Calculating the Lévy Area of Deterministic Dynamics

by

Oliver Crook

M4K9 Dissertation
Mathematics Institute
The University of Warwick
April, 2016

1 Introduction

Statistical limit laws are a fundamental part of probability theory and its applications. Amongst many, these include the Central Limit Theorem (CLT), the Weak Invariance Principle (WIP) (also known as the Functional Central Limit Theorem) and the Strong Law of Large Numbers (SLLN). These theorems are often proved under the i.i.d (*independent and identically distributed*) assumption. In the dynamical systems setting, and in applications to many physical systems, the assumption of independence is unnatural and unfavorable; however because we work with an invariant measure we maintain the assumption of stationarity.

Consider the canonical example of the SLLN. This is traditionally proved in the situation of i.i.d random variables $(X_k)_{k=1,2,...}$ each with finite and equal expectation. However, this can be seen as a special case of the Ergodic Theorem which is often proved in the setting of an integrable observable v and transformation $f : \Lambda \to \Lambda$ with an underlying invariant measure. As such, in this setting, the sequence $(v \circ f^k)_{k=0,1,...}$ is stationary but not, in general, independent.

The project is centered around the relatively recent limit law: the Iterated Weak Invariance Principle (Iterated WIP) which is closely related to calculating the Lévy Area from Stochastic Analysis. In recent years cutting-edge methods have been developed in probability theory (the so-called Maxwell-Woodroofe condition) and we seek to explore the application of these methods to the Iterated WIP.

To this end, we present the broad theory surrounding limit laws for dynamical systems and, in particular, Section 2 formulates the general framework under which we will work. The introduction of the Transfer Operator helps to unify our proofs and the main dynamical assumption on the decay of the correlation function. Under sensible assumptions we obtain a powerful tool: the martingale coboundary decomposition. We proceed to use this tool in Sections 3 and 4 to prove a CLT and WIP respectively, introducing the background as we progress.

In Section 5, using Stochastic Differential Equations (SDEs), we motivate the study of the Iterated WIP before giving a sketch proof of the main elements of the results. Again Section 5 works under the general framework from the earlier sections.

Section 6 proposes a new framework and we explore some of the literature motivating the so called Maxwell-Woodroofe condition which would greatly relax the assumptions introduced in Section 2. We proceed to prove a CLT under this assumption and state results concerning the WIP. The techniques used to prove this CLT present a possible method to extend the Iterated WIP.

In the final section we seek to extend the simpler result of the Iterated CLT. We present results which highlight that under the naïve assumption of the Maxwell-Woodroofe condition we run into complications and have some unfavorable assumptions to check. We show that the strengthening of some of the standard dynamical assumptions made in our theorem reduces us to one assumption left to be understood. We conjecture that this assumption is in fact unnecessary and we are working with the optimal set of assumption which is justified in the final section.

This project uses techniques in Ergodic Theory, Probability Theory and Stochastic Analysis.

2 Background

In this section we introduce some of the required background theory needed in our proofs that are not standard in a measure-theoretic Probability or Ergodic Theory course. Consider a probability space (Λ, μ) and a transformation on this probability space $f : \Lambda \to \Lambda$ where μ is invariant under f. That is, if $A \subset \Lambda$ then $\mu(f^{-1}(A)) = \mu(A)$.

Definition 2.1. The *Koopman operator* is defined as the map $U : L^1(\Lambda) \to L^1(\Lambda)$ given by $Uv = v \circ f$.

Several nice properties of the map follow directly from the definition:

Proposition 2.2. *Elementary Properties of the Koopman Operator*

$(U1)$ $U1 = 1$.

$(U2)$ $\displaystyle\int_\Lambda Uv\, d\mu = \int_\Lambda v\, d\mu$.

$(U3)$ *for any* $p \geq 1$ *and* $v \in L^p(\Lambda)$ *then* $\|Uv\|_p = \|v\|_p$.

$(U4)$ *for any* $v \in L^1(\Lambda)$ *and* $w \in L^\infty(\Lambda)$ $\displaystyle\int_\Lambda UvUw\, d\mu = \int_\Lambda vw\, d\mu$.

Proof. These proofs are routine since μ is an invariant measure. □

It is worth remarking that $(U3)$ means that the Koopman operator is an isometry. This is an unfavorable property when studying statistical limit laws. Hence, we study the corresponding dual operator: the transfer operator. The transfer operator is the main approach used in the modern study of statistical limit laws [13,17]. We state the definition in the general setting but note that in L^2 the transfer operator is the adjoint of the Koopman operator.

Definition 2.3. The *transfer operator* is defined as the map $P : L^1(\Lambda) \to L^1(\Lambda)$ such that $\int_\Lambda Pv\, w\, d\mu = \int_\Lambda v\, Uw\, d\mu$ for every $w \in L^\infty(\Lambda)$.

Again, there are corresponding elementary properties for the transfer operator. We state them without proof:

Proposition 2.4. *Elementary Properties of the Transfer Operator*

$(P1)$ $P1 = 1$.

$(P2)$ $\displaystyle\int_\Lambda Pv\, d\mu = \int_\Lambda v\, d\mu$.

$(P3)$ *for any* $p \geq 1$ *and* $v \in L^p(\Lambda)$ *then* $\|Pv\|_p \leq \|v\|_p$.

$(P4)$ $PU = I$.

We remark that $(P3)$ means that the transfer operator is a contraction on L^p. This is a desirable property when discussing statistical limit laws because we get decay upon iteration. The following are two examples of transfer operators for dynamical systems.

Example 2.5. *Transfer Operator for Dynamical Systems*
1) If $f : [0,1] \to [0,1]$ *is* $f(x) = 2x \mod 1$, *then* $(Pv)(y) = \frac{1}{2}\left[v(\frac{y}{2}) + v(\frac{y+1}{2})\right]$.
2) If $f : [0,1] \to [0,1]$ *is* $f(x) = \{\frac{1}{x}\}$, *that is* f *is the Gauss map, then the transfer operator is* $(Pv)(y) = \sum_{n=1}^{\infty} \frac{1}{(x+n)^2} v(\frac{1}{x+n})$.

The following interpretation of the transfer operator bridges the gap between the earlier CLT and WIP in this project and than the ones presented later.

Remark 2.6. Consider $g \in L^1(\Lambda)$, $f : \Lambda \to \Lambda$ a measure preserving transformation and μ the underlying invariant measure for f. Then $UPg = (Pg) \circ f = E[g|f^{-1}\mathcal{B}]$.

The first section of this project is dedicated to proving results under sufficiently rapid decay of correlations. We discuss this notion briefly and its relation to mixing. Recalling that f is mixing (with respect to the underlying invariant measure μ) if and only if $\mu(A \cap f^{-n}B) \to \mu(A)\mu(B)$ for every $A, B \in \mathcal{B}$ where $\mathcal{B}$ is the underlying σ-algebra. It is a standard result that this convergence can happen arbitrarily slowly. In terms of the correlation function mixing is equivalent to the following: $Cov(v, w \circ f^n) = E(vw \circ f^n) - E(v)E(w) = \int_\Lambda vw \circ f^n \, d\mu - \int_\Lambda v \, d\mu \int_\Lambda w \, d\mu \to 0$ for all $v, w \in L^2(\Lambda)$.

Consider the following framework: let $f : \Lambda \to \Lambda$ be an ergodic transformation with respect to μ and let $v : \Lambda \to \mathbb{R}$ be a mean-zero observable which lies in $L^\infty(\Lambda)$. Suppose that there exists a $\beta > 1$ and a $C > 0$ such that for all $w \in L^1(\Lambda)$ and sufficiently large $n \geq 1$ we have:

$$|\int_\Lambda vw \circ f^n \, d\mu| \leq Cn^{-\beta}\|w\|_1. \tag{1}$$

Remark 2.7. The decay described in (1) is an assumption on the dynamics of f, the underlying measure and the observable v. Such an assumption cannot hold for general $v \in L^\infty(\Lambda)$. The main tool following from this estimate will be the Martingale Coboundary Decomposition:

Proposition 2.8. *Martingale Coboundary Decomposition*
Assume that our observable v satisfies (1)*. Then there exists $m, \chi \in L^\infty(\Lambda)$ such that m has mean zero, $Pm = 0$ and,*

$$v = m + \chi \circ f - \chi. \tag{2}$$

Proof. This proof follows Melbourne-Török Proposition 1 and Lemma 1 [13]. First note by definition of the transfer operator that $\int_\Lambda vw \circ f^n \, d\mu = \int_\Lambda P^n vw \, d\mu$. Now let $\epsilon > 0$ and suppose there is a set A of positive measure on which $|P^n v| \geq Cn^{-\beta} + \epsilon$. Choose $w = \mathbb{1}_A sgn(P^n v)$ and we see that $\int_\Lambda |w| \, d\mu \leq \mu(A) \leq 1$ so w is in L^1. Then

$$|\int_\Lambda P^n v \mathbb{1}_A sgn(P^n v) \, d\mu| = \int_A |P^n v| \, d\mu \geq (Cn^{-\beta} + \epsilon)\mu(A)$$

by our assumption. On the other hand, we have that $|\int_\Lambda vw \circ f^n \, d\mu| \leq Cn^{-\beta}\|w\|_1$ for every $w \in L^1(\Lambda)$. Thus, for this choice of w we obtain the following inequality: $(Cn^{-\beta} + \epsilon)\mu(A) \leq Cn^{-\beta}\mu(A)$. Hence, we conclude $\mu(A) = 0$ and so $\|P^n v\|_\infty \leq Cn^{-\beta}$. Since $Cn^{-\beta}$ is summable, it follows that $\|P^n v\|_\infty$ is summable. Now define $\chi : \Lambda \to \mathbb{R}$ by $\chi = \sum_{n=1}^\infty P^n v$ and $m = v + \chi - \chi \circ f$. The fact that m is mean zero follows directly from using that μ is an invariant measure:

$$\int_\Lambda m \, d\mu = \int_\Lambda v + (\chi - \chi \circ f) \, d\mu = \int_\Lambda v \, d\mu + \int_\Lambda \chi - \chi \, d\mu = 0.$$

Moreover $m \in \text{Ker } P$ follows by applying P to the definition of m:

$$Pm = Pv - PU\chi + P\chi = Pv - \sum_{n=1}^\infty P^n v + \sum_{n=2}^\infty P^n v = 0.$$

□

Remark 2.9. Later we will need the decomposition (2) in L^2. Precisely, let $f : \Lambda \to \Lambda$ be ergodic with respect to μ and let $v : \Lambda \to \mathbb{R}$ be a mean-zero observable lying in $L^2(\Lambda)$. Suppose that there exists $\beta > 1$ and $C > 0$ such that for all $w \in L^2(\Lambda)$ and sufficiently large $n \geq 1$ we have:

$$|\int_\Lambda vw \circ f^n \, d\mu| \leq Cn^{-\beta}\|w\|_2.$$

Under these assumptions the decomposition (2) holds in $L^2(\Lambda)$. We see this by choosing $w = P^n v$ which is clearly in $L^2(\Lambda)$ because $v \in L^2(\Lambda)$ and P a contraction on $L^2(\Lambda)$. Then substituting $w = P^n v$ in our assumption of the decay of correlations above gives $\|P^n v\|_2^2 \leq Cn^{-\beta}\|P^n v\|_2$. Thus, $\|P^n v\|_2 \leq Cn^{-\beta}$ and so $\|P^n v\|_2$ is summable. Defining $\chi = \sum_{n=1}^{\infty} P^n v$ we observe that the decomposition (2) holds in $L^2(\Lambda)$.

In later sections it will be clear in which space we are considering the Martingale Coboundary Decomposition by the set up of the problem. For the remainder of this section we discuss some of the properties of the resulting decomposition (2). These properties lead us to to prove results about m which extend to v. We refer to m as the martingale part of v, but it will not be fully clear until later sections in what sense m is a martingale. First, we recall $v_n = \sum_{j=0}^{n-1} v \circ f^j$ and define $m_n = \sum_{j=0}^{n-1} m \circ f^j$. Thus, from telescoping sums, $v_n = m_n + \chi \circ f^n - \chi$. It follows that $n^{-1/2}(v_n - m_n) \to 0$ *a.e.* because $\|\chi \circ f^n - \chi\|_\infty \leq \|2\chi\|_\infty$. This leads us to the following result:

Lemma 2.10. *Suppose that Y is a random variable. Then $n^{-1/2}v_n \to_d Y$ if and only if $n^{-1/2}m_n \to_d Y$.*

Lemma 2.11. *Orthogonality*
The sequence $\{m \circ f^j : j \geq 0\}$ is orthogonal. That is $\int_\Lambda m \circ f^{j_1} m \circ f^{j_2} \dots m \circ f^{j_r} \, d\mu = 0$ for all $0 \leq j_1 < \cdots < j_r$ $r \geq 1$.

Proof. If $r = 1$ the result follows as m is mean-zero. For $r \geq 2$,

$$\begin{aligned}\int_\Lambda m \circ f^{j_1} m \circ f^{j_2} \dots m \circ f^{j_r} \, d\mu &= \int_\Lambda m \, U(m \circ f^{j_2 - j_1 - 1} \dots m \circ f^{j_r - j_1 - 1}) \, d\mu \\ &= \int_\Lambda Pm(m \circ f^{j_2 - j_1 - 1} \dots m \circ f^{j_r - j_1 - 1}) \, d\mu = 0.\end{aligned}$$

□

We now state two theorems from standard probability theory which will be used in the following sections.

Theorem 2.12. *Bounded Convergence Theorem*
Suppose that $f_n \in L^\infty(\Lambda), n \geq 1$. If $\|f_n\|_\infty$ is bounded and $f_n \to f$ a.e., then $\int_\Lambda f_n \, d\mu \to \int_\Lambda f d\mu$.

Theorem 2.13. *Lévy's Continuity Theorem*
Suppose that $Y_n, n = 1, 2, ...$ and Y are random variables with values in $\mathbb{R}^k$. Then $Y_n \to Y$ in $\mathbb{R}^k$ if and only if $\lim_{n\to\infty} E[e^{<it,Y_n>}] = E[e^{<it,Y>}]$ for every $t \in \mathbb{R}^k$.

3 The Central Limit Theorem

In this section, we prove the CLT and its multi-dimensional counterpart. The standard CLT assumes independence or sufficient control of higher moments of the random variables. Here we use the set-up of the previous section; in particular, our main tool will be the Martingale Coboundary Decomposition (MCD). The proofs follow McLeish [11].

Theorem 3.1. *The Central Limit Theorem*
Let $f : \Lambda \to \Lambda$ be a transformation of a probability space (Λ, μ) where μ is assumed to be invariant and ergodic with respect to f. Suppose that our observable $v : \Lambda \to \mathbb{R}$ lies in $L^\infty(\Lambda)$ and is mean-zero. Further assume that the decomposition (2) is valid and define v_n as before. Then the following hold:

a) The limit $\sigma^2 = \lim_{n\to\infty} \int_\Lambda (n^{-1/2} v_n)^2 \, d\mu$ exists and $\sigma^2 = \int_\Lambda m^2 \, d\mu$.
b) $n^{-1/2} v_n \to_d N(0, \sigma^2)$ as $n \to \infty$.

Furthermore, $\sigma^2 = 0$ if and only if $v = h \circ f - h$ for some $h \in L^\infty(\Lambda)$.

Proof. Firstly, we prove (a) and identify $\sigma^2 = \int_\Lambda m^2 \, d\mu$. From the orthogonality proved in Lemma 2.11, it follows that $\int_\Lambda m_n^2 = \sum_{j=0}^{n-1} \int_\Lambda (m \circ f^j)^2 \, d\mu$. By invariance of the measure μ, we see that $n^{-1} \int_\Lambda m_n^2 = \int_\Lambda m^2 \, d\mu$. Since $v_n = m_n + \chi \circ f^n - \chi$, we have by the triangle inequality:

$$|\|v_n\|_2 - \|m_n\|_2| \le \|v_n - m_n\|_2 = \|\chi \circ f^n - \chi\|_2 \le 2\|\chi\|_2.$$

Hence $n^{-1/2}(\|v_n\|_2 - \|m_n\|_2) \to 0$ a.e. Squaring this equation we see

$$\lim_{n\to\infty} \int_\Lambda (n^{-1/2} v_n)^2 \, d\mu = \lim_{n\to\infty} \int_\Lambda (n^{-1/2} m_n)^2 \, d\mu = \int_\Lambda m^2 \, d\mu.$$

Thus we have proven (a).

Next we prove that $n^{-1/2} v_n \to_d N(0, \sigma^2)$ as $n \to \infty$. By Lemma 2.10 it is sufficient to prove $n^{-1/2} m_n \to_d N(0, \sigma^2)$. Then by Theorem 2.13 it is enough to show that for every fixed $t \in \mathbb{R}$ $\lim_{n\to\infty} \int_\Lambda e^{itn^{-1/2} m_n} \, d\mu = e^{-\frac{(t\sigma)^2}{2}}$. Since $\log(1+ix) = ix + \frac{1}{2}x^2 + O(|x|^3)$, it follows that $e^{ix} = (1 + ix) \exp\{-\frac{1}{2}x^2 + O(|x|^3)\}$. Then using this and recalling $\exp\{itn^{-1/2} m_n\} = \exp\{itn^{-1/2} \sum_{j=0}^{n-1} m \circ f^j\}$, we have

$$\exp\{itn^{-1/2} m_n\} = \prod_{j=0}^{n-1} \exp\{itn^{-1/2} m \circ f^j\} = T_n e^{-U_n},$$

where

$$T_n = \prod_{j=0}^{n-1} (1 + itn^{-1/2} m \circ f^j)$$
$$U_n = \frac{1}{2} t^2 n^{-1} \sum_{j=0}^{n-1} m^2 \circ f^j + O\left(n^{-3/2} \sum_{j=0}^{n-1} m^3 \circ f^j \right) = \frac{1}{2} t^2 n^{-1} \sum_{j=0}^{n-1} m^2 \circ f^j + O(n^{-1/2}).$$

The last equality follows because there are n terms in the sum.
Now $T_n = 1 + \{$Terms of the form: $Cm \circ j_1 m \circ j_2 \ldots m \circ j_r\}$ for some $C \in \mathbb{C}$. Thus, by orthogonality $\int_\Lambda T_n \, d\mu = 1$. Also, by the Ergodic Theorem:

$$U_n \to \frac{1}{2} t^2 \int_\Lambda m^2 \, d\mu = \frac{1}{2} t^2 \sigma^2 \ a.e.$$

Hence,

$$\int_\Lambda e^{itn^{-\frac{1}{2}} m_n} \, d\mu - e^{-\frac{1}{2} t^2 \sigma^2} = \int T_n (e^{-U_n} - e^{-\frac{1}{2} t^2 \sigma^2}) \, d\mu$$

where the second integrand convergences to 0 a.e. We claim that the integrand is uniformly bounded and thus part (b) follows from Theorem 2.12. We see that

$$|T_n| = |\prod_{j=0}^{n-1}(1 + itn^{-1/2} m \circ f^j)| = \prod_{j=0}^{n-1}(1 + t^2 n^{-1}|m \circ f^j|^2)^{1/2}$$

and so clearly $|T_n| \geq 1$. Since $e^{-U_n} = e^{itn^{-\frac{1}{2}} m_n}/T_n$, we see $\|e^{-U_n}\|_\infty \leq 1$. Consider the following:

$$d_n = \prod_{j=0}^{n-1}(1 + t^2 n^{-1}\|m\|_\infty^2)^{1/2} = (1 + n^{-1}t^2\|m\|_\infty^2)^{\frac{n}{2}} \to \exp(\frac{1}{2}t^2\|m\|_\infty^2).$$

Hence d_n is a convergent sequence and thus bounded for all n, by say M. It follows that $\|T_n\|_\infty \leq M$ and so is uniformly bounded, which completes the proof of part (b).

For non-degeneracy, as we identified $\sigma^2 = \int_\Lambda m^2 \, d\mu$ it follows that if $\sigma^2 = 0$ then $m = 0$ a.e. and so $v = h \circ f - h$ with $h = \chi$. For the converse, that is assuming $v = h \circ f - h$, we have $v_n = h \circ f^n - h$. Thus $\|v_n\|_\infty \leq 2\|h\|_\infty$ is bounded and hence $\sigma^2 = \lim_{n \to \infty} \int_\Lambda (n^{-1/2} v_n)^2 \, d\mu = 0$. This concludes the proof. □

Remark 3.2. The essential assumption in this theorem is that $\|P^n v\|_\infty$ is summable (which allowed us to use the MCD). This applies to many uniformly expanding and many non-uniformly expanding dynamical systems though some additional work is required in the latter setting. However, later in this project we will relax this assumption to $\sum_{n=1}^\infty n^{-3/2}\|\sum_{k=0}^{n-1} P^k v\|_2 < \infty$, which is due to Tyran-Kamińska [17] based on ideas proved by Maxwell and Woodroofe [10].

We have proved convergence in distribution for real-valued random variables and so it is natural to ask whether this extends to multi-dimensional results. In fact, for random vectors in $\mathbb{R}^d$ the proof is straight forward using a technique called the Cramer-Wold device, which follows directly from the Lévy Continuity Theorem. A later section will deal with the case where we take random elements in the Banach space $C([0,1], \mathbb{R}^d)$. Note that all the statements about which we have already spoken can be thought of componentwise. In particular the transfer operator acts coordinate wise on the vectors.

Proposition 3.3. *Cramer-Wold device*
Suppose that $(Y_n)_{n \geq 1}$ and Y are random vectors in $\mathbb{R}^d$. Then $Y_n \to_d Y$ in $\mathbb{R}^d$ if and only if $\langle c, Y_n \rangle \to_d \langle c, Y \rangle$ in $\mathbb{R}$ for all $c \in \mathbb{R}^d$.

Proof. Suppose that $\langle c, Y_n \rangle \to_d \langle c, Y \rangle$ in $\mathbb{R}$ for all $c \in \mathbb{R}^d$. It follows that $E[e^{i\langle c, Y_n \rangle}] \to E[e^{i\langle c, Y \rangle}]$ and so by Lévy continuity $Y_n \to_d Y$ in $\mathbb{R}^d$. Now suppose $Y_n \to_d Y$ in $\mathbb{R}^d$ and for any $c \in \mathbb{R}^d$ let $X_n = c^T Y_n$ and $X = c^T Y$. Now let $t \in \mathbb{R}$ and so we have

$$E[e^{itX_n}] = E[e^{i\langle tc, Y_n \rangle}] \to E[e^{i\langle tc, Y \rangle}] = E[e^{itX}].$$

Thus, again by Lévy continuity $X_n \to X$ in $\mathbb{R}$ and hence $\langle c, Y_n \rangle \to_d \langle c, Y \rangle$ in $\mathbb{R}$. □

Theorem 3.4. *multi-dimensional Central Limit Theorem*
Let $f : \Lambda \to \Lambda$ be a transformation of a probability space (Λ, μ) where μ is assumed to be invariant and ergodic with respect to f. Suppose that our observables $v^i : \Lambda \to \mathbb{R}$ lie in $L^\infty(\Lambda)$ and are mean-zero. Further assume that the decomposition (2) *is valid and define v_n as before but v is now a vector. Also define the covariance matrix $\Sigma = \int_\Lambda mm^T \, d\mu$. Then the following hold:*

a) The limit $\Sigma = \lim_{n\to\infty} n^{-1} \int_\Lambda v_n v_n^T \, d\mu$ *exists.*
b) $n^{-1/2} v_n \to_d N(0,\Sigma)$ *as* $n \to \infty$.

Furthermore, $\det \Sigma = 0$ *if and only if there exists a nonzero* $c \in \mathbb{R}^d$ *such that* $c^T v = h \circ f - h$ *for some* $h \in L^\infty(\Lambda)$.

Proof. For part (a) it suffices to work in coordinates and show that: $\lim_{n\to\infty} n^{-1} \int_\Lambda v_n^i v_n^j \, d\mu = \int_\Lambda m^i m^j \, d\mu$. For the diagonal terms, that is $i = j$, the result follows from part (a) of the 1-dimensional CLT. Now take $i \neq j$ and recall, by the parallelogram identity, that:

$$v_n^i v_n^j = \frac{1}{4}\left((v_n^i + v_n^j)^2 - (v_n^i - v_n^j)^2\right).$$

Noting that $v_n^i + v_n^j = m_n^i + m_n^j + \chi^i \circ f^n - \chi^i + \chi^j \circ f^n - \chi^j = m_n^i + m_n^j + (\chi^i + \chi^j) \circ f^n - (\chi^i + \chi^j)$, which is in the form exploited in the 1-dimensional CLT and likewise $v_n^i - v_n^j = m_n^i - m_n^j + (\chi^i - \chi^j) \circ f^n - (\chi^i - \chi^j)$. Hence,

$$\frac{1}{4n}\int_\Lambda \left((v_n^i + v_n^j)^2 - (v_n^i - v_n^j)^2\right) d\mu \to \frac{1}{4}\int_\Lambda (m^i + m^j)^2 - (m^i - m^j)^2 \, d\mu = \int m^i m^j \, d\mu$$

as required.

For (b) we apply the Cramer-Wold device with $Y_n = n^{-1/2} v_n$ and $Y \sim N(0,\Sigma)$. It is standard that $\langle c, Y\rangle = c^T Y \sim N(0, c^T \Sigma c)$ and hence $E[e^{it\langle c,y\rangle}] = \exp\{-\frac{1}{2} t^2 c^T \Sigma c\}$. Meanwhile, $Y_n = n^{-1/2} v_n$ and $\langle c, Y_n\rangle = n^{-1/2} \langle c, v\rangle_n$ where $\langle c, v\rangle_n = n^{-1} \sum_{j=0}^{n-1} \langle c, v\rangle \circ f^j$ which is a 1-dimensional Birkhoff sum. Recalling the MCD, we have: $\langle c, v\rangle = \langle c, m\rangle + h \circ f - h$ where $h = \langle c, \chi\rangle$. It follows by the 1-dimensional CLT that:

$$\langle c, Y_n\rangle = n^{-1/2} \langle c, v\rangle_n \to N(0, \sigma_c^2),$$

where $\sigma_c^2 = \int_\Lambda \langle c, m\rangle^2 \, d\mu$. To conclude the theorem, it remains to show $\sigma_c^2 = c^T \Sigma c$ for every c. We have

$$c^T \Sigma c = c^T \int_\Lambda m m^T \, d\mu \, c = \int_\Lambda \langle c, m\rangle \langle m, c\rangle \, d\mu = \int_\Lambda \langle c, m\rangle^2 \, d\mu = \sigma_c^2,$$

as we needed.

For non-degeneracy of the CLT we notice that $\det \Sigma = 0$ is equivalent to there existing a $c \neq 0$ such that $\Sigma c = 0$. However $c^T \Sigma c = \int_\Lambda \langle c, m\rangle^2 \, d\mu$ and so $\Sigma c = 0$ is equivalent to $\langle c, m\rangle = 0$ in which case $\langle c, v\rangle = h \circ f - h$ where $h = \langle c, \chi\rangle$. This concludes the proof. □

4 A Weak Invariance Principle

Considering our original set up, let $v : \Lambda \to \mathbb{R}$ be a Hölder mean-zero observable. By the CLT we have that $n^{-1/2} v_n \to_d N(0, \sigma^2)$. Now define a sequence of continuous time processes $W_n : [0,1] \to \mathbb{R}$ in the following way. Set $W_n(t) = n^{-1/2} v_{nt}$ for $t = j/n$, $j = 0, 1, \ldots, n$. Then linearly interpolate to obtain a continuous function $W_n \in C[0,1]$. First notice $W_n(t) = n^{-1/2} v_{[nt]} + O(n^{-1/2})$ uniformly and that for fixed t we have that $W_n(t) \to_d N(0, t\sigma^2)$. The function W_n is a random element in the Banach Space $C[0,1]$ and so our result will be more complex because $C[0,1]$ is infinite dimensional. We state the main theorem of this section below; then cover some preliminary material, before ending this section with a proof of the Weak Invariance Principle (WIP) also known as the Functional Central Limit Theorem.

Theorem 4.1. *Weak Invariance Principle*
Let $f : \Lambda \to \Lambda$ be a transformation of a probability space (Λ, μ) where μ is assumed to be invariant and ergodic with respect to f. Suppose that our observable $v : \Lambda \to \mathbb{R}$ lies in $L^\infty(\Lambda)$ and is mean-zero. Further assume that the decomposition (2) is valid and define $W_n(t)$ as above. Then $W_n(t) \to W(t)$ in $C[0,1]$ where $W(t)$ is a Brownian motion with variance $\sigma^2 = \int_\Lambda m^2 \, d\mu$.

Remark 4.2. Since we are still assuming the martingale coboundary decomposition, we are proving the WIP in the context that $\|P^n v\|_\infty$ is summable. As with the CLT, we will relax this assumption in a later section.

We state the following theorems, which will be of use in the proof of Theorem 4.1.

Theorem 4.3. *Continuous Mapping Theorem*
Suppose that $(Y_n)_{n\geq 1}$ and Y are random elements in $C[0,1]$. Then the following are equivalent:
a) $Y_n \to_w Y$ in $C[0,1]$.
b) $\psi(Y_n) \to_d \psi(Y)$ in $\mathbb{R}$ for every continuous map $\psi : C[0,1] \to \mathbb{R}$.

We will need tightness of our random elements since it is a necessary condition for convergence in distribution. To work with tightness can be rather tricky but Prokhorov's theorem shows us how important it is.

Definition 4.4. A sequence of random elements $Y_n \in C[0,1]$ is *tight* if for any $\epsilon > 0$ there is a compact subset $K \subset C[0,1]$ such that $P(Y_n \in K) > 1 - \epsilon$ for every $n \geq 1$.

We can already see the problem with working with tightness in $C[0,1]$ because compact sets are complicated in $C[0,1]$. The Arzelà-Ascoli Theorem characterizes compact set in $C[0,1]$ as those which are closed, bounded (in the sup-norm topology) and equicontinuous. Note that equicontinuity is stronger than uniform continuity. We now state Prokhorov's theorem (for a proof see Billingsley [1]) and then we prove the WIP.

Theorem 4.5. *Prokhorov*
Let $Y_n, Y \in C[0,1]$ be random elements. Then $Y_n \to_w Y$ if and only if both the following hold:
a) Convergence of finite-dimensional distributions

$$(Y_n(t_1), ..., Y_n(t_k)) \to_d (Y(t_1), \ldots, Y(t_k)) \text{ as } n \to \infty$$

for every $t_1, ..., t_k \in [0,1], k \geq 1$.
b) Tightness: The sequence $\{Y_n\}$ is tight.

The following is the proof of Theorem 4.1.

Proof. We apply Prokhorov's Theorem and so it suffices to check convergence of finite distributions and tightness. As before it is sufficient to prove the result for $M_n(t) = n^{-1/2} m_{[nt]} + O(n^{-1/2})$ where we have the decomposition (2). For convergence of finite distribution we can apply the Cramer-Wold device and show:

$$\exp\{\langle itc, (M_n(t_1)), \ldots, M_n(t_k)\rangle\} \to \exp\{\langle itc, (W(t_1)), \ldots, W(t_k)\rangle\}$$

for every $t \in \mathbb{R}$ and for all $c \in \mathbb{R}^k$. By setting $t_0 = 0$ it is enough to show: $e^{itY_n} \to e^{itY}$ for all $t \in \mathbb{R}$ and $c \in \mathbb{R}^k$ where

$$Y_n = \langle c, (M_n(t_1) - M_n(t_0), \ldots, M_n(t_k) - M_n(t_{k-1})\rangle$$

$$Y = \langle c, (W(t_1) - W(t_0), \ldots, W(t_k) - W(t_{k-1})\rangle.$$

By definition $Y = c^T N(0,\Sigma) = N(0,\sigma_c^2)$ where $\Sigma = \sigma^2 diag(t_1 - t_0, \dots, t_k - t_{k-1})$ and $\sigma_c^2 = c^T \Sigma c = \sigma^2(c_1^2(t_1 - t_0) + \cdots + c_k^2(t_k - t_{k-1}))$. Also

$$\begin{aligned} Y_n =& n^{-1/2}\left[c_1(m_{[nt_1]} - m_{[nt_0])} + \cdots + c_k(m_{[nt_k]} - m_{[nt_{k-1}]})\right] \\ =& n^{-1/2}\sum_{l=1}^{k} c_l \left(\sum_{j=0}^{[nt_l]-1} m\circ f^j - \sum_{j=0}^{[nt_{l-1}]-1} m\circ f^j\right) + O(n^{-1/2}) \\ =& n^{-1/2}\sum_{l=1}^{k} c_l \left(\sum_{j=[nt_{l-1}]}^{[nt_l]-1} m\circ f^j\right) = n^{-1/2}\sum_{j=0}^{[nt_k]-1} d(j,n)\, m\circ f^j + O(n^{-1/2}) \end{aligned}$$

where $d(j,n) \in \{c_1, \dots, c_k\}$ and in the last equality we switched the order of summation. Now writing $e^{itY_n} = T_n e^{-U_n}$ we have:

$$\begin{aligned} \exp\left[itY_n\right] =& \exp\left[it\left(\sum_{j=0}^{[nt_k]-1} d(j,n)\, m\circ f^j + O(n^{-1/2})\right)\right] \\ =& \prod_{j=0}^{[nt_k]-1} \exp\left[itn^{-1/2} d(j,n) m\circ f^j + O(n^{-1/2})\right]. \end{aligned}$$

Thus,

$$\begin{aligned} \exp\left[itY_n\right] =& \prod_{j=0}^{[nt_k]-1}\left(1 + itn^{-1/2}d(j,n)m\circ f^j + O(n^{-1/2})\right) \\ & \exp\left[-\frac{t^2 d(j,n)^2}{2n} m^2\circ f^j + O\left(-\frac{t^3 d(j,n)^3}{n^{3/2}} m^3\circ f^j\right)\right]. \end{aligned}$$

Hence,

$$\begin{aligned} T_n =& \prod_{j=0}^{[nt_k]-1}\left(1 + itn^{-1/2}d(j,n)m\circ f^j + O(n^{-1/2})\right) \\ U_n =& \frac{t^2}{2n}\sum_{j=0}^{[nt_k]-1} d(j,n)^2 m^2\circ f^j + O(n^{-1/2}). \end{aligned}$$

Recalling the form of T_n and the orthogonality deduced in Lemma 2.11, we have $\int_\Lambda T_n\, d\mu = 1$ and as before $\|T_n\|_\infty \le K$ and $\|e^{-U_n}\|_\infty \le 1$. We compute that,

$$U_n = \frac{t^2}{2n}\sum_{j=0}^{[nt_k]-1} d(j,n)^2 m^2\circ f^j + O(n^{-1/2}) = n^{-1/2}\sum_{l=1}^{k} c_l^2 \left(\sum_{j=[nt_{l-1}]}^{[nt_l]-1} m^2\circ f^j\right) + O(n^{-1/2}).$$

Now,

$$
\begin{aligned}
n^{-1}\sum_{j=[nt_{l-1}]}^{[nt_l]-1} m^2\circ f^j =& n^{-1}\left(\sum_{j=0}^{[nt_l]-1} m^2\circ f^j - \sum_{j=0}^{[nt_{l-1}]-1} m^2\circ f^j\right)\\
=&\frac{1}{nt_l}t_l\left(\sum_{j=0}^{[nt_l]-1} m^2\circ f^j\right) - \frac{1}{nt_{l-1}}t_{l-1}\left(\sum_{j=0}^{[nt_{l-1}]-1} m^2\circ f^j\right)\\
\to&(t_l-t_{l-1})\int_\Lambda m^2\,d\mu = (t_l-t_{l-1})\sigma^2 \ a.e.
\end{aligned}
$$

by the Ergodic Theorem. Hence,

$$
\begin{aligned}
U_n \to \frac{t^2}{2}\sum_{l=1}^{k} c_l^2(t_l-t_{l-1})\sigma^2 =& \frac{t^2}{2}(c_1^2(t_1-t_0)+c_k^2(t_k-t_{k-1}))\sigma^2\\
=&\frac{t^2\sigma_c^2}{2}\ a.e.
\end{aligned}
$$

Thus by a similar argument in the CLT via the bounded convergence theorem we conclude that:

$$\int_\Lambda T_n e^{-U_n}\,d\mu \to e^{-\frac{t^2\sigma_c^2}{2}}.$$

Having checked convergence of finite distributions it remains to check tightness. From Billigsley, [1] we quote the following result:

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\lambda^2\mu(\max_{k\le n}|m_k|\ge\lambda\sqrt{n})=0 \implies \{M_n\} \text{ is tight.}$$

Now we show the sequence $\{M_n\}$ is tight in $C[0,1]$. Let $p>2$, we have there exists a $C>0$ s.t $\int_\Lambda \max_{j\le n}|m_j|^p\,d\mu\le Cn^{p/2}$ for all $n\ge1$, by Burkholder's inequality [3]. By Markov's inequaility:

$$\mu(\max_{j\le n}|m_j|\ge\lambda\sqrt{n}) = \mu(\max_{j\le n}|m_j|^p\ge\lambda^p n^{p/2}) \le \int_\Lambda \max_{j\le n}|m_j|^p\,\frac{d\mu}{\lambda^p n^{p/2}} \le C\lambda^{-p}$$

hence,

$$\limsup_{n\to\infty}\mu(\max_{j\le n}|m_j|\ge\lambda\sqrt{n})\le C\lambda^{-p}$$

and

$$\lambda^2\limsup_{n\to\infty}\mu(\max_{j\le n}|m_j|\ge\lambda\sqrt{n})\le C\lambda^{2-p}\to 0 \text{ as } \lambda\to\infty$$

so $\{M_n\}$ is tight in $C[0,1]$ and the WIP follows. □

Remark 4.6. The WIP extends, using the Cramer-Wold device and tightness (componentwise), to vector valued observables $v:\Lambda\to\mathbb{R}^d$.

4.1 Diversion: Martingales

We divert briefly to discuss the reason for calling $m\in\ker P$ the martingale part of the decomposition described in (2). Let $(\Lambda,\mathcal{B},\mu)$ be the underlying probability space. We see that $\mathcal{B}\supset f^{-1}\mathcal{B}\supset f^{-2}\mathcal{B}\supset\ldots$ is a decreasing sequence of σ-algebras, and $m\circ f^n$ is $f^{-n}\mathcal{B}$-measurable for every $m\in L^1(\Lambda)$ and every $n\ge0$. Supposing $f:\Lambda\to\Lambda$ were invertible we could define the increasing filtration $\mathcal{F}_n=f^n\mathcal{B}$ and consider the backward Birkhoff sums $m_n^-=\sum_{j=-n}^{-1}m\circ f^j$. However, f is not necessarily invertible so to give meaning to this backward sum we state the following proposition from a standard ergodic theory course.

Proposition 4.7. *Natural Extension*
Assume that $f : \Lambda \to \Lambda$ is a surjective transformation and that μ is an f-invariant probability measure. Then there exists an invertible transformation $\tilde{f} : \tilde{\Lambda} \to \tilde{\Lambda}$ with an $\tilde{f}$-invariant probability measure $\tilde{\mu}$, as well as a measure preserving projection $\pi : \tilde{\Lambda} \to \Lambda$ such that $\pi \circ \tilde{f} = f \circ \pi$. Furthermore, if μ is ergodic then $\tilde{\mu}$ can be chosen to be ergodic as well.

Let $\mathcal{F}_0 = \pi^{-1}\mathcal{B}$ and define $\mathcal{F}_n = \tilde{f}^n\mathcal{F}_0$. Firstly notice, $\mathcal{F}_{-1} = \tilde{f}^{-1}\mathcal{F}_0 = \tilde{f}^{-1}\pi^{-1}\mathcal{B} = (\pi \circ \tilde{f})^{-1}\mathcal{B} = (f \circ \pi)^{-1}\mathcal{B} = \pi^{-1}f^{-1}\mathcal{B} \subset \pi^{-1}\mathcal{B} = \mathcal{F}_0$ and so $\mathcal{F}_{-1} \subset \mathcal{F}_0$. Repeating the process gives us $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \in \mathbb{Z}$. For example, to see that $\mathcal{F}_{-2} \subset \mathcal{F}_{-1}$ we proceed as follows: $\mathcal{F}_{-2} = \tilde{f}^{-2}\pi^{-1}\mathcal{B} = \tilde{f}^{-1}(\tilde{f}^{-1} \circ \pi^{-1})\mathcal{B} = \tilde{f}^{-1}\pi^{-1}f^{-1}\mathcal{B} \subset \tilde{f}^{-1}\pi^{-1}\mathcal{B} = \mathcal{F}_{-1}$. Thus we conclude that the sequence of σ-algebras $\{\mathcal{F}_n, n \geq 1\}$ defines a filtration. Now, let $m : \Lambda \to \mathbb{R}$ be a mean zero observable lying in $L^2(\Lambda)$. Lifting this observable to $\tilde{m} = m \circ \pi : \Lambda \to \mathbb{R}$ we define the lifted forward and backward Birkhoff sums:

$$\tilde{m}_n = \sum_{j=0}^{n-1} \tilde{m} \circ \tilde{f}^j, \quad \tilde{m}_n^- = \sum_{j=-n}^{-1} \tilde{m} \circ \tilde{f}^j.$$

Since π is measure-preserving, it follows that $\tilde{m}$ is also mean-zero and that $\int_{\tilde{\Lambda}} \tilde{m}^2\, d\tilde{\mu} = \int_{\Lambda} m^2\, d\mu$. We have already seen that $\{\mathcal{F}_n, n \geq 1\}$ defines a filtration and that $\tilde{m} \circ \tilde{f}^{-j}$ is $\mathcal{F}_j$ measurable. Supposing further that $m \in \ker P$ we see that $E[\tilde{m} \circ \tilde{f}^{-j} | \mathcal{F}_{-(j+1)}] = E[m \circ f^j | f^{-(j+1)}\mathcal{B}] \circ \pi = E[m | f^{-1}\mathcal{B}] \circ (f^j \circ \pi) = 0$. Hence it follows that $\tilde{m}_n^{-1}$ is an ergodic stationary martingale. The following result, due to Brown [2], shows that the CLT and WIP holds for $\tilde{m}_n^{-1}$. Then as $\tilde{m}_n = \tilde{m}_n^- \circ f^n$ and the measure μ is invariant the CLT and WIP holds for $\tilde{m}_n$. Moreover as π is measure preserving the result also holds for m_n.

Theorem 4.8. *Let $f : \Lambda \to \Lambda$ be an ergodic measure-preserving transformation, and let $Y \in L^2(\Lambda)$ with $E[Y] = 0$. Suppose that $S_n = \sum_{j=0}^{n-1} Y \circ f^j$ is a martingale. Then the CLT and WIP are valid. That is $n^{-1/2}S_n \to_d N(0, \sigma^2)$ where $\sigma^2 = E[Y^2]$ and if we have $Q_n \in C[0,1]$ such that $Q_n(t) = n^{-1/2}S_{[nt]} + O(n^{-1/2})$, then $Q_n \to_w W$ where W is a Brownian motion with variance σ^2.*

Remark 4.9. It follows that the assumptions in the decomposition (2) can be relaxed to holding in $L^2(\Lambda)$. We need only show $\sup_{t\in[0,1]}|W_n(t) - M_n(t)| \to 0$ a.e. if v, m and χ lie L^2. To prove this we need the following technical result from real analysis and the consequent corollary.

Proposition 4.10. *Let a_n be a sequence of real numbers and suppose that $n^{-1/2}a_n \to 0$ as $n \to \infty$ then $n^{-1/2}\max_{k\leq n}|a_k| \to 0$ as $n \to \infty$.*

Proof. Given $\epsilon > 0$, we know that there exists $N_1 \geq 1$ such that $n^{-1/2}|a_n| \leq \epsilon$ for every $n \geq N_1$. Now let $M = \max_{k\leq N_1}\{|a_k|\}$ and pick sufficiently large N_2 such that $N_2 \geq N_1$ and $N_2^{-1/2}M \leq \epsilon$. Then, for $n \geq N_2$, we see that

$$n^{-1/2}\max_{k\leq N_1}|a_k| \leq N_2^{-1/2}M \leq \epsilon$$

and

$$n^{-1/2}\max_{N_1\leq k\leq n}|a_k| \leq \max_{N_1\leq k\leq n} k^{-1/2}|a_k| \leq \max_{N_1\leq k} k^{-1/2}|a_k| \leq \epsilon.$$

Thus $n^{-1/2}\max_{k\leq n}|a_k| \leq \epsilon$ as required. □

Corollary 4.11. *Assume that v, m and χ lie in $L^2(\Lambda)$, then $\sup_{t\in[0,1]}|W_n(t) - M_n(t)| \to 0$ a.e.*

Proof. Since

$$|W_n(t) - M_n(t)| = |W_n(t) - n^{-1/2}v_{[nt]} + n^{-1/2}v_{[nt]} - n^{-1/2}m_{[nt]} + n^{-1/2}m_{[nt]} - M_n(t)|,$$

we have

$$\begin{aligned}
&\sup_{t\in[0,1]} |W_n(t) - M_n(t)| \\
&\quad \leq \sup_{t\in[0,1]} |W_n(t) - n^{-1/2}v_{[nt]}| + \sup_{t\in[0,1]} |n^{-1/2}v_{[nt]} - n^{-1/2}m_{[nt]}| + \sup_{t\in[0,1]} |n^{-1/2}m_{[nt]} - M_n(t)| \\
&\quad \leq n^{-1/2} \max_{k\leq n} |v \circ f^k| + 2n^{-1/2} \max_{k\leq n} |\chi \circ f^k| + n^{-1/2} \max_{k\leq n} |m \circ f^k|.
\end{aligned}$$

Now, let us consider a general observable w lying in $L^2(\Lambda)$. Since w^2 lies in $L^1(\Lambda)$, it is an easy consequence of the Ergodic Theorem that:

$$n^{-1}w^2 \circ f^n = n^{-1}\sum_{k=1}^{n} w^2 \circ f^k - n^{-1}\sum_{k=1}^{n-1} w^2 \circ f^k \to \int_\Lambda w^2 \, d\mu - \int_\Lambda w^2 \, d\mu = 0 \ a.e.$$

It follows that $n^{-1/2}|w \circ f^n| \to 0$ a.e. as $n \to \infty$, by taking square roots. Thus, by Proposition 4.10 with $a_n = w \circ f^n$ we see that $n^{-1/2}\max_{k\leq n}|w \circ f^k| \to 0$ a.e. Then, since v, m and χ are in $L^2(\Lambda)$, we confirm that $\sup_{t\in[0,1]}|W_n(t) - M_n(t)| \to 0$ a.e. This concludes the proof and justifies Remark 4.9. □

5 An Iterated Weak Invariance Principle

This section discusses a relatively recent statistical limit law which follows from $\sum_{k=1}^{\infty}\|P^k v\|_\infty < \infty$. The result is due to Kelly and Melbourne [7], and here we give an overview of the main arguments rather than full proofs. Firstly, we present motivation for an Iterated WIP using Stochastic Differential Equations.

5.1 Motivation: Stochastic Differential Equations

Let X be a d-dimensional Itô process defined by a Stochastic Differential Equation (SDE) of the form:

$$dX = a(X)dt + b(X)dW \tag{3}$$

where $a : \mathbb{R}^d \to \mathbb{R}^d$ is C^{1+}, $b : \mathbb{R}^d \to \mathbb{R}^{d\times e}$ is C^{2+}, and W is an e-dimensional Brownian Motion with $e \times e$-dimensional covariance matrix Σ. Given a sequence of e-dimensional processes W_n with smooth sample paths, we consider the sequence of Ordinary Differential Equations (ODEs)

$$dX_n = a(X_n)dt + b(X_n)dW_n \tag{4}$$

where $dW_n = \dot{W}_n dt$. We suppose that there is an initial condition $\eta \in \mathbb{R}^n$ such that $x(0) = X_n(0) = \eta$. Suppose W_n satisfies the the WIP discussed in the previous section but on $C([0,T], \mathbb{R}^e)$ for $T > 0$. Recalling that (3) is simply notation for an integral equation an important question in stochastic analysis is whether $X_n \to_w X$ in $C([0,T], \mathbb{R}^d)$ for a suitable interpretation of this integral equation. Explicitly we ask how should we interpret $\int b(X) \star dW$, where we use $\star$ to denote the fact the interpretation has yet to be chosen. Wong and Zakai [19] gave sufficient conditions under which one should give the stochastic integral the Stratonovich interpretation. These conditions are satisfied in the one-dimensional case but there are counterexamples in higher dimensions due to McShane [12] and Sussman [16]. Below, we give an illustrative example:

let $d = e = 2$ take a to be identically 0 and $b(x_1, x_2) = \begin{pmatrix} 1 & 0 \\ 0 & x_1 \end{pmatrix}$.
The ODE expressed in (4) becomes

$$dX_n^1 = dW_n^1, \quad dX_n^2 = X_n^1 dW_n^2.$$

Thus assuming the initial condition $\eta = 0$ we obtain $X_n^1(t) = W_n^1(t)$ for all t and hence $X_n^2(t) = \int_0^t W_n^1 dW_n^2$. In general the weak convergence $W_n \to_w W$ does not determine the weak convergence of $\int_0^t W_n^1 dW_n^2$. However, due to rough path theory [9], this is the key hurdle in solving the motivating problem. In a general framework, consider the family of smooth process $\mathbb{W}_n \in C([0,\infty), \mathbb{R}^{e\times e})$,

$$\mathbb{W}_n^{ij}(t) = \int_0^t W_n^i dW_n^j, 1 \leq i, j \leq e.$$

Under some additional work the weak limit of $(W_n, \mathbb{W}_n)$ determines the weak limit of X_n defined in (4) and the correct interpretation for the underlying integral in equation (3). Kelly and Melbourne proved an Iterated WIP [7] and hence gave a solution to the problem under mild assumptions. In this project we do not concern ourselves with the details of rough path theory but give an exposition of the results in [7] proving the Iterated WIP and seek to find methods to extend their results.

5.2 Statement of Results:

Here we give the statement of the main results and a discussion of elements required in the proofs. We denote by $D[0,\infty)$ the space of cádlág function endowed with the standard uniform topology, see Billingsley [1] for further details of this space. We work in this space, rather than $C[0,\infty)$, for technical reasons. Furthermore, we define $\mathbb{W}_n(t)$ as the following; where the true definition should be taken as the iterated sum

$$\mathbb{W}_n^{ij}(t) = \int_0^t W_n^i dW_n^j = n^{-1} \sum_{1\leq k<l\leq [nt]-1} v^i \circ f^k v^j \circ f^l.$$

Theorem 5.1. *Kelly-Melbourne*
Let $f : \Lambda \to \Lambda$ be a mixing transformation on the probability space (Λ, μ). Suppose that the MCD (2) holds in $L^2(\Lambda)$. Then $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ in $D([0,\infty), \mathbb{R}^d \times \mathbb{R}^{d\times d})$. Here, W is a d-dimensional Brownian Motion with covariance Σ where $\Sigma^{ij} = \int_\Lambda v^i v^j \, d\mu + \sum_{k=1}^\infty \int_\Lambda v^i v^j \circ f^k + v^j v^i \circ f^k \, d\mu$ and

$$\mathbb{W}^{ij}(t) = \int_0^t W^i dW^j + t \sum_{m=1}^\infty \int_\Lambda v^i v^j \circ f^m \, d\mu$$

and we give the stochastic integral the Itô interpretation.

Definition 5.2. We say $v, \hat{v} : \Lambda \to \mathbb{R}^d$ are L^2*-cohomologous* if there exists $\chi : \Lambda \to \mathbb{R}^d$ such that $\chi \in L^2(\Lambda)$ and $v = \hat{v} + \chi \circ f - \chi$.

Firstly, we need a cohomological invariance result for iterated integrals.

Theorem 5.3. *Suppose that $f : \Lambda \to \Lambda$ is a mixing transformation and that $v, \hat{v} \in L^2(\Lambda, \mathbb{R}^e)$ are L^2 cohomologous and mean-zero. Let, $1 \leq i, j \leq e$. Then the limit*

$$\lim_{n\to\infty} \sum_{m=1}^n \int_\Lambda v^i v^j \circ f^m - \hat{v}^i \hat{v}^j \circ f^m \, d\mu$$

exists and

$$\mathbb{W}_n^{ij}(t) - \widehat{\mathbb{W}}_n^{ij}(t) \to t \sum_{m=1}^{\infty} \int_{\Lambda} v^i v^j \circ f^m - \hat{v}^i \hat{v}^j \circ f^m \, d\mu \ a.e$$

as $n \to \infty$, uniformly on compact subsets of $[0, \infty)$. In particular, the weak limits of the processes

$$\mathbb{W}_n^{ij}(t) - \sum_{m=1}^{\infty} \int_{\Lambda} v^i v^j \circ f^m \, d\mu, \quad \widehat{\mathbb{W}}_n^{ij}(t) - \sum_{m=1}^{\infty} \int_{\Lambda} \hat{v}^i \hat{v}^j \circ f^m \, d\mu$$

coincide.

Corollary 5.4. *Suppose that $f : \Lambda \to \Lambda$ is a mixing transformation and that $v, \hat{v} \in L^2(\Lambda, \mathbb{R}^e)$ are L^2 cohomologous and mean-zero. Suppose that $(\widehat{W}_n, \widehat{\mathbb{W}}_n) \to_w (\widehat{W}, \widehat{\mathbb{W}})$ in $D([0, \infty), \mathbb{R}^e \times \mathbb{R}^{e \times e})$ as $n \to \infty$. Then $(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$ in $D([0, \infty), \mathbb{R}^e \times \mathbb{R}^{e \times e})$, where $W = \widehat{W}$ and*

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

This demonstrates that Theorem 5.1 reduces to proving a result for the martingale part of the MCD. To this end, define the following cádlág processes $M_n \in D([0, \infty), \mathbb{R}^e)$ and $\mathbb{M}_n \in D([0, \infty), \mathbb{R}^{e \times e})$ by

$$M_n(t) = n^{-1/2} \sum_{l=0}^{[nt]-1} m \circ f^l, \ \mathbb{M}_n^{ij} = \int_0^t M_n^i dM_n^j = n^{-1} \sum_{0 \le k < l \le [nt]-1} m^i \circ f^k m^j \circ f^l.$$

Then Theorem 5.1 follows from the proceeding lemma:

Lemma 5.5. *Suppose that f is an ergodic transformation and that $m \in L^2(\Lambda, \mathbb{R}^e)$ with $Pm = 0$. Then $(M_n, \mathbb{M}_n) \to_w (W, I)$ in $D([0, \infty), \mathbb{R}^e \times \mathbb{R}^{e \times e})$, as $n \to \infty$, where W is an e-dimensional Brownian Motion with covariance matrix $\Sigma = \int_{\Lambda} mm^T \, d\mu$ and $I^{ij}(t) = \int_0^t W^i dW^j$.*

Here we show that Theorem 5.1 can be deduced from Corollary 5.4 and Lemma 5.5.

Proof. We apply Corollary 5.4 with $\hat{v} = m$. First note that for all $k \geq 1$

$$\int_{\Lambda} mm^T \circ f^k \, d\mu = \int_{\Lambda} mU^k m^T \, d\mu = \int_{\Lambda} P^k mm^T \, d\mu = 0,$$

because of duality and the fact that $Pm = 0$. It follows from Theorem 5.3 that $\sum_{k=1}^{n} \int vv^T \circ f^k \, d\mu$ is convergent. Then, by Corollary 5.4 we have that:

$$(W_n, \mathbb{W}_n) \to_w (W, \mathbb{W})$$

where $W(t)$ is an e-dimensional Brownian Motion with covariance $\Sigma = \int_{\Lambda} mm^T \, d\mu$ and $\mathbb{W}(t) = I(t) + t \sum_{k=1}^{\infty} \int vv^T \circ f^k \, d\mu$. It remains to prove that:

$$\Sigma^{ij} = \int_{\Lambda} v^i v^j \, d\mu + \sum_{k=1}^{\infty} \int_{\Lambda} v^i v^j \circ f^k + v^j v^i \circ f^k \, d\mu.$$

Define v_n and m_n as before. Then using the fact that μ is invariant we deduce that:

$$\int_{\Lambda} m_n m_n^T \, d\mu = \sum_{0 \le l, k \le n-1} \int_{\Lambda} m \circ f^l m^T \circ f^k \, d\mu = \sum_{0 \le l, k \le n-1} \int_{\Lambda} mm^T \, d\mu = n\Sigma.$$

Notice also that $c^T\Sigma c = n^{-1}\int_\Lambda (c^T m_n)^2$ for all $c \in \mathbb{R}^e$ and $n \geq 1$. Thus $n^{1/2}(c^T\Sigma c)^{1/2} = \|c^T m_n\|_2$. Using the MCD described in (2) we have $v_n - m_n = \chi \circ f^n - \chi$. Then the following uses that μ is an invariant measure for f and the reverse triangle inequality:

$$||\|c^T v_n\|_2 - n^{1/2}(c^T\Sigma c)^{1/2}| = ||\|c^T v_n\|_2 - \|c^T m_n\|_2| \leq \|c^T(v_n - m_n)\|_2 \leq 2\|c^T\chi\|_2.$$

That is $\lim_{n\to\infty} n^{-1/2}\|c^T v_n\|_2 = (c^T\Sigma c)^{1/2}$ which is equivalent to:

$$\Sigma = \lim_{n\to\infty} n^{-1}\int_\Lambda v_n v_n^T \, d\mu = \lim_{n\to\infty}\int_\Lambda W_n(1)W_n(1)^T \, d\mu. \tag{5}$$

Now, let $a_r = \int_\Lambda v \circ f^r \, d\mu$ and $s_k = \sum_{r=1}^k a_r$. We notice that

$$\sum_{k=1}^n s_k = \sum_{k=1}^n\sum_{r=1}^k a_r = \sum_{r=1}^{n-1} na_r - ra_r.$$

Then we compute that:

$$\sum_{0\leq l<p\leq n-1}\int_\Lambda v\circ f^{p-l}v^T \, d\mu = \sum_{1\leq r<n}(n-r)\int_\Lambda v\circ f^r v^T \, d\mu = \sum_{1\leq r<n}(n-r)a_r = \sum_{k=1}^n s_k$$

where we let $r = p - l$. Hence it follows that,

$$\lim_{n\to\infty} n^{-1}\sum_{0\leq l<p\leq n-1}\int_\Lambda v\circ f^{p-l}v^T \, d\mu = \lim_{n\to\infty} n^{-1}\sum_{k=1}^n s_k = \lim_{n\to\infty} s_n = \sum_{k=1}^\infty\int_\Lambda v\circ f^k v^T \, d\mu \tag{6}$$

where we used the fact that the average of a convergence sequence converges to the same limit as the convergent sequence itself, that is the Stolz-Cesàro Theorem. Likewise,

$$\lim_{n\to\infty} n^{-1}\sum_{0\leq l<p\leq n-1}\int_\Lambda v(v\circ f^{l-p})^T \, d\mu = \sum_{k=1}^\infty\int_\Lambda v(v\circ f^k)^T \, d\mu. \tag{7}$$

Writing,

$$\begin{aligned} n^{-1}\int_\Lambda v_n v_n^T \, d\mu &= n^{-1}\sum_{0\leq l,k\leq n-1}\int_\Lambda v\circ f^l(v\circ f^k)^T \, d\mu \\ &= \int_\Lambda vv^T \, d\mu + n^{-1}\sum_{0\leq l<p\leq n-1}\int_\Lambda v\circ f^{p-l}v^T \, d\mu + n^{-1}\sum_{0\leq l<p\leq n-1}\int_\Lambda v(v\circ f^{l-p})^T \, d\mu. \end{aligned} \tag{8}$$

Thus taking the limit as $n \to \infty$ in (8) and using the results obtained at (6) and (7) we see that, from (5):

$$\Sigma = \int_\Lambda vv^T \, d\mu + \sum_{k=1}^\infty\int_\Lambda v\circ f^k v^T + v(v\circ f^k)^T \, d\mu,$$

as required, concluding the proof. □

5.3 Discussion and Proofs of Results

The previous section presented two results: Theorem 5.3 and Lemma 5.5 which led us to conclude the last section with a proof of Theorem 5.1. This was the main theorem in this section and the central result of the project. In the following discussion of Theorem 5.3 and Lemma 5.5 we present a full proof of Theorem 5.3 and outline the elements of the proof of Lemma 5.5. For a full proof of Lemma 5.5 see Subsection 4.2 of [7]. Here is the proof of Theorem 5.3:

Proof. Define $v = \hat{v} + a$, denoting $a = \chi \circ f - \chi$. Then let

$$A_n(t) = n^{-1/2} \sum_{n=1}^{[nt]-1} a \circ f^j$$

and notice:

$$\mathbb{W}_n^{ij}(t) - \widehat{\mathbb{W}}_n^{ij}(t) = \int_0^t W_n^i dW_n^j - \int_0^t \widehat{W}_n^i d\widehat{W}_n^j = \int_0^t A_n^i dW_n^j + \int_0^t \widehat{W}_n^i dA_n^j$$

where the first equality followed by definition and the second by writing out the sums and substituting for a. Now, using telescoping sums:

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

Hence

$$\int_0^t A_n^i dW_n^j \to \int_\Lambda \chi^i v^j \, d\mu - \chi^i \int_\Lambda v^j \, d\mu = \int_\Lambda \chi^i v^j \, d\mu \ a.e.$$

as $n \to \infty$, using the Ergodic Theorem and fact that v is a mean-zero observable. Likewise

$$\begin{aligned}\int_0^t \widehat{W}_n^i dA_n^j =& n^{-1} \sum_{k=1}^{[nt]-1} \sum_{l=0}^{k-1} \hat{v}^i \circ f^l a^j \circ f^k = n^{-1} \sum_{l=1}^{[nt]-1} \sum_{k=l+1}^{[nt]-2} \hat{v}^i \circ f^l a^j \circ f^k \\ =& n^{-1} \sum_{l=1}^{[nt]-1} \hat{v}^i \circ f^l (\chi^j \circ f^{[nt]-2} - \chi^j \circ f^{l+1}) \\ =& \chi \circ f^{[nt]-2} \sum_{l=1}^{[nt]-1} \hat{v}^i \circ f^l - n^{-1} \sum_{l=1}^{[nt]-1} (\hat{v}^i U \chi^j) \circ f^l\end{aligned}$$

and so,

$$\int_0^t \widehat{W}_n^i dA_n^j \to - \int_\Lambda \hat{v}^i U \chi^j \, d\mu \ a.e.$$

as $n \to \infty$, again by the Ergodic Theorem. Thus we have shown:

$$\mathbb{W}_n^{ij}(t) - \widehat{\mathbb{W}}_n^{ij}(t) \to \int_\Lambda \chi^i v^j - \hat{v}^i U \chi^j \, d\mu \ a.e. \tag{9}$$

Now,

$$v^i v^j \circ f^k - \hat{v}^i \hat{v}^j \circ f^k = (v^i - \hat{v}^i) v^j \circ f^k + \hat{v}^i (v^j - \hat{v}^j) \circ f^k = a^i v^j \circ f^k + \hat{v}^i a^j \circ f^k$$

hence it follows that

$$\begin{aligned}
\sum_{k=1}^{n} \int_\Lambda v^i v^j \circ f^k \, d\mu - &\sum_{k=1}^{n} \int_\Lambda \hat{v}^i \hat{v}^j \circ f^k \, d\mu \\
&= \sum_{k=1}^{n} \int_\Lambda a^i v^j \circ f^k + \sum_{k=1}^{n} \int_\Lambda \hat{v}^i a^j \circ f^k \, d\mu \\
&= \{\text{Using the fact that } \mu \text{ is an invariant measure for } f\} \\
&= \sum_{k=1}^{n} \int_\Lambda (\chi^i \circ f^{n-k+1} - \chi^i \circ f^{n-k}) v^j \circ f^n + \hat{v}^i (\chi \circ f^{j+1} - \chi \circ f^j) \, d\mu \\
&= \{\text{Swapping integral and sum}\} \\
&= \int_\Lambda \chi^i v^j \, d\mu - \int_\Lambda \hat{v}^i U \chi^j \, d\mu + L_n. \qquad (10)
\end{aligned}$$

Then,

$$L_n = \int_\Lambda (\hat{v}^i \chi^j \circ f^{n+1} - \chi^i v^j \circ f^n) \, d\mu \to \int_\Lambda \hat{v}^i \, d\mu \int_\Lambda \chi^j \, d\mu - \int_\Lambda \chi^i \, d\mu \int_\Lambda v^j \, d\mu = 0$$

due to mixing. The result then follows from (9) and (10). □

Recall from the previous section that Lemma 5.5 was central to proving the Iterated WIP. Though there are many elements to the proof, none of which are particularly difficult to prove, it is indeed a considerable amount of work. Instead, we give an indication of what is needed to be proved at the same time as giving a full proof of the $M_n \to W$ part of Lemma 5.5, as this should indicated some of the steps required. First let us recall the discussion of martingales in Subsection 4.1. Supposing that M_n was a martingale we would wish to apply the following well-known result by Kurtz and Protter, which is Theorem 2.2 in [8]:

Theorem 5.6. *Kurtz and Protter*
For each n, let (X_n, Y_n) be an adapted process in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{d\times e})$ and let X_n be a semi-martingale. Fix any $\delta > 0$ (allowing the case $\delta = \infty$) and define:

$$X_n^\delta = X_n - J_\delta(X_n) = X_n - \sum_{s\le t} h_\delta(|(X_n(s) - X_n(s-))|)(X_n(s) - X_n(s-))$$

where $h_\delta(r) = (1-\delta/r)^+$. (Note that X_n^δ is also a semi-martingale). Let $X_n^\delta = M_n^\delta + A_n^\delta$ be the semi-martingale decomposition, where M_n^δ is an adapted local martingale and A_n^δ a process with finite variation. Now make the following assumption: for all $\alpha > 0$, there exist stopping times $\{\tau_n^\alpha\}$ such that

$$P(\tau_n^\alpha \le \alpha) \le \frac{1}{\alpha} \text{ and } \sup_n E[[M_n^\delta]_{t\wedge\tau_n^\alpha} + T_{t\wedge\tau_n^\alpha}(A_n^\delta)] < \infty \qquad (11)$$

where for every $t > 0$ we have $T_t(A) = \sup \sum |A(t_{i+1}) - A(t_i)|$ and the supremum is taken over partitions of $[0,t]$. If $(X_n, Y_n) \to_w (X, Y)$ in the Skorohod topology $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{d\times e})$, then X is a semi-martingale with respect to the filtration to which X and Y are adapted and $\int_0^t Y_n dX_n \to_w \int_0^t Y dX$ in the Skorohod topology on $D([0,\infty), \mathbb{R}^d)$. If we replace weak convergence in the assumption by convergence in probability then the integral converges in probability.

Unfortunately, M_n is not a martingale. However recalling that by passing to the natural extension and working with backward Birkhoff sums (see Subsection 4.1) we have that

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

is an ergodic stationary martingale. Likewise define

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

where $(\widetilde{M}_n, \widetilde{\mathbb{M}}_n) = (M_n, \mathbb{M}_n) \circ \pi$. Since π was measure preserving, proving $(M_n, \mathbb{M}_n) \to (W, I)$ is equivalent to proving $(\widetilde{M}_n, \widetilde{\mathbb{M}}_n) \to (W, I)$. We claim that we can relate the convergence of $(\widetilde{M}_n, \widetilde{\mathbb{M}}_n)$ to $(\widetilde{M}_n^-, \widetilde{\mathbb{M}}_n^-)$. This is useful because we can use Kurtz and Protter's Theorem on the pair $(\widetilde{M}_n^-, \widetilde{\mathbb{M}}_n^-)$, which we verify in the following proposition.

Proposition 5.7. *$(\widetilde{M}_n^-, \widetilde{\mathbb{M}}_n^-) \to_w (W, I)$ in $D([0,\infty), \mathbb{R}^e \times \mathbb{R}^{e\times e})$.*

Proof. Our proof is in the setting of Theorem 5.6 with $\delta = \infty$ and A_n identically 0. Hence, we are in the simpler case of $h_\delta = 0$ and so $X_n^\delta = X_n$. We have already shown that $\widetilde{M}_n^-$ is a martingale (see Subsection 4.1) and hence a semi-martingale. Moreover, we see that

$$E[\widetilde{M}_n^{j,-}(t)^2] = n^{-1} \| \sum_{k=[nt]-1}^{-1} \tilde{m}^j \circ \tilde{f}^k \|_2^2 = t \int_{\tilde{\Lambda}} (\tilde{m}^j)^2 \, d\tilde{\mu}$$

which follows from the calculation performed in proof of Theorem 5.1, and notably didn't require the conclusion of Lemma 5.5 for consistency. Clearly $E[\widetilde{M}_n^{j,-}(t)^2] = t \int_{\tilde{\Lambda}} (\tilde{m}^j)^2 \, d\tilde{\mu}$ is independent of n and so assumption (11) is satisfied in a trivial way. Then we have the WIP for ergodic stationary L^2 martingales, that is Theorem 4.8, gives $\widetilde{M}_n^- \to_w W$ in $D([0,\infty), \mathbb{R}^e)$. This tells us $(\widetilde{M}_n^{i,-}, \widetilde{M}_n^{j,-}) \to_w (W^i, W^j)$ in $D([0,\infty), \mathbb{R}^2)$. Thus by Theorem 5.6 with $(X_n, Y_n) = (\widetilde{M}_n^{j,-}, \widetilde{M}_n^{i,-})$ we have that

$$\widetilde{\mathbb{M}}_n^{ij,-}(t) = \int_0^t \widetilde{M}_n^{i,-} d\widetilde{W}_n^{j,-} \to_w \int_0^t W^i dW^j$$

as $n \to \infty$ in $D[0,\infty)$, which concludes the proof. ☐

We now show that $M_n \to W$. Note for each result we prove there is analogous statement for $\mathbb{M}_n$ but, in general, the proofs will require considerably more work. We refer the reader to [7] for the results concerning $\mathbb{M}_n$.

Lemma 5.8. *Let $g(u)(t) = u(T) - u(T-t)$. Then $\widetilde{M}_n \circ \tilde{f}^{-nT} = g(\widetilde{M}_n^-) + F_n$ where $\sup_{t\in[0,T]} F_n(t) \to 0$ a.e.*

Proof. First we compute that:

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

Now, since $F_n(t)$ contains at most one term, we may see that $|F_n(t)| \le n^{-1/2} |\max_{k=1,\dots,[nT]} \tilde{m} \circ \tilde{f}^{-k}|$. Without loss of generality we can assume $e = 1$ by working component-wise. By the Ergodic Theorem, we have that $n^{-1} \sum_{k=1}^n \tilde{m}^2 \circ \tilde{f}^k \to \int_{\tilde{\Lambda}} \tilde{m}^2 \, d\tilde{\mu}$ and so $n^{-1} \tilde{m}^2 \circ \tilde{f}^{-n} \to 0$. It follows that $n^{-1} \max_{k=1,\dots,[nT]} \tilde{m}^2 \circ \tilde{f}^{-k} \to 0$ a.e. and so $\sup_{t\in[0,T]} F_n(t) \to 0$ as required. ☐

We state the following proposition without proof. We refer the reader to Skorohod [15] and Kelly-Melbourne [7] for the details:

Proposition 5.9. *Denote by $\tilde{D}([0,T],\mathbb{R}^q)$ the space of cáglád functions with the standard Skorohod $\mathcal{J}_1$ topology. Suppose $A_n = B_n + F_n$ where $A_n \in D([0,T],\mathbb{R}^q)$, $B_n \in \tilde{D}([0,T],\mathbb{R}^q)$ and $F_n \to 0$ uniformly in probability. If Z has continuous sample paths and $B_n \to_w Z$ in $\tilde{D}([0,T],\mathbb{R}^q)$, then $A_n \to_w Z$ in $D([0,T],\mathbb{R}^q)$.*

Corollary 5.10. *$\widetilde{M}_n \to_w g(W)$ in $D([0,T],\mathbb{R}^e)$ as $n \to \infty$.*

Proof. Clearly $g(u)$ is a continuous functional. Hence it follows by Proposition 5.7 and the Continuous Mapping Theorem that $g(\widetilde{M}_n^-) \to_w g(W)$ in $\tilde{D}([0,T],\mathbb{R}^e \times \mathbb{R}^{e\times e})$. Then by Lemma 5.8, $\widetilde{M}_n \circ \tilde{f}^{-nT} = g(\widetilde{M}_n^-) + F_n$ where F_n converges uniformly in probability to 0. Thus by Proposition 5.9 the result follows. □

The preceding corollary and the following lemma complete the proof that $M_n \to W$.

Lemma 5.11. *$g(W) =_d W$ in $D([0,T],\mathbb{R}^e)$.*

Proof. $g(W) = W(T) - W(T-t)$ and $W(t)$ are both Gaussian with continuous sample paths. $g(W)(0) = W(T) - W(T) = 0$ and $W(0) = 0$ so $g(W)(0) = W(0)$. Then

$$\begin{aligned} Cov(g(W)(t_1), g(W)(t_2)) =& E[(g(W)(t_1) - E[g(W)(t_1)])(g(W)(t_2) - E[g(W)(t_2))] \\ =& E[(W(t_1) - E[W(t_1)])(W(t_2) - E[W(t_2))] \\ =& (t_1 \wedge t_2)\Sigma. \end{aligned}$$

Hence, $Cov(g(W)(t_1), g(W)(t_2)) = (t_1 \wedge t_2)\Sigma$ for all $t_1, t_2 \in [0,T]$. Thus we conclude $g(W) =_d W$. □

We conclude this section with a number of remarks:

Remark 5.12. The results presented here hold true only for non-invertible maps. However Kelly and Melbourne show in Section 5 of [7] that the results can be extend to invertible maps as well.

Remark 5.13. In the context of Proposition 2.8 the result we proved holds for a class of non-uniformly expanding maps with sufficiently rapid decay of correlations.

Remark 5.14. The MCD was the main tool used to prove statistical limit laws so far. It proved especially useful in the Iterated WIP because the coboundary telescoped under one of the sums. In the CLT and WIP this led to something that was asymptotic negligible but this is no longer true in the iterated case. Theorem 5.3 indeed shows that the coboundary is not negligible but highlights that our processes can be adjusted appropriately by a drift to take into account the coboundary. Despite this, our intuition suggests that the MCD is a some what rigid tool.

The precise form of the coboundary is what enabled us to prove our theorems but in essence we relied on the fact that there was a sufficient amount of cancellation due to telescoping sums. This leads us to think about how the theorem could possibly be generalized, if indeed at all. In the following sections we will discuss recent general assumptions for the CLT and WIP which do not require a MCD. Our exploration of these results will lead us to discuss whether any of these methods could be used to prove an Iterated WIP.

6 The Maxwell-Woodroofe Condition

This section is purely a discussion of the literature that motivates the rest of this project. Recently, Maxwell and Woodroofe [10] presented cutting-edge techniques and proved a CLT for $(v \circ f^k)_{k=0,1,\ldots}$ under the following condition:

$$\sum_{n=1}^{\infty} n^{-3/2} \| \sum_{k=1}^{n-1} P^k v \|_2 < \infty. \tag{12}$$

To be more precise, their condition was in the setting of Markov-Chains but this is how the assumption reads in the context of ergodic transformations. This assumption is far more relaxed than the MCD setting. The CLT was proved under condition (12) in the setting of ergodic transformations in 2004 by M.Tyran-Kamińska in [17] who used the approach of Maxwell and Woodroofe and analyzed the perturbed solution of the Poisson Equation. We present these results in the next section in the hope that it will give us a new approach to proving an Iterated WIP.

In 2005 Peligrad and Utev proved, under condition (12), a new maximal inequality for stationary sequences [18], from which they deduced the WIP under this condition. Their approach was entirely different since they analyzed small blocks and applied their maximal inequality to the sums of random variables on these blocks. From this, they deduced that each of these small blocks could be approximated by stationary martingale differences. Using a counterexample from Markov-Chain theory they deduce that, in some sense, the assumption (12) is optimal. In 2013 Merlevéde and Peligrad further extended the maximal inequality [14].

In 2012 Cuny and Merlevéde proved the quenched CLT and WIP [5] under condition (12) again using martingale differences. Furthermore, in 2014, Cuny proved that under this condition the Law of the Iterated Logarithm (LIL) holds and an almost-sure invariance principle (ASIP) [4] - again using martingale differences. By further adapting the Markov-Chain counterexample he also proves optimality of his results.

The wealth of statistical limit laws proved under the Maxwell-Woodroofe condition suggest it is worth investigating the Iterated WIP under this condition. Moreover there is a suggestion that the above condition is in some sense optimal. We conjecture the following results:

Conjecture 6.1. *Let $f : \Lambda \to \Lambda$ be a mixing transformation on the probability space $(\Lambda, \mathcal{F}, \mu)$. Assume that v is a mean zero observable and $v \in L^2(\Lambda, \mathbb{R}^e)$. Suppose also that* (12) *holds. Then,*

$$\lim_{n\to\infty} \sum_{m=1}^{n} \int_\Lambda v^i v^j \circ f^m \, d\mu$$

exists and $\mathbb{W}_n \to_w \mathbb{W}$ in $D([0,\infty), \mathbb{R}^{e\times e})$ where,

$$\mathbb{W}^{ij}(t) = \int_0^t W^i dW^j + t \sum_{m=1}^{\infty} \int_\Lambda v^i v^j \circ f^m \, d\mu.$$

Conjecture 6.2. *Optimality*
For any non-negative sequence $a_n \to 0$ and mixing transformation f on the probability space $(\Lambda, \mathcal{F}, \mu)$ we can find a mean zero observable v lying in $L^2(\Lambda)$ such that

$$\sum_{n=1}^{\infty} a_n n^{-3/2} \| \sum_{k=0}^{n-1} P^k v \|_2 < \infty$$

but $\limsup_{n\to\infty} n^{-1}\sum_{0\le l<k\le[nt]-1} v\circ f^l v\circ f^k = \infty$. *That is* $\limsup_{n\to\infty}\mathbb{W}_n = \infty$.

7 Perturbing the Poisson Equation: Another CLT

This section is dedicated to presenting the CLT proved by Tyran-Kamińska in [17], which uses methods from [10]. Recall the assumptions under which we proved the CLT in Section 3. The requirement $Pv = 0$ is a strong assumption and, in general, might not hold. To overcome this the general approach is to write v as a sum of two functions: one for which the assumption of Theorem 7.1 are satisfied; the other which is irrelevant for the CLT to hold. That is the approach taken in this setting. We will find $\tilde{v}$ satisfying the assumption of Theorem 7.1 and also show that the sequence $(n^{-1/2}\sum_{k=0}^{n-1}(v-\tilde{v})\circ f^k)_{n=1,2,\ldots}$ converges in probability to 0. We state the main theorem of this section:

Theorem 7.1. *Let f be an ergodic transformation on the probability space $(\Lambda,\mathcal{B},\mu)$ and let v a mean-zero observable lying in $L^2(\Lambda)$. Suppose the Maxwell-Woodroofe condition* (12) *holds. Then there exists $\tilde{v}\in L^2$ such that $P\tilde{v}=0$ and $(n^{-1/2}\sum_{k=0}^{n-1}(v-\tilde{v})\circ f^k)_{n=1,2,\ldots}$ converges to 0 in $L^2(\Lambda)$ as $n\to\infty$. In particular the CLT holds for v.*

For completeness, we prove the key ideas from [10] which are used in the proof of Theorem 7.1. Let $\epsilon>0$ and define $g_\epsilon = \sum_{k=1}^{\infty}\frac{P^{k-1}v}{(1+\epsilon)^k}$, which we notice solves the perturbed Poisson equation $(1+\epsilon)g = Pg + v$. Moreover, define $v_\epsilon = g_\epsilon - UPg_\epsilon$. Note that in the literature $g = Pg+v$ is referred to as the Poisson Equation.

Lemma 7.2. *Suppose the Maxwell-Woodroofe condition* (12) *holds, then*

$$\sum_{k=1}^{\infty}\sqrt{2^{-k}}\sup_{2^{-k}\le\epsilon\le 2^{-(k-1)}}\|g_\epsilon\|_2 < \infty,$$

and so $\sqrt{\epsilon}\|g_\epsilon\|_2\to 0$ as $\epsilon\to 0$.

Proof. First, we show that:

$$g_\epsilon = \epsilon\sum_{n=1}^{\infty}\frac{\sum_{k=0}^{n-1}P^k v}{(1+\epsilon)^{n+1}}. \tag{13}$$

We notice, by changing order of summation, that

$$\epsilon\sum_{n=1}^{\infty}\frac{\sum_{k=0}^{n-1}P^k v}{(1+\epsilon)^{n+1}} = \sum_{k=0}^{\infty}P^k v\sum_{k+1}^{\infty}\frac{\epsilon}{(1+\epsilon)^{n+1}}.$$

This is equal to:

$$\sum_{k=0}^{\infty}P^k v\left(\frac{\epsilon}{(1+\epsilon)^{k+2}}\right)\sum_{n=0}^{\infty}\frac{1}{(1+\epsilon)^n} = \sum_{k=0}^{\infty}P^k v\frac{\epsilon}{(1+\epsilon)^{k+2}}\frac{1+\epsilon}{\epsilon} = \sum_{k=1}^{\infty}\frac{P^{k-1}v}{(1+\epsilon)^k}.$$

Then setting $\delta_k = 2^{-k}$ we see that for $\delta_k\le\epsilon\le\delta_{k-1}$ we have:

$$\|g_\epsilon\|_2 \le \delta_{k-1}\sum_{n=1}^{\infty}\frac{\|\sum_{k=0}^{n-1}P^k v\|_2}{(1+\delta_k)^n}.$$

Since $2\delta_k = \delta_{k-1}$, we deduce that

$$\sum_{k=1}^{\infty} \sqrt{\delta_k} \sup_{\delta_k \le \epsilon \le \delta_{k-1}} \|g_\epsilon\|_2 \le 2 \sum_{n=1}^{\infty} \left[\sum_{k=1}^{\infty} \frac{\delta_k^{3/2}}{(1+\delta_k)^n} \right] \| \sum_{l=0}^{n-1} P^l v \|_2.$$

We can compare the inner sum $\sum_{k=1}^{\infty} \frac{\delta_k^{3/2}}{(1+\delta_k)^n}$ with $\int_0^\infty 2^{-3y/2} \left(1+2^{-y}\right)^{-n} dy$. Then making the substitution $x = 2^{-y}$, the comparison is hence with the following integral:

$$C \int_0^1 \frac{\sqrt{x}}{(1+x)^n} dx.$$

The constant C takes in account that the function $\frac{\sqrt{x}}{(1+x)^n}$ has a unique maximum in the domain $[0,1]$ and so the comparison is permitted. The following calculation demonstrates that the integral (and hence the sum) is $O(n^{-3/2})$:

$$\begin{aligned}
&\int_0^1 \frac{\sqrt{x}}{(1+x)^n} dx \\
&= \int_0^{\frac{1}{n}} \frac{\sqrt{x}}{(1+x)^n} dx + \int_{\frac{1}{n}}^1 \frac{\sqrt{x}}{(1+x)^n} dx \\
&\le \int_0^{\frac{1}{n}} \sqrt{x} dx + \int_{\frac{1}{n}}^1 \frac{\sqrt{x}}{(1+x)^n} dx \\
&= \{\text{Integrating by parts}\} \\
&= \frac{2}{3n^{3/2}} + \left[\frac{2^{1-n}}{1-n} + \frac{n^{-1/2}}{(n-1)} \left(1+\frac{1}{n}\right)^{1-n} \right] + \frac{1}{2(n-1)} \int_{\frac{1}{n}}^1 \frac{x^{-1/2}}{(1+x)^{n-1}} dx \\
&\le \frac{2}{3n^{3/2}} + \left[\frac{2^{1-n}}{1-n} + \frac{n^{-1/2}}{(n-1)} \left(1+\frac{1}{n}\right)^{1-n} \right] + \frac{n^{1/2}}{2(n-1)} \int_{\frac{1}{n}}^1 \frac{1}{(1+x)^{n-1}} dx \\
&= \frac{2}{3n^{3/2}} + \left[\frac{2^{1-n}}{1-n} + \frac{n^{-1/2}}{(n-1)} \left(1+\frac{1}{n}\right)^{1-n} \right] + \frac{n^{1/2}}{2(n-1)(2-n)} \left(2^{2-n} - \left(1+\frac{1}{n}\right)^{2-n} \right).
\end{aligned}$$

Clearly the first, third and fifth terms are $O(n^{-3/2})$ by using the standard result that $(1+\frac{1}{n})^n \to e$. The remaining two terms decay exponentially and so are trivially $O(n^{-3/2})$. Thus we have shown that $\sum_{k=1}^{\infty} \frac{\delta_k^{3/2}}{(1+\delta_k)^n}$ is $O(n^{-3/2})$. This concludes the lemma by the Maxwell-Woodroofe condition. □

Lemma 7.3. *For any* $\epsilon, \delta > 0$, $\|v_\epsilon - v_\delta\|_2^2 \le (\epsilon + \delta) \left(\|g_\epsilon\|_2^2 + \|g_\delta\|_2^2 \right).$

Proof. First recall $v_\epsilon = g_\epsilon - UPg_\epsilon$ so,

$$\begin{aligned}
\int v_\epsilon v_\delta \, d\mu &= \int (g_\epsilon - UPg_\epsilon)(g_\delta - UPg_\delta) \, d\mu \qquad (14) \\
&= \int g_\epsilon g_\delta \, d\mu - \int UPg_\epsilon g_\delta \, d\mu - \int g_\epsilon UPg_\delta \, d\mu + \int UPg_\epsilon UPg_\delta \, d\mu.
\end{aligned}$$

Then it follows by the elementary properties of the Koopman and transfer operators that:

$$\int v_\epsilon v_\delta \, d\mu = \int g_\epsilon g_\delta \, d\mu - \int Pg_\epsilon Pg_\delta \, d\mu.$$

Moreover, since g_ϵ solves the equation $Pg = (1+\epsilon)g - v$, we have that

$$\int Pg_\epsilon Pg_\delta \, d\mu = \int((1+\epsilon)g_\epsilon - v)((1+\delta)g_\delta - v)$$
$$=(1+\epsilon)(1+\delta)\int g_\epsilon g_\delta \, d\mu - (1+\epsilon)\int g_\epsilon v \, d\mu - (1+\delta)\int g_\delta v \, d\mu + \int v^2 \, d\mu.$$

Hence it follows that:

$$\int v_\epsilon v_\delta \, d\mu = -(\epsilon + \delta + \epsilon\delta)\int g_\epsilon g_\delta \, d\mu + \left[(1+\epsilon)\int g_\epsilon v \, d\mu + (1+\delta)\int g_\delta v \, d\mu - \int v^2 \, d\mu\right].$$

From which we have that:

$$\int v_\epsilon^2 \, d\mu = -(2\epsilon + \epsilon^2)\int g_\epsilon^2 \, d\mu + \left[2(1+\epsilon)\int g_\epsilon v \, d\mu - \int v^2 \, d\mu\right]$$

and:

$$\int v_\delta^2 \, d\mu = -(2\delta + \delta^2)\int g_\delta^2 \, d\mu + \left[2(1+\delta)\int g_\delta v \, d\mu - \int v^2 \, d\mu\right].$$

Then, compute that:

$$\|v_\epsilon - v_\delta\|_2^2 = \int (v_\epsilon - v_\delta)(v_\epsilon - v_\delta) \, d\mu = \|v_\epsilon\|_2^2 - 2\int v_\epsilon v_\delta \, d\mu + \|v_\delta\|_2^2.$$

Furthermore, by the above calculations we have:

$$\begin{aligned}
\|v_\epsilon - v_\delta\|_2^2 =& -(2\epsilon + \epsilon^2)\|g_\epsilon\|_2^2 + 2(\epsilon + \delta + \epsilon\delta)\int g_\epsilon g_\delta \, d\mu - (2\delta + \delta^2)\|g_\delta\|_2^2 \\
\leq& -\epsilon^2\|g_\epsilon\|_2^2 + 2(\epsilon + \delta + \epsilon\delta)\int g_\epsilon g_\delta \, d\mu - \delta^2\|g_\delta\|_2^2 \\
\leq& 2(\epsilon + \delta)\|g_\epsilon\|_2\|g_\delta\|_2 - (\epsilon^2\|g_\epsilon\|_2^2 + \delta^2\|g_\delta\|_2^2 - 2\epsilon\delta\|g_\epsilon\|_2\|g_\delta\|_2) \\
\leq& 2(\epsilon + \delta)\|g_\epsilon\|_2\|g_\delta\|_2 - (\epsilon\|g_\epsilon\|_2 - \delta\|g_\delta\|_2)^2 \\
\leq& (\epsilon + \delta)(\|g_\epsilon\|_2^2 + \|g_\delta\|_2^2).
\end{aligned}$$

In the third line we applied Hölder's inequality and the last inequality follows because $0 \leq (\|g_\epsilon\|_2 - \|g_\delta\|_2)^2 = \|g_\epsilon\|_2^2 - 2\|g_\epsilon\|_2\|g_\delta\|_2 + \|g_\delta\|_2^2$. This proves the lemma. □

Proposition 7.4. *Suppose that the Maxwell-Woodroofe condition* (12) *holds then* $\tilde{v} = \lim_{\epsilon\to 0} v_\epsilon$ *exists in* $L^2(\Lambda)$.

Proof. As before, let $\delta_k = 2^{-k}$. Then

$$\|v_{\delta_k} - v_{\delta_{k-1}}\|_2^2 \leq (\delta_k + \delta_{k-1})(\|g_{\delta_k}\|_2^2 + \|g_{\delta_{k-1}}\|_2^2)$$

for every $k \geq 1$, by Lemma 7.3. Then using that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for positive a, b it follows:

$$\sum_{k=1}^{\infty}\|v_{\delta_k} - v_{\delta_{k-1}}\|_2 \leq \sum_{k=1}^{\infty}(\sqrt{\delta_k} + \sqrt{\delta_{k-1}})(\|g_{\delta_k}\|_2 + \|g_{\delta_{k-1}}\|_2) \leq \sum_{k=0}^{\infty}2((1+\sqrt{2})\sqrt{\delta_k})\|g_{\delta_k}\|_2$$

which is finite by Lemma 7.2. Thus it follows that $(v_{\delta_k})_{k\geq 0}$ is a Cauchy sequence in L^2 and since L^2 is complete we deduce that the $\lim_{k\to\infty} v_{\delta_k} := \tilde{v}$ exists in $L^2(\Lambda)$. For $0 < \epsilon < 1$ there is a unique $k = k(\epsilon)$ for which $\delta_k \leq \epsilon < \delta_{k-1}$. With this k we have

$$\|v_\epsilon - v_{\delta_{k-1}}\|_2^2 \leq (\epsilon + \delta_{k-1})(\|g_\epsilon\|_2^2 + \|g_{\delta_{k-1}}\|_2^2) \leq C\delta_k \sup_{\delta_k \leq \epsilon < \delta_{k-1}} \|g_\epsilon\|_2^2$$

and the right hand side converges to 0 as $\epsilon \to 0$ and so the $\lim_{\epsilon\to 0} v_\epsilon = \tilde{v}$ in $L^2(\Lambda)$. □

We now have everything we need to prove Theorem 7.1.

Proof. Let's us remind ourselves of the setup. Given $\epsilon > 0$, define $g_\epsilon = \sum_{k=1}^{\infty} \frac{P^{k-1}v}{(1+\epsilon)^k}$ where P denotes the transfer operator. Since $\|Pv\|_2 \leq \|v\|_2$, it follows that $g_\epsilon \in L^2(\Lambda)$ from the definition of g_ϵ. Recalling that g_ϵ solves the perturbed Poisson equation gives $v = (1+\epsilon)g_\epsilon - Pg_\epsilon$. Defining $v_\epsilon = g_\epsilon - UPg_\epsilon$ it is an easy consequence that $Pv_\epsilon = Pg_\epsilon - Pg_\epsilon = 0$. Moreover, we establish the decomposition:

$$v = v_\epsilon + \epsilon g_\epsilon + UPg_\epsilon - Pg_\epsilon. \tag{15}$$

Then by Proposition 7.4 we know that $\tilde{v} = \lim_{\epsilon \to 0} v_\epsilon$ exists in $L^2(\Lambda)$ and as $Pv_\epsilon = 0$ for every $\epsilon > 0$ it follows that $P\tilde{v} = 0$. As a result of the decomposition (15) we obtain:

$$\sum_{k=0}^{n-1}(v - \tilde{v}) \circ f^k = \sum_{k=0}^{n-1}(v_\epsilon - \tilde{v}) \circ f^k + \epsilon \sum_{k=0}^{n-1} g_\epsilon \circ f^k + U^n Pg_\epsilon - Pg_\epsilon \tag{16}$$

having telescoped the coboundary terms. Now $P(v_\epsilon - \tilde{v}) = 0$, hence by calculations from the CLT in Section 3 we have that:

$$n^{-1/2}\|\sum_{k=0}^{n-1}(v_\epsilon - \tilde{v}) \circ f^k\|_2 = \|v_\epsilon - \tilde{v}\|_2.$$

Now, we notice that there are n terms in the second sum of (16). Further as U is an isometry on $L^2(\Lambda)$ and P is a contraction on this space it follows that:

$$n^{-1/2}\|\sum_{k=0}^{n-1}(v - \tilde{v}) \circ f^k\|_2 \leq \|v_\epsilon - \tilde{v}\|_2 + \epsilon n^{1/2}\|g_\epsilon\|_2 + 2n^{-1/2}\|g_\epsilon\|_2. \tag{17}$$

Now choosing $\epsilon_n = \frac{1}{n}$, we deduce from (17) that:

$$n^{-1/2}\|\sum_{k=0}^{n-1}(v - \tilde{v}) \circ f^k\|_2 \leq \|v_{\epsilon_n} - \tilde{v}\|_2 + 3\epsilon_n^{1/2}\|g_{\epsilon_n}\|_2$$

and the right hand side of this inequality converges to 0 as $n \to \infty$, which concludes the proof. □

Remark 7.5. Firstly, the proof we have given here obtains a slightly sharper final bound than the one presented in [17]. It is important to notice the similarity of some of the steps used in this proof with proofs already presented in this project. Indeed, the decomposition constructed for v (15) is somewhat similar to the MCD (2). In both cases we exploited the telescoping coboundary terms. We remark that (15) adds useful flexibility opposed to the rigid MCD. However, this flexibility may be too much and as such impossible to prove an Iterated WIP in this relaxed setting.

Remark 7.6. Having covered the case of the CLT under condition (12), it is worth mentioning that the WIP is proved under this condition by Peligrad and Utev [18]. Their approach is different to the one presented in the preceding section. Their main result is a maximal inequality which generalizes classical results by Doob. Then, by an application of this maximal inequality, they prove the WIP. The idea is to split up the random variable into blocks and approximate each block by a stationary martingale.

8 Extending the Iterated Central Limit Theorem

Though our original goal was to generalize results about the iterated WIP; this section presents lemmas leading to a theorem and proposition which seek to relax the assumptions of the iterated CLT. In particular, we wish to look at the case where $\|P^k v\|_2$ is non-summable. The simpler setting of the Iterated CLT means that all our random variables are finite-dimensional and the convergence takes place in $\mathbb{R}$. Using the notation and ideas established in Section 7, we progress towards a useful set of assumptions to prove the Iterated CLT. The proposition shows that under standard dynamical assumption there is only one difficult assumption to check.

Recall that $g_\epsilon = \sum_{k=1}^{\infty} \frac{P^{k-1}v}{(1+\epsilon)^k}$ solves the perturbed Poisson equation $(1+\epsilon)g = Pg + v$ and that $\chi_{\epsilon_n} = Pg_{\epsilon_n}$ where $\epsilon_n = n^{-1}$, as in the proof of Theorem 7.1. We recall that $v_\epsilon = g_\epsilon - UPg_\epsilon$ and $v = z_\epsilon + \epsilon g_\epsilon$, where $z_\epsilon = v_\epsilon + U\chi_\epsilon - \chi_\epsilon$. We introduce the notation $a_n \ll b_n$ to mean that there is some constant $K \in \mathbb{R}$ such that $a_n < Kb_n$ for all $n \geq 1$.

Lemma 8.1. *Let $f : \Lambda \to \Lambda$ be a mixing transformation on the probability space $(\Lambda, \mathcal{F}, \mu)$. Assume that v is a mean zero observable and $v \in L^2(\Lambda, \mathbb{R}^e)$ and suppose the Maxwell-Woodroofe condition* (12) *holds. Then*

$$n^{-1} \sum_{0\leq k<l<n} v^i_{\epsilon_n} \circ f^k v^j_{\epsilon_n} \circ f^l \to_d \int_0^1 W^i dW^j \text{ as } n \to \infty \tag{18}$$

and

$$n^{-1} \sum_{0\leq k<l<n} \tilde{v}^i \circ f^k \tilde{v}^j \circ f^l \to_d \int_0^1 W^i dW^j \text{ as } n \to \infty. \tag{19}$$

Proof. Since $f : \Lambda \to \Lambda$ is mixing and $\tilde{v} \in \ker P$, the limit (19) follows from Lemma 5.5. For (18) we seek to apply Theorem 2.6 from Jakubowski et al. [6]. We note that $v_\epsilon \in \ker P$ and now we show that the CLT holds for v_ϵ. Recall from Section 7 that the CLT holds for $\tilde{v}$. Hence, since $P(v_\epsilon - \tilde{v}) = 0$, we have that:

$$\|n^{-1/2} \sum_{k=1}^{n-1} (v_\epsilon - \tilde{v}) \circ f^k\|_2 = \|v_\epsilon - \tilde{v}\|_2.$$

The right hand side of this converges to 0 as $\epsilon \to 0$ and thus the CLT holds for v_ϵ. To satisfy the technical condition of Theorem 2.6 [6] we appeal to Proposition 3.2(a) of [6]. That is, we need to show that $\sup_n E[\sup_{k<n} |n^{-1/2} v_{\epsilon_n} \circ f^k|] < \infty$. Clearly, the following inequalities hold:

$$E[\sup_{k<n} |n^{-1/2} v_{\epsilon_n} \circ f^k|] \leq \|\sup_{k<n} |n^{-1/2} v_{\epsilon_n} \circ f^k|\|_2 \leq 2\|n^{-1/2} \sum_{k=1}^{n} v_{\epsilon_n} \circ f^k\|_2 \leq 2\|v_{\epsilon_n}\|_2.$$

The last inequality was achieved using Burkholder's inequality. Then since $v_{\epsilon_n} \to \tilde{v}$ in L^2, there exists M such that $\|v_{\epsilon_n}\|_2 \leq M$. Thus, it follows that $\sup_n E[\sup_{k<n} |n^{-1/2} v_{\epsilon_n} \circ f^k|] \leq 2M$ and this concludes the proof. □

The following lemma shows that the error term ϵg_ϵ in the decomposition is irrelevant for the Iterated CLT to hold.

Lemma 8.2. *Let $f : \Lambda \to \Lambda$ be a mixing transformation on the probability space $(\Lambda, \mathcal{F}, \mu)$. Assume that v is a mean zero observable and $v \in L^2(\Lambda, \mathbb{R}^e)$ and suppose the Maxwell-Woodroofe condition* (12) *holds. Then*

$$n^{-1} \sum_{0\le k<l<n} v^i \circ f^k v^j \circ f^l - \sum_{k=1}^{n} \int_\Lambda v^i v^j \circ f^k \, d\mu \to_d \int_0^1 W^i dW^j \text{ as } n \to \infty \tag{20}$$

if and only if

$$n^{-1} \sum_{0\le k<l<n} z^i_{\epsilon_n} \circ f^k z^j_{\epsilon_n} \circ f^l - \sum_{k=1}^{n} \int_\Lambda z^i_{\epsilon_n} z^j_{\epsilon_n} \circ f^k \, d\mu \to_d \int_0^1 W^i dW^j \text{ as } n \to \infty. \tag{21}$$

Proof. It is sufficient to prove the difference between equations (20) and (21) converges to 0 in L^1. Momentarily fixing $\epsilon > 0$, we notice that the difference:

$$\begin{aligned} n^{-1} &\sum_{0\le k<l<n} v^i \circ f^k v^j \circ f^l - \sum_{k=1}^{n} \int_\Lambda v^i v^j \circ f^k \, d\mu \\ &- n^{-1} \sum_{0\le k<l<n} z^i_\epsilon \circ f^k z^j_\epsilon \circ f^l + \sum_{k=1}^{n} \int_\Lambda z^i_\epsilon z^j_\epsilon \circ f^k \, d\mu = (I) - (II), \end{aligned} \tag{22}$$

where

$$\begin{aligned} (I) =& n^{-1}\epsilon \sum_{0\le k<l<n} g^i_\epsilon \circ f^k z^j_\epsilon \circ f^l + n^{-1}\epsilon \sum_{0\le k<l<n} z^i_\epsilon \circ f^k g^j_\epsilon \circ f^l + n^{-1}\epsilon^2 \sum_{0\le k<l<n} g^i_\epsilon \circ f^k g^j_\epsilon \circ f^l \\ =& (Ia) + (Ib) + (Ic). \end{aligned} \tag{23}$$

That is $(Ia), (Ib)$ and (Ic) express the following:

$(Ia) = n^{-1}\epsilon \sum_{0\le k<l<n} g^i_\epsilon \circ f^k z^j_\epsilon \circ f^l$.

$(Ib) = n^{-1}\epsilon \sum_{0\le k<l<n} z^i_\epsilon \circ f^k g^j_\epsilon \circ f^l$.

$(Ic) = n^{-1}\epsilon^2 \sum_{0\le k<l<n} g^i_\epsilon \circ f^k g^j_\epsilon \circ f^l$.

For (Ic) we see that, by setting $\epsilon = \epsilon_n$,

$$\|n^{-1}\epsilon_n^2 \sum_{0\le k<l<n} g^i_{\epsilon_n} \circ f^k g^j_{\epsilon_n} \circ f^l\|_1 \ll \epsilon_n \|g^i_{\epsilon_n}\|_2 \|g^j_{\epsilon_n}\|_2 \to 0$$

as $n \to \infty$. Then the term (Ib) splits into two terms using the definition of z_ϵ:

$$n^{-1}\epsilon \sum_{0\le k<l<n} z^i_\epsilon \circ f^k g^j_\epsilon \circ f^l = n^{-1}\epsilon \sum_{0\le k<l<n} v^i_\epsilon \circ f^k g^j_\epsilon \circ f^l + n^{-1}\epsilon \sum_{0\le k<l<n} (U\chi^i_\epsilon - \chi^i_\epsilon) \circ f^k g^j_\epsilon \circ f^l.$$

Since the second term has a telescoping coboundary, it follows that:

$$\|n^{-1}\epsilon_n \sum_{0\le k<l<n} (U\chi^i_{\epsilon_n} - \chi^i_{\epsilon_n}) \circ f^k g^j_{\epsilon_n} \circ f^l\|_1 \ll \epsilon_n \|\chi^i_{\epsilon_n}\|_2 \|g^j_{\epsilon_n}\|_2 \to 0$$

as $n \to \infty$. Then we see that:

$$n^{-1}\epsilon \sum_{0 \le k < l < n} v_\epsilon^i \circ f^k g_\epsilon^j \circ f^l = n^{-1}\epsilon \sum_{l=1}^{n-1} \left[\sum_{k=0}^{l-1} v_\epsilon^i \circ f^k \right] g_\epsilon^j \circ f^l.$$

Thus, by Burkholder's inequality, we have that:

$$\begin{aligned} \|n^{-1}\epsilon \sum_{0 \le k < l < n} v_\epsilon^i \circ f^k g_\epsilon^j \circ f^l \|_1 \le n^{-1}\epsilon \sum_{l=1}^{n-1} \left\| \sum_{k=0}^{l-1} v_\epsilon^i \circ f^k \right\|_2 \left\| g_\epsilon^j \circ f^l \right\|_2 \le & n^{-1}\epsilon \sum_{l=1}^{n-1} l^{1/2} \|g_\epsilon^j\|_2 \\ \ll & n^{1/2}\epsilon \|g_\epsilon^j\|_2. \end{aligned}$$

Hence,

$$\|n^{-1}\epsilon_n \sum_{0 \le k < l < n} v_{\epsilon_n}^i \circ f^k g_{\epsilon_n}^j \circ f^l \|_1 \ll \epsilon_n^{1/2} \|g_{\epsilon_n}^j\|_2 \to 0$$

as $n \to \infty$. The term (Ia) can be dealt with in a similar way. That is, it splits into two terms: one which contains a telescoping coboundary; the other which Burkholder's inequality can be applied to after switching the order of summation. We see that:

$$n^{-1}\epsilon \sum_{0 \le k < l < n} g_\epsilon^i \circ f^k z_\epsilon^j \circ f^l = n^{-1}\epsilon \sum_{0 \le k < l < n} g_\epsilon^i \circ f^k v_\epsilon^j \circ f^l + n^{-1}\epsilon \sum_{0 \le k < l < n} g_\epsilon^i \circ f^k (U\chi_\epsilon^j - \chi_\epsilon^j) \circ f^l.$$

Then we deduce that:

$$\|n^{-1}\epsilon \sum_{0 \le k < l < n} g_{\epsilon_n}^i \circ f^k (U\chi_{\epsilon_n}^j - \chi_{\epsilon_n}^j) \circ f^l \|_1 \ll \epsilon_n \|g_{\epsilon_n}^i\|_2 \|\chi_{\epsilon_n}^j\|_2 \to 0$$

as $n \to \infty$. Moreover, as before, using Burkholder's inequality:

$$\|n^{-1}\epsilon \sum_{0 \le k < l < n} g_\epsilon^i \circ f^k v_\epsilon^j \circ f^l \|_1 \le n^{-1}\epsilon \sum_{k=0}^{n-1} \left\| \sum_{l=k+1}^{n-1} v_\epsilon^j \circ f^l \right\|_2 \|g_\epsilon^i \circ f^k\|_2 \ll n^{1/2}\epsilon \|g_\epsilon^i\|_2.$$

Hence we confirm that:

$$\|n^{-1}\epsilon_n \sum_{0 \le k < l < n} g_{\epsilon_n}^i \circ f^k v_{\epsilon_n}^j \circ f^l \|_1 \ll \epsilon_n^{1/2} \|g_{\epsilon_n}^i\|_2 \to 0$$

as $n \to \infty$. Hence we have shown the terms in equation (23) go to 0 as $n \to \infty$ in L^1. It remains to show the same holds for (II). That is we need to show:

$$\sum_{k=1}^{n} \int_\Lambda v^i v^j \circ f^k \, d\mu - \sum_{k=1}^{n} \int_\Lambda z_{\epsilon_n}^i z_{\epsilon_n}^j \circ f^k \, d\mu$$

converges to 0 in L^1 as $n \to \infty$. Again, fixing $\epsilon > 0$ momentarily we compute that,

$$\begin{aligned} (II) =& \sum_{k=1}^{n} \int_\Lambda v^i v^j \circ f^k \, d\mu - \sum_{k=1}^{n} \int_\Lambda z_{\epsilon_n}^i z_{\epsilon_n}^j \circ f^k \, d\mu \\ =& \epsilon \sum_{k=1}^{n} \int_\Lambda g_\epsilon^i z_\epsilon^j \circ f^k \, d\mu + \epsilon \sum_{k=1}^{n} \int_\Lambda z_\epsilon^i g_\epsilon^j \circ f^k \, d\mu + \epsilon^2 \sum_{k=1}^{n} \int_\Lambda g_\epsilon^i g_\epsilon^j \circ f^k \, d\mu. \end{aligned}$$

That is $(II) = (IIa) + (IIb) + (IIc)$, where:

$$(IIa) = \epsilon \sum_{k=1}^{n} \int_{\Lambda} g_\epsilon^i z_\epsilon^j \circ f^k \, d\mu.$$

$$(IIb) = \epsilon \sum_{k=1}^{n} \int_{\Lambda} z_\epsilon^i g_\epsilon^j \circ f^k \, d\mu.$$

$$(IIc) = \epsilon^2 \sum_{k=1}^{n} \int_{\Lambda} g_\epsilon^i g_\epsilon^j \circ f^k \, d\mu.$$

Considering term (IIc), it is easy to see that

$$\|\epsilon_n^2 \sum_{k=1}^{n} \int_{\Lambda} g_{\epsilon_n}^i g_{\epsilon_n}^j \circ f^k \|_1 \ll \epsilon_n \|g_{\epsilon_n}^j\|_2 \|g_{\epsilon_n}^j\|_2 \to 0$$

as $n \to \infty$. For term (IIb) we obtain:

$$\epsilon \sum_{k=1}^{n} \int_{\Lambda} z_\epsilon^i g_\epsilon^j \circ f^k \, d\mu = \epsilon \sum_{k=1}^{n} \int_{\Lambda} v_\epsilon^i g_\epsilon^j \circ f^k \, d\mu + \epsilon \sum_{k=1}^{n} \int_{\Lambda} (U\chi_\epsilon^i - \chi_\epsilon^i) g_\epsilon^j \circ f^k \, d\mu.$$

The first term of which is identically 0 by duality. For the second, we apply invariance of the measure μ and swap summation and integration:

$$\begin{aligned}\epsilon \sum_{k=1}^{n} \int_{\Lambda} (U\chi_\epsilon^i - \chi_\epsilon^i) g_\epsilon^j \circ f^k \, d\mu =& \epsilon \int_{\Lambda} \sum_{k=1}^{n} (U\chi_\epsilon^i - \chi_\epsilon^i) \circ f^{n-k} g_\epsilon^j \circ f^n \, d\mu \\ =& \epsilon \int_{\Lambda} (U^n \chi_\epsilon^i - \chi_\epsilon^i) g_\epsilon^j \circ f^n \, d\mu.\end{aligned}$$

Hence:

$$\|\epsilon_n \sum_{k=1}^{n} \int_{\Lambda} z_{\epsilon_n}^i g_{\epsilon_n}^j \circ f^k \, d\mu\|_1 \ll \epsilon_n \|\chi_{\epsilon_n}^i\|_2 \|g_{\epsilon_n}^j\|_2 \to 0$$

as $n \to \infty$. For the term (IIa) we have:

$$\epsilon \sum_{k=1}^{n} \int_{\Lambda} v_\epsilon^i g_\epsilon^j \circ f^k \, d\mu = \epsilon \sum_{k=1}^{n} \int_{\Lambda} g_\epsilon^i v_\epsilon^j \circ f^k \, d\mu + \epsilon \sum_{k=1}^{n} \int_{\Lambda} g_\epsilon^i (U\chi_\epsilon^j - \chi_\epsilon^j) \circ f^k \, d\mu.$$

The second part of this equation can be dealt with, as above, by telescoping the coboundary. For the first term we swap summation and integration

$$\epsilon \sum_{k=1}^{n} \int_{\Lambda} g_\epsilon^i v_\epsilon^j \circ f^k \, d\mu = \epsilon \int_{\Lambda} g_\epsilon^i \sum_{k=1}^{n} v_\epsilon^j \circ f^k \, d\mu.$$

Thus applying Burkholder's inequality we have:

$$\|\epsilon_n \sum_{k=1}^{n} \int_{\Lambda} g_{\epsilon_n}^i v_{\epsilon_n}^j \circ f^k \, d\mu\|_1 \ll \epsilon_n^{1/2} \|g_{\epsilon_n}^i\|_2 \|v_{\epsilon_n}^j\|_2 \to 0$$

as $n \to \infty$. Hence

$$\|\epsilon_n \sum_{k=1}^{n} \int_{\Lambda} v_{\epsilon_n}^i g_{\epsilon_n}^j \circ f^k \, d\mu\|_1 \to 0$$

as $n \to \infty$. Thus (II) converges to 0 in L^1 as $n \to \infty$. This concludes the proof. □

The following lemma establishes the iterated CLT for an ϵ-family of L^2 coboundaries.

Lemma 8.3. *Let $f:\Lambda\to\Lambda$ be a mixing transformation on the probability space $(\Lambda,\mathcal{F},\mu)$. Assume that v is a mean zero observable and $v\in L^2(\Lambda,\mathbb{R}^e)$. Suppose the following conditions hold:*

(a) $\sum_{n=1}^{\infty} n^{-3/2}\|\sum_{k=0}^{n-1} P^k v\|_2 < \infty$*(Maxwell-Woodroofe Condition).*

(b) $n^{-1}\sum_{k=0}^{n-1}(\chi^i_{\epsilon_n} v^j)\circ f^l - E[\chi^i_{\epsilon_n} v^j] \to_p 0$ *as* $n\to\infty$.

(c) $n^{-1}\sum_{k=1}^{n-1}(v^i_{\epsilon_n} U\chi^j_{\epsilon_n})\circ f^k \to_p 0$ *as* $n\to\infty$.

(d) $E[\chi^i_\epsilon v^j\circ f^n]\to_p 0$ *as* $n\to\infty$.

Then,

$$n^{-1}\sum_{0\le k<l<n} z^i_{\epsilon_n}\circ f^k z^j_{\epsilon_n}\circ f^l - \sum_{k=1}^{n}\int_\Lambda z^i_{\epsilon_n} z^j_{\epsilon_n}\circ f^k\, d\mu \to_d \int_0^1 W^i dW^j.$$

Proof. Let us first calculate the terms left in the drift. Formally expanding, we see that:

$$\sum_{k=1}^{n}\int_\Lambda z^i_\epsilon z^j_\epsilon\circ f^k\, d\mu = \sum_{k=1}^{n}\int_\Lambda (v^i_\epsilon + U\chi^i_\epsilon - \chi^i_\epsilon)(v^j_\epsilon + U\chi^j_\epsilon - \chi^j_\epsilon)\circ f^k\, d\mu.$$

This is equal to:

$$\sum_{k=1}^{n}\left(\int_\Lambda v^i_\epsilon(v^j_\epsilon + U\chi^j_\epsilon - \chi^j_\epsilon)\circ f^k + (U\chi^i_\epsilon - \chi^i_\epsilon)(v^j_\epsilon + U\chi^j_\epsilon - \chi^j_\epsilon)\circ f^k\, d\mu\right) \tag{24}$$

in which the first term is 0 by duality. The final term of (24) is equal to following. We use invariance of the measure; we switch order of integration; we telescope sums, and we substitute (15) to obtain:

$$\int_\Lambda (U^n\chi^i_\epsilon - \chi^i_\epsilon)(v^j - \epsilon g^j_\epsilon)\circ f^n\, d\mu = \int_\Lambda (U^n\chi^i_\epsilon - \chi^i_\epsilon)v^j\circ f^n\, d\mu - \epsilon\int_\Lambda (U^n\chi^i_\epsilon - \chi^i_\epsilon)g^j_\epsilon\, d\mu.$$

The final term converges to 0 as $\epsilon\to 0$ in the 1-norm by applying Hölder's inequality and previous arguments. The first term splits into the following two terms:

$$\int_\Lambda (\chi^i_\epsilon v^j)\circ f^n\, d\mu - \int_\Lambda \chi^i_\epsilon v^j\circ f^n\, d\mu,$$

which we will return to later. Recalling (18) we have

$$n^{-1}\sum_{0\le k<l<n} v^i_{\epsilon_n}\circ f^k v^j_{\epsilon_n}\circ f^l \to_d \int_0^1 W^i dW^j.$$

It remains to show the drift terms are the only term that remain in the difference. We must establish the convergence of the following. Again fix $\epsilon>0$ momentarily and consider

$$n^{-1}\sum_{0\le k<l<n} z^i_\epsilon\circ f^k z^j_\epsilon\circ f^l - n^{-1}\sum_{0\le k<l<n} v^i_\epsilon\circ f^k v^j_\epsilon\circ f^l.$$

The above equals:

$$n^{-1}\sum_{0\le k<l<n} v_\epsilon^i\circ f^k(U\chi_\epsilon^j-\chi_\epsilon^j)\circ f^l+n^{-1}\sum_{0\le k<l<n}(U\chi_\epsilon^i-\chi_\epsilon^i)\circ f^k v_\epsilon^j\circ f^l \tag{25}$$
$$+n^{-1}\sum_{0\le k<l<n}(U\chi_\epsilon^i-\chi_\epsilon^i)\circ f^k(U\chi_\epsilon^j-\chi_\epsilon^j)\circ f^l.$$

Calculating the final term containing the double coboundary, we obtain:

$$n^{-1}\sum_{0\le k<l<n}(U\chi_\epsilon^i-\chi_\epsilon^i)\circ f^k(U\chi_\epsilon^j-\chi_\epsilon^j)\circ f^l=n^{-1}\left[\sum_{l=1}^{n-1}U^l\chi_\epsilon^i(U\chi_\epsilon^j-\chi_\epsilon^j)\circ f^l\right]+n^{-1}\chi_\epsilon^i(\chi_\epsilon^j-U^n\chi_\epsilon^j),$$

where the final term was achieved by telescoping sums twice. The final term above clearly goes to 0 in L^1 as $n\to\infty$, with our choice of ϵ_n, with the following estimate:

$$\|n^{-1}\chi_\epsilon^i(\chi_\epsilon^j-U^n\chi_\epsilon^j)\|_1\ll\epsilon_n\|\chi_\epsilon^i\|_2\|\chi_\epsilon^j\|_2.$$

Now considering the first sum:

$$n^{-1}\sum_{l=1}^{n-1}U^l\chi_\epsilon^i(U\chi_\epsilon^j-\chi_\epsilon^j)\circ f^l=n^{-1}\sum_{l=1}^{n-1}U^l\chi_\epsilon^i(v^j-v_\epsilon^j)\circ f^l-n^{-1}\epsilon\sum_{l=1}^{n-1}U^l\chi_\epsilon^i g_\epsilon^j\circ f^l \tag{26}$$

the final term of which converges to 0 as $n\to\infty$ in L^1 by previous arguments. It remains to analyze the following two terms of (25). These are:

$$n^{-1}\sum_{0\le k<l<n}v_\epsilon^i\circ f^k(U\chi_\epsilon^j-\chi_\epsilon^j)\circ f^l+n^{-1}\sum_{0\le k<l<n}(U\chi_\epsilon^i-\chi_\epsilon^i)\circ f^k v_\epsilon^j\circ f^l. \tag{27}$$

Here, the second term telescopes to:

$$n^{-1}\sum_{l=1}^{n-1}(U^l\chi_\epsilon^i-\chi_\epsilon^i)v_\epsilon^j\circ f^l=n^{-1}\sum_{l=1}^{n-1}(\chi_\epsilon^i v_\epsilon^j)\circ f^l-n^{-1}\chi_\epsilon^i\sum_{l=1}^{n-1}v_\epsilon^j\circ f^l.$$

Then estimating the second term in the 1-norm we have:

$$\|n^{-1}\chi_{\epsilon_n}^i\sum_{l=1}^{n-1}v_{\epsilon_n}^j\circ f^l\|_1\le n^{-1}\|\chi_{\epsilon_n}^i\|_2\left\|\sum_{l=1}^{n-1}v_{\epsilon_n}^j\circ f^l\right\|_2\ll n^{-1/2}\|\chi_{\epsilon_n}^i\|_2\ll\epsilon_n^{1/2}\|\chi_{\epsilon_n}^i\|_2.$$

The right hand side converges to 0 as $n\to\infty$. The first term of (27) can be dealt with in a similar way:

$$\begin{aligned}n^{-1}\sum_{0\le k<l<n}v_\epsilon^i\circ f^k(U\chi_\epsilon^j-\chi_\epsilon^j)\circ f^l=&n^{-1}\sum_{k=0}^{n-1}\sum_{l=k+1}^{n-1}v_\epsilon^i\circ f^k(U^{l+1}\chi_\epsilon^j-U^l\chi_\epsilon^j)\\=&n^{-1}\sum_{k=0}^{n-1}v_\epsilon^i\circ f^k(U^n\chi_\epsilon^j-U^{k+1}\chi_\epsilon^j).\end{aligned} \tag{28}$$

The first term of which goes to 0 as $n \to \infty$ in L^1 by earlier arguments. Hence we are left with the following terms:

$$n^{-1}\sum_{l=1}^{n-1}(\chi_\epsilon^i v_\epsilon^j)\circ f^l - n^{-1}\sum_{k=0}^{n-1}(v_\epsilon^i U\chi_\epsilon^j)\circ f^k.$$

Further, recalling the terms left from (26) are:

$$n^{-1}\sum_{l=1}^{n-1}(\chi_\epsilon^i v^j)\circ f^l - (\chi_\epsilon^i v_\epsilon^j)\circ f^l.$$

Thus after cancellation we are left with:

$$n^{-1}\sum_{l=1}^{n-1}(\chi_\epsilon^i v^j)\circ f^l - n^{-1}\sum_{k=1}^{n-1}(v_\epsilon^i U\chi_\epsilon^j)\circ f^k.$$

It remains to establish the convergence of:

$$n^{-1}\sum_{l=0}^{n-1}(\chi_\epsilon^i v^j)\circ f^l - n^{-1}\sum_{k=1}^{n-1}(v_\epsilon^i U\chi_\epsilon^j)\circ f^k - \int_\Lambda(\chi_\epsilon^i v^j)\circ f^n\,d\mu + \int_\Lambda \chi_\epsilon^i v^j\circ f^n\,d\mu. \quad (29)$$

These converge to 0 in probability by assumptions $(b), (c), (d)$. Hence, the proof is complete. □

Theorem 8.4. *Let $f:\Lambda\to\Lambda$ be a mixing transformation on the probability space $(\Lambda,\mathcal{F},\mu)$. Assume that v is a mean zero observable and $v\in L^2(\Lambda,\mathbb{R}^e)$. Suppose the following conditions hold:*

(a) $\sum_{n=1}^{\infty} n^{-3/2}\|\sum_{k=0}^{n-1}P^k v\|_2 < \infty$*(Maxwell-Woodroofe Condition).*

(b) $n^{-1}\sum_{k=0}^{n-1}(\chi_{\epsilon_n}^i v^j)\circ f^l - E[\chi_{\epsilon_n}^i v^j] \to_p 0$ *as* $n\to\infty$.

(c) $n^{-1}\sum_{k=1}^{n-1}(v_{\epsilon_n}^i U\chi_{\epsilon_n}^j)\circ f^k \to_p 0$ *as* $n\to\infty$.

(d) $E[\chi_\epsilon^i v^j\circ f^n]\to_p 0$ *as* $n\to\infty$.

Then, $\lim_{n\to\infty}\sum_{m=1}^n\int_\Lambda v^i v^j\circ f^m\,d\mu$ *exists and* $\mathbb{W}_n\to_d\mathbb{W}$ *where,*

$$\mathbb{W}^{ij} = \int_0^1 W^i dW^j + \sum_{m=1}^{\infty}\int_\Lambda v^i v^j\circ f^m\,d\mu.$$

Proof. Under our assumptions Lemma 8.2 holds and establishes the convergence:

$$\mathbb{W}_n - \sum_{m=1}^{n}\int_\Lambda v^i v^j\circ f^m\,d\mu \to_d \int_0^1 W^i dW^j$$

if and only if

$$n^{-1}\sum_{0\le k<l<n} z_{\epsilon_n}^i\circ f^k z_{\epsilon_n}^j\circ f^l - \sum_{k=1}^{n}\int_\Lambda z_{\epsilon_n}^i z_{\epsilon_n}^j\circ f^k\,d\mu \to_d \int_0^1 W^i dW^j \text{ as } n\to\infty. \quad (30)$$

Furthermore the convergence of (30) holds because the assumptions of Lemma 8.3 are satisfied. This concludes the proof. □

This final proposition shows that under some additional dynamical assumptions we can remove some of the probabilistic assumptions of Lemma 8.3 and Theorem 8.4.

Proposition 8.5. *Let $f : \Lambda \to \Lambda$ be a mixing transformation on the probability space $(\Lambda, \mathcal{F}, \mu)$. Assume that v is a mean zero observable and $v \in L^\infty(\Lambda, \mathbb{R}^e)$. Suppose the following condition holds:*

(a') $\sum_{k=1}^{\infty} \|P^k v\|_1 < \infty$*(Summable in 1-norm)*

Then the conditions (b) and (d) of Lemma 8.3 and Theorem 8.4 are satisfied.

Proof. It is enough to prove that the assumptions (a') and $v \in L^\infty$ imply assumptions (b) and (d). First, let us consider:

$$n^{-1} \sum_{l=0}^{n-1} (\chi_\epsilon^i v^j) \circ f^l - \int_\Lambda (\chi_\epsilon^i v^j) \circ f^n \, d\mu.$$

It is standard that under assumption (a') that $\chi_\epsilon \to \chi$ as $\epsilon \to 0$ in L^1, where $\chi = \sum_{k=1}^{\infty} P^n v$. Thus consider the following estimate of the above

$$\begin{aligned}
&\|n^{-1} \sum_{l=0}^{n-1} (\chi_\epsilon^i v^j) \circ f^l - \int_\Lambda (\chi_\epsilon^i v^j) \circ f^n \, d\mu\|_1 \\
\leq &\|v^j\|_\infty \left\| n^{-1} \sum_{l=0}^{n-1} \chi_\epsilon^i \circ f^l - \int_\Lambda \chi_\epsilon^i \, d\mu \right\|_1 \\
= &\|v^j\|_\infty \left\| n^{-1} \sum_{l=0}^{n-1} (\chi_\epsilon^i - \chi^i) \circ f^l + n^{-1} \sum_{l=0}^{n-1} \chi^i \circ f^l - \int_\Lambda \chi^i \, d\mu + \int_\Lambda \chi^i - \chi_\epsilon^i \, d\mu \right\|_1 \\
\leq &\|v^j\|_\infty \left(n^{-1} \sum_{l=0}^{n-1} \|\chi_\epsilon^i - \chi^i\|_1 + \left\| n^{-1} \sum_{l=0}^{n-1} \chi^i \circ f^l - \int_\Lambda \chi^i \, d\mu \right\|_1 + \left\| \int_\Lambda \chi^i - \chi_\epsilon^i \, d\mu \right\|_1 \right) \\
= &\|v^j\|_\infty \left(2\|\chi_\epsilon^i - \chi^i\|_1 + \left\| n^{-1} \sum_{l=0}^{n-1} \chi^i \circ f^l - \int_\Lambda \chi^i \, d\mu \right\|_1 \right).
\end{aligned}$$

The first of these term in the final estimate converges to 0 as $\epsilon \to 0$, since $\chi_\epsilon \to \chi$, with no dependence on n. Now, since $\chi \in L^1$, we can apply the Ergodic Theorem and the second term converges to 0 as $n \to \infty$. Now considering the final term of (29), and letting $\epsilon = \epsilon_n$, we see that:

$$\left\| \int_\Lambda \chi_{\epsilon_n}^i v^j \circ f^n \, d\mu - \int_\Lambda \chi^i v^j \circ f^n \right\|_1 \leq \|v^j\|_\infty \|\chi_{\epsilon_n}^i - \chi^i\|_1 \to 0$$

as $n \to \infty$. Moreover, because of mixing:

$$\int_\Lambda \chi^i v^j \circ f^n \, d\mu \to \int \chi^i \, d\mu \int v^j \, d\mu = 0$$

as $n \to \infty$. Hence, allowing $n \to \infty$, we have established that

$$\int_\Lambda \chi^i_{\epsilon_n} v^j \circ f^n \to 0,$$

which concludes the proof. □

Remark 8.6. The assumptions of our theorem are easy to check except for (c). A Weak Law of Large Numbers for L^1 martingales arrays under our assumptions would remove this assumption; however, we cannot find a suitable result in the literature. Further, we note that if we assume the set up of earlier sections in which we assumed $\sum_{k=1}^\infty \|P^k v\|_2 < \infty$, then the hypothesis (c) would be removed by simple arguments. We make the final conjecture that the iterated CLT holds under the assumptions of Proposition 8.5 along with the Maxwell-Woodroofe condition. That is, the assumption (c) is unnecessary.

9 References

[1] P. Billingsley. Convergence in Probability Measures, second ed., *Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley and Sons Inc,.* New York, 1999, A Wiley-Interscience Publication.

[2] B.M. Brown. Martingale Central Limit Theorems, *Annals of Math. Statist.* **42** (1971), 59-66.

[3] D.L. Burkholder. Distribution Function Inequalities for martingales, *Annals of Probability.* **1** (1973), 19-42.

[4] C. Cuny. Limit Theorems under the Maxwell-Woodroofe Condition in Banach Spaces, *Annals of Probability.* To appear.

[5] C. Cuny and F. Merlevéde. On Martingale Approximations and the Quenched Weak Invariance Principle, *Annals of Probability.* **42** (2014), 760-793.

[6] A. Jakubowski, J. Mémin, G. Pages. Convergence en loi des suites d'intégrales stochastiques sur l'espace D^1 de Shorokhod, *Probab. Th. Rel. Fields.* **81** (1989), 111-137.

[7] D. Kelly and I. Melbourne. Smooth Approximations of Stochastic Differential Equations, *Annals of Probability.* **44** (2016), 479-520.

[8] T.G. Kurtz and P. Protter. Weak Limit Theorems of Stochastic Integrals and Stochastic Differential Equations, *Annals of Probability* **19** (1991), 1035-1070.

[9] A. Lejay and T. Lyons. On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization, *Current Trends in potential theory*, Theta Ser. Adv. Math. **4** (2005), Theta, Bucharest, 63-84.

[10] M. Maxwell and M. Woodroofe. Central limit Theorems for Additive Functionals of Markov Chains, *Annals of Probability.* **28** (2000), 713-724.

[11] D.L. McLeish. Dependent Central Limit Theorems and Invariance Principles, *Annals of Probability.* **4** (1974), 620-628.

[12] E.J. McShane. Stochastic Differential Equations and model of random processes, *Proceedings of the sixth Berkeley Symposium on Mathematical Statistics and Probability: Vol. III: Probability Theory*, Univ. California Press, (1972) 263-294.

[13] I. Melbourne and A. Torok. Central limit Theorems and Invariance Principles for Time-One Maps of Hyperbolic Flows, *Commun. Math. Phys.* **229** (2002), 57-71.

[14] F. Merlevéde and M. Peligrad. Rosenthal-Type Inequalities for the Maximum of Partial Sums of Stationary Processes and Examples, *Annals of Probability.* **41** (2013), 914-960.

[15] A.V. Skorohod. Limit Theorems for Stochastic Processes, *Teor.Veroyatnost.i Primenen*, **1** (1956), 289-319.

[16] H.J. Sussmann. On The Gap Between Deterministic and Stochastic Differential Equations, *Annals of Probability.* **6** (1978), 19-41.

[17] M. Tyran-Kaminǹska. An Invariance Principle for Maps with Polynomial Decay of Correlations, *Commun. Math. Phys.* **260** (2005), 1-15.

[18] M. Peligrad and S.Utev. A New Maximal Inequality and Invariance Principle for Stationary Sequences, *Annals of Probability.* **33** (2005), 798-815.

[19] E. Wong and M. Zakai. On the convergence of ordinary integrals to stochastic integrals, *Annals of Math. Statist.* **36** (1965), 1560-1564.