 2.3].
When $T$ is not necessarily modelled by a Young tower, the iterated WIP holds for $\alpha, \beta > 2$ [29, Theorem 5.2]. In this thesis we improve this to $\alpha > 1, \beta > 2$. This result may have room for further improvement as the CLT/WIP holds when $\alpha, \beta > 1$. We give an example not covered by previous work.

**Thesis layout**

In Chapter 2 we overview some basic preliminary material on mixing.

Chapter 3 contains the limit theorems for non-invertible dynamical systems: Section 3.2 proves the CLT and recovers part of a well known result by Liverani [32, Theorem 1.1]. We use a different decomposition due to Melbourne [36, Equation (2.1)], yielding a more elementary proof. Section 3.3 proves the WIP using this decomposition, setting up some preliminary calculations needed for the iterated WIP. Section 3.4 proves the iterated WIP, and gives an example which is not covered by previous known work.

In Chapter 4 we prove the corresponding invertible limit theorems and provide an example not covered by previous work.

We write $a_n \ll b_n$ if there is a constant $C > 0$ such that $a_n \leq C b_n$ for all $n \geq 1$. The letter $C$ is reserved for some positive constant.
Chapter 2

Preliminary Material

2.1 Rates of Mixing

Let \((\Lambda, \mathcal{A}, \mu)\) be a probability space and \(T : \Lambda \to \Lambda\) be a measurable map. We say \(T\) is measure-preserving with respect to \(\mu\) if for every \(A \in \mathcal{A}\) we have \(\mu(T^{-1}A) = \mu(A)\). An equivalent definition of measure-preserving is that \(\int v \circ T \, d\mu = \int v \, d\mu\) for every \(v \in L^1(\Lambda)\). We say \(T\) is ergodic with respect to \(\mu\) if for all \(A \in \mathcal{A}\) such that \(T^{-1}A = A\) we have \(\mu(A) = 0\) or \(1\).

Let \(v : \Lambda \to \mathbb{R}\) be a measurable function. We call \(v\) an observable. The observable \(v\) can be thought of as taking ‘observations’ of a system as it evolves, \((v(x), v(Tx), v(T^2x), \ldots)\). Let

\[ v_n := \sum_{j=0}^{n-1} v \circ T^j \]

be the \(n\)th Birkhoff sum of \(v\) (a subscript \(n\) after any observable denotes the \(n\)th Birkhoff sum).

**Proposition 2.1.1** (Birkhoff’s Ergodic Theorem [44, Theorem 2.3]). Let \(T : \Lambda \to \Lambda\) be an ergodic, measure-preserving transformation over the probability space \((\Lambda, \mathcal{A}, \mu)\), and let \(v \in L^1(\Lambda)\). Then

\[ \frac{1}{n} \sum_{j=0}^{n-1} v \circ T^j \to_{a.s.} \int v \, d\mu. \]

\( \square \)

Suppose \(T\) is a measure-preserving transformation with respect to \(\mu\). One method for measuring how unpredictable the dynamics of a system are is to see how
well correlated a ‘measurement at time zero’ $v(x)$ and a ‘measurement at time $n$’ $v(T^nx)$ are. We measure this using the correlation function

$$C_n(v, w) := \int vw \circ T^n d\mu - \int v d\mu \int w d\mu.$$ 

The map $T$ is called mixing with respect to $\mu$ if for every $v, w \in L^2(\Lambda)$,

$$C_n(v, w) \to 0 \quad \text{as } n \to \infty.$$ 

Mixing can be arbitrarily slow with respect to $L^\infty(\Lambda)$ observables, so we restrict ourselves to consider observables in smaller Banach spaces $\mathcal{L}$, such as bounded variation or Hölder observables. We say that $T : \Lambda \to \Lambda$ mixes exponentially with respect to $\mathcal{L}$ if there exists constants $K > 0$, $c > 0$ and a suitable test space $K$ such that $C_n(v, w) \leq Ke^{-cn}$ for every $n \geq 1$, $v \in \mathcal{L}$ and $w \in K$. We say that $T : \Lambda \to \Lambda$ mixes polynomially with respect to $\mathcal{L}$ if there exists constants $K > 0$, $\beta > 0$ and a suitable test space $K$ such that $C_n(v, w) \leq Kn^{-\beta}$ for every $n \geq 1$, $v \in \mathcal{L}$ and $w \in K$. The rate of mixing indicates how quickly a system ‘forgets’ its initial condition.

Let $p \in [1, \infty]$ and $q^{-1} = 1 - p^{-1}$ be its conjugate, with $q = 1$ when $p = \infty$. Define the Koopman operator

$$U : \mathcal{L}^p(\Lambda) \to \mathcal{L}^p(\Lambda), \quad Uv := v \circ T.$$ 

The Koopman operator enjoys the following properties:

• $\int Uv d\mu = \int v d\mu$.

• $\|Uv\|_p = \|v\|_p$ for all $v \in \mathcal{L}^p(\Lambda)$, $p \in [1, \infty]$.

• $\int UvUw d\mu = \int vw d\mu$ for all $v \in \mathcal{L}^p(\Lambda)$, $w \in \mathcal{L}^q(\Lambda)$.

The transfer operator $P : \mathcal{L}^q(\Lambda) \to \mathcal{L}^q(\Lambda)$ is defined as the unique operator satisfying

$$\int vw \circ T d\mu = \int Pv w d\mu \quad \text{for all } v \in \mathcal{L}^q(\Lambda), \ w \in \mathcal{L}^p(\Lambda).$$

The transfer operator has the following properties:

• $P1 = 1$.

• $\int Pvd\mu = \int vd\mu$.

• $\|Pv\|_p \leq \|v\|_p$ for all $p \in [1, \infty]$.
• \( PU = I \).

**Proposition 2.1.2.** Let \( T : \Lambda \rightarrow \Lambda \) be a measure-preserving map. If \( T \) is invertible, then \( UP = I \) and \( \| P^nv \|_p = \| v \|_p \) for all \( v \in L^p(\Lambda) \) and \( p \geq 1 \). In general we have \( UP = \mathbb{E}(\cdot \mid T^{-1}A) \).

**Proof.** Let \( v \in L^1(\Lambda) \) and \( w \in L^\infty(\Lambda) \). If \( T \) is invertible, then
\[
\int P^nv\, w\, d\mu = \int v \circ T^{-1} \circ T \circ w \circ T \, d\mu = \int v \circ T^{-1} \, w \, d\mu.
\]
Therefore \( P^nv = v \circ T^{-1} \) and \( UP = I \). The fact \( \| P^nv \|_p = \| v \|_p \) for any \( p \geq 1 \) now follows easily from invariance of \( T \).

Next suppose \( T \) is a general measure-preserving map and let \( A \in \mathcal{A} \). By uniqueness of conditional expectation we must show
\[
\int T^{-1}A \, UPv \, d\mu = \int T^{-1}A \, v \, d\mu
\]
and that \( UPv \) is \( T^{-1}A \)-measurable. The first condition follows from the following calculation:
\[
\int T^{-1}A \, UPv \, d\mu = \int \mathbb{1}_A \, UPv \, d\mu = \int \mathbb{1}_A \, Pv \, d\mu = \int T^{-1}A \, v \, d\mu.
\]
Next, since \( P : L^1(\Lambda) \rightarrow L^1(\Lambda) \), we see \( P^nv \) is \( \mathcal{A} \)-measurable and hence \( UPv \) is \( T^{-1}A \)-measurable. \( \square \)

For any \( n \geq 1 \), the transfer operator for \( T^n \) is \( P^n \), since
\[
\int v \circ T^n \, d\mu = \int P^nv \circ T^{n-1} \, d\mu = \cdots = \int P^n v \, d\mu.
\]

**Proposition 2.1.3.** Suppose \( T : \Lambda \rightarrow \Lambda \) is a measure-preserving map and let \( \gamma_n \geq 0 \), for \( n \geq 1 \). Then for any \( v \in L^1(\Lambda) \) with \( \int v \, d\mu = 0 \) the following statements are equivalent:

• \( \| P^n v \|_1 \leq \gamma_n \),

• For every \( w \in L^\infty(\Lambda) \) we have \( |C_n(v, w)| \leq \| w \|_\infty \gamma_n \).

**Proof.** First suppose \( \| P^n v \|_1 \leq \gamma_n \). Then for any \( w \in L^\infty(\Lambda) \) we have
\[
|C_n(v, w)| = \left| \int v \circ T^n \, d\mu \right| = \left| \int P^n v \, w \, d\mu \right| \leq \| P^n v \|_1 \| w \|_\infty \leq \| w \|_\infty \gamma_n.
\]

Next suppose that \( |C_n(v, w)| \leq \| w \|_\infty \gamma_n \) for any \( w \in L^\infty(\Lambda) \). Then,
\[
\| P^n v \|_1 = \left| \int P^n v \, \text{sign}(P^n v) \, d\mu \right| \leq \| \text{sign}(P^n v) \|_\infty \gamma_n = \gamma_n.
\]
Rates of mixing depend on the regularity of the observable, as we see from the two following examples.

**Example 2.1.4.** Let $T : [0, 1] \rightarrow [0, 1]$ be the doubling map, $Tx = 2x \mod 1$, and let $v : [0, 1] \rightarrow \mathbb{R}$ be a Lipschitz continuous observable. For any $n \geq 1$ we have

$$\|P^n v - \int v \, d\mu\|_\infty \leq |v|_{\text{Lip}} 2^{-n},$$

where $|v|_{\text{Lip}} := \sup_{x \neq y} \frac{|v(x) - v(y)|}{|x - y|}$ is the Lipschitz constant of $v$.

To see this, first note that $(P v)(x) = \frac{1}{2} \left[ v\left(\frac{x}{2}\right) + v\left(\frac{x + 1}{2}\right)\right]$, so

$$|(P v)(x) - (P v)(y)| = \frac{1}{2} \left[ v\left(\frac{x}{2}\right) - v\left(\frac{y}{2}\right)\right] + \frac{1}{2} \left[ v\left(\frac{x + 1}{2}\right) - v\left(\frac{y + 1}{2}\right)\right]$$

$$\leq \frac{1}{2} |v|_{\text{Lip}} \left| \frac{x}{2} - \frac{y}{2} \right| + \frac{1}{2} |v|_{\text{Lip}} \left| \frac{x + 1}{2} - \frac{y + 1}{2} \right|$$

$$= |v|_{\text{Lip}} \left| \frac{x}{2} - \frac{y}{2} \right| + \frac{1}{2} |v|_{\text{Lip}} |x - y|.$$

Therefore $|P v|_{\text{Lip}} \leq \frac{1}{2} |v|_{\text{Lip}}$. Next note that for any $w : \Lambda \rightarrow \mathbb{R}$ Lipschitz, we have

$$\left\| w(x) - \int w \, d\mu \right\| \leq \int |w(x) - w(y)| \, d\mu(y) \leq |w|_{\text{Lip}} \int |x - y| \, d\mu(y)$$

$$\leq |w|_{\text{Lip}} \sup_y |x - y| \mu(\Lambda) = |w|_{\text{Lip}} \sup_y |x - y|.$$

Hence

$$\|P^n v - \int v \, d\mu\|_\infty = \|P^n v - \int P^n v \, d\mu\|_\infty \leq \sup_x \left( |P^n v|_{\text{Lip}} \sup_y |x - y| \right)$$

$$= |P^n v|_{\text{Lip}} \text{diam} \Lambda = |P^n v|_{\text{Lip}} \leq |v|_{\text{Lip}} 2^{-n}.$$

Although mixing is exponential for this map with respect to Lipschitz observables, we can construct $L^\infty$ observables that mix as slow as we desire:

**Example 2.1.5.** Let $\Sigma := \{0, 1\}^\mathbb{N}$ be the space of sequences over two symbols and let $\sigma : \Sigma \rightarrow \Sigma$ be the left-shift, $\sigma((x_0 x_1 x_2 \cdots)) = (x_1 x_2 \cdots)$. Define the cylinders $[a_0 \cdots a_n; m] := \{ x \in \Sigma : x_{m+i} = a_i, i = 0, \ldots, n \}$ and let $\mu$ be the $(1/2, 1/2)$-Bernoulli measure. That is, $\mu$ is the extension of the function that assigns value $2^{-n}$ to every cylinder of length $n$. Let $\gamma_n \geq 0$ be any sequence such that $\gamma_n \rightarrow 0$ and
choose \( N \) such that \( \gamma_N < 1/4 \). Let \( A = [1; 0], B = [0; N] \) and note
\[
|\mu(\sigma^{-N}A \cap B) - \mu(A)\mu(B)| = \left| \mu([1; N] \cap [0; N]) - \frac{1}{4} \right| = \frac{1}{4} > \gamma_N.
\]

2.2 Auxiliary Results

Here we collect some results that will prove useful for the rest of the thesis.

**Proposition 2.2.1.** Let \((b_{r,n})_{r,n \geq 1}\) be a sequence of real numbers such that the following conditions hold:

(a) \( \sum_{r=1}^{\infty} \sup_n |b_{r,n}| < \infty \)

(b) for every \( r \geq 1 \), \( \lim_{n \to \infty} b_{r,n} = 0 \).

Then
\[
\sum_{r=1}^{\infty} b_{r,n} \to 0,
\]
as \( n \to \infty \).

**Proof.** Let \( \epsilon > 0 \). By condition (a) choose \( R \geq 1 \), independently of \( n \), such that
\[
\sum_{r=R+1}^{\infty} |b_{r,n}| < \epsilon.
\]
For each \( r \geq 1 \), condition (b) guarantees us an \( N_r \) such that \( |b_{r,n}| < \epsilon/R \) for all \( n > N_r \). Therefore \( \sum_{r=1}^{R} |b_{r,n}| < \epsilon \) for \( n > \max_{r \leq R} N_r \). Hence
\[
\sum_{r=1}^{\infty} |b_{r,n}| < 2\epsilon.
\]

**Proposition 2.2.2.** If \( c_n \) is a sequence of real numbers such that \( n^{-1}c_n \to c \), then \( \sup_{0 \leq t \leq 1} |n^{-1}c_{[nt]} - tc| \to 0 \).

**Proof.** Suppose without loss that \( c > 0 \) and note that \(-ntc \leq -[nt]c\). Calculate that
\[
\sup_{0 \leq t \leq 1} |n^{-1}c_{[nt]} - tc| = n^{-1} \sup_t |c_{[nt]} - ntc|
\leq n^{-1} \sup_t |c_{[nt]} - [nt]c| \leq n^{-1} \max_{0 \leq j \leq n} |c_j - jc|.
\]
Denoting $d_j = |c_j - jc|$, we see that $n^{-1}d_n \to 0$ and $\max_{0 \leq j \leq n} d_j = d_{m(n)}$ for some non-decreasing sequence $m(n) \leq n$. Note that $\sup_k |n^{-1}c_{[nt]} - tc| = n^{-1}d_{m(n)}$.

Finally, either $m(n)$ is bounded and so $n^{-1}d_{m(n)} \to 0$ as $n \to \infty$, or

$$\sup_t |n^{-1}c_{[nt]} - tc| = n^{-1}d_{m(n)} \leq m(n)^{-1}d_{m(n)} \to 0,$$

as $m(n) \to \infty$. 

\begin{proof}
By the ergodic theorem, we have

$$\frac{1}{n}(f^p)_n \to_{a.s.} \int f^p \, d\mu.$$

It follows that

$$\frac{1}{n}f^p \circ T^{n-1} = \frac{1}{n}((f^p)_n - (f^p)_{n-1}) = \frac{1}{n}(f^p)_n - \left(\frac{n}{n-1}\right) \frac{1}{n-1}(f^p)_{n-1} \to_{a.s.} 0.$$

Noting $\max_{0 \leq j \leq n-1} |f|^p \circ T^j = \sup_{0 \leq t \leq 1} |f|^p \circ T^{[(n-1)t]}$, Proposition 2.2.2 shows $n^{-1} \max_{0 \leq j \leq n-1} |f|^p \circ T^j \to_{a.s.} 0$ and therefore $n^{-1/p} \max_{0 \leq j \leq n-1} |f|^p \circ T^j \to_{a.s.} 0$. 

\end{proof}

**Proposition 2.2.3.** Suppose $T : \Lambda \to \Lambda$ is an ergodic, measure-preserving map over the probability space $(\Lambda, \mathcal{A}, \mu)$, and let $f \in L^p(\Lambda)$ for some $p \geq 1$. Then

$$n^{-1/p} \max_{0 \leq j \leq n-1} |f|^p \circ T^j \to_{a.s.} 0.$$

**Proof.** By the ergodic theorem, we have

$$\frac{1}{n}(f^p)_n \to_{a.s.} \int f^p \, d\mu.$$

It follows that

$$\frac{1}{n}f^p \circ T^{n-1} = \frac{1}{n}((f^p)_n - (f^p)_{n-1}) = \frac{1}{n}(f^p)_n - \left(\frac{n}{n-1}\right) \frac{1}{n-1}(f^p)_{n-1} \to_{a.s.} 0.$$

Noting $\max_{0 \leq j \leq n-1} |f|^p \circ T^j = \sup_{0 \leq t \leq 1} |f|^p \circ T^{[(n-1)t]}$, Proposition 2.2.2 shows $n^{-1} \max_{0 \leq j \leq n-1} |f|^p \circ T^j \to_{a.s.} 0$ and therefore $n^{-1/p} \max_{0 \leq j \leq n-1} |f|^p \circ T^j \to_{a.s.} 0$. 

**Remark 1.** If $T$ is invertible, then a near identical proof shows $n^{-1/p} \max_{1 \leq j \leq n} |f|^p \circ T^{-j} \to_{a.s.} 0$.

We call a sequence of random variables $X_n : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ **tight** if for all $\epsilon > 0$ there exists a $K > 0$ such that $\max_{n \geq 1} \mathbb{P}(|X_n| \geq K) < \epsilon$.

**Proposition 2.2.4.** Let $X_n, Y_n : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be random variables such that $X_n$ is a tight sequence and $Y_n \to_p 0$. Then $X_nY_n \to_p 0$.

**Proof.** Let $\lambda, K > 0$ and note

$$\mathbb{P}(|X_nY_n| > \lambda) = \mathbb{P}(|X_nY_n| > \lambda, |X_n| \leq K) + \mathbb{P}(|X_nY_n| > \lambda, |X_n| > K)$$

$$\leq \mathbb{P}(|Y_n| \geq \frac{\lambda}{K}) + \mathbb{P}(|X_n| > K).$$

Let $\epsilon > 0$ and choose $K$ large enough so $\mathbb{P}(|X_n| > K) < \epsilon$. Now take $n$ large enough.
so \( P(\{|Y_n| \geq \frac{1}{R}\}) < \epsilon \). We now have

\[
P(|X_n Y_n| > \lambda) \leq 2\epsilon.
\]

\[\square\]

**Proposition 2.2.5.** Define \( h : \mathbb{R}^n \to \mathbb{R}, \ h(a) := \max\{0, \sum_{i=1}^{1} a_i, \ldots, \sum_{i=1}^{n} a_i\}. \)

Then

\[
|h(a) - h(b)| \leq \sum_{i=1}^{n} |a_i - b_i|.
\]

**Proof.** If \( h(a) = h(b) = 0 \), then the result is clear. Next suppose \( h(a) = \sum_{i=1}^{n_a} a_i \), \( h(b) = \sum_{i=1}^{n_b} b_i \) for some \( n_a, n_b \leq n \). Without loss assume \( h(a) \geq h(b) \). Then

\[
|h(a) - h(b)| = h(a) - h(b) = \sum_{i=1}^{n_a} a_i - \sum_{i=1}^{n_b} b_i.
\]

By definition of \( h \), we have \( h(b) = \sum_{i=1}^{n_b} b_i \geq \sum_{i=1}^{p} b_i \) for any \( p = 1, \ldots, n \). In particular, \(-\sum_{i=1}^{n_b} b_i \leq -\sum_{i=1}^{n_a} b_i \) and so

\[
|h(a) - h(b)| \leq \sum_{i=1}^{n_a} (a_i - b_i) \leq \sum_{i=1}^{n_a} |a_i - b_i| \leq \sum_{i=1}^{n} |a_i - b_i|.
\]

\[\square\]
Chapter 3

Limit Theorems for Non-Invertible Maps

Throughout this chapter, we suppose \( T : \Lambda \to \Lambda \) is a ergodic, measure-preserving map over the probability space \((\Lambda, \mathcal{A}, \mu)\), with associated transfer operator \( P \). Further assume \( v \in L^\infty(\Lambda) \) is mean zero and satisfies the following mixing assumption:

\[
\sum_{n=1}^\infty \|P^n v\|_1 < \infty. \tag{A}
\]

**Remark 1.** If \( T \) is invertible, then Proposition 2.1.2 shows \( P : L^p(\Lambda) \to L^p(\Lambda) \) is an isometry for every \( p \geq 1 \). Therefore Assumption (A) (and hence the hypothesis of Theorem 3.1.2) can never be non-trivially satisfied.

### 3.1 A Martingale Coboundary Decomposition

We say \( v \) admits a **martingale coboundary decomposition** if \( v = m + \chi \circ T - \chi \) for some measurable observable \( \chi \) with \( m \in \text{Ker} \ P \). We call \( \chi \circ T - \chi \) the **coboundary** and say \( m \) and \( v \) are **cohomologous**. We call \( \chi \circ T - \chi \) an \( L^p(\Lambda) \) coboundary if \( \chi \in L^p(\Lambda) \).

**Proposition 3.1.1.** Let \( k \geq 1 \) and define \( \chi_1^k := \sum_{n=1}^k P^n v \). We have the decomposition \( v = m^{(k)} + \chi_1^k \circ T - \chi_1^k + P^k v \), where \( \chi_1^k, m^{(k)} \in L^\infty(\Lambda) \) and \( m^{(k)} \in \text{Ker} \ P \).

**Proof.** For \( k \geq 1 \) define \( m^{(k)} := v - \chi_1^k \circ T + \chi_1^k - P^k v \). Then \( \chi_1^k, m^{(k)} \in L^\infty(\Lambda) \) follows easily since \( \|P^n v\|_\infty \leq \|v\|_\infty < \infty \) for any \( n \geq 1 \). We see \( m^{(k)} \in \text{Ker} \ P \) by
considering

\[ Pm^{(k)} = Pv - PU\chi^k + P\chi^k - P^{k+1}v = Pv - \chi^k + P\chi^k - P^{k+1}v \]
\[ = Pv - P^{k+1}v + \sum_{n=1}^{k} (P^{n+1}v - P^n v) = 0. \]

\[ \square \]

By Assumption (A), \( \chi := \sum_{n=1}^{\infty} P^n v \) is well defined. Further, \( \chi^k \to \chi \) in \( L^1(\Lambda) \) since \( \| \chi - \chi^k \|_1 \leq \sum_{n=k+1}^{\infty} \| P^n v \|_1 \to 0. \)

**Theorem 3.1.2.** We have the martingale coboundary decomposition \( v = m + \chi \circ T - \chi \), where \( \chi \in L^1(\Lambda) \). Further, \( m \in \text{Ker}(P) \), \( m \in L^2(\Lambda) \) and \( m^{(k)} \to m \) in \( L^2(\Lambda) \).

**Proof.** Define \( m := v - \chi \circ T + \chi \). Then \( \chi \in L^1(\Lambda) \) follows directly from Assumption (A) and we see \( m \in \text{Ker} P \) since

\[ Pm = Pv - PU\chi + P\chi = Pv - \chi + P\chi = Pv + \sum_{n=1}^{\infty} (P^{n+1}v - P^n v) = 0. \]

It remains to show \( m^{(k)} \to m \) in \( L^2(\Lambda) \). Assumption (A) shows \( m^{(k)} \to m \) in \( L^1(\Lambda) \) since

\[ \| m - m^{(k)} \|_1 = \left\| \chi^\infty_{k+1} - \chi^\infty_{k+1} \circ T + P^k v \right\|_1 = \left\| \sum_{n=k+1}^{\infty} (P^n v - UP^n v) + P^k v \right\|_1 \]
\[ \leq 2 \sum_{n=k}^{\infty} \| P^n v \|_1 \to 0. \]

By uniqueness of limits in \( L^1(\Lambda) \) it suffices to show \( m^{(k)} \) is a Cauchy sequence in \( L^2(\Lambda) \). Without loss suppose \( \ell > k \) and calculate that

\[ m^{(\ell)} - m^{(k)} = \chi^\ell_1 - \chi^\ell_1 \circ T - P^\ell v - \chi^k_1 + \chi^k_1 \circ T + P^k v \]
\[ = \chi^\ell_{k+1} - \chi^\ell_{k+1} \circ T + P^k v - P^\ell v = \chi^{\ell-1}_k - \chi^{\ell}_k \circ T. \]
Hence, we see $m^{(k)}$ is a Cauchy sequence by considering

$$\int (m^{(ℓ)} - m^{(k)})^2 \, d\mu = \int (m^{(ℓ)} - m^{(k)}) (\chi^{ℓ-1}_k - \chi^{ℓ}_k \circ T) \, d\mu$$

$$= \int \chi^{ℓ-1}_k (m^{(ℓ)} - m^{(k)}) \, d\mu - \int \chi^{ℓ}_k P (m^{(ℓ)} - m^{(k)}) \, d\mu = \int \chi^{ℓ-1}_k (\chi^{ℓ-1}_k - \chi^{ℓ}_k \circ T) \, d\mu$$

$$= \int \chi^{ℓ-1}_k \, d\mu - \int (\chi^{ℓ-1}_k - \chi^{ℓ}_k) \, d\mu = \int \chi^{ℓ-1}_k (\chi^{ℓ-1}_k + \chi^{ℓ}_k) \, d\mu$$

$$= \int (P^ℓ v - P^k v) (\chi^{ℓ-1}_k + \chi^{ℓ}_k) \, d\mu \leq \|P^ℓ v - P^k v\|_\infty \|\chi^{ℓ-1}_k + \chi^{ℓ}_k\|_1$$

$$\leq 2\|v\|_\infty (\|\chi^{ℓ-1}_k\|_1 + \|\chi^{ℓ}_k\|_1) \ll \sum_{n=k}^ℓ \|P^n v\|_1 \to 0,$$

as $ℓ, k \to \infty$. □

This technique is a standard part of the martingale method, first initialised by Gordin [18]. See also [21, 32].

**Proposition 3.1.3** ($L^2$ orthogonality of martingale differences). Suppose $f, g : \Lambda \to \mathbb{R}$ are square integrable observables such that $f, g \in \text{Ker } P$. Then for every $i, j \geq 0$,

$$\int f \circ T^i g \circ T^j \, d\mu = \delta_{ij} \int f g \, d\mu,$$

where $\delta_{ij}$ is the Kronecker delta.

**Proof.** When $i = j$ the result follows from invariance of $T$. If $i < j$, then

$$\int f \circ T^i g \circ T^j \, d\mu = \int (f g \circ T^{j-i}) \circ T^i \, d\mu = \int f g \circ T^{j-i} \, d\mu$$

$$= \int P f g \circ T^{j-i-1} \, d\mu = 0,$$

where we have used $f \in \text{Ker } P$. The case $i > j$ follows from the same argument and the fact that $g \in \text{Ker } P$. □

For square integrable observables $f \in \text{Ker } P$, Proposition 3.1.3 shows

$$\left\| \sum_{j=0}^{n-1} f \circ T^j \right\|_2 = n^{1/2}\|f\|_2.$$
3.2 Central Limit Theorem

3.2.1 Scalar-Valued Observables

We say a sequence of random variables $Z_n : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ satisfies a central limit theorem (CLT) with mean zero and variance $\sigma^2$ if

$$\mathbb{P}(Z_n \leq a) \rightarrow \int_{-\infty}^{a} \left(2\pi\sigma^2\right)^{-1/2} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx,$$

for any $a \in \mathbb{R}$. This convergence is denoted by $Z_n \rightarrow_d N(0, \sigma^2)$. The classical Lindeberg-Lévy CLT states that if $Z_n = n^{-1/2} \sum_{j=1}^{n} X_j$ for $X_j$ a sequence of independent and identically distributed square integrable random variables with mean zero, then $Z_n \rightarrow_d N(0, \sigma^2)$ with $\sigma^2 = \text{Var}(X_1)$. We are interested in the case $X_j = v \circ T^j$, which form a dependent sequence.

**Theorem 3.2.1 (CLT).** We have that $\sum_{r=1}^{\infty} \int v \circ T^r \, d\mu$ converges and

$$n^{-1/2}v_n \rightarrow_d N(0, \sigma^2),$$

where $\sigma^2 = \int v^2 \, d\mu + 2\sum_{r=1}^{\infty} \int v \circ T^r \, d\mu$. Moreover, $\sigma^2 = 0$ if and only if $v$ is an $L^1(\Lambda)$ coboundary.

**Remark 1.** This result is not new; a stronger result was shown by Liverani [32, Theorem 1.1].

**Remark 2.** Zweimüller [54, Corollary 1] proved that the CLT also holds with respect to every probability measure $\nu$ which is absolutely continuous with respect to $\mu$.

We now give an example of a slowly mixing map which satisfies the hypothesis of Theorem 3.2.1.

**Example 3.2.2.** Suppose $0 < \alpha < 1$ and let $T : [0, 1] \rightarrow [0, 1]$ be an intermittent map of the form

$$T x = \begin{cases} 
  x(1 + 2^{\alpha}x^{\alpha}) & \text{if } 0 \leq x \leq 1/2 \\
  2x - 1 & \text{if } 1/2 < x \leq 1
\end{cases} \quad (3.1)$$

Figure 3.1 shows a graph of $T$ for $\gamma = 0.9$. Liverani, Saussol and Vaienti [33, Lemma 2.3] showed there exists a unique ergodic invariant probability measure $\mu$ which is absolutely continuous with respect to Lebesgue. Let $v$ be a mean zero Hölder observable. Young [53, Theorem 5(c)] showed $|C_n(v, w)| \leq Cn^{-1/\alpha+1}$ for all $w \in L^\infty(\Lambda)$. By Proposition 2.1.3, assumption (A) and therefore the CLT holds when $0 < \alpha < 1/2$. 

15
When the correlations are not summable, the CLT generally fails [20, Theorem 3]: Take $1/2 < \alpha < 1$ in the above example (Example 3.2.2) and suppose $v$ is Hölder with $v(0) \neq 0$. Then $n^{-\alpha}v_n \to_d Z$, where $Z$ is a stable law [16, Chapter VI.1].

To prove Theorem 3.2.1, we first show that a CLT holds for the Birkhoff sums $n^{-1/2}m_n$ defined by the cohomologous observable $m$ in Theorem 3.1.2, and then relate this to convergence of $n^{-1/2}v_n$. This technique was initialised by Gordin [18]. To show that the sequence $n^{-1/2}m_n$ satisfies a CLT, we show that it forms a martingale and apply a suitable martingale CLT result.

We call an increasing sequence of sub $\sigma$-algebras $A_1 \subseteq A_2 \subseteq \cdots \subseteq A$ a filtration. We say that a sequence of random variables $Z_n$ is a martingale [15, Section 2.2] with respect to a filtration $A_n$ if for each $n \geq 1$,

1. $Z_n \in L^1$
2. $Z_n$ is $A_n$-measurable
3. $\mathbb{E}(Z_{n+1} \mid A_n) = Z_n$

We call $X_j$ a sequence of martingale differences if $Z_n = \sum_{j=1}^n X_j$ is a martingale. In this case, condition 3 is equivalent to $\mathbb{E}(X_{j+1} \mid A_j) = 0$ for all $j \geq 1$. The following proposition originates from Billingsley [5] (and proven independently by Ibragimov [25]).

**Proposition 3.2.3** (Brown’s Theorem [8]). Let $f : X \to X$ be an ergodic, measure-preserving map, and let $\phi \in L^2(\Lambda)$ with $\mathbb{E}\phi = 0$. Suppose that $S_n := \sum_{j=0}^{n-1} \phi \circ f^j$ is a martingale. Then $n^{-1/2}S_n \to_d \mathcal{N}(0, \sigma^2)$, where $\sigma^2 = \mathbb{E}\phi^2$. $\square$
For our Birkhoff sums \( n^{-1/2} m_n \), the natural candidate for a filtration is \( F_n = T^{-n} A \). However \( F_n \) forms a sequence of decreasing \( \sigma \)-algebras since \( T^{-1} A \subseteq A \). The solution is to lift to an invertible map in the following way:

**Proposition 3.2.4** ([44, pg. 13]). Suppose that \( T : (\Lambda, \mathcal{A}, \mu) \rightarrow (\Lambda, \mathcal{A}, \mu) \) is a surjective and measure-preserving map over the probability space \((\Lambda, \mathcal{A}, \mu)\). Then there exists a probability space \((\Lambda, \mathcal{A}, \mu)\), an invertible, measure-preserving map \( \tilde{T} : \Lambda \rightarrow \Lambda \) and a measure-preserving projection \( \pi : \Lambda \rightarrow \Lambda \) such that \( \pi \circ \tilde{T} = T \circ \pi \). If \( \mu \) is ergodic, then \( \tilde{\mu} \) can be taken to be ergodic too. \( \square \)

We say the two systems \( T, \tilde{T} \) in Proposition 3.2.4 are semi-conjugate. Using Proposition 3.2.4, define the lifted observable \( \tilde{m} = m \circ \pi : \Lambda \rightarrow \mathbb{R} \), and the forward and backward Birkhoff sums

\[
\tilde{m}_n := \sum_{j=0}^{n-1} \tilde{m} \circ \tilde{T}^j, \quad \tilde{m}^-_n := \sum_{j=1}^{n} \tilde{m} \circ \tilde{T}^{-j}.
\]  

Define the \( \sigma \)-algebras

\[
\mathcal{A}_n := \tilde{T}^{n-1} \mathcal{A}.
\]

Since \( \mathcal{A}_{-1} = \tilde{T}^{-1} \mathcal{A}_0 = \tilde{T}^{-1} \pi^{-1} \mathcal{A} = \pi^{-1} T^{-1} \mathcal{A} \), we see \( \mathcal{A}_{-1} \subseteq \mathcal{A}_0 \). Hence \( \mathcal{A}_n \subseteq \mathcal{A}_{n+1} \) for all \( n \in \mathbb{Z} \) and \( (\mathcal{A}_n)_{n \geq 0} \) defines a filtration.

**Proposition 3.2.5.** Suppose \( f \in L^1(\Lambda) \) is such that \( f \in \text{Ker} P \), and define \( \bar{f} := f \circ \pi \). Then \( \bar{f}_n^- := \sum_{j=1}^{n} \bar{f} \circ \tilde{T}^{-j} \) is a martingale with respect to the filtration \( \mathcal{A}_n = \tilde{T}^{-n} \pi^{-1} \mathcal{A} \).

**Proof.** Let \( n \geq 1 \). Firstly, we see \( \| \bar{f}_n^- \|_1 = n \| f \|_1 < \infty \). To see \( \bar{f}_n^- = \sum_{j=1}^{n} \bar{f} \circ \tilde{T}^{-j} \) is \( \mathcal{A}_n \)-measurable, note that for every Borel subset \( E \subseteq \mathbb{R} \) we have

\[
(f \circ \pi \circ \tilde{T}^{-j})^{-1} E = \tilde{T}^j \pi^{-1} (f^{-1} E) \in \mathcal{A}_j \subseteq \mathcal{A}_n.
\]

It remains to show \( \mathbb{E} \left( \bar{f} \circ \tilde{T}^{-n} \mid \tilde{T}^{n-1} \pi^{-1} \mathcal{A} \right) = 0 \). Note that

\[
\mathbb{E} \left( \bar{f} \circ \tilde{T}^{-n} \mid \tilde{T}^{n-1} \pi^{-1} \mathcal{A} \right) = \mathbb{E} \left( \bar{f} \mid \tilde{T}^{-1} \pi^{-1} \mathcal{A} \right) \circ \tilde{T}^{-n} = \mathbb{E} \left( f \mid \pi^{-1} T^{-1} \mathcal{A} \right) \circ \tilde{T}^{-n} = \mathbb{E} \left( f \mid T^{-1} \mathcal{A} \right) \circ \pi \circ \tilde{T}^{-n}.
\]

Proposition 2.1.2 shows \( \mathbb{E}(f \mid T^{-1} \mathcal{A}) = (Pf) \circ T \) and therefore

\[
\mathbb{E} \left( \bar{f} \circ \tilde{T}^{-n} \mid \tilde{T}^{n-1} \pi^{-1} \mathcal{A} \right) = \mathbb{E} \left( f \mid T^{-1} \mathcal{A} \right) \circ \pi \circ \tilde{T}^{-n} = (Pf) \circ T \circ \pi \circ \tilde{T}^{-n} = 0,
\]

as required. \( \square \)

17
Proposition 3.2.6 (Martingale CLT). The normalised Birkhoff sums $n^{-1/2}\tilde{m}_n$ satisfy a CLT with mean zero and variance $\sigma^2 = \int m^2 \, d\mu$.

Proof. Since $m$ is a mean zero, $L^2(\Lambda)$ observable by Theorem 3.1.2 and $\pi$ is measure-preserving, we see $\tilde{m}$ is a mean zero, $L^2(\tilde{\Lambda})$ observable. Proposition 3.2.5 shows that $\tilde{m}_n - n$ is a martingale with respect to the filtration $A_n = \tilde{T}^n \pi^{-1} A$. Proposition 3.2.3 now shows $n^{-1/2}\tilde{m}_n - n \to_d N(0, \sigma^2)$, where $\sigma^2 = \int \tilde{m}^2 \, d\tilde{\mu} = \int m^2 \, d\mu$. □

In the following two propositions, we obtain the variance claimed in Theorem 3.2.1.

Proposition 3.2.7 (Green-Kubo Formula). The Green-Kubo formula holds:

$$\lim_{n \to \infty} \frac{1}{n} \int v_n^2 \, d\mu = \int v^2 \, d\mu + 2 \sum_{r=1}^\infty \int v \circ T^r \, d\mu.$$ 

Proof. First note that

$$\frac{1}{n} \int v_n^2 \, d\mu = \frac{1}{n} \sum_{i,j=0}^{n-1} \int v \circ T^i v \circ T^j \, d\mu.$$ 

We split the sum into the diagonal $i = j$, the lower triangle $i < j$ and upper triangle $i > j$. The lower and upper triangle are the same, so we obtain

$$\frac{1}{n} \int v_n^2 \, d\mu = \int v^2 \, d\mu + \frac{2}{n} \sum_{0 \leq i < j < n} \int v \circ T^i v \circ T^j \, d\mu.$$ 

Next calculate that

$$\frac{1}{n} \int v_n^2 \, d\mu = \int v^2 \, d\mu + \frac{2}{n} \sum_{0 \leq i < j < n} \int v \circ T^i \, d\mu.$$ 

Let $a_{r,n} := 1_{(r \leq n-1)}(1-r/n) \int v \circ T^r \, d\mu$ and $a_r := \int v \circ T^r \, d\mu$. We conclude that the limit $\lim_{n \to \infty} n^{-1} \int v_n^2 \, d\mu$ exists and equals $\int v^2 \, d\mu + 2 \sum_{r=1}^\infty \int v \circ T^r \, d\mu$ by showing $\sum_{r=1}^\infty a_{r,n} \to \sum_{r=1}^\infty a_r$, as $n \to \infty$. We verify the conditions of Proposition 2.2.1 with

$$b_{r,n} := a_{r,n} - a_r = (1_{(r \leq n-1)}(1-r/n) - 1) \int v \circ T^r \, d\mu.$$
For condition (a), first note that $|b_{r,n}| \leq 2 \mid \int vv \circ T^r d\mu \mid$. The condition now follows from the following calculation:

$$\sum_{r=1}^{\infty} \sup_n |b_{r,n}| \leq 2 \sum_{r=1}^{\infty} \mid \int P^r v v d\mu \mid \leq 2 \|v\|_{\infty} \sum_{r=1}^{\infty} \|P^r v\|_1 < \infty.$$ 

Condition (b) holds since for each fixed $r \geq 1$,

$$\lim_{n \to \infty} \frac{1}{n} \int vv \circ T^r d\mu = \int vv \circ T^r d\mu.$$

Therefore

$$\lim_{n \to \infty} \frac{1}{n} \int v_n^2 d\mu = \int v^2 d\mu + 2 \sum_{r=1}^{\infty} a_r = \int v^2 d\mu + 2 \sum_{r=1}^{\infty} \int vv \circ T^r d\mu.$$

Proposition 3.2.8. We have that

$$\int m^2 d\mu = \lim_{n \to \infty} \frac{1}{n} \int v_n^2 d\mu.$$

Proof. Let $m^{(k)} = v - \chi_{1} \circ T + \chi_{1} - P^k v$ be defined as in Proposition 3.1.1. Let $\epsilon > 0$ and choose $k$ large enough so $\|m - m^{(k)}\|_2 < \epsilon$ and $\sum_{n=k}^{\infty} \|P^n v\|_1 < \epsilon^2$. By orthogonality of martingale differences (Proposition 3.1.3) we have

$$\|m\|_2 - n^{-1/2}\|v_n\|_2 = \|m\|_2 - \|m^{(k)}\|_2 + \|m^{(k)}\|_2 - n^{-1/2}\|v_n\|_2$$

$$= \|m\|_2 - \|m^{(k)}\|_2 + n^{-1/2}\|m^{(k)}\|_2 - \|m^{(k)}\|_2 - n^{-1/2}\|v_n\|_2$$

$$\leq \|m - m^{(k)}\|_2 + n^{-1/2}\|m^{(k)} - v_n\|_2$$

$$\leq \epsilon + n^{-1/2}\|v_n - m^{(k)}_n\|_2.$$

Next use the decomposition of $v$ to calculate that

$$\|v_n - m^{(k)}_n\|_2 = \left\| \chi_{1} \circ T^n - \chi_{1} + \sum_{j=0}^{n-1} (P^j v) \circ T^j \right\|_2$$

$$\leq 2 \|\chi_{1}\|_2 + \left\| (P^k v)_n \right\|_2 \leq 2 k \|v\|_{\infty} + \left\| (P^k v)_n \right\|_2.$$

Since $P^k v \in L^\infty(\Lambda)$ and $\sum_{n=1}^{\infty} \|P^n v\|_1 < \infty$, we use the Green-Kubo formula
(Proposition 3.2.7) with \( v \) replaced by \( P^k v \) to obtain

\[
\lim_{n \to \infty} \frac{1}{n} \int (P^k v)^2 d\mu = \int (P^k v)^2 d\mu + 2 \sum_{r=1}^{\infty} \int P^k v P^k v \circ T^r d\mu
\]

\[
\leq \|v\|_\infty \|P^k v\|_1 + 2 \sum_{r=1}^{\infty} \|v\|_\infty \|P^r v\|_1 \leq 2 \|v\|_\infty \sum_{r=k}^{\infty} \|P^r v\|_1 < 2 \|v\|_\infty \epsilon^2.
\]

Therefore

\[
\limsup_{n \to \infty} \left( \|m\|_2 - n^{-1/2}\|v_n\|_2 \right) \leq \limsup_{n \to \infty} \left( \epsilon + n^{-1/2}(2k\|v\|_\infty) + n^{-1/2}(\|P^k v\|_2) \right)
\]

\[
\leq \left( 1 + \sqrt{2\|v\|_\infty} \right) \epsilon,
\]

and the result is shown. \( \square \)

**Proof of Theorem 3.2.1.** By Proposition 3.2.6 we have

\[
n^{-1/2} \tilde{m}^\to \to_d N(0, \sigma^2),
\]

where \( \sigma^2 = \int m^2 d\mu \). Propositions 3.2.7 and 3.2.8 show \( \sigma^2 = \int v^2 d\mu + 2 \sum_{r=1}^{\infty} \int v \circ T^r d\mu \).

Since \( \tilde{m}^\to = \tilde{m}^\to \circ T^n = \tilde{m}_n = m_n \circ \pi = \to_d m_n \), we have \( n^{-1/2}m_n \to_d N(0, \sigma^2) \). The central limit theorem for \( n^{-1/2}v_n \) now follows easily, since

\[
n^{-1/2}\|v_n - m_n\|_1 = n^{-1/2}\|\chi \circ T^n - \chi\|_1 \leq 2n^{-1/2}\|\chi\|_1 \to 0.
\]

Finally, we see \( \sigma^2 = \|m\|_2^2 = 0 \) if and only if \( m = 0 \). \( \square \)

### 3.2.2 Vector-Valued Observables

Suppose \( v : \Lambda \to \mathbb{R}^d \) is a vector-valued observable, with each coordinate \( v^i \in L^\infty(\Lambda) \).

Define the \( L^p(\Lambda, \mathbb{R}^d) \) norm

\[
\|v\|_p := \left\| |v|_{\mathbb{R}^d} \right\|_{L^p(\Lambda, \mathbb{R})},
\]

where \( |\cdot|_{\mathbb{R}^d} \) is any norm on \( \mathbb{R}^d \). We define the transfer operator \( P \) as acting coordinatewise on vector-valued observables. Assumption (A) becomes \( \sum_{n=1}^{\infty} \|P^n v^i\|_1 < \infty \) for each \( i = 1, \ldots, d \) and we immediately obtain the decompositions in Proposition 3.1.1 and Theorem 3.1.2 for vector-valued observables. Let \( v_n \) denote the \( n \)th Birkhoff sum, and note \( (v_n)^i = (v^i)_n \). For two vectors \( f = (f^1, \ldots, f^d) \) and \( g = (g^1, \ldots, g^d) \) define the outer product \( (f g^T)_{ij} := f^i g^j \).
We say a sequence of random vectors $Z_n : (\Omega, \mathcal{F}, P) \to \mathbb{R}^d$ satisfies a central limit theorem (CLT) with mean zero and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ if for any rectangle $I = [a_1, b_1] \times \cdots \times [a_d, b_d]$ we have

$$P(Z_n \in I) \to \int_I \det(2\pi\Sigma)^{-1/2} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) \, dx.$$ 

We denote this convergence by $Z_n \to_d \mathcal{N}(0, \Sigma)$. A result which will prove useful is the continuous mapping theorem.

**Proposition 3.2.9** (Continuous Mapping Theorem [27, Lemma 4.3, Theorem 4.27]). For any two metric spaces $S$ and $T$, let $X_1, X_2, \ldots$ be random elements in $S$ with $X_n \to_w X$, and consider a continuous map $f : S \to T$. Then $f(X_n) \to_w f(X)$. Moreover, if $X_n \to_p X$, then $f(X_n) \to_p f(X)$.

A well known way of relating distributional convergence in $\mathbb{R}^d$ to distributional convergence in $\mathbb{R}$ is the Cramér-Wold device:

**Proposition 3.2.10** (Cramér-Wold device [27, Corollary 4.5]). Let $X_n, X$ be random vectors in $\mathbb{R}^d$, for $n \geq 1$. Then $X_n \to_d X$ if and only if $c^T X_n \to_d c^T X$ in $\mathbb{R}$, for any $c \in \mathbb{R}^d$.

For convergence in probability, the picture is a little simpler:

**Proposition 3.2.11.** Let $X_n, X$ be random vectors in $\mathbb{R}^d$, for $n \geq 1$. Then $X_n \to_p X$ if and only if $X^i_n \to_p X^i$ for each coordinate $i = 1, \ldots, d$.

**Proof.** If $X_n \to_p X$, then $X^i_n \to_p X^i$ for each $i = 1, \ldots, d$ by the continuous mapping theorem. Next suppose $X^i_n \to_p X^i$ for every $i = 1, \ldots, d$. Then we have

$$P\left(\left|\left(X^1_n, \ldots, X^d_n\right) - (X^1, \ldots, X^d)\right| > a\right) \leq P\left(\sum_{i=1}^d |X^i_n - X^i| > a\right) \leq \sum_{i=1}^d P\left(|X^i_n - X^i| > \frac{a}{d}\right) \to 0,$$

as $a \to \infty$. \qed

The main theorem for this section is the following:

**Theorem 3.2.12** (Vector CLT [6, Theorem 29.5]). We have that $\sum_{r=1}^\infty \int v \left((v \circ T^r)^T + (v \circ T^r) v^T\right) d\mu$ converges and

$$n^{-1/2} v_n \to_d \mathcal{N}(0, \Sigma),$$

21
where $\Sigma = \int v v^T d\mu + \sum_{r=1}^{\infty} \int v (v \circ T^r)^T + (v \circ T^r) v^T d\mu$. For the degenerate case, $\det \Sigma = 0$ if and only if there exists a non-zero $c \in \mathbb{R}^d$ such that $c^T v$ is an $L^1(\Lambda)$ coboundary.

The following two propositions give the correct covariance claimed in Theorem 3.2.12.

**Proposition 3.2.13 (Vector Green-Kubo Formula).** The following Green-Kubo formula holds:

$$
\lim_{n \to \infty} \frac{1}{n} \int v_n v_n^T d\mu = \int v v^T d\mu + \sum_{r=1}^{\infty} \int (v \circ T^r)^T + (v \circ T^r) v^T d\mu.
$$

**Proof.** We show the coordinate wise statement. That is, for each $i, j = 1, \ldots, d$,

$$
\lim_{n \to \infty} \frac{1}{n} \int v^i_n v^j_n d\mu = \int v^i v^j d\mu + \sum_{r=1}^{\infty} \int v^i v^j \circ T^r + v^j v^i \circ T^r d\mu.
$$

Using the polarisation identity ($4 \int fg = \|f + g\|^2 - \|f - g\|^2$) we see

$$
\frac{1}{n} \int v^i_n v^j_n d\mu = \frac{1}{4n} \int (v^i_n + v^j_n)^2 d\mu - \frac{1}{4n} \int (v^i_n - v^j_n)^2 d\mu. \quad (3.4)
$$

Since $v^i \pm v^j \in L^\infty(\Lambda)$ and $\sum_{k=1}^{\infty} \|P^k (v^i \pm v^j)\|_1 \leq \sum_{k=1}^{\infty} \|P^k v^i\|_1 + \|P^k v^j\|_1 < \infty$ we use the scalar Green-Kubo formula (Proposition 3.2.7) with $v$ replaced by $v^i \pm v^j$ to see

$$
\lim_{n \to \infty} \frac{1}{4n} \int (v^i \pm v^j)_n d\mu = \frac{1}{4} \int (v^i \pm v^j)^2 d\mu + \frac{2}{4} \sum_{r=1}^{\infty} \int (v^i \pm v^j) (v^i \pm v^j) \circ T^r d\mu.
$$

Therefore, using the polarisation identity once more and taking $\lim_{n \to \infty}$ in (3.4) gives

$$
\lim_{n \to \infty} \frac{1}{n} \int v^i_n v^j_n d\mu
$$

$$
= \int v^i v^j d\mu + \frac{2}{4} \sum_{r=1}^{\infty} \int (v^i + v^j) (v^i + v^j) \circ T^r - (v^i - v^j) (v^i - v^j) \circ T^r d\mu
$$

$$
= \int v^i v^j d\mu + \sum_{r=1}^{\infty} \int v^i v^j \circ T^r + v^j v^i \circ T^r d\mu.
$$

\qed
Using the polarisation identity (4 for the scalar-valued observables $v$ with $\sigma$ note we have the martingale coboundary decompositions Proposition 3.2.13 now shows $\Sigma = c^T \Sigma c$). We show the coordinate wise statement. That is, for each $i, j = 1, \ldots, d$,

$$\lim_{n \to \infty} \frac{1}{n} \int v_n^i v_n^j d\mu = \int m^i m^j d\mu.$$ 

Using the polarisation identity ($4 \int fg = \|f + g\|_2^2 - \|f - g\|_2^2$) we see

$$\frac{4}{n} \int v_n^i v_n^j d\mu = \frac{1}{n} \int (v_n^i + v_n^j)^2 - (v_n^i - v_n^j)^2 d\mu = \frac{1}{n} \int (v^i + v^j)^2 - (v^i - v^j)^2_2 d\mu.$$

Note we have the martingale coboundary decompositions

$$v^i \pm v^j = (m^i \pm m^j) + (\chi^i \pm \chi^j) \cdot T - (\chi^i \pm \chi^j),$$

for the scalar-valued observables $v^i \pm v^j$. Since $v^i \pm v^j \in L^\infty(\Lambda)$ and $\sum_{k=1}^\infty \|P^k(v^i \pm v^j)\|_1 \leq \sum_{k=1}^\infty \|P^k v^i\|_1 + \|P^k v^j\|_1 < \infty$ we use the scalar case (Proposition 3.2.8) with $v$ replaced by $v^i \pm v^j$ to see

$$\lim_{n \to \infty} \frac{4}{n} \int v_n^i v_n^j d\mu = \lim_{n \to \infty} \frac{1}{n} \int (v^i + v^j)^2 - (v^i - v^j)^2 d\mu = \int (m^i + m^j)^2 - (m^i - m^j)^2 d\mu = 4 \int m^i m^j d\mu.$$

Proof of Theorem 3.2.12. By the Cramér-Wold device (Theorem 3.2.10) it suffices to show $c^T (n^{-1/2} v_n) = n^{-1/2} (c^T v)_n \to_d c^T N(0, \Sigma) = N(0, c^T \Sigma c)$ for any $c \in \mathbb{R}^d$.

Note that $c^T v \in L^\infty(\Lambda)$, $\int c^T v d\mu = 0$ and $\sum_{n=1}^\infty \|P^n (c^T v)\|_1 < \infty$. By the scalar CLT (Theorem 3.2.1), $n^{-1/2} (c^T v)_n \to_d N(0, \sigma^2)$ with $\sigma^2 = \lim_{n \to \infty} n^{-1} \int (c^T v)_n^2 d\mu$.

Next, we see $\sigma^2 = c^T \Sigma c$, since

$$c^T \Sigma c = c^T \left( \lim_{n \to \infty} \frac{1}{n} \int v_n v_n^T d\mu \right) c = \lim_{n \to \infty} \frac{1}{n} \int c^T (v_n v_n^T) c d\mu = \lim_{n \to \infty} \frac{1}{n} \int (c^T v)_n^2 d\mu = \sigma^2.$$

Proposition 3.2.13 now shows $\Sigma = \int v v^T d\mu + \sum_{r=1}^\infty \int v \int (v \circ T^r) + (v \circ T^r) v^T d\mu$.

For the degenerate case, first suppose $\det \Sigma = 0$. Then there exists a non-zero $c \in \mathbb{R}^d$ such that $\sigma^2 = \Sigma c = 0$ and hence $c^T \Sigma c = 0$. From the above calculation
we have $n^{-1/2}(c^T v) \to_d N(0, c^T \Sigma c)$ and so by the scalar CLT (Theorem 3.2.1), $c^T v$ is an $L^1(\Lambda)$ coboundary. Finally, suppose there exists a non-zero $c \in \mathbb{R}^d$ such that $c^T v$ is an $L^1(\Lambda)$ coboundary. Then $c^T \Sigma c = 0$ by the scalar CLT. Since $\Sigma$ is positive semi-definite we have $\Sigma = A^2$ for some symmetric matrix $A$. Then $0 = c^T \Sigma c = c^T A^T A c = \|Ac\|^2$ and so $Ac = 0$. Therefore $A = 0$ and $\Sigma = (\det A)^2 = 0$. □

### 3.3 Weak Invariance Principle

Here we consider vector-valued observables $v : \Lambda \to \mathbb{R}^d$. Define the process

$$W_n : [0, 1] \to \mathbb{R}^d, \quad W_n(t) := n^{-1/2}v[nt] = n^{-1/2} \sum_{j=0}^{[nt]-1} v \circ T^j,$$

(3.5)

with the convention that $\sum_{j=0}^{-1} v \circ T^j = 0$. Figure 3.2 illustrates a typical sample path of $W_n$ for $d = 1$.

Suppose $X_n : (\Lambda, \mathcal{A}, \mu) \to S, \ n \geq 1$, and $X : (\Omega, \mathcal{F}, \mathbb{P}) \to S$ are random elements of some metric space $S$. We say $X_n$ converges weakly to $X$, denoted $X_n \to_w X$, if $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ for every bounded, continuous function $f : S \to \mathbb{R}$ [7, Section 3].

One choice for $S$ is the Skorohod space $D([0, 1], \mathbb{R}^d)$, the space of $\mathbb{R}^d$-valued functions with left limits that are right continuous. We call a function $X \in D([0, 1], \mathbb{R}^d)$ a càdlàg process or càdlàg function. Any càdlàg process is bounded [7, Lemma 12.1], and we can endow the Skorohod space $D([0, 1], \mathbb{R}^d)$ with the supremum norm $\|\cdot\|_{\infty}$ to obtain a non-separable Banach space [7, Section 15]. Clearly $C([0, 1], \mathbb{R}^d) \subset D([0, 1], \mathbb{R}^d)$. We have $W_n \in D([0, 1], \mathbb{R}^d)$. An important fact that we have already used is the following:

**Proposition 3.3.1** ([7, Theorem 3.1]). Suppose $X_n, Y_n : S \to \mathbb{R}, \ n \geq 1$, and $X, Y : S \to \mathbb{R}$ are random elements of a separable metric space $S$. Assume that $X_n \to_w X$, $Y_n \to_w Y$ and that $Y$ is constant almost surely. Then $X_n + Y_n \to_w X + Y$.

Since $(D([0, 1], \mathbb{R}^d), \|\cdot\|_{\infty})$ is not separable, we require a workaround to use this proposition: We can endow $D([0, 1], \mathbb{R}^d)$ with a weaker topology, the Skorohod topology $\mathcal{J}_1$. The Skorohod space $(D([0, 1], \mathbb{R}^d), \mathcal{J}_1)$ is separable and complete under this topology [7, Theorem 12.2].

**Proposition 3.3.2** ([7, Section 15]). If $X_n \to_w X$ in $(D([0, 1], \mathbb{R}^d), \|\cdot\|_{\infty})$, then convergence also holds in $(D([0, 1], \mathbb{R}^d), \mathcal{J}_1)$. If $X_n \to_w X$ in $(D([0, 1], \mathbb{R}^d), \mathcal{J}_1)$ and $X \in C([0, 1], \mathbb{R}^d)$, then $X_n \to_w X$ in $(D([0, 1], \mathbb{R}^d), \|\cdot\|_{\infty})$. 

24
Remark. This proposition allows us to work in the $J_1$ topology when required, and ‘upgrade’ convergence to the $\| \cdot \|_\infty$ topology.

We call $W : (\Omega, \mathcal{F}, \mathbb{P}) \to C([0, 1], \mathbb{R}^d)$ a $d$ dimensional Brownian motion with covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ if $W(0) = 0$ and

1. (Independent increments) The increments $W(t_1) - W(t_0), \ldots, W(t_k) - W(t_{k-1})$ are independent for all $0 = t_0 \leq t_1 \leq \cdots \leq t_k \leq 1$.

2. (Normal increments) For all $0 \leq s \leq t \leq 1$ we have $W(t) - W(s) =_d \mathcal{N}(0, \Sigma(t - s))$.

We say a sequence of random functions $Z_n$ satisfies a weak invariance principle (WIP) if $Z_n \to_w W$ in $(C([0, 1], \mathbb{R}^d), \| \cdot \|_\infty)$ or $(D([0, 1], \mathbb{R}^d), \| \cdot \|_\infty)$.

**Theorem 3.3.3 (WIP).** We have the convergence

$W_n \to_w W$

in $(D([0, 1], \mathbb{R}^d), \| \cdot \|_\infty)$, where $W$ is $d$ dimensional Brownian motion with mean zero and covariance $\Sigma = \int v v^T d\mu + \sum_{r=1}^{\infty} \int v \left( v \circ T^r \right)^T + \left( v \circ T^r \right) v^T d\mu$.

Remark 1. This was first proven by Dedecker and Rio [11, Theorem 1] under similar hypotheses: They work with invertible maps $T : (\Lambda, \mathcal{A}, \mu) \to (\Lambda, \mathcal{A}, \mu)$ and assume
existence of a sub-sigma algebra \( A_0 \subset A \) such that \( T^{-1}A_0 \supset A_0 \). Their observable \( v : \Lambda \to \mathbb{R} \) is taken to be square integrable and \( A_0 \)-measurable. The measurability assumption is an important one, and is what causes their result to be analogous to the non-invertible case. They also assume the mixing condition

\[
\sum_{n=0}^{\infty} v \mathbb{E}(v \circ T^{-n} | A_0) \text{ converges in } L^1(\Lambda).
\]

For more details, see the discussion at the beginning of Chapter 4 (note that the first term in Assumption (A\text{	ext{inv}}) dissapears when \( v \) is \( A_0 \)-measurable. Also, Dedecker & Rio have time running backwards with respect to our convention).

**Remark 2.** As in Theorem 3.2.12, det \( \Sigma = 0 \) if and only if there exists a non-zero \( c \in \mathbb{R}^d \) such that \( c^Tv \) is an \( L^1(\Lambda) \) coboundary.

**Remark 3.** We could equally define our processes in \( D([0,K], \mathbb{R}^d) \) for any integer \( K \geq 1 \), and the result still holds. Hence the result is true if we define our processes in \( D([0,\infty), \mathbb{R}^d) \) by [7, Section 16].

**Remark 4.** Zweimüller [54, Corollary 3] proved that the WIP also holds with respect to every probability measure \( \nu \) which is absolutely continuous with respect to \( \mu \).

**Example 3.3.4.** For \( 0 < \alpha < 1 \), let \( T : [0,1] \to [0,1] \) be the intermittent map defined by equation (3.1), introduced in Example 3.2.2. Let \( v : [0,1] \to \mathbb{R} \) be a mean zero Hölder observable. Recall \( \|P_n v\|_1 \leq Cn^{-1/\alpha+1} \) and so Theorem 3.3.3 gives weak convergence of the process \( W_n \) to Brownian motion \( W \) in \( (D([0,1], \mathbb{R}), \| \cdot \|_\infty) \) for \( 0 < \alpha < 1/2 \).

For the above example (Example 3.3.4), the WIP can fail if the correlations are not summable: If \( 1/2 < \alpha < 1 \) and \( v : \Lambda \to \mathbb{R} \) is a mean-zero observable such that \( v(0) \neq 0 \), then Melbourne and Zweimüller [41, Theorem 1.1] proved that the process \( n^{-\alpha}v_{[nt]} \) converges to an \( \alpha \)-stable Lévy process [27, Pg. 291–292] in some suitable topology over \( D([0,1], \mathbb{R}) \).

To prove Theorem 3.3.3, we follow a similar procedure as in Section 3.2. First, we setup and state the martingale WIP (Proposition 3.3.5). Next, we show that the lifted backwards process satisfies the martingale WIP. Finally, we prove that the lifted backwards process has the same weak limit as \( W_n \).

We call a sequence of sigma algebras \( \{A_t | 0 \leq t \leq 1 \} \) a \text{filtration} if for every \( 0 \leq s \leq t \leq 1 \) we have \( A_s \subseteq A_t \). A continuous time stochastic process \( \{X(t) : \Omega \to \mathbb{R}^d | 0 \leq t \leq 1 \} \) is said to be a \text{adapted} to a filtration \( \mathcal{F}_t \) if \( X_t \) is \( \mathcal{F}_t \)-measurable for every \( t \in [0,1] \). We say \( X_t \) is a \text{martingale} [15, Section 2.2] with respect to the filtration \( \{A_t | 0 \leq t \leq 1 \} \) if for all \( 0 \leq s \leq t \leq 1 \) we have
1. \(X(t)\) is \(\mathcal{A}_t\)-measurable
2. \(\|X(t)\|_1 < \infty\)
3. \(\mathbb{E}(X(t) \mid \mathcal{A}_s) = X(s)\) a.s.

For \(X \in D([0,1],\mathbb{R})\), define \(X(t-) := \lim_{s \uparrow t} X(s)\). Let

\[
J(X, t) := \sup \left\{ |X(s) - X(s-) : 0 < s \leq t \right\}
\]

be the largest jump up to time \(t\). Let \(t \in [0,1]\) and suppose \(\Delta \ell\) is a sequence of refining partitions of \([0,t]\) such that \(\max_{\ell} |t_{k+1} - t_k| \to 0\) as \(\ell \to \infty\). For \(Y \in D([0,1],\mathbb{R})\), define the quadratic covariation between \(X\) and \(Y\) as

\[
[X, Y](t) := \lim_{\ell \to \infty} \sum_{k=0}^{\ell-1} (X(t_{k+1}) - X(t_k))(Y(t_{k+1}) - Y(t_k)),
\]

where the limit is taken in probability, whenever it exists. Let \(\Delta X(t) := X(t) - X(t-\).

If \(X\) and \(Y\) are piecewise constant, then

\[
[X, Y](t) = X(0)Y(0) + \sum_{0 \leq s \leq t} \Delta X(s)\Delta Y(s). \quad (3.6)
\]

The statement of the following proposition has been taken from [51, Theorem 2.1], which is an adaptation of [15, Section 7.1].

**Proposition 3.3.5** (Martingale WIP [51, Theorem 2.1]). For each \(n \geq 1\), let \(F_n \in D([0,1],\mathbb{R}^d)\), and suppose \(F_n := (F_n^1, \ldots, F_n^d)\) is a martingale with respect to the filtration \(\{\mathcal{A}_{n,t} \mid 0 \leq t \leq 1\}\) such that \(F_n(0) = 0\). Let \(\Sigma\) be a \(d \times d\) symmetric and positive semi-definite matrix of real numbers. Assume

(i) for each \(0 \leq t \leq 1\), we have \(\lim_{n \to \infty} \mathbb{E}(J(F_n, t)) = 0\)

(ii) for each coordinate \(\beta, \gamma = 1, \ldots, d\) and every \(0 \leq t \leq 1\) we have \([F_n^\beta, F_n^\gamma](t) \to_d t\Sigma_{\beta\gamma}\)

Then

\(F_n \to_w W\)

in \((D([0,1],\mathbb{R}^d), \|\cdot\|_\infty)\), where \(W\) is a \(d\) dimensional Brownian motion with mean zero and covariance \(\Sigma\).

By Proposition 3.2.4, lift the probability space \((\Lambda, \mathcal{A}, \mu)\) and map \(T\) to obtain a probability space \((\tilde{\Lambda}, \tilde{\mathcal{A}}, \tilde{\mu})\), where \(\tilde{\mu}\) is an ergodic, invariant measure with respect to
Proposition 3.3.6. The quadratic variation of \( R \) and \( \sum \) respectively, and follows from (3.6).

Define the backwards lifted processes \( \tilde{\pi} \) and the backwards lifted process \( \tilde{\beta}, \gamma \) for each \( m \). \( \tilde{\pi} \) satisfies

\[
\tilde{\pi} \circ \tilde{\beta} = \tilde{\pi} \circ \tilde{\gamma},
\]

Note that the jumps of \( \tilde{\pi} \) are \( 1 \).

Proof. \( \tilde{\pi} \) is \( \tilde{\pi} \)-measurable, note that for any Borel subset \( A \) of \( \Lambda \),

\[
\tilde{\pi} \circ \tilde{\beta} \circ (\tilde{\pi} \circ \tilde{\gamma}) = \tilde{\pi} \circ \tilde{\beta} \circ \tilde{\gamma}.
\]

Moreover, each jump is given by \( \Delta \tilde{\pi} \circ \tilde{\beta} \circ \tilde{\gamma} = 1 \).

Proposition 3.3.7. Suppose \( f \in L^1(\Lambda, \mathbb{R}^d) \) is such that \( f \in \text{Ker} P \), and define \( \tilde{f} := f \circ \pi \). Then \( \tilde{f} \circ \tilde{\pi} \circ \tilde{\beta} \circ \tilde{\gamma} \) is a martingale with respect to the filtration \( \mathcal{A}_{n,t} = \{ \tilde{T}^{[nt]} \pi^{-1} A \mid 0 \leq t \leq 1 \} \).

Proof. Let \( n \geq 1 \). Firstly, \( \tilde{f} \circ \tilde{\pi} \circ \tilde{\beta} \circ \tilde{\gamma} \in L^1(\Lambda, \mathbb{R}^d) \) since

\[
\| \tilde{f} \circ \tilde{\pi} \circ \tilde{\beta} \circ \tilde{\gamma} \|_1 = \left\| \sum_{j=1}^{[nt]} \tilde{f} \circ \tilde{\pi} \circ \tilde{\beta} \circ \tilde{\gamma} \right\|_1 < \infty.
\]

To see \( \tilde{f} \circ \tilde{\pi} \circ \tilde{\beta} \circ \tilde{\gamma} \) is \( \mathcal{A}_{n,t} \)-measurable, note that for any Borel subset \( E \) of \( \mathbb{R}^d \) and all \( s \leq t \) we have

\[
((f \circ \pi) \circ \tilde{\beta} \circ \tilde{\gamma})^{-1} E = \tilde{T}^{[nt]} \pi^{-1} (f^{-1} E) \in \mathcal{A}_{n,s} \subseteq \mathcal{A}_{n,t}.
\]
It remains to show
\[ \mathbb{E}(\tilde{f}_n^{-}(t) \mid A_{n,s}) = \tilde{f}_n^{-}(s), \]
for all \( s \leq t \). If \([nt] = [ns]\), then \( \tilde{f}_n^{-}(t) = \tilde{f}_n^{-}(s) \) and we are done. Suppose \([nt] > [ns]\) and note that
\[
\mathbb{E}\left( \sum_{j=1}^{[nt]} \tilde{f} \circ \tilde{T}^{-j} \mid A_{n,s} \right) = \tilde{f}_n^{-}(s) + \sum_{j=[ns]+1}^{[nt]} \mathbb{E}(\tilde{f} \circ \tilde{T}^{-j} \mid \tilde{T}^{[ns]} \pi^{-1} A).
\]

Let \( j = [ns]+1, \ldots, [nt] \) be fixed, and note it suffices to show \( \mathbb{E}(\tilde{f} \circ \tilde{T}^{-j} \mid \tilde{T}^{[ns]} \pi^{-1} A) = 0 \).
Write \( \tilde{T}^{[ns]} \pi^{-1} A = \tilde{T}^{j-k} \pi^{-1} A \), where \( k = j - [ns] > 0 \). Since \( f \in \text{Ker} P \),
\[
\mathbb{E}(\tilde{f} \circ \tilde{T}^{-j} \mid \tilde{T}^{j-k} \pi^{-1} A) = \mathbb{E}(\tilde{f} \mid \tilde{T}^{j-k} \pi^{-1} A) \circ \tilde{T}^{-j} = \mathbb{E}(f \circ \pi \mid \pi^{-1} T^{-k} A) \circ \tilde{T}^{-j}
= \mathbb{E}(f \mid T^{-k} A) \circ \pi \circ \tilde{T}^{-j} = (P^k f) \circ T^k \circ \pi \circ \tilde{T}^{-j} = 0.
\]

\[ \square \]

**Proposition 3.3.8** (Martingale WIP). We have
\[
\tilde{M}_n^{-} \to_w W
\]
in \( (D([0,1],\mathbb{R}^d), \| \cdot \|_\infty) \), where \( W \) is a \( d \) dimensional Brownian motion with mean zero and covariance \( \Sigma = \int m m^T d\mu \).

**Proof.** By Proposition 3.3.7, \( \tilde{M}_n^{-} \) is a martingale. Since \( \tilde{M}_n^{-}(0) = 0 \), it remains to verify conditions (i) and (ii) of Proposition 3.3.5. We begin by noting that for \( 0 \leq t \leq 1 \) we have
\[
J(\tilde{M}_n^{-}, t) = n^{-1/2} \max_{1 \leq j \leq [nt]} |\tilde{m} \circ \tilde{T}^{-j}| \leq n^{-1/2} \max_{1 \leq j \leq n} |\tilde{m} \circ \tilde{T}^{-j}|
\]
For condition (i), we consider the square of the expected value of the jumps:
\[
\left[ \mathbb{E}(J(\tilde{M}_n^{-}, t)) \right]^2 = \|J(\tilde{M}_n^{-}, t)\|_2^2 \leq \|J(\tilde{M}_n^{-}, t)\|_2^2 \leq \int \max_{1 \leq j \leq n} n^{-1} \tilde{m}^2 \circ \tilde{T}^{-j} d\mu. \tag{3.7}
\]
We use the fact that for any \( a > 0 \) we have
\[
n^{-1} \tilde{m}^2 \circ \tilde{T}^{-j} \leq a^2 + n^{-1} \tilde{m}^2 \circ \tilde{T}^{-j} \mathbb{1}\{n^{-1} \tilde{m}^2 \circ \tilde{T}^{-j} \geq a^2\}. \tag{3.8}
\]
Take the maximum over \( j = 1, \ldots, n \) and integrate in (3.8) to obtain
\[
\int \max_{1 \leq j \leq n} \tilde{m}^2 \circ \tilde{T}^{-j} \, d\tilde{\mu} \leq a^2 + n^{-1} \int \max_{1 \leq j \leq n} \tilde{m}^2 \circ \tilde{T}^{-j} \mathbb{1}\{\tilde{m}^2 \circ \tilde{T}^{-j} \geq na^2\} \, d\tilde{\mu}
\]
\[
\leq a^2 + n^{-1} \sum_{j=1}^{n} \tilde{m}^2 \circ \tilde{T}^{-j} \mathbb{1}\{\tilde{m}^2 \circ \tilde{T}^{-j} \geq na^2\} \, d\tilde{\mu}
\]
\[
= a^2 + \int \tilde{m}^2 \mathbb{1}\{\tilde{m}^2 \geq na^2\} \, d\tilde{\mu}.
\]
(3.9)

Combining (3.7) and (3.9), and taking \( \limsup_{n \to \infty} \) shows
\[
\limsup_{n \to \infty} \left[ \mathbb{E}(J(\tilde{\mu}_n, t)) \right]^2 \leq a^2.
\]

Since \( a > 0 \) was arbitrary we see condition (i) is satisfied.

Condition (ii) follows from Proposition 3.3.6 and the ergodic theorem, since
\[
[\tilde{M}_{n\beta}, \tilde{M}_{n\gamma}] (t) = \frac{1}{nt} \sum_{j=1}^{[nt]} (\tilde{m}^\beta \tilde{m}^\gamma) \circ \tilde{T}^{-j} \to_{a.s.} t \int m^\beta m^\gamma \, d\mu = t \Sigma_{\beta\gamma}.
\]

Therefore Proposition 3.3.5 gives the convergence \( \tilde{M}_n \to_w W \) in \( (D([0,1], \mathbb{R}^d), \| \cdot \|_\infty) \),
where \( \Sigma = \int \tilde{m} \tilde{m}^T \, d\mu = \int m m^T \, d\mu. \)

For a sequence \( S_n = \sum_{j=1}^{n} X_j \) of random vectors \( X_j \) define \( \overline{S}_n = \max_{1 \leq \ell \leq n} |S_\ell| \).
Note that
\[
\overline{v_n} = \max_{1 \leq \ell \leq n} |v_\ell| = \max_{1 \leq \ell \leq n} \left| \sum_{j=1}^{\ell} v \circ T^j \right|.
\]

To show \( \tilde{M}_n \) has the same weak limit as \( W_n \), we will need the two following maximal inequalities.

**Proposition 3.3.9** (Doob’s inequality [21, Theorem 2.2]). Let \( S_n, \ n \geq 1 \) be a sequence of square integrable random variables that form a martingale. Then
\[
\|S_n\|_2 \leq \|\overline{S}_n\|_2 \leq 2\|S_n\|_2.
\]

**Proposition 3.3.10** (Rio’s inequality [48]). Let \( X_j \) be a sequence of square integrable random variables adapted to some filtration \( \mathcal{F}_j \), for \( j \geq 1 \), and define
\( S_n := \sum_{j=1}^n X_j \). Then
\[
\mathbb{E}(\mathbf{S}_n^2) \leq 16 \sum_{j=1}^n \max_{j \leq t \leq n} b_{j,t},
\]
where \( b_{j,t} = \left\| X_{j} \sum_{u=j}^t \mathbb{E}(X_u | F_j) \right\|_1 \).
\[\square\]

**Proposition 3.3.11.** We have that \( \tilde{W}_n^- \) has the same weak limit as \( \tilde{M}_n^- \), so \( \tilde{W}_n^- \to_w W \) in \( (D([0,1], \mathbb{R}^d), \| \cdot \|_\infty) \), where \( W \) is a \( d \) dimensional Brownian motion with mean zero and covariance \( \Sigma = \int m m^T \, d\mu \).

**Proof.** Proposition 3.3.8 shows \( \tilde{M}_n^- \to_w W \), where \( W \) is a \( d \) dimensional Brownian motion with mean zero and covariance \( \Sigma = \int m m^T \, d\mu \). Therefore, it suffices to show
\[
\sup_{0 \leq t \leq 1} |\tilde{M}_n^-(t) - \tilde{W}_n^-(t)| \leq \sum_{\beta=1}^d \sup_t |\tilde{M}_n^{\beta,2} - \tilde{W}_n^{\beta,2} - (t)| \to_p 0.
\]
Without loss we assume \( \tilde{W}_n^-, \tilde{M}_n^- \) are scalar-valued processes and show
\[
\sup_{0 \leq t \leq 1} |\tilde{M}_n^-(t) - \tilde{W}_n^-(t)| \to_p 0.
\]
Let \( \epsilon > 0 \) and choose \( k \) large enough so that \( \|m - m^{(k)}\|_2 = \|\hat{m} - \hat{m}^{(k)}\|_2 < \epsilon \) and \( \sum_{j=k} \|P^j v\|_1 < \epsilon^2 \). Note that
\[
\left\| \sup_{0 \leq t \leq 1} |\tilde{M}_n^-(t) - \tilde{W}_n^-(t)| \right\|_2 \leq \left\| \sup_t |\tilde{M}_n^-(t) - \tilde{M}_n^{(k)} - (t)| \right\|_2 + \left\| \sup_t |\tilde{W}_n^-(t) - \tilde{M}_n^{(k)} - (t)| \right\|_2.
\]
Since \( \tilde{W}_n^-(t), \tilde{M}_n^-(t), \tilde{M}_n^{(k)} - (t) \) are constant on each interval \([\frac{j}{n}, \frac{j+1}{n}]\) for \( j = 0, \ldots, n-1 \), the supremum is attained at one of the points \( t = 0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1 \). Hence \( \sup_{0 \leq t \leq 1} |\sum_{j=1}^n \tilde{v} \circ \tilde{T}^{-j}| = \max_{1 \leq j \leq n} |\sum_{t=1}^j \tilde{v} \circ \tilde{T}^{-j}| \) and
\[
\left\| \sup_{0 \leq t \leq 1} |\tilde{M}_n^-(t) - \tilde{W}_n^-(t)| \right\|_2 \leq n^{-1/2} \left\| (\hat{m} - \hat{m}^{(k)}) \right\|_2 + n^{-1/2} \left\| (\hat{v} - \hat{m}^{(k)}) \right\|_2.
\]
(3.10)

We deal with the two terms in (3.10) separately. For the first term, note that \( (\hat{m} - \hat{m}^{(k)}) \circ \tilde{T}^{-j} \) is a martingale difference sequence by Proposition 3.3.7. Hence apply Doob’s inequality (Proposition 3.3.9) and use the orthogonality of \( (\hat{m} - \hat{m}^{(k)}) \circ \tilde{T}^{-j} \)
(Proposition 3.1.3) to calculate that

\[
\begin{align*}
    n^{-1/2} \left\| \frac{1}{n} \sum_{j=1}^{n} (\tilde{m} - \tilde{m}(k)) \circ T^{-j} \right\|_2 & \leq 2n^{-1/2} \left\| \sum_{j=1}^{n} (\tilde{m} - \tilde{m}(k)) \circ T^{-j} \right\|_2 \\
    & = 2\|\tilde{m} - \tilde{m}(k)\|_2 < 2\epsilon.
\end{align*}
\]

(3.11)

Next we consider the second term of (3.10). Use the martingale coboundary decomposition \( v = m(k) + \chi_1^k \circ T - \chi_1 \circ P^k v \) to calculate that

\[
\begin{align*}
    n^{-1/2} \left\| \max_{1 \leq \ell \leq n} \left\| \sum_{j=1}^{\ell} (\bar{v} - \tilde{m}(k)) \circ T^{-j} \right\|_2 \right\|_2 & = n^{-1/2} \left\| \max_{1 \leq \ell \leq n} \left\| \sum_{j=1}^{\ell} (\chi_1^k \circ T - \chi_1 \circ P^k v) \circ T^{-j} \right\|_2 \right\|_2 \\
    & = \max_{1 \leq \ell \leq n} \left\| \sum_{j=1}^{\ell} (\chi_1^k \circ \pi \circ T - \chi_1 \circ \pi \circ P^k v \circ \pi) \circ T^{-j} \right\|_2 \\
    & \leq 2n^{-1/2}\|\chi_1^k\|_\infty + n^{-1/2} \left\| \sum_{j=1}^{\ell} (P^k v) \circ \pi \circ T^{-j} \right\|_2 \\
    & \leq 2n^{-1/2}\|v\|_\infty + n^{-1/2} \left\| \sum_{j=1}^{\ell} (P^k v) \circ \pi \circ T^{-j} \right\|_2.
\end{align*}
\]

(3.12)

Next apply Rio’s inequality (Proposition 3.3.10) to the square of the second term in (3.12) to see

\[
\begin{align*}
    n^{-1} \left\| \max_{1 \leq \ell \leq n} \left\| \sum_{j=1}^{\ell} (P^k v) \circ \pi \circ T^{-j} \right\|_2 \right\|_2^2 & = n^{-1} \mathbb{E} \left( \max_{1 \leq \ell \leq n} \left\| \sum_{j=1}^{\ell} (P^k v) \circ \pi \circ T^{-j} \right\|_2^2 \right) \\
    & \leq 16n^{-1} \sum_{j=1}^{n} \max_{1 \leq \ell \leq n} b_{j,\ell},
\end{align*}
\]

(3.13)

where

\[
b_{j,\ell} = \left\| (P^k v) \circ \pi \circ T^{-j} \sum_{u=j}^{\ell} \mathbb{E}( (P^k v) \circ \pi \circ T^{-u} | A_j ) \right\|_1.
\]
Note that the observable $P^k v$ has a martingale coboundary decomposition

$$P^k v = r^{(k)} + \left( \sum_{n=1}^{\infty} P^n P^k v \right) \circ T - \sum_{n=1}^{\infty} P^n P^k v = r^{(k)} + \chi_{k+1}^\infty \circ T - \chi_{k+1}^\infty,$$

with $r^{(k)} \in \text{Ker} \ P$, $r^{(k)} \in L^2(\Lambda, \mathbb{R}^d)$ and $\chi_{k+1}^\infty \in L^1(\Lambda, \mathbb{R}^d)$. The proofs of these facts are identical to the ones proven for $m$ in Theorem 3.1.2. Moreover, Proposition 3.3.7 shows $\tilde{r}^{(k)} \circ \tilde{T}^{-j} := r^{(k)} \circ \pi \circ \tilde{T}^{-j}$ forms a martingale difference sequence. That is, $\mathbb{E}(\tilde{r}^{(k)} \circ \tilde{T}^{-u} | \mathcal{A}_j) = 0$ for every $u > j$, and $\mathbb{E}(\tilde{r}^{(k)} \circ \tilde{T}^{-j} | \mathcal{A}_j) = \tilde{r}^{(k)} \circ \tilde{T}^{-j}$. Calculate that

$$b_{j, \ell} \leq \| v \|_\infty \left\| \sum_{u=j}^{\ell} \mathbb{E} \left( (r^{(k)} + \chi_{k+1}^\infty \circ T - \chi_{k+1}^\infty) \circ \pi \circ \tilde{T}^{-u} \big| \mathcal{A}_j \right) \right\|_1$$

$$= \| v \|_\infty \left\| r^{(k)} \circ \pi \circ \tilde{T}^{-j} + \mathbb{E} \left( \sum_{u=j}^{\ell} \chi_{k+1}^\infty \circ \pi \circ \tilde{T}^{-u+1} - \chi_{k+1}^\infty \circ \pi \circ \tilde{T}^{-u} \big| \mathcal{A}_j \right) \right\|_1$$

$$\leq \| v \|_\infty \left( \| r^{(k)} \|_1 + 2 \| \chi_{k+1}^\infty \|_1 \right).$$

Returning to (3.13) we have

$$n^{-1} \left\| \max_{1 \leq \ell \leq n} \left( \sum_{j=1}^{\ell} (P^k v) \circ \pi \circ \tilde{T}^{-j} \right) \right\|_2^2 \ll n^{-1} \sum_{j=1}^{n} \max_{j \leq \ell \leq n} \left( \| r^{(k)} \|_1 + 2 \| \chi_{k+1}^\infty \|_1 \right)$$

$$= \| P^k v + \chi_{k+1}^\infty \circ T - \chi_{k+1}^\infty \|_1 + 2 \| \chi_{k+1}^\infty \|_1 \leq \| P^k v \|_1 + 4 \| \chi_{k+1}^\infty \|_1 \leq 4 \sum_{n=k}^{\infty} \| P^n v \|_1 \ll \epsilon^2.$$

Therefore (3.12) becomes

$$n^{-1/2} \left\| (\bar{v} - \bar{m}(k))_n \right\|_2 \ll n^{-1/2} k + \epsilon.$$

Combining this and (3.11) with (3.10), we see that

$$\limsup_{n \to \infty} \left\| \sup_{0 \leq t \leq 1} |\tilde{W}_n^- (t) - \tilde{M}_n^- (t)| \right\|_2 \ll \limsup_{n \to \infty} \left( 2 \epsilon + n^{-1/2} k + \epsilon \right) = 3 \epsilon.$$

Let $D([0, 1], \mathbb{R}^d)$ denote the space of $\mathbb{R}^d$-valued functions that are left continuous with right limits. This is the space of càdlàg (not càdàg) functions. This space and the following Proposition is needed because of our time reversal.
Proposition 3.3.12 ([29, Proposition 4.9]). Suppose that \( A_n = B_n + F_n \), where \( A_n \in D([0,1],\mathbb{R}^d) \), \( B_n \in \hat{D}([0,1],\mathbb{R}^d) \) and \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \). If \( B_n \to_w Z \) in \( (\hat{D}([0,1],\mathbb{R}^d), \mathcal{J}_1) \) and \( Z \) has continuous sample paths, then \( A_n \to_w Z \) in \( (D([0,1],\mathbb{R}^d), \| \cdot \|_\infty) \). \( \square \)

The time reversal argument in the following proof is borrowed heavily from [29].

Proof of Theorem 3.3.3. Defining the map

\[
g : (D([0,1],\mathbb{R}^d), \| \cdot \|_\infty) \to (\hat{D}([0,1],\mathbb{R}^d), \| \cdot \|_\infty), \quad g(u)(t) := u(1) - u(1-t),
\]
we see \( g \) is continuous since \( \|g(u) - g(v)\|_\infty \leq 2\|u - v\|_\infty \). Therefore Proposition 3.3.11 and the continuous mapping theorem (Proposition 3.2.9) shows \( g(\hat{W}_n^-) \to_w g(W) \) in \( (\hat{D}([0,1],\mathbb{R}^d), \mathcal{J}_1) \), where \( W \) is a \( d \) dimensional Brownian motion with mean zero and covariance \( \Sigma = \int m m^T \, d\mu \). By Propositions 3.2.13 and 3.2.14, \( \Sigma = \int v v^T \, d\mu + \sum_{j=1}^\infty \int v (v \circ T^j)^T + (v \circ T^j) v^T \, d\mu \). The result follows from Proposition 3.3.12 and the following two claims:

1. \( \hat{W}_n(t) \circ \bar{T}^{-n} = g(\hat{W}_n^-)(t) + F_n(t) \), where \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \)

2. \( g(W) =_d W \)

Assume the claims are true and denote \( A_n = \hat{W}_n(t) \circ \bar{T}^{-n} \), \( B_n = g(\hat{W}_n^-)(t) \). We have \( B_n \to_w g(W) =_d W \) in \( (\hat{D}([0,1],\mathbb{R}^d), \mathcal{J}_1) \), and \( \sup_t |F_n(t)| \to_p 0 \). Since \( W \) has continuous sample paths we see by Proposition 3.3.12 that \( W_n =_d \hat{W}_n =_d W_n \circ \bar{T}^{-n} \to_w W \) in \( (D([0,1],\mathbb{R}^d), \| \cdot \|_\infty) \). It remains to verify the two claims.

To prove the first claim, consider

\[
\hat{W}_n(t) \circ \bar{T}^{-n} = n^{-1/2} \sum_{j=0}^{[nt]-1} \bar{v} \circ \bar{T}^j \circ \bar{T}^{-n} = n^{-1/2} \sum_{j=-n}^{[nt]-1-n} \bar{v} \circ \bar{T}^j = n^{-1/2} \sum_{j=-[nt]}^{n-[nt]+1} \bar{v} \circ \bar{T}^{-j}
\]

\[
= n^{-1/2} \sum_{j=1}^{n-[nt]} \bar{v} \circ \bar{T}^{-j} - n^{-1/2} \sum_{j=1}^{n-[nt]} \bar{v} \circ \bar{T}^{-j} = \hat{W}_n^- (1) - n^{-1/2} \sum_{j=1}^{n-[nt]} \bar{v} \circ \bar{T}^{-j}.
\]
Define $F_n(t) := n^{-1/2}(\sum_{j=1}^{[n-nt]} - \sum_{j=1}^{n-[nt]} )\tilde{v} \circ \tilde{T}^{-j}$. Then

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

$$= \tilde{W}_n^{-1}(1) + n^{-1/2} \sum_{j=1}^{[n-nt]} \tilde{v} \circ \tilde{T}^{-j} + F_n(t)$$

$$= \tilde{W}_n^{-1}(1) - \tilde{W}_n^{-1}(1-t) + F_n(t) = g(\tilde{W}_n^{-1})(t) + F_n(t).$$

If $nt$ is an integer, then $F_n(t) = 0$. Otherwise $[n-nt] = n - [nt] - 1$ and $F_n(t) = -n^{-1/2}\tilde{v} \circ \tilde{T}^{[nt]-n}$. It follows from Proposition 2.2.3 that $\sup_{0 \leq t \leq 1} |F_n(t)| = n^{-1/2} \max_{0\leq j \leq n} |\tilde{v} \circ \tilde{T}^{-j}| \rightarrow_{p} 0$.

Finally, for the second claim, we show $g(W)(t) = W(1) - W(1-t)$ is a Brownian motion with mean zero and covariance $\Sigma$. Clearly $g(W)(0) = 0$. Next, $g(W)$ has continuous sample paths since $W$ has continuous sample paths and $g$ is a continuous function. Let $0 \leq t_0 \leq \cdots \leq t_k \leq 1$. Independent increments of $g(W)$ follows from independent increments of $W$, since

$$(g(W)(t_1) - g(W)(t_0), \ldots, g(W)(t_k) - g(W)(t_{k-1})) = (W(1 - t_0) - W(1 - t_1), \ldots, W(1 - t_{k-1}) - W(1 - t_k)).$$

Finally, note that for $0 \leq s \leq t \leq 1$ we have

$$g(W)(t) - g(W)(s) = W(1 - s) - W(1 - t) =_{d} N(0, \Sigma(t-s)).$$

Therefore $g(W) =_{d} W$. 

\[\square\]

### 3.4 Iterated Weak Invariance Principle

Recall the standing assumptions that $T : \Lambda \rightarrow \Lambda$ is an ergodic, measure-preserving map over the probability space $(\Lambda, \mathcal{A}, \mu)$, and our observable $v \in L^\infty(\Lambda, \mathbb{R}^d)$ is mean zero and satisfies the mixing assumption $\sum_{n=1}^{\infty} \|P^n v^i\|_1 < \infty$ for every $i = 1, \ldots, d$. 

35
For coordinates $\beta, \gamma = 1, \ldots, d$, define the process
\[
\mathcal{W}_n : [0, 1] \to \mathbb{R}^{d \times d}, \quad \mathcal{W}_n^{\beta \gamma}(t) := n^{-1} \sum_{j=0}^{\lfloor nt \rfloor - 1} \sum_{i=0}^{j-1} v^\beta \circ T^i v^\gamma \circ T^j.
\]

**Definition 3.4.1** (Itô integral, [31]). Let $X, Y \in D([0, 1], \mathbb{R})$. For $0 < t \leq 1$, let $\Delta_\ell$ be a sequence of refining partitions of $[0, t]$ such that $\max_{k} |t_{k+1} - t_k| \to 0$.

We define the **Itô integral** to be the following limit of Riemann-Stieltjes like sums, whenever it exists:
\[
\int_0^t X dY = \lim_{\ell \to \infty} \sum_{k=0}^{\ell-1} X(t_k) (Y(t_{k+1}) - Y(t_k)),
\]
where the limit is taken in probability.

**Remark.** If $Y$ has paths of finite variation on compact subsets of $[0, 1]$, then the Itô integral coincides with the Lebesgue-Stieltjes integral almost surely [46, Theorem 2.17].

Suppose $X$ and $Y$ are $n$ dimensional vector-valued processes such that $\int X^\beta dY^\gamma$ exists, for every $\beta, \gamma = 1, \ldots, n$. Define the $n \times n$ dimensional matrix-valued process $\int X \otimes dY$ as
\[
\left( \int X \otimes dY \right)^{\beta \gamma} := \int X^\beta dY^\gamma,
\]
for $\beta, \gamma = 1, \ldots, n$.

**Theorem 3.4.2** (Iterated WIP). We have that
\[
(W_n, \mathcal{W}_n) \to_w (W, \mathcal{W})
\]
in $D([0, 1], \mathbb{R}^d \times \mathbb{R}^{d \times d}, \| \cdot \|_\infty)$, where $W$ is a $d$ dimensional Brownian motion with mean zero and covariance $\Sigma = \int v v^T d\mu + \sum_{r=1}^{\infty} \int (v \circ T^r)^T (v \circ T^r) v^T d\mu$, and
\[
\mathcal{W}(t) = \int_0^t W \otimes dW + t \sum_{r=1}^{\infty} \int v (v \circ T^r)^T d\mu,
\]
for $t \in [0, 1]$.

**Remark 1.** As in Theorem 3.2.12, $\det \Sigma = 0$ if and only if there exists a non-zero $c \in \mathbb{R}^d$ such that $c^T v$ is an $L^1(\Lambda)$ coboundary.

**Remark 2.** We could equally define our processes in $D([0, K], \mathbb{R}^d)$ for any integer.
\( K \geq 1 \), and the result still holds. Hence the result is true if we define our processes in \( D([0, \infty), \mathbb{R}^d) \) by [7, Section 16].

**Remark 3.** If we replace Assumption (A) with the stronger condition \( \sum_{n=1}^{\infty} \| P^n v \|_2 < \infty \), then the iterated WIP is already known by Kelly and Melbourne [29, Theorem 4.3].

**Remark 4.** If \( T \) is modelled by a Young tower [52, 53], then the iterated WIP is already known under Assumption (A) by Kelly and Melbourne [29, Theorem 10.2].

Remarks 3 and 4 tell us that Theorem 3.4.2 gives new examples for maps which are not modelled by a Young tower and that mix sufficiently slowly. Time-one maps of slowly mixing semi-flows are a rich source of such examples.

### 3.4.1 Suspensions and Examples

Suppose \( T_t : M \to M \) is a semi-flow over some metric space \((M, d)\) and that \( v, w : M \to \mathbb{R} \) are some observables. For \( 0 < \eta \leq 1 \) define \( |v|_{C^\eta} := \operatorname{sup}_{x \neq x'} |v(x) - v(x')|/d(x, x')^\eta \) and

\[
C^\eta(M) := \{ v : M \to \mathbb{R} : \|v\|_{C^\eta} := |v|_{C^\eta} + \|v\|_\infty < \infty \}.
\]

Let \( |v|_{C^0, \eta} := \operatorname{sup}_{x \in M, t > 0} |v(T_t x) - v(x)|/t^\eta \) and define the set of observables

\[
C^{0, \eta}(M) := \{ v : M \to \mathbb{R} : \|v\|_{C^\eta} + |v|_{C^0, \eta} < \infty \}.
\]

Such observables are Hölder, and *Hölder in the flow direction*. Further define the *derivative in the flow direction* \( \partial_t w := \lim_{t \to 0} (w \circ T_t - w)/t \), and let

\[
L^{\infty, m}(M) := \{ w : M \to \mathbb{R} : \|w\|_{L^{\infty, m}} := \sum_{j=0}^{m} \| \partial_t^j w \|_\infty < \infty \}
\]

be the set of observables that have \( m \) derivatives in the flow direction that are \( L^{\infty}(M) \).

The following example outlines a generic situation where the semi-flow attains the same mixing rate as a suitable Poincaré map.

**Example 3.4.3.** Suppose \( T_t : M \to M \) is a semi-flow over an ambient, bounded metric space \((M, d)\). Let \( \Lambda \subseteq M \) be a Borel subset and suppose there exists a Hölder function \( h : \Lambda \to \mathbb{R}^+ \), with \( \inf h > 0 \), such that the map

\[
T_x := T_{h(x)} : \Lambda \to \Lambda
\]
is modelled by a Young tower with polynomial tails [53]. For example \( \Lambda \) can be a Poincaré section for the semi-flow. Further assume that for any \( C_1 > 0 \) there exists \( C_2 > 0 \) such that

\[
d(T_t x, T'_t x') \leq C_2 d(x, x')^q \quad \text{for all } t \in [0, C_1] \text{ and } x, x' \in M.
\]

Finally, assume absence of approximate eigenfunctions [37, Definition 3.4]. [37, Section 5] gives three different sufficient conditions for absence of approximate eigenfunctions. The first condition is imposed on periodic data of the semi-flow, and the set of roof functions that satisfy this condition is a prevalent [24] set. Prevalence is an analogue of almost-everywhere suitable for infinite dimensional spaces, and so absence of approximate eigenfunctions is satisfied with ‘probability one’.

The second condition is \( C^\infty \)-dense and \( C^2 \)-open; i.e. if there exists approximate eigenfunctions, we can make a \( C^\infty \) perturbation to the roof function \( h \) to ensure absence of approximate eigenfunctions. Let \( v \in C^{0,\eta}(M) \) be a mean zero observable. Then [37, Theorem 3.11] guarantees a \( \beta > 0, m \geq 1 \) and \( C > 0 \) such that

\[
\left| \int v \circ T_t \, d\mu \right| \leq C (\|v\|_{C^\eta} + |v|_{C^{0,\eta}}) \|w\|_{L_{m,1}} t^{-\beta},
\]

for all \( w \in L_{m,1}(M) \). The rate \( \beta \) depends only on the polynomial tail of the the Young tower. The iterated WIP now holds for the time one map \( T_1 \) when \( \beta > 1 \). This gives new results for \( \beta \in (1, 2] \). The case \( \beta > 2 \) follows from Kelly and Melbourne [29, Theorem 4.3].

We now outline a method of constructing semi-flows from a discrete map \( T : \Lambda \to \Lambda \), and give an example where the mixing rate of \( T \) passes over to the semi-flow.

Assume \( (\Lambda, d) \) is a metric space and consider an integrable function \( h : [0, 1] \to \mathbb{R}^+ \).

Define the space \( \Lambda^h := (\Lambda \times \mathbb{R}^+) / \sim \), where we make the identification \((x, h(x)) \sim (Tx, 0)\). We call \( h \) a roof function and define the suspension semi-flow \( T_t \) over \( h \) as

\[
T_t : \Lambda^h \to \Lambda^h
T_t(x, u) = (x, u + t) \mod \sim
\]

Figure 3.3 depicts a graphical representation of \( T_t \). Defining \( \bar{h} := \int h \, d\mu \) we have that \( \mu^h := \mu \times \text{Leb} / \bar{h} \) is a probability measure on \( \Lambda^h \).

**Proposition 3.4.4.** If \( T \) is measure-preserving with respect to \( \mu \) on \( \Lambda \), then \( T_t \) is measure-preserving with respect to \( \mu^h \) on \( \Lambda^h \).
Proof. Let \( f \in L^1(\Lambda^h) \) and calculate that
\[
\int f \circ T_t \, d\mu^h = \bar{h}^{-1} \int_0^{h(x)} \int_\Lambda f(T_t(x, s)) \, d\mu(x) \, ds = \bar{h}^{-1} \int_0^{h(x)} \int_\Lambda f(x, s + t) \, d\mu(x) \, ds
\]
\[
= \bar{h}^{-1} \int_t^{h(x)+t} \int_\Lambda f(x, s) \, d\mu(x) \, ds
\]
\[
= \bar{h}^{-1} \left( \int_0^{h(x)} + \int_{h(x)}^{h(x)+t} - \int_0^t \right) \int_\Lambda f(x, s) \, d\mu(x) \, ds
\]
\[
= \int f \, d\mu^h + \bar{h}^{-1} \left( \int_{h(x)}^{h(x)+t} - \int_0^t \right) \int_\Lambda f(x, s) \, d\mu(x) \, ds.
\]
We see that the second term is zero since
\[
\int_{h(x)}^{h(x)+t} \int_\Lambda f(x, s) \, d\mu(x) \, ds = \int_0^t \int_\Lambda f(T_t(x, s)) \, d\mu(x) \, ds = \int_0^t \int_\Lambda f(x, s) \, d\mu(x) \, ds,
\]
and the result follows.

Example 3.4.5. Let \( 0 < \alpha < 1 \) and let \( T \) be the intermittent map introduced in Example 3.2.2:
\[
T_x = \begin{cases} 
    x(1 + 2^\alpha x^\alpha) & \text{if } 0 \leq x \leq 1/2 \\
    2x - 1 & \text{if } 1/2 < x \leq 1.
\end{cases}
\]
Recall that for mean zero Hölder observables \( v \in C^\eta([0, 1]) \) we have \( \|P^n v\|_1 \leq C n^{-1/\alpha + 1} \), and so Assumption (A) holds when \( 0 < \alpha < 1/2 \). Since \( T \) is modelled by a Young tower, the iterated WIP already holds by [29]. To obtain new examples, we lift to a suitable semi-flow and consider the time-one map.
Let \( h : [0, 1] \to \mathbb{R}^+ \) be an \( \eta \)-Hölder roof function such that \( \inf h > 0 \) and let \( T_t : \Lambda^h \to \Lambda^h \) be the corresponding suspension flow. Further assume absence of approximate eigenfunctions for \( T_t \) and assume condition (3.14) holds with respect to the ‘metric’ induced on \( \Lambda^h \), 
\[
d((x, u), (y, t)) := d(x, y) + |u - t|.
\]
Note that \( d \) is not really a metric on \( \Lambda^h \), since 
\[
d((x, h(x)), (Tx, 0)) = d(x, Tx) + h(x).
\]
Let \( v \in C^{0, \eta}(\Lambda^h) \) be a mean zero observable. Then [37, Theorem 3.11] guarantees an \( m \geq 1 \) and \( C > 0 \) such that 
\[
\left| \int v w \circ T_t d\mu^h \right| \leq C \left( \|v\|_{C^\eta} + |v|_{C^{0, \eta}} \right) \|w\|_{L^\infty, m} t^{-1/\alpha + 1},
\]
for all \( w \in L^\infty, m(\Lambda^h) \). Proposition 2.1.3 shows assumption (A) holds for the time-one map \( T_1 \) when \( 0 < \alpha < 1/2 \), and therefore Theorem 3.4.2 gives the iterated WIP for \( T_1 \). Since \( T_1 \) is not modelled by a Young tower, this gives new results in the range \( 1/3 \leq \alpha < 1/2 \).

Remark. Any (semi-)flow with a neutral direction different to the flow direction cannot be modelled by a Young tower.

### 3.4.2 Proof of the Iterated WIP

Define 
\[
\mathcal{X}_n : [0, 1] \to \mathbb{R}^{d \times d}, \quad \mathcal{X}_n^{\beta\gamma}(t) := n^{-1} \sum_{j=0}^{[nt]-1} \sum_{i=0}^{j-1} m^\beta \circ T_i v^\gamma \circ T_j.
\]

**Proposition 3.4.6** (Martingale iterated WIP). We have that 
\[
(W_n, \mathcal{X}_n) \to_w (W, \mathcal{X})
\]
in \( D([0,1], \mathbb{R}^d \times \mathbb{R}^{d \times d}, \|\cdot\|_\infty) \), where \( W \) is a \( d \) dimensional Brownian motion with mean zero and covariance \( \Sigma = \int m m^T d\mu \), and 
\[
\mathcal{X}(t) = \int_0^t W \otimes dW
\]
for \( t \in [0, 1] \).

The proof of this proposition is postponed until the next subsection. We first prove the iterated WIP:

**Proof of Theorem 3.4.2.** Let \( \Gamma^{\beta\gamma}(t) := t \sum_{r=1}^\infty \int v^\beta v^\gamma \circ T^r d\mu \). Note that \( (W_n, \mathcal{W}_n) = (W_n, \mathcal{X}_n) + (0, \mathcal{W}_n - \mathcal{X}_n) \). By Proposition 3.4.6, \( (W_n, \mathcal{X}_n) \to_w (W, \mathcal{X}) \) with \( \Sigma = \int m m^T d\mu \). Propositions 3.2.13 and 3.2.14 give the correct covariance. By Proposition 3.3.1 it remains to show \( \mathcal{W}_n - \mathcal{X}_n \to_p \Gamma \). This is equivalent to the component-wise
statement:
\[
\sup_{0 \leq t \leq 1} \left| \mathbb{W}_{n}^{\beta \gamma}(t) - \mathbb{X}_{n}^{\beta \gamma}(t) - t \sum_{r=1}^{\infty} \int v^\beta v^\gamma \circ T^r \, d\mu \right| \to_{p} 0,
\]
for each \( \beta, \gamma = 1, \ldots, d \). Instead of convergence in probability, we prove convergence almost surely.

By Proposition 2.2.2 it suffices to show
\[
\mathbb{W}_{n}^{\beta \gamma}(1) - \mathbb{X}_{n}^{\beta \gamma}(1) \to_{a.s} \sum_{j=1}^{\infty} \int v^\beta v^\gamma \circ T^j \, d\mu,
\]
which follows from the following calculation:
\[
\mathbb{W}_{n}^{\beta \gamma}(1) - \mathbb{X}_{n}^{\beta \gamma}(1) = n^{-1} \sum_{j=0}^{n-1} \sum_{i=0}^{j-1} \left( v^\beta \circ T^i v^\gamma \circ T^j - m^\beta \circ T^i v^\gamma \circ T^j \right)
\]
\[
= n^{-1} \sum_{j=0}^{n-1} \sum_{i=0}^{j-1} \left( \chi^\beta \circ T - \chi^\beta \right) \circ T^i v^\gamma \circ T^j = n^{-1} \sum_{j=0}^{n-1} \left( \chi^\beta \circ T^j - \chi^\beta v^\gamma \circ T^j \right)
\]
\[
= n^{-1} \sum_{j=0}^{n-1} \left( \chi^\beta v^\gamma \circ T^j - \chi^\beta \right) \circ T^j = n^{-1} \sum_{j=0}^{n-1} \left( \chi^\beta v^\gamma \circ T^j = \int \chi^\beta v^\gamma \, d\mu \to_{a.s.} \int \chi^\beta v^\gamma \circ T^r \, d\mu. \right.
\]

\[\square\]

**Proof of Proposition 3.4.6**

Here we follow a technique similar to one used in [29]. We first introduce the lifted space to consider the backwards processes.

By Proposition 3.2.4, lift the probability space \((\Lambda, \mathcal{A}, \mu)\) and map \(T\) to obtain a probability space \((\tilde{\Lambda}, \tilde{\mathcal{A}}, \tilde{\mu})\), where \(\tilde{\mu}\) is an ergodic, invariant measure with respect to the invertible map \(\tilde{T} : \tilde{\Lambda} \to \tilde{\Lambda}\). Associated to this is a measure-preserving projection \(\pi : \tilde{\Lambda} \to \Lambda\) such that \(\pi \circ \tilde{T} = T \circ \pi\). Define the lifted observables \(\tilde{v} := v \circ \pi : \tilde{\Lambda} \to \mathbb{R}^d\) and \(\tilde{m} = m \circ \pi\).

Next define the lifted forwards process \(\tilde{W}_n\) the same as the forward processes \(W_n\) but with \(v, T\) replaced with \(\tilde{v}, \tilde{T}\) respectively. Define the lifted backwards process \(\tilde{W}_n^- : [0, 1] \to \mathbb{R}^d\),
\[
\tilde{W}_n^-(t) = n^{-1/2} \sum_{j=1}^{[nt]} \tilde{v} \circ \tilde{T}^{-j},
\]
and define \(\tilde{M}_n^- (t) = n^{-1/2} \sum_{j=1}^{[nt]} \tilde{m} \circ \tilde{T}^{-j}. \) Recall that \(\sum_{j=1}^{0} \tilde{v} \circ \tilde{T}^{-j} = \sum_{j=1}^{0} \tilde{m} \circ \tilde{T}^{-j} = 0. \) We define our iterated processes as follows. For coordinates \(\beta, \gamma = 1, \ldots, d\) define
the lifted process
\[ \tilde{x}_n : [0, 1] \to \mathbb{R}^{d \times d}, \quad \tilde{x}_n^{\beta \gamma}(t) := n^{-1} \sum_{j=0}^{\lfloor nt \rfloor - 1} \sum_{i=0}^{j-1} \tilde{m}^\beta \circ \tilde{T}^i \tilde{v}^\gamma \circ \tilde{T}^j, \]
and the lifted backwards process
\[ \tilde{x}_n^{-} : [0, 1] \to \mathbb{R}^{d \times d}, \quad \tilde{x}_n^{-\beta \gamma}(t) := n^{-1} \sum_{j=1}^{\lfloor nt \rfloor - 1} \sum_{i=j+1}^{\lfloor nt \rfloor} \tilde{v}^\beta \circ \tilde{T}^{-j} \tilde{m}^\gamma \circ \tilde{T}^{-i}. \]

To prove Proposition 3.4.6 we verify the hypothesis of Proposition 3.3.12 with \( A_n = (W_n, \chi_n) \) and \( B_n = \chi(W_n^{-}, \tilde{M}_n^{-}, \tilde{x}_n^{-}) \) for some continuous function \( \chi \). The remainder of this section is as follows: First, we verify that \( A_n = B_n + F_n \), where \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \). Next, we show \( (W_n^{-}, \tilde{M}_n^{-}, \tilde{x}_n^{-}) \to_w (W, W, \chi) \). Finally, we show \( B_n = \chi(W_n^{-}, \tilde{M}_n^{-}, \tilde{x}_n^{-}) \to_w \chi(W, W, \chi) =: Z \) and \( Z \equiv_d (W, \chi) \).

Verifying \( A_n = B_n + F_n \)

Define
\[ g : D([0, 1], \mathbb{R}^d) \to \tilde{D}([0, 1], \mathbb{R}^d), \quad g(u)(t) := u(1) - u(1 - t), \]
\[ h : D([0, 1], \mathbb{R}^d) \times D([0, 1], \mathbb{R}^d) \to \tilde{D}([0, 1], \mathbb{R}^{d \times d}), \quad h(u, v)(t) := u(1 - t)(v(1) - v(1 - t))^T. \]

**Proposition 3.4.7.** We have that
\[ (W_n, \chi_n) =_d \left( g(W_n^{-}), \left( g(\tilde{x}_n^{-}) - h(W_n^{-}, \tilde{M}_n^{-}) \right)^T \right) + F_n, \]
where \( F_n \) is such that \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \).

**Proof.** Since \( \pi \) and \( \tilde{T} \) are measure-preserving, we see
\[ (W_n, \chi_n) =_d (\tilde{W}_n, \tilde{x}_n) \circ \tilde{T}^{-n}. \]
Hence it suffices to show
\[ (\tilde{W}_n, \tilde{x}_n) \circ \tilde{T}^{-n} = \left( g(\tilde{W}_n^{-}), \left( g(\tilde{x}_n^{-}) - h(\tilde{W}_n^{-}, \tilde{M}_n^{-}) \right)^T \right) + F_n, \tag{3.16} \]
where \( F_n \) is to be defined and \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \). We now drop the tilde notation.

The first coordinate of (3.16) has already been shown in the proof of Theorem 3.3.3. Since convergence in probability can be shown coordinatewise, it remains
to show
\[ \mathcal{X}_n^\gamma(t) \circ T^{-n} = g(\mathcal{X}_n^\gamma(t)) - h(W_n^-, M_n^-)^\gamma(t) + F_n(t), \]

and \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \). Consider

\[ \mathcal{X}_n^\gamma(t) \circ T^{-n} = n^{-1} \sum_{j=0}^{[nt]-1} \sum_{i=0}^{j-1} m^\beta \circ T^{i-n} v^\gamma \circ T^j \]

\[ = n^{-1} \sum_{j=-n}^{[nt]-1} \sum_{i=-n}^{j-1} m^\beta \circ T^{i-n} v^\gamma \circ T^j \]

\[ = n^{-1} \sum_{j=-n}^{n} \sum_{i=j+1}^{n} m^\beta \circ T^{-i} v^\gamma \circ T^{-j} \]

\[ = n^{-1} \left( \sum_{j=1}^{n-1} + \sum_{j=n}^{n-1-[nt]} - \sum_{j=1}^{n-1-[nt]} - \sum_{j=[nt]}^{n-1} \right) \sum_{i=j+1}^{n} m^\beta \circ T^{-i} v^\gamma \circ T^{-j} \]

\[ = \mathcal{X}_n^\gamma(t)^-1 + F_1^1(t) - A_n(t) - F_2^2(t), \quad (3.17) \]

where

\[ F_1^1(t) = n^{-1} v^\gamma \circ T^{-n} \sum_{i=n+1}^{n} m^\beta \circ T^{-i} = (n^{-1} v^\gamma \circ T^{-n}) \left( \sum_{i=1}^{0} m^\beta \circ T^{-i} \right) \circ T^{-n} = 0, \]

\[ A_n(t) = n^{-1} \sum_{j=1}^{[nt]-1} \sum_{i=j+1}^{n} m^\beta \circ T^{-i} v^\gamma \circ T^{-j}, \]

\[ F_2^2(t) = n^{-1} \sum_{j=[nt]-1}^{n} \sum_{i=j+1}^{n} m^\beta \circ T^{-i} v^\gamma \circ T^{-j}. \]
Next, calculate that

\[ A_n(t) = n^{-1} \sum_{j=1}^{[n-nt]-1} \sum_{i=j+1}^{n} m^\beta \circ T^{-i} v^\gamma \circ T^{-j} \]

\[ = n^{-1} \sum_{j=1}^{[n-nt]-1} \left( \sum_{i=j+1}^{n} + \sum_{i=[n-nt]+1}^{n} \right) m^\beta \circ T^{-i} v^\gamma \circ T^{-j} \]

\[ = X_n^{\gamma \beta, -}(1 - t) \]

\[ + \left( n^{-1/2} \sum_{i=[n-nt]+1}^{n} m^\beta \circ T^{-i} \right) \left( n^{-1/2} \sum_{j=1}^{[n-nt]} v^\gamma \circ T^{-j} - n^{-1/2} v^\gamma \circ T^{-[n-nt]} \right) \]

\[ = X_n^{\gamma \beta, -}(1 - t) + \left( M_n^{\beta, -}(1) - M_n^{\beta, -}(1 - t) \right) W_n^{\gamma, -}(1 - t) - F_3^\beta(t), \]

where

\[ F_3^\beta(t) = n^{-1/2} v^\gamma \circ T^{-[n-nt]} \left( M_n^{\beta, -}(1) - M_n^{\beta, -}(1 - t) \right). \]

Combining equations (3.17) and (??), we have

\[ X_n^{\gamma \beta}(t) \circ T^{-n} = X_n^{\gamma \beta, -}(1) - A_n(t) - F_2^\beta(t) \]

\[ = X_n^{\gamma \beta, -}(1 - t) - X_n^{\gamma \beta, -}(1 - t) - W_n^{\gamma, -}(1 - t) \left( M_n^{\beta, -}(1) - M_n^{\beta, -}(1 - t) \right) - F_2^\beta(t) + F_3^\beta(t) \]

\[ = g(X_n)^\gamma(1) - h(W_n, M_n)^\gamma(t) - F_2^\beta(t) + F_3^\beta(t). \]

Finally, we show \( \sup_t |F_2^\beta(t)|, \sup_t |F_3^\beta(t)| \to_p 0 \). For the first part, since \([n - nt] \geq \]

44
For \( n - [nt] - 1 \) we see

\[
|F^2_n(t)| = n^{-1} \left| \sum_{j=[nt]}^{n-[nt]} \sum_{i=j+1}^{n} m^\beta \circ T^{-i} \circ \gamma \circ T^{-j} \right|
\]

\[
\leq n^{-1} \sum_{j=-[nt]-1}^{n-[nt]} \left| \sum_{i=j+1}^{n} m^\beta \circ T^{-i} \circ \gamma \circ T^{-j} \right|
\]

\[
= n^{-1} \left| \gamma \circ T^{[nt]-n+1} \right| \sum_{i=-[nt]}^{n} m^\beta \circ T^{-i}
\]

\[
+ n^{-1} \left| \gamma \circ T^{[nt]-n} \right| \sum_{i=-[nt]+1}^{n} m^\beta \circ T^{-i}
\]

\[
\leq \left( n^{-1/2} \left| \gamma \circ T^{[nt]-n+1} \right| \right) \left( n^{-1/2} \max_{0 \leq j \leq n} |m^\beta \circ T^{-j}| + \sup_{0 \leq t \leq 1} |M^\beta_n(t) \circ T^{-n}| \right)
\]

\[
+ \left( n^{-1/2} \left| \gamma \circ T^{[nt]-n} \right| \right) \left( |M^\beta_n(t) \circ T^{-n}| \right).
\]

Hence

\[
\sup_{0 \leq t \leq 1} |F^2_n(t)| \leq n^{-1/2} \max_{0 \leq j \leq n} |\gamma| \circ T^{-j+1} \left( n^{-1/2} \max_{0 \leq j \leq n} |m^\beta \circ T^{-j}| + \sup_{0 \leq t \leq 1} |M^\beta_n(t) \circ T^{-n}| \right)
\]

\[
+ n^{-1/2} \max_{0 \leq j \leq n} |\gamma| \circ T^{-j} \sup_{0 \leq t \leq 1} |M^\beta_n(t) \circ T^{-n}|
\]

\[
\leq 2n^{-1/2} \max_{0 \leq j \leq n} |\gamma| \circ T^{-j} \sup_{0 \leq t \leq 1} |M^\beta_n(t) \circ T^{-n}|
\]

\[
+ n^{-1} \max_{0 \leq j \leq n} |m^\beta \circ \gamma \circ T| \circ T^{-j}.
\]  \hspace{1cm} (3.18)

For \( F^3_n \), calculate that

\[
\sup_{0 \leq t \leq 1} |F^3_n(t)| = \sup_{0 \leq t \leq 1} \left| n^{-1/2} \gamma \circ T^{-[n-nt]} \left( M^\beta_n(-1) - M^\beta_n(-(1-t)) \right) \right|
\]

\[
\leq 2n^{-1/2} \max_{1 \leq j \leq n} |\gamma| \circ T^{-j} \sup_{0 \leq t \leq 1} |M^\beta_n(t) - M^\beta_n(t)|. \hspace{1cm} (3.19)
\]

Since \( \gamma \in L^2(\Lambda) \) and \( m^\beta \circ \gamma \circ T \in L^1(\Lambda) \), Proposition 2.2.3 shows

\[
n^{-1/2} \max_{1 \leq j \leq n} |\gamma| \circ T^{-j} \to_{a.s} 0, \quad n^{-1} \max_{1 \leq j \leq n} |m^\beta \circ \gamma \circ T| \circ T^{-j} \to_{a.s} 0.
\]

Therefore the second term of (3.18) converges to zero almost surely. Finally, we show (3.19) and the first term of (3.18) converges to zero in probability. Note that by Proposition 3.3.8 we have \( M^\beta_n \to_w W^\beta \), and so the continuous map-
Proposition 3.4.9. Showing

\[ (\tilde{\omega}) \text{ theorem shows } \sup_t |M^{\delta}_{n,-}(t)|, \sup_t |M^\beta_{n}(t)| \to_w \sup_t |W^\beta(t)|. \]
In particular, \( \left( \sup_t |M^{\delta}_{n,-}(t)| \right)_{n \geq 1}, \left( \sup_t |M^\beta_{n}(t)| \right)_{n \geq 1} \) are tight sequences. Proposition 2.2.4 now shows

\[
\begin{align*}
&n^{-1/2} \max_{1 \leq j \leq n} |v^\gamma| \circ T^{-j} \sup_{0 \leq t \leq 1} |M^{\delta}_{n,-}(t)| \to_p 0, \\
&n^{-1/2} \max_{1 \leq j \leq n} |v^\gamma| \circ T^{-j} \sup_{0 \leq t \leq 1} |M^\beta_{n}(t) \circ T^{-n}| \to_p 0,
\end{align*}
\]
and therefore \( \sup_t |F^2_n(t)|, \sup_t |F^3_n(t)| \to_p 0. \)

\[ \square \]

Showing \( (\tilde{W}_n^{-}, \tilde{M}_n^{-}, \tilde{X}_n^{-}) \to_w (W, W, X) \)

To show \( (\tilde{W}_n^{-}, \tilde{M}_n^{-}, \tilde{X}_n^{-}) \to_w (W, W, X) \), we need a suitable convergence result:

**Proposition 3.4.8** ([26], [31, Theorem 2.2]). For each \( n \geq 1 \), let \( (X_n, Y_n) \) be an \( \mathcal{A}_{n,t} \)-measurable process with sample paths in \( D([0,1], \mathbb{R}^d \times \mathbb{R}^d) \) and let \( Y_n \) be a martingale with respect to \( \mathcal{A}_{n,t} \). Suppose \( \sup_n \mathbb{E}(|Y_n^\beta, Y_n^{\beta_1}|(t)) < \infty \) for each \( \beta = 1, \ldots, d \) and \( t \in [0,1] \), and \( (X_n, Y_n) \to_w (X, Y) \) in \( D([0,1], \mathbb{R}^d \times \mathbb{R}^d), \| \cdot \|_\infty \). Then

\[
\left(X_n, Y_n, \int X_n \otimes dY_n\right) \to_w \left(X, Y, \int X \otimes dY\right)
\]

in \( D([0,1], \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d}), \| \cdot \|_\infty \). \[ \square \]

**Proposition 3.4.9.** For \( \beta, \gamma = 1, \ldots, d \) and \( t \in [0,1] \) we have that the process \( \tilde{X}_{n}^{-} \) is given by the following Riemann-Stieltjes integral:

\[
\tilde{X}^{-}_{n}(t) = \int_{0}^{t} \tilde{W}_{n}^{-} \otimes d\tilde{M}_{n}^{-}.
\]

**Proof.** In this proof we drop the tilde notation. Consider a sequence of partitions \( \Delta_\ell \) of \( [0, t] \) such that \( \max_{\Delta_\ell} |t_{k+1} - t_k| \to 0 \) as \( \ell \to \infty \). By definition,

\[
\int_{0}^{t} W^{\beta}_{n,-} \, dM^{\gamma}_{n,-} = \lim_{\ell \to \infty} S_\ell,
\]

where

\[
S_\ell = \sum_{k=0}^{\ell-1} W^{\beta}_{n,-}(t_k) (M^{\gamma}_{n,-}(t_{k+1}) - M^{\gamma}_{n,-}(t_k)).
\]

Take \( \ell \) large enough so that \( \max_{k<\ell} |t_{k+1} - t_k| < \frac{1}{n} \). For \( j = 0, \ldots, \lfloor nt \rfloor - 1 \), the map \( s \mapsto \sum_{j=1}^{\lfloor n(s) \rfloor} v \circ T^{-j} \) is constant on each interval \( \left[ \frac{j}{n}, \frac{j+1}{n} \right) \), and so \( S_\ell \) depends only

46
on each jump $v_j$, for $j = 0, 1, \ldots, [nt] - 1$. Thus
\[
S_t = n^{-1} \sum_{j=0}^{[nt]-1} v_j^\beta (m_{j+1}^\gamma - m_j^\gamma) = n^{-1} \sum_{j=0}^{[nt]-1} \left( \sum_{i=1}^{j} v_i^\beta \circ T^{-i} \right) m_j^\gamma \circ T^{-(j+1)}
\]
\[
= n^{-1} \sum_{j=1}^{[nt]} \sum_{i=1}^{j-1} v_i^\beta \circ T^{-i} m_j^\gamma \circ T^{-j} = n^{-1} \sum_{1 \leq i < j \leq [nt]} v_i^\beta \circ T^{-i} m_j^\gamma \circ T^{-j}.
\]
This can be written as
\[
n^{-1} \sum_{1 \leq i < j \leq [nt]} v_i^\beta \circ T^{-i} m_j^\gamma \circ T^{-j} = n^{-1} \sum_{i=1}^{[nt]} \sum_{j=i+1}^{[nt]} v_i^\beta \circ T^{-i} m_j^\gamma \circ T^{-j}
\]
\[
= n^{-1} \sum_{i=1}^{[nt]-1} \sum_{j=i+1}^{[nt]} v_i^\beta \circ T^{-i} m_j^\gamma \circ T^{-j} = \mathbb{X}_n^\gamma(t),
\]
since taking $i = [nt]$ in the outer sum gives $n^{-1} v_i^\beta \circ T^{-[nt]} \sum_{j=[nt]+1}^{[nt]} m_j^\gamma \circ T^{-j} = 0$. Thus we obtain $\int_0^t W_n^- \otimes dM_n^- = \mathbb{X}_n(t)$ as required.

**Proposition 3.4.10** (Backward iterated WIP). We have the following convergence:

\[
(\tilde{W}_n^-, \tilde{M}_n^-, \tilde{X}_n^-) \to_w (W, W, X)
\]
in $(D([0,1], \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d}), \|\cdot\|_\infty)$.

**Proof.** For this proof we drop the tilde notation. We verify the hypothesis of Proposition 3.4.8 with

\[
X_n = W_n^-, \quad Y_n = M_n^-.
\]
We see $Y_n$ is a martingale by Proposition 3.3.7. The condition $\sup_n \mathbb{E}(|Y_n^{\beta, \gamma}(t)|) < \infty$ follows from Proposition 3.3.6 since

\[
\mathbb{E}(|M_n^{\beta, \gamma}(t)|) = \frac{1}{n} \sum_{j=1}^{[nt]} \int (m_j^\beta)^2 \circ T^{-j} d\mu = \frac{[nt]}{n} \int (m_j^\beta)^2 d\mu \leq t\|m_j^\beta\|_2^2. \quad (3.20)
\]
To verify weak convergence of $(X_n, Y_n)$, first note that $M_n^- \to_w W$ by Corollary 3.3.8, which implies $(M_n^-, M_n^-) \to_w (W, W)$ by the continuous mapping theorem. Moreover, the proof of Proposition 3.3.11 shows $\sup_{0 \leq t \leq 1} |W_n^-(t) - M_n^-(t)| \to_p 0$, so $(W_n^-, M_n^-) \to_w (W, W)$. Proposition 3.4.8 now applies and we obtain the
desired convergence
\[ (W_n^-, M_n^- \cdot \int W_n^- \otimes dM_n^-) \to_w (W, W, \int W \otimes dW), \]
in \( (\mathcal{D}([0, 1], \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d}), \| \cdot \|_{\infty}) \).

Checking \( B_n \to_w Z \) and \( Z =_d (W, X) \)

**Proposition 3.4.11.** We have that
\[ \left( g(\bar{W}_n^-), \left( g(\bar{X}_n^-) - h(\bar{W}_n^-, \bar{M}_n^-) \right)^T \right) \to_w \left( g(W), (g(X) - h(W, W))^T \right) \]
in \( (\hat{\mathcal{D}}([0, 1], \mathbb{R}^d \times \mathbb{R}^{d \times d}), \| \cdot \|_{\infty}) \).

**Proof.** We easily obtain the desired convergence by Proposition 3.4.10 and the continuous mapping theorem once we establish continuity of \( g \) and \( h \). It was shown in the proof of Theorem 3.3.3 that \( g \) is continuous. We see \( h \) is continuous since \( h^{\beta \gamma} \) is the product of two continuous maps, for each \( \beta, \gamma = 1, \ldots, d \).

**Proposition 3.4.12.** We have that
\[ \left( g(W), (g(X) - h(W, W))^T \right) =_d (W, X). \]

**Proof.** Fix \( t \in [0, 1] \) and coordinates \( \beta, \gamma = 1, \ldots, d \). Let \( C^{\beta \gamma}(t) := \int_0^t g(W)^\beta \, dg(W)^\gamma \).
We show \( (W, X) =_d (g(W), C) \) and \( C = (g(X) - h(W, W))^T \).

To see \( (W, X) =_d (g(W), C) \), note that
\[ X(t) = \lim_{n \to \infty} X_n(t), \quad C(t) = \lim_{n \to \infty} Y_n(t), \]
where the limits are taken in probability and
\[
X_n^{\beta \gamma}(t) = \sum_{k=0}^{[nt]-1} W^\beta \left( \frac{k}{n} \right) \left[ W^\gamma \left( \frac{k+1}{n} \right) - W^\gamma \left( \frac{k}{n} \right) \right],
\]
\[
Y_n^{\beta \gamma}(t) = \sum_{k=0}^{[nt]-1} \left[ g(W)^\beta \left( \frac{k}{n} \right) \left[ g(W)^\gamma \left( \frac{k+1}{n} \right) - g(W)^\gamma \left( \frac{k}{n} \right) \right].
\]
Denote the identity map by $I$. Since the map

$$f_n : D([0, 1], \mathbb{R}^d) \to D([0, 1], \mathbb{R}^{d \times d}),$$

$$(f_n(X))^{\beta \gamma}(t) = \sum_{k=0}^{[nt]-1} X^{\beta} \left(\frac{k}{n}\right) \left[ X^{\gamma} \left(\frac{k + 1}{n}\right) - X^{\gamma} \left(\frac{k}{n}\right) \right]$$

is continuous, $W =_d g(W)$ and weak equality is preserved under continuous maps, we have

$$(W, X_n) = (I, f_n)W =_d (I, f_n)g(W) = (g(W), Y_n).$$

Moreover since $X_n \to_p \mathbb{X}$ and $Y_n \to_p \mathbb{G}$, we see

$$(W, X_n) \to_w (W, \mathbb{X}), \quad (g(W), Y_n) \to_w (g(W), \mathbb{G}).$$

By uniqueness of weak limits, $(W, \mathbb{X}) =_d (g(W), \mathbb{G})$.

It remains to show $G^{\beta \gamma}(t) = g(\mathbb{X})^{\gamma \beta}(t) - h(W, W)^{\gamma \beta}(t)$. Recall that

$$G^{\beta \gamma}(t) = \int_0^t g(W)^\beta \, dg(W)^\gamma = \lim_{n \to \infty} Y_n^{\beta \gamma}(t),$$

where the limit is taken in probability and

$$Y_n^{\beta \gamma}(t) = \sum_{k=0}^{[nt]-1} g(W)^\beta \left(\frac{k}{n}\right) \left[ g(W)^\gamma \left(\frac{k + 1}{n}\right) - g(W)^\gamma \left(\frac{k}{n}\right) \right]$$

for $\mathbb{X}$. For $\mathbb{G}$, we have

$$Y_n^{\beta \gamma}(t) = \sum_{k=0}^{[nt]-1} \left[ W^{\beta} \left(1 - \frac{k}{n}\right) - W^{\beta} \left(1 - \frac{k + 1}{n}\right) \right] \left[ W^{\gamma} \left(1 - \frac{k}{n}\right) - W^{\gamma} \left(1 - \frac{k + 1}{n}\right) \right]$$

for $\mathbb{X}$. For $\mathbb{G}$, we have

$$Y_n^{\beta \gamma}(t) = \sum_{k=0}^{[nt]-1} \left[ W^{\beta} \left(1 - \frac{j}{n}\right) - W^{\beta} \left(1 - \frac{j + 1}{n}\right) \right] \left[ W^{\gamma} \left(1 - \frac{k}{n}\right) - W^{\gamma} \left(1 - \frac{k + 1}{n}\right) \right]$$

for $\mathbb{X}$. For $\mathbb{G}$, we have

$$Y_n^{\beta \gamma}(t) = \sum_{j=0}^{[nt]-1} \left[ W^{\beta} \left(1 - \frac{j}{n}\right) - W^{\beta} \left(1 - \frac{j + 1}{n}\right) \right] \left[ W^{\gamma} \left(1 - \frac{j}{n}\right) - W^{\gamma} \left(1 - \frac{j + 1}{n}\right) \right]$$

for $\mathbb{X}$. For $\mathbb{G}$, we have

$$Y_n^{\beta \gamma}(t) = \sum_{j=0}^{[nt]-1} \left[ W^{\beta} \left(1 - \frac{j}{n}\right) - W^{\beta} \left(1 - \frac{j + 1}{n}\right) \right] \left[ W^{\gamma} \left(1 - \frac{j}{n}\right) - W^{\gamma} \left(1 - \frac{j + 1}{n}\right) \right].$$
Now consider

\[
(g(X) - h(W, W))^{\gamma \beta}(t) = X^{\gamma \beta}(1) - X^{\gamma \beta}(1 - t) - W^{\gamma}(1 - t)\left(W^{\beta}(1) - W^{\beta}(1 - t)\right)
\]

\[
= \int_0^1 W^{\gamma} dW^{\beta} - \int_0^{1-t} W^{\gamma} dW^{\beta} - W^{\gamma}(1 - t)\left(W^{\beta}(1) - W^{\beta}(1 - t)\right)
\]

\[
= \int_{1-t}^{1} (W^{\gamma} - W^{\gamma}(1 - t)) \, dW^{\beta} = \lim_{n \to \infty} Z_{n}^{\beta \gamma}(t),
\]

where the limit is taken in probability and

\[
Z_{n}^{\beta \gamma}(t) = \sum_{k=\lfloor n - nt \rfloor}^{n-1} \left[ W^{\gamma}\left(\frac{k}{n}\right) - W^{\gamma}(1 - t) \right] \left[ W^{\beta}\left(\frac{k+1}{n}\right) - W^{\beta}\left(\frac{k}{n}\right) \right].
\]

With the change of coordinates \(j = n - k - 1\) we see

\[
Z_{n}^{\beta \gamma}(t) = \sum_{j=0}^{n - \lfloor n - nt \rfloor - 1} \left[ W^{\gamma}\left(1 - \frac{j + 1}{n}\right) - W^{\gamma}(1 - t) \right] \left[ W^{\beta}\left(1 - \frac{j}{n}\right) - W^{\beta}\left(1 - \frac{j + 1}{n}\right) \right].
\]  

(3.21)

By uniqueness of limits in probability, the result will follow once we establish that \(Z_{n} - Y_{n} \to_p 0\). We instead show the stronger statement \(\sup_{0 \leq t \leq 1} |Z_{n}^{\beta \gamma}(t) - Y_{n}^{\beta \gamma}(t)| \to_{a.s.} 0\). First suppose \(nt\) is not an integer, so \(\lfloor n - nt \rfloor = n - \lfloor nt \rfloor - 1\) and the upper limit in the sum in equation (3.21) is now \(n - n - \lfloor nt \rfloor - 1 = \lfloor nt \rfloor\). Denoting \(W_{j}^{\beta} := W^{\beta}(1 - \frac{j}{n}) - W^{\beta}(1 - \frac{j + 1}{n})\) we have

\[
Z_{n}^{\beta \gamma}(t) - Y_{n}^{\beta \gamma}(t) = \sum_{j=0}^{\lfloor nt \rfloor} W_{j}^{\beta} \left[ W^{\gamma}\left(1 - \frac{j + 1}{n}\right) - W^{\gamma}(1 - t) \right]
\]

\[
- \sum_{j=0}^{\lfloor nt \rfloor - 1} W_{j}^{\beta} \left[ W^{\gamma}\left(1 - \frac{j + 1}{n}\right) - W^{\gamma}\left(1 - \frac{\lfloor nt \rfloor}{n}\right) \right]
\]

\[
= \sum_{j=0}^{\lfloor nt \rfloor - 1} W_{j}^{\beta} \left[ W^{\gamma}\left(1 - \frac{\lfloor nt \rfloor}{n}\right) - W^{\gamma}(1 - t) \right]
\]

\[
+ \sum_{j=\lfloor nt \rfloor}^{\lfloor nt \rfloor} W_{j}^{\beta} \left[ W^{\gamma}\left(1 - \frac{j + 1}{n}\right) - W^{\gamma}(1 - t) \right]
\]

\[
= A_{n}(t) + B_{n}(t),
\]

50
where

\[ A_n(t) = \sum_{j=0}^{[nt]-1} W_j^\beta \left[ W^\gamma \left( 1 - \frac{[nt]}{n} \right) - W^\gamma (1 - t) \right] \]

\[ = \left[ W^\beta (1) - W^\beta \left( 1 - \frac{[nt]}{n} \right) \right] \left[ W^\gamma \left( 1 - \frac{[nt]}{n} \right) - W^\gamma (1 - t) \right], \]

\[ B_n(t) = \sum_{j=[nt]}^{[nt]+1} W_j^\beta \left[ W^\gamma \left( 1 - \frac{j}{n} \right) - W^\gamma (1 - t) \right] \]

\[ = \left[ W^\beta \left( 1 - \frac{[nt]}{n} \right) - W^\beta \left( 1 - \frac{[nt]+1}{n} \right) \right] \left[ W^\gamma \left( 1 - \frac{[nt]+1}{n} \right) - W^\gamma (1 - t) \right]. \]

Since \( \frac{[nt]}{n}, \frac{[nt]+1}{n} \to t \) and \( W \) is almost surely uniformly continuous, we have \( \sup_t |A_n(t)|, \sup_t |B_n(t)| \to_{a.s.} 0 \) and hence \( \sup_t |Z_n^{\beta\gamma}(t) - Y_n^{\beta\gamma}(t)| \to_{a.s.} 0 \). When \( nt \) is an integer, the upper limit in the sum in equation (3.21) is \( n-[n-nt]-1 = [nt]-1 \), so \( B_n = 0 \) and the above convergence still holds. \( \square \)

**Proof of Proposition 3.4.6**

*Proof of Proposition 3.4.6.* We verify the hypothesis of Proposition 3.3.12 with

\[ A_n = (W_n, \tilde{X}_n), \quad B_n = \left( g(W_n^-), \left( g(\tilde{X}_n^-) - h(\tilde{W}_n^-, \tilde{M}_n^-) \right)^T \right). \]

Proposition 3.4.7 shows \( A_n = B_n + F_n \) and \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \). Proposition 3.4.11 shows \( B_n \to_w Z \) for some process \( Z \) and Proposition 3.4.12 shows \( Z =_d (W, \tilde{X}) \). Therefore, Proposition 3.3.12 shows

\[ (W_n, \tilde{X}_n) \to_w (W, \tilde{X}), \]

in \( (D([0,1], \mathbb{R}^d \times \mathbb{R}^{d \times d}), \| \cdot \|_\infty) \) as desired. \( \square \)
Chapter 4

Limit Theorems for Invertible Maps

For this chapter, suppose $T : \Lambda \to \Lambda$ is an invertible, ergodic and measure-preserving map over the probability space $(\Lambda, \mathcal{A}, \mu)$ and let $v : \Lambda \to \mathbb{R}$ be a mean zero $L^\infty(\Lambda)$ observable.

We do not need to lift our map $T$ to obtain an invertible one, and hence the backwards Birkhoff sums $v_n^{-} = \sum_{j=1}^{n} v \circ T^{-j}$ are always defined. We denote the $n$th backwards Birkhoff sum with a subscript $n$ and a minus superscript. Unfortunately, however, the transfer operator is $Pv = v \circ T^{-1}$ (see the proof of Proposition 2.1.2). As a consequence of $T$ being measure-preserving, the mixing condition $\sum_{n=1}^{\infty} \|P^n v\|_1 < \infty$ from Section 3 can never be non-trivially satisfied. To remedy this, we suppose there exists a sub-$\sigma$ algebra $\mathcal{A}_0 \subset \mathcal{A}$ such that $T^{-1} \mathcal{A}_0 \subseteq \mathcal{A}_0$. Then $\mathbb{E}(v \circ T^{-1} \mid \mathcal{A}_0)$ plays the role of the transfer operator. Many authors [11, 18, 32, 42] assume existence of such a sigma algebra.

If there exists a partition $W^s = \{W^s(x) : x \in \Lambda\}$ of $\Lambda$—where $W^s(x)$ is a measurable set containing $x$—such that $T W^s(x) \subseteq W^s(Tx)$ for all $x \in \Lambda$, then we say $W^s$ is a stable foliation for the map $T$ and we call $W^s(x)$ a local stable leaf. We say $T$ contracts stable leaves exponentially if there exists $C > 0$ and $\gamma \in (0, 1)$ such that for every $x \in \Lambda$ and $x' \in W^s(x)$ we have $d(T^n x', T^n x) \leq C \gamma^n$. The sigma algebra generated by the partition, $\mathcal{A}_0 := \sigma(W^s)$, is a sub-$\sigma$ algebra such that $T^{-1} \mathcal{A}_0 \subseteq \mathcal{A}_0$.

Let $L^p(\Lambda, \mathcal{A}_0) := \{v \in L^p(\Lambda) : v \text{ is } \mathcal{A}_0\text{-measurable}\}$. For $n \in \mathbb{Z}$, define $\mathcal{A}_n := T^{-n} \mathcal{A}_0$ and $\mathbb{E}_n v := \mathbb{E}(v \mid \mathcal{A}_n)$. For any $\ell, p \in \mathbb{Z}$ we have

$$\mathbb{E}_\ell \mathbb{E}_p v = \mathbb{E}(\mathbb{E}(v \mid T^{-\ell} \mathcal{A}_0) \mid T^{-p} \mathcal{A}_0) = \mathbb{E}(v \mid T^{-\max(\ell,p)} \mathcal{A}_0) = \mathbb{E}_{\max(\ell,p)} v,$$
\[ (\mathbb{E}_\ell v) \circ T^p = \mathbb{E}(v \mid T^{-\ell}A_0) \circ T^p = \mathbb{E}(v \circ T^p \mid T^{-\ell-p}A_0) = \mathbb{E}_{\ell+p}(v \circ T^p). \]

A standard probabilistic assumption is

\[ \sum_{n=0}^{\infty} \| \mathbb{E}_0(v \circ T^n) - v \circ T^n \|_1 + \sum_{n=1}^{\infty} \| \mathbb{E}_0(v \circ T^{-n}) \|_1 < \infty, \quad (A_{\text{inv}}) \]

and we assume this holds throughout the following chapter, unless stated otherwise.

### 4.0.1 Sufficient Criteria for Assumption \( (A_{\text{inv}}) \)

We do not assume \( (A_{\text{inv}}) \) in this subsection.

**Proposition 4.0.1** ([12]). Assume there exists \( \beta > 1 \) and \( C > 0 \) such that for all \( n \geq 1 \),

\[ \left| \int v w \circ T^n d\mu \right| \leq C \| w \|_\infty n^{-\beta} \text{ for all } w \in L^\infty(\Lambda, A_0). \]

Then

\[ \| \mathbb{E}_0(v \circ T^{-n}) \|_1 \leq C \| v \|_\infty n^{-\beta}. \]

**Proof.** Define \( \zeta := \text{sgn} \left( \mathbb{E}_n v \right) \). Clearly \( \zeta \) is \( A_n \)-measurable. Using the hypothesis and the fact that \( \| \mathbb{E}_0(v \circ T^{-n}) \|_1 = \| \mathbb{E}_n v \|_1 \) we see

\[ \| \mathbb{E}_n v \|_1 = \int \mathbb{E}_n v \zeta d\mu = \int \mathbb{E}_n (v \zeta) d\mu = \int v \zeta d\mu = \int v \text{sgn} \left( \mathbb{E}_n v \right) \circ T^{-n} d\mu \leq C \| v \|_\infty n^{-\beta}. \]

\[ \square \]

**Proposition 4.0.2** ([12]). Assume that \( (\Lambda, d) \) is a metric space. The observable \( v : \Lambda \to \mathbb{R} \) only needs to be mean zero and measurable here. Suppose \( T \) admits a stable foliation \( W^s \) and let \( Y \subseteq \Lambda \) be a positive measure subset that is a union of stable leaves in \( W^s \). Define the first return time \( \tau : Y \to \mathbb{N} \), \( \tau(y) := \inf \{ n \geq 1 : T^n y \in Y \} \) and induced map \( F := T^\tau : Y \to Y \). Let \( h_n(x) := \# \{ 0 \leq j \leq n : T^j x \in Y \} \).

Suppose that \( \mu(\tau > n) \leq C n^{-\alpha} \) for some \( \alpha > 1 \), and that there are constants \( C \geq 1, \gamma \in (0, 1) \) such that

\[ \left| \text{diam} \left( v(T^n W^s) \right) \right| \leq C \gamma^{h_n(x)} \text{ for all } W^s \in \mathcal{W}^s, n \geq 1. \quad (4.1) \]
Then
\[ \|E_0(v \circ T^n) - v \circ T^n\|_1 \leq Cn^{-\alpha + 1}. \]

Remark. Suppose \( S : \Lambda \to \Lambda \) is another invertible, ergodic and measure-preserving map that admits the same stable foliation as \( T \). The conclusion of Proposition 4.0.2 remains valid when equation (4.1) is replaced with
\[ \text{diam} \left(v(S^n W^s)\right) \leq C_{\gamma} h_n(x) \text{ for all } W^s \in W^s, n \geq 1, \]
where \( h_n \) remains unchanged.

The following definition is a subset of the original definition of Young [52], but it suffices for our purposes.

**Definition 4.0.3** (Young tower, [40, Page 3]). We say \( T \) is modelled by a Young tower with tail \( \tau \) if there is a subset \( Y \) and a countable partition \( Y_j \) of \( Y \subseteq \Lambda \) such that \( \tau \) is constant on each \( Y_j \). There is a measurable partition \( W^s \) of \( Y \) (stable leaves) and each \( Y_j \) is a union of elements of \( W^s \). Let \( W^u \) denote a measurable subset that intersects each stable leaf only once. The induced map \( F = T^\tau : Y \to Y \) satisfies \( F(W^s(y)) \subseteq W^s(Fy) \), and we define the separation time \( s(y, y') \) as the least integer \( n \geq 0 \) such that \( F^n y, F^n y' \) lie in distinct partition elements. Further suppose there exists constants \( C \geq 1, \gamma \in (0, 1) \) such that
\[ d(T^\ell F^j y, T^\ell F^j y') \leq C_{\gamma}^j \text{ for all } y' \in W^s(y), y \in Y, \]
\[ d(T^\ell F^j y, T^\ell F^j y') \leq C_{\gamma}^s(y, y')^{-j} \text{ for all } y, y' \in W^s, \]
for all \( j \geq 0, 0 \leq \ell \leq r(F^j y) \). Let \( \bar{Y} = Y/ \sim \), where \( y \sim y' \) if \( y' \in W^s(y) \). Since \( F(W^s(y)) \subseteq W^s(Fy) \), we obtain a well defined quotient map \( \bar{F} : \bar{Y} \to \bar{Y} \). Assume \( \bar{F} \) is Gibbs-Markov [1, Chapter 4] and assume \( \tau \in L^1(Y) \).

**Proposition 4.0.4** ([12]). Assume \( (\Lambda, d) \) is a metric space and \( v : \Lambda \to \mathbb{R} \) is a Hölder continuous and mean zero observable. If \( T \) is modelled by a Young tower with return time \( \tau \), and tail rate \( \mu(\tau > n) \leq C n^{-\beta} \) for some \( C > 0 \) and \( \beta > 1 \), then
\[ \|E_0(v \circ T^{-n})\|_1 \leq C n^{-\beta + 1}, \quad \|E_0(v \circ T^n) - v \circ T^n\|_1 \leq C n^{-\beta + 1}. \]

\]
4.0.2 An Example – Intermittent Baker’s map

For $0 < e, k < 1$, let $T_{ek} : \Lambda \to \Lambda$, $\Lambda := [0, 1]^2$ be the intermittent Baker’s map

$$T_{ek}(x, y) = \begin{cases} (f_e(x), f_e^{-1}(y)), & (x, y) \in [0, 1/2) \times [0, 1] \\ (2x - 1, (y + 1)/2), & (x, y) \in [1/2, 1] \times [0, 1] \end{cases}$$

(4.3)

where $f_e(x) = x(1 + 2^e x^e)$. Note that $T_{ek}(x, y) = (T^{(1)}_e x, T^{(2)}_k y)$ has a skew-product structure. The map $T^{(1)}_e$ is the intermittent map from Example 3.2.2 and sees only expansion, whereas the map $T^{(2)}_k$ sees only contraction. The stable leaves are vertical lines. A smaller parameter corresponds to stronger expansion/contraction (taking the parameter $e = 0$ in $T^{(1)}_e$ gives the doubling map). Let $v : \Lambda \to \mathbb{R}$ be a Hölder continuous, mean zero observable.

The case $k = e$ is considered by Melbourne and Varandas [40, Example 4.1], and following this example, there is an ergodic invariant measure $\mu$ for $T := T_{ek}$. Moreover defining $Y := [1/2, 1] \times [0, 1]$, $T$ is modelled by a Young tower with return time map $\tau_e(\underline{x}) := \inf\{n \geq 1 : T^n\underline{x} \in Y\}$. Since $Y$ extends fully in the $y$-direction, the return time $\tau$ is identical to return time of the intermittent map $T^{(1)}_e$ to $[1/2, 1]$ and so [53, Proof of Theorem 5(c)] shows $\mu(\tau > n) \leq Cn^{-1/e}$. Theorem 4.0.4 now shows

$$\|E_0(v \circ T^{-n})\|_1 \leq Cn^{-1/e+1}, \quad \|E_0(v \circ T^n) - v \circ T^n\|_1 \leq Cn^{-1/e+1},$$

for mean zero, Hölder observables. Assumption (A_{inv}) now holds for all $0 < e = k < 1/2$.

For the iterated WIP, it will be necessary to construct an example with stronger contraction than expansion:

**Proposition 4.0.5.** Suppose $0 < k < e < 1/2$. Let $T_{ek}$ be the map defined in (4.3), and let $v : \Lambda \to \mathbb{R}$ be a Hölder continuous, mean zero observable. Then we have

$$\|E_0(v \circ T^{-n})\|_1 \leq Cn^{-1/e+1}, \quad \|E_0(v \circ T^n) - v \circ T^n\|_1 \leq Cn^{-1/k+1}.$$

**Proof.** Denote $T := T_{ek}$ and define $Y := [1/2, 1] \times [0, 1]$, $\tau_e(\underline{x}) := \inf\{n \geq 1 : T^n\underline{x} \in Y\}$. We first show $T$ is modelled by a Young tower with tail $\tau_e$. Since $T$ has a product structure and the stable leaves are the same for every choice of parameters $0 < k, e < 1/2$, the only condition we need to check is the exponential contraction on returns to $Y$ (condition (4.2)). This follows from the claim that $\tau_k \leq \tau_e$ for every $k < e$. Assuming the claim, we have $\tau_e = \tau_k + \tilde{\tau}$ for the function

55
Let \( S \) see \( \mathcal{N} \) leaves, it follows that there exists a \( T \) see \( \mathcal{N} \) from the product structure and the fact that \( \tau \) is constant on stable leaves, it follows that there exists a \( p \geq 0 \), depending on the value of \( \ell + j \tau \), such that

\[
d(T^\ell+j\tau, T^\ell\tau) \leq C\gamma^{-j-p} \leq C\gamma^{-j},
\]

for some \( C > 1 \), \( \gamma \in (0, 1) \), and so (4.2) is satisfied. Therefore \( T \) is modelled by a Young tower with tail \( \tau \) and Theorem 4.0.4 shows

\[
\| E_0(u \circ T^{-n}) \|_1 \leq Cn^{-1/e+1}. 
\]

We now prove the claim: Denote \( S_e := T^{(1)}_e \) with return time \( R_e(x) := \inf\{n \geq 1 : S^n_e x \in [1/2, 1]\} \). Since \( Y \) extends fully in the \( y \)-direction and \( T \) has a product structure, \( \tau = R_e \). Therefore it suffices to show that \( R_k \leq R_e \) for \( k < e \).

Recall that

\[
S_k x = \begin{cases} 
  x(1 + 2^k x^k), & x \in [0, 1/2) \\
  2x - 1 & x \in [1/2, 1]
\end{cases}
\]

Let \( y \in [1/2, 1] \). Since the second branch does not depend on the parameter \( k \), we see \( S_k y = S_e y =: x_0 \). If \( x_0 \in [1/2, 1] \), then \( R_k(y) = R_e(y) = 1 \) and there is nothing to show. Suppose \( x_0 \in [0, 1/2) \). We now want to show that the orbit of \( x_0 \) under \( S_k \) hits \([1/2, 1] \) before the orbit of \( x_0 \) under \( S_e \) does. Since \( S_k, S_e \) are increasing functions, it suffices to show that for any \( x \in [0, 1/2) \) we have \( S_k x \geq S_e x \). This follows easily since \( 2^k x^k \geq 2^e x^e \). Therefore \( \tau_k \leq \tau_e \).

For the second condition, we show (4.1') holds and apply Proposition 4.0.2. Note that for any \( x_0 \in [0, 1] \times [0, 1] \) we have

\[
\left| \operatorname{diam} \left( v(T^n W^s(x_0)) \right) \right| = \sup_{x, x' \in W^s(x_0)} |v(T^n x) - v(T^n x')| \\
\leq |v|_C \sup_{x, x' \in W^s(x_0)} d(T^n x, T^n x') \quad (4.4)
\]

Due to the product structure, the contraction on stable leaves does not depend on the first coordinate of \( T = T_{ek} \), so we can replace \( T \) on the right hand side with \( T_{kk} \).
to obtain

\[ \left| \text{diam} \left( v(T^n W^s(x_0)) \right) \right| \leq |v|_{C^0} \sup_{x, x' \in W^s(x_0)} d(T^n_{kk} x, T^n_{kk} x') \] (4.5)

Since \( T^n_{kk} \) contracts stable leaves exponentially, we see (4.1') holds and so

\[ \|E_0(v \circ T^n) - v \circ T^n\|_1 \leq Cn^{-1/k + 1}. \]

\[ \square \]

### 4.1 A Martingale Coboundary Decomposition

For \( \ell > k \geq 0 \) define the forwards part of the transfer function

\[ \chi^\ell_k := \sum_{n=k}^{\ell} E_0(v \circ T^n) - v \circ T^n, \]

and for \( \ell > k \geq 1 \) define the backwards part as

\[ \chi^{-\ell}_{-k} := \sum_{n=-k}^{-\ell} E_0(v \circ T^n). \]

Now, for \( k < 0 < \ell \) we define \( \chi^\ell_k := \chi^{-1}_k + \chi^\ell_0 \). By (A\text{inv}), \( \chi := \chi^{-\infty}_0 \) is well-defined and lies in \( L^1(\Lambda) \).

**Proposition 4.1.1.** We have the following martingale coboundary decomposition:

\[ v = m + \chi \circ T - \chi, \]

where \( m \in L^1(\Lambda) \), \( m \) is \( A_0 \)-measurable, and \( E_1 m = 0 \).

**Proof.** Denoting \( V_n := v \circ T^n \), define

\[ m := V_0 + \chi - \chi \circ T = V_0 + \chi_0^\infty - \chi_0^\infty \circ T + \chi^{-1}_{-\infty} - \chi^{-1}_{-\infty} \circ T \]

\[ = V_0 + \sum_{n=0}^{\infty} E_0 v_n - v_n - (E_0 v_n) \circ T + V_{n+1} + E_0 V_{n-1} - (E_0 V_{n-1}) \circ T \]

\[ = \sum_{n=0}^{\infty} E_0 v_n - E_1 V_{n+1} + E_0 V_{n-1} - E_1 V_n. \]

It is clear that \( m \in L^1(\Lambda) \) since \( v, \chi \in L^1(\Lambda) \). Since \( A_1 \subseteq A_0 \), we also see that \( m \) is
$\mathcal{A}_0$-measurable. It remains to show that $\mathbb{E}_1 m = 0$. This follows from the calculation

$$E_1 m = \mathbb{E}_1 \left( \sum_{n=0}^{\infty} E_0 V_n - E_1 V_{n+1} + E_0 V_{-n-1} - E_1 V_{-n} \right) = 0.$$ 

An important hypothesis in the martingale limit theorems that we apply is that the random variables are square integrable. The next two propositions introduce an ‘intermediate’ truncated decomposition, and then use this truncated decomposition to show $m \in L^2(\Lambda)$.

**Proposition 4.1.2.** For each $k \geq 1$, $v$ has the decomposition

$$v = m^{(k)} + \chi^{-k} - \chi^{-k} \circ T - E_0 (v \circ T^{-k}) - E_0 (v \circ T^{k+1}) + v \circ T^{k+1},$$

where $m^{(k)} \in L^\infty(\Lambda)$, $m^{(k)}$ is $\mathcal{A}_0$-measurable, and $\mathbb{E}_1 m^{(k)} = 0$. Moreover, $m^{(k)} \to m$ in $L^1(\Lambda)$.

**Proof.** Denote $V_n := v \circ T^n$. For $k \geq 1$ define

$$m^{(k)} := V_0 - V_{k+1} + \sum_{n=0}^{k} (E_0 V_n - V_n - (E_0 V_n) \circ T + V_{n+1})$$

$$+ \sum_{n=1}^{k} (E_0 V_{-n} - (E_0 V_{-n}) \circ T) + E_0 V_{k+1} - E_0 V_k$$

$$= E_0 V_{k+1} - E_0 V_k + \sum_{n=0}^{k} (E_0 V_n - E_1 V_{n+1}) + \sum_{n=1}^{k} (E_0 V_{-n} - E_1 V_{-n+1}). \quad (4.6)$$

It follows that $m^{(k)}$ is $\mathcal{A}_0$-measurable. To see $m^{(k)} \in L^\infty(\Lambda)$, use the fact that $\|\chi^{-k}\|_\infty \leq (3k + 2)\|v\|_\infty$ to calculate

$$\|m^{(k)}\|_\infty \leq \|V_0 - E_0 V_{-k} + E_0 V_{k+1} - V_{k+1}\|_\infty + (6k + 4)\|v\|_\infty \leq (6k + 8)\|v\|_\infty < \infty.$$
Next, $E_1 m^{(k)} = 0$ follows from the fact that $E_1 E_0 = E_1$ and (4.6), since

$$E_1 m^{(k)} = E_1 \left( E_0 V_{k+1} - E_0 V_{-k} + \sum_{n=0}^{k} (E_0 V_n - E_1 V_{n+1}) + \sum_{n=1}^{k} (E_1 V_n - E_1 V_{n+1}) \right)$$

$$= E_1 V_{k+1} - E_1 V_{-k} + \sum_{n=0}^{k} (E_1 V_n - E_1 V_{n+1}) + \sum_{n=1}^{k} (E_1 V_n - E_1 V_{n+1})$$

$$= E_1 V_{k+1} - E_1 V_{-k} + E_1 V_0 - E_1 V_{1} + E_1 V_{-k} - E_1 V_0 = 0.$$

Finally, we see $m^{(k)} \rightarrow m$ in $L^1(\Lambda)$ by calculating

$$\|m - m^{(k)}\|_1 = \left\| \chi - \chi^k - \chi \circ T + \chi^k \circ T + E_0(V_{-k}) - (E_0(V_{k+1}) - V_{k+1}) \right\|_1$$

$$= \left\| \chi^{k-1} + \chi_{k+1} - \chi^{k-1} \circ T - \chi^{k+1} \circ T + E_0(V_{-k}) - (E_0(V_{k+1}) - V_{k+1}) \right\|_1$$

$$\leq \left\| \chi^{k-1} - \chi^{k-1} \circ T \right\|_1 + \left\| \chi_{k+1} - \chi^{k+1} \circ T \right\|_1 \leq 2\|\chi^k\|_{L^1} + 2\|\chi_{\infty}\|_{L^1}$$

$$\leq 2 \sum_{n=k}^{\infty} (\|E_0(v \circ T^{-n})\|_1 + \|E_0(v \circ T^n)\|_1 - v \circ T^n) \rightarrow 0.$$

\[\square\]

**Proposition 4.1.3.** We have $m \in L^2(\Lambda)$ and $m^{(k)} \rightarrow m$ in $L^2(\Lambda)$.

**Proof.** By Proposition 4.1.2, $m^{(k)} \rightarrow m$ in $L^1(\Lambda)$. Therefore by uniqueness of limits in $L^1(\Lambda)$ it suffices to show that $m^{(k)}$ is a Cauchy sequence in $L^2(\Lambda)$. For clarity denote $V_n := v \circ T^n$ and

$$a_n = \begin{cases} E_0 V_n - V_n, & n \geq 0 \\ E_0 V_n, & n < 0 \end{cases}.$$  

Assume without loss that $\ell > k > 0$. With this notation, we have

$$m^{(\ell)} = v - \chi^\ell \circ T + \chi_{-\ell} a_{-\ell} + a_{\ell+1},$$

and $\chi^k = \sum_{j=-k}^{k} a_j$. Calculate that

$$m^{(\ell)} - m^{(k)} = (\chi^\ell - \chi^k) - (\chi^\ell \circ T - \chi^k \circ T) + a_{\ell+1} - a_{-\ell} - (a_{k+1} - a_{-k})$$

$$= (\chi_{-\ell} - \chi^k) - (\chi_{-\ell} \circ T - \chi^k \circ T) + a_{\ell+1} - a_{-\ell} - (a_{k+1} - a_{-k})$$

$$= (\chi^{k-1} + \chi_{k+1} - \chi^{k-1} \circ T + \chi_{k+1} \circ T) + a_{\ell+1} - a_{-\ell} - (a_{k+1} - a_{-k})$$

$$= (\chi^{k-1} - \chi_{-\ell} - a_{-\ell} a_{k+1} + \chi_{k+1} \circ T + a_{\ell+1} - a_{k+1})$$

$$= (\chi^{k-1} - \chi_{-\ell} \circ T) + (\chi_{k+1} \circ T + a_{\ell+1} - a_{k+1}).$$
Hence
\[ \|m^{(\ell)} - m^{(k)}\|_2 \leq \|\chi_{-\ell+1}^k - \chi_{-\ell}^{-k-1} \circ T\|_2 + \|\chi_{k+1}^\ell - \chi_{k+1}^\ell \circ T\|_2. \quad (4.7) \]

We deal with these two terms separately. For the first term of (4.7), calculate that
\[
\|\chi_{-\ell+1}^{-k} - \chi_{-\ell}^{-k-1} \circ T\|_2^2 = \mathbb{E} \left( (\chi_{-\ell+1}^{-k} - \chi_{-\ell}^{-k-1} \circ T)^2 \right) \\
= \mathbb{E} \left( (\chi_{-\ell+1}^{-k} - 2\chi_{-\ell+1}^{-k-1} \circ T + (\chi_{-\ell}^{-k-1})^2 \right). \quad (4.8)
\]

Considering the cross term in (4.8), we see
\[
\mathbb{E} \left( \chi_{-\ell+1}^{-k} \chi_{-\ell}^{-k-1} \circ T \right) = \mathbb{E} \mathbb{E}_1 \left( \chi_{-\ell+1}^{-k} \chi_{-\ell}^{-k-1} \circ T \right) = \mathbb{E} \left( \mathbb{E}_1 \left( \chi_{-\ell+1}^{-k} \chi_{-\ell}^{-k-1} \circ T \right) \right).
\]
Noting \( \mathbb{E}_1 \left( \chi_{-\ell+1}^{-k} \right) = \mathbb{E}_1 \left( \sum_{j=-\ell+1}^{-k} \mathbb{E}_0 V_j \right) = \sum_{j=-\ell+1}^{-k} (\mathbb{E}_0 V_{j-1}) \circ T = \chi_{-\ell}^{-k-1} \circ T, \) it follows that the cross term in (4.8) is
\[
\mathbb{E} \left( \chi_{-\ell+1}^{-k} \chi_{-\ell}^{-k-1} \circ T \right) = \mathbb{E} \left( \chi_{-\ell}^{-k-1} \circ T \chi_{-\ell}^{-k-1} \circ T \right) = \mathbb{E} \left( (\chi_{-\ell}^{-k-1})^2 \right).
\]

Therefore returning to (4.8) we obtain
\[
\|\chi_{-\ell+1}^{-k} - \chi_{-\ell}^{-k-1} \circ T\|_2^2 = \mathbb{E} \left( (\chi_{-\ell+1}^{-k})^2 - (\chi_{-\ell}^{-k-1})^2 \right) \\
= \mathbb{E} \left( (\chi_{-\ell+1}^{-k} - \chi_{-\ell}^{-k-1}) (\chi_{-\ell+1}^{-k} + \chi_{-\ell}^{-k-1}) \right) \\
= \mathbb{E} \left( (\chi_{-\ell+1}^{-k} - \chi_{-\ell}^{-k-1}) (\chi_{-\ell+1}^{-k} + \chi_{-\ell}^{-k-1}) \right) \leq 2\|v\|_\infty \|\chi_{-\ell+1}^{-k} + \chi_{-\ell}^{-k-1}\|_1 \\
\ll \|\chi_{-\ell+1}^{-k}\|_1 + \|\chi_{-\ell}^{-k-1}\|_1 \leq 2 \sum_{n=-\ell}^{-k} \|\mathbb{E}_0 V_n\|_1 \to 0,
\]
as \( k, \ell \to \infty. \) It remains to show that the second term of (4.7) converges to zero.

Note that
\[
\|\chi_{k+2}^{\ell+1} - \chi_{k+1}^\ell \circ T\|_2^2 = \mathbb{E} \left( (\chi_{k+2}^{\ell+1} - \chi_{k+1}^\ell \circ T)^2 \right) \\
= \mathbb{E} \left( \chi_{k+2}^{\ell+1} (\chi_{k+2}^{\ell+1} - \chi_{k+1}^\ell \circ T) \right) - \mathbb{E} \left( \chi_{k+1}^\ell \circ T (\chi_{k+2}^{\ell+1} - \chi_{k+1}^\ell \circ T) \right) \\
= \mathbb{E} \left( \chi_{k+2}^{\ell+1} (\chi_{k+2}^{\ell+1} - \chi_{k+1}^\ell \circ T) \right) + \mathbb{E} \left( (\chi_{k+1}^\ell)^2 - \chi_{k+1}^\ell \circ T \chi_{k+2}^{\ell+1} \right). \quad (4.9)
\]

We claim that the first term of (4.9) is zero. Since \( \mathbb{E}_0 \chi_{k+2}^{\ell+1} = 0, \) the claim easily
follows by showing \( \chi_{k+1}^{\ell+1} - \chi_{k+1}^{\ell} \circ T \) is \( \mathcal{A}_0 \)-measurable. Indeed we have

\[
\chi_{k+2}^{\ell+1} - \chi_{k+1}^{\ell} \circ T = \sum_{j=k+1}^{\ell} \mathbb{E}_0 V_{j+1} - V_{j+1} - (\mathbb{E}_0 V_j) \circ T + V_{j+1} = \sum_j \mathbb{E}_0 V_{j+1} - (\mathbb{E}_0 V_j) \circ T,
\]

which proves the claim. Next, for the second term of (4.9), add and subtract \( \chi_{k+1}^{\ell} \chi_{k+2}^{\ell+1} \) to obtain

\[
\mathbb{E} \left( (\chi_{k+1}^{\ell})^2 - \chi_{k+1}^{\ell} \circ T \chi_{k+2}^{\ell+1} \right) = \mathbb{E} \left( (\chi_{k+1}^{\ell})^2 - \chi_{k+1}^{\ell} \chi_{k+2}^{\ell+1} + \chi_{k+1}^{\ell} \chi_{k+2}^{\ell+1} - \chi_{k+1}^{\ell} \circ T \chi_{k+2}^{\ell+1} \right)
\]

\[
= \mathbb{E} \left( \chi_{k+1}^{\ell} (\chi_{k+1}^{\ell} - \chi_{k+2}^{\ell+1}) + (\chi_{k+1}^{\ell} - \chi_{k+1}^{\ell} \circ T) \chi_{k+2}^{\ell+1} \right)
\]

\[
= \mathbb{E} \left( \chi_{k+1}^{\ell} (a_{k+1} - a_{\ell+1}) \right) + \mathbb{E} \left( (\chi_{k+1}^{\ell} - \chi_{k+1}^{\ell} \circ T) \chi_{k+2}^{\ell+1} \right).
\]

Equation (4.9) now becomes

\[
\|\chi_{k+2}^{\ell+1} - \chi_{k+1}^{\ell} \circ T\|^2 = \mathbb{E} \left( \chi_{k+1}^{\ell} (a_{k+1} - a_{\ell+1}) \right) + \mathbb{E} \left( (\chi_{k+1}^{\ell} - \chi_{k+1}^{\ell} \circ T) \chi_{k+2}^{\ell+1} \right)
\]

\[
\leq 4\|v\|_\infty \|\chi_{k+1}^{\ell}\|_1 + \mathbb{E} \left( (\chi_{k+1}^{\ell} - \chi_{k+1}^{\ell} \circ T) \chi_{k+2}^{\ell+1} \right).
\]

(4.10)

Next we focus on the second term of (4.10). Note that

\[
\chi_{k+1}^{\ell} \circ T - \chi_{k+1}^{\ell} = \sum_{j=k+1}^{\ell} ((\mathbb{E}_0 V_j) \circ T - \mathbb{E}_0 V_j - V_{j+1} + V_j)
\]

\[
= \sum_{j=k+1}^{\ell} ((\mathbb{E}_0 V_j) \circ T - \mathbb{E}_0 V_j) + V_{j+1} - V_{k+1} =: W + V_{\ell+1} - V_{k+1},
\]

where \( W \) is \( \mathcal{A}_0 \)-measurable. Since \( \mathbb{E}_0 \chi_{k+2}^{\ell+1} = 0 \) we see the second term of (4.10) is

\[
\mathbb{E} \left( (\chi_{k+1}^{\ell} \circ T - \chi_{k+1}^{\ell}) \chi_{k+2}^{\ell+1} \right) = \mathbb{E}(W \chi_{k+2}^{\ell+1}) + \mathbb{E} \left( (V_{\ell+1} - V_{k+1}) \chi_{k+2}^{\ell+1} \right)
\]

\[
= \mathbb{E} \left( (V_{\ell+1} - V_{k+1}) \chi_{k+2}^{\ell+1} \right).
\]

Therefore (4.10) is now

\[
\|\chi_{k+2}^{\ell+1} - \chi_{k+1}^{\ell} \circ T\|^2 \leq 4\|v\|_\infty \|\chi_{k+1}^{\ell}\|_1 + \mathbb{E} \left( (V_{k+1} - V_{\ell+1}) \chi_{k+2}^{\ell+1} \right)
\]

\[
\leq 4\|v\|_\infty \|\chi_{k+1}^{\ell}\|_1 + 2\|v\|_\infty \|\chi_{k+2}^{\ell+1}\|_1 \ll \sum_{n=k+1}^{\ell+1} \|\mathbb{E}_0 V_n - V_n\|_1 \to 0,
\]

61
as \( k, \ell \to \infty \). Hence \( m^{(k)} \) is a Cauchy sequence in \( L^2(\Lambda) \) and so \( m^{(k)} \to m \) in \( L^2(\Lambda) \) and \( m \in L^2(\Lambda) \).

**Proposition 4.1.4** (*L^2* orthogonality of martingale differences). Suppose \( f : \Lambda \to \mathbb{R} \) is an \( L^2(\Lambda) \) observable which is \( A_0 \)-measurable and \( \mathbb{E}_1 f = 0 \). Then

\[
\int f \circ T^i f \circ T^j d\mu = \delta_{ij} \int f^2 d\mu
\]

for every \( i, j \in \mathbb{Z} \), where \( \delta_{ij} \) is the Kronecker delta.

**Proof.** The case \( i = j \) is clear from invariance of \( T \). Without loss assume \( i > j \) and note \( f \circ T^{i-j} \) is \( A_1 \)-measurable. Now calculate

\[
\int f \circ T^i f \circ T^j d\mu = \int f f \circ T^{i-j} d\mu = \int \mathbb{E}_1 (f f \circ T^{i-j}) d\mu = \int \mathbb{E}_1 f f \circ T^{i-j} d\mu = 0,
\]

and the result is shown. \( \square \)

For square integrable, \( A_0 \)-measurable observables \( f : \Lambda \to \mathbb{R} \) such that \( \mathbb{E}_1 f = 0 \), Proposition 4.1.4 shows

\[
\left\| \sum_{j=0}^{n-1} f \circ T^j \right\|_2 = n^{1/2} \left\| f \right\|_2.
\]

**Preliminary Calculations**

The invertible limit theorems follow the same proof structure as their non-invertible counterparts, with only a few calculations being different. We do these calculations now for ease of exposition.

Recall that \( A_0 \) is a sub-sigma algebra of \( A \) that coarsens under \( T \): With the notation \( A_n = T^{-n} A_0 \), \( \mathbb{E}_n v = \mathbb{E}(v | A_n) \) we have \( \cdots \subseteq A_1 \subseteq A_0 \subseteq A_{-1} \subseteq \cdots \) and for any \( m, \ell \in \mathbb{Z} \),

\[
(\mathbb{E}_m v) \circ T^\ell = \mathbb{E}_{m+\ell}(v \circ T^\ell), \quad \mathbb{E}_m \mathbb{E}_\ell v = \mathbb{E}_{\max\{m,\ell\}} v.
\]

**Proposition 4.1.5.** We have that

\[
\lim_{k \to \infty} \sup_{n \geq 1} n^{-1/2} \left\| \left[ \mathbb{E}_0 (v \circ T^{-k}) \right]^{-1} \right\|_2 = 0.
\]

**Proof.** Since \( \mathbb{E}_0 (v \circ T^{-1}) \) plays the role that the transfer operator did in the non-invertible case, the proof used for the non-invertible analogue (Proposition 3.3.11)
carries over with a few minor changes. Fix \( n \geq 1 \) and consider the decomposition
\[
\mathbb{E}_0(v \circ T^{-k}) = r^{(k)} + \psi \circ T - \psi,
\]
where
\[
\psi = \sum_{\ell=1}^{\infty} \mathbb{E}_0(\mathbb{E}_0(v \circ T^{-k}) \circ T^{-\ell}) = \sum_{\ell} \mathbb{E}_{\ell+k} \mathbb{E}_k v \circ T^{-\ell-k} = \sum_{\ell} \mathbb{E}_{\ell+k} v \circ T^{-\ell-k} = \sum_{\ell=1}^{\infty} \mathbb{E}_0(v \circ T^{-\ell}).
\]

It follows that \( r^{(k)} \) is \( \mathcal{A}_0 \)-measurable and we see
\[
\mathbb{E}_1 r^{(k)} = 0
\]
since
\[
\mathbb{E}_1 r^{(k)} = \mathbb{E}_1(\mathbb{E}_0(v \circ T^{-k}) + \psi - \psi \circ T) = \mathbb{E}_1(v \circ T^{-k}) + \mathbb{E}_1 \left( \sum_{\ell=1}^{\infty} \mathbb{E}_0(v \circ T^{-\ell}) - \mathbb{E}_0(v \circ T^{-\ell}) \circ T \right) = \mathbb{E}_1(v \circ T^{-k}) + \sum_{\ell=k+1}^{\infty} \mathbb{E}_1(v \circ T^{-\ell}) - \mathbb{E}_1(v \circ T^{-\ell+1}) = 0.
\]

As \( \mathbb{E}_0(v \circ T^{-k}) \circ T^{-j} \) is \( \mathcal{A}_{-j} \)-measurable, we apply Rio’s inequality (Proposition 3.3.10) with the filtration \( \mathcal{A}_{-j}, \ j \geq 1 \). Denoting
\[
b_{j,\ell} = \mathbb{E}_0(v \circ T^{-k}) \circ T^{-j} \sum_{u=j}^{\ell} \mathbb{E}(\mathbb{E}_0(v \circ T^{-k}) \circ T^{-u} \mid \mathcal{A}_{-j})
\]
we have
\[
\left\| \mathbb{E}_0(v \circ T^{-k})_n \right\|_2^2 = \mathbb{E}_1(\mathbb{E}_0(v \circ T^{-k})_n)^2 \leq C \sum_{j=1}^{n} \max_{1 \leq \ell \leq n} b_{j,\ell}.
\]

Note that
\[
b_{j,\ell} \leq \|v\|_{\infty} \left\| \sum_{u=j}^{\ell} \mathbb{E}_{-j} \left( r^{(k)} + \psi \circ T - \psi \right) \circ T^{-u} \right\|_1
\]
\[
\leq \|v\|_{\infty} \left\| \sum_{u=j}^{\ell} \left( \mathbb{E}_{-j+u} r^{(k)} \right) \circ T^{-u} \right\|_1 + \|v\|_{\infty} \left\| \sum_{u=j}^{\ell} \left( \psi \circ T - \psi \right) \circ T^{-u} \right\|_1
\]
\[
\leq \|v\|_{\infty} \sum_{u=j}^{\ell} \left\| \mathbb{E}_{-j+u} r^{(k)} \right\|_1 + \|v\|_{\infty} \left\| \psi \circ T^{-j+1} - \psi \circ T^{-\ell} \right\|_1.
\]

For any \( p \geq 1 \) we have \( \mathbb{E}_p r^{(k)} = \mathbb{E}_p \mathbb{E}_1 r^{(k)} = 0 \), and so

63
\[ \sum_{u=j}^\ell \| \mathbb{E}_{-j+u} r^{(k)} \|_1 = \| \mathbb{E}_0 r^{(k)} \|_1 = \| r^{(k)} \|_1. \] Therefore
\[ b_{j,\ell} \leq \| v \|_\infty (\| r^{(k)} \|_1 + 2\| \psi \|_1). \]

Returning to (4.11) we now have
\[ \left\| \left( \mathbb{E}_0(v \circ T^{-k}) \right)_n \right\|_2^2 \ll \sum_{j=1}^n \max_{1 \leq \ell \leq n} \left( \| r^{(k)} \|_1 + 2\| \psi \|_1 \right) \]
\[ = n \left( \| \mathbb{E}_0(v \circ T^{-k}) - \psi \circ T + \psi \|_1 + 2\| \psi \|_1 \right) \]
\[ \leq n \left( \| \mathbb{E}_0(v \circ T^{-k}) \|_1 + 4 \left\| \sum_{\ell=k+1}^\infty \mathbb{E}_0(v \circ T^{-\ell}) \right\|_1 \right) \]
\[ \ll n \sum_{\ell=k}^\infty \| \mathbb{E}_0(v \circ T^{-\ell}) \|_1. \]

Hence
\[ \sup_n n^{-1/2} \left\{ \left( \mathbb{E}_0(v \circ T^{-k})_n \right)_n \right\}_2 \ll \left( \sum_{\ell=k}^\infty \| \mathbb{E}_0(v \circ T^{-\ell}) \|_1 \right)^{1/2} \rightarrow 0, \]
as \( k \rightarrow \infty. \)

The following proposition does not require the random variables to be adapted to any filtration.

**Proposition 4.1.6** ([11, Equation (3.4)]). For a sum \( S_n = \sum_{j=1}^n X_j \) of square integrable random variables we have
\[ \mathbb{E}((S_n^*)^2) \leq 4\mathbb{E}(S_n^2) - 4 \sum_{j=1}^n \mathbb{E}(X_j S_j^*). \]

\[ \square \]

**Proposition 4.1.7.** We have that
\[ \lim \sup_{k \rightarrow \infty} n^{-1/2} \left\{ \left( \mathbb{E}_0(v \circ T^k) - v \circ T^k \right)_n \right\}_2 = 0. \]

The proof of Proposition 4.1.7 is similar to an argument presented in [11, Proposition 4].

**Proof.** Let \( k \geq 1, \ V_j := v \circ T^j, \ X_j := \mathbb{E}_{-j-k}V_j - V_j \) and \( S_n := \sum_{j=1}^n X_j. \) Without loss assume that \( n > k + 1. \) Define \( S_n^* := \max\{0, S_1, \ldots, S_n\}, \ S_{n,\ast} := \)
max\{0, -S_1, \ldots, -S_n\} and recall that \( \overline{n} = \max_{0 \leq \ell \leq n} |S_\ell| \). We want to show
\[
\lim_{n \to \infty} n^{-1} \mathbb{E} \left( \overline{n}^2 \right) = 0.
\]

Since \( \overline{n}^2 = \left( \max\{S_1^*, S_n^*\}\right)^2 = \max\{(S_1^*)^2, (S_n^*)^2\} \leq (S_1^*)^2 + (S_n^*)^2 \), it suffices to show
\[
\lim_{n \to \infty} n^{-1} \left[ \mathbb{E} \left( (S_1^*)^2 \right) + \mathbb{E} \left( (S_n^*)^2 \right) \right] = 0.
\]

We first bound \( n^{-1} \mathbb{E} ((S_1^*)^2) \). By Proposition 4.1.6,
\[
\mathbb{E} \left( (S_1^*)^2 \right) \leq 4 \mathbb{E}(S_1^2) - 4 \sum_{j=1}^{n} \mathbb{E}(X_j S_{j-1})^* \leq 4 \mathbb{E}(S_1^2) + 4 \sum_{j=1}^{n} \mathbb{E}(X_j S_{j-1})^*.
\] (4.12)

Note that \( X_j = (E_{-k} v - v) \circ T^{-j} \). For the first term of (4.12), the calculations from the proof of Proposition 3.2.7 show
\[
\mathbb{E}(S_1^2) = \sum_{j=1}^{n} \int X_j^2 d\mu + 2 \sum_{r=1}^{n} (n - r) \int X_0 X_{-r} d\mu
\]
\[
\leq 2n||v||_\infty ||E_{-k} v - v||_1 + 2n \sum_{r=1}^{n} \int (E_{-k} v - v) (E_{-k} v - v) \circ T^r d\mu.
\] (4.13)

Next, note that \((E_{-k} v) \circ T^r\) is \(\mathcal{A}_{-k+r}\)-measurable and hence \(\mathcal{A}_{-k}\)-measurable. Similarly \((E_{-k} v) \circ T^r\) is \(\mathcal{A}_{-k}\)-measurable. Therefore
\[
\int (E_{-k} v - v) (E_{-k} v - v) \circ T^r d\mu = \int E_{-k} (E_{-k} v - v) (E_{-k} v - v) \circ T^r d\mu = 0.
\]

In particular, \(\int (E_{-k} v - v) (E_{-k} v) \circ T^r d\mu = \int (E_{-k} v - v) (E_{-k} v) \circ T^r d\mu\) and, returning to (4.13), we have
\[
\mathbb{E}(S_1^2) \ll n ||E_{-k} v - v||_1 + n \sum_{r=1}^{n} \left| \int (E_{-k} v - v) (E_{-k} v - v) \circ T^r d\mu \right|
\]
\[
\leq n ||E_{-k} v - v||_1 + 2 ||v||_\infty n \sum_{r=1}^{n} \|E_{-k} v - v\|_1 \ll n \sum_{r=1}^{\infty} ||E_{-k} v - v||_1.
\] (4.14)

Next, we bound the second term of (4.12). Define \(Y_{i,j} := E_{-k-i} V_{-i} - E_{-k-j} V_{-i}\) and \(T_{p,j} := \sum_{i=1}^{p} Y_{i,j}\). For every \(i < j\) we have \(\mathcal{A}_{-k-i} \subseteq \mathcal{A}_{-k-j}\) and so \(Y_{i,j}\) is \(\mathcal{A}_{-k-j}\)-measurable. Therefore \(T_{j-1}^* := \max\{0, T_{1,j}, \ldots, T_{j-1,j}\}\) is \(\mathcal{A}_{-k-j}\)-
measurable. Note that $E(X_j T^*_j) = E(E_{k-j}(X_j T^*_j)) = E(E_{k-j}(E_{k-j} V_{-j}) T^*_j) = 0$. Hence

$$
\sum_{j=1}^n |E(X_j S^*_j)| = \sum_j |E\left(X_j(S^*_j - T^*_j-1)\right)| \leq 2\|v\|_\infty \sum_j |S^*_j - T^*_j-1|.
$$

Now, by Proposition 2.2.5 we have

$$
|S^*_j - T^*_j-1| \leq \sum_{i=1}^{j-1} |X_i - Y_{i,j}| = \sum_i |E_{k-j} V_{-i} - V_{-i}|.
$$

Therefore we obtain

$$
\sum_{j=1}^n |E(X_j S^*_j)| \leq 2\|v\|_\infty \sum_{j=1}^{j-1} \sum_{i=1} E|E_{k-j} V_{-i} - V_{-i}| = 2\|v\|_\infty \sum_{1 \leq i < j \leq n} ||E_{k-j+i}v - v||_1
$$

$$
= 2\|v\|_\infty \sum_{r=1}^n (n - r) ||E_{k-r}v - v||_1
$$

(4.15)

Combining (4.12) with (4.14) and (4.15) we see that

$$
n^{-1}E\left((S_n^*)^2\right) \ll n^{-1}\left( n \sum_{r=k}^\infty ||E_{r}v - v||_1 + 2\|v\|_\infty \sum_{r=1}^n (n - r) ||E_{k-r}v - v||_1\right)
$$

$$
\ll \sum_{r=k}^\infty ||E_{r}v - v||_1.
$$

To obtain a bound for $(S_{n,*})^2$, we note that changing $X_j$ to $-X_j$ changes $S_n$ to $-S_n$ and $S_n^*$ to $S_{n,*}$. Therefore, the corresponding bound for $(S_{n,*})^2$ is also

$$
n^{-1}E\left((S_{n,*})^2\right) \ll \sum_{r=k}^\infty ||E_{r}v - v||_1.
$$

The result now follows since

$$
\sup_{n \geq 1} \left[n^{-1}E\left((S_n^*)^2\right) + n^{-1}E\left((S_{n,*})^2\right)\right] \ll \sum_{r=k}^\infty ||E_{r}v - v||_1 \to 0,
$$

as $k \to \infty$.

\[ \square \]

**Proposition 4.1.8.** We have that

$$
\lim_{n \to \infty} n^{-1/2} \left\|\left(\frac{v - m}{n}\right)\right\|_2 = 0.
$$
Proof. First note that
\[ n^{-1/2} \left\| (v - m)_n \right\|_2 \leq n^{-1/2} \left\| (m - m^{(k)})^-_n \right\|_2 + n^{-1/2} \left\| (v - m^{(k)})_n^- \right\|_2. \tag{4.16} \]

Note \((m - m^{(k)}) \circ T^{-j}\) forms a sequence of martingale differences by Proposition 4.2.1.

For the first term of (4.16), use Doob’s inequality (Proposition 3.3.9) and the orthogonality of martingale differences (Proposition 4.1.4) to see
\[ n^{-1/2} \left\| (m - m^{(k)})^-_n \right\|_2 \leq 2n^{-1/2} \left\| (m - m^{(k)})^-_n \right\|_2 = 2\|m - m^{(k)}\|_2. \]

Proposition 4.1.3 shows \(\|m - m^{(k)}\|_2 \to 0\) as \(k \to \infty\) and so it remains to bound the second term of (4.16). To do this, we use the truncated decomposition \(v = m^{(k)} + \chi_k^{k} \circ T - \chi_j^{k,j} + \mathbb{E}_0 (v \circ T^{-j}) - \mathbb{E}_0 (v \circ T^{k+1}) + v \circ T^{k+1}\) and take \(\limsup_{n \to \infty}\) to see
\[
\limsup_{n \to \infty} n^{-1/2} \left\| (v - m^{(k)})^-_n \right\|_2 \leq \limsup_{n \to \infty} n^{-1/2} \left\| (\mathbb{E}_0 (v \circ T^{k+1}) - v \circ T^{k+1})^-_n \right\|_2 \\
+ \limsup_{n \to \infty} n^{-1/2} \left\| (\mathbb{E}_0 (v \circ T^{-j}))^-_n \right\|_2 \\
+ \limsup_{n \to \infty} n^{-1/2} \left\| (\chi_k^{k} \circ T - \chi_j^{k,j})^-_n \right\|_2.
\]

The first two terms converge to zero as \(k \to \infty\) by Propositions 4.1.5 and 4.1.7. It remains to show
\[
\lim_{k \to \infty} \limsup_{n \to \infty} n^{-1/2} \left\| (\chi_j^{k} \circ T - \chi_k^{k,j})^-_n \right\|_2 = 0.
\]

For any \(k \geq 1\) we have \(|\chi_j^{k,j}| \leq (3k + 2)\|v\|_\infty\), and so
\[
\limsup_{n \to \infty} n^{-1/2} \left\| (\chi_j^{k} \circ T - \chi_k^{k,j})^-_n \right\|_2 = \limsup_{n \to \infty} n^{-1/2} \left\| \max_{1 \leq \ell \leq n} \sum_{j=1}^{\ell} (\chi_j^{k,j} \circ T - \chi_j^{k,j}) \circ T^{-j} \right\|_2 \\
= \limsup_{n \to \infty} n^{-1/2} \left\| \max_{\ell} |\chi_j^{k} - \chi_k^{k,j} \circ T^{-\ell}| \right\|_2 \leq \limsup_{n \to \infty} n^{-1/2} (6k + 4)\|v\|_\infty = 0.
\]

\[\square\]

4.2 Central Limit Theorem

Proposition 4.2.1. Suppose \(f \in L^1(\Lambda)\) is \(\mathcal{A}_0\)-measurable and \(\mathbb{E}_1 f = 0\). Then \(f_n^- = \sum_{j=1}^{n} f \circ T^{-j}\) is a martingale with respect to the filtration \(\mathcal{A}_{-n} = T^n \mathcal{A}_0\).
Proof. Firstly we note that \( \|f_n^-\|_1 \leq n\|f\|_1 < \infty \). Secondly, \( f_n^- \) is \( A_{-n} \)-measurable since \( f \circ T^{-j} \) is \( A_{-j} \)-measurable and \( A_{-j} \subseteq A_{-n} \) for all \( j \leq n \). Finally, we have

\[
\mathbb{E}(f \circ T^{-j} | A_{-j+1}) = \mathbb{E}_{-j+1}(f \circ T^{-j}) = \mathbb{E}_1(f) \circ T^{-j} = 0,
\]

so \( f \circ T^{-j} \) forms a martingale difference sequence and \( f_n^- \) is a martingale with respect to \( A_{-n} \).

Proposition 4.2.2 (Variance). We have that

\[
\lim_{n \to \infty} \frac{1}{n} \int v_n^2 \, d\mu = \int m^2 \, d\mu.
\]

Proof. By orthogonality of martingale differences (Proposition 4.1.4) and Proposition 4.1.8 we have

\[
\left| n^{-1/2}\|v_n\|_2 - \|m\|_2 \right| = n^{-1/2}\|v_n\|_2 - n^{-1/2}\|m_n\|_2 \leq n^{-1/2}\|(v - m)_n\|_2 \\
= n^{-1/2}\|(v - m)_n \circ T^{-n}\|_2 = n^{-1/2}\|(v - m)_n^-\|_2 \\
\leq n^{-1/2}\|(v - m)_n^-\|_2 \to 0.
\]

Theorem 4.2.3 (Scalar CLT). We have

\[
n^{-1/2}v_n \to_d N(0, \sigma^2),
\]

where \( \sigma^2 := \lim_{n \to \infty} n^{-1} \int v_n^2 \, d\mu \). Moreover, \( \sigma^2 = 0 \) if and only if \( v \) is an \( L^1(\Lambda) \) coboundary.

Remark 1. This result is not new; Instead of Assumption \( (A_{inv}) \), Liverani [32, Theorem 1.2] assumes the conditions

\[
(i) \sum_{n=0}^{\infty} \left| \int v \circ T^n \, d\mu \right| < \infty,
\]

\[
(ii) \sum_{n=0}^{\infty} \left| \mathbb{E}_0(v \circ T^{-n}) \right| \text{ converges in } L^1(\Lambda),
\]

\[
(iii) \text{ There exists } \alpha > 1 \text{ such that } \sup_{k \geq 1} k^{\alpha} \|\mathbb{E}_0(v \circ T^n) - v \circ T^n\|_1 < \infty.
\]

Remark 2. Zweimüller [54, Corollary 1] proved that the CLT also holds with respect to every probability measure \( \nu \) which is absolutely continuous with respect to \( \mu \).
Proof of Theorem 4.2.3. We see $m_n^-$ is a martingale due to Proposition 4.2.1. Proposition 3.2.6 now shows $n^{-1/2}m_n^- \rightarrow_d \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \int m^2 \, d\mu$. Proposition 4.2.2 shows $\sigma^2 = \lim_n n^{-1} \int v_n^2 \, d\mu$. The CLT for $n^{-1/2}v_n$ now follows since $m_n^- = m_n \circ T^{-n} = m_n$ and $n^{-1/2}\|v_n - m_n\|_1 = n^{-1/2}\|\chi \circ T^n - \chi\|_1 \leq 2n^{-1/2}\|\chi\|_1 \rightarrow 0$ as $n \rightarrow \infty$. We easily see $\sigma^2 = \|m\|_2^2 = 0$ if and only if $m = 0$, so $v$ is an $L^1(\Lambda)$ coboundary.

Example 4.2.4. Suppose $0 < \alpha < 1/2$, $\Lambda = [0,1] \times [0,1]$ and let $T_{\alpha} : \Lambda \rightarrow \Lambda$ be the intermittent Baker’s map defined in (4.3). Let $v : \Lambda \rightarrow \mathbb{R}$ be a Hölder continuous and mean zero observable. Proposition 4.0.5 shows

$$\|E_0(v \circ T_{\alpha}^{-n})\|_1 \leq Cn^{-1/\alpha+1}, \quad \|E_0(v \circ T_{\alpha}^n) - v \circ T_{\alpha}^n\|_1 \leq Cn^{-1/\alpha+1},$$

and so assumption $(A_{inv})$ is satisfied for all $0 < \alpha < 1/2$. The CLT now follows from Theorem 4.2.3.

As in the non-invertible case, Assumption $(A_{inv})$ and the decompositions in Proposition 4.1.1 and 4.1.2 easily carry over to vector-valued observables $v \in L^\infty(\Lambda, \mathbb{R}^d)$ by considering them component wise.

Proposition 4.2.5 (Covariance). For $v \in L^\infty(\Lambda, \mathbb{R}^d)$ we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int v_n v_n^T \, d\mu = \int m m^T \, d\mu.$$

Proof. The proof for non-invertible maps (Proposition 3.2.14) carries over verbatim after replacing the non-invertible scalar variance (Proposition 3.2.8) with its invertible counterpart (Proposition 4.2.2).

Theorem 4.2.6 (Vector CLT). We have

$$n^{-1/2}v_n \rightarrow_d \mathcal{N}(0, \Sigma),$$

where $\Sigma := \lim_{n \rightarrow \infty} n^{-1} \int v_n v_n^T \, d\mu$. For the degenerate case, $\det \Sigma = 0$ if and only if there exists a non-zero $c \in \mathbb{R}^d$ such that $c^Tv = \chi \circ T - \chi$ for some $\chi \in L^1(\Lambda)$.

Proof. The covariance $\Sigma = \int m m^T \, d\mu = \lim_{n \rightarrow \infty} n^{-1} \int v_n v_n^T \, d\mu$ is now due to Proposition 4.2.5. For any $c \in \mathbb{R}^d$, we see $c^Tv \in L^\infty(\Lambda)$, $\int c^Tv \, d\mu = 0$ and $\sum_{j \geq 0} \left\|E_0((c^Tv) \circ T^{-j-1})\right\|_1 + \left\|E_0((c^Tv) \circ T^j) - (c^Tv) \circ T^j\right\|_1 < \infty$. Therefore the scalar CLT (Theorem 4.2.3) holds for $n^{-1/2}(c^Tv)_n$ and the proof of the non-invertible vector CLT (Theorem 3.2.12) carries over.

69
4.3 Weak Invariance Principle

Define the forward càdlàg process

\[ W_n(t) := n^{-1/2} \sum_{j=0}^{[nt]-1} v \circ T^j, \]

and the backwards càdlàg processes

\[ W_n^-(t) := n^{-1/2} \sum_{j=1}^{[nt]} v \circ T^{-j}, \quad M_n^-(t) := n^{-1/2} \sum_{j=1}^{[nt]} m \circ T^{-j}. \]

**Proposition 4.3.1.** For each \( n \geq 1 \), \( M_n^- : [0, 1] \to \mathbb{R}^d \) is a martingale with respect to the filtration \( \mathcal{A}_{n,t} = \{ T^{[nt]} \mathcal{A}_0 \mid 0 \leq t \leq 1 \} \).

**Proof.** Let \( n \geq 1 \). Firstly, \( M_n^-(t) \in L^1(\Lambda, \mathbb{R}^d) \) since

\[ \| M_n^-(t) \|_1 = n^{-1/2} \left\| \sum_{j=1}^{[nt]} m \circ T^{-j} \right\|_1 \leq \frac{[nt]}{n^{1/2}} \| m \|_1 < \infty. \]

Secondly, \( M_n^- \) is \( \mathcal{A}_{n,t} \)-measurable since \( m \circ T^{-j} \) is \( \mathcal{A}_{-j} \)-measurable and \( \mathcal{A}_{-j} \subseteq \mathcal{A}_{n,t} \) for all \( j \leq [nt] \). It remains to show that

\[ \mathbb{E}(M_n^-(t) \mid \mathcal{A}_{n,s}) = M_n^-(s), \]

for all \( s \leq t \). If \( [nt] = [ns] \), then we have \( M_n^-(t) = M_n^-(s) \) and we are finished.

Suppose \( [nt] > [ns] \) and note that

\[ \mathbb{E}(n^{-1/2} \sum_{j=1}^{[nt]} m \circ T^{-j} \mid \mathcal{A}_{n,s}) = M_n^-(s) + n^{-1/2} \sum_{j=[ns]+1}^{[nt]} \mathbb{E}(m \circ T^{-j} \mid T^{[ns]} \mathcal{A}_0). \]

Let \( j = [ns] + 1, \ldots, [nt] \) be fixed, and calculate that

\[ \mathbb{E}(m \circ T^{-j} \mid T^{[ns]} \mathcal{A}_0) = \mathbb{E}(m \mid T^{[ns]-j} \mathcal{A}_0) \circ T^{-j} = (\mathbb{E}_{j-[ns]} m_{\Lambda}) \circ T^{-j} = 0. \]

Therefore \( \mathbb{E}(M_n^-(t) \mid \mathcal{A}_{n,s}) = M_n^-(s) \) and \( M_n^- \) is a martingale.

**Theorem 4.3.2 (Invertible WIP).** We have

\[ W_n \xrightarrow{w} W, \]

70
in \((D([0,1],\mathbb{R}^d), \| \cdot \|_{\infty})\), where \(W\) is a \(d\) dimensional Brownian Motion with mean zero and covariance \(\Sigma = \lim_{n \to \infty} n^{-1} \int v_n v_n^T d\mu\).

**Remark 1.** This was first proven by Dedecker and Rio [11, Corollary 4(b)].

**Remark 2.** As in Theorem 4.2.6, \(\det \Sigma = 0\) if and only if there exists a non-zero \(c \in \mathbb{R}^d\) such that \(c^T v\) is an \(L^1(\Lambda)\) coboundary.

**Remark 3.** We could equally define our processes in \(D([0,\infty),\mathbb{R}^d)\) for any integer \(K \geq 1\), and the result still holds. Hence the result is true if we define our processes in \(D([0, \infty), \mathbb{R}^d)\) by [7, Section 16].

**Remark 4.** Zweimüller [54, Corollary 3] proved that the WIP also holds with respect to every probability measure \(\nu\) which is absolutely continuous with respect to \(\mu\).

**Proof of Theorem 4.3.2.** By Proposition 4.3.1, \(M_n\) is a martingale. Since \(T\) is invertible, lifting is not needed and Proposition 3.3.8 carries over to the invertible case by simply dropping the tilde notation. Therefore we have \(M_n \to_w W\) in \((D([0,1],\mathbb{R}^d), \| \cdot \|_{\infty})\), where \(W\) is a \(d\) dimensional Brownian motion with mean zero and covariance \(\Sigma = \int m m^T d\mu\). Proposition 4.2.5 shows \(\Sigma = \lim_{n} n^{-1} \int v_n v_n^T d\mu\). Proposition 4.1.8 extends to vector-valued observables since for \(f = (f_1, \ldots, f_d)\) we have

\[
\|f\|_2 = \|f\|_{L^2(\Lambda,\mathbb{R})} = \left\| \sum_{i=1}^d |f_i| \right\|_{L^2(\Lambda,\mathbb{R})} \leq \sum_{i} \|f_i\|_2.
\]

Therefore \(\lim_{n \to \infty} n^{-1/2} \left\| (v - m)_{\pi}^{-n} \right\|_2 = 0\) and

\[
\sup_{t \in [0,1]} |W_n^{-}(t) - M_n^{-}(t)| = n^{-1/2} \max_{1 \leq \ell \leq n} \left\| \sum_{j=1}^\ell (v - m) \circ T^{-j} \right\| = n^{-1/2} (v - m)_{\pi}^{-n} \to_p 0.
\]

This gives the convergence \(W_n^{-} \to_w W\). Now, the proof of the non-invertible WIP (Theorem 3.3.3) carries over by dropping the tilde notation and we see \(W_n \to_w W\) as required.

**Example 4.3.3.** Suppose \(0 < \alpha < 1/2\), \(\Lambda = [0,1] \times [0,1]\) and let \(T_{\alpha \alpha} : \Lambda \to \Lambda\) be the intermittent Baker’s map defined in (4.3). Let \(v : \Lambda \to \mathbb{R}\) be a Hölder continuous and mean zero observable. Proposition 4.0.5 shows

\[
\|E_0(v \circ T_{\alpha \alpha}^{-n})\|_1 \leq C n^{-1/\alpha + 1}, \quad \|E_0(v \circ T_{\alpha \alpha}^{-n}) - v \circ T_{\alpha \alpha}^{-n}\|_1 \leq C n^{-1/\alpha + 1},
\]

and so assumption \((A_{inv})\) is satisfied for all \(0 < \alpha < 1/2\). The WIP now follows from Theorem 4.3.2.
4.4 Iterated Weak Invariance Principle

Recall the standing assumptions from the start of the chapter: Suppose $T : \Lambda \to \Lambda$ is an invertible, ergodic and measure-preserving transformation over the probability space $(\Lambda, \mathcal{A}, \mu)$ and $v \in L^\infty(\Lambda, \mathbb{R}^d)$ is a mean zero observable. In this section, we replace Assumption (A_{inv}) with the following stronger one:

\[ \sum_{n=0}^\infty \| \mathbb{E}_0(v \circ T^n) - v \circ T^n \|_2 < \infty, \quad \sum_{n=1}^\infty \| \mathbb{E}_0(v \circ T^{-n}) \|_1 < \infty. \quad (A_{\text{strong}}) \]

**Theorem 4.4.1** (Invertible Iterated WIP). Assume that $T : \Lambda \to \Lambda$ is also mixing. We have that

\[ (W_n, \mathcal{W}_n) \to_w (W, \mathcal{W}) \]

in $\left( D([0,1], \mathbb{R}^d \times \mathbb{R}^{d d}), \| \cdot \|_\infty \right)$, where $W$ is a Brownian motion with mean zero and covariance $\Sigma = \lim_{n \to \infty} n^{-1} \int v_n v_n^T d\mu$, and

\[ \mathcal{W}(t) = \int_0^t W \otimes dW + t \sum_{j=1}^\infty \int v(v \circ T^j)^T d\mu, \]

for $t \in [0,1]$.

**Remark 1.** This iterated WIP is already known by Kelly and Melbourne [29, Theorem 5.2] under the more restrictive assumption

\[ \sum_{n=0}^\infty \| \mathbb{E}_0(v \circ T^n) - v \circ T^n \|_2 < \infty, \quad \sum_{n=1}^\infty \| \mathbb{E}_0(v \circ T^{-n}) \|_1 < \infty. \]

**Remark 2.** If $T$ is modelled by a Young tower with polynomial tails $\mu(r > n) \leq Cn^{-\beta}$ for some $\beta > 2$ (and so assumption (A) holds by Theorem 4.0.4), then the iterated WIP is already known by Melbourne and Varandas [40, Corollary 2.3].

**Remark 3.** As in Theorem 3.2.12, $\det \Sigma = 0$ if and only if there exists a non-zero $c \in \mathbb{R}^d$ such that $c^T v$ is an $L^1(\Lambda)$ coboundary.

**Remark 4.** We could equally define our processes in $D([0, K], \mathbb{R}^d)$ for any integer $K \geq 1$, and the result still holds. Hence the result is true if we define our processes in $D([0, \infty), \mathbb{R}^d)$ by [7, Section 16].

**Remark 5.** It is still not known if the iterated WIP holds under the weaker assumption

\[ \sum_{n=0}^\infty \| \mathbb{E}_0(v \circ T^n) - v \circ T^n \|_1 < \infty, \quad \sum_{n=1}^\infty \| \mathbb{E}_0(v \circ T^{-n}) \|_1 < \infty. \]
Next we give an example where the iterated WIP holds, but is not covered by [29, 40]. Similar to the previous chapter, we first outline a generic situation where the mixing conditions for a suitable suspension is inherited from the underlying map which is modelled by a Young tower.

### 4.4.1 Suspensions and Examples

Suppose $T : \Lambda \to \Lambda$ is an invertible, ergodic and measure-preserving map over the probability space $(\Lambda, A, \mu)$, and that $T$ has a stable foliation $W^s_T = \{ W^s_T(x) : x \in \Lambda \}$. Define the suspension flow $T^t : \Lambda_h \to \Lambda_h$ as in (3.15) over an $\eta$-Hölder roof function $h$ such that $\inf h > 0$. Further assume that the roof function $h$ is constant on local stable leaves for $T$.

**Proposition 4.4.2.** For any $t > 0$, the map $T_t$ admits a partition into stable leaves given by $W^s(x,u) := W^s_T(x) \times \{ u \}$.

**Proof.** For any $(x,u) \in \Lambda_h$, define $W^s(x,u) := W^s_T(x) \times \{ u \}$. Clearly this forms a partition of $\Lambda_h$. Let $N = N(x,u,t)$ be the lap number, i.e. the integer $N$ satisfying $u + t \in [h_N(x), h_{N+1}(x))$. Then $T_t(x,u) = (T^N x, u + t - h_N(x))$. Since $h$ is constant on stable leaves, $h(T^j x) = h(T^j x)$ and $h_n(x') = h_n(x)$ for every $x' \in W^s_T(x)$ and $n \geq 1$. Therefore $N(x',u,t) = N(x,u,t)$. Suppose that $(x',u) \in W^s(x,u) = W^s_T(x) \times \{ u \}$ and $t > 0$. This partition is invariant under $T_t$ since

$$T_t(x',u) = (T^N x', u + t - h_N(x')) \in W^s_T(T^N x) \times \{ u + t - h_N(x) \}$$

$$= W^s((T^N x, u + t - h_N(x)))$$

$$= W^s(T_t(x,u)).$$

Moreover, we have

$$d(T_t(x',u), T_t(x,u)) = d(T^N(x',u,t)x', T^N(x,u,t)x) + |u + t - h_N(x') - (u + t - h_N(x))|$$

$$= d(T^N(x,u,t)x', T^N(x,u,t)x) \to 0,$$

since $x' \in W^s_T(x)$. 

### Verifying Assumption (A\textunderscore strong)

**Proposition 4.4.3 ([12]).** Let $v : \Lambda^h \to \mathbb{R}$ be a Hölder continuous, mean zero observable. Suppose $T$ satisfies the hypothesis of Proposition 4.0.2, and assume
inf \gamma h > 1 \text{ and } h \in L^\infty(\Lambda). \text{ Then}

\|E_0(v \circ T^n) - v \circ T^n\|_1 \leq Cn^{-\alpha+1}.

\square

Remark. Roughly speaking, we define our inducing set for \(T_1\) as \(Y^h = \{(y, u) : y \in Y, u \in [0, h(y)]\}\). The assumption \(\inf \gamma h > 1\) ensures that the first return \(\tau^h\) of \(T_1\) to \(Y^h\) has tails comparable to the first return \(\tau\) of \(T\) to \(Y\). Without this assumption, \(h\) can be very small on \(Y\) and the time one map may ‘pass over’ the inducing set \(Y^h\), meaning we cannot control the tail of \(\tau^h\). We can drop the assumption \(\inf h > 1\) if we replace the time-one map \(T_1\) with the time-(\(\inf h - \epsilon\)) map \(T_{\inf h - \epsilon}\), for any \(\epsilon > 0\).

Proof. We verify the hypothesis of Proposition 4.0.2 for \(T_1\). Define \(Y^h := \{(y, u) : y \in Y, u \in [0, 1]\} \subset \Lambda^h\). By Proposition 4.4.2, \(T\) admits a stable foliation and \(Y^h\) is union of stable leaves for \(T_1\). Let \(\tau^h\) be the first return of \(T_1\) to \(Y^h\). It remains to show that \(\mu^h(\tau^h > n) \leq Cn^{-\alpha}\) and condition (4.1).

For the first part, note that

\[\tau^h(y, u) \leq \tau^h(y, 0) \leq \sum_{j=0}^{\tau(y)-1} h(T^jy) + 1 \leq \|h\|_\infty \tau(y) + 2.\]

Therefore

\[\mu^h(\tau^h > n) \leq \mu(\|h\|_\infty \tau + 2 > n) \leq Cn^{-\alpha}.\]

For the second part, since \(v\) is Hölder, it suffices to show \(T_1^{\tau^h}\) contracts stable leaves exponentially (see equation (4.4)). Let \((x, u) \in \Lambda^h\) and define the lap number \(N = N(x, u, t)\) as the integer such that \(u + t \in [h_N(x), h_{N+1}(x)]\). Then \(T_1^{\tau^h}(x, u) = T_{\tau^h}(x, u) = (T^N x, u + t - h_N(x)) = (T^\tau x, u + t - h_N(x)) = (Fx, u + t - h_N(x)).\) Let \(n \geq 1\) and note that \((T_1^{\tau^h})^n(x, u) = (F^n x, g(x, u, n))\) for some \(g(x, u, n) \in [0, 1]\). Therefore for any \((x', u) \in W^s(x, u)\) we have

\[d(T_1^{\tau^h}(x', u), T_1^{\tau^h}(x, u)) = d(F^n x', F^n x) + |g(x', u, n) - g(x, u, n)| \leq C\gamma^n + 1 = \tilde{C}\gamma^n,
\]

and so the induced map \(T_1^{\tau^h}\) contracts stable leaves exponentially as desired. \(\square\)

Define \(C^{n,m}(M) := \{w : M \to \mathbb{R} : \|w\|_{C^{n,m}} := \sum_{j=0}^m \|\partial_i^j w\|_{C^n} < \infty\}\).

**Proposition 4.4.4.** Suppose \(T : \Lambda \to \Lambda\) is modelled by a Young tower with polynomial tail \(\tau\), so \(\mu(\tau > n) \leq Cn^{-\beta}\) for some \(\beta > 1\). Assume condition (3.14) is

74
satisfied and absence of approximate eigenfunctions holds for $T$ [37, Definition 3.4].
Suppose $v \in C^n(\Lambda^h) \cap C^{0,\eta}(\Lambda^h)$ is a mean zero observable. Then,
\[
\|E_0(v \circ T^{-n})\|_1 \leq Cn^{-\beta}.
\]

Proof. [2, Corollary 8.1] guarantees an $m \geq 1$ and $C > 0$ such that
\[
\left| \int v w \circ T^t \mu^h \right| \leq C \left( \|v\|_{C^0} + |v|_{C^{0,\eta}} \right) \|w\|_{C^{0,m}t^{-\beta}},
\]
for every $w \in C^{n,m}(\Lambda^h)$. Therefore $\left| \int v w \circ T^n \mu^h \right| \leq C \|w\|_{C^{0,m}n^{-\beta}}$. Denote the set of continuous functions, constant on stable leaves, by $C^0(A_0)$. Note that $C^0(A_0) \subset C^{0,m}(\Lambda^h)$, which shows $\left| \int v w \circ T^n \mu^h \right| \leq C \|w\|_{C^0(A_0)n^{-\beta}}$ for every $w \in C^0(A_0)$. Let $w' \in L^\infty(A_0)$. For any $\epsilon > 0$ there exists $w \in C^0(A_0)$ such that $\|w\|_{C^0(A_0)} \leq \|w'\|_{L^\infty(A_0)}$ and $\int |w' - w| \mu < \epsilon$. Therefore
\[
\left| \int v w' \circ T^n \mu^h \right| \leq \left| \int v w \circ T^n \mu^h \right| + \left| \int v (w' - w) \circ T^n \mu^h \right|
\leq C \|w\|_{C^0(A_0)n^{-\beta}} + \|v\|_\infty \|w' - w\|_1
\leq C \|w\|_{L^\infty(A_0)n^{-\beta}} + \|v\|_\infty \epsilon.
\]

Theorem 4.0.1 now shows
\[
\|E_0(v \circ T^{-n})\|_1 \leq Cn^{-\beta}.
\]

\[\square\]

**An Example**

For $0 < \epsilon, k < 1$, let $T_{ek} : \Lambda \to \Lambda$, $\Lambda = [0,1]^2$ be the intermittent Baker’s map defined in (4.3),
\[
T_{ek}(x, y) = \begin{cases} (f_\epsilon(x), f_\epsilon^{-1}(y)), & (x, y) \in [0,1/2) \times [0,1] \\ (2x - 1, (y + 1)/2), & (x, y) \in [1/2,1] \times [0,1], \end{cases}
\]
where $f_\epsilon(x) = x(1 + 2^\epsilon x^\epsilon)$. Denote $T := T_{ek}$. Define $Y := [1/2] \times [0,1]$ and recall from the proof of Proposition 4.0.5 that $T$ is modelled by a Young tower with tail $\tau_\epsilon := \inf\{n \geq 1 : T_{ek}^n \in Y\}$, and $\mu(\tau_\epsilon > n) \leq Cn^{-1/\epsilon}$. Therefore the iterated WIP already holds for $0 < k, \epsilon < 1/2$ by [40, Corollary 2.3]. To obtain a new example, we lift our map to a suspension flow and consider the time-one map.
Let $h$ be a Hölder continuous and bounded roof function such that $\inf h > 0$, $\inf_Y h > 1$ and $h$ is constant on stable leaves for $T$. Let $T_1 : \Lambda^h \to \Lambda^h$ be the corresponding suspension flow, defined by (3.15), and assume $T_1$ satisfies absence of approximate eigenfunctions and condition (3.14). Suppose $v \in C^0(\Lambda^h) \cap C^{0,\delta}(\Lambda^h)$ is a mean zero observable. Proposition 4.4.4 now shows

$$\|E_0(v \circ T_1^{-n})\|_1 \leq C n^{-1/c+1}.$$ 

For the remaining condition of $(A_{\text{strong}})$, we mimic the the argument from the proof of Proposition 4.0.5. The argument will show condition (4.1) holds and we can apply Proposition 4.0.2. Denoting $S := T_{kk}$, let $S_t : \Lambda^h \to \Lambda^h$ be the corresponding suspension flow and $\tau^h_T(x, u) := \inf\{n \geq 1 : S^n_t(x, u) \in Y^h\}$. The stable foliation for $S_1$ is the same as the stable foliation for $T_1$. For the proof of Proposition 4.4.3 to carry over, we must show that $S_1^{\tau^h_T}$ contracts stable leaves exponentially and that equation (4.5) holds when we replace $T_{ek} = T$ with $T_1$ and $T_{kk}$ with $S_1$. Exponential contraction along stable leaves for $S_1$ was shown in the proof of Proposition 4.4.3. To show equation (4.5) holds, it suffices to show that distances on stable leaves with respect to $T_1$ and $S_1$ agree. Let $(x, u) \in \Lambda^h$, $(x', u) \in W^s(x, u)$ and let $N = N(x, u, n)$ be the lap number, i.e. the integer $N$ such that $u + n \in [h_N(x), h_{N+1}(x))$. Since $h$ is constant on stable leaves we have $h_m(x') = h_m(x)$ for all $m \geq 1$ and $N(x', u, n) = N(x, u, n)$. Calculate that

$$d(T^n_1(x', u), T^n_1(x, u)) = d \left( (T^N x', u + n - h_N(x')) , (T^N x, u + n - h_N(x)) \right)$$

$$= d(T^N x', T^N x) + |u + n - h_N(x') - (u + n - h_N(x))|$$

Again, since the contraction along stable leaves does not depend on the first coordinate, we can replace $T = T_{ek}$ with $S = T_{kk}$ to obtain

$$d(T^n_1(x', u), T^n_1(x, u)) = d(S^N(x', u), S^N(x)) + |u + n - h_N(x') - (u + n - h_N(x))|$$

$$= d(S^n_1(x', u), S^n_1(x, u)).$$

Therefore equation (4.5) holds and the proof of Proposition 4.4.3 carries over. We now have

$$\|E_0(v \circ T_1^n) - v \circ T_1^n\|_1 \leq C n^{-1/c+1}.$$ 

Assumption $(A_{\text{strong}})$ and hence the iterated WIP (Theorem 4.4.1) holds for all $0 < c < 1/3$ and $0 < e < 1/2$. This extends the previous best known result by Kelly and Melbourne [29, Theorem 5.2] significantly, which was previously known for all
$0 < c, e < 1/3$.

### 4.4.2 Proof of the Iterated WIP

For the proof of the iterated WIP, we will need a second coboundary decomposition

$$v = \hat{v} + \chi_0^\infty \circ T - \chi_0^\infty,$$

(4.17)

where we recall $\chi_0^\infty = \sum_{n=0}^{\infty} E_0(v \circ T^n) - v \circ T^n$. Under the assumption (A$_{\text{strong}}$) we have $\chi_0^\infty \in L^2(\Lambda, \mathbb{R}^d)$. Moreover, $\hat{v}$ is $\mathcal{A}_0$-measurable and $\hat{v} \in L^2(\Lambda, \mathbb{R}^d)$.

To use Proposition 3.4.8 for the approximation of the stochastic integral

$$\int X_n \otimes dY_n \rightarrow_w \int X \otimes dY, X_n$$

must be an adapted process. To this end, define the processes

$$\hat{W}_n := n^{-1/2} \sum_{j=0}^{[nt]-1} \hat{v} \circ T^j, \quad \hat{W}_n^- := n^{-1/2} \sum_{j=1}^{[nt]} \hat{v} \circ T^{-j}.$$

We see that $\hat{W}_n^-$ is an adapted process with respect to $\{A_{n,t} = T^{[nt]} \mathcal{A}_0 | 0 \leq t \leq 1\}$. Further define the processes

$$\mathcal{X}_n^{\beta\gamma}(t) = n^{-1} \sum_{j=0}^{[nt]-1} \sum_{i=0}^{j-1} m^\beta \circ T^i \hat{v}^\gamma \circ T^j,$$

$$\mathcal{X}_n^{\beta\gamma,-}(t) = n^{-1} \sum_{j=1}^{[nt]} \sum_{i=j+1}^{[nt]} \hat{v}^\beta \circ T^{-j} m^\gamma \circ T^{-i} = \int_0^t \hat{W}_n^{\beta,-} dM_n^{\gamma,-},$$

for coordinates $\beta, \gamma = 1, \ldots, d$.

Recall that for random vectors $X_j$ and $S_n = \sum_{j=1}^n X_j$ we have $\overline{S_n} = \max_{1 \leq t \leq n} |S_t|$. In particular,

$$\sup_{0 \leq t \leq 1} \left| W_n^- (t) - \hat{W}_n^- (t) \right| = \sup_t n^{-1/2} \sum_{j=1}^{[nt]} (v - \hat{v}) \circ T^{-j} = n^{-1/2} \max_{1 \leq t \leq n} \left| \sum_{j=1}^t (v - \hat{v}) \circ T^{-j} \right| = n^{-1/2} (v - \hat{v})^-_n.$$

**Proposition 4.4.5.** We have that

$$(W_n, \mathcal{X}_n) \rightarrow_w (W, \mathcal{X})$$

in $(D([0, 1], \mathbb{R}^d \times \mathbb{R}^{d \times d}), \| \cdot \|_\infty)$, where $\mathcal{X}^{\beta\gamma}(t) := \int_0^t W^\beta dW^\gamma$ for $t \in [0, 1], \beta, \gamma =$
Proof. We first verify the hypothesis of Proposition 3.4.8 with \( X_n = \hat{W}_n^-, Y_n = M_n^- \) to obtain convergence of the backwards process \((\hat{W}_n^-, M_n^-) \xrightarrow{w} (W,W,X)\). As \( \hat{\sigma} \) and \( m \) are \( \mathcal{A}_0 \)-measurable, we see \((\hat{W}_n^-, M_n^-)\) is a measurable process with respect to \( \{\mathcal{A}_{n,t} = T[\alpha t], \mathcal{A}_0 \} | 0 \leq t \leq 1\). Moreover, \( M_n^- \) is a martingale by Proposition 4.3.1. The estimate \( \sup_n \mathbb{E} \left( \left[ M_n^{\beta,-}, M_n^{\beta,-} \right](t) \right) < \infty \) follows from the corresponding non-invertible calculation (equation (3.20)). It remains to show convergence of the pair \((\hat{W}_n^-, M_n^-) \xrightarrow{w} (W,W)\). The martingale WIP (Proposition 3.3.5) was shown in the non-invertible case in Proposition 3.3.8, and the proof carries over to the invertible case by dropping the tilde notation. That is, \( M_n^- \xrightarrow{w} W \), where \( W \) is a mean zero Brownian motion with covariance \( \Sigma = \int m m^T d\mu \). By Proposition 4.2.5, \( \Sigma = \lim_{n \to \infty} n^{-1} \int v_n v_n^T d\mu \). The continuous mapping theorem (Proposition 3.2.9) now shows \((M_n^-, M_n^-) \xrightarrow{w} (W,W)\). Thus \((W_n^-, M_n^-) \xrightarrow{w} (W,W)\) by Proposition 4.1.8.

It now remains to show

\[
\sup_{0 \leq t \leq 1} \left| W_n^-(t) - \hat{W}_n^-(t) \right| = n^{-1/2} \max_{1 \leq \ell \leq n} \left| \sum_{j=1}^{\ell} (v - \hat{v}) \circ T^{-j} \right| = n^{-1/2} \max_{\ell} \left| \chi^+ - \chi^+ \circ T^{-\ell} \right| \leq n^{-1/2} |\chi^+| + n^{-1/2} \max_{\ell} \left| \chi^+ \circ T^{-\ell} \right| \to_p 0.
\]

The first term clearly converges to zero, and the second term converges to zero by Proposition 2.2.3. Hence we have \((\hat{W}_n^-, M_n^-) \xrightarrow{w} (W,W)\) and Proposition 3.4.8 now applies, so

\[(\hat{W}_n^-, M_n^-, X_n^-) \xrightarrow{w} (W,W,X) \quad (4.18)\]

in \((D([0,1], \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d}), \| \cdot \|_{\infty})\).

It now remains to complete the time reversal procedure. Recall that \( g(u)(t) = u(1) - u(1 - t) \) and \( h(u,v)(t) = u(1 - t) \left( v(1) - v(1 - t) \right)^T \) are continuous maps. The time reversal argument (Proposition 3.4.7) carries over to the invertible case by replacing \( \hat{W}_n^- \) with \( \hat{W}_n^- \) in the Stieltjes integral and dropping the tilde notation throughout. This gives

\[
(W_n, X_n) = \text{d} \left( g(W_n^-), \left( g(X_n^-) - h(\hat{W}_n^-, M_n^-) \right)^T \right) + F_n,
\]

where \( F_n \) is such that \( \sup_{0 \leq t \leq 1} |F_n(t)| \to_p 0 \). By Proposition 3.3.12 it now suffices to show

\[
\left( g(W_n^-), \left( g(X_n^-) - h(\hat{W}_n^-, M_n^-) \right)^T \right) \xrightarrow{w} Z,
\]

where \( Z \) is a process with continuous sample paths and \( Z = \text{d} (W,X) \). It follows
from (4.18) that \((W_n^{-}, \hat{W}_n^{-}, M_n^{-}, X_n^{-}) \rightarrow_{w} (W, W, W, X)\). The continuous mapping theorem (Proposition 3.2.9) now shows
\[
\left( g(W_n^{-}), (g(X_n^{-}) - h(W_n^{-}, M_n^{-}))^T \right) \rightarrow_{w} \left( g(W), (g(X) - h(W, W))^T \right).
\]

Proposition 3.4.12 shows \( (g(W), (g(X) - h(W, W))^T) =_{d} (W, X) \) and so we have shown \((W_n, X_n) \rightarrow_{w} (W, X)\) as required. \(\square\)

**Proof of Theorem 4.4.1.** For this proof we denote \(\chi_0^\infty = \chi^+\), with coordinates \(\chi^{\beta, +}\) for \(\beta = 1, \ldots, d\). With this notation, our secondary coboundary decomposition (4.17) is \(v = \hat{v} + \chi^+ \circ T - \chi^+\).

We first show that
\[
\Gamma^{\beta, \gamma}(t) = \sum_{j=1}^{\infty} \int v^\beta v^\gamma \circ T^j d\mu = \int \chi^\beta v^\gamma - m^\beta \chi^{\gamma, +} \circ T d\mu.
\]

Consider
\[
\sum_{j=1}^{n} \int v^\beta v^\gamma \circ T^j d\mu = \sum_{j} \int (m^\beta + \chi^\beta \circ T - \chi^\beta) v^\gamma \circ T^j d\mu
\]
\[
= \sum_{j} \int m^\beta (\hat{v}^\gamma + \chi^{\gamma, +} \circ T - \chi^{\gamma, +}) \circ T^j + \int (\chi^\beta - \chi^\beta \circ T^{-n}) v^\gamma d\mu
\]
\[
= \sum_{j} \int m^\beta \hat{v}^\gamma \circ T^j + \int m^\beta (\chi^{\gamma, +} \circ T^{n+1} \circ T - \chi^{\gamma, +} \circ T) d\mu + \int (\chi^\beta - \chi^\beta \circ T^{-n}) v^\gamma d\mu
\]
\[
= \sum_{j} \int m^\beta \hat{v}^\gamma \circ T^j + \int (\chi^\beta v^\gamma - m^\beta \chi^{\gamma, +} \circ T) d\mu + \int m^\beta \chi^{\gamma, +} \circ T^{n+1} d\mu
\]
\[
- \int \chi^\beta v^\gamma \circ T^j d\mu.
\]

The first term is zero since for every \(j \geq 1\), \(\hat{v} \circ T^j\) is \(A_j\)-measurable and \(E_j m = E_j E_1 m = 0\). Thus \(\int m^\beta \hat{v}^\gamma \circ T^j d\mu = \int E_j (m^\beta \hat{v}^\gamma \circ T^j) d\mu = \int E_j m^\beta \hat{v}^\gamma \circ T^j d\mu = 0\). The third and fourth term converge to zero by mixing and the fact that \(\int m^\beta d\mu = \int v^\gamma d\mu = 0\). Therefore \(\sum_{j=1}^{\infty} \int v^\beta v^\gamma \circ T^j d\mu \rightarrow_{a.s.} \int \chi^\beta v^\gamma - m^\beta \chi^{\gamma, +} \circ T d\mu\).

Next, note that
\[
v^\beta \circ T^i v^\gamma \circ T^j = m^\beta \circ T^i v^\gamma \circ T^j + (\chi^\beta \circ T - \chi^\beta) \circ T^i v^\gamma \circ T^j
\]
\[
= m^\beta \circ T^i \hat{v}^\gamma \circ T^j + m^\beta \circ T^i (\chi^{\gamma, +} \circ T - \chi^{\gamma, +}) \circ T^j + (\chi^\beta \circ T - \chi^\beta) \circ T^i v^\gamma \circ T^j.
\]
It follows that

\[
\mathbb{W}_n^\beta \gamma (t) - X_n^\beta \gamma (t) = n^{-1} \sum_{0 \leq i < j \leq \lfloor nt \rfloor - 1} \left( v^\beta \circ T^i \circ \gamma \circ T^j - m^\beta \circ T_i \circ \hat{\gamma} \circ T^j \right)
\]

\[
= n^{-1} \sum_{0 \leq i < j \leq \lfloor nt \rfloor - 1} \left( m^\beta \circ T^i (\chi_\gamma^+ \circ T - \chi_\gamma^+) \circ T^j + (\chi^\beta \circ T - \chi^\beta) \circ T_i \circ v^\gamma \circ T^j \right)
\]

\[
= n^{-1} \sum_{i=0}^{\lfloor nt \rfloor - 1} \sum_{j=i+1}^{\lfloor nt \rfloor - 1} m^\beta \circ T^i (\chi_\gamma^+ \circ T - \chi_\gamma^+) \circ T^j + n^{-1} \sum_{j=0}^{\lfloor nt \rfloor - 1} \sum_{i=0}^{\lfloor nt \rfloor - 1} (\chi^\beta \circ T - \chi^\beta) \circ T_i \circ v^\gamma \circ T^j
\]

\[
= n^{-1} \sum_{i=0}^{\lfloor nt \rfloor - 1} \sum_{j=i+1}^{\lfloor nt \rfloor - 1} (\chi^\beta v^\gamma - m^\beta \chi_\gamma^+ \circ T) \circ T^j + n^{-1} \sum_{i=0}^{\lfloor nt \rfloor - 1} m^\beta \circ T_i \circ n - n^{-1} \sum_{i=0}^{\lfloor nt \rfloor - 1} \chi^\beta \sum_{i=0}^{\lfloor nt \rfloor - 1} v^\gamma \circ T^i
\]

\[
= A_n^\beta \gamma (t) + B_n^\beta \gamma (t) - C_n^\beta \gamma (t)
\]

We see that \( A_n^\beta \gamma (1) \rightarrow a.s. \int \chi^\beta v^\gamma - m^\beta \chi_\gamma^+ \circ T \ d\mu = \Gamma^\beta \gamma (1) \) by the Ergodic Theorem, and so \( \sup_t |A_n^\beta \gamma (t) - \Gamma^\beta \gamma (t)| \rightarrow p 0 \) by Proposition 2.2.2. Similarly, \( C_n^\beta \gamma (1) \rightarrow a.s. \chi^\beta \int v^\gamma \ d\mu = 0 \) and so \( \sup_t |C_n^\beta \gamma (t)| \rightarrow p 0 \). Finally, we see that \( \sup_t |B_n^\beta \gamma (t)| \rightarrow p 0 \) since

\[
\left\| \sup_t n^{-1} \chi_\gamma^+ \circ T^n \sum_{i=0}^{\lfloor nt \rfloor - 1} m^\beta \circ T^i \right\| \leq n^{-1} \| \chi_\gamma^+ \| \| 1 \leq n^{-1} \| \chi_\gamma^+ \| \| 2 \leq \left\| \sup_t \sum_{i=0}^{\lfloor nt \rfloor - 1} m^\beta \circ T^i \right\| \| 2
\]

\[
\leq n^{-1} n^{1/2} \| \chi_\gamma^+ \| \| m^\beta \| \| 2 \rightarrow 0,
\]

where we have used Doob’s inequality (Proposition 3.3.9) and Proposition 4.1.4. \( \square \)
Bibliography


