2003

Ergodic transformations that generate the Carathéodory definition of measurable sets

Amos Nathan Koeller
University of Wollongong

Follow this and additional works at: https://ro.uow.edu.au/theses

Recommended Citation

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
Ergodic transformations that generate the Carathéodory definition of measurable sets.

A thesis submitted in partial fulfilment of the requirements for the award of the degree

MASTERS BY RESEARCH

from

UNIVERSITY OF WOLLONGONG

by

AMOS NATHAN KOELLER,

BACHELOR OF MATHEMATICS AND ECONOMICS (HONOURS)

SCHOOL OF MATHEMATICS AND APPLIED STATISTICS

2003
Contents

1 Introduction 8

2 A naturally arising condition leading to Carathéodory. 15
  2.1 Notes ....................................................................................................... 20

3 Ergodicity, Mixing and the Spectral Theorem 21
  3.1 Birkoff's Individual Ergodic Theorem ........................................... 26
  3.2 Mixing ..................................................................................................... 42
  3.3 Mixing and Continuous Spectrum .................................................. 63
  3.4 Notes ..................................................................................................... 74

4 Rotations and Orbits 75
  4.1 Rotations ............................................................................................. 77
  4.2 Orbits ................................................................................................... 93
  4.3 Notes ......................................................................................................... 129

5 Generalisations, Towers and the rise of Splinters 130
  5.1 Generalisations ..................................................................................... 131
  5.2 Towers .................................................................................................... 159
  5.3 Notes ......................................................................................................... 199

6 An Ergodic Theorem for Measurable Sets: The Carathéodory
definition generated 200
  6.1 $T$ generates if $T$ is outer measure preserving and ergodic . . . . 213
  6.2 Outer measure preserving transformations ....................................... 229
  6.3 Towers and ergodic outer measure preserving transformations 245
  6.4 Notes .......................................................................................................... 252

7 Non-measurable Sets 253
7.1 General results on non-measurable sets ........................................254
7.2 The outer measure of sets invariant under transformations that generate the Carathéodory definition of measurable sets . . . . 271
7.3 Examples of non-measurable sets ....................................................277
7.4 Transformation invariant non-measurable sets and tower extensions .................................................................303
7.5 Notes ............................................................................ 306

8 Bibliography 307
Abstract

Motivated by a pedagogical argument about the existence of intuitive interpretations of the important definition of measurable sets given by Carathéodory [7] in 1914 this thesis defines the concept of a transformation which generates the definition of Carathéodory, and describes transformations which posses this property. In doing so, particular attention is paid to the relatively uninvestigated behaviour of the outer measures of sets invariant under ergodic transformations.

The investigation of sets invariant under ergodic transformations also considers the ergodic tool of tower extensions of transformations and their effect on transformation invariant sets, the transformation properties of invariance, mixing (in the ergodic sense) and the generation of Carathéodory's definition. Specific examples of transformations coming within the orbit of the discussion are irrational rotations on the unit circle, Kakutani's transformation, Chacon's tower transformation and generalisations of some of these.

It is shown that a transformation will generate Carathéodory's definition only if it is ergodic and outer measure preserving. The condition of outer measure invariance, rarely mentioned in current mathematical literature, is also explored. It is shown that a transformation is outer measure invariant if the images of Borel sets under the transformation are measurable, in the sense of Carathéodory. Some discussion of the conditions required for this to occur is also given.

In order for transformations to be considered to have generated Carathéodory's definition, that is to say that the notion has actual content, it is necessary to show that there are transformation invariant non-measurable
sets in the Carathéodory sense. Some results are given, providing illus-
tration as to the properties required for a transformation to allow
invariant non-measurable sets. Finally, in discussing these properties
some interesting group theoretic results arise including, for example,
the unexpected result that, for Kakutani’s transformation $T$, the $T$-
orbit of any point in $[0,1)$ is a coset of the dyadic rationals.
ACKNOWLEDGEMENTS

In the course of my research, I have received support that I would be remiss not to acknowledge. First and foremost, my supervisors, Associate Professors Rodney Nillsen and Graham Williams, who have provided inspirational encouragement, support and assistance throughout my studies. Secondly, the School of Mathematics and Applied statistics, especially Mrs. Carolyn Silveri and Mrs. Kerri Gamble who have ensured that all of my administrative needs were met.

I would also like to thank the University of Wollongong for the financial support given in the form of a University Postgraduate Award, without which, the attention that I have given my work would not have been possible.
1 Introduction

In mathematics, and particularly in the theory of integration, the idea of measurement plays a vital role. It follows that, given a method of measurement (i.e. a measure), the question of exactly which objects may be reasonably measured by the measure is an important one. That is, we would like to know, given a measure and an object, what condition, when satisfied, will ensure that the object is similar enough to the objects on which a measure is defined to allow a satisfactory result when measured by the given measure.

In this thesis we will be considering the above in the context of measure spaces so that for a measure space \((X, \mathcal{B}, \mu)\), the measure is \(\mu\), the objects on which the definition of the measure is based are the elements of \(\mathcal{B}\) and the set of objects, from which we wish to distinguish between those that can be reasonably measured by \(\mu\) and those that cannot, is the set of all subsets of \(X\). In this context, perhaps the most famous condition is that given by Carathéodory [7] in 1914 where he gives a definition of measurable sets based on the concept of outer measure. An outer measure is, in a sense, a type of measure that is based on an original measure defined on basic subsets of some set (or space), say \(X\), that gives a value to all subsets of \(X\). By convention, the properties of an outer measure, say \(\mu_*\) are

(i) \(\mu_*(\emptyset) = 0\),

(ii) \(A \subset B \Rightarrow \mu_*(A) \leq \mu_*(B)\),

where \(\emptyset\) denotes the null set, and also that \(\mu_*\) must be countably subadditive. That is, if \(A_1, A_2, \ldots\) is a countable sequence of sets then

\[
\mu_* \left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mu_*(A_i).
\]

The problem is that under the definition of the outer measure, the desirable property of the size (outer measure) of the union of a sequence of disjoint
sets is not necessarily equal to the sum of the sizes of the sets. Hence we wish to find a condition that enables us to find which sets will allow this property (known as countable additivity) to occur in all cases. It is the sets that allow for the countable additivity property that may be regarded as sets that may be reasonably, or satisfactorily, measured. Carathéodory’s condition has proven to be very successful and is commonly accepted as the definition of which sets may be measured. That is, it is commonly accepted as the definition of measurable sets.

Unfortunately, given the importance of the ability to define what a measurable set is, Carathéodory’s definition does not at all make the reason why the condition given is used at all obvious. In fact, as Edwin Hewitt and Karl Stromberg [18] write: "How Carathéodory came to think of this definition seems mysterious as it is not in the least intuitive." This is due, in part, as Halmos [15] states, to it being, under this definition of measurability "difficult to get an understanding of the meaning of ... measurability except through familiarity with its implications." This is, in turn, because, as Nillsen [32] writes in a deeper discussion on the motivation of Carathéodory’s definition, "The definition of a measurable set appears to be unusually remote from the context which gives it its mathematical importance."

As we have mentioned, Carathéodory’s definition is based on the notion of outer measure. While we have mentioned the properties of an outer measure we have not described how to define one. There are several ways of defining an outer measure, however, in this thesis, to remain consistent with the motivating discussion, we will only use outer measure to describe outer measures that arise from one particular method of definition. Given a measure that is originally defined on a collection of ‘basic’ subsets, say $\mathcal{B}$ of a particular
set, say $X$, sets, $\mu$, we can define an outer measure, $\mu_*$, associated with the measure $\mu$, by

$$\mu_*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(B_n) : A \subseteq \bigcup_{n=1}^{\infty} B_n, B_n \in \mathcal{B} \text{ for each } n \right\}.$$ 

In this way we are using the measure $\mu$ to attempt to measure any given subset of the space. We would like to know when the resulting value can give a satisfactory indication of size, especially in comparing the size of a set to other subsets of $X$. Carathéodory's condition states that a set, $A$, will receive a satisfactory measure from this process, and hence be a measurable set, if for all subsets $B$ of $X$

$$\mu_* (B) = \mu_* (A \cap B) + \mu_* (A^c \cap B) \tag{1}$$

where $A^c$ denotes the complement of $A$. It is not obvious why the ability of $A$ and its complement to separate all subsets of $X$ in this neat fashion should be a good definition of when $A$ will receive a reasonable measure from $\mu_*$. 

Note that the term 'basic sets' is used mostly to emphasise the role of the sets without giving them any particular properties at this stage. In most cases, including our Definition 2.2, these basic sets will in fact be a $\sigma$-algebra on which the measure is defined.

Certainly, one would expect the sets that are 'well behaved' enough with respect to the measure will satisfy this separation condition, though it is not at all immediately clear why the sets satisfying this condition will be the sets that we are looking for. While an intuitive interpretation can be given (Halmos [15]), the 'remoteness' of the definition remains, especially in the sense that there does not appear to be any previous mathematics which allows one to naturally arrive at this definition. Nillsen [32], addressed this problem in
showing that in considering sets invariant under an irrational rotation on the circle group, when the sets satisfy a further condition, then the outer measure of the set must be 0 or 1. The condition in question arises naturally from the consideration of the invariance under the rotation and is similar to the condition in Carathéodory’s definition. Also, importantly, it was shown that this line of investigation does allow for non-measurable sets (in the Carathéodory sense) to be considered and further that non-measurable sets are rejected by the condition given by Nillsen. Hence Nillsen’s condition also provided for a distinction between sets for which the measure (through the outer-measure) gave a reasonable value and those that did not. Nillsen’s condition can also be shown to lead directly to Carathéodory’s definition. In this way an irrational rotation on the circle group can be said to ‘generate’ Carathéodory’s definition.

In the same work Nillsen also noted that it was certainly not true that all transformations could ‘generate’ Carathéodory’s definition in the same manner and conjectured that the property of ‘generating’ Carathéodory’s definition in this manner depended on their classification with regards to the ergodic concept of mixing properties.

In this thesis, we aimed to establish a relationship between the property of ‘generating’ Carathéodory’s definition and some mixing property. In our investigation we found that, while it was certainly necessary for the transformation to be ergodic, the property of ‘generating’ Carathéodory’s definition does not depend on the transformations mixing properties. Instead, we found that it was sufficient that the transformation be outer measure preserving.

The later and more general results in this thesis become more abstracted
from considering specific transformations and, if considered alone, would lose, to some extent, any kind of natural progression of ideas to Carathéodory's definition. For this reason, we present the work chronologically in the sense of the time at which results were established. In this way the progression of ideas, definitions and results can be easily followed from the original consideration of Nillsen's investigation of irrational rotations, through to the most general results toward the end of the thesis.

The structure of the thesis is as follows.

Chapter 2 introduces Nillsen's condition and how it may be considered to lead to Carathéodory's definition, further we give an initial definition of a transformation that generates Carathéodory's definition of measurable sets.

In this thesis, familiarity with measure theory is assumed, as is a basic knowledge of functional analysis and group theory. However, no knowledge of ergodic theory is assumed. Hence, as the concept of ergodicity is vital for the thesis and the concepts of mixing necessary for addressing Nillsen's conjecture, Chapter 3 introduces ergodicity and the ergodic results that will be necessary in proving the other results discussed later in the thesis.

Chapter 4 commences the discussion on the mathematical tools necessary for consideration of Nillsen's condition and whether or not a transformation generates Carathéodory's definition. This is done firstly through discussing how we can prove that the irrational rotation 'generates' the Carathéodory definition. It is done secondly through the discussion of how an example of a transformation with mixing properties of particular interest to us, given by Kakutani [21], can be proven to 'generate' Carathéodory's definition. This
second result is our own, though it was originally published in [25].

Chapter 5 consists mostly of original work in which the tools discussed in Chapter 4 are extended and generalised. Further transformations, and some general types of transformations are considered and proven to 'generate' Carathéodory’s definition. The proofs of the results in this chapter are strongly suggestive as to which features are absolutely necessary to ensure that a transformations ‘generates’ Carathéodory’s definition and how to use these features. We also give some general conditions as to which extensions or products of transformations known to ‘generate’ Carathéodory’s definition will also generate Carathéodory’s definition.

Our most general results on the ‘generation’ of Carathéodory’s definition are presented in Chapter 6. The strongest being that any ergodic outer measure preserving transformation will generate Carathéodory’s definition. As the requirement that a transformation be outer measure preserving, as opposed to being measure preserving, is an unusual one, we give some consideration to which transformations are outer measure preserving and what properties they have. Importantly, we show that not all measure preserving transformations are outer measure preserving, thus preventing Nillsen’s condition from necessarily being equivalent to ergodicity. Apart from some reference and supporting results, the results in Chapter 6 are original.

As mentioned above, it was important in the original case of irrational rotations that the new condition be able to identify non-measurable sets in the Carathéodory sense so as to show that the new condition can meaningfully give some distinction between those sets that can be adequately measured by the given measure and those which can not. In Chapter 7 we investigate
and give some indication of which transformations will also be able to identify non-measurable sets in the Carathéodory sense. While the approach for constructing the basic non-measurable set used in Chapter 7 is well known, the specific adaptations are all our own though some appear in [27].

Each chapter will be concluded with a section noting other work in the area dealt with by the chapter and where the interested reader should look. This section, in each chapter, also gives the origin, original or otherwise, of each of the results presented in the chapter.
2 A naturally arising condition leading to Carathéodory.

This chapter establishes the condition that we, in this thesis, will be investigating in terms of finding identifying properties of the transformations that satisfy the condition. Before doing so, we formally state the sense in which we will be using outer measure and Carathéodory’s condition itself.

Definition 2.1

Let \((X, B, \mu)\) be a measure space. Then we define the outer measure associated with \(\mu\) (or simply the outer measure), \(\mu_*\) by

\[
\mu_*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(B_n) : A \subseteq \bigcup_{n=1}^{\infty} B_n, B_n \in B \text{ for each } n \right\}.
\]

As mentioned in the introduction, we will only be using outer measures in this sense as derived from a previously designated measure.

Definition 2.2 (Carathéodory’s definition of measurable sets)

Let \((X, B, \mu)\) be a measure space. Then a subset \(A\) of \(X\) is measurable if for all subsets \(B\) of \(X\)

\[
\mu_*(B) = \mu_*(A \cap B) + \mu_*(A^c \cap B).
\]

It is a common convention to say that the elements of \(B\) are the measurable sets. Carathéodory’s definition will certainly include all of the elements of \(B\), but it may (depending on the measure) also include other sets. Throughout this thesis, a measurable set will be understood to be a set that satisfies the condition in Carathéodory’s definition.

In the introduction we discussed that measurable sets were those for which a value of reasonable meaning could be established using a measure that was not necessarily defined for that set. One example of a measure is Lebesgue
measure defined on the unit circle $T$. This measure is originally defined on the arcs which are subsets of the circle. The definition is then extended to the remaining measurable sets, which include the Borel sets and sets of outer measure zero. As each of these sets will be measurable in the Carathéodory sense, it is accepted that the values for these sets given by the outer measure will be satisfactory values and hence the value is called and denoted as the measure of these sets instead of just the outer measure. (This process gives the $\sigma$-algebra for the Lebesgue measure space on $T$.)

We then wish to find what other sets will receive satisfactory values from this measure. These sets will certainly be those that are 'well behaved' when compared to the sets (arcs) on which the measure is defined in the sense of a separation condition similar to that given by Carathéodory. If we take a set in which we are interested in and translate the set to all parts of the circle we can then see if the set will be 'well behaved'. We can move a set around in this manner by way of use of an ergodic transformation.

**Definition 2.3**

Let $X$ be a set and $T$ be a transformation defined on $X$. Then a subset $B$ of $X$ is said to be $T$-invariant if $T^{-1}(B) = B$.

**Definition 2.4**

Let $(X, B, \mu)$ be a measure space and $T$ be a transformation on the space. If for each $B \in B$, $\mu(T^{-1}(B)) = \mu(B)$ then $T$ is called measure preserving. $T$ is called ergodic if $T$ is measure preserving and for each $B \in B$ that is $T$-invariant $\mu(B) \in \{0, \mu(X)\}$.

We look more into what this definition means in Chapter 3, for now we
only need to note that the definition of ergodicity (as will be proven in Theorem 3.1) is equivalent to saying that for each pair of subsets, $A$ and $B$, of $X$ with positive measure there is a natural number $n$, $(n \in \mathbb{N})$ such that $\mu(T^{-n}(B) \cap A) > 0$. Hence any set of positive measure is moved to intersect every other set of positive measure. As each arc has positive measure, an ergodic transformation therefore moves any set of positive measure to intersect each set of positive measure on which the measure on $T$ is defined. It follows that if $T$ is ergodic and $A$ is a $T$-invariant subset of positive measure of $X$, then $A$ intersects every arc. In this case we can clearly compare $A$ to every arc as we would like to be able to do.

If we take $T$ to be an irrational rotation (which is known to be ergodic as proven in Theorem 4.2), Nillsen [32] proved the following Theorem.

**Theorem 2.1**

*Let $T$ be an irrational rotation on the Lebesgue measure space on the unit circle $(T, \mathcal{B}, \mu)$ and let $A$ be a $T$-invariant subset of $T$. Suppose that there is a $\theta = \theta(A) < 2$ such that for all arcs $J$ of $T$,

$$\mu_*(A \cap J) + \mu_*(A^c \cap J) < \theta \mu(J).$$

(2)

Then either $A$ or $A^c$ is a set of outer measure zero.*

Hence, we have some information about the value that $\mu$ will give to any $T$-invariant set that satisfies the weaker 'separation' property (2), (thus being in some sense 'well behaved'). Clearly, the "simplest" way to satisfy (2) is to satisfy the stronger separation condition

$$\mu_*(A \cap J) + \mu_*(A^c \cap J) = \mu(J).$$

(3)
However, in this case we have the Proposition 2.1 which is due to Craven [9].
We note here that throughout this thesis we will use \( \{J_n\}_{n=1}^{\infty} \) to denote an infinite sequence of sets \( J_1, J_2, \ldots \) and \( \{J_n\}_{n=1}^{k} \) to denote a finite sequence of sets \( J_1, J_2, \ldots, J_k \). In each case \( \{J_n\} \) will be used to denote the sequence if no ambiguity about the indexation will result.

**Proposition 2.1**

Let \( A \subseteq \mathbb{T} \). If for each arc \( J \) in \( \mathbb{T} \),

\[
\mu_*(A \cap J) + \mu_*(A^c \cap J) = \mu(J), \tag{4}
\]

then for each \( W \subseteq \mathbb{T} \)

\[
\mu_*(A \cap W) + \mu_*(A^c \cap W) = \mu_*(W).
\]

**Proof:**

By the definition of \( \mu_* \) and by the fact that \( \mathbb{T} \) is of finite measure, for any \( \varepsilon > 0 \) we can find a sequence \( \{J_n\} \) of arcs such that \( W \subseteq \bigcup_{n=1}^{\infty} J_n \) and that

\[
\sum_{n=1}^{\infty} \mu(J_n) < \mu_*(W) + \varepsilon.
\]

Then

\[
\mu_*(W) = \mu_*((W \cap A) \cup (W \cap A^c)) \\
\leq \mu_*(W \cap A) + \mu_*(W \cap A^c) \\
\leq \mu_* \left( \left( \bigcup_{n=1}^{\infty} J_n \right) \cap A \right) + \mu_* \left( \left( \bigcup_{n=1}^{\infty} J_n \right) \cap A^c \right) \\
\leq \sum_{n=1}^{\infty} \mu_*(J_n \cap A) + \sum_{n=1}^{\infty} \mu_*(J_n \cap A^c) \\
= \sum_{n=1}^{\infty} \mu_*(J_n \cap A) + \mu_*(J_n \cap A^c) \\
= \sum_{n=1}^{\infty} \mu(J), \quad \text{by (4),} \\
< \mu_*(W) + \varepsilon.
\]
Since this is true for any \( \varepsilon > 0 \) the result follows.

Thus, if the simplest way for (2) to be satisfied is satisfied, we actually have Carathéodory's condition.

In addition we note that if (2) is satisfied, we then have that either \( A \) or \( A^c \) has outer measure zero and hence is measurable in the Carathéodory sense. Thus the only way for (2) to be satisfied is the simplest way which gives Carathéodory's condition. Hence, we have that Carathéodory's condition follows naturally from Theorem 2.1. In this way, the irrational rotation used in Theorem 2.1 can be considered to have 'generated' Carathéodory's condition. From this line of argument we can form the following definition.

**Definition 2.5**

*Let \( T \) be a transformation on the Lebesgue measure space of an interval in \( \mathbb{R} \), \( (X, B, \mu) \). Suppose that for any \( T \)-invariant subset, \( A \), of \( X \), for which there is a \( \theta = \theta(A) < 2 \) such that for all subintervals \( J \) of \( X \)

\[
\mu_*(A \cap J) + \mu_*(A^c \cap J) \leq \theta \mu(J),
\]

either \( A \) or \( A^c \) is a set of outer measure zero. Then, \( T \) is said to generate Carathéodory's definition of measurable sets or simply to generate Carathéodory's definition.*

From observation of Definition 2.5 there does not seem to be a particular reason to restrict our spaces \( X \) to intervals. However, we wait until what the more general form should be is more obvious before making the more general definition. In particular, our later generalisation allows the unit circle so that Theorem 2.1 will again be considered to be demonstrating a transformation that generates Carathéodory's definition, whereas, by this definition, it is not.
It is the properties of transformations that generate Carathéodory's definition that this thesis, especially Chapters 4, 5 and 6, investigates.

2.1 Notes

This chapter, like the introduction discusses the motivation and direction of the thesis. Much of the discussion is based on Nillsen [32]. A proof for Theorem 2.1 can also be found in [32]. We used Craven [9] for proposition 2.1. As mentioned in the introduction, a knowledge of measure theory is assumed. A standard text for measure theory is Federer [11]. A less compacted discussion of the assumed knowledge can be found in Bartle [3].
3 Ergodicity, Mixing and the Spectral Theorem

As we are aiming to investigate the relationship between the generation of the Carathéodory definition of measurable sets, levels of mixing and ergodicity, we will need to establish the ergodic results through which we intend to establish the relationship. It has been mentioned that we will find that all outer measure preserving ergodic transformations generate the Carathéodory definition and that all transformations that generate the Carathéodory definition are ergodic. In demonstrating this result and in describing the path of ideas that leads to the result several equivalent forms of ergodicity and mixing conditions are used. It is therefore necessary to show that these various conditions are indeed equivalent to each other and to the definitions that will be given. It is assumed that the reader is familiar with all of the necessary measure theoretic concepts, though important measure theoretic results that we use are stated without proof for convenience. The results presented in this chapter are relatively standard results. The proofs presented follow various other works specified in the notes.

Due to its importance to this chapter, we restate the definition of ergodicity.

Definition 2.4

Let \((X, B, \mu)\) be a measure space and \(T\) be a transformation on the space. If for each \(B \in B\), \(\mu(T^{-1}(B)) = \mu(B)\) then \(T\) is called measure preserving. \(T\) is called ergodic if \(T\) is measure preserving and for each \(B \in B\) that is \(T\)-invariant \(\mu(B) \in \{0, \mu(X)\}\).

Often ergodicity is specifically defined on a measure space of measure 1 (That
is, for \((X, B, \mu)\), we have \(\mu(X) = 1\). Such a space is commonly referred to as a probability space. When we are particularly considering a measure space of measure 1 we will follow this convention and call the measure space a probability space.

Some of the conditions equivalent to ergodicity that we use can be proved relatively directly from the definition. These conditions are shown in Theorem 3.1, which follows the necessary technical result proven in Lemma 3.1.

To simplify notation we will use the symbol \(\Delta\) for the symmetric difference of two sets. That is, for two sets \(A\) and \(B\) we have

\[
A \Delta B = (A \cap B^c) \cup (A^c \cap B).
\]

**Lemma 3.1**

*Let \(X\) be a space with a transformation \(T\). If \(B \subseteq X\), then for all \(n \in \mathbb{N}\)

\[
T^{-n} B \Delta B \subseteq \bigcup_{i=0}^{n-1} (T^{-(i+1)} B \Delta T^{-i} B).
\]

**Proof:**

Proceeding by induction, for \(n = 1\) it is clear that

\[
T^{-1} B \Delta B = \bigcup_{i=0}^{1-1} (T^{-(i+1)} B \Delta T^{-i} B).
\]

Then, to prove the inductive step, suppose that the hypothesis is true for \(n = k\) so that

\[
T^{-k} B \Delta B \subseteq \bigcup_{i=0}^{k-1} (T^{-(i+1)} B \Delta T^{-i} B).
\]

Now let \(x \in T^{-(k+1)} B \Delta B\).

If \(x \in B - T^{-(k+1)} B\), either \(x \in B - T^{-k} B\) which gives \(x \in B \Delta T^{-k} B\)
or \( x \in T^{-k}B - T^{-(k+1)}B \) so that \( x \in T^{-k}B \triangle T^{-(k+1)}B \).

If \( x \in T^{-(k+1)}B - B \), either \( x \in T^{-(k+1)}B - T^{-k}B \) which gives \( x \in T^{-(k+1)}B \triangle T^{-k}B \), or \( x \in T^{-k}B - B \) so that \( x \in T^{-k}B \triangle B \).

In either case we now know that for all \( x \in T^{-(k+1)}B \triangle B \),

\[
x \in (B \triangle T^{-k}B) \cup (T^{-k}B \triangle T^{-(k+1)}B)
\]

\[
\subseteq \left( \bigcup_{i=0}^{k-1} (T^{-(i+1)}B \triangle T^{-i}B) \right) \cup (T^{-k}(B) \triangle T^{-(k+1)}B).
\]

Thus

\[
T^{-(k+1)}B \triangle B \subseteq \bigcup_{i=0}^{(k+1)-1} (T^{-(i+1)}B \triangle T^{-i}B).
\]

\[\diamondsuit\]

**Theorem 3.1**

*If \( T : X \rightarrow X \) is a measure-preserving transformation of the measure space \( (X, \mathcal{B}, \mu) \) of finite measure, then the following statements are equivalent.*

1. \( T \) is ergodic.

2. The only members \( B \) of \( \mathcal{B} \) with \( \mu(T^{-1}B \triangle B) = 0 \) are those with \( \mu(B) \in \{0, \mu(X)\} \).

3. For every \( A \in \mathcal{B} \) with \( \mu(A) > 0 \) we have \( \mu(\bigcup_{n=1}^{\infty} T^{-n}A) = \mu(X) \).

4. For every \( A, B \in \mathcal{B} \) with \( \mu(A) > 0, \mu(B) > 0 \) there exists an \( n \in \mathbb{N} \) with \( \mu(T^{-n}A \cap B) > 0 \).

**Proof:**

We prove the theorem by showing

\((1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1)\).
For (1) ⇒ (2),

Let $B \in B$ and $\mu(T^{-1}B \Delta B) = 0$. We show that there is a $T$-invariant set $A \in B$ with $\mu(A \Delta B) = 0$ which gives the result. By Lemma 3.1 we have for each $n \in \mathbb{N}$ that

$$
\begin{align*}
\mu(T^{-n}B \Delta B) &\leq \mu \left( \bigcup_{i=0}^{n-1} T^{-(i+1)}B \Delta T^{-i}B \right) \\
&\leq \sum_{i=0}^{n-1} \mu(T^{-(i+1)}B \Delta T^{-i}B) \\
&= \sum_{i=0}^{n-1} \mu(T^{-i}(T^{-1}B \Delta B)) \\
&= \sum_{i=0}^{n-1} \mu(T^{-1}B \Delta B),
\end{align*}
$$

and hence that $\mu(T^{-n}B \Delta B) \leq n \mu(T^{-1}B \Delta B) = 0$.

Let

$$A = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} T^{-i}B.$$

By the above we know that for each $n \in \mathbb{N}$

$$
\begin{align*}
\mu \left( B \Delta \bigcup_{i=n}^{\infty} T^{-i}B \right) &\leq \mu \left( \bigcup_{i=n}^{\infty} (B \Delta T^{-i}B) \right) \\
&\leq \sum_{i=n}^{\infty} \mu(B \Delta T^{-i}B) \\
&= 0.
\end{align*}
$$

Since $\{\bigcup_{i=n}^{\infty} T^{-i}B\}$ is a decreasing sequence with the property that

$$\mu(B \Delta \bigcup_{i=n}^{\infty} T^{-i}B) = 0 \quad \text{for each } n,$$

we have

$$
\mu(B \Delta A) = \mu \left( B \Delta \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} T^{-i}B \right).
$$
Also

\[ T^{-1}A = T^{-1} \left( \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} T^{-i}B \right) \]
\[ = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} T^{-i}B \]
\[ = \bigcap_{n=0}^{\infty} \bigcup_{i=n+1}^{\infty} T^{-i}B \]
\[ = A. \]

Hence, by ergodicity \( \mu(A) \in \{0, \mu(X)\} \). Therefore, since \( \mu(A \Delta B) = 0 \), that is \( \mu(A) = \mu(B) \), we have that \( \mu(B) \in \{0, \mu(X)\} \).

For (2) \( \Rightarrow \) (3),

let \( A \in \mathcal{B} \) where \( \mu(A) > 0 \), and let \( A_1 = \bigcup_{n=1}^{\infty} T^{-n}A \). It is immediate that \( T^{-1}A_1 \subset A_1 \) and since \( T \) is measure preserving, \( \mu(T^{-1}A_1) = \mu(A_1) \). Therefore \( \mu(T^{-1}A_1 \Delta A_1) = 0 \). By (2) we then have that \( \mu(A_1) \in \{0, \mu(X)\} \).

However

\[ \mu(A_1) = \mu(\bigcup_{n=1}^{\infty} T^{-n}A) \geq \mu(T^{-1}A) = \mu(A) > 0. \]

Thus \( \mu(A_1) = \mu(X) \).

For (3) \( \Rightarrow \) (4),

Let \( A, B \in \mathcal{B}, \mu(A) > 0, \mu(B) > 0 \). By (3) we have that \( \mu(\bigcup_{n=1}^{\infty} T^{-n}A) = \mu(X) \)
so that

\[
0 < \mu(B) = \mu \left( B \cap \bigcup_{n=1}^{\infty} T^{-n}A \right) = \mu \left( \bigcup_{n=1}^{\infty} B \cap T^{-n}A \right) \leq \sum_{n=1}^{\infty} \mu(B \cap T^{-n}A).
\]

We must therefore have \( \mu(B \cap T^{-n}A) > 0 \) for some \( n \in \mathbb{N} \).

Finally, for \( (4) \Rightarrow (1) \),

suppose that \( B \in \mathcal{B} \) and \( T^{-1}B = B \). If \( 0 < \mu(B) < \mu(X) \) then \( \mu(B^c) > 0 \) and

\[
0 = \mu(B \cap B^c) = \mu(T^{-n}B \cap B^c)
\]

for each \( n \in \mathbb{N} \), which contradicts (4).

To continue with further results that we use, we must first establish Birkhoff’s ergodic Theorem.

### 3.1 Birkhoff’s Individual Ergodic Theorem

Birkhoff’s Individual ergodic theorem, established in 1931, was one of the first major results in ergodic theory. We present it as, like much of the work in ergodic theory, our results, at least implicitly, rely on it. The result provides conditions under which Boltzmann’s original ergodic hypothesis will hold. That is, if we consider the action of the iterations of a transformation on a point as the passage of that point through time, the theorem provides conditions to ensure that the ‘average amount of time’ spent in any set is equal
to the proportion of the entire space that the set occupies. Our treatment of Birkoff’s Individual Ergodic Theorem mainly follows Walters [46].

Our main use for this theorem comes from Theorem 3.3 which is vital in proving Theorem 6.3, which along with Theorem 6.4 is a central theorem in this work.

First we need to establish a technical result. This result demonstrates a property of operators with respect to measure preserving transformations.

**Definition 3.1**

Let $N$ be a normed linear space. We call a continuous linear transformation of $N$ into itself an **operator**.

**Definition 3.2**

Let $(X, B, \mu)$ be a probability space with a measure preserving transformation $T$. Then the induced operator $U_T : L^0(X, B, \mu) \rightarrow L^0(X, B, \mu)$ is defined by $(U_T(f))(x) = f(T(x))$, for each $f \in L^0(X, B, \mu)$ and for all $x \in X$. We use $L^0(X, B, \mu)$ to denote the space of all real valued measurable functions of $(X, B, \mu)$.

**Lemma 3.2**

Let $(X, B, \mu)$ be a measure space of finite measure and let $T$ be a measure preserving transformation on $X$, and let $p \geq 1$. Then

$$||U_T f||_p = ||f||_p$$

for each $f \in L^p(X, B, \mu)$.

**Proof:**

Consider a simple real valued $f \in L^1(X, B, \mu)$, so that for some finite sequence
\( \{a_i\}_{i=1}^k \) of real numbers and finite disjoint sequence \( \{E_i\}_{i=1}^k \) of measurable subsets of \( X \), we have
\[
f = \sum_{i=1}^{k} a_i \chi_{E_i}.
\]

Then using that \( T \) is measure-preserving
\[
\int_X U_T f \, d\mu = \int_X f \circ T \, d\mu
\]
\[
= \int_X \sum_{i=1}^{k} a_i \chi_{T^{-1}E_i} \, d\mu
\]
\[
= \sum_{i=1}^{\infty} a_i \mu(T^{-1}E_i)
\]
\[
= \sum_{i=1}^{\infty} a_i \mu(E_i)
\]
\[
= \int_X f \, d\mu.
\]

Now let \( f \in L^p(X, B, \mu) \) and \( \{f_n\} \) be a sequence of simple functions, with \( |f_n| \leq |f| \) for each \( n \in \mathbb{N} \), in \( L^p(X, B, \mu) \) converging to \( f \). We now have that \( \{|f_n|^p\} \) is a sequence of simple real valued functions in \( L^1(X, B, \mu) \) converging to \( |f|^p \in L^1(X, B, \mu) \) with \( |f_n|^p \leq |f|^p \) for all \( n \in \mathbb{N} \). We therefore also have that \( \{U_T(|f_n|^p)\} \) is a sequence of simple functions converging to \( U_T(|f|^p) \) that are bounded above by \( U_T(|f|^p) \). Using that
\[
\int_X U_T(|f_n|^p) \, d\mu = \int_X |f_n|^p \, d\mu
\]
for each \( n \in \mathbb{N} \), we therefore have
\[
||U_T(f)||_p = \int_X U_T(|f|^p) \, d\mu
\]
\[
= \lim_{n \to \infty} \int_X U_T(|f_n|^p) \, d\mu
\]
\[
= \lim_{n \to \infty} \int_X |f_n|^p \, d\mu
\]
\[
= \int_X |f|^p \, d\mu
\]
\[
= ||f||_p.
\]
We now present the result in a fairly general form so as to provide for the uses that we wish to make of it. We present the theorem in the form of two technical results and the theorem itself. The first Lemma is called the Maximal Ergodic Theorem.

Lemma 3.3 (Maximal Ergodic Theorem)

Let \((X, \mathcal{B}, \mu)\) be the usual Lebesgue measure space on \(X = \mathbb{R}\). Let

\[
U : L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{B}, \mu)
\]

be a positive linear operator with \(\|U\| \leq 1\). Let \(N > 0\) be an integer and let \(f \in L^1(X, \mathcal{B}, \mu)\). Define \(f_0 = 0\), \(f_n = f + Uf + \ldots + U^{n-1}f\) for \(n \geq 1\) and \(F_N = \max_{0 \leq n \leq N} f_n \geq 0\). Then

\[
\int_{\{z : F_N(z) > 0\}} f \, d\mu \geq 0.
\]

Proof:

\(f \in L^1(X, \mathcal{B}, \mu) \Rightarrow Uf \in L^1(X, \mathcal{B}, \mu)\) and hence we have that \(F_N \in L^1(X, \mathcal{B}, \mu)\) for each \(N\).

Also for \(0 \leq n \leq N\) we have, by definition, that \(F_n \geq f_n\) so that \(UF_N \geq Uf_n\) by positivity.

Now for each \(0 \leq n \leq N\)

\[
f_{n+1} = f + Uf + \ldots + U^{n+1}f
= f + U(f + Uf + \ldots + U^{n}f)
= f + U(f_n)
\leq f + UF_N.
\]

Since \(F_N(x) > f_0(x) = 0\) whenever \(F_N(x) > 0\) we now have

\[
UF_N(x) + f(x) \geq \max_{1 \leq n \leq N} f_n(x)
\]
= \max_{0 \leq n \leq N} f_n(x) \\
= F_N(x).

This implies that \( f \geq F_N - UF_N \) on \( A = \{ x : F_N(x) > 0 \} \). Hence, noting that \( F_N = 0 \) on \( A^c \) and that since \( F_N \geq 0, UF_N \geq 0 \) we have

\[
\int_A f d\mu \geq \int_A F_N d\mu - \int_A UF_N d\mu \\
= \int_X F_N d\mu - \int_A UF_N d\mu \\
\geq \int_X F_N d\mu - \int_X UF_N d\mu \\
\geq 0,
\]

where the final relation follows from the fact that \(||U|| \leq 1\).

\hfill \Box

A corollary of the Maximal Ergodic Theorem that we use in proving Birkoff's Theorem is now presented.

**Corollary 3.1**

Let \((X, \mathcal{B}, \mu)\) be a measure space and let \( T : X \rightarrow X \) be measure-preserving. If \( A \in \mathcal{B} \) is such that \( T^{-1}(A) \subset A \) and \( \mu(A) < \infty \), \( g \in L^1(X, \mathcal{B}, \mu) \) and

\[
B_\alpha = \left\{ x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i(x)) > \alpha \right\},
\]

then

\[
\int_{B_\alpha \cap A} g d\mu \geq \alpha \mu(B_\alpha \cap A).
\]

**Proof:**

Note that \( \mu(A) < \infty \) and that since \( T^{-1}(A) \subset A \) we can define a transformation, \( T|A \) on \( A \), by

\[
(T|A)(x) = T(x) \quad \text{for all } x \in A.
\]
Note that $T|A$ is measure-preserving on subsets of $A$. Let $\mathcal{B}_A$ be the $\sigma$-algebra $\{B \cap A : B \in \mathcal{B}\}$, $\mu_A$ be the measure $\mu$ restricted to $A$ and consider the operator $U_{T|A} : L^1(A, \mathcal{B}_A, \mu_A) \to L^1(A, \mathcal{B}_A, \mu_A)$ defined by

$$U_{T|A}(f) = f(T|A).$$

Then clearly $U_{T|A}$ is positive and linear, and using Lemma 3.2,

$$||U_{T|A}f|| = ||f||$$

for any $f \in L^1(A, \mathcal{B}_A, \mu_A)$ so that $||U_{T|A}|| \leq 1$.

Let $f$ be defined on $A$ by $f = g - \alpha$, and, as in the Maximal Ergodic Theorem, let

$$f_n = f + U_{T|A}f + \ldots + U_{T|A}^{n-1}f$$

and

$$F_N(x) = \begin{cases} \max_{0 \leq n \leq N} f_n(x), & x \in A, \\ 0, & x \notin A \end{cases}$$

Now note that

$x \in B_\alpha \cap A \iff \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} g(T|A^i(x)) > \alpha$

$\iff \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f((T|A)^i(x)) > 0$

$\iff \sup_{n \geq 1} \sum_{i=0}^{n-1} f((T|A)^i(x)) > 0$

$\iff \sum_{i=0}^{N-1} f((T|A)^i(x)) > 0$, for some $N \geq 1,$

$\iff \sum_{i=0}^{N-1} U_{T|A}^i(f)(x) > 0$, for some $N \geq 1,$

$\iff x \in \{x : F_N(x) > 0\}$, for some $N \geq 1,$

$\iff x \in \bigcup_{N=0}^{\infty} \{x : F_N(x) > 0\}$.
(Note that in lines 4, 5 and 6 of the above it suffices to use the same $N$, though to ensure that reading the implications in both directions makes sense we use "for some $N \geq 1" in each case.)

Now, note that $\{x : F_N(x) > 0\}_{N=0}^{\infty}$ is an increasing sequence of sets for which (by the Maximal Ergodic Theorem) $\int_{\{x:F_N(x)>0\}} f d\mu \geq 0$ for each $N$, as we also have that $\{\bigcup_{i=0}^{N} \{x:F_i(x)>0\}\}$ is bounded by $f \chi_A$ and

$$\lim_{N \to \infty} f \chi_{\bigcup_{i=0}^{N} \{x:F_i(x)>0\}} = f \chi_{\bigcup_{N=0}^{\infty} \{x:F_N(x)>0\}},$$

(with pointwise convergence) so that by Lebesgue's Dominated Convergence Theorem

$$\int_{B_\alpha \cap A} f d\mu = \int_{\bigcup_{N=0}^{\infty} \{x:F_N(x)>0\} \cap A} f d\mu$$

$$= \int_A f \chi_{\bigcup_{N=0}^{\infty} \{x:F_N(x)>0\}} d\mu$$

$$= \lim_{N \to \infty} \int_A f \chi_{\{x:F_N(x)>0\}} d\mu$$

$$\geq 0.$$

Therefore

$$\int_{B_\alpha \cap A} g d\mu = \int_{B_\alpha \cap A} \alpha d\mu + \int_{B_\alpha \cap A} f d\mu$$

$$= \alpha \mu(B_\alpha \cap A) + \int_{B_\alpha \cap A} f d\mu$$

$$\geq \alpha \mu(B_\alpha \cap A).$$

We can now present Birkoff's Theorem. The Theorem can be validated for spaces of infinite measure, however, we only need the result for spaces of finite measure and hence shall only prove the theorem for this case. Also, we
have so far only considered real valued functions, but now consider complex valued functions. This does not present difficulties in using the previous results as we shall prove the Theorem by considering a complex valued function as being made of its real and imaginary parts.

**Theorem 3.2 (Birkhoff's Individual Ergodic Theorem)**

Suppose that \((X, \mathcal{B}, \mu)\) is a measure space, \(T : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{B}, \mu)\) is measure preserving, \(\mu(X) < \infty\) and \(f \in L^1(X, \mathcal{B}, \mu)\). Then there is an \(f^* \in L^1(X, \mathcal{B}, \mu)\) such that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = f^* \quad \mu - \text{a.e.}
\]

Also \(f^* \circ T = f^*\) \(\mu\)-a.e. and

\[
\int f^* d\mu = \int f d\mu.
\]

**Proof:**

We first consider \(f\) being real valued, from which the general result will follow easily. For such an \(f\) let

\[
f^*(x) = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))
\]

and

\[
f_*(x) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)).
\]

We need to show that \(f^* = f_*\) \(\mu\)-a.e. In doing so we will show that \(f^* = f^* \circ T\) \(\mu\)-a.e. To show that \(f^* = f_*\) \(\mu\)-a.e. for real numbers \(\alpha\) and \(\beta\), let

\[
E_{\alpha, \beta} = \{x \in X : f_*(x) < \beta \quad \text{and} \quad \alpha < f^*(x)\}.
\]

Since

\[
\{x : f_*(x) < f^*(x)\} = \bigcup \{E_{\alpha, \beta} : \beta < \alpha; \beta, \alpha \in \mathbb{Q}\},
\]

33
if we can show that $\mu(E_{\alpha,\beta}) = 0$ whenever $\beta < \alpha$, then, as $f_* \leq f^*$ always, we will then have that $f^* = f_*$ almost everywhere.

We will need to show that $E_{\alpha,\beta} \in B$ for each choice of $\alpha$ and $\beta$. For each $n \in \mathbb{N}$ let

$$a_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T_i(x)), \quad \text{for all } x \in X.$$  

Also let

$$E_{\alpha_n} = \{x \in X : a_n(x) > \alpha\}$$

and

$$E_{\beta_n} = \{x \in X : a_n(x) < \beta\}.$$  

Now note that $x \in \{x \in X : f^*(x) > \alpha\}$ if and only if for each $n \in \mathbb{N}$ there exists an $n_0 > n$ such that $a_{n_0}(x) > \alpha$. That is

$$\{x \in X : f^*(x) > \alpha\} = \bigcap_{n \in \mathbb{N}} \bigcup_{j=n}^{\infty} E_{\alpha_j}.$$  

Also $x \in \{x \in X : f_*(x) < \beta\}$ if and only if there is a $n_0 \in \mathbb{N}$ such that for all $n > n_0$, $a_n(x) < \beta$. That is

$$\{x \in X : f_*(x) < \beta\} = \bigcup_{n \in \mathbb{N}} \bigcap_{j=n}^{\infty} E_{\beta_j}.$$  

Since $f \in L^1(X, B, \mu)$ it is clear that $a_n$ is measurable for each $n \in \mathbb{N}$ and hence $E_{\alpha_n}$ and $E_{\beta_n}$ are in $B$ for each $n \in \mathbb{N}$. Therefore

$$E_{\alpha,\beta} = \bigcup_{n \in \mathbb{N}} \bigcap_{j=n}^{\infty} E_{\beta_j} \cap \bigcap_{n \in \mathbb{N}} \bigcup_{j=n}^{\infty} E_{\alpha_j} \in B.$$  

We now show that $f^* = f^* \circ T \mu$-a.e. Since

$$f^* = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T_i(x)) = \limsup_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n+1-1} f(T_i(x)),$$
we can note that

\[ f^* \circ T = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^{i+1}(x)) \]

\[ = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(T^i(x)) \]

\[ = \limsup_{n \to \infty} \left( \frac{n+1}{n} \left( \frac{1}{n+1} \sum_{i=0}^{n} f(T^i(x)) \right) - \frac{f(x)}{n} \right). \]

Now for any \( \varepsilon > 0 \) there is an \( n_0 \in \mathbb{N} \) such that for all \( n > n_0 \),

\[ 1 < \frac{n+1}{n} < 1 + \varepsilon \quad \text{and} \quad \left| \frac{f(x)}{n} \right| < \varepsilon. \]

In this case, for any \( \varepsilon > 0 \),

\[ f^* - \varepsilon = \limsup_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} f(T^i(x)) - \varepsilon \]

\[ \leq \limsup_{n \to \infty} \left( \frac{n+1}{n} \left( \frac{1}{n+1} \sum_{i=0}^{n} f(T^i(x)) \right) - \frac{f(x)}{n} \right) \]

\[ \leq (1 + \varepsilon) \left( \limsup_{n \to \infty} \frac{1}{n+1} \sum_{i=0}^{n} f(T^i(x)) \right) + \varepsilon \]

\[ = (1 + \varepsilon) f^* + \varepsilon. \]

That is, for any \( \varepsilon > 0 \)

\[ f^* - \varepsilon \leq f^* \circ T \leq (1 + \varepsilon) f^* + \varepsilon \]

so that

\[ f^* \circ T = f^*. \]

In a similar manner we can show that

\[ f_* \circ T = f_. \]

Now if \( x \in E_{\alpha, \beta} \), then for any \( x_1 \in T^{-1}(x) \) we have that

\[ f^*(x) = f^*T(x_1) = f^*(T(x_1)) = f^*(x) > \alpha. \]
Similarly, $f_*(x_1) < \beta$ and hence $T^{-1}(E_{\alpha,\beta}) \subseteq E_{\alpha,\beta}$.

Further, if we set
\[
B_{\alpha} = \left\{ x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) > \alpha \right\},
\]
clearly $E_{\alpha,\beta} \subseteq B_{\alpha}$ and hence $E_{\alpha,\beta} \cap B_{\alpha} = E_{\alpha,\beta}$. Thus, by Corollary 3.1 we have that
\[
\int_{E_{\alpha,\beta}} f \, d\mu = \int_{E_{\alpha,\beta} \cap B_{\alpha}} f \, d\mu \geq \alpha \mu(E_{\alpha,\beta} \cap B_{\alpha}) = \alpha \mu(E_{\alpha,\beta}),
\]
so that $\int_{E_{\alpha,\beta}} f \, d\mu \geq \alpha \mu(E_{\alpha,\beta})$.

If we then replace $f$, $\alpha$ and $\beta$ by $-f$, $-\beta$ and $-\alpha$ respectively, we get that
\[
E_{-\beta,-\alpha} \cap B_{-\beta} = E_{-\beta,-\alpha}
\]
and that
\[
E_{-\beta,-\alpha} = \{ x \in X : (-f)_* < -\alpha; (-f)^* > -\beta \}
\]
\[
= \{ x \in X : -f^* < -\alpha; -f_* > -\beta \}
\]
\[
= \{ x \in X : f^* > \alpha; f_* < \beta \}
\]
\[
= E_{\alpha,\beta}.
\]

Hence, $T^{-1}(E_{-\beta,-\alpha}) = E_{-\beta,-\alpha}$. Again using Corollary 3.1 we get
\[
\int_{E_{\alpha,\beta}} -f \, d\mu = \int_{E_{-\beta,-\alpha}} -f \, d\mu \geq -\beta \mu(E_{-\beta,-\alpha}) = -\beta \mu(E_{\alpha,\beta}),
\]
so that
\[
\int_{E_{\alpha,\beta}} f \, d\mu \leq \beta \mu(E_{\alpha,\beta}).
\]

36
We now have that
\[ \alpha \mu(E_{\alpha, \beta}) \leq \beta \mu(E_{\alpha, \beta}) \]
so that if \( \beta < \alpha \) we must have \( \mu(E_{\alpha, \beta}) = 0 \). This, as discussed above, must give that \( f^* = f_* \mu \)-a.e. Therefore
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = f^* \mu \text{-a.e.}
\]
We now wish to show that \( f^* \in L^1(X, \mathcal{B}, \mu) \).

Let
\[
g_n(x) = \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \right|,
\]
which is clearly in \( L^1(X, \mathcal{B}, \mu) \) for each \( n \in \mathbb{N} \). Then by Lemma 3.2
\[
\int g_n d\mu = \int \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i) \right| d\mu
\leq \int \frac{1}{n} \sum_{i=0}^{n-1} |f(T^i)| d\mu
= \frac{1}{n} \sum_{i=0}^{n-1} \int |f(T^i)| d\mu
= \frac{1}{n} \sum_{i=0}^{n-1} \int |f| d\mu
= \int |f| d\mu,
\]
for each \( n \in \mathbb{N} \), so that we can apply Fatou's Lemma to give us
\[
\int |f^*| d\mu = \int \lim_{n \to \infty} |g_n| d\mu
\leq \liminf_{n \to \infty} \int |g_n| d\mu
\leq \int |f| d\mu,
\]
37
so that \( |f^*| \in L^1(X, B, \mu) \) and therefore \( f^* \in L^1(X, B, \mu) \).

It now only remains to show that \( \int f \, d\mu = \int f^* \, d\mu \). Let

\[
D^n_k = \left\{ x \in X : \frac{k}{n} \leq f^*(x) < \frac{k+1}{n} \right\},
\]

where \( k \in \mathbb{Z}, n \in \mathbb{N} \). (Note that as \( f^* \in L^1(X, B, \mu) \), \( f^* \) is measurable and hence \( D^n_k \) must be in \( B \) for each \( k \) and \( n \).) Then, for each small \( \varepsilon \), we must have that if \( x \in D^n_k \) then

\[
\sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \geq \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \geq \frac{k}{n} > \frac{k}{n} - \varepsilon,
\]

so that \( x \in B(k/n-\varepsilon) \). That is, \( D^n_k \subset B(k/n-\varepsilon) \) so that \( B(k/n-\varepsilon) \cap D^n_k = D^n_k \).

Hence, by use of Corollary 3.1 we have that

\[
\int_{D^n_k} f \, d\mu = \int_{D^n_k \cap B(k/n-\varepsilon)} f \, d\mu \geq \left( \frac{k}{n} - \varepsilon \right) \mu(D^n_k),
\]

so that, since \( \varepsilon \) is arbitrarily small

\[
\int_{D^n_k} f \, d\mu \geq \frac{k \mu(D^n_k)}{n}.
\]

By the above inequality we have that

\[
\int_{D^n_k} f^* \, d\mu \leq \frac{k+1}{n} \mu(D^n_k) \leq \frac{\mu(D^n_k)}{n} + \int_{D^n_k} f \, d\mu.
\]

Summing over all \( k \) we then have

\[
\int_X f^* \, d\mu \leq \frac{\mu(X)}{n} + \int_X f \, d\mu.
\]

Since this is true for any \( n \in \mathbb{N} \), we can allow \( n \) to become arbitrarily large so that, since \( \mu(X) < \infty \), we have

\[
\int f^* \, d\mu \leq \int f \, d\mu.
\]
Applying the same argument to $-f$ gives that
\[ \int (-f)^* d\mu \leq \int -fd\mu, \]
so that
\[ \int f_\ast d\mu \geq \int fd\mu. \]
Since $f^* = f_\ast \mu\text{-a.e.}$ we must have
\[ \int f^* d\mu = \int fd\mu. \]
For a complex valued $f$, we consider $f = f_1 + if_2$ with $f_1$ and $f_2$ real valued.
Then,
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f_1(T^i(x)) + i \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f_2(T^i(x)).
\]
As $f_1$ and $f_2$ are real valued it follows that there are $f_1^*$ and $f_2^*$ such that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f_1(T^i(x)) = f_1^* \mu\text{-a.e.}
\]
and
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f_2(T^i(x)) = f_2^* \mu\text{-a.e.}
\]
Further $f_1^* \circ T = f_1^*$, $f_2^* \circ T = f_2^*$, $\int f_1^* d\mu = \int f_1 d\mu$ and $\int f_2^* d\mu = \int f_2 d\mu$.
Setting $f^* = f_1^* + if_2^*$, we therefore have that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = f_1^* + if_2^* = f^* \mu\text{-a.e.},
\]
\[
f^* \circ T = f_1^* \circ T + if_2^* \circ T = f_1^* + if_2^* = f^*
\]
and
\[
\int f^* d\mu = \int f_1^* d\mu + i \int f_2^* d\mu = \int f_1 d\mu + i \int f_2 d\mu = \int fd\mu. \]
The following is a Corollary of Theorem 3.2 that is important to this re-
search and so is written as a Theorem.

**Theorem 3.3**

Let \((X, \mathcal{B}, \mu)\) be a probability space and \(T : X \to X\) be a measure preserving transformation. Then \(T\) is ergodic if and only if for all \(A, B \in \mathcal{B}\)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i} A \cap B) = \mu(A) \mu(B).
\]

**Proof:**

Suppose that \(T\) is ergodic. Put \(f = \chi_A\) so that by Theorem 3.2 we have that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x)) = \chi^*_A
\]

for some \(\chi^*_A \in L^1(X, \mathcal{B}, \mu)\) such that \(\chi^*_A \circ T = \chi^*_A\).

Now suppose that \(\chi^*_A\) is not constant \(\mu\)-a.e. then for some \(a \in \mathbb{R}\), if we put

\[
C = \{ x \in X : \chi^*_A(x) > a \}
\]

and

\[
D = \{ x \in X : \chi^*_A(x) \leq a \},
\]

we will have \(\mu(C) > 0\) and \(\mu(D) > 0\). Then by Theorem 3.1 we know that there is an \(i \in \mathbb{N}\) such that \(\mu(T^{-i} C \cap D) > 0\). Let \(x \in T^{-i} C \cap D\), so that \(T^i(x) \in C\), say \(T^i(x) = y \in C\) and \(x \in D\). In this case we have that \(\chi^*_A(x) \leq a\) and as \(\chi^*_A \circ T = \chi^*_A\)

\[
\chi^*_A(x) = \chi^*_A(T(x))
\]
... 

= \chi_A^*(T^i(x)) \\
= \chi_A^*(y) \\
> a.

This contradiction implies that \( \chi_A^* \) is constant \( \mu \)-a.e. Since, by Theorem 3.2, \( \int \chi_A d\mu = \int \chi_A^* d\mu \) we have that

\[
\chi_A^* = \mu(X)\chi_A = \int \chi_A^* d\mu = \int \chi_A d\mu = \mu(A).
\]

Now multiplying both sides of (5) by \( \chi_B \) we have that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x))\chi_B = \mu(A)\chi_B.
\]

We have immediately that

\[
\int \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x))\chi_B d\mu = \int \mu(A)\chi_B d\mu
\]

\[
= \mu(A)\mu(B).
\]

Also, as for all \( x \in X \),

\[
(\chi_A(T^i)\chi_B)(x) = \chi_{T^{-i}A \cap B}(x)
\]

and as for all \( n \in \mathbb{N} \)

\[
\frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x))\chi_B \leq 1
\]

we have, by the Dominated Convergence Theorem,

\[
\int \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i(x))\chi_B d\mu = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \int \chi_A(T^i(x))\chi_B d\mu
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \int \chi_{T^{-i}A \cap B} d\mu
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B).
\]
Therefore, we have that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A) \mu(B).
\]

Conversely, suppose that the convergence property holds and suppose \( T^{-1} E = E \) for some \( E \in \mathcal{B} \). Set \( A = B = E \) in the convergence property to get
\[
\mu(E) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(E) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}E \cap E) = \mu(E)^2,
\]
so that \( \mu(E) \) must be either 1 or 0. We therefore have, by definition, that \( T \) must be ergodic.

\[\diamondsuit\]

### 3.2 Mixing

For much of this Thesis we are investigating the relationship between the generation of the Carathéodory definition of measurable sets and various 'strengths' of ergodicity. The varying levels of strength are defined in terms of the strength of convergence that can be obtained when altering the desired type of convergence in Corollary 3.2. These 'strengths' are defined as levels of mixing. We picture this concept by thinking of the process of iterating \( T \) as a mixing mechanism. Then if \( T \) mixes strongly, any given starting set should be distributed evenly throughout, or dissolved in, the space \( X \). If \( T \) doesn't mix strongly then there should be parts of a starting set that stay together, allowing the mixed space to remain lumpy. We define the levels of mixing formally as follows.
Definition 3.3

Let $T$ be a measure preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. Then

1. $T$ is **weak mixing** if for all $A, B \in \mathcal{B}$

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0
$$

2. $T$ is **strong mixing** if for all $A, B \in \mathcal{B}$

$$
\lim_{n \to \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).
$$

The investigation of the relationship between mixing and the generation of the Carathéodory definition of measurable sets requires the use of some of the properties of mixing transformations. The remainder of this chapter is used to establish these properties.

Our first proposition establishes a chain of implications that justifies the adjectives used in the definitions of the levels of mixing.

**Proposition 3.1**

*If $T$ is strong mixing, it is weak mixing. If $T$ is weak mixing, it is ergodic.*

**Proof:**

Suppose that $T$ is strong mixing, so that for all $A, B \in \mathcal{B}$

$$
\lim_{n \to \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).
$$

Select any two sets $A, B \in \mathcal{B}$ and any $\varepsilon > 0$, then there is an $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$ we have

$$
|\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| < \frac{\varepsilon}{2}.
$$
Also, there is an \( n_1 \), such that for all \( n \geq n_1 \)

\[
\frac{1}{n} \sum_{i=0}^{n_0-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| < \frac{\varepsilon}{2}.
\]

Thus, setting such an \( n_1 > n_0 \), for all \( n > n_1 \)

\[
\frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = \frac{1}{n} \sum_{i=0}^{n_0-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| + \frac{1}{n} \sum_{i=n_0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)|
\]

\[
< \frac{\varepsilon}{2} + \frac{1}{n} \sum_{i=n_0}^{n-1} \frac{\varepsilon}{2} < \varepsilon.
\]

As this is true for any \( \varepsilon > 0 \) and for any \( A, B \in \mathcal{B}, T \) must be weak mixing.

(This argument is in fact a special case of

\[
\lim_{n \to \infty} a_n = 0 \Rightarrow \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |a_i| = 0
\]

when \( \{a_n\} \) is a bounded sequence. However, we do not prove this more general result until Theorem 3.5.)

Now suppose that \( T \) is weak mixing, so that for all \( A, B \in \mathcal{B} \) we have

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0
\]

Then

\[
0 \leq \limsup_{n \to \infty} \left| \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) \right| \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0.
\]
Therefore
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) = 0
\]
and hence
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B) = 0
\]
so that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B).
\]
Thus by Theorem 3.3, \(T\) is ergodic.

It can be shown that in Theorem 3.3 and in the definitions of weak and strong mixing we need only consider convergence of the relevant limits over sets from a semi-algebra that generates \(\mathcal{B}\). However, we only require the result for weak mixing. We prove this assertion below.

**Definition 3.4** A collection of subsets \(S\) of a set \(X\) is a **semi-algebra** if the following three conditions hold.

(i) \(\emptyset \in S\).

(ii) If \(A, B \in S\), then \(A \cap B \in S\).

(iii) If \(A \in S\) then \(X - A = \bigcup_{i=1}^{n} E_i\) where \(E_1, E_2, \ldots, E_n \in S\) are pairwise disjoint subsets of \(X\).

The **\(\sigma\)-algebra generated by \(S\)** is the smallest \(\sigma\)-algebra containing \(S\).

**Theorem 3.4**

Let \((X, \mathcal{B}, \mu)\) be a measure space and let \(S\) be a semi-algebra that generates \(\mathcal{B}\). Let \(T : X \to X\) be a measure preserving transformation. Then \(T\) is weak
mixing if and only if for all \( A, B \in S \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0.
\]

**Proof:**

Each member of the algebra generated by \( S \), \( \mathcal{A}(S) \), can be written as a finite disjoint union of members of \( S \). So if the convergence property holds for all pairs of members of \( S \) then we can sum the various pairings of the elements of \( S \) that make up any given pair of elements in \( \mathcal{A}(S) \) as follows.

Let \( A = \bigcup_{i=1}^{a} A_i \), where \( A \in \mathcal{A}(S) \) and \( A_1, \ldots, A_a \in S \). Also let \( B = \bigcup_{i=1}^{b} B_i \), where \( B \in \mathcal{A}(S) \) and \( B_1, \ldots, B_b \in S \). Then

\[
0 \leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)|
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \mu \left( \bigcup_{j=1}^{a} T^{-i}A_j \right) \cap \left( \bigcup_{k=1}^{b} B_k \right) \right| - \mu \left( \bigcup_{j=1}^{a} A_j \right) \mu \left( \bigcup_{k=1}^{b} B_k \right)
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \mu \left( \bigcup_{j=1}^{a} T^{-i}A_j \right) \cap \left( \bigcup_{k=1}^{b} B_k \right) \right| - \sum_{j=1}^{a} \sum_{k=1}^{b} \mu(A_j)\mu(B_k)
\]

\[
\leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=1}^{a} \sum_{k=1}^{b} \mu((T^{-i}A_j) \cap B_k) - \sum_{j=1}^{a} \sum_{k=1}^{b} \mu(A_j)\mu(B_k)
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=1}^{a} \sum_{k=1}^{b} \mu((T^{-i}A_j) \cap B_k) - \mu(A_j)\mu(B_k)
\]

\[
\leq \sum_{j=1}^{a} \sum_{k=1}^{b} \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu((T^{-i}A_j) \cap B_k) - \mu(A_j)\mu(B_k)
\]

\[
= \sum_{j=1}^{a} \sum_{k=1}^{b} 0
\]

\[
= 0.
\]

We therefore have that if the convergence property holds for all pairs of ele-
ments of $S$ then the convergence property holds for all pairs of elements of $A(S)$.

Now let $\varepsilon > 0$ and $A, B$ be any pair of elements in $B$. We can now choose some $A_0, B_0 \in A(S)$ such that $\mu(A \Delta A_0) < \varepsilon$ and $\mu(B \Delta B_0) < \varepsilon$. Then for any $i \geq 0$ we have

$$(T^{-i} A \cap B) \Delta (T^{-i} A_0 \cap B_0) \subset (T^{-i} A \Delta T^{-i} A_0) \cup (B \Delta B_0)$$

so that, since $T$ is measure preserving

$$|\mu(T^{-i} A \cap B) - \mu(A)\mu(B)| \leq |\mu(T^{-i} A \cap B) - \mu(T^{-i} A_0 \cap B_0)|$$
$$+ |\mu(T^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)|$$
$$+ |\mu(A_0)\mu(B_0) - \mu(A)\mu(B)|$$
$$+ |\mu(A)\mu(B_0) - \mu(A)\mu(B)|$$
$$< 2\varepsilon + |\mu(T^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)|$$
$$+ |\mu(B_0)(\mu(A_0) - \mu(A))|$$
$$+ |\mu(A)(\mu(B_0) - \mu(B))|$$
$$< 4\varepsilon + |\mu(T^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)|.$$

Then as the convergence property holds for $A_0, B_0$ we have

$$0 \leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i} A \cap B) - \mu(A)\mu(B)|$$
$$< \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 4\varepsilon + |\mu(T^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)|$$
$$= 4\varepsilon + \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i} A_0 \cap B_0) - \mu(A_0)\mu(B_0)|$$
$$= 4\varepsilon.$$

Since $\varepsilon$ is arbitrary the desired result follows.
The converse is obvious as if $T$ is weak mixing, then the convergence property holds for all pairs of elements of $B$ which includes all pairs of elements from $S$.

The following two theorems show that a weak mixing transformation is almost a strong mixing transformation in the sense that the convergence required for strong mixing will occur for each $A, B \in B$ once a relatively 'small set' of the $\mu(T^{-i}A \cap B)$ are removed. They also assist in investigating the relationship between the mixing properties of $T$ and $T \times T$. This relationship is very important in the heuristic development of this work.

We first define the concept necessary to specify what a 'small set' is in $\mathbb{N}$. We denote the cardinality of a subset $J$ of $\mathbb{Z}$ by $|J|$ and we denote $\mathbb{N} \cup \{0\}$ by $\mathbb{Z}^+$.

**Definition 3.5**

A subset $J$ of $\mathbb{Z}^+$ is said to have **density zero** if

$$\lim_{n \to \infty} \frac{|J \cap \{0, 1, 2, \ldots, n - 1\}|}{n} = 0.$$

**Theorem 3.5**

If $\{a_n\}$ is a bounded sequence of real numbers then the following are equivalent:

1. $$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0.$$

2. There exists a subset $J$ of $\mathbb{Z}^+$ of density zero such that

$$\lim_{n \to \infty, n \notin J} a_n = 0.$$
3. 

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i|^2 = 0. \]

Proof:

If \( M \subset \mathbb{Z}^+ \) let \( \alpha_M(n) \) denote the cardinality of \( M \cap \{0, 1, \ldots, n-1\} \).

We prove the theorem by showing (1) \( \Rightarrow \) (2) \( \Rightarrow \) (1) and (2) \( \Leftrightarrow \) (3).

(1) \( \Rightarrow \) (2)

Suppose that

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0. \]

Let

\[ J_k = \left\{ n \in \mathbb{Z}^+: |a_n| \geq \frac{1}{k} \right\}, \quad (k > 0). \]

Then \( J_1 \subset J_2 \subset \ldots \) and also we must have that for each \( k \)

\[ \frac{1}{n} \sum_{i=0}^{n-1} |a_i| \geq \frac{1}{k} \alpha_k(n). \]

Consequently, since

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0, \]

\[ \lim_{n \to \infty} \frac{1}{n} \alpha_k(n) = 0, \]

so that \( J_k \) must be of density zero for each \( k \in \mathbb{N} \). Therefore using the convergence of \( \alpha_{J_k}(n)/n \) there must exist integers \( 0 = l_0 < l_1 < l_2 < \ldots \) such that for \( n \geq l_k \),

\[ \frac{1}{n} \alpha_{J_{k+1}}(n) < \frac{1}{k + 1}. \]

Set

\[ J = \bigcup_{k=0}^{\infty} (J_{k+1} \cap [l_k, l_{k+1})). \]
We now show that $J$ has density zero. Since $J_1 \subset J_2 \subset \ldots$, we have that if $l_k \leq n < l_{k+1}$

$$J \cap [0, n) = (J \cap [0, l_k)) \cup (J \cap [l_k, n))$$

$$= \bigcup_{i=0}^{k-1} (J \cap [l_i, l_{i+1})) \cup (J \cap [l_k, n))$$

$$= \bigcup_{i=0}^{k-1} (J_i \cap [l_i, l_{i+1})) \cup (J \cap [l_k, n))$$

$$\subseteq \bigcup_{i=0}^{k-1} (J_k \cap [l_i, l_{i+1})) \cup (J_{k+1} \cap [l_k, n))$$

$$= (J_k \cap [0, l_k)) \cup (J_{k+1} \cap [l_k, n))$$

$$\subseteq (J_k \cap [0, l_k)) \cup (J_{k+1} \cap [0, n))$$

and therefore

$$\frac{1}{n} \alpha_J(n) \leq \frac{1}{n} (\alpha_{J_k}(l_k)) + \frac{1}{n} \alpha_{J_{k+1}}(n)$$

$$\leq \frac{1}{n} (\alpha_{J_k}(n) + \alpha_{J_{k+1}}(n))$$

$$< \frac{1}{k} + \frac{1}{k + 1}.$$ 

Hence, since $k \to \infty$ as $n \to \infty$ we have that

$$\lim_{n \to \infty} \frac{\alpha_J(n)}{n} = \lim_{k \to \infty} \frac{1}{k} + \frac{1}{k + 1} = 0,$$

and therefore $J$ has density zero.

Now, if $n > l_k$ and $n \not\in J$, then $n \not\in J_{k+1}$ and hence $|a_n| < 1/(k + 1)$ so that, again, because $k \to \infty$ as $n \to \infty$ we have

$$\lim_{n \to \infty, n \not\in J} |a_n| = 0.$$

(2) $\Rightarrow$ (1)

Since $\{a_n\}$ is bounded, we may suppose that $|a_n| < K$ for all $n \in \mathbb{N}$. Let
\( \varepsilon > 0 \). Then there exists \( N_\varepsilon \in \mathbb{N} \) and a set \( J \) of density zero such that for all \( n > N_\varepsilon \) we have

\[
\frac{\alpha_J(n)}{n} < \frac{\varepsilon}{K + 1}
\]

and, for \( n \notin J \),

\[
|a_n| < \frac{\varepsilon}{K + 1}.
\]

Thus \( n > N_\varepsilon \) implies

\[
\frac{1}{n} \sum_{i=0}^{n-1} |a_i| = \frac{1}{n} \left( \sum_{i \in J \cap \{0, 1, \ldots, n-1\}} |a_i| + \sum_{i \notin J \cap \{0, 1, \ldots, n-1\}} |a_i| \right) < \frac{K \alpha_J(n)}{n} + \frac{\varepsilon}{K + 1} < \frac{K \varepsilon}{K + 1} + \frac{\varepsilon}{K + 1} = \frac{K + 1}{K + 1} \varepsilon = \varepsilon.
\]

Therefore

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0.
\]

(2) \( \Leftrightarrow \) (3)

Note that if \( \{a_n\} \) is bounded then so too is \( \{a_n^2\} \) and that \( \lim_{n \to \infty, n \notin J} |a_n| = 0 \) if and only if \( \lim_{n \to \infty, n \notin J} |a_n^2| = 0 \). Hence applying the above arguments to \( \{a_n^2\} \), gives the result.

\( \diamond \)

**Theorem 3.6**

If \( T \) is a measure preserving transformation of a probability space \((X, \mathcal{B}, \mu)\) then the following are equivalent:

1. \( T \) is weak mixing.
2. For every pair of elements $A, B \in B$ there is a subset $J(A, B)$ of $\mathbb{Z}^+$ of density zero such that
\[
\lim_{n \to \infty, n \notin J(A, B)} \mu(T^{-n} A \cap B) = \mu(A)\mu(B).
\]

3. For every pair of elements $A, B \in B$ we have
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i} A \cap B) - \mu(A)\mu(B)|^2 = 0.
\]

**Proof:**
For any $A, B \in B$, let $a_n = \mu(T^{-n} \cap B) - \mu(A)\mu(B)$. Note that $|a_n| \leq 1$ for all $n \in \mathbb{N}$, that $T$ is weak mixing if and only if
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i} A \cap B) - \mu(A)\mu(B)| = 0,
\]
and that
\[
\lim_{n \to \infty, n \notin J(A, B)} \mu(T^{-n} A \cap B) = \mu(A)\mu(B),
\]
if and only if
\[
\lim_{n \to \infty, n \notin J(A, B)} |\mu(T^{-n} A \cap B) - \mu(A)\mu(B)| = 0.
\]
Thus the Theorem requires, for each $A, B \in B$, that the following are equivalent:

1. 
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i| = 0.
\]

2. There exists a subset $J$ of $\mathbb{Z}^+$ such that
\[
\lim_{n \to \infty, n \notin J} a_n = 0.
\]

3. 
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i|^2 = 0.
\]
which follows directly from Theorem 3.5.

We can now establish some equivalences regarding the transformation $T$ and its cross product $T \times T$.

**Theorem 3.7**

*If $T$ is a measure preserving transformation on a probability space $(X, B, \mu)$ then the following are equivalent:*

1. $T$ is weak mixing.
2. $T \times T$ is ergodic.
3. $T \times T$ is weak mixing.

**Proof:**

We prove the theorem by showing that $(1) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1)$.

$(1) \Rightarrow (3)$

Let $A, B, C, D \in B$. Then by Theorem 3.6 there exist subsets $J_1, J_2$ of $\mathbb{Z}^+$ of density zero such that

$$\lim_{n \to \infty, n \notin J_1} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$$

and

$$\lim_{n \to \infty, n \notin J_2} \mu(T^{-n}C \cap D) = \mu(C)\mu(D).$$

Then

$$\lim_{n \to \infty, n \notin J_1 \cup J_2} (\mu \times \mu)((T \times T)^{-n}(A \times C) \cap (B \times D))$$

$$= \lim_{n \to \infty, n \notin J_1 \cup J_2} \mu(T^{-n}A \cap B)\mu(T^{-n}C \cap D)$$

$$= \mu(A)\mu(B)\mu(C)\mu(D)$$

$$= (\mu \times \mu)(A \times C)(\mu \times \mu)(B \times D).$$
So that

$$\lim_{n \to \infty, n \notin J_1 \cup J_2} |(\mu \times \mu)((T \times T)^{-n}(A \times C) \cap (B \times D)) - (\mu \times \mu)(A \times C)(\mu \times \mu)(B \times D)| = 0.$$  

Hence by Theorem 3.5

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |(\mu \times \mu)((T \times T)^{-n}(A \times C) \cap (B \times D)) - (\mu \times \mu)(A \times C)(\mu \times \mu)(B \times D)| = 0.$$  

Since the measurable rectangles form a semi-algebra that generates $\mathcal{B} \times \mathcal{B}$, Theorem 3.6 gives us that $T \times T$ is weak mixing.

(3) $\Rightarrow$ (2)

If $T \times T$ is weak mixing then it follows from Proposition 3.1 that $T \times T$ is ergodic.

(2) $\Rightarrow$ (1)

Since $T \times T$ is ergodic we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} (\mu \times \mu)((T \times T)^{-i}(A \times X) \cap (B \times X))$$

$$= (\mu \times \mu)(A \times X)(\mu \times \mu)(B \times X)$$

$$= \mu(A)\mu(B).$$

We also have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B)^2 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} (\mu \times \mu)((T \times T)^{-i}(A \times A) \cap (B \times B))$$

$$= (\mu \times \mu)(A \times A)(\mu \times \mu)(B \times B)$$

$$= \mu(A)^2\mu(B)^2.$$  

Thus

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} (\mu(T^{-i}A \cap B) - \mu(A)\mu(B))^2$$
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} (\mu(T^{-i} \cap B)^2 - 2\mu(T^{-i} \cap B)\mu(A)\mu(B) + \mu(A)^2\mu(B)^2) \]
\[ = \mu(A)^2\mu(B)^2 - 2\mu(A)\mu(B)\mu(A)\mu(B) + \mu(A)^2\mu(B)^2 \]
\[ = 0. \]

That is
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \mu(T^{-i} \cap B) - \mu(A)\mu(B) \right|^2 = 0 \]
and hence by Theorem 3.6 \( T \) is weak mixing.

For later results it is useful to consider a condition that investigates, not iterations of \( T \) on points or pairs of sets, but rather iterations of the operator \( U_T \) on pairs of functions in \( L^2(X, B, \mu) \). As can be seen from the Theorem 3.8 in which we consider these properties, the type of conditions that we find are of a similar form to those already discussed.

In order to present Theorem 3.8 (and some of the following results) it is now necessary to note that we will be using \( < \cdot, \cdot > \) to denote the inner product on the space \( L^2(X, B, \mu) \). That is, for all \( f, g \in L^2(X, B, \mu) \) we have
\[ < f, g > = \int_X f(x)\overline{g(x)}d\mu(x). \]

**Theorem 3.8**

Suppose that \( (X, B, \mu) \) is a probability space, \( T : X \to X \) is measure preserving and \( U_T \) is as defined in Definition 3.1. The following are equivalent:

1. \( T \) is weak mixing.
2. For all \( f, g \in L^2(X, B, \mu) \),
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | < U_T^i f, g > - < f, 1 > < 1, g > | = 0. \]
3. For all \( f \in L^2(X, \mathcal{B}, \mu) \),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | \langle U_T^i f, f \rangle - \langle f, 1 \rangle < 1, f > | = 0.
\]

4. For all \( f \in L^2(X, \mathcal{B}, \mu) \),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | \langle U_T^i f, f \rangle - \langle f, 1 \rangle < 1, f > |^2 = 0.
\]

**Proof:**

We prove the result by showing \((2) \Rightarrow (1) \Rightarrow (3) \Rightarrow (2)\) and \((3) \Leftrightarrow (4)\).

\((2) \Rightarrow (1)\)

For any \( A, B \in \mathcal{B} \) we put \( f = \chi_A \) and \( g = \chi_B \). So that we have

\[
\langle U_T^n f, g \rangle = \langle U_T^n \chi_A, \chi_B \rangle = \int U_T^n \chi_A \chi_B d\mu = \int \chi_T^{-n} \chi \chi_B d\mu = \int \chi_T^{-n} \chi \chi_B d\mu = \mu(T^{-n} A \cap B)
\]

and

\[
\langle f, 1 \rangle < 1, g \rangle = \langle \chi_A, 1 \rangle < 1, \chi_B \rangle = \int \chi_A d\mu \int \chi_B d\mu = \mu(A)\mu(B).
\]

Since for all \( f, g \in L^2(X, \mathcal{B}, \mu) \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | \langle U_T^i f, g \rangle - \langle f, 1 \rangle < 1, g > | = 0,
\]

56
we have

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| \]

\[ = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T \chi_A, \chi_B > - \chi_A, 1 > < 1, \chi_B > | \]

\[ = 0, \]

so that \( T \) is weak mixing.

\((1) \Rightarrow (3)\)

Since \( T \) is weak mixing, for any \( A, B \in B \) we have, by equations (6) and (7) that

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T \chi_A, \chi_B > - \chi_A, 1 > < 1, \chi_B > | \]

\[ = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| \]

\[ = 0. \]

Then, by fixing \( B \), we have that for any simple function

\[ h = \sum_{i=1}^{\infty} a_i \chi_{E_i}, \]

\[ 0 \leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T h, \chi_B > - h, 1 > < 1, \chi_B > | \]

\[ = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T \sum_{j=1}^{\infty} a_j \chi_{E_j}, \chi_B > - \sum_{j=1}^{\infty} a_j \chi_{E_j}, 1 > < 1, \chi_B > | \]

\[ = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \sum_{j=1}^{\infty} a_j < U^i_T \chi_{E_j}, \chi_B > - \sum_{j=1}^{\infty} a_j \chi_{E_i}, 1 > < 1, \chi_B > \right| \]

\[ = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \sum_{j=1}^{\infty} a_j ( < U^i_T \chi_{E_j}, \chi_B > - \chi_{E_i}, 1 > < 1, \chi_B > ) \right| \]
\[
\leq \sum_{i=j}^{\infty} |a_j| \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \langle U_T^n \chi_{E_j}, \chi_B \rangle - \langle \chi_{E_j}, 1 \rangle < 1, \chi_B \rangle \\
= 0,
\]
and hence
\[
\lim_{n \to \infty} \sum_{i=0}^{n-1} \langle U_T^n h, \chi_B \rangle - \langle h, 1 \rangle < 1, \chi_B \rangle = 0.
\]
Similarly, we can fix \(h\) to get
\[
\lim_{n \to \infty} \sum_{i=0}^{n-1} \langle U_T^n h, h \rangle = \langle h, 1 \rangle < 1, h \rangle
\]
\[
= \lim_{n \to \infty} \sum_{i=0}^{n-1} \sum_{j=1}^{\infty} a_j (\langle U_T^n h, \chi_{E_j} \rangle - \langle h, 1 \rangle < 1, \chi_{E_j} \rangle)
\]
\[
= 0.
\]
Now suppose \(f \in L^2(X, \mathcal{B}, \mu)\), and let \(\varepsilon > 0\). Choose a simple function \(h\) so that \(||f - h||_2 < \varepsilon\). Then for each \(n \in \mathbb{N}\), we use the Schwarz inequality to find
\[
\langle U_T^n f, f \rangle - \langle f, 1 \rangle < 1, f \rangle
\]
\[
\leq \langle U_T^n f, f \rangle - \langle U_T^n h, f \rangle + \langle U_T^n h, f \rangle - \langle U_T^n h, h \rangle
\]
\[
+ \langle U_T^n h, h \rangle - \langle h, 1 \rangle < 1, h \rangle + \langle h, 1 \rangle < 1, h \rangle
\]
\[
\leq \langle U_T^n (f - h), f \rangle + \langle U_T^n h, f - h \rangle
\]
\[
+ \langle U_T^n h, h \rangle - \langle h, 1 \rangle < 1, h \rangle + \langle 1, h \rangle \| < h - f, 1 \rangle
\]
\[
+ \langle f, 1 \rangle \| < 1, h - f \rangle
\]
\[
\leq ||f - h||_2 \|f\|_2 + ||f - h||_2 \|h\|_2 + \langle U_T^n h, h \rangle - \langle h, 1 \rangle < 1, h \rangle
\]
\[
+ \langle h \|_2 \|f - h\|_2 + \|f\|_2 \|h - f\|_2
\]
\[
\leq \varepsilon \|f\|_2 + \varepsilon (||f||_2 + \varepsilon) + \langle U_T^n h, h \rangle - \langle h, 1 \rangle < 1, h \rangle
\]
\[
+ (||f||_2 + \varepsilon) \varepsilon + \varepsilon \|f\|_2.
\]
Hence

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^n f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle| \\
\leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} (\varepsilon \|f\|_2 + \varepsilon (\|f\|_2 + \varepsilon) + |\langle U_T^n h, h \rangle - \langle h, 1 \rangle \langle 1, h \rangle| + (\|f\|_2 + \varepsilon)\varepsilon + \varepsilon \|f\|_2)
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^n h, h \rangle - \langle h, 1 \rangle \langle 1, h \rangle| + \varepsilon \|f\|_2 + \varepsilon (\|f\|_2 + \varepsilon) + (\|f\|_2 + \varepsilon)\varepsilon + \varepsilon \|f\|_2
\]

\[
= \varepsilon \|f\|_2 + \varepsilon (\|f\|_2 + \varepsilon) + (\|f\|_2 + \varepsilon)\varepsilon + \varepsilon \|f\|_2.
\]

Therefore, since \(\varepsilon\) is arbitrarily small, we have that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^n f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle| = 0.
\]

(3) \(\Rightarrow\) (2)

Let \(f \in L^2(X, \mathcal{B}, \mu)\) and let \(\mathcal{H}_f\) denote the smallest closed subspace of \(L^2(X, \mathcal{B}, \mu)\) containing \(f\) and the constant functions and satisfying

\[
U_T \mathcal{H}_f \subset \mathcal{H}_f.
\]

We show that the set

\[
\mathcal{F}_f = \left\{ g \in L^2(X, \mathcal{B}, \mu) : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^n f, g \rangle - \langle f, 1 \rangle \langle 1, g \rangle| = 0 \right\}
\]

is a closed subspace of \(L^2(X, \mathcal{B}, \mu)\) containing \(f\), the constant functions and satisfying

\[
U_T \mathcal{F}_f \subset \mathcal{F}_f.
\]

Suppose \(g_1, g_2 \in \mathcal{F}_f\), then

\[
0 \leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\langle U_T^n f, g_1 + g_2 \rangle - \langle f, 1 \rangle \langle 1, g_1 + g_2 \rangle|
\]

59
Next suppose that \( g \in F_f \) and \( a \in C \) then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T f, ag > - < f, 1 > < 1, g > |
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\overline{a} < U^i_T f, g > - \overline{a} < f, 1 > < 1, g > |
\]

\[
= |a| \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T f, g > - < f, 1 > < 1, g > |
\]

\[
= 0.
\]

To finish showing that \( F_f \) is closed we show that it is closed. We do this by showing that if \( \{ g_n \} \) is a convergent sequence in \( F_f \) then \( g = \lim_{n \to \infty} g_n \in F_f \).

Let \( \{ g_n \} \) be a convergent sequence in \( F_f \) and let \( g = \lim_{n \to \infty} g_n \). Then for any \( \varepsilon > 0 \) there is an \( n \in \mathbb{N} \) such that \( ||g - g_n||_2 < \varepsilon \) and hence using the Cauchy Schwarz inequality we have

\[
0 \leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T f, g > - < f, 1 > < 1, g > |
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T f, g_n > + < U^i_T f, g - g_n >
\]

\[
- < f, 1 > < 1, g_n > - < f, 1 > < 1, g - g_n > |
\]

\[
\leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T f, g_n > - < f, 1 > < 1, g_n > |
\]

\[
+ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T f, g - g_n > - < f, 1 > < 1, g - g_n > |
\]

\[
\leq 0 + \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U^i_T f, g - g_n > | + | < f, 1 > < 1, g - g_n > |
\]
\[
\begin{align*}
\leq & \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \|U_T^i f\|_2 ||g - g_n||_2 + ||f||_2 \|g - g_n||_2 \\
< & \varepsilon \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 2||f||_2 \\
= & 2\varepsilon ||f||_2.
\end{align*}
\]
Since this is true for all \( \varepsilon > 0 \) we have that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-1} |< U_T^i f, g > - < f, 1 > | < 1 > | = 0
\]
and hence that \( g \in \mathcal{F}_f \) from which we have that \( \mathcal{F}_f \) is a closed subspace of \( L^2(X, \mathcal{B}, \mu) \). Thus \( \mathcal{F}_f \) is closed. As we are assuming that (3) holds, \( f \in \mathcal{F}_f \).

If \( g \) is a constant, say \( c \) then by using that \( T \) is a measure preserving transformation we note that for any \( n \in \mathbb{N} \)
\[
|< U_T^n f, 1 >| = \int U_T^n f d\mu = \int f d\mu = |< f, 1 >|
\]
so that we have
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U_T^i f, g > - < f, 1 > | < 1 > | = 0.
\]
Also, if \( g \in \mathcal{F}_f \), then
\[
0 \leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |< U_T^i f, U_T g > - < f, 1 > | < 1 > | U_T g > |
\]

61
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<U_T^i f, U_T g > - < f, 1 > < 1, U_T g>| \\
+ \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-1} |<U_T^i f, U_T g > - < f, 1 > < 1, U_T g>| \\
= 0 + \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n-1} |<U_T^{i-1} f, g > - < f, 1 > < 1, g>| \\
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-2} |<U_T^i f, g > - < f, 1 > < 1, g>| \\
\leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<U_T^i f, g > - < f, 1 > < 1, g>| \\
= 0,
\]
so that \( U_T g \in \mathcal{F}_f \) and hence \( U_T \mathcal{F}_f \subset \mathcal{F}_f \).

We therefore have that \( \mathcal{H}_f \subset \mathcal{F}_f \). Now suppose \( g \in \mathcal{H}_f^\perp \). Since \( f \in \mathcal{H}_f \), and \( U_T \mathcal{H}_f \subset \mathcal{H}_f \) we have that

\[
U_T f \in \mathcal{H}_f, U_T(U_T f) = U_T^2 f \in \mathcal{H}_f, ..., U_T^n f \in \mathcal{H}_f
\]
for all \( n \in \mathbb{N} \). Thus \( < U_T^n f, g > = 0 \) for all \( n \in \mathbb{N} \). Also, as the constant functions are in \( \mathcal{H}_f \) we have that \( < 1, g > = 0 \). Therefore

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<U_T^i f, g > - < f, 1 > < 1, g>| = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |0 - < f, 1 > 0| \\
= 0.
\]
Thus \( \mathcal{H}_f^\perp \subset \mathcal{F}_f \) and hence \( \mathcal{F}_f = L^2(X, \mathcal{B}, \mu) \).

(3) \( \iff \) (4)

By setting \( a_n = < U_T^n f, f > - < f, 1 > < 1, f > \) this equivalence follows directly from Theorem 3.5.

\( \diamond \)
3.3 Mixing and Continuous Spectrum

To support our research there is one final ergodic theorem that we need to consider. In order to prove this result we will need the spectral theorem for unitary operators. This theorem is not at all trivial and requires a lot of preliminary work. As the theorem is not in the main line of the argument of this thesis, we will not provide a proof. We will however provide definitions and statements of results sufficient to understand the statement of the spectral theorem. The notes at the conclusion of this section provide references for the interested reader on where to find further reading on the spectral theorems.

We will first need to define a unitary operator. The definition of a unitary operator rests on the definition of other types of operator and so we also state these definitions.

**Definition 3.6**

Let $\mathcal{H}$ be a Hilbert space and $T$ be an operator on $\mathcal{H}$. Then the **adjoint** $T^*$ of $T$ is an operator on $\mathcal{H}$ satisfying

$$<Tx, y> = <x, T^*y>$$

for each pair of vectors $x$ and $y$ in $\mathcal{H}$.

Note, it can be shown that the adjoint, $T^*$ exists for any bounded linear operator $T$.

**Definition 3.7**

An operator $T$ on a Hilbert space is called **self adjoint** if $T = T^*$.
Definition 3.8
An operator $T$ on a Hilbert space is called unitary if

$$TT^* = T^*T = 1.$$  

While the concept of a self adjoint operator is not necessary to define a unitary operator we need the definition of a self adjoint operator in order to define the other necessary concepts and objects in stating the spectral theorem. These are, the concepts of a spectrum, a spectral measure and a spectral integral.

Definition 3.9
Let $X$ be a Hilbert space and $T : X \to X$ be a bounded linear operator. The spectrum of $T$, $\sigma(T)$ is the set of complex valued scalars $\lambda$ such that $T - \lambda I$ is not invertible, where $I$ is the identity operator.

Definition 3.10
Let $(X, \mathcal{B})$ be a measurable space and $\mathcal{H}$ be a Hilbert space. A spectral measure on $X$ is a function $E$ whose domain is $\mathcal{B}$ and whose values are idempotent, (i.e. $E^2 = E$) self-adjoint operators (projections) on $\mathcal{H}$ such that $E(X) = 1$ and

$$E \left( \bigcup_{n=1}^{\infty} M_n \right) = \sum_{n=1}^{\infty} E(M_n)$$  

whenever $\{M_n\}$ is a disjoint sequence of sets in $\mathcal{B}$.

A particular property of such an $E$, of which we make note, is listed in the following proposition.

Proposition 3.2
Let $(X, \mathcal{B})$ be a measurable space and $E$ be a spectral measure with domain $\mathcal{B}$. 

64
Then for each pair of vectors \( x, y \), the complex valued set function \( \mu \) defined for every \( M \in \mathcal{B} \) by \( \mu(M) = \langle E(M)x, y \rangle \) is a measure.

The above proposition now allows us to, in a sense, integrate over projections to get an operator.

**Proposition 3.3**

If \( E \) is a spectral measure and if \( f \) is a complex-valued, bounded, measurable function on \( X \), then there exists a unique operator \( A \) such that

\[
\langle Ax, y \rangle = \int f(\lambda)d\langle E(\lambda)x, y \rangle
\]

for each pair of vectors \( x, y \).

The dependence of \( A \) on \( E \) and \( f \) is denoted by writing

\[
A = \int fdE.
\]

Such an expression (resultant from the process of integrating to get an operator for a value) is known as a spectral integral.

The Spectral Theorem for unitary operators is a converse for Proposition 3.3. That is, given a unitary operator \( A \), we wish to show that there is a unique spectral measure \( E \) such that \( A = \int \lambda dE(\lambda) \).

**Theorem 3.9 (The Spectral Theorem for Unitary Operators)**

If \( A \) is unitary operator on \( \mathcal{H} \), then there exists a (necessarily real and unique) compact, complex spectral measure \( E \) such that

\[
A = \int \lambda dE(\lambda).
\]
That is, we have that for any unitary operator $A$ on a Hilbert space there is a unique spectral measure $E$ such that

$$\langle Ax, y \rangle = \int \lambda d \langle E(\lambda)x, y \rangle$$

for each pair of vectors $x$ and $y$. Since $\langle E(M)x, y \rangle$ is a Borel measure, we now have that for any unitary operator $A$ and each vector $x$ there is a unique Borel measure $\mu_x$ such that

$$\langle A^n x, x \rangle = \int \lambda^n d\mu_x$$

for each $n \in \mathbb{N}$. To appropriately apply the spectral theorem we need one more property.

**Definition 3.11**

Let $(X, B, \mu)$ be a probability space, then $\mu$ is non-atomic if for each set $B \in B$ such that $\mu(B) > 0$ there is an $A \in B$ such that $A \subset B$ and $0 < \mu(A) < \mu(B)$.

(Note that a non-atomic measure is sometimes defined by saying that for each $x \in X$, $\mu(x) = 0$; however this definition and Definition 3.11 are equivalent.)

**Proposition 3.4**

If $T$ is an invertible measure preserving transformation on a measure space $(X, B, \mu)$ then $U_T$ is unitary and if we denote by $\mu_f$, the unique Borel measure corresponding to $f \in L^2(X, B, \mu)$ such that

$$\langle U_T f, f \rangle = \int \lambda f d\mu_f(\lambda),$$

then if $T$ has continuous spectrum and $\langle f, 1 \rangle = 0$, $\mu_f$ is non-atomic.
We are now finally able to conclude the chapter with the proof of a theorem important to this work. It shows that a transformation is weak mixing if and only if it has a continuous spectrum.

**Definition 3.12**

Let $T$ be a measure preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. We say a complex number, $\lambda$, is an **eigenvalue** for $T$ if there is an $f \in L^2(X, \mathcal{B}, \mu)$, $f \neq 0$ such that $U_T f = \lambda f$. Such an $f$ is called an **eigenfunction** corresponding to $\lambda$.

**Definition 3.13**

We will say that a measure preserving transformation $T$ on a probability space $(X, \mathcal{B}, \mu)$ has **continuous spectrum** if 1 is the only eigenvalue for $T$ and the only eigenfunctions are the constants.

The name continuous spectrum is derived from a corresponding definition for operators which is not immediately apparent in this context. We note that since

$$||f||^2 = ||U_T f||^2 = \langle U_T f, U_T f \rangle = \langle \lambda f, \lambda f \rangle = |\lambda|^2 ||f||^2$$

we must have that $|\lambda| = 1$ whenever $\lambda$ is an eigenvalue for $T$.

For the final theorem in this chapter we need one final preliminary result. A measure theoretic result on a property of non-atomic measures. It is presented in the form of two propositions and a lemma. The two propositions are standard results (see, for example, Halmos [14]) and so are presented without proof. The Lemma is the relevant result to this discussion and so we present a proof.
Proposition 3.5
Let $(X, B, \mu)$ be a probability space with a non-atomic measure. Then, for any $B \in B$ such that $\mu(B) > 0$ and any $\varepsilon > 0$ there are subsets $B_1$ and $B_2$ of $B$ such that $B_1, B_2 \in B$, $0 \neq \mu(B_1) < \varepsilon$ and $\mu(B_2) > \mu(B) - \varepsilon$.

Proposition 3.6
Let $(X, B, \mu)$ be a probability space with a non-atomic measure and let $B \in B$ be such that $\mu(B) > 0$. Then for any $a \in (0, \mu(B))$ there is an $A \in B$, $A \subset B$ such that $\mu(A) = a$.

Lemma 3.4
Let $(X, B, \mu)$ be a probability space where $\mu$ is a nonatomic measure. Then if $D = \{(x, x) : x \in X\} \subset X \times X$, $(\mu \times \mu)(D) = 0$.

Proof:
Let $n \in \mathbb{N}$. Then, by Proposition 3.6, there is an $E_1 \in B$ such that $\mu(E_1) = 1/n$. Also $\mu(X - E_1) = 1 - 1/n$, (and clearly $X - E_1 \in B$) so that by Proposition 3.6 we can find an $E_2 \subset X - E_1$ such that $\mu(E_2) = 1/n$. Then inductively we can select

$$E_i \subset X - \bigcup_{n=1}^{i-1} E_n$$

such that $E_i \in B$, $\mu(E_i) = 1/n$ for each $i \in \{2, 3, ..., n - 1\}$.

Note then that the sequence $\{E_i\}_{i=1}^{n-1}$ is a disjoint sequence of sets in $B$ and that $\mu(X - \bigcup_{i=1}^{n-1} E_i) = 1/n$. Set $E_n = X - \bigcup_{i=1}^{n-1} E_i \in B$ so that we have a disjoint sequence of sets $\{E_i\}_{i=1}^{n}$ in $B$ such that $X = \bigcup_{i=1}^{n} E_i$ and that $\mu(E_i) = 1/n$ for each $n$.

Note that $\{E_i \times E_i\}_{i=1}^{n}$ is a disjoint sequence of subsets of $X \times X$, that
$D \subset \bigcup_{i=1}^{n} E_i \times E_i$ and that for each $i \in \{1, 2, ..., n\}$,

$$(\mu \times \mu)(E_i \times E_i) = \mu(E_i)\mu(E_i) = \frac{1}{n^2}.$$  

Hence

$$(\mu \times \mu)_*(D) \leq \sum_{i=1}^{n} (\mu \times \mu)(E_i \times E_i)$$

$$= \sum_{i=1}^{n} \frac{1}{n^2}$$

$$= \frac{1}{n}.$$  

Since this is true for each $n \in \mathbb{N}$ we have that $(\mu \times \mu)_*(D) \leq 0$. Also, it is obvious that $(\mu \times \mu)_*(D) \geq 0$ and hence $(\mu \times \mu)_*(D) = 0$. Further as $D$ is in the Borel sets (By taking $\cap_{n \in \mathbb{N}} \bigcup_{i=1}^{n} (E_n \times E_n)$) $D \in \mathcal{B}$ and so $(\mu \times \mu)(D) = (\mu \times \mu)_*(D) = 0$.  

\[\Diamond\]

We can now prove the final theorem in the chapter which relates the property of weak mixing to that of having a continuous spectrum.

**Theorem 3.10**

An invertible measure preserving transformation, $T$, of a probability space $(X, \mathcal{B}, \mu)$ is weak mixing if and only if it has continuous spectrum.

**Proof:**

Suppose that $T$ is weak mixing and let $U_T f = \lambda f$, $f \in L^2(X, \mathcal{B}, \mu)$.

If $\lambda \neq 1$ then as

$$0 = U_T f - \lambda f,$$

$$0 = \int U_T f - \lambda f \, d\mu$$

$$= \int U_T f \, d\mu - \int \lambda f \, d\mu$$

$$= \int (1 - \lambda) f \, d\mu$$
and so we must have that \( <f, 1> = \int f\,d\mu = 0 \). By Theorem 3.8

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<U_t^i f, f> - <f, 1><1, f>| = 0
\]

so that we now have

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<U_T^i f, f>| = 0
\]

and hence

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<\lambda^i f, f>| = 0.
\]

Now

\[
|<\lambda^i f, f>| = \left| \int \lambda^i f \overline{f} \,d\mu \right|
\]

\[
= |\lambda^i| \left| \int f \overline{f} \,d\mu \right|
\]

\[
= |\lambda^i| \left| \int \overline{f} f \,d\mu \right|
\]

\[
= \left| \int f \overline{f} \,d\mu \right|
\]

\[
= |<f, f>|,
\]

and hence

\[
0 \leq ||f||^2
\]

\[
= <f, f>
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<f, f>| 
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |<\lambda^i f, f>| 
\]

\[
= 0,
\]
so that $f = 0 \mu$-a.e.

If $\lambda = 1$ then we show that $f$ must be constant almost everywhere using the ergodicity of $T$. Suppose that $f$ is not constant almost everywhere then there is a $B \in \mathcal{B}$ and an $A \subset \mathbb{C}$ such that $\mu(B) > 0$, $\mu(B^c) > 0$, $f(x) \in A$ for all $x \in B$ and $f(x) \notin A$ for all $x \in B^c$. Then, as $T$ is ergodic $\mu(T(B) \cap B^c) > 0$. As we have that $U_T f = f$, we then have that for all $x \in (T(B) \cap B^c)$ $f(x) \in A$ and $f(x) \notin A$. Thus $f$ is constant almost everywhere.

As whenever $T$ is weakly mixing $\lambda = 1$ and the only eigenfunctions are the constants, if $T$ is weak mixing then $T$ has continuous spectrum.

Now suppose that $T$ has continuous spectrum. We show that if $f \in L^2(X, \mathcal{B}, \mu)$ then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | < U_T^i f, f > - < f, 1 > < 1, f > |^2 = 0.$$ 

This is clear if $f$ is a constant, say $z$, as in this case

$$< U_T^i f, f > = < f, f >$$
$$= \int f \overline{f} d\mu$$
$$= zz$$
$$= \int f d\mu \int \overline{f} d\mu$$
$$= < f, 1 > < 1, f >,$$

so that $| < U_T^i f, f > - < f, 1 > < 1, f > | = 0$ for each $i \in \mathbb{N} \cup \{0\}$. For a general $f \in L^2(X, \mathcal{B}, \mu)$ we can write $f = I + g$ where $I = \int f d\mu$ and $g$ is such that $< g, 1 > = 0$ and hence $< 1, g > = 0$ and $< U_T^i g, 1 > = 0$ for each $i \in \mathbb{N}$. Using this we note that for each $i \in \mathbb{N}$

$$| < U_T^i f, f > - < f, 1 > < 1, f > | = | < U_T^i I, I > + < U_T^i I, g > + < U_T^i g, I >$$
\[ + \langle U^i_T g, g \rangle - \langle I, 1 \rangle \langle 1, I \rangle \\
- \langle I, 1 \rangle \langle 1, g \rangle - \langle g, 1 \rangle \langle 1, I \rangle \\
- \langle g, 1 \rangle \langle 1, g \rangle \]
\[ = \langle I, I \rangle + I \langle 1, g \rangle + I \langle U^i_T g, 1 \rangle \\
+ \langle U^i_T g, g \rangle - I^2 - 0 - 0 - 0 \]
\[ = I^2 + 0 + 0 + \langle U^i_T g, g \rangle - I^2 \]
\[ = \langle U^i_T g, g \rangle , \]

so that
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | \langle U^i_T g, g \rangle |^2 = 0 \]
implies that
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | \langle U^i_T f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle |^2 = 0. \]

Hence in general to prove the result it is sufficient to prove that \(< f, 1 > = 0 \)
implies that
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} | \langle U^i_T f, f \rangle |^2 = 0. \]

By the Spectral Theorem there exists a unique non-atomic Borel measure \( \mu_f \)
such that
\[ \langle U^i_T f, f \rangle = \int \lambda^i d\mu_f(\lambda) \]
so that to prove the result it suffices to prove that if \( \mu_f \) is a non-atomic
measure on \( \sigma(U_T) \) then
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \int \lambda^i d\mu_f(\lambda) \right|^2 = 0. \]

Recalling that for all \( \lambda \in \sigma(U_T), |\lambda| = 1 \), we can now use Fubini’s Theorem
to obtain, for any \( n \in \mathbb{N} \),
\[ \frac{1}{n} \sum_{i=0}^{n-1} \left| \int \lambda^i d\mu_f(\lambda) \right|^2 = \frac{1}{n} \sum_{i=0}^{n-1} \left( \int \lambda^i d\mu_f(\lambda) \int \lambda^{-i} d\mu_f(\lambda) \right) \]

72
\[
\begin{align*}
&= \frac{1}{n} \sum_{i=0}^{n-1} \left( \int \lambda^i d\mu_f(\lambda) \int \tau^{-i} d\mu_f(\tau) \right) \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \left( \int \int \lambda^i d\mu_f(\lambda) \tau^{-i} d\mu_f(\tau) \right) \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \left( \int \int \lambda^i \tau^{-i} d\mu_f(\lambda) d\mu_f(\tau) \right) \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \int \int_{\sigma(U_T) \times \sigma(U_T)} (\lambda \tau)^i d(\mu_f \times \mu_f)(\lambda, \tau) \\
&= \int \int_{\sigma(U_T) \times \sigma(U_T)} \left( \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i \right) d(\mu_f \times \mu_f)(\lambda, \tau).
\end{align*}
\]

If \((\lambda, \tau)\) is not in the diagonal of \(\sigma(U_T) \times \sigma(U_T),\) \(D,)\) then
\[
\begin{align*}
\left| \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i \right| &= \left| \frac{1}{n} \left( 1 - (\lambda \tau)^n \right) \right| \\
&\leq \left| \frac{2}{n(1 - (\lambda \tau))} \right|
\end{align*}
\]
so that
\[
\lim_{n \to \infty} \left| \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i \right| = 0
\]
and hence
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i = 0.
\]
Since \(\mu_f \times \mu_f\) is non-atomic, \((\mu_f \times \mu_f)(D) = 0\) and hence
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i = 0 \quad (\mu_f \times \mu_f) - \text{a.e.}
\]
Since
\[
\left| \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i \right| \leq 1
\]
we can apply the Bounded Convergence Theorem to get that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \int \lambda^i d\mu_f(\lambda) \right|^2 = \lim_{n \to \infty} \int \int_{\sigma(U_T) \times \sigma(U_T)} \left( \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i \right) d(\mu_f \times \mu_f)(\lambda, \tau)
\]

73
\[
\begin{align*}
\int \int_{\sigma(U_T) \times \sigma(U_T)} & \lim_{n \to \infty} \left( \frac{1}{n} \sum_{i=0}^{n-1} (\lambda \tau)^i \right) d(\mu_f \times \mu_f)(\lambda, \tau) \\
= & \int \int_{\sigma(U_T) \times \sigma(U_T)} 0 d(\mu_f \times \mu_f)(\lambda, \tau) \\
= & 0.
\end{align*}
\]

3.4 Notes

Lemma 3.1 is due to Nillsen [32]. Apart from Lemma 3.2, due to Rudin [38], the discussion in up to Theorem 3.8 follows Walters [46] which is an excellent source for introductory Ergodic Theory. Theorem 3.2 (Birkoff’s Ergodic Theorem) is due to Birkoff [5]. For the discussion on the Spectral Theorem, our source was Halmos [15], although Rudin [38] also presents a good discussion of the Spectral Theorem. Further reading on Ergodic Theory can be found in Peterson [35] and an interesting approach (using towers, as we discuss later) to the proofs given for the ergodic results can be found in Friedman [13]. Non-atomic measures are a standard part of measure theory and further reading on Propositions 3.5 and 3.6 can be found in Halmos [14].
4 Rotations and Orbits

We have discussed what it means to generate the Carathéodory definition of measurable sets, why we are interested in generating it and have established some preliminary results. We will now pursue further the interesting results that have come about in pursuing exactly what conditions allow the generation of Carathéodory’s definition.

As was mentioned in the introduction we will present the results in approximately chronological order with respect to when we established the results. In this way, the motivation for each result, the path to the most general result and the origin of the ideas behind the proofs should be made clear. Fortunately, this general method of presentation also allows the clear and separate development of each of the important ideas necessary in this research.

We begin with the first result on generating Carathéodory’s definition which was given by Nillsen [32]. As has been discussed, we need to be able to make a comparison between arbitrary sets and subintervals by examining the proportion of each subinterval made up of the arbitrary set in the sense of outer measure. This requires two things, finding a tool of comparison, and finding a way of moving the set throughout the space to make the necessary comparisons. The importance of Nillsen’s work is three fold. Firstly, it established that transformations generating Carathéodory’s definition do indeed exist. Secondly, the tool for comparison is developed. Based loosely on the common measure theoretic idea of points of density, the idea of an interval of density is introduced. This allows us to find a part of any arbitrary set of positive outer measure that is ‘approximately’ an interval, that is, a part that is sufficiently approximated by an interval. Thirdly, it established the use of ‘rotations’ of intervals (and more importantly of intervals of density)
by a transformation. These move the intervals of density through a space to be compared to other intervals in the space. Provided that the transformation is ergodic it will also allow our necessary comparisons with sufficiently many subintervals of the space. The results presented in [32] were also the results that led to the conjecture that the properties of a transformation that generates the Carathéodory definition were closely related to the mixing properties of the transformation. Both of the important tools developed in [32], intervals of density and rotations, will, of course, be formally defined at the appropriate time.

The transformations used in [32] were the irrational rotations on the unit circle. To briefly describe these rotations; let $\mathbf{T}$ be the set $\{z \in \mathbb{C} : |z| = 1\}$, then a transformation $\rho$ is a rotation on $\mathbf{T}$ if there is an $\nu \in \mathbf{T}$ such that $\rho(z) = \nu z$ for all $z \in \mathbf{T}$. In this setting the use of the term ‘rotation’, which will be formally defined later, has a clear interpretation. The use of the term ‘rotation’ in association with other transformations does not always have so intuitively clear a meaning. However, we will show that in some sense any transformation can be considered to act as a ‘rotation’. The idea of looking at relationships with mixing properties led us to consider a weakly but not strongly mixing transformation provided by Kakutani in [21]. This transformation is one for which the idea of ‘rotation’ is not entirely clear. However, our investigation of this transformation showed that something similar could be used. The new interpretation is the interpretation in which all transformations can, analogously, be considered to ‘rotate’ sets. The concept, to maintain consistency, is still called a rotation, but to make this idea work we needed to consider more closely the orbits, not only of points, but also of intervals, and of sets in general. Section 4.2 looks at the Kakutani example and how it uses orbits. As with intervals of density and rotations, orbits will
be defined later.

4.1 Rotations

We first make a more formal construction of the transformation that is being considered here. There are two ways of considering this type of transformation. Firstly, we will describe the one from which its name arises. Let \( T \) be the set \( \{z \in \mathbb{C} : |z| = 1\} \). A rotation, \( \rho \) on \( T \) is a transformation for which there exists \( z \in T \) such that

\[
\rho(\omega) = z\omega, \text{ for all } \omega \in T.
\]

To allow identification of which rotation we are using, we will denote the rotation \( \rho \) as \( \rho_z \) whenever \( \rho(\omega) = z\omega \), for all \( \omega \in \mathbb{T} \) for some \( z \in \mathbb{T} \).

Geometrically, a rotation can be interpreted as rotating the points in \( \mathbb{T} \) anticlockwise through an angle of \( 2\pi \theta \) where \( z = e^{2\pi i \theta} \). We then go on to define a rational rotation as a rotation for which \( z \) is a root of unity and an irrational rotation as a rotation for which \( z \) is not a root of unity. A problem that immediately arises is that our, at this stage, definition of generating the Carathéodory definition relies on the space having intervals and so, at this point, must be defined on an interval in \( \mathbb{R}^1 \). To deal with this problem we note that each \( w \in \mathbb{T} \) can be written as \( w = e^{2\pi i \alpha} \) for some unique \( \alpha \in [0, 1) \). We then define a mapping \( h : \mathbb{T} \to [0, 1) \) by \( h(w) = h(e^{2\pi i \alpha}) = \alpha \). Denoting \([0, 1)\) by \( \mathbb{I} \) we can then consider the transformation equivalent to \( \rho_z \), \( P_z \) on \( \mathbb{I} \) defined by

\[
P_z(x) = h(\rho_z(h^{-1}(x)))
\]

\[
= h(\rho_z(e^{2\pi ix}))
\]

\[
= h(e^{2\pi i(x+\theta)})
\]

\[
= (x + \theta)(\text{mod } 1).
\]
for all $x \in \mathbb{I}$.

It has been mentioned that $P_z$ is defined on $[0,1)$. The dynamical system we will be considering for any given $P_z$ is $([0,1), \mathcal{B}, \mu, P_z)$ where $\mathcal{B}$ is the $\sigma$-algebra of Borel sets and $\mu$ is Lebesgue measure. It is with this transformation that we wish to generate the Carathéodory definition of measurable sets. It is clear that $P_z$ is one to one and onto. Also, as $P_z$ and $P_z^{-1}$ are simply translations in $\mathbb{R}$ modulo 1 and as Lebesgue measure is translation invariant, $P_z$ must be measure preserving (and importantly, as we find out later, outer measure preserving). In order to show that $P_z$ generates Carathéodory's definition we must show that the transformation satisfies definition 2.5. That is we must prove a theorem worded similarly to the following.

Let $z$ be an element of $\mathbb{I}$ which is irrational, and let $A$ be a $P_z$-invariant subset of $X$. Suppose that there is a $\theta = \theta(A) < 2$ such that for all subintervals $J$ of $\mathbb{I}$,

$$\mu_*(A \cap J) + \mu_*(A^c \cap J) \leq \theta \mu(J).$$

Then either $A$ or $A^c$ is a set of outer measure zero.

In fact the above result will be the final theorem in this section. It is immediately obvious that finding a particular interval that is made up 'mostly' of $A$ in $\mathbb{T}$ will more readily allow us to use the condition regarding the make up of intervals in terms of $A$ and its complement $A^c$. We first define an interval of density and then describe how it is used.
Definition 4.1
Let $B_0$ denote the σ-algebra of Borel subsets of $\mathbb{R}$. Let $(X, \mathcal{B}, \mu)$ be a measure space where $X \subseteq \mathbb{R}$,

\[ \mathcal{B} = \{ B \cap X : B \in B_0 \}, \]

and $\mu$ is Lesbesgue measure restricted to $X$. Let $B \subseteq X$ have positive outer measure and let $\varepsilon > 0$. Then an interval $J \subseteq X$ is called an interval of density to within $\varepsilon$ for $B$ if

\[ \mu^*(B \cap J) > (1 - \varepsilon) \mu(J). \]

The way we make use of this concept is as follows. For the case where $X = \mathbb{I}$, we first show that any set of positive outer measure in $\mathbb{R}$ must have intervals of density, then we show that provided that the intervals are small enough, these intervals may be chosen to be the measure of our choice, provided that the measure is sufficiently small. From this it follows that if $B$ and $B^c$ both have positive outer measure then we must be able to find, for any $\varepsilon > 0$, two intervals of density to within $\varepsilon$ of the same length; one for $B$ and one for $B^c$. Then we use iterations of the transformation and the transformation invariance of $B$ to rotate one interval of density, at least approximately, onto the other. For $\varepsilon$ small enough this will contradict the condition given on intervals for the interval of density of $B^c$ which gives the result. The fundamental importance of being able to rotate the intervals is obvious and so we now give a formal definition of a rotation.

Definition 4.2
Let $(X, \mathcal{B}, \mu, T)$ be a dynamical system, with $T$ a measure preserving transformation and let $B \subseteq X$. To rotate $B$ by $T$ through $X$ is to iteratively apply $T^{-1}$ to the set $B$ $n$ times, obtaining the set $T^{-n}B$ for $n \in \mathbb{N}$ of the same outer measure as $B$. To rotate $B$ onto $A$ to within $\varepsilon$ for some $\varepsilon > 0$
and some subset \( A \) of \( X \) of equal measure to \( B \) is to rotate \( B \) through \( X \) by \( T \) and obtain a set \( T^{-1}B \) such that \( \mu_*(T^{-1}B \Delta A) \leq \varepsilon \). To rotate \( B \) onto \( A \) is to rotate \( B \) onto \( A \) to within 0. That is \( \mu_*(T^{-1}B \Delta A) = 0 \).

We now go about proving that the heuristic argument just given does indeed work. First we prove that there are intervals of density and that they can be chosen to be any length provided the length is sufficiently small. The proofs will be given in the form of two lemmas, Lemmas 4.1 and 4.2 which follow Lemmas 4.1 and 4.2, in [32]. In the following results \( T, \mathbb{I}, \rho_z \) and \( P_z \) will be the complex unit circle, the unit interval and the two forms of the irrational rotation transformations described above.

**Lemma 4.1**

Let \( A \subset \mathbb{I} \), let \( \mu_*(A) > 0 \) and let \( \varepsilon > 0 \). Then there is a non-empty interval \( J \) of \( \mathbb{I} \) such that

\[
\mu_*(A \cap J) > (1 - \varepsilon)\mu(J).
\]

**Proof:**

We prove the result by contradiction, so, assume that the result is false. Then for any non-empty interval \( W \) of \( \mathbb{I} \)

\[
\mu_*(A \cap W) \leq (1 - \varepsilon)\mu(W).
\]

Now, by the definition of \( \mu_*(A) \), for any \( \delta > 0 \) there is a sequence of intervals \( \{W_n\} \) in \( \mathbb{I} \) such that

\[
A \subset \bigcup_{n=1}^{\infty} W_n \quad \text{and} \quad \sum_{n=1}^{\infty} \mu(W_n) < \mu_*(A) + \delta.
\]

Then, we note that

\[
A \subset \bigcup_{n=1}^{\infty} (A \cap W_n),
\]

80
so that, using the countable subadditivity of the outer measure and the fact that \( \mu_*(A \cap W) \leq (1 - \varepsilon)\mu(W) \) for each interval \( W \subset \mathbb{I} \), we have

\[
\mu_*(A) \leq \sum_{n=1}^{\infty} \mu_*(A \cap W_n) \\
\leq (1 - \varepsilon) \sum_{n=1}^{\infty} \mu_*(W_n) \\
= (1 - \varepsilon) \sum_{n=1}^{\infty} \mu(W_n) \\
\leq (1 - \varepsilon)(\mu_*(A) + \delta).
\]

Hence

\[
\mu_*(A) \leq \frac{\delta(1 - \varepsilon)}{\varepsilon}.
\]

Since this is true for any \( \delta > 0 \), we have that \( \mu_*(A) = 0 \). This contradicts the fact that \( \mu_*(A) > 0 \) so that there must exist a non-empty interval, say \( J \), such that

\[
\mu_*(A \cap J) > (1 - \varepsilon)\mu(J).
\]

Lemma 4.2

Let \( A \subseteq \mathbb{I} \), let \( \mu_*(A) > 0 \) and let \( \varepsilon > 0 \). Then there is a number \( \theta > 0 \) which has the following property: if \( \eta \in (0, \theta) \), there is an interval \( J \) in \( \mathbb{I} \) such that

\[
\mu(J) = \eta \text{ and } \mu_*(A \cap J) > (1 - \varepsilon)\mu(J).
\]

Proof:

If \( \varepsilon \geq 1 \) then \( 1 - \varepsilon \leq 0 \) and in this case the result is obviously true as Lebesgue measure is a non-negative measure. It may therefore be assumed that \( 0 < \varepsilon < 1 \). By Lemma 4.1, there must exist an interval \( J_1 \subset \mathbb{I} \) such that

\[
\mu_*(A \cap J_1) > \left(1 - \frac{\varepsilon}{2}\right)\mu(J_1).
\]

81
Then the $\theta$ that we will use is

$$\theta = \frac{\mu(J_1)}{2}.$$ 

Note that in this case we have

$$2\theta = \mu(J_1) > 0.$$ 

Now let $\eta$ be any positive number with $\eta < \theta$. Then

$$0 < \eta < \theta < 2\theta = \mu(J_1).$$ 

Hence we can find $m \in \mathbb{N}$ and $r \in [0, \eta)$ such that

$$\mu(J_1) = m\eta + r.$$ 

We now note that we can split $J_1$ up into consecutive disjoint intervals $Z_1, Z_2, \ldots, Z_m, Z$ in $I$ such that

$$\mu(Z_i) = \eta \text{ for each } i \in \{1, 2, \ldots, m\}, \quad \mu(Z) = r$$

and

$$J_1 = \left( \bigcup_{j=1}^{m} Z_j \right) \cup Z.$$

Now we show that at least one of these $Z_i$'s must be an interval of density to within $\varepsilon$. Suppose that there is no such $Z_i$, so that for each $i \in \{1, 2, \ldots, m\}$

$$\mu_*(A \cap Z_i) < (1 - \varepsilon)\eta.$$ 

Then we have

$$\mu(J_1) \left( 1 - \frac{\varepsilon}{2} \right) < \mu_*(A \cap J_1)$$

$$= \mu_*(A \cap \left( (\bigcup_{j=1}^{m} Z_j) \cup Z \right))$$

$$= \mu_*( (A \cap Z) \cup \left( \bigcup_{j=1}^{m} (A \cap Z_j) \right))$$

$$\leq \mu_*(A \cap Z) + \sum_{j=1}^{m} \mu_*(A \cap Z_j)$$
\[ \mu(Z) + \sum_{j=1}^{m} (1 - \varepsilon)\eta \]
\[ = (1 - \varepsilon) \left( \mu(Z) + \sum_{j=1}^{m} \eta \right) + \varepsilon \mu(Z) \]
\[ = (1 - \varepsilon) \left( \mu(Z) + \sum_{j=1}^{m} \mu(Z_j) \right) + \varepsilon \mu(Z) \]
\[ = (1 - \varepsilon) \mu(Z \cup \bigcup_{j=1}^{m} Z_j) + \varepsilon \mu(Z) \]
\[ = (1 - \varepsilon) \mu(J_1) + \varepsilon \mu(Z). \]

This inequality implies that
\[
\mu(J_1) \leq 2\mu(Z) \\
= 2r \\
< 2\eta \\
< 2\theta \\
= \mu(J_1).
\]

Thus the assumption that \( \mu_*(A \cap Z_j) < (1 - \varepsilon)\eta \) for each \( j \in \{1, 2, \ldots, m\} \) cannot hold and thus there is an \( Z_j \) such that
\[
\mu_*(A \cap Z_j) \geq (1 - \varepsilon)\eta
\]
for some \( j \in \{1, 2, \ldots, m\} \). That is, there is an interval of density for \( A \) to within \( \varepsilon \) that has measure \( \eta \) for any \( \eta \in (0, \theta) \) which proves the result. \( \diamond \)

We now wish to prove that we can rotate any interval onto any other interval of the same length to within any \( \varepsilon > 0 \). In order to do this we first need to prove a technical preliminary result. The result is due to Kronecker [26], though we present our own proof to avoid the need for too much new notation and further preliminary results. We do, however, find it necessary
to introduce two new notations. These are used throughout the work and hence are described formally.

Definition 4.3

The **integer part** of a real number $x$, $[x]$, is the integer $a$ such that $a \leq x < a + 1$.

The **fractional part** of $x$, $\{x\}$ is then defined as $\{x\} = x - [x]$.

Note that with this notation, $x \pmod{1} = \{x\}$.

Proposition 4.1

For each $z \in \mathbb{I} \cap \mathbb{Q}^c$, \{nz\(\pmod{1} : n \in \mathbb{N}\} is dense in $\mathbb{I}$.

**Proof:**

We prove this result by showing that if $W$ is an arbitrary non-empty open interval in $\mathbb{I}$ then it contains a point in \{nz\(\pmod{1} : n \in \mathbb{N}\}. Let $W$ be an arbitrary non-empty open interval $(a, b)$. Then there is an $n \in \mathbb{N}$ such that $b - a > 1/n$.

Next, we also know that for all $p, q \in \mathbb{N}$, $p \neq q$, $pz \neq qz \pmod{1}$ as otherwise there is an $s \in \mathbb{N}$ such that

$$|pz - qz| = s \Rightarrow z = \frac{s}{|p - q|} \in \mathbb{Q}.$$

It then follows that \{pz\(\pmod{1} : p \in \mathbb{N}, 0 < p \leq 2n\} is a set of $2n$ distinct points in $(0, 1)$ so that $|pz\(\pmod{1} - qz\(\pmod{1}| < 1$ for each choice of $0 < p, q \leq 2n$. Order \{pz\(\pmod{1} : 0 < p \leq 2n\} in ascending order as points denoted $p_1, p_2, ..., p_{2n}$. If $|pz_i - pz_j| \geq 1/n$ for each $i \neq j$

$$|pz_{2n} - p_1| \geq \frac{2n - 1}{n} \geq 1$$

which is impossible. Therefore

$$\min \{|pz\(\pmod{1} - qz\(\pmod{1}| : p, q \in \mathbb{N}, p \neq q0 < p, q \leq 2n\} < \frac{1}{n}.$$
Let

\[
\min\{|pz(\mod 1) - qz(\mod 1)| : p, q \in \mathbb{N}, p \neq q, 0 < p, q \leq 2n\} = |p_1z(\mod 1) - q_1z(\mod 1)|
\]

for some \(0 < p_1, q_1 \leq 2n\). Then without loss of generality suppose \(p_1 > q_1\).

If \(p_1z(\mod 1) > q_1z(\mod 1)\), we have

\[
\zeta = p_1z(\mod 1) - q_1z(\mod 1) = p_1z - q_1z + [p_1z] - [q_1z]
\]

\[
= (p_1 - q_1)z + [p_1z] - [q_1z]
\]

\[
= ((p_1 - q_1)z) \mod 1 + [(p_1 - q_1)z] + [p_1z] - [q_1z].
\]

As \(\zeta \in [0, 1)\), \([(p_1 - q_1)z] + [p_1z] - [q_1z] = 0\) so that

\[
p_1z(\mod 1) - q_1z(\mod 1) = (p_1z - q_1z)(\mod 1).
\]

Now we know that there is a \(c \in \mathbb{N} \cup \{0\}\) and \(r \in [0, \zeta)\) such that

\[
a = c\zeta + r
\]

and since

\[
b - a > 1/n > \zeta,
\]

\[(c + 1)\zeta \in \mathbb{W}.
\]

Thus

\[(c + 1)((p_1 - q_1)z)(\mod 1) \in \mathbb{W}.
\]

We note that if \(a \in \mathbb{Z}\) then for any \(b \in \mathbb{R}\),

\[
a((b)(\mod 1)) \equiv (ab)(\mod 1)
\]

85
and hence as $c + 1 \in \mathbb{Z}$ we know that

$$(c + 1)(((p_1 - q_1)z)(\text{mod } 1)) \equiv ((c + 1)(p_1 - q_1)z)(\text{mod } 1).$$

As

$$(c + 1)(((p_1 - q_1)z)(\text{mod } 1)) \in W \subset [0, 1)$$

and clearly

$$((c + 1)(p_1 - q_1)z)(\text{mod } 1) \in [0, 1),$$

we must have that

$$(c + 1)(((p_1 - q_1)z)(\text{mod } 1)) = ((c + 1)(p_1 - q_1)z)(\text{mod } 1) \in W$$

and thus $W \cap \{nz(\text{mod } 1) : n \in \mathbb{N}\} \neq \emptyset$.

Finally, suppose that $p_1z(\text{mod } 1) < q_1z(\text{mod } 1)$, then

$$-\zeta = p_1z(\text{mod } 1) - q_1z(\text{mod } 1),$$

so that $((p_1 - q_1)z)(\text{mod } 1) = -\zeta(\text{mod } 1)$. Similarly to the previous case we know that there are $d \in \mathbb{N} \cup \{0\}$ and $s \in [0, \zeta)$ such that $b - 1 = -d\zeta + s$. As $b - a > \zeta$ we know $-d\zeta \in (a - 1, b - 1)$. That is $-d\zeta + 1 \in (a, b)$ so that there is a $\gamma \in (a, b)$ such that $d(-\zeta) \equiv \gamma(\text{mod } 1)$. Hence, as $\gamma \in W \subset [0, 1)$ and clearly $(d(-\gamma))(\text{mod } 1) \in [0, 1)$ we have that

$$(d(p_1 - q_1)z)(\text{mod } 1) = \gamma \in W$$

Therefore with $d(p - q) = m$, say, we have that $m \in \mathbb{N}$ and

$$(mz)(\text{mod } 1) \in W,$$

so that

$$W \cap \{nz(\text{mod } 1) : n \in \mathbb{N}\} \neq \emptyset.$$
We have now considered all the necessary cases and so the result is proven. ◇

For the following result we use the notation that if $J$ is a subinterval of $I$ then for $a \in \mathbb{R}$

$$J + ^{'} a = \{(x + a)(\text{mod } 1) : x \in J\}.$$

Also we will say that two subintervals of $I$ are of the same form if they are both open, both closed, both half open with the closed end greater than the open end or both half open with the closed end smaller than the open end.

**Lemma 4.3**

Let $z \in \mathbb{I} - \mathbb{Q}$, let $J, K$ be intervals of equal length and form in $\mathbb{I}$, and let $\varepsilon > 0$. Then there is some $n \in \mathbb{N}$ such that

$$
\mu((P_n^z J) \Delta K) = \mu((J + ^{'} nz) \Delta K) < \varepsilon.
$$

**Proof:**

As $J, K$ have equal length, there is $\omega \in \mathbb{I}$ such that $J + ^{'} \omega = K$. Also if $V = [1 - \varepsilon/2, 1) \cup [0, \varepsilon/2)$ then

$$
\mu((J + ^{'} y) \Delta J) < \varepsilon \text{ for all } y \in V.
$$

(That is, if we move $J$ by a small enough amount, either left or right, then the measure of the symmetric difference will be less than $\varepsilon$.) Now, by Proposition 4.1 \{nz(\text{mod } 1) : n \in \mathbb{N}\} is dense in $\mathbb{I}$ so that there is an $n \in \mathbb{N}$ such that $nz(\text{mod } 1) \in V + ^{'} \omega$ that is $(nz - \omega)(\text{mod } 1) \in V$. We now have, using the fact that Lebesgue measure is translation invariant

$$
\mu((J + ^{'} nz) \Delta K) = \mu((J + ^{'} nz) \Delta (J + ^{'} \omega))
$$

$$
= \mu(((J + ^{'} nz) \Delta (J + ^{'} \omega)) + ^{'} (-\omega))
$$

$$
= \mu((J + ^{'} (nz - \omega)) \Delta J)
$$

< $\varepsilon$,  

87
which proves the Lemma.

We have thus described the origins of the fundamental tools that are used in this research. We now show how they are put together in proving that irrational rotations generate the Carathéodory definition. The form of proof is also important as throughout the work any proof of transformations generating the Carathéodory definition is fundamentally the same.

**Theorem 4.1**

*Any given irrational rotation on $\mathbb{I}$ with the Borel $\sigma$-algebra and Lebesgue measure generates the Carathéodory definition of measurable sets.*

**Proof:**

Let $P$ be an irrational rotation. Suppose that $A$ is $P$-invariant, as $P$ is one-to-one and onto

$$P^{-1}A = A \Rightarrow PA = A$$

so that $A$ is $P^{-1}$-invariant. Also since, like $P$, $P^{-1}$ is simply a translation in $\mathbb{I}$, $P^{-1}$ is measure, and outer measure, preserving. Let $A \subset \mathbb{I}$ be a $P$-invariant set with the property that there is a $\theta < 2$ such that

$$\mu_*(A \cap J) + \mu_*(A^c \cap J) \leq \theta \mu(J),$$

for all intervals $J \subset \mathbb{I}$. Suppose that both $A$ and $A^c$ have strictly positive outer measure. Then by Lemmas 4.1 and 4.2 we can find intervals of density of equal length $J_1$ and $J_2$ for $A$ and $A^c$ respectively to within $\varepsilon = (2 - \theta)/3$. Also, by Lemma 4.3 we can find an $n \in \mathbb{N}$ such that

$$\mu(P^n J_1 \triangle J_2) < \varepsilon \mu(J_2)$$

and so

$$\mu(J_1 \triangle P^{-n} J_2) < \varepsilon \mu(J_2).$$
We then observe that

\[
\mu_*(A \cap J_2) = \mu_*(P^{-n}(A \cap J_2)) \\
= \mu_*(P^{-n}A \cap P^{-n}J_2) \\
= \mu_*(A \cap P^{-n}J_2) \\
\geq \mu_*(A \cap J_1) - \mu(J_1 \triangle P^{-n}J_2) \\
> \mu_*(A \cap J_1) - \varepsilon \mu(J_2) \\
\geq (1 - \varepsilon)\mu(J_1) - \varepsilon \mu(J_2) \\
= (1 - \varepsilon)\mu(J_2) - \varepsilon \mu(J_2) .
\]

Therefore, we now have

\[
\mu_*(A \cap J_2) + \mu_*(A^c \cap J_2) > (1 - \varepsilon)\mu(J_2) - \varepsilon \mu(J_2) + (1 - \varepsilon)\mu(J_2) \\
= (2 - 3\varepsilon)\mu(J_2) \\
= (2 - 3\frac{(2 - \theta)}{3})\mu(J_2) \\
= (2 - (2 - \theta))\mu(J_2) \\
= \theta \mu(J_2) .
\]

This contradiction of the property of \( A \) means that either \( A \) or \( A^c \) has zero outer measure, which means that \( P \) has satisfied the conditions to become a transformation that generates the Carathéodory definition of measurable sets.

\[\diamondsuit\]

While the above results are interesting in themselves, it was not without a counterclaim that these results led to further work. It was claimed in [32] that there was another ergodic transformation on the unit circle for which the Carathéodory definition could not be generated in the same way as it was for irrational rotations. This transformation is \( T(z) = z^2 \) for all \( z \in \mathbb{T} \). Certainly directly rotating arcs (the circle form of intervals in \( \mathbb{I} \)) through
appropriate angles is impossible as, with $T$, forward iterations of the transformation are not measure preserving, and backwards iterations make two smaller copies of the set. It was then noted that the irrational rotation is not mixing, while $T$ is strong mixing. The questions

1. Is there a defining characteristic that determines whether or not a transformation will generate the Carathéodory definition?,

2. Is such a characteristic related to the level of mixing that a transformation has?

arose from this fact. These two questions, as mentioned in the introduction are two of the central questions in this thesis. The first work done in attempting to answer these questions was to find a weak but not strong mixing transformation and determine whether or not it would generate the definition. The fruits of this work are presented in the next section, concentrating on the contribution that the work made to the final answers to the questions listed above. To conclude this section we prove, for completeness of the presentation, the claim that an irrational rotation is ergodic but non-mixing. We will prove that $T$ is strongly mixing, however, due to the necessity of some concepts yet to be introduced we delay the proof until the end of this chapter, at which point all of the necessary concepts will have been established. While the proofs presented are original, these results are already known. Alternative proofs are mentioned in the notes.

**Theorem 4.2**

An irrational rotation is ergodic but not weak mixing

**Proof:**

To prove that an irrational rotation $P_z$ is ergodic, we use Theorem 3.1. Let $P_z$ be an arbitrary irrational rotation. Let $A, B \in \mathcal{B}$ with $\mu(A) > 0$ and
\( \mu(B) > 0 \). Set some \( 0 < \varepsilon < 1/3 \), then by Lemmas 4.1 and 4.2 we can find intervals of density of equal length for \( A \), and \( B \) to within \( \varepsilon \), say \( J_1 \) and \( J_2 \). Then by Lemma 4.3 there is an \( n \in \mathbb{N} \) such that

\[
\mu(P_z^{-n}J_1 \triangle J_2) \leq \varepsilon \mu(J_2)
\]

and it follows that

\[
\mu(P_z^{-n}A \cap J_2) \geq \mu(P_z^{-n}A \cap P_z^{-n}J_1) - \mu(P_z^{-n}A \cap (J_2 \triangle P_z^{-n}J_1))
\]
\[
> \mu(P_z^{-n}A \cap P_z^{-n}J_1) - \varepsilon \mu(J_2)
\]
\[
= \mu(P_z^{-n}(A \cap J_1)) - \varepsilon \mu(J_2)
\]
\[
= \mu(A \cap J_1) - \varepsilon \mu(J_2)
\]
\[
> (1 - \varepsilon)\mu(J_1) - \varepsilon \mu(J_2)
\]
\[
= (1 - \varepsilon)\mu(J_2) - \varepsilon \mu(J_2).
\]

Now using the fact that \( A, B \in \mathcal{B} \) we note that

\[
\mu((P_z^{-n}A \triangle B) \cap J_2) \leq \mu(J_2)
\]
\[
< (2 - 3\varepsilon)\mu(J_2)
\]
\[
= (1 - \varepsilon)\mu(J_2) + (1 - \varepsilon)\mu(J_2) - \varepsilon \mu(J_2)
\]
\[
< \mu(B \cap J_2) + \mu(P_z^{-n}A \cap J_2)
\]
\[
= \mu(B \cap J_2 \cap (P_z^{-n}A)) + \mu(B \cap J_2 \cap (P_z^{-n}A)^c)
\]
\[
+ \mu((P_z^{-n}A) \cap J_2 \cap B) + \mu((P_z^{-n}A) \cap J_2 \cap B^c)
\]
\[
= \mu((P_z^{-n}A \triangle B) \cap J_2) + 2\mu(P_z^{-n}A \cap B \cap J_2).
\]

So that \( \mu(P_z^{-n}A \cap B) > 0 \) from which it follows, by Theorem 3.1, that an irrational rotation is ergodic.

Now select

\[
\delta = \frac{\min\{z, 1-z\}}{2} < \frac{1}{4}
\]
and set

\[ A = [0, \delta) = B. \]

Then if \( P_z^{-i}A \cap B \neq \emptyset \) we must have

\[ P_z^{-i}A \subseteq (1 - \delta, 1) \cup [0, 2\delta) \]
\[ \subseteq (1 - \delta, 1) \cup [0, z), \]

so that \( P_z^{-i-1}A \subseteq (1 - z - \delta, 1) \). Then as

\[ \delta \leq \frac{1 - z}{2} \]
\[ = 1 - z - \frac{1 - z}{2} \]
\[ \leq 1 - z - \delta, \]

\( P_z^{-i-1}A \cap B = \emptyset \). We therefore have that at least one in every two \( i \in \mathbb{N} \),
\( P_z^{-i}A \cap B = \emptyset \). Denote by \( P_N \), \( \{ n \in \mathbb{N} : n \leq N, P_z^{-n}A \cap B = \emptyset \} \) and using
\( |P_N| \) to denote the cardinality of \( P_N \) we have

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |\mu(P_z^{-i}A \cap B) - \mu(A)\mu(B)| \]
\[ = \lim_{n \to \infty} \frac{1}{n} \sum_{i \in P_n} |\mu(P_z^{-i}A \cap B) - \mu(A)\mu(B)| \]
\[ + \lim_{n \to \infty} \frac{1}{n} \sum_{i \in P_n, i \leq n} |\mu(P_z^{-i+1}A \cap B) - \mu(A)\mu(B)| \]
\[ = \lim_{n \to \infty} \frac{|P_n|\delta^2}{n} + \lim_{n \to \infty} \frac{1}{n} \sum_{i \in P_n, i \leq n} |\mu(P_z^{-i}A \cap B) - \mu(A)\mu(B)| \]
\[ \geq \frac{\delta^2}{2} \]
\[ > 0. \]

Comparison with the definition of weak mixing transformations then gives us that \( P_z \) is not weak mixing. \( \diamond \)
4.2 Orbits

This section discusses, as mentioned at the conclusion of the previous section, the initial results obtained when investigating an example of a weakly but not strongly mixing transformation. The transformation that we use for this was first described by Kakutani in [21]. We find, that this transformation does indeed generate the Carathéodory definition of measurable sets. Along the way to proving this result some interesting technical results concerning the characteristics of the transformation arise. However, while we will look at these interesting results, properties specific to Kakutani’s transformation are not the main contribution of this research to the central pursuit of the thesis. The main contributions are firstly that there is a weak but not strong mixing transformation that generates the Carathéodory definition. Secondly, the investigation of Kakutani’s transformation establishes the first ‘generalisation’ of a rotation. That is, the use of orbits of intervals and, more generally the orbits of sets are established in the sense of this work. The results in this section are our own (except where noted) though they are incorporated into [25].

We start by describing the transformation that we will be considering. It is an example of the type of transformation known as either a tower or skyscraper construction. The general form of tower transformations is of great interest to us. However, we will pay particular attention to the derived two level tower transformation of which Kakutani’s transformation is an example and leave the general form of tower transformations to the next chapter where their effect on this research is considered. Let $(X, B, \mu)$ be a measure space and $T$ be a transformation on $X$. We now construct the derived two level tower transformation from this space which, in this situation, is called the primitive space. Let $A$ be a non-empty subset of $X$ and let $A'$ be a set disjoint
from $X$ for which there exists a one-to-one onto mapping $\tau : A \rightarrow A'$. Let
\[
\tilde{X} = X \cup A'
\]
\[\tilde{\mathcal{B}} = \left\{ B : B \subseteq \tilde{X}, B \cap X \in \mathcal{B} \text{ and } \tau^{-1}(B \cap A') \in \mathcal{B} \right\}, \text{ and}\]
\[\tilde{\mu}(B) = \mu(B \cap X) + \mu(\tau^{-1}(B \cap A')), \text{ for all } B \in \tilde{\mathcal{B}}.\]

$(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ is the measure space for the derived tower transformation. To make sense of the term tower we think of $A'$ being placed directly above $A$. Then define the tower transformation $\tilde{T}$ on $\tilde{X}$ by
\[
\tilde{T} = \begin{cases} 
\tau(x) & \text{if } x \in A, \\
T(x) & \text{if } x \in X \cap A^c, \text{ and} \\
T(\tau^{-1}(x)) & \text{if } x \in A'.
\end{cases}
\]
It is easy to see that $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ is a measure space and that $\tilde{T}$ is a transformation on $\tilde{X}$ that is measure preserving if $T$ is. We show later that the outer measure preserving property is also retained. As well as the measure we will also need to consider the outer measures of both the primitive and derived spaces. The outer measure $\mu_*$ associated with $\mu$ is defined for all $B \subseteq X$ by
\[\mu_*(B) = \inf\{\mu(C) : B \subseteq C, C \in \mathcal{B}\}.
\]
Similarly the outer measure associated with the derived space, $\tilde{\mu}_*$ is defined for all $B \subseteq \tilde{X}$ by
\[
\tilde{\mu}_*(B) = \inf\{\mu(C) : B \subseteq C, C \in \tilde{\mathcal{B}}\} = \mu_*(B \cap X) + \mu_*(\tau^{-1}(B \cap A')).
\]
So that, just as in the case of the measures, the outer measures $\mu_*$ and $\tilde{\mu}_*$ are identical on $X$.

Throughout the Thesis an extension on a dynamical system $(X, \mathcal{B}, \mu, T)$ constructed as we have just described will be referred to as a **usual two level tower extension**.
Definition 4.4

A point in the interval [0, 1) which is of the form \( k/p^n \) for some \( n \in \mathbb{N} \) and some \( k \in \{0, 1, 2, ..., p^n - 1\} \) is called a p-adic rational. A subinterval \( J \) of [0, 1) is called a p-adic interval of order \( n \) if there is a \( q \in \{0, 1, 2, ..., p^n - 1\} \) such that

\[
J = \left( \frac{q}{p^n}, \frac{q + 1}{p^n} \right).
\]

In the special case where \( p = 2 \), we exchange the adjective p-adic with dyadic.

Originally, Kakutani's example was defined on the unit interval with the dyadic rationals removed. Indeed some of the concepts we are trying to introduce are more cleanly illustrated with the dyadic rationals removed. However, for the purposes of convenience and consistency in the second part of this thesis, the dyadic rationals have been included. Therefore, for Kakutani's two level tower transformation we start with the primitive space space \((X, \mathcal{B}, \mu)\) of \( X = [0, 1) \), with the usual Lebesgue measure space on \( X \). Also for the purposes of defining the transformation we let

\[
I_n = \left[ 1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}} \right),
\]

for each \( n \in \{0\} \cup \mathbb{N} \). We note that \( X = \bigcup_{n=0}^{\infty} I_n \) and define \( \psi \) on \( X \) as

\[
\psi(x) = x - 1 + \frac{1}{2^n} + \frac{1}{2^{n+1}}, \text{ for } x \in I_n.
\] (9)

To complete the definition of Kakutani's transformation we let

\[
A = \bigcup_{n=0}^{\infty} I_{2n},
\]

and define \( \tilde{\psi} \) on \((\bar{X}, \bar{\mathcal{B}}, \bar{\mu})\) by the general rules for the construction of a two level tower transformation, where we let the disjoint set be denoted \( A' \) and the one-to-one onto mapping between \( A \) and \( A' \) be denoted by \( \tau \).
We can see from the definitions that \( \psi \) and \( \tilde{\psi} \) involve rearranging dyadic intervals. To fully comprehend the mechanism by which the transformation rearranges dyadic intervals we need an additional definition.

**Definition 4.5**

Suppose that \( J \) is a dyadic interval

\[
J = \left( \frac{p}{2^n}, \frac{p+1}{2^n} \right)
\]

for some \( n \in \mathbb{N} \cup \{0\} \), \( p \in \{0, 1, 2, ..., 2^n - 1\} \). Then the upper half of \( J \), denoted by \( J_u \), is the dyadic interval

\[
J_u = \left( \frac{p + \frac{1}{2}}{2^n}, \frac{p + 1}{2^n} \right) = \left( \frac{2p + 1}{2^{n+1}}, \frac{p + 1}{2^n} \right).
\]

Similarly the lower half of \( J \), denoted by \( J_l \), is the dyadic interval

\[
J_l = \left( \frac{p}{2^n}, \frac{p + \frac{1}{2}}{2^n} \right) = \left( \frac{p}{2^n}, \frac{2p + 1}{2^n} \right).
\]

In the more general case, if

\[
J = \left( \frac{p}{b^n}, \frac{p+1}{b^n} \right)
\]

for some \( b \in \mathbb{N} \) and some \( p \in \{0, 1, ..., b^n-1\} \), then for each \( i \in \{0, 1, ..., b-1\} \) we will write

\[
J_i = \left( \frac{p}{b^n} + \frac{i}{b^{n+1}}, \frac{p}{b^n} + \frac{i+1}{b^{n+1}} \right).
\]

The transformation \( \psi \) is not weak mixing, whereas the tower extension \( \tilde{\psi} \) of \( \psi \) is weak mixing on \( \tilde{X} \). As in the previous section the mixing properties of the transformations in the present discussion will be discussed at the end of the section.

Clearly, our aim in considering this Kakutani example of a weak but not strong mixing transformation is to show that it generates the Carathéodory definition of measurable sets. That is we wish to prove the following result.
Theorem 4.3
Let $B \subseteq \tilde{X}$ and let $B$ be $\tilde{\psi}$-invariant. Suppose that there is a $\theta = \theta(B) \in [0, 2)$ such that for all dyadic intervals, $J$ of $[0, 1)$,

$$\bar{\mu}_*(B \cap J) + \bar{\mu}_*(B^c \cap J) \leq \theta \mu(J).$$

Then $\theta \in [1, 2)$ and either $\bar{\mu}_*(B) = 0$ or $\bar{\mu}_*(B^c) = 0$.

We aim to use the same or a similar method to that for the irrational rotations. The existence of the intervals of density carries over immediately, though we later find that there is a further property of density of intervals in $[0, 1)$ that needs to be proven in this case. The rotation of sets in this example is somewhat more difficult. However, from the definition of the transformation we can see that dyadic intervals are intimately involved. Just how thoroughly intertwined dyadic intervals and points are in the properties of the transformation is quite surprising. (See, for example, Theorem 7.7 (ii).) It is from these relations that our ability to rotate intervals of density will follow.

To compensate for the loss of rotations in the sense of an irrational rotation on the unit circle or its analogy in the unit interval, we define orbits.

Definition 4.6
Let $X$ be a set and $T : X \to X$ be a (not necessarily invertible) transformation. For any subset $D$ of $X$ the $T$-orbit of $D$ is

$$O_T(D) = \{T^n(D) : n \in \mathbb{Z}\},$$

or, if $T$ is understood, $O_T(D)$ is simply the orbit of $D$. A similar definition is made for the orbit of a point $x \in X$. In the case where $T^{-n}(x)$ exists (and
is a point) in $X$ for all $n \in \mathbb{N}$ we define the $T$-orbit of $x$ or simply the orbit of $x$, if $T$ is understood, to be

$$O_T(x) = \{T^n(x) : n \in \mathbb{Z}\}.$$ 

Otherwise let $j > 0$ be the smallest integer such that either $T^{-j}(x)$ does not exist or $T^{-j}(x)$ is a set of more than one point, then we define the $T$-orbit of $x$ or simply the orbit of $x$ if $T$ understood, to be

$$O_T(x) = \{T^n(x) : n \in \mathbb{Z}, n > -j\}.$$ 

Two directly related definitions follow.

**Definition 4.7**

Let $(X, \mathcal{B}, \mu)$ be a measure space. A subset $D$ of $X$ is said to essentially equal to another subset $B$ of $X$ if $\mu(B \Delta D) = 0$, and we write this as $D \equiv B$.

**Definition 4.8**

Let $X$ be a set and $T : X \to X$ be a transformation on $X$. Then a subset $D$ of $X$ is said to have a cyclic orbit if for some $n \in \mathbb{Z}$, $T^n(D) = D$, in which case the sets of sets $O_T(D)$ is finite. A subset $D$ of $X$ is said to have an essentially cyclic orbit if for some $n \in \mathbb{Z} - 0$, $T^n(D)$ essentially equals $D$. In this case the $T$-orbit of $D$ is said to essentially equal $\{D, T(D), ..., T^{n-1}(D)\}$ if $n > 0$ and $\{D, T^{-1}D, ..., T^{n+1}D\}$ if $n < 0$. We write

$$O_T(D) \equiv \{D, T(D), ..., T^{n-1}(D)\},$$

and

$$O_T(D) \equiv \{D, T^{-1}D, ..., T^{n+1}D\}$$

for these respective cases to denote that $O_T(D)$ is essentially equal to $\{D, T(D), ..., T^{n-1}(D)\}$ or $\{D, T^{-1}D, ..., T^{n+1}D\}$ in the respective cases of $n > 0$ and $n < 0$. 

98
Orbits come in particularly useful with Kakutani’s $\psi$ transformation because we find that the orbit of a dyadic interval is essentially cyclic and that the sets in the orbit are, essentially, each of the dyadic intervals of the same order as the original interval. Provided that we can prove that intervals of density can always be chosen to be dyadic intervals of any high enough order, it is clear how we would prove that $\psi$ generates the Carathéodory definition. For any set $B$ satisfying the necessary conditions, we would find dyadic intervals of density for $B$ and $B^c$ of the same order, and then move one through its orbit, until the image under consideration essentially equals the other interval of density. It is in this sense that we are ‘rotating’ the dyadic intervals of density.

The main results presented regarding Kakutani’s $\tilde{\psi}$ transformation are, as previously mentioned, original results incorporated into [25]. These results are now presented in Lemma’s 4.4, 4.5, 4.6, 4.7 and Theorem 4.4.

We now prove the result concerning the orbits of dyadic intervals just discussed in Lemma 4.4. The Lemma is in four parts. The first shows that the image under $\psi$ of a dyadic interval is essentially another dyadic interval of the same order. The second describes the effect of iterations of $\psi$ on the internal structure of a dyadic interval, which is necessary to prove part three, which is the result discussed above. The final part pays particular attention to the orbit of the dyadic interval $[0, 1/2^n)$. This is a vital part of extending the ideas just discussed from the primitive transformation $\psi$ to the weakly mixing derived transformation $\tilde{\psi}$. 
Lemma 4.4

Let $J$ be a dyadic subinterval of $X$ and let $n \in \mathbb{N}$, $q \in \{0, 1, 2, \ldots, 2^n - 1\}$ be such that

$$J = \left( \frac{q}{2^n}, \frac{q + 1}{2^n} \right).$$

Then the following hold.

(i) If $q \in \{0, 1, 2, \ldots, 2^n - 2\}$, then there is a $k \in \{0, 1, 2, \ldots, 2^n - 1\}$, $k \neq q$ such that

$$\psi(J) = \left( \frac{k}{2^n}, \frac{k + 1}{2^n} \right).$$

If $q = 2^n - 1$, then

$$\psi(J) = \left( 0, \frac{1}{2n+1} \right) \cup \left( \frac{1}{2n+1}, \frac{1}{2n} \right).$$

(ii) If $q \in \{0, 1, \ldots, 2^n - 2\}$ then

$$\psi(J_u) = (\psi(J))_u \text{ and } \psi(J_l) = (\psi(J))_l.$$ 

If $q = 2^n - 1$ then

$$\psi(J_l) = (\psi(J))_u \text{ and } \psi(J_u) \equiv (\psi(J))_l.$$ 

(iii) $O_{\psi}(J)$ is essentially equal to the set of all dyadic intervals of order $n$.

(iv) If $q = 0$ and $p \in \{0, 1, 2, \ldots, 2^n - 1\}$ then

$$\psi_p(J) = \left( \frac{2^n - 1}{2^n}, 1 \right) \iff p = 2^n - 1.$$ 

Proof:

(i) As $q/2^n \in [0, 1)$ there is an $n_0 \in \{0\} \cup \mathbb{N}$ such that

$$\frac{q}{2^n} \in I_{n_0} = \left[ 1 - \frac{1}{2^n_0}, 1 - \frac{1}{2^n_{0+1}} \right].$$ 

If $n < n_0$, then $1/2^n > 1/2^{n_0}$ and so

$$1 \geq \frac{q + 1}{2^n} \geq 1 - \frac{1}{2^{n_0}} + \frac{1}{2^n} > 1,$$
a contradiction.  Hence \( n \geq n_0 \). There are two cases remaining.

CASE I.

Consider when \( n > n_0 \). In this case

\[
1 - \frac{1}{2^{n_0}} \leq \frac{q}{2^n} < 1 - \frac{1}{2^{n_0+1}} = \frac{2^n - 2^{n-n_0-1}}{2^n},
\]

so that

\[
1 - \frac{1}{2^{n_0}} \leq \frac{q}{2^n} < \frac{q + 1}{2^n} \leq \frac{2^n - 2^{n-n_0-1}}{2^n} = 1 - \frac{1}{2^{n_0+1}}.
\]

Thus

\[
J = \left( \frac{q}{2^n}, \frac{q + 1}{2^n} \right) \subseteq \left[ 1 - \frac{1}{2^{n_0}}, 1 - \frac{1}{2^{n_0+1}} \right) = I_{n_0}.
\]

It now follows from the definition of \( \psi \) that

\[
\psi \left( \frac{q}{2^n} \right) = \frac{q}{2^n} - 1 + \frac{1}{2^{n_0}} + \frac{1}{2^{n_0+1}} = \frac{q - 2^n + 2^{n-n_0} + 2^{n-n_0-1}}{2^n}, \quad \text{and}
\]

\[
\lim_{x \to (q+1)/2^n} \psi(x) = \frac{q + 1}{2^n} - 1 + \frac{1}{2^{n_0}} + \frac{1}{2^{n_0+1}} = \frac{q - 2^n + 2^{n-n_0} + 2^{n-n_0-1} + 1}{2^n},
\]

where \( x \) approaches \((q + 1)/2^n\) from below. Now, \( \psi \) is linear and increasing on \( I_{n_0} \). Thus, if we put \( k = q - 2^n + 2^{n-n_0} + 2^{n-n_0-1} \), we have

\[
\psi(J) = \psi \left( \left( \frac{q}{2^n}, \frac{q + 1}{2^n} \right) \right) = \left( \frac{k}{2^n}, \frac{k + 1}{2^n} \right),
\]

where necessarily \( k \neq q \) and \( k \in \{0, 1, 2, \ldots, 2^n - 1\} \). Thus (i) holds in this case.

CASE II.

Consider when \( n = n_0 \). Then \( n_0 > 0, q > 0 \) and \( q/2^n \in [1 - 1/2^n, 1 - 1/2^{n+1}) \).

However the only dyadic rational of order \( n \) in this interval is \( 1 - 1/2^n \). Hence

\[
q = 2^n - 1 \quad \text{so that}
\]

\[
\frac{q}{2^n} = 1 - \frac{1}{2^n}, \quad \text{and} \quad \frac{q + 1}{2^n} = 1.
\]
The definition of \( I_n \) now gives us that

\[
J = \left( 1 - \frac{1}{2^n}, 1 \right) = \left( 1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}} \right) \bigcup \left( \bigcup_{k=n+1}^{\infty} I_k \right).
\]

Hence,

\[
\psi(J) = \psi \left( \left( 1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}} \right) \bigcup \left( \bigcup_{k=n+1}^{\infty} I_k \right) \right)
= \psi \left( \left( 1 - \frac{1}{2^n}, 1 - \frac{1}{2^{n+1}} \right) \bigcup \left( \bigcup_{k=n+1}^{\infty} \psi(I_k) \right) \right)
= \left( \frac{1}{2^{n+1}}, \frac{1}{2^n} \right) \bigcup \left( \bigcup_{k=n+1}^{\infty} \left[ \frac{1}{2^{k+1}}, \frac{1}{2^k} \right] \right)
= \left( 0, \frac{1}{2^{n+1}} \right) \bigcup \left( \frac{1}{2^{n+1}}, \frac{1}{2^n} \right).
\]

This shows that the conclusion of (i) also holds in this case.

(ii) If \( q \in \{0, 1, 2, \ldots, 2^n - 2\} \) then the situation is as in the first case considered in (i). We have that for some \( n_0 \),

\[
J \subseteq \left[ 1 - \frac{1}{2^{n_0}}, 1 - \frac{1}{2^{n_0+1}} \right] = I_{n_0}.
\]

As \( \psi \) is linear and increasing on \( I_{n_0} \), \( \psi \) is linear and increasing on \( J \), so (ii) holds for the case when \( q \in \{0, 1, 2, \ldots, 2^n - 2\} \).

If \( q = 2^n - 1 \), then \( J = (1 - 1/2^n, 1) \) and by part (i) we have that

\[
\psi(J) = (0, 1/2^{n+1}) \cup (1/2^{n+1}, 1/2^n)
\]

so that

\[
(\psi(J))_I = (0, 1/2^{n+1}).
\]

Now \( J_u = (1 - 1/2^{n+1}, 1) \). It follows that

\[
\psi(J_u) = \left( 0, \frac{1}{2^{n+2}} \right) \bigcup \left( \frac{1}{2^{n+2}}, \frac{1}{2^{n+1}} \right),
\]

102
so that

$$\psi(J_u) \equiv (\psi(J))_u.$$ 

Also, $J_i = (1 - 1/2^{n+1}, 1 - 1/2^{n+1})$ and so we have

$$\psi(J_i) = \left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right) = (\psi(J))_u.$$ 

Thus (ii) holds when $q = 2^n - 1$.

(iii) We proceed using mathematical induction on $n$. For $n = 1$, consider when $J = (0, 1/2)$. Then $\psi(J) = (1/2, 1)$, and

$$\psi^2(J) = \psi(1/2, 1) = (0, 1/4) \cup (1/4, 1/2) \equiv J.$$ 

Thus the orbit $O_\psi(J)$ of $J$ is essentially cyclic with each element essentially equal to one of the 2 dyadic intervals of order 1, $(0, 1/2)$ or $(1/2, 1)$. A similar argument applies for $J = (1/2, 1)$ and hence the claim is true for $n = 1$.

Now assume that the claim is true for $n = k$. That is, if $q \in \{0, 1, 2, ..., 2^k - 1\}$ and $J = (q/2^k, (q+1)/2^k)$, the orbit $O_\psi(J)$ of $J$ is essentially cyclic consisting of

$$\left(0, \frac{1}{2^k}\right), \left(\frac{1}{2^k}, \frac{2}{2^k}\right), ..., \left(\frac{2^k - 1}{2^k}, 1\right)$$

with $\psi^{2^k}(J) \equiv J$.

Let $J$ be the dyadic interval of order $k$ given by

$$J = \left(\frac{2^k - 1}{2^k}, 1\right),$$

and let $K$ be the dyadic interval of order $k + 1$ given by

$$K = \left(\frac{2^{k+1} - 1}{2^{k+1}}, 1\right) = \left(\frac{2^k - 1}{2^k}, 1\right)_u = J_u.$$
By the inductive assumption, the orbit of $J$ is essentially cyclic, consisting of all of the dyadic intervals of order $k$. It follows by part (ii) that

$$\psi(K) = \psi(J) \equiv (\psi(J))_l,$$

and that

$$\psi^i(K) = (\psi^i(J))_l,$$

for all $i = 1, 2, ..., 2^k - 1$. Thus, the first $2^k$ sets in the $\psi$-orbit of $K$ respectively are equal to $K = J_u, (\psi(J))_l, (\psi^2(J))_l, ..., (\psi^{2^k-1}(J))_l$.

Now,

$$\psi^{2^k}(K) = \psi \left( \psi^{2^k-1}(K) \right)$$
$$\equiv \psi \left( (\psi^{2^k-1}(J))_l \right)$$
$$\equiv \left( \psi(\psi^{2^k-1}(J)) \right)_l$$
$$\equiv \left( \psi^{2^k}(J) \right)_l$$
$$\equiv J_l.$$

But by part (ii) $\psi(J_l) = (\psi(J))_u$ and the previous argument may now be repeated with $(\psi(J))_u$ in place of $(\psi(J))_l$. We deduce that the next $2^k$ elements of the orbit of $K$ are essentially equal to the upper parts of the intervals for which the previous $2^k$ elements were the lower parts. That is, the next $2^k$ elements in the orbit of $K$ are $J_u, (\psi(J))_u, ..., (\psi^{2^k-1}(J))_u$ respectively, and that $\psi^{2^k}(J_l) = J_u$. In summary, we now see that the $\psi$-orbit of $K$ is essentially cyclic with elements

$$J_u, (\psi(J))_l, (\psi^2(J))_l, ..., (\psi^{2^k-1}(J))_l, J_l, (\psi(J))_u, (\psi^2(J))_u, ..., (\psi^{2^k}(J))_u.$$

However, by the inductive assumption, these sets are essentially equal to all the upper and lower parts of all the dyadic intervals of order $k$. Since the
collection of all of the upper and lower parts of the dyadic intervals of order \( k \) is the set of all dyadic intervals of order \( k + 1 \), we have that the \( \psi \)-orbit of \( K \) is essentially cyclic consisting of all of the dyadic intervals of order \( k + 1 \). But, since the \( \psi \)-orbit of \( K \) essentially equals the set of all the dyadic intervals of order \( k + 1 \), any dyadic interval of order \( k + 1 \) will have an essentially cyclic orbit consisting of all of the dyadic intervals of order \( k + 1 \). So, if the result is true for \( k \), it is also true for \( k + 1 \) which proves (iii).

(iv) The fact that \( (0, 1/2^n) \) has an essentially cyclic orbit consisting of all of the dyadic intervals of order \( N \) is a consequence of part (iii), and by observing that \( \psi(1 - 1/2^n, 1) \equiv (0, 1/2^n) \equiv \psi^2(0, 1/2^n) \) since \( O_\psi((0, 1/2^n) \) is essentially cyclic, it follows that \( \psi^p((0, 1/2^n) \equiv (1 - 1/2^n, 1) \iff p = 2^n - 1 \) from which the result is a direct consequence.

An important note that follows from part (i) is that if \( J \) is a dyadic subinterval of \( X \) not of the form \([1 - 1/2^n, 1]\) then \( J \subseteq I_m \) for some \( m \in \mathbb{N} \).

Lemma 4.4, as proven, is perfect for considering \( \psi \) as a transformation that generates the Carathéodory definition. However, we wish to particularly consider the orbits of intervals under \( \tilde{\psi} \). Fortunately, we are able to show that \( O_\psi((0, 1/2^n)) \subset O_\tilde{\psi}((0, 1/2^n)) \) for each \( n \in \mathbb{N} \) so that by placing any two dyadic intervals of the same order in the orbit of the dyadic interval \((0, 1/2^n)\) of the same order we are still able to rotate one onto the other. This particularly useful result is proven in Lemma 4.6 after we prove, in Lemma 4.5, a necessary technical result.
Lemma 4.5

Let $J$ be a dyadic subinterval of $[0,1)$ and let $n \in \mathbb{N}, \ q \in \{0,1,...,2^n - 2\}$ be such that

$$J = \left( \frac{q}{2^n}, \frac{q+1}{2^n} \right).$$

Then either $J \cap A = \emptyset$ in which case $\tilde{\psi}(J) = \psi(J)$, or $J \subseteq A$ in which case $\tilde{\psi}(J) = \tau(J) \subseteq A'$.

Proof:

By the definition of the $I_m$ there is a unique $m \in \{0,1,2,...\}$ such that $q/2^n \in I_m$. As in the proof of part (i) of Lemma 4.4, it is easy to check that $n \geq m$. Now, recall that

$$I_m = \left[ \frac{2^m - 1}{2^n}, \frac{2^{m+1} - 1}{2^n + 1} \right) = \left[ \frac{2^n - 2^n - m}{2^n}, \frac{2^n - 2^{n-m-1}}{2^n} \right) = \bigcup_{j=2^n-2^n-m}^{2^n-2^n-m-1} \left[ \frac{j}{2^n}, \frac{j+1}{2^n} \right),$$

and it follows that $q \in \{2^n-2^n-m,...,2^n-2^n-m-1\}$. As $J = (q/2^n, (q+1)/2^n)$, this makes it clear that $J \subseteq I_m$. So by the definition of $A$ we now have that $J \subseteq A$ or $J \cap A = \emptyset$. Further, by the definition of $\tilde{\psi}$ we have

$$\tilde{\psi}(J) = \begin{cases} \tau(J), & \text{if } m \text{ is even (if } J \subseteq A), \\ \psi(J), & \text{if } m \text{ is odd (if } J \cap A = \emptyset). \end{cases}$$

\[ \diamond \]

Lemma 4.6

Let $n \in \mathbb{N}$ and $J = (0,1/2^n)$. Then for all $j \in \{0,1,2,...,2^n - 1\},$

$$\psi^j(J) \in O_{\tilde{\psi}}(J).$$
Proof:

We keep $n$ fixed, and use an induction argument on $j$. For $j = 0$,

$$\psi^0(J) = J = \bar{\psi}^0(J) \in O_\bar{\psi}(J),$$

so that the claim is true.

Now, assume that the claim is true for some $j \in \{0, 1, 2, ..., 2^n - 2\}$. Then there is a $q \in \{0, 1, 2, ...\}$ such that $\psi^j(J) = \bar{\psi}^q(J)$, and we know that $\psi^j(J)$ is a dyadic interval of order $n$. Further because $j < 2^n - 1$, part $(iv)$ of Lemma 4.4 shows that $\psi^j(J) \neq (1 - 1/2^n, 1)$. Thus Lemma 4.5 applies to the dyadic interval $\psi^j(J)$, and we have

$$\bar{\psi}(\psi^j(J)) = \psi(\psi^j(J)) \text{ or } \bar{\psi}(\psi^j(J)) = \tau(\psi^j(J)).$$

In the former case we have

$$\bar{\psi}^{q+1}(J) = \bar{\psi}(\bar{\psi}^q(J)) = \bar{\psi}(\psi^j(J)) = \psi^{j+1}(J),$$

so that $\psi^{j+1}(J) \in O_\bar{\psi}(J)$. In the latter case we have

$$\bar{\psi}^{q+2}(J) = \bar{\psi}^2(\bar{\psi}^q(J)) = \bar{\psi}^2(\psi^j(J)) = \bar{\psi}(\bar{\psi}(\psi^j(J))) = \bar{\psi}(\tau(\psi^j(J))) = \psi(\psi^j(J)) = \psi^{j+1}(J),$$

so that again $\psi^{j+1}(J) \in O_\bar{\psi}(J)$. The argument now covers both cases and the proof is complete. \hfill \Diamond
The proof that the desired mechanism for rotating sets exists is now complete. Moreover, we can see how the use of orbits of intervals is an excellent substitute for rotation. On showing that using the above lemmas we can generate the Carathéodory definition with $\psi$ we will have demonstrated the usefulness of the orbits of intervals.

To show that we can indeed now generate Carathéodory’s definition we must first show that the consideration of dyadic intervals advances our argument in showing that we can find dyadic intervals of density. As we have not placed many restrictions on the structure of $A'$, we do not know that there are dyadic intervals (or intervals at all) in $A'$, hence in proving the existence of dyadic intervals of density, which will be done in the space $X$, we use $\tilde{\psi}$-invariance to ensure that any set of positive outer measure will also have an intersection with $X$ of positive outer measure. The fact that a $\tilde{\psi}$ set of positive outer measure must have an intersection with $X$ of positive outer measure requires proof. The proof is given, along with the existence of dyadic intervals of density in the following lemma.

**Lemma 4.7**

Let $B \subseteq \tilde{X}$ be $\tilde{\psi}$-invariant, let $\tilde{\mu}(B) > 0$ and let $\varepsilon > 0$. Then there is an $n_0 \in \mathbb{N}$ which has the following property: if $n > n_0$, there is an dyadic interval $J$ in $X$ of order $n$ such that

$$\tilde{\mu}_*(B \cap J) > (1 - \varepsilon)\tilde{\mu}(J).$$

**Proof:**

We first need to show that $\tilde{\mu}_*(B \cap X) > 0$. To this end, observe that

$$\tilde{\mu}_*(B) = \mu_*(B \cap X) + \mu_*(\tau^{-1}(B \cap A')).$$

As $\tilde{\mu}_*(B) > 0$, either $\mu_*(B \cap X) > 0$ or $\mu_*(\tau^{-1}(B \cap A')) > 0$. In the former
case there is nothing to verify while in the latter case we have

\[ 0 < \mu_*(\tau^{-1}(B \cap A')) \]

\[ = \tilde{\mu}_*(B \cap A') \]

\[ = \tilde{\mu}_*(\tilde{\psi}^{-1}(B \cap A')) \]

\[ = \tilde{\mu}_*(\tilde{\psi}^{-1}(B) \cap \tilde{\psi}^{-1}(A')) \]

\[ = \tilde{\mu}_*(B \cap A) \]

\[ \leq \tilde{\mu}_*(B \cap X). \]

Thus, in either case \( \tilde{\mu}_*(B \cap X) > 0. \)

Now, by Lemma 4.1, with \( \varepsilon \) replaced by \( \varepsilon/2 \), we know that we can find an interval \( J = (a, b) \) in \( X \) such that

\[ \mu_*(B \cap J) > (1 - \varepsilon/2)\mu(J). \]

Define

\[ \theta = \mu(J)/2 = (b - a)/2 \]

and then choose

\[ n_0 \in \mathbb{N} \text{ such that } 1/2^{n_0} < \theta. \]

Let \( n_1 \) be any natural number greater than \( n_0 \) then there exist \( p, q \in \mathbb{N} \) and \( r, s \in \mathbb{R} \) with \( 0 \leq r < 1/2^{n_1} \), \( 0 \leq s < 1/2^{n_1} \) and

\[ a = p/2^{n_1} + r \quad \text{and} \quad b = q/2^{n_1} + s. \]

Further by the choice of \( n_1 \) (and \( n_0 \)) we have \( p < q - 1. \)

\( J \) can now essentially be separated into \((q - p + 1)\) distinct intervals

\[ J = W \cup Z_1 \cup \ldots \cup Z_{q-p-1} \cup V \]
where

\[ W = \left( a, \frac{p + 1}{2^{n_1}} \right) \text{ and } \mu(W) = \delta_1, \quad 0 < \delta_1 < 1/2^{n_1}, \]

\[ Z_i = \left( \frac{p + i}{2^{n_1}}, \frac{p + i + 1}{2^{n_1}} \right) \text{ and } \mu(Z_i) = 1/2^{n_1}, \quad i = 1, 2, \ldots, (q - p - 1), \]

\[ V = \left( \frac{q}{2^{n_1}}, b \right) \text{ and } \mu(V) = \delta_2, \quad 0 < \delta_2 < 1/2^{n_1}. \]

Note that the intervals \( Z_i \) are dyadic, and so it is enough to show that at least one of these is an interval of density. Suppose that this is not the case, then

\[ \mu_*(B \cap Z_i) \leq (1 - \varepsilon)/2^{n_1} \text{ for } i = 1, 2, \ldots, q - (p + 1). \]

By the choice of \( J \), letting \( q_1 = q - (p + 1) \)

\[
(1 - \varepsilon/2)\mu(J) < \mu_*(B \cap J)
\]

\[ = \mu_*(B \cap \left( \bigcup_{j=1}^{q_1} Z_j \cup W \cup V \right))
\]

\[ = \mu_*(\bigcup_{j=1}^{q_1} (B \cap Z_j) \cup (B \cap W) \cup (B \cap V))
\]

\[ \leq \sum_{j=1}^{q_1} \mu_*(B \cap Z_j) + \mu_*(B \cap W) + \mu_*(B \cap V)
\]

\[ \leq \sum_{j=1}^{q_1} (1 - \varepsilon)/(2^{n_1}) + \mu(W) + \mu(V)
\]

\[ = (1 - \varepsilon) \left( \sum_{j=1}^{q_1} 1/2^{n_1} + \mu(W) + \mu(V) \right) + \varepsilon (\mu(W) + \mu(V))
\]

\[ = (1 - \varepsilon) \left( \sum_{j=1}^{q_1} \mu(Z_j) + \mu(W) + \mu(V) \right) + \varepsilon (\mu(W) + \mu(V))
\]

\[ = (1 - \varepsilon) \mu \left( \bigcup_{j=1}^{q_1} Z_j \cup W \cup V \right) + \varepsilon (\mu(W) + \mu(V))
\]

\[ = (1 - \varepsilon) \mu(J) + \varepsilon (\mu(W) + \mu(V))
\]

This implies

\[ \mu(J) < 2(\mu(W) + \mu(V)) \]
\[ = 2(\delta_1 + \delta_2) \]
\[ < 4/2^{n_1} \]
\[ < 2\theta \]
\[ = \mu(J), \]

which is a contradiction. Therefore there is at least one interval \( Z \) of the form
\[ Z = \left( \frac{k}{2^{n_1}}, \frac{k+1}{2^{n_1}} \right) \]
for some \( k \in \{0, 1, \ldots, 2^{n_1} - 1\} \), such that
\[ \mu_*(B \cap Z) > (1 - \varepsilon)\mu(Z). \]

As \( n_1 \) was any natural number greater than \( n_0 \) it is clear that the dyadic interval of density can be chosen to be of any length of the form \( 1/2^n \) provided that \( n \) is greater than the relevant \( n_0 \).

The reason that this result is sufficient for \( \tilde{\psi} \) is that if a \( \tilde{\psi} \)-invariant set, \( A \), has positive outer measure in \( \tilde{X} \) then \( A \cap X \) will also have positive outer measure. This property also gives us that for each set of positive outer measure in \( \tilde{X} \) will have dyadic intervals of density in \( X \). Lemma 4.6 then allows us to prove Theorem 4.3 in the same way as we did Theorem 4.1. We now give a proof of Theorem 4.3.

**Theorem 4.3**

Let \( B \subseteq \tilde{X} \) and let \( B \) be \( \tilde{\psi} \)-invariant. Suppose that there is a \( \theta = \theta(B) \in [0, 2) \) such that for all dyadic intervals, \( J \) of \([0, 1)\),
\[ \tilde{\mu}_*(B \cap J) + \tilde{\mu}_*(B^c \cap J) \leq \theta \mu(J). \]

Then \( \theta \in [1, 2) \) and either \( \tilde{\mu}_*(B) = 0 \) or \( \tilde{\mu}_*(B^c) = 0 \).
Proof:
The fact that \( \theta \in [1, 2) \) is immediate from the observation that

\[
\mu(J) = \tilde{\mu}(J) \leq \tilde{\mu}_*(B \cap J) + \tilde{\mu}_*(B^c \cap J) \leq \theta \mu(J).
\]

Now suppose that \( \tilde{\mu}_*(B) > 0 \) and \( \tilde{\mu}_*(B^c) > 0 \). Set \( \varepsilon = (2 - \theta)/2 \). Then as \( \tilde{\mu}_*(B) > 0, \tilde{\mu}_*(B^c) > 0, \) and as \( B \) and thus \( B^c \) are \( \tilde{\psi} \)-invariant, by Lemma 4.7 there exist dyadic intervals of density \( J_1 \) and \( J_2 \) of the same order for \( B \) and \( B^c \) respectively to within \( \varepsilon \). That is,

\[
\tilde{\mu}_*(B \cap J_1) > (1 - \varepsilon)\mu(J_1) \quad \text{and} \quad \tilde{\mu}_*(B^c \cap J_2) > (1 - \varepsilon)\mu(J_2).
\]

Since \( J_1 \) and \( J_2 \) are dyadic intervals of the same order \( n \), say, Lemma 3.3 shows that both \( J_1 \) and \( J_2 \) belong to \( O_{\tilde{\psi}}((0, 1/2^n)) \). Thus, there are \( r, s \in \{0, 1, 2, \ldots, 2^n - 1\} \) such that

\[
J_1 = \tilde{\psi}^r \left( \left(0, \frac{1}{2^n} \right) \right) \quad \text{and} \quad J_2 = \tilde{\psi}^s \left( \left(0, \frac{1}{2^n} \right) \right).
\]

Without loss of generality we may suppose that \( r \leq s \), in which case we have

\[
\tilde{\psi}^{s-r}(J_1) = \tilde{\psi}^{s-r} \left( \tilde{\psi}^r \left( \left(0, \frac{1}{2^n} \right) \right) \right) = \tilde{\psi}^s \left( \left(0, \frac{1}{2^n} \right) \right) = J_2.
\]

Now, \( \mu(J_1) = \mu(J_2) \), \( \tilde{\mu}_* \) is \( \tilde{\psi} \)-invariant as an outer measure and \( B^c \) is a \( \tilde{\psi} \)-invariant set. Hence

\[
(1 - \varepsilon)\mu(J_1) = (1 - \varepsilon)\mu(J_2)
\]

\[
< \tilde{\mu}_*(B^c \cap J_2)
\]

\[
= \tilde{\mu}_*(\tilde{\psi}^{-(s-r)}(B^c \cap J_2))
\]

\[
= \tilde{\mu}_*(\tilde{\psi}^{-(s-r)}(B^c) \cap \tilde{\psi}^{-(s-r)}(J_2))
\]

\[
= \mu_*(B^c \cap J_1).
\]

Thus, using that \( \theta \geq 1 \)

\[
\tilde{\mu}_*(B \cap J_1) + \tilde{\mu}_*(B^c \cap J_1) > 2(1 - \varepsilon)\mu(J_1)
\]

\[
= \frac{2}{3}(1 + \theta)\mu(J_1)
\]

\[
> \mu(J_1).
\]
Comparing this with the given condition (10) gives an immediate contradiction. Hence, having both \( \tilde{\mu}_*(B) > 0 \) and \( \tilde{\mu}_*(B^c) > 0 \) is impossible and therefore \( \tilde{\mu}_*(B) = 0 \) or \( \tilde{\mu}_*(B^c) = 0 \).

\[ \diamond \]

This result, using orbits, shows the use of orbits and why we intend to continue to use them. It should also be apparent that it will not necessarily be the case that the orbits of intervals will be helpful. For example, we can not guarantee that there will not be a transformation \( T \) for which, for some \( \varepsilon > 0 \), some arbitrary interval \( J \), and for all intervals \( K \) such that \( \mu(K) = \mu(J) \),

\[ \mu(T^{-n}J \triangle K) < (1 - \varepsilon)\mu(J), \]

for some \( n \in \mathbb{N} \). Certainly, for such transformations, our current method of determining whether a transformation generates the Carathéodory definition or not is inappropriate.

At this stage we are still trying to prove the conjecture that the generation of Carathéodory's definition is related to the mixing properties of transformations. Since the fact that \( \tilde{\psi} \) is weakly mixing relies on the derived tower construction, and since it was the tower aspect of \( \tilde{\psi} \) that prevented the essentially cyclic nature of dyadic intervals from extending to \( \tilde{\psi} \) from \( \psi \), the potential of the orbits method of failing in derived tower constructions is of interest. The nature of the relationship between the property of generating Carathéodory's definition in a primitive space and in the derived spaces therefore became an area known to require investigation. The results of the investigation were interesting in that they lead to the development of the concept of 'splintering' an interval during its rotation through its orbit. These results are discussed in the next chapter. We conclude this section, and chapter, by proving the mixing properties of \( \psi \) and \( \tilde{\psi} \).

This part of the chapter is based on Kakutani [21].
We first show that $\psi$ is ergodic but not weakly mixing. This important result was mentioned without proof in [21]. Determining that the inclusion of a proof in this work was desirable we have included our own.

**Theorem 4.4**

$\psi$ is ergodic but not weak mixing.

**Proof:**

First we prove that $\psi$ is ergodic. Let $A, B \in \mathcal{B}$ with $\mu(A), \mu(B) > 0$. Then, for any selected $0 < \varepsilon < 1/3$, by Lemma 4.7 there are dyadic intervals of density $J_1$ and $J_2$ of equal order, say $m$, for $A$ and $B$ respectively. Also, using Lemma 4.4 there exists an $n \in \mathbb{N}$ such that $\psi^{-n}J_1 \equiv J_2$ and for this $n$ we have

$$\mu(\psi^{-n}A \cap J_2) \geq \mu(\psi^{-n}(A \cap J_1)) = \mu(A \cap J_1) > (1 - \varepsilon)\mu(J_1) = (1 - \varepsilon)\mu(J_2).$$

Then as $A, B, J_1, J_2$ are measurable

$$\mu(J_2) = \mu(B \cap \psi^{-n}A \cap J_2) + \mu((B \cap J_2) - \psi^{-n}A) + \mu((J_2 - B) \cap \psi^{-n}A)$$

$$\quad + \mu((J_2 - B) - \psi^{-n}A) \leq \mu(B \cap \psi^{-n}A) + \mu(J_2 - \psi^{-n}A) + \mu(J_2 - B) + \mu(J_2 - B)$$

$$\leq \mu(B \cap \psi^{-n}A) + 3\varepsilon \mu(J_2)$$

so that $\mu(B \cap \psi^{-n}A) > 0$. Thus, by Theorem 3.1, $\psi$ is ergodic.

To show that $\psi$ is not weakly mixing, let $A = (0, 1/2)$ and $B = (1/2, 1)$. Then, using Lemma 4.4, we have that $\psi^{-1}(A) \equiv B$ and $\psi^{-2}(B) \equiv B$ so that $\psi^{-2n-1}(A) \equiv B$ for all $n \in \mathbb{N}$. Similarly $\psi^{-2n}(A) \equiv A$ for all $n \in \mathbb{N}$. Hence for all $n \in \mathbb{N}$

$$|\mu(\psi^{-2n}(A) \cap B) - \mu(A)\mu(B)| = \left|0 - \frac{1}{4}\right| = \frac{1}{4}.$$
and
\[ |\mu(\psi^{-2n-1}(A) \cap B) - \mu(A)\mu(B)| = \left| \frac{1}{2} - \frac{1}{4} \right| = \frac{1}{4}. \]

We therefore have that
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |\mu(\psi^{-i}(A) \cap B) - \mu(A)\mu(B)| = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{4} = \frac{1}{4} \neq 0, \]
so that by the definition of weakly mixing transformations \( \psi \) is not weakly mixing. \( \Diamond \)

In proving that \( \tilde{\psi} \) is weakly mixing we need two technical results.

**Lemma 4.8**

Let \((X, \mathcal{B}, \mu)\) be a measure space defined on the interval \( X = [0,1) \) and \( b \in \mathbb{N} \).

Let \( \psi : X \to X \) be defined such that for any \( p \in \mathbb{N} \) and any \( b \)-adic interval of order \( p \)
\[ O_{\psi}(J) \equiv \left\{ \left( \frac{k}{bp}, \frac{k+1}{bp} \right) : k \in \{0,1,\ldots, b^p - 1\} \right\}. \]

Then for any \( f \in L^2(X, \mathcal{B}, \mu) \),
\[ \lim_{p \to \infty} \int_X |f \circ \psi^{bp} - f|^2 d\mu = 0. \]

**Proof:**

We show that
\[ \lim_{p \to \infty} \|f \circ \psi^{bp} - f\|_2 = 0. \]

Let \( \varepsilon > 0 \). Let \( f \in L^2([0,1]) \). Let \( g \in C([0,1]) \) be such that
\[ \|f - g\|_2 < \varepsilon, \]

115
which we know exists as \( f \in L^2(X, \mathcal{B}, \mu) \). (for a proof that \( C([0,1]) \) is dense in \( L^2(X, \mathcal{B}, \mu) \) see, for example, Federer [10] or Weierstrass [45].) As \( g \) is continuous on \([0,1]\), there is \( n_0 \in \mathbb{N} \) such that: if \( x, y \in [0,1] \) and \(|x - y| < 1/b_0^n\), then \(|g(x) - g(y)| < \varepsilon\).

Now, we know that if \( J \) is a dyadic interval of order \( n_0 \), then

\[
\psi^{b_{n_0}}(J) \equiv J.
\]

So, for such a \( J \), for any \( p \geq n_0 \),

\[
\psi^{b_{p}}(J) \equiv J.
\]

Now, let \( p \geq n_0 \). Then,

\[
||g \circ \psi^{b_p} - g||^2_2 = \int_X |g(\psi^{b_p}(x)) - g(x)|^2dx
\]

\[
= \int_0^1 |g(\psi^{b_p}(x)) - g(x)|^2dx
\]

\[
= \sum_{j=0}^{b_{n_0}-1} \int_{j/b_{n_0}}^{(j+1)/b_{n_0}} |g(\psi^{b_p}(x)) - g(x)|^2dx
\]

\[
\leq \sum_{j=0}^{b_{n_0}-1} \int_{j/b_{n_0}}^{(j+1)/b_{n_0}} \varepsilon dx
\]

\[
= \varepsilon.
\]

It follows that if \( p \geq n_0 \) then by Lemma 3.2

\[
||f \circ \psi^{b_p} - f||^2_2 \leq ||f \circ \psi^{b_p} - g \circ \psi^{b_p}||^2_2 + ||g \circ \psi^{b_p} - g||^2_2 + ||g - f||^2_2
\]

\[
\leq ||(f - g) \circ \psi^{b_p}||^2_2 + \varepsilon + ||g - f||^2_2
\]

\[
= ||U_{\psi^{b_p}}(f - g)||^2_2 + \varepsilon + ||g - f||^2_2
\]

\[
= ||g - f||^2_2 + \varepsilon + ||g - f||^2_2
\]

\[
< \varepsilon + \varepsilon + \varepsilon
\]

116
Thus,

\[ \lim_{p \to \infty} \| f \circ \psi^p - f \|_2 = 0, \]

for all \( f \in L^2([0,1]) \) and hence

\[ \lim_{p \to \infty} \| f \circ \psi^p - f \|_2^2 = 0. \]

for all \( f \in L^2([0,1]) \). That is

\[ \lim_{p \to \infty} \int_X |f \circ \psi^p - f|^2 d\mu = 0 \]

for all \( f \in L^2([0,1]) \).

\[ \diamondsuit \]

**Lemma 4.9**

Let \( T \) be an ergodic transformation on a probability space \((X, B, \mu)\) and \((\tilde{X}, \tilde{B}, \tilde{\mu}, \tilde{T})\) be a usual finite level tower extension for \((X, B, \mu, T)\). Suppose that there is an eigenvalue, \( \lambda \) with an eigenfunction \( f \in L^2(\tilde{X}, \tilde{B}, \tilde{\mu}) \) so that

\[ U_T(f)(x) = e^{2\pi i \lambda} f(x). \]

Then \( |f| \) is constant \( \tilde{\mu} \)-a.e.

**Note:** As yet only the usual two level tower extension has been defined. Hence, for the purposes of this chapter the Lemma should only be read with the two level extension in mind. However, the proof is identical for higher levels and so, to prevent repetition in later chapters, the general case is presented now.

**Proof:**

Note that for each \( x \in X \) and \( n \in \mathbb{N} \) there is a \( p(n, x) \in \mathbb{N} \) such that

\[ f(T^n(x)) = f(\tilde{T}^{p(n,x)}(x)) = e^{2\pi i \lambda} f(\tilde{T}^{p(n,x)-1}(x)) = \ldots = e^{2\pi i p(n,x) \lambda} f(x). \]
Now suppose that $|f|$ is not constant $\bar{\mu}$-a.e. Then there is an $a \in \mathbb{R}$ with two associated subsets $A$ and $C$ of $\bar{X}$ with

$$\bar{\mu}(A) > 0, \bar{\mu}(C) > 0,$$

$$|f(x)| \leq a \quad \text{for all } x \in A$$

and

$$|f(x)| > a \quad \text{for all } x \in C.$$  

As $\bar{\mu}(A) > 0$ there must be a level of the tower $B_i$ such that $\bar{\mu}(A \cap B_i) > 0$. Then

$$\bar{T}^{-l}(A \cap B_i) = \sigma^{l-1}_{i=0} \tau_i^{-1}(A \cap B_i) \subset X.$$  

We then have that

$$\mu(\bar{T}^{-l}(A \cap B)) = \bar{\mu}(\sigma^{l-1}_{i=0} \tau_i^{-1}(A \cap B_i))$$

$$= \bar{\mu}(A \cap B_i)$$

$$> 0.$$  

Also, for all $x \in \bar{T}^{-l}(A \cap B_i)$, $\bar{T}^l(x) \in A \cap B_i$ so that as

$$f(\bar{T}^l(x)) = e^{2\pi ip(l,x)\lambda}f(x),$$

$$|f(\bar{T}^l(x))| = |e^{2\pi ip(l,x)\lambda}f(x)| = |f(x)| \leq a.$$  

We therefore have that there is a set $A_1$, $T^{-l}(A \cap B_i) \subseteq A_1$, such that $A_1 \subset X$, $\mu(A_1) > 0$ and $|f(x)| \leq a$ for all $x \in A_1$.

Similarly, due to the fact that $\bar{\mu}(C) > 0$, we can find a set $C_1 \subset X$ such that $\mu(C_1) > 0$ and $|f(x)| > a$ for all $x \in C_1$.

Now, as $T$ is ergodic, there is an $n \in \mathbb{N}$ such that

$$\mu(T^{-n}A_1 \cap C_1) > 0.$$
so that there is an \( x \in C_1 \) such that \( T^n(x) \in A_1 \). Thus, by the properties of \( C_1 \), \( |f(x)| > a \), and by the properties of \( A_1 \), \( |f(T^n(x))| \leq a \). However, we also know that

\[
|f(T^n(x))| = |e^{2\pi i p(n,x) \lambda} f(x)| = |f(x)| > a.
\]

This contradiction proves that \( |f| \) must be constant \( \tilde{\mu} \)-a.e.

\[\Diamond\]

**Theorem 4.5**

\( \tilde{\psi} \) is weak but not strong mixing

**Proof:**

We will prove that \( \tilde{\psi} \) is weak mixing via the use of Theorem 3.10. That is, we will show that \( \tilde{\psi} \) has a continuous spectrum.

Suppose that there is an eigenvalue, \( \lambda_0 \), of \( \tilde{\psi} \) not equal to one, with an eigenfunction \( f \in L^2(\tilde{X}, \tilde{B}, \tilde{\mu}) \). As we know \( |\lambda_0| \neq 1 \) there is a \( \lambda \in [0, 1) \) such that \( \lambda_0 = e^{2\pi i \lambda} \). As \( \lambda_0 \neq 1 \), \( \lambda \neq 0 \). We therefore have

\[
U_{\tilde{\psi}}(f) = \lambda_0 \cdot f
\]

and hence

\[
f(\tilde{\psi}(x)) = e^{2\pi i \lambda} f(x).
\]

Then by Lemma 4.9 the magnitude of \( f \) is constant. The actual magnitude thus has no effect on our argument and so we can assume \( |f(x)| = 1 \) for all \( x \in \tilde{X} \).

We now define a sequence of functions that will tell us how many iterations of \( \tilde{\psi} \) are required to have the same effect as \( n \)-iterations of \( \psi \) on a point in \( X \). For \( n = 1 \) we define

\[
u = \chi_x + \chi_A
\]
so that if $x$ is in $A$, $u(x) = 2$ and

$$\tilde{\psi}^2(x) = \tilde{\psi}(\tau(x)) = \psi(\tau^{-1}(\tau(x))) = \psi(x).$$

Also, if $x \in X - A$, $u(x) = 1$ and $\psi(x) = \psi(x)$. Thus, $u$ is the required function for $n = 1$. That is $\tilde{\psi}^u(x)(x) = \psi(x)$ for each $x \in X$. Then, by simply adding together $u(\psi^m(x))$ for each $0 \leq m \leq n - 1$ we can obtain a function $u_n : X \rightarrow \mathbb{N}$ that gives us the number of iterations of $\tilde{\psi}$ required to obtain $\psi^n(x)$. That is, we define

$$u_n(x) = \sum_{m=0}^{n-1} u(\psi^m(x)),$$

and note that $\tilde{\psi}^{u_n(x)}(x) = \psi^n(x)$ for each $x \in X$. Hence

$$f(\psi^n(x)) = f(\tilde{\psi}^{u_n(x)}) = e^{2\pi i \lambda} f(\tilde{\psi}^{u_n(x)} - 1)$$

$$= e^{2\pi i u_n(x) \lambda} f(x). \quad (11)$$

Now we show that for any given $p \in \mathbb{N}$, $u_{2^p}$ can take only two values over all $x \in X$. First note that there are $2^{2p}$ dyadic intervals of order $2p$ in $X$ and, by Lemma 4.5, all but the dyadic interval $(1 - 1/2^{2p}, 1)$ are either a subset of $A$ or are disjoint from $A$. Further

$$\left(1 - \frac{2}{2^{2p}}, 1 - \frac{1}{2^{2p}}\right) = \left(1 - \frac{1}{2^{2p-1}}, 1 - \frac{1}{2^{2p}}\right) = I_{2p-1},$$

so that $((2^{2p} - 2)/2^{2p}, (2^{2p} - 1)/2^{2p})$ is disjoint from $A$. Thus

$$A - \left(\frac{2^{2p} - 1}{2^{2p}}, 1\right) = \bigcup_{n=0}^{p-1} I_{2n}$$

$$= \bigcup_{n=0}^{p-1} \left(1 - \frac{2^{2(p-n)}}{2^{2p}}, 1 - \frac{2^{2(p-n)-1}}{2^{2p}}\right)$$

$$= \bigcup_{n=0}^{p-1} \bigcup_{k=2^{2p-2^{2(p-n)-1}}} \left(\frac{k}{2^{2p}}, \frac{k+1}{2^{2p}}\right).$$
Similarly
\[ A^c = \left( \frac{2^{2p} - 1}{2^{2p}}, 1 \right) = \bigcup_{n=0}^{p-1} I_{2n+1} \]
\[ = \bigcup_{n=0}^{p-1} \left( 1 - \frac{2^{2(p-n)-1}}{2^{2p}}, 1 - \frac{2^{2(p-n-1)}}{2^{2p}} \right) \]
\[ = \bigcup_{n=0}^{p-1} \bigcup_{k=2^{2p-2(2(p-n)-1)}}^{2^{2p}-2^2(p-n)} \left( \frac{k}{2^{2p}}, \frac{k+1}{2^{2p}} \right) . \]

These unions imply that the number of dyadic intervals of order 2\( p \) contained in \( A \) is
\[
(2^{2p} - 2^{2(p-0)-1} - (2^{2p} - 2^{2(p-0)})) + \ldots + (2^{2p} - 2^{2(p-(p-1))-1} - (2^{2p} - 2^{2(p-(p-1))}))
\]
\[ = 2^{2p-1} + 2^{2p-3} + \ldots + 2 \]
and that the number of dyadic intervals of order 2\( p \) disjoint from \( A \) is
\[
(2^{2p} - 2^{2(p-0)-1} - (2^{2p} - 2^{2(p-0)})) + \ldots + (2^{2p} - 2^{2(p-(p-1))-1} - (2^{2p} - 2^{2(p-(p-1))-1}))
\]
\[ = 2^{2p-2} + 2^{2p-4} + \ldots + 1. \]

Denoting the number of intervals contained in \( A \) by \( I_A \) and the number disjoint from \( A \) by \( I_{A^c} \) we can see that
\[ I_A = 2^{2p-1} + 2^{2p-3} + \ldots + 2 \]
and
\[ I_{A^c} = 2^{2p-2} + 2^{2p-4} + \ldots + 1 \]
and since \( 2^{2p-1} + 2^{2p-3} + \ldots + 2 = 2(2^{2p-2} + 2^{2p-4} + \ldots + 1) \) we have that \( I_A = 2I_{A^c} \). That is \( 2p - 1 = I_A + I_{A^c} = 3I_{A^c} \) and hence \( I_A = 2(2p-1)/3 \) and \( I_{A^c} = (2p-1)/3 \). Since all of the \( 2p - 1 \) dyadic intervals of order \( 2p \) excluding \( (1-1/2^{2p}, 1) \) are either a subset of \( A \) or disjoint from \( A \), we now have that of these \( 2p - 1 \) dyadic intervals of order \( 2p \), two thirds are contained in \( A \) (on
which $u(x) = 2$) and one third are disjoint from $A$ (on which $u(x) = 1$). As the first $2^p$ iterations of any $x \in X - \{0, 1/2^p, ..., 1 - 1/2^p\}$ intersects each of these dyadic intervals exactly once we have that $u_{2^p}$ is equal to the sum of the value of $u$ on each of the dyadic intervals plus whatever value $u_{2^p}(x)$ takes on when the orbit of $x$ intersects $(1 - 1/2^p, 1)$ (which will be either a 1 or a 2). That is

$$u_{2^p}(x) \in \{2I_A + I_{A^c} + 1, 2I_A + I_{A^c} + 2\} = \left\{\frac{2(2^p - 1)}{3} + \frac{2^p - 1}{3} + 1, 2\frac{2(2^p - 1)}{3} + \frac{2^p - 1}{3} + 2\right\} = \left\{\frac{5}{3}(2^p - 1) + 1, \frac{5}{3}(2^p - 1) + 2\right\}.$$

As the orbit of each of the points $\{0, 1/2^p, ..., 1 - 1/2^p\}$ is

$$\{0, 1/2^p, ..., 1 - 1/2^p\}$$

the same argument will apply for these points. Let $M_p = (5(2^p - 1))/3$ so that for each $p \in \mathbb{N}$ and each $x \in X$ we have

$$u_{2^p}(x) \in \{M_p + 1, M_p + 2\}.$$

Now note that $(1 - 1/2^p, 1 - 1/2^{p+1}) = I_{2^p} \subset A$, hence $I_{2n} \subset A \cap (1 - 1/2^p, 1)$ for each $n \geq p$. Therefore

$$\mu\left(A \cap \left(1 - \frac{1}{2^p}, 1\right)\right) = \frac{1}{2} \sum_{i=0}^{\infty} \frac{\mu\left((1 - \frac{1}{2^p}, 1)\right)}{4^i} = \mu\left((1 - \frac{1}{2^p}, 1)\right) \frac{2}{3}.$$

Hence

$$\mu\{x : u_{2^p}(x) = M_p + 1\} = \frac{1}{3}$$

and

$$\mu\{x : u_{2^p}(x) = M_p + 2\} = \frac{2}{3}.$$
Using Lemma 4.8 and (11) we now have that

\[ 0 = \int_X \left| f(\psi^{2p}(x)) - f(x) \right|^2 d\mu \]
\[ = \int_X \left| e^{2\pi i \lambda \mu_2 p(x)} - 1 \right|^2 |f(x)|^2 d\mu \]
\[ = \left| e^{2\pi i \lambda(M_p + 1)} - 1 \right|^2 (1/3) + \left| e^{2\pi i \lambda(M_p + 2)} - 1 \right|^2 (2/3). \]

For this equation to hold we need \( \lambda (M_p + 1) \equiv 0 \pmod{1} \) and \( \lambda (M_p + 2) \equiv 0 \pmod{1} \).
Clearly this requires \( \lambda \equiv 0 \pmod{1} \) which is a contradiction of our assumption that \( \lambda \neq 0 \). Therefore \( \lambda = 0 \) and hence \( \lambda_0 = 1 \). To show that \( \tilde{\psi} \) has a continuous spectrum, giving \( \tilde{\psi} \) as weak mixing we now only need show that \( f \) must be a constant function.

Suppose that \( f \) is an eigenfunction so that since \( \lambda_0 = 1 \) we have

\[ f(\tilde{\psi}(x)) = f(x) \text{ } \tilde{\mu} - \text{a.e.} \]

Now suppose that \( f \) is not constant \( \tilde{\mu} \)-a.e. then there exists a set, \( B \), such that \( \tilde{\mu}(B) > 0 \) and \( \tilde{\mu}(B^c) > 0 \) with the property that \( f(B) \cap f(B^c) = \emptyset \).

Without loss of generality suppose that \( \tilde{\mu}(B \cap X) > 0 \).

If \( \tilde{\mu}(B^c \cap X) = 0 \), then \( \tilde{\mu}(B^c \cap A^c) > 0 \) so that \( \tilde{\mu}(\tilde{\psi}^{-1}B^c \cap X) > 0 \). However, as all \( x \in \tilde{\psi}^{-1}B^c \) have the property that \( \tilde{\psi}(x) \in B^c \) and hence that \( f(x) = f(\tilde{\psi}(x)) \in f(B^c) \) we must have that \( x \in B^c \) so that \( \tilde{\mu}(B^c \cap X) > 0 \).

Since \( \psi \) is ergodic there exists an \( n \in \mathbb{N} \) such that \( \mu(\psi^{-n}B^c \cap B) > 0 \).
Let \( x \in \psi^{-n}B^c \cap B \), then \( \psi^n(x) \in B \) and hence \( \tilde{\psi}^{\mu_n}(x) \in B \). Thus as \( f(\tilde{\psi}^{\mu_n}(x)) = f(x) \), \( x \in B \), \( \tilde{\psi}^{\mu_n}(x) \in B^c \) and hence \( f(B) \cap f(B^c) \neq \emptyset \).

This contradiction means that \( f \) must be a constant function.
To show that \( \tilde{\psi} \) is not strong mixing, we note that from the above definitions of \( u_n \) and \( M_p \)

\[
\tilde{\psi}^{M_p}(x) = \tilde{\psi}^{-1}(\psi^{2^{2p}}(x)) \text{ or } \tilde{\psi}^{-2}(\psi^{2^{2p}}(x)).
\]

Thus, for any \( B \in B \)

\[
\tilde{\psi}^{M_p}(B) \subseteq \tilde{\psi}^{-1}(\psi^{2^{2p}}(B)) \cup \tilde{\psi}^{-2}(\psi^{2^{2p}}(B)).
\]

Now set \( B = (1/2, 1/4) \). Then for all \( p \in \mathbb{N} \) we know from Lemma 4.4 that \( \psi^{2^{2p}}(B) \equiv B \) so that \( \tilde{\psi}^{M_p}(B) \subset (0, 1/4) \cup \tau((0, 1/4)) \) and hence

\[
\mu(\tilde{\psi}^{M_p}(B) \cap B) = 0 \neq \mu(B)^2.
\]

Then, as \( \tilde{\psi} \) is \( \tilde{\mu} \) measure preserving and one-to-one, \( \mu(B) = \mu(\tilde{\psi}(B)) \) for all \( B \in B \). Hence

\[
\lim_{n \to \infty} \mu(\tilde{\psi}^{-n}(B) \cap B) = \lim_{n \to \infty} \mu(\tilde{\psi}^{-n}(\tilde{\psi}^{-n}B \cap B)) \\
= \lim_{n \to \infty} \mu(B \cap \tilde{\psi}^n(B)) \\
\neq \mu(B)\mu(B).
\]

Thus, by definition \( \tilde{\psi} \) is not strong mixing. \( \diamond \)

We, now that we are in a position to give the proof that \( T(z) = z^2 \) on \( T \) is strong mixing. We do this in the form of a Lemma and a Theorem.

When \( A \) is a set, we will use the notation \( A + s = \{x + s : x \in A\} \) and \( rA = \{rx : x \in A\} \).

**Lemma 4.10**

Let the transformation \( T \) be defined on the unit interval \( \mathbb{I} \) with the Borel \( \sigma \)-algebra \( B \) and Lebesgue measure \( \mu \) by

\[
T(x) = 2x \text{ (mod 1)} \quad \text{for all } x \in \mathbb{I}.
\]
Then, for each \( A \in \mathcal{B} \) and each \( n \in \mathbb{N} \),

\[
T^{-n}A = \bigcup_{j=0}^{2^n-1} (2^{-n}A + j2^{-n}).
\]

Consequently, for each dyadic interval \( J \) of order \( n \)

\[
\mu(T^{-n}A \cap J) = \mu(A)\mu(J)
\]

Proof:

We prove this Lemma inductively. For \( n = 1 \) we note that for each \( x \in [0,1) \),

\( T^{-1}(x) = \{ x/2, (x+1)/2 \} \). Hence

\[
T^{-1}A = \left\{ \frac{x}{2} : x \in A \right\} \cup \left\{ \frac{x}{2} + \frac{1}{2} : x \in A \right\}
\]

\[
= \frac{1}{2}A \cup \left( \frac{1}{2}A + \frac{1}{2} \right)
\]

\[
= \left( T^{-1}A \cap \left[ 0, \frac{1}{2} \right) \right) \cup \left( T^{-1}A \cap \left[ \frac{1}{2}, 1 \right) \right)
\]

We also note that

\[
\mu \left( T^{-1}A \cap \left[ 0, \frac{1}{2} \right) \right) = \mu \left( \frac{1}{2}A \right)
\]

\[
= \frac{\mu(A)}{2}
\]

\[
= \mu(A)\mu \left( \left[ 0, \frac{1}{2} \right) \right),
\]

and

\[
\mu \left( T^{-1}A \cap \left[ \frac{1}{2}, 1 \right) \right) = \mu \left( \frac{1}{2}A \right)
\]

\[
= \frac{\mu(A)}{2}
\]

\[
= \mu(A)\mu \left( \left[ \frac{1}{2}, 1 \right) \right),
\]
so that the claim is true for \( n = 1 \). Now assume that the claim is true for \( n = k \), so that

\[
T^{-k}A = \bigcup_{i=0}^{2^k-1} (2^{-k}A + i2^{-k}).
\]

Thus, using the fact that the claim is true for \( n = 1 \), we have

\[
T^{-k-1}A = T^{-1}(T^{-k}A)
\]

\[
= T^{-1} \left( \bigcup_{i=0}^{2^{k-1}} (2^{-k}A + i2^{-k}) \right)
\]

\[
= \frac{1}{2} \left( \bigcup_{i=0}^{2^{k-1}} (2^{-k}A + i2^{-k}) \right) \cup \left( \frac{1}{2} \left( \bigcup_{i=0}^{2^{k-1}} (2^{-k}A + i2^{-k}) \right) + \frac{1}{2} \right)
\]

\[
= \bigcup_{i=0}^{2^{k-1}} (2^{-k-1}A + i2^{-k-1}) \cup \bigcup_{i=0}^{2^{k-1}} (2^{-k-1}A + (i + 2^{k})2^{-k-1})
\]

\[
= \bigcup_{i=0}^{2^{k+1}-1} (2^{-k-1}A + i2^{-k-1})
\]

\[
= \bigcup_{i=0}^{2^{k-1}} A_i,
\]

where \( A_i = 2^{-k-1}A + 2^{-k-1} \). Note that each \( A_i \subset [i/2^{k+1}, (i + 1)/2^{k+1}) \) and thus disjoint from each of the other dyadic intervals of order \( k + 1 \) so that

\[
\mu \left( T^{-k-1}A \cap \left[ \frac{i}{2^{k+1}}, \frac{i+1}{2^{k+1}} \right) \right) = \mu(A_i)
\]

\[
= \mu(2^{-k-1}A + 2^{-k-1})
\]

\[
= 2^{-k-1} \mu(A).
\]

\[
= \mu(A) \mu \left( \left[ \frac{i}{2^{k+1}}, \frac{i+1}{2^{k+1}} \right) \right).
\]

Therefore the hypothesis holds for \( n = k + 1 \) proving the Lemma.

\[\diamondsuit\]

We are now able to prove that the transformation \( T(z) = z^2 \) on \( \mathbb{T} \) is strongly mixing. This theorem concludes the chapter.
Theorem 4.6

The transformation, $T$, defined on the complex unit circle $\mathbb{T}$ by

$$T(z) = z^2$$

is strong mixing.

Proof:

Let $A, B \in \mathcal{B}$. Then

$$\mu(A) = \inf \left\{ \sum_{j=1}^{k} \mu(I_j), A \subseteq \bigcup_{j=1}^{k} I_j, I_j \text{ are intervals} \right\}$$

$$= \inf \left\{ \sum_{j=1}^{k} \mu(J_j), A \subseteq \bigcup_{j=1}^{k} J_j, J_j \text{ are dyadic intervals of the same order} \right\}.$$  

Suppose that $\varepsilon > 0$, then by the above observation there must exist disjoint dyadic intervals, all of the same order, say $n_0$, such that if $C = \bigcup_{j=1}^{k} J_j$ then

$$B \subseteq C$$

and

$$\mu(B) \leq \mu(C) \leq \mu(B) + \varepsilon.$$  

Since $C \in \mathcal{B}$ is measurable and contains $B$, we note that

$$\mu(C) = \mu(B) + \mu(C - B)$$

and hence

$$\mu(C - B) < \varepsilon.$$  

Also since any dyadic interval can be written as the disjoint union of dyadic intervals of any higher order common to each interval, we can assume that the common order is $n$ for any $n > n_0$.  

127
By Lemma 4.10 we know that

\[ \mu(T^{-n}A \cap J) = \mu(A)\mu(J) \]

if \( J \) is a dyadic interval of order \( n \). Hence

\[
\mu(T^{-n}A \cap C) = \sum_{j=1}^{k} \mu(T^{-n}A \cap J_j)
\]

\[
= \mu(A) \sum_{j=1}^{k} \mu(J_j)
\]

\[
= \mu(A)\mu(C).
\]

Since \( B \subseteq C \) this gives us that

\[
\mu(T^{-n}A \cap B) < \mu(T^{-n}A \cap C)
\]

\[
= \mu(A)\mu(C)
\]

\[
\leq \mu(A)(\mu(B) + \varepsilon)
\]

\[
\leq \mu(A)\mu(B) + \varepsilon.
\]

As this is true for all \( \varepsilon > 0 \) we have that

\[
\limsup_{n \to \infty} \mu(T^{-n}A \cap B) \leq \mu(A)\mu(B). 
\]  \hspace{1cm} (12)

Next, we note that

\[
\mu(T^{-n}A \cap B) + \mu(T^{-n}A \cap (C - B)) = \mu(T^{-n}A \cap C)
\]

and that

\[
\mu(T^{-n}A \cap (C - B)) \leq \mu(C - B) < \varepsilon,
\]

and therefore that

\[
\mu(T^{-n}A \cap B) > \mu(T^{-n}A \cap C) - \varepsilon
\]

\[
= \mu(A)\mu(C) - \varepsilon
\]

\[
\geq \mu(A)\mu(B) - \varepsilon.
\]
Since this is true for any $\varepsilon > 0$

$$\liminf_{n \to \infty} \mu(T^{-n} A \cap B) \geq \mu(A)\mu(B),$$

so that combining with (12) we have

$$\mu(A)\mu(B) \leq \liminf_{n \to \infty} \mu(T^{-n} A \cap B) \leq \limsup_{n \to \infty} \mu(T^{-n} A \cap B) \leq \mu(A)\mu(B).$$

Therefore

$$\lim_{n \to \infty} \mu(T^{-n} A \cap B) = \mu(A)\mu(B).$$

As $A, B$ were arbitrarily chosen from $\mathcal{B}$ the Theorem is proved. \hfill \diamond

4.3 Notes

The discussion on irrational rotations is based on Nillsen [32]. Lemmas 4.1 and 4.2 and Theorem 4.1 come from the same work and Lemma 4.7 is a generalisation of Lemma 4.2 in [32]. The example of a weak but not strong mixing transformation was given by Kakutani [21] who proves that it is weak mixing but not strong mixing though the proof we give (Theorem 4.5) is a variation and significant extension of his proof. Lemma 4.8 used in Theorem 4.5 was developed in communication with Rodney Nillsen. The proof for Proposition 4.1 is our own, though the result is not original, see, for example, Berberian [4]. Similarly, the proofs for Theorems 4.2 and 4.6 are our own, but the results are already known, a discussion of these and further similar results can be found in Katok [24]. The remaining results, Lemmas 4.3, 4.4, 4.5, 4.6, 4.9, 4.10 and Theorems 4.3 and 4.4 are original, although Lemmas 4.3, 4.5 and 4.6 are first presented in [25] and [26]. As mentioned in the chapter, the concept of an interval of density is based loosely on the idea of a point of density. For further discussion on the original points of density, see Federer [11] or Morgan[29].
5 Generalisations, Towers and the rise of Splinters

In this chapter we look at the emergence of the final important tool to be used in proofs relating to the generation of Carathéodory's definition. That is, we discuss how sets are 'splintered' under iterations (or inverse iterations) of a transformation. The occurrence of 'splintering' has already been seen in the breaking up of dyadic intervals under iterations of Kakutani's $\tilde{\psi}$.

We arrive at the investigation of splinters through the investigation of towers after presenting some generalisations of the transformations considered so far that assist in refining and highlighting the important aspects of the transformations in terms of what allows these transformations to generate Carathéodory's definition. We find in the initial generalisations that intervals of density and orbits are indeed important.

We also find that tower extensions appear to have little effect on whether a transformation will generate the Carathéodory definition or not, but have a significant effect on the mixing properties of the transformation. The effect of general tower extensions is therefore investigated. It is from this investigation that 'splinters' are shown to be important. Recalling that the extension to the tower transformation $\tilde{\psi}$ from the primitive transformation $\psi$ was the cause of the important 'splintering' of sets it should be no surprise that the concept of 'splintering' is developed further when the effect of derived and induced tower or skyscraper transformations is considered.

The concept of 'splintering' arises in the midst of a number of other results that were found during the same stage of our research that we consider interesting in their own right. These other results are almost all our own.
(those that are not will be noted as such) and, as mentioned, come from
the first generalised results or results allowing generalisation instead of the
consideration of specific examples.

5.1 Generalisations

The first and most obvious choice with which to commence our investigation
is to consider the other weak but not strong mixing transformation given in
the same paper as the first. Especially as it is itself an extension of a special
case of the irrational rotation already discussed.

Part of the construction depends on a particular number, \( \alpha \), being what
is known as a Liouville number; a real number \( x \) is a Liouville number if for
each \( m \in \mathbb{N} \) there exists a rational number \( p/q \) such that

\[
0 < \left| x - \frac{p}{q} \right| < \frac{1}{q^m}.
\]

While Liouville numbers are an interesting field of study, the fact that \( \alpha \) is
Liouville is only of interest in this research because a Liouville number is irra­
tional as this property allows strong comparison between the transformation
currently under investigation and the irrational rotation already discussed.

We can now construct the example as follows.

Let \( (X, \mathcal{B}, \mu) \) be a measure space where \( X = [0, 1) \), and \( (X, \mathcal{B}, \mu) \) is the
usual Lebesgue measure space on \( X = [0, 1) \). Let

\[
\alpha = \sum_{k=1}^{\infty} 10^{-n_k}
\]

where

\[
\{n_k : k \in \mathbb{N}\}
\]
is an increasing sequence of positive integers such that

\[
\lim_{k \to \infty} (n_{k+1} - 2n_k) = +\infty.
\]

Such an \(\alpha\) is Liouville and hence irrational. (For a proof of the fact that \(\alpha\) is Liouville and therefore irrational see, for example, [33].) Thus when we let \(\psi : X \to X\) be defined by

\[
\psi(x) = x + \alpha \pmod{1}
\]

we note that \(\psi = P_\alpha\) in the notation of Chapter 4. Therefore, for consistency of notation we will denote the transformation by \(P_\alpha\).

We then 'extend' the function to a two level tower transformation as we did for Kakutani's first example as follows. Let

\[
\beta = \sum_{n=1}^{\infty} b_n 10^{-n}
\]

where \(b_n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}\) and \(b_{n_{k+1}} = 5\), for all \(k \in \mathbb{N}\). Then let \(A = \{0, \beta\}\) and extend the original function in the usual way. That is, let \(A'\) be a set such that \(X \cap A' = \emptyset\) and let \(\tau : A \to A'\) be a 1-1 and onto function. Let \(\bar{X} = X \cup A'\) and

\[
\bar{B} = \{ B : B \subseteq \bar{X}, B \cap X \in B \text{ and } \tau^{-1}(B \cap A') \in B \}.
\]

If \(B \subseteq \bar{X}\), let \(\tilde{\mu}(B) = \mu(B \cap X) + \mu(\tau^{-1}(B \cap A'))\) and let \(\tilde{P}_\alpha : \bar{X} \to \bar{X}\) be defined by

\[
\tilde{P}_\alpha(x) = \begin{cases} 
P_\alpha(x) & x \in X \cap A^c, \\
\tau(x) & x \in A, \\
P_\alpha(\tau^{-1}(x)) & x \in A'.
\end{cases}
\]

Also let \(\mu^*\) and \(\tilde{\mu}^*\) be the associated outer measures for \(\mu\) and \(\tilde{\mu}\) respectively.

As \(\beta\) is not even a particular given constant, it is not at all clear how
we might be able to use the rotating argument in proving that $P_\alpha$ generates Carathéodory's definition. While it is possible to use, as we did for Kakutani's first example, a truncation of the orbit of an interval under the primitive transformation $P_\alpha$, we use an alternative method using properties of tower transformations that demonstrates how considering transformations as tower transformations may in fact simplify matters. The properties that we are particularly referring to are proven below and allow us to show that, at least in a two level situation, it is in fact sufficient to consider the primitive transformation when trying to prove that a transformation generates Carathéodory's definition.

**Lemma 5.1**

Let $(X, \mathcal{B}, \mu)$ be a measure space of finite measure and $T$ be a measure preserving invertible transformation. Let $(\hat{X}, \hat{\mathcal{B}}, \hat{\mu})$ and $\hat{T}$ be a usual two level tower extension of $(X, \mathcal{B}, \mu, T)$. If $B \subseteq \hat{X}$, $B$ is $\hat{T}$-invariant and $\hat{\mu}_*(B) > 0$ then $\mu_*(B \cap X) > 0$.

**Proof:**

Let the tower be defined using the same notation as in the preceding discussion. That is, let \( \hat{X} = X \cup A' \) where \( X \cap A' = \emptyset \) and there is a one to one and onto function $\tau : A \mapsto A'$ where $A \in \mathcal{B}$. Then

\[
0 < \hat{\mu}_*(B) = \hat{\mu}_*(B \cap (X \cup A'))
\]

\[
= \mu_*(B \cap X) + \hat{\mu}_*(B \cap A').
\]

Thus, at least one of $\mu_*(B \cap X)$ and $\hat{\mu}_*(B \cap A')$ are strictly positive. If $\mu_*(B \cap X) > 0$ there is nothing more to prove. Otherwise, as $\hat{T}$ is measure preserving and $B$ is $\hat{T}$-invariant we have that

\[
\mu_*(B \cap X) \geq \mu_*(B \cap T(A))
\]

\[
= \hat{\mu}_*(B \cap \hat{T}(A')).
\]
\[
\begin{align*}
= \mu_*(\tilde{T}(B \cap A')) \\
= \mu_*(B \cap A') \\
> 0.
\end{align*}
\]

Lemma 5.2

Let \((X, B, \mu)\) be a measure space of finite measure and \(T\) be a measure preserving invertible transformation. Let \((\tilde{X}, \tilde{B}, \tilde{\mu})\) and \(\tilde{T}\) be the usual two level tower extension as described above. If \(B\) is \(\tilde{T}\)-invariant then \(B \cap X\) is \(T\) - invariant. That is

\[
\tilde{T}^{-1}(B) = B \Rightarrow T^{-1}(B \cap X) = B \cap X.
\]

Proof:

Let \(B\) be a \(\tilde{T}\)-invariant subset of \(\tilde{X}\). That is, \(\tilde{T}^{-1}(B) = B\). Also as \(\tilde{T}\) is one-to-one and onto, \(B\) is also \(\tilde{T}^{-1}\)-invariant. That is \(\tilde{T}(B) = B\).

Let \(x \in T^{-1}(B \cap X)\) so that \(T(x) \in B \cap X\).

Then, clearly, \(x \in X\) and if \(x \in (X \cap A^c)\), \(\tilde{T}(x) = T(x) \in B \cap X \subset B\). Thus as \(B\) is \(\tilde{T}\)-invariant, \(x \in B\). That is \(x \in B \cap X\). Also if \(x \in A\), clearly \(x \in X\) and also \(\tilde{T}^2(x) = T(x) \in B \cap X \subset B\) and thus \(x \in B\) as \(B\) is \(\tilde{T}\) - invariant. That is, we now have that \(x \in T^{-1}(B \cap X) \Rightarrow x \in B \cap X\), meaning that \(T^{-1}(B \cap X) \subset B \cap X\).

Conversely, if \(x \in B \cap X\) then \(x \in X\) and hence \(T(x) \in X\). Also \(x \in B\) and hence if \(x \in A\), \(T(x) = \tilde{T}^2(x) \in B\) as \(B\) is \(\tilde{T}^{-1}\)-invariant. Finally, with \(x \in B\), if \(x \notin A\), \(T(x) = \tilde{T}(x) \in B\). That is

\[
x \in B \cap X \Rightarrow T(x) \in B \cap X \Rightarrow x \in T^{-1}(B \cap X),
\]
meaning that $B \cap X \subset T^{-1}(B \cap X)$.

We now have that $T^{-1}(B \cap X) = B \cap X$. \hfill \diamondsuit

In the following theorem it becomes clear that the above lemmas mean that whether this two level tower transformation generates Carathéodory's definition or not depends only on the primitive transformation. Since we know from Chapter 4 that the primitive transformation, which in the present context is $P_\alpha$, does generate the definition, so must $\tilde{P}_\alpha$.

**Theorem 5.1**

*Let $B$ be a $\tilde{P}_\alpha$-invariant subset of $\tilde{X}$. Suppose that there is a $\theta < 2$ such that for all subintervals $V$ of $\tilde{X}$

$$\tilde{\mu}_* (B \cap V) + \tilde{\mu}_* (B^c \cap V) \leq \theta \tilde{\mu}(V).$$

Then, either $B$ or $B^c$ is a set of measure zero.*

**Proof:**

Suppose that $B$ is a $\tilde{P}_\alpha$-invariant subset of $\tilde{X}$ such that $\tilde{\mu}_*(B) > 0$ and that $\tilde{\mu}_*(B^c) > 0$. Then, by Lemma 5.1, $\mu_*(B \cap X) > 0$ and $\mu_*(B^c \cap X) > 0$. Then using Lemma 4.1 and Lemma 4.2, we see that for any $\epsilon > 0$ there exist intervals $J_\epsilon$ and $K_\epsilon$ with $\mu(J_\epsilon) = \mu(K_\epsilon)$ such that

$$\mu_*(B \cap X \cap J_\epsilon) > (1 - \epsilon) \mu(J_\epsilon)$$

and

$$\mu_*(B^c \cap X \cap K_\epsilon) > (1 - \epsilon) \mu(K_\epsilon).$$

Further, by Lemma 4.3, for any $\delta > 0$ there is an $n \in \mathbb{N}$ such that

$$\mu(P^n_\alpha(J_\epsilon) \Delta K_\epsilon) < \delta.$$
Now we observe that

\[ B^c \cap X \cap K_\epsilon \subseteq (B^c \cap X \cap P_\alpha^{[-n]}(J_\epsilon)) \cup (P_\alpha^{[-n]}(J_\epsilon) \Delta K_\epsilon), \]

so that

\[ \tilde{\mu}_*(B^c \cap X \cap K_\epsilon) \leq \tilde{\mu}_*(B^c \cap X \cap P_\alpha^{[-n]}(J_\epsilon)) + \tilde{\mu}(P_\alpha^{[-n]}(J_\epsilon) \Delta K_\epsilon). \]

Using this fact we now have that

\[
\begin{align*}
\tilde{\mu}_*(B^c \cap X \cap J_\epsilon) &= \tilde{\mu}_*(P_\alpha^{[-n]}(B^c \cap X \cap J_\epsilon)) \quad \text{as } P_\alpha \text{ is measure preserving,} \\
&= \tilde{\mu}_*(P_\alpha^{[-n]}(B^c \cap X) \cap P_\alpha^{[-n]}(J_\epsilon)) \\
&= \tilde{\mu}_*((B^c \cap X) \cap P_\alpha^{[-n]}(J_\epsilon)) \\
&\geq \tilde{\mu}_*(B^c \cap X \cap K_\epsilon) - \tilde{\mu}(P_\alpha^{[-n]}J_\epsilon \Delta K_\epsilon) \\
&\geq \tilde{\mu}_*(B^c \cap X \cap K_\epsilon) - \delta \\
&> (1 - \epsilon)\tilde{\mu}(K_\epsilon) - \delta \\
&= (1 - \epsilon)\tilde{\mu}(J_\epsilon) - \delta.
\end{align*}
\]

As this is true for any \( \delta > 0 \) we have that

\[ \tilde{\mu}_*(B^c \cap X \cap J_\epsilon) \geq (1 - \epsilon)\tilde{\mu}(J_\epsilon). \]

Thus

\[ \tilde{\mu}_*(B \cap X \cap J_\epsilon) + \tilde{\mu}_*(B^c \cap X \cap J_\epsilon) \geq 2(1 - \epsilon)\tilde{\mu}(J_\epsilon). \]

Therefore, choosing \( \epsilon = \theta/4 \) contradicts the hypothesis, hence we cannot have both \( \tilde{\mu}_*(B) > 0 \) and \( \tilde{\mu}_*(B^c) > 0 \). That is, either \( \tilde{\mu}_*(B) = 0 \) or \( \tilde{\mu}_*(B^c) = 0 \). \( \diamond \)

Notice that the proof of Theorem 5.1 does not depend, in any way, on the subtleties of the choices of \( \alpha \) and \( \beta \). On observing how strongly the proof that \( P_\alpha \) is weak mixing depends on \( \alpha \) and \( \beta \) we can see that we may in fact be able to find a transformation capable of generating Carathéodory’s definition.
with stronger mixing properties than originally aimed for. Alternatively, it may be that $\alpha$ and $\beta$ are chosen in such a way as to make it impossible to get a transformation of the same form that has a stronger mixing property and that generally such a transformation is not mixing. Following the technical lemma, Lemma 5.3, which is similar to Lemma 4.8, we give the proof that $\tilde{P}_\alpha$ is weak mixing.

To make the proof of Lemma 5.3 easier we need to make the following definition.

**Definition 5.1**

A real valued function $f$ defined on a metric space $X$ is said to be **Lipschitz** if there is an $M$ such that

$$|f(x) - f(y)| \leq M|x - y|$$

for each $x, y \in X$. The least such constant $M$ is known as the **Lipschitz constant** for $f$ and is denoted by $\text{Lip}_f$.

**Lemma 5.3**

Let $(X, \mathcal{B}, \mu)$ be the usual Lebesgue measure space on $X = [0, 1)$. Then, for any $f \in L^2(X, \mathcal{B}, \mu)$,

$$\lim_{k \to \infty} \int_X |f \circ P^{10^n}_\alpha(x) - f(x)|^2 d\mu(x) = 0.$$

**Proof:**

Let $\varepsilon > 0$ and $f \in L^2(X, \mathcal{B}, \mu)$. and let $g$ be a Lipschitz function such that

$$\|f - g\|_2 < \varepsilon,$$

We know that such a $g$ exists as a polynomial is Lipschitz and the Weierstrass Approximation Theorem (see for example [47]) then gives us that there must
exist an appropriate polynomial, which we could select as \( g \). As \( g \) is Lipschitz, we know that if \( x, y \in [0, 1) \) and
\[
|x - y| < \frac{\varepsilon}{\text{Lip } g},
\]
then
\[
|g(x) - g(y)| < \varepsilon.
\]
Now, since \( \lim_{k \to \infty}(n_k - 2n_{k-1}) = \infty \) and \( \{n_k\} \) is increasing, we note, using the fractional part notation \( \{\} \), defined in Definition 4.3
\[
\lim_{k \to \infty} \{10^{n_k} \alpha\} = \lim_{k \to \infty} 10^{n_k} \sum_{j=k+1}^{\infty} 10^{-nj} \\
\leq \lim_{k \to \infty} 10^{n_k} \sum_{j=k+1}^{\infty} 10^{-2n_k-j} \\
= \lim_{k \to \infty} \sum_{j=k+1}^{\infty} 10^{-n_k-j} \\
= 0.
\]
So that we can find a \( k_0 \in \mathbb{N} \) such that for each \( k > k_0 \), \( \{10^{n_k} \alpha\} < \varepsilon/\text{Lip } g \).

Therefore for each \( k > k_0 \) and each \( x < 1 - \frac{\varepsilon}{\text{Lip } g} \),
\[
P_{\alpha}^{10^{n_k}}(x) = (x + 10^{n_k} \alpha)(\text{mod } 1) = x + 10^{n_k} \sum_{j=k+1}^{\infty} 10^{-nj}
\]
so that
\[
|P_{\alpha}^{10^{n_k}}(x) - x| < \frac{\varepsilon}{\text{Lip } g},
\]
and hence
\[
|g \circ P_{\alpha}^{10^{n_k}}(x) - g(x)| < \varepsilon.
\]
Now denoting \( [0, 1 - \varepsilon/\text{Lip } g) \) by \( \varepsilon_g \) we therefore have for each \( k > k_0 \),
\[
\|g \circ P_{\alpha}^{10^{n_k}} - g\|_2^2 = \int_X |g \circ P_{\alpha}^{10^{n_k}}(x) - g(x)|^2 d\mu(x)
\]
\begin{align*}
&= \int_{\epsilon_{g}} |g \circ P_{\alpha}^{10n^{k}}(x) - g(x)|^{2}d\mu(x) \\
&\quad + \int_{X - \epsilon_{g}} |g \circ P_{\alpha}^{10n^{k}}(x) - g(x)|^{2}d\mu(x) \\
&< \int_{\epsilon_{g}} \varepsilon^{2} + \int_{X - \epsilon_{g}} |g \circ P_{\alpha}^{10n^{k}}(x) - g(x)|^{2}d\mu(x) \\
&< \varepsilon^{2} + \int_{X - \epsilon_{g}} |g \circ P_{\alpha}^{10n^{k}}(x) - g(x)|^{2}d\mu(x) \\
&< \varepsilon^{2} + \frac{\varepsilon}{Lipg}(4 \sup |g(x)|^{2}).
\end{align*}

Hence,

\[
\lim_{k \to \infty} \|g \circ P_{\alpha}^{10n^{k}} - g\|_{2}^{2} = 0.
\]

Therefore there is a \( k_1 \in \mathbb{N} \) such that for each \( k > k_1 \),

\[
\|g \circ P_{\alpha}^{10n^{k}} - g\|_{2} < \varepsilon.
\]

Hence, for each \( k > k_1 \),

\[
\|f \circ P_{\alpha}^{10n^{k}} - f\|_{2} \leq \|f \circ P_{\alpha}^{10n^{k}} - g \circ P_{\alpha}^{10n^{k}}\|_{2} + \|g \circ P_{\alpha}^{10n^{k}} - g\|_{2} + \|g - f\|_{2} \\
\leq \|(f - g) \circ P_{\alpha}^{10n^{k}}\|_{2} + \varepsilon + \|f - g\|_{2} \\
= \|U_{P_{\alpha}^{10n^{k}}}(f - g)\|_{2} + \varepsilon + \|f - g\|_{2} \\
= \|g - f\|_{2} + \varepsilon + \|g - f\|_{2} \\
< \varepsilon + \varepsilon + \varepsilon \\
= 3\varepsilon,
\]

where \( U_{P_{\alpha}} \) is the operator induced by \( P_{\alpha} \) defined in Definition 3.2 so that the equality following \( U_{P_{\alpha}} \)'s use follows from Lemma 3.2. Thus

\[
\lim_{k \to \infty} \|f \circ P_{\alpha}^{10n^{k}} - f\|_{2} = 0
\]

for all \( f \in L^{2}(X, B, \mu) \), and hence

\[
\lim_{k \to \infty} \|f \circ P_{\alpha}^{10n^{k}} - f\|_{2}^{2} = 0
\]
for all \( f \in L^2(X, B, \mu) \). That is, for all \( f \in L^2(X, B, \mu) \)

\[
\lim_{k \to \infty} \int_X |f \circ P_\alpha^{10^n k}(x) - f(x)|^2 d\mu(x) = 0.
\]

\[\diamondsuit\]

**Theorem 5.2**

\( \tilde{P}_\alpha \) is weak but not strong mixing.

**Proof:**

We will again be using Theorem 3.10 to give us the result. That is, we will show that \( \tilde{P}_\alpha \) has a continuous spectrum. Also in this proof, the numbers \( \alpha \) and \( \beta \) will refer to the important values used in the definition of \( \tilde{P}_\alpha \) defined in (13) and (14).

Suppose that there is an eigenvalue \( \lambda \neq 1 \) with an eigenfunction \( f \in L^2(X, B, \mu) \) so that

\[ f(\tilde{P}_\alpha(x)) = e^{2\pi i \lambda} f(x). \]

Then by Theorem 4.2 \( P_\alpha \) is ergodic, so that Lemma 4.9 gives us that \( |f| \) is constant \( \tilde{\mu}\text{-a.e.} \), and hence for the remainder of this proof we may assume that \( |f(x)| = 1 \) for all \( x \in X \).

Now, just as in Theorem 4.6 we wish to define a function \( u_n : X \to \mathbb{N} \) that will define, for each \( x \in X \), the number of iterations of \( \tilde{P}_\alpha \) that are required to give \( P_\alpha^n(x) \). This is done by defining

\[ u(x) = \begin{cases} 
1, & x \in X - A, \\
2, & x \in A,
\end{cases} \]

and

\[ u_n(x) = \sum_{m=0}^{n-1} u(P_\alpha^m(x)). \]
This gives us that \( P^n_\alpha(x) = \tilde{P}_\alpha u_n(x) \) for each \( x \in X \), and hence that

\[
\begin{align*}
    f(P^n_\alpha(x)) &= f(\tilde{P}_\alpha u_n(x)) \\
    &= e^{2\pi i \lambda} f(\tilde{P}_\alpha u_n(x)^{-1}(x)) \\
    &\quad \vdots \\
    &= e^{2\pi i u_n(x)\lambda} f(x).
\end{align*}
\]

(15)

We now show that \( u_{10^n k} \) can take one of only two values over all \( x \in X \). That is, we need to show that the number of elements in the first \( 10^n k \) elements of \( O_{P_\alpha}(x) \) in \( A = [0, \beta) \) can be only one of two values for sufficiently large \( k \). \( k \) is considered to be sufficiently large when, for each \( k_0 \geq k \),

\[
n_{k_0} - 2n_{k_0-1} > 0.
\]

(16)

In this way, for each \( 1 \leq n \leq 10^n k \)

\[
0 < n \sum_{j=k+1}^{\infty} 10^{-n_j} < 2 \cdot 10^n k 10^{-n_k+1} \leq 2 \cdot 10^n k - 2n_k - 1 < 10^{-n_k}.
\]

(17)

The reason why this condition is important is revealed later.

To demonstrate that \( u_{10^n k} \) takes one of only two values for sufficiently large \( k \), we consider that

\[
\begin{align*}
u_{10^n k}(x) &= \sum_{m=0}^{10^n k - 1} u(P^n_\alpha(x)) \\
&= 10^n k + \sum_{m=0}^{10^n k - 1} \chi_{[0,\beta)}(P^n_\alpha(x)) \\
&= 10^n k + \sum_{m=0}^{10^n k - 1} \chi_{[0,\beta)}((x + m\alpha)(\text{mod } 1))
\end{align*}
\]

\[ 141 \]
Hence we need to find the cardinality of

\[ \{(x + m\alpha)(\mod 1) : 0 \leq m \leq 10^{n_k} - 1\} \cap [0, \beta) \].

It is here that we need to use (17). As, by (17) we know that for each element

\[ y \in \{(m\alpha)(\mod 1) : 0 \leq m \leq 10^{n_k} - 1\} \]

we can express \( y \) as

\[ y = (m(\alpha - (\alpha)(\mod 10^{-n_k}))) + m((\alpha)(\mod 10^{-n_k}))) \mod 1 \]

with \( m(\alpha - (\alpha)(\mod 10^{-n_k}))) < 10^{-n_k} \) for each \( 0 \leq m \leq 10^{n_k} - 1 \). Hence for each \( y \in \{(x + m\alpha)(\mod 1) : 0 \leq m \leq 10^{n_k} - 1\} \) we can consider \( y \) as \( y = y_m + y_b \), where \( y_m = m((\alpha)(\mod 10^{-n_k})) \) and \( y_b \) represents the small remaining parts \((m(\alpha - (\alpha)(\mod 10^{-n_k})))\).

Note that for each \( x \in [0, 1) \), \((x)(\mod 10^{-n_k})\) is a rational with denominator \(10^{n-k}\) which we will denote by \( x_d \) so that for each \( x \in X \) we can express \( x \) as

\[ x = x_d + x_b \]

where \( x_d \) is a rational with denominator \(10^{n_k}\) and \( 0 \leq x_b < 10^{-n_k} \).

To consider \( u_{10^{n_k}} \) we approach the problem by finding the cardinality of

\[ R_x = \{(x)(\mod 10^{-n_k}) + m((\alpha)(\mod 10^{-n_k}))(\mod 1) : 0 \leq m < 10^{n_k}\} \]

intersected with \([0, \beta)\). That is, we find the cardinality of

\[ R_x \cap [0, \beta) \]

and consider the effects of the remaining parts (the \( x_b \)'s) later.

As \( y_{10^{n_k}} = 0 = y_0 \),

\[ \{y_m : 0 \leq m < 10^{n_k}\} \]
is equal to

\[ \{ y_m : 1 \leq m \leq 10^{n_k} \} \]

and hence

\[ R_x = \{ (x_d + y_m)(\text{mod } 1) : 1 \leq m \leq 10^{n_k} \}. \]

Now, to find the cardinality of \( R_x \cap [0, \beta) \) we first find the cardinality of

\[ \{ y_m : 1 \leq m \leq 10^{n_k} \} \cap [0, \beta). \]

Note that in decimal (10-adic) notation we can write \( y_1 \) as

\[ y_1 = 0.a_1a_2...a_{n_k} \]

where \( a_i \in \{0,1\}, 1 \leq i \leq n_k - 1 \) and \( a_{n_k} = 1 \). (This follows from the definition of \( \alpha \) and the fact that \( y_1 = \alpha - (\alpha)(\text{mod } 10^{-n_k})) \). Hence, finding the cardinality of \( \{ y_m : 1 \leq m \leq 10^{-n_k} \} \) is the same as finding the cardinality of

\[ C_{n_k} = \{ 0.c_1c_2...c_{n_k} : 0.c_1c_2...c_{n_k} = (m \times 0.a_1...a_{n_k})(\text{mod } 1), 1 \leq m \leq 10^{n_k} \}. \]

Finding the elements of this set is equivalent to finding the elements of the modulo class, or subgroup, of \( \mathbb{Z}_{n_k}, < a_1a_2...a_{n_k} > \), generated by the element \([a_1a_2...a_{n_k}]\) and then dividing each element by \( 10^{n_k} \). As \( a_{n_k} = 1 \) we know that

\[ \gcd(a_1a_2...a_{n_k}, 10^{n_k}) = 1 \]

so that \( < a_1a_2...a_{n_k} > = \mathbb{Z}_{n_k} \) and thus

\[ C_{n_k} = \{ 0.c_1c_2...c_{n_k} : c_i \in \{0, 1, ..., 9\}, 1 \leq i \leq n_k \}. \]

That is,

\[ \{ y_m : 1 \leq m \leq 10^{n_k} \} = \{ 0.c_1c_2...c_{n_k} : c_i \in \{0, 1, ..., 9\}, 1 \leq i \leq n_k \}. \] (18)
That is, \( \{y_m : 1 \leq m \leq 10^{nk}\} \) is the set of \( 10^{nk} \)-adic rationals in \([0, 1)\). Thus clearly the cardinality of \( \{y_m : 1 \leq m \leq 10^{nk}\} \cap [0, \beta) \) is the set of rationals in \([0, \beta)\) with denominator \( 10^{nk} \) of which there are

\[
\left(10^{nk} \sum_{n=1}^{nk} b_n 10^{-n}\right) + 1,
\]

these being the rationals of denominator \( 10^{nk} \) that can be expressed as

\[
0.x_1x_2...x_{nk}
\]
such that

\[
0.00...0 \leq 0.x_1x_2...x_{nk} \leq 0.b_1b_2...b_{nk}.
\]

It is also clear that as \( \{y_m : 1 \leq m \leq 10^{nk}\} \) is the set of \( 10^{nk} \)-adic rationals modulo 1, for any rational of denominator \( 10^{nk} \), \( d \),

\[
\{(y_m + d) \pmod{1} : 1 \leq m \leq 10^{nk}\} = \{y_m : 1 \leq m \leq 10^{nk}\}.
\]

Hence, as \( x_d \) is a rational of denominator \( 10^{nk} \) for each \( x \in [0, 1) \), we have that, for each \( x \in [0, 1) \)

\[
R_x = \{y_m : 1 \leq m \leq 10^{nk}\}
\]

and hence the cardinality of \( R_x \cap [0, \beta) \) is

\[
\left(10^{nk} \sum_{n=1}^{nk} b_n 10^{-n}\right) + 1
\]

for each \( x \in [0, 1) \).

We now know that for each \( x \in [0, 1) \), \( u_{10^{nk}}(x) = 10^{nk}(1 + \sum_{n=1}^{nk} b_n 10^{-n}) + 1 \) plus the effect that the remainder parts, \( y_b \) and \( x_b \), that we discussed earlier have on the cardinality of

\[
\{(x + m\alpha) \pmod{1} : 0 \leq m \leq 10^{nk} - 1\} \cap [0, \beta).
\]
Hence we now look at the effect that these remaining parts can have.

By (18) we know that for each $x \in [0, 1)$ and $n \in \{0, 1, \ldots, 10^n k - 1\}$ there is an $m(n) \in \mathbb{N}$ such that
\[
y_{m(n)} + x_d = \frac{n}{10^n k}.
\]

Next, for each $x \in [0, 1)$ we write
\[
b_n(x) = (x - ((x) \text{mod} \ 10^{-n_k})) + m(n)(\alpha - ((\alpha) \text{mod} \ 10^{-n_k}))
\]
(which are the remaining parts, that is $x_b + y_b$) and
\[
e_n(x) = y_{m(n)} + x_d + b_n(x)
\]
(which expresses actual points that occur in $O_{\alpha}(x)$).

Note that
\[
(x - ((x) \text{mod} \ 10^{-n_k})) < 10^{-n_k}
\]
and by (17)
\[
m(n)(\alpha - ((\alpha) \text{mod} \ 10^{-n_k})) < 2 \cdot 10^{-n_k - 1}
\]
for each $0 \leq m(n) \leq 10^n k - 1$ so that we also have $b_n(x) < 1.2 \cdot 10^{-n_k}$ for each $x \in [0, 1)$.

We are now in a position to consider the effects that these remaining parts have on $u_{10^n k}(x)$.

Firstly, it is clear that $e_0(x) \in [0, \beta)$ and that
\[
0 < e_{b_1 \ldots b_{n_k} - 1}(x) \leq 0. b_1 b_2 \ldots b_{n_k} 2 < \beta
\]
so that $e_{b_1 \ldots b_{n_k} - 1}(x) \in [0, \beta)$ and hence, since $e_0(x) < e_i(x) < e_{b_1 \ldots b_{n_k} - 1}(x)$ for $i \in \{1, 2, \ldots, b_{n_k} - 2\}$, $e_0(x), \ldots, e_{b_1 \ldots b_{n_k} - 1}(x) \in [0, \beta)$.  

145
Similarly, it is clear that \( e_{b_1 \ldots b_{n_k+1}}(x), \ldots, e_{10^{n_k-2}}(x) \not\in [0, \beta) \).

Whether or not \( e_{10^{n_k-1}}(x) \) and \( e_{b_1 \ldots b_{n_k}}(x) \) are in \([0, \beta)\) depend on the value of \( x \) and there are three cases to consider.

**CASE I**

It is possible that \( e_{10^{n_k-1}}(x) \in [0, \beta) \). In this case

\[
x - ((x)(\mod 10^{-n_k})) > 9 \times 10^{-n_k-1}
\]

and so in this case \( e_{b_1 \ldots b_{n_k}}(x) > 0.b_1b_2 \ldots b_{n_k}9 > \beta \).

Therefore in this case \( e_{10^{n_k-1}}(x), e_0(x), e_1(x), \ldots, e_{b_1 \ldots b_{n_k-1}}(x) \) are the elements of \( O_{P_a}(x) \) that are in \([0, \beta)\) and hence

\[
u_{10^{n_k}}(x) = 10^{n_k} \left( 1 + \sum_{n=1}^{n_k} b_n 10^{-n} \right) + 1.
\]

**CASE II**

The case when \( e_{10^{n_k-1}}(x) \not\in [0, \beta) \) and that \( e_{b_1 \ldots b_{n_k}}(x) \in [0, \beta) \). This case will definitely occur when

\[
x - ((x)(\mod 10^{-n_k})) \in [0, 3 \cdot 10^{-n_k-1}),
\]

so that \( 0 < e_{b_1 \ldots b_{n_k}}(x) < 0.b_1b_2 \ldots b_{n_k}5 < \beta \). In this case

\[
O_{P_a}(x) \cap [0, \beta) = \{ e_0(x), \ldots, e_{b_1 \ldots b_{n_k}}(x) \}
\]

so that

\[
u_{10^{n_k}}(x) = 10^{n_k} \left( 1 + \sum_{n=1}^{n_k} b_n 10^{-n} \right) + 1.
\]

**CASE III**

Neither CASE I or CASE II hold. In this case \( e_{10^{n_k}}(x), e_{b_1 \ldots b_{n_k}}(x) \not\in [0, \beta) \).
This case will definitely occur when

\[ x - ((x) \text{(mod } 10^{-n_k})) \in [6 \cdot 10^{-n_k}, 8 \cdot 10^{-n_k-1}), \]

so that \( \beta < 0.b_1b_2...b_{n_k}6 < e_{b_1...b_{n_k}}(x) < 0.99...98 + 0.000...02 = 1. \) In this case

\[ O_{P\alpha}(x) \cap [0, \beta) = \{e_0(x), ..., e_{b_1...b_{n_k}-1}(x)\} \]

so that

\[ u_{10^{n_k}}(x) = 10^{n_k} \left( 1 + \sum_{n=1}^{n_k} b_n 10^{-n} \right). \]

Therefore, in all cases

\[ u_{10^{n_k}}(x) \in \{N_k, N_k + 1\} \]

where

\[ N_k = 10^{n_k} \left( 1 + \sum_{j=1}^{n_k} b_j 10^{-j} \right). \]

Also we note that

\[ \mu(\{x : u_{10^{n_k}}(x) = N_k\}) \geq \mu(\{x : x - ((x) \text{(mod } 10^{-n_k})) \in [0.6, 0.8]\}) = 0.2, \]

and

\[ \mu(\{x : u_{10^{n_k}}(x) = N_k + 1\}) \geq \mu(\{x : x - ((x) \text{(mod } 10^{-n_k})) \in [0, 0.3]\}) = 0.3. \]

By Lemma 5.3 and (15) we now have

\[ 0 = \lim_{k \to \infty} \int_X |f \circ P_{\alpha}^{10^{n_k}}(x) - f(x)|^2 d\mu(x) \]

\[ = \lim_{k \to \infty} \int_X |e^{2\pi i \lambda u_{10^{n_k}}}(x) - 1|^2 d\mu(x) \]

\[ = \lim_{k \to \infty} \int_X |f(x)|^2 |e^{2\pi i \lambda u_{10^{n_k}}}(x) - 1|^2 d\mu(x) \]

147
\[ \begin{align*}
&= \lim_{k \to \infty} \int_X |e^{2\pi i \lambda u_{10^k}(x)} - 1|^2 d\mu(x) \\
&= \lim_{k \to \infty} \left( |e^{2\pi i \lambda N_k^k} - 1|^2 \mu(\{x : u_{10^k}(x) = N_k\}) \\
&\quad + |e^{2\pi i \lambda (N_k^k + 1)} - 1|^2 \mu(\{x : u_{10^k}(x) = N_k + 1\}) \right),
\end{align*} \]
which implies that
\[ \lim_{k \to \infty} N_k \lambda \equiv 0 \pmod{1} \]
and that
\[ \lim_{k \to \infty} (N_k + 1) \lambda \equiv 0 \pmod{1}. \]
In turn, we have the implication that \( \lambda \equiv 0 \pmod{1} \). This contradicts the assumption that \( \lambda \neq 1 \) hence the only eigenvalue for \( \tilde{P}_\alpha \) is 1 and Lemma 5.3 ensures that the only eigenfunctions are the constants, therefore \( \tilde{P}_\alpha \) has continuous spectrum and thus is weak mixing.

To show that \( \tilde{P}_\alpha \) is not strong mixing we first recall that \( \tilde{P}_\alpha^{u_n(x)} = P_n^\alpha(x) \) for all \( x \in X, n \in \mathbb{N} \). We let \( n = 10^k \) and note that by the arguments used in the proof that \( \tilde{P}_\alpha \) was weak mixing we have that
\[ P^{10^k}_\alpha(x) = \tilde{P}_\alpha^{u_{10^k}}(x) \in \{ \tilde{P}_\alpha^{N_k}(x), \tilde{P}_\alpha^{N_k+1}(x) \} \]
so that
\[ \tilde{P}_\alpha^{N_k}(x) \in \{ P^{10^k}_\alpha(x), \tilde{P}_\alpha^{-1}(P^{10^k}_\alpha(x)) \} \]
for each \( x \in X \). Thus
\[ \tilde{P}_\alpha^{N_k}(B) \subset P^{10^k}_\alpha(B) \cup \tilde{P}_\alpha^{-1}(P^{10^k}_\alpha(B)) \]
for any \( B \in \mathcal{B} \) and \( k \in \mathbb{N} \).

Next note that if \( B \) is a subinterval of \( X \) then using the fact that
\[ \lim_{k \to \infty} n_k - 2n_{k-1} = \infty \]

148
and thus
\[ \lim_{k \to \infty} 10^{\alpha_k} \alpha \equiv 0 \pmod{1} = 0 \]
we know that
\[ \lim_{k \to \infty} \mu(P_{\alpha}^{10^{\alpha_k}}(B) \triangle B) \leq \lim_{k \to \infty} 2(10^{\alpha_k} \alpha \equiv 0 \pmod{1}) = 0. \]
Hence
\[ \lim_{k \to \infty} \mu(P_{\alpha}^{10^{\alpha_k}}(B) \cap B) = \mu(B) \]
and
\[ \lim_{k \to \infty} \mu(P_{\alpha}^{10^{\alpha_k}}(B) \cap B^c) = 0. \]
Similarly
\[ \lim_{k \to \infty} \mu(\tilde{P}_{\alpha}^{-1}(P_{\alpha}^{10^{\alpha_k}}(B)) \triangle \tilde{P}_{\alpha}^{-1}(B)) \leq \lim_{k \to \infty} \mu(\tilde{P}_{\alpha}^{10^{\alpha_k}}(B) \triangle B)) = 0. \]
Hence
\[ \lim_{k \to \infty} \mu(\tilde{P}_{\alpha}^{-1}(P_{\alpha}^{10^{\alpha_k}}(B)) \cap \tilde{P}_{\alpha}^{-1}(B)) = \mu(B) \]
and
\[ \lim_{k \to \infty} \mu(\tilde{P}_{\alpha}^{-1}(P_{\alpha}^{10^{\alpha_k}}(B)) \cap \tilde{P}_{\alpha}^{-1}(B)^c) = 0. \]
Choose \( B \) to be an interval such that \( 0 < \mu(B) < \bar{\mu}(\tilde{X})/2 \). Then clearly
\[ \lim_{k \to \infty} \mu(\tilde{P}_{\alpha}^{N_k}(B) \cap (\tilde{X} - (B \cup \tilde{P}_{\alpha}^{-1}(B)))) = 0 \]
and
\[ \mu(B)\mu(\tilde{X} - (B \cup \tilde{P}_{\alpha}^{-1}(B))) > 0, \]
Hence
\[ \lim_{k \to \infty} \mu(\tilde{P}_{\alpha}^{N_k}(B) \cap (\tilde{X} - (B \cup \tilde{P}_{\alpha}^{-1}(B)))) \neq \mu(B)\mu(\tilde{X} - (B \cup \tilde{P}_{\alpha}^{-1}(B))) \]
and thus $\tilde{P}_c$ cannot be strong mixing.

The next specific example we consider is a generalisation of Kakutani’s first example examined in Chapter 4. We wish in the generalisation to highlight that the important feature in Kakutani’s first example in terms of generating Carathéodory’s definition is the orbits of intervals of density in the primitive space $(X)$. Further reduction to using transformations defined only in terms of orbits is possible, though it requires more preliminary work.

The definition of the type of transformation is based almost entirely on orbits of intervals. For now we continue to consider the usual Lebesgue measure space, $(X, B, \mu)$ on the half open unit interval $X = [0,1)$. We then allow in the generalisation any transformation, $T$, that satisfies the following requirements

1. $T$ is piecewise linear with
   $$\frac{dT}{dx} = 1$$
on each piece.

2. There is a natural number $b \geq 2$ such that for each $n \in \mathbb{N}$
   $$O_T((0, b^{-n})) \equiv \{(kb^{-n}, (k + 1)b^{-n}) : k \in \{0, 1, 2, \ldots, b - 1\}\}.$$

In this thesis we will understand a transformation, $T$, on $[0,1)$ being piecewise linear to mean that there is a countably infinite sequence of intervals $\{J_i\}_{i=1}^\infty$ such that $T$ is linear on each interval and further that for each $x \in [0,1)$ there is a half open interval $[a(x), b(x))$ containing $x$ such that $T$ is linear on $[a(x), b(x))$.

From this definition we can see that any such $T$ must be measure preserving.
and almost invertible in the sense that there must be a set of $X_T$ of measure zero such that $T$ is invertible on $X - X_T$. Kakutani's example is clearly an example of such a transformation. Any such transformation will be known as a Kakutani I type transformation. To show that Kakutani's example is not the only Kakutani I type transformation we provide a general form that we will show gives Kakutani I type transformations.

Let $b \in \mathbb{N}$ and define

$$I_k = \left[1 - \frac{1}{b^k}, 1 - \frac{1}{b^{k+1}}\right].$$

Then for each $j \in \{0, 1, \ldots, b - 2\}$ define

$$I_{kj} = \left[1 - \frac{1}{b^k} + \frac{j}{b^{k+1}}, 1 - \frac{1}{b^k} + \frac{j+1}{b^{k+1}}\right].$$

Note that

$$\bigcup_{j=0}^{b-2} I_{kj} = I_k$$

and

$$\bigcup_{k=0}^{\infty} I_k = [0, 1)$$

so that

$$\bigcup_{k=0}^{\infty} \bigcup_{j=0}^{b-2} I_{kj} = [0, 1).$$

We then define a general form of Kakutani I type transformations by

$$P(x) = x - \frac{b^{k+1} - b + j}{b^{k+1}} + \sum_{i=1}^{k} \frac{p - 1}{b^i} + \frac{(p + j)(\text{mod } b)}{b^{k+1}}.$$  

for all $x \in I_{kj}$ whenever $k \in \mathbb{N}$, $j \in \{0, 1, \ldots, b - 2\}$, $p, b$ and gcd$(p, b) = 1$.

The motivation for the general coming from noting that by setting $b = 2$ and $p = 1$ we have kakutani’s original transformation.
Clearly, for each $I_{kj}$, $P$ is of the form

$$P(x) = x - c(kj) \quad x \in I_{kj},$$

where $c(kj)$ is a $b$-adic rational of order $k + 1$ dependent only on $kj$. Note that by the definition of $P$ and the fact that $\gcd(p, b) = 1$ we have that $c(kj)$ is never equal to 0. Importantly, from this form it is clear that $P$ is linear and of slope 1 on each of the countable set of intervals

$$\{I_{kj} : k \in \mathbb{N}, j \in \{0, 1, ..., b - 2\},$$

so that the first requirement for $P$ to be a kakutani I type transformation is satisfied.

We show that the second requirement is also fulfilled in the following Lemma.

**Lemma 5.4**

Let $P$ be a transformation of the form described above. Suppose that $J$ is the $b$-adic interval

$$J = \left( \frac{q}{b^n}, \frac{q + 1}{b^n} \right),$$

where $n \in \mathbb{N}$ and $q \in \{0, 1, 2, ..., b^n - 1\}$. Then the following statements hold.

1. $P(J) = (k/b^n, (k+1)/b^n)$ for some $k \in \{0, 1, 2, ..., b^n - 1\}$.

2. If $q \neq b^n - 1$, then $(P(J))_i = P(J_i), \quad i \in \{0, 1, 2, ..., b - 1\}$; but if $q = b^n - 1$, $P(J_i) = (P(J))_j$ where $j \equiv (i + p)(\text{mod } b)$.

3. $J$ has a cyclic orbit of $b^n$ distinct elements which consists of all the $b$-adic intervals in $X$ of order $n$. 

152
Proof:

For this lemma we will write

$$\rho_k = \sum_{i=1}^{k} \frac{p - 1}{b^i}.$$ 

(i) For some \(n_0 \in \mathbb{N}, j_0 \in \{0, 1, \ldots, b - 2\}\)

$$q/b^n \in \left[1 - \frac{1}{b^{n_0}}, 1 - \frac{1}{b^{n_0+1}} + \frac{j_0 + 1}{b^{n_0+1}}\right].$$

If \(n < n_0\), then \(1/b^n > 1/b^{n_0}\) and so

$$\frac{q + 1}{b^n} \geq 1 - \frac{1}{b^{n_0}} + \frac{1}{b^n} > 1.$$ 

This is impossible and so \(n \not< n_0\).

If \(n = n_0\) then \(q/b^n \in [1 - 1/b^n, 1 - 1/b^{n+1}]\). However, the only \(b\)-adic rational of order \(n\) in this interval is \(1 - 1/b^n\) and so

$$\frac{q}{b^n} = 1 - \frac{1}{b^n}, \quad \frac{(q + 1)}{b^n} = 1, \quad \text{and} \quad J = \left(1 - \frac{1}{b^n}, 1\right).$$

Using the definition of the \(I_{kj}\),

$$J = \bigcup_{k=n}^{\infty} \bigcup_{j=0}^{b-2} I_{kj}$$

and since \(P\) is increasing and linear on each of the \(I_{kj}\) we have, by the definition of \(P\) over the \(I_{kj}\), that

$$P(I_{kj}) = \left[\rho_k + \frac{(p + j)(\text{mod } b)}{b^{k+1}}, \rho_k + \frac{(p + j)(\text{mod } b) + 1}{b^{k+1}}\right]$$

so that for each \(k \in \mathbb{N}\)

$$P(I_k) = \bigcup_{j=0}^{b-2} P(I_{kj})$$

$$= \bigcup_{j=0}^{b-2} \left[\rho_k + \frac{(p + j)(\text{mod } b)}{b^{k+1}}, \rho_k + \frac{(p + j)(\text{mod } b) + 1}{b^{k+1}}\right]$$

$$= \left[\rho_k, \rho_k + \frac{1}{b^k}\right] - \left[\rho_k + \frac{(p + b - 1)(\text{mod } b)}{b^{k+1}}, \rho_k + \frac{(p + b - 1)(\text{mod } b) + 1}{b^{k+1}}\right]$$

$$= \left[\rho_k, \rho_k + \frac{1}{b^k}\right] - \left[\rho_k, \rho_k + \frac{1}{b^{k+1}}\right].$$
Hence

\[ P(J) = \bigcup_{k=n}^{\infty} P(I_k) \]
\[ = \bigcup_{k=n}^{\infty} \left[ \rho_k, \rho_k + \frac{1}{b^k} \right) - \left[ \rho_k, \rho_k + \frac{1}{b^{k+1}} \right) \]
\[ = \left[ \rho_k, \rho_k + \frac{1}{b^n} \right), \]

which is a $b$-adic interval of order $n$ and hence the result is proved for $n = n_0$.

The final case is for $n < n_0$ so that $n = n_0 + a$, for some $a \in \mathbb{N}$. Then

\[ \frac{b^n - b^a}{b^n} \leq \frac{q}{b^n} < \frac{b^n - b^{a-1}}{b^n} \]

and so

\[ \frac{b^n - b^a}{b^n} < \frac{q}{b^n} < \frac{b^n - b^{a-1}}{b^n}. \]

Thus

\[ J = \left[ \frac{q}{b^n}, \frac{(q+1)}{b^n} \right) \subseteq \left( 1 - \frac{1}{b^{n_0}}, 1 - \frac{1}{b^{n_0+1}} \right) = I_{n_0}. \]

That is, $J$ is a $b$-adic interval of order $n$ ($n > n_0$) that is a subset of $I_{n_0}$. It follows that $J$ is a subset of $I_{n_0j}$ for some $j \in \{0, 1, ..., b-2\}$.

As $P$ is linear on $I_{n_0j}$ we need only calculate the values of $J$'s endpoints. For this we use the form of the definition of $P$

\[ P(x) = x + c(n_0j) \quad x \in I_{n_0j} \]

where $c(n_0j)$ is a non-zero $b$-adic rational of order $n_0 + 1 \leq n$. Then

\[ P \left( \frac{q}{b^n} \right) = \frac{q}{b^n} + c(n_0j) \]

and

\[ P \left( \frac{(q+1)}{b^n} \right) = \frac{(q+1)}{b^n} + c(n_0j). \]
Hence it is clear that $P(J) = (q/b^n + c(n_0j), (q + 1)/b^n + c(n_0j))$ which is a $b$-adic rational of order $n$. As $c(n_0j) \neq 0$ we have the required result. (ii) If $q = b^n - 1$, then $J = [1 - 1/b^n, 1)$ and by previous calculations we know that

$$P(J) = \left[ \rho_k, \rho_k + \frac{1}{b^n} \right].$$

Now, noting that for each $i \in \{0, 1, \ldots, b - 2\}, J_i = I_{ni}$ we have

$$P(J_i) = P(I_{ni}) = \left[ \rho_n + \frac{(p + j)(\text{mod } b)}{b^{n+1}}, \rho_n + \frac{(p + j)(\text{mod } b + 1)}{b^{n+1}} \right] = (P(J))_{(p+j)(\text{mod } b)}.$$

Then, as

$$J_{b-1} = \left[ 1 - \frac{1}{b^{n+1}} \right]$$

we know from previous calculations that

$$P(J_{b-1}) = \left[ \rho_{n+1}, \rho_{n+1} + \frac{1}{b^{n+1}} \right] = \left[ \rho_n + \frac{p - 1}{b^{n+1}}, \rho_n + \frac{p}{b^{n+1}} \right] = (P(J))_{(p-1)(\text{mod } b)} = (P(J))_{(p+b-1)(\text{mod } b)},$$

which completes the proof of (ii)

(iii) Mathematical induction on $n$ is used.

For $n = 1$, let $J = (0, 1/b)$ then for $n = 1, 2, \ldots, b$

$$P^n(J) = \left( \frac{(n + p)(\text{mod } b)}{b^n}, \frac{(n + p)(\text{mod } b + 1)}{b^n} \right),$$

and as $\gcd(p, b) = 1$ this means that

$$\{P^n(J) : n = 1, 2, \ldots, b\} = \left\{ \left( \frac{k}{b}, \frac{(k+1)}{b} \right) : k = 0, 1, 2, \ldots, b - 1 \right\}$$
meaning that the orbit of each $b$-adic interval of order 1 is precisely the set of the $b$ $b$-adic elements of order 1, proving the claim for $n = 1$.

Now assume that the claim is true for $n = k$ so that for each $b$-adic interval, $J$, of order $k$ $O_P(J)$ consists essentially of the set the $b^k$ $b$-adic intervals of order $k$. Now choose $J$ so that $O_P(J)_b^k \equiv (1 - 1/b^k, 1)$.

Define $K$ to be $J_0$ which is a $b$-adic interval of order $k + 1$. From (ii) we know that

$$O_P(K)_i \equiv (O_P(J)_i)_0 \quad i = 1, 2, ..., b^k$$

and that using the hypothesis of the cyclic orbit

$$O_P(K)_{b^k+1} \equiv (O_P(J)_1)_{(p)(\text{mod } b)}.$$ 

Again using (ii) we continue through the orbit of $K$ and find that

$$O_P(K)_{b^k+i} \equiv (O_P(J)_i)_{(p)(\text{mod } b)} \quad i = 1, 2, ..., b^k,$$

which are $b$-adic intervals of order $k + 1$ distinct from any previously found to be in the orbit of $K$. Repeat this process $b$ times to find that there are (using the fact that $p$ is relatively prime to $b$) $b^{k+1}$ elements in $O_P(K)$ concluding thus far with the interval

$$O_P(K)_{b^k+1} \equiv ((O_P(J)_{b^k})_{((b-1)p)(\text{mod } b)}$$

so that

$$O_P(K)_{b^k+1+1} \equiv (O_P(J)_1)_{(bp)(\text{mod } b)} \equiv (O_P(J)_1)_0 = J_0 = K.$$ 

Thus the orbit of $K$ consists precisely of the $b^{k+1}$ $b$-adic orbits of order $k + 1$. Thus the claim being true for $n = k$ implies that it is true for $n = k + 1$ and therefore the result follows by induction. \diamond
Still restricting ourselves to two level towers we will now show that a Kakutani I type transformation with a usual two level tower extension will generate the Carathéodory definition of measurable sets. The only additional result we require to prove this fact is the existence of $b$-adic intervals of density of any sufficiently large order for any set of positive outer measure. The corresponding dyadic result for Kakutani's original transformation is Lemma 4.7 and the proof for the current variation of Lemms 4.7 is identical to that for Lemma 4.7 only changing some 2's to $b$'s. For this reason we will formally state the required result but omit the proof.

**Lemma 5.5**

Let $B \subseteq \tilde{X}$ be $\tilde{P}$-invariant for a usual two level tower extension of some Kakutani I type transformation $P$ on the space $(X, \mathcal{B}, \mu)$ and let $\varepsilon > 0$. Then there is an $n_0 \in \mathbb{N}$ which has the following property: if $n > n_0$ there is a $b$-adic interval, $J$, of order $n$ in $X$ such that

$$\tilde{\mu}_*(B \cap J) > (1 - \varepsilon)\tilde{\mu}(J).$$

The following theorem now gives us that if $\tilde{P}$ is a usual two level tower extension of a Kakutani I type transformation then $\tilde{P}$ generates the Carathéodory definition of measurable sets.

**Theorem 5.3**

Let $P$ be a Kakutani I type transformation on $(X, \mathcal{B}, \mu)$ and let $\tilde{P}$ be any usual two level tower extension on the extended space $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$. Let $B$ be a $\tilde{P}$-invariant subset of $\tilde{X}$. Suppose that there is a $\theta < 2$ such that for all subintervals $V$ of $\tilde{X}$

$$\tilde{\mu}_*(B \cap V) + \tilde{\mu}_*(B^c \cap V) \leq \theta \tilde{\mu}(V).$$
Then, either $B$ or $B^c$ is a set of measure zero.

Proof:
Suppose that $\bar{\mu}_*(B) > 0$ and $\bar{\mu}_*(B^c) > 0$. Then by Lemmas 5.1 and 5.2 we know that $\mu_*(B \cap X) > 0$, $\mu_*(B^c \cap X) > 0$ and that both $B \cap X$ and $B^c \cap X$ are $P$-invariant. Since $P$ is 1 to 1 and onto, this also means that $B \cap X$ and $B^c \cap X$ are $P^{-1}$-invariant. Set $\epsilon$ as $1 - \theta/2 > 0$. Then we know, by Lemma 5.5 that we can find two $b$-adic intervals $Z_1$ and $Z_2$ of equal length such that

$$\mu_*(B \cap X \cap Z_1) = \mu_*(B \cap Z_1) > (1 - \epsilon)\mu(Z_1) = \theta\mu(Z_1)/2,$$

and

$$\mu_*(B^c \cap X \cap Z_2) = \mu_*(B^c \cap Z_2) > (1 - \epsilon)\mu(Z_2) = \theta\mu(Z_2)/2.$$

By the specification of $P$ we have that $Z_1 \in o(Z_2)$ so that for some $p \in \mathbb{N}$

$Z_1 = P^p(Z_2)$. Then,

$$\mu_*(B^c \cap Z_1) = \mu_*(B^c \cap X \cap Z_1)$$
$$= \mu_*(B^c \cap X \cap P^p(Z_2))$$
$$= \mu_*(P^p(B^c \cap X) \cap P^p(Z_2)) \quad \text{as } B^c \cap X \text{ is } P^{-1} \text{-invariant}$$
$$= \mu_*(P^p(B^c \cap X \cap Z_2))$$
$$= \mu_*(B^c \cap X \cap Z_2)$$
$$> \frac{\theta\mu(Z_2)}{2}$$
$$= \frac{\theta\mu(Z_1)}{2}.$$

Thus

$$\bar{\mu}_*(B \cap Z_1) + \bar{\mu}_*(B^c \cap Z_1) > \frac{\theta\mu(Z_1)}{2} + \frac{\theta\mu(Z_1)}{2} = \theta\mu(Z_1).$$

This contradiction means that either $\bar{\mu}_*(B) = 0$ or $\bar{\mu}_*(B^c) = 0$. \hfill \Diamond
Note that we have not discussed the mixing properties of $P$. This is because it depends greatly on the set $A$ used in the tower extension. While this means that there could be transformations that have stronger mixing properties than weak mixing, we don’t actually have an example. However, after considering towers in a more general sense we do show that for any Kakutani I type transformation, $P$, and any $n \in \mathbb{N}$ there is a $n$ level tower extension of $P$ that is weak mixing.

5.2 Towers

It now appears that the only essential features of transformations that generate Carathéodory's definition are the existence of intervals of density and an appropriate form of orbits of these intervals. Since we would like to see where these transformations fit in in the sense of mixing, and since the most important feature in affecting mixing levels was tower extensions, we would like to see what effect tower extensions have on the generation of Carathéodory's definition. The intention being to then try and control mixing levels by choosing tower extensions that will give the desired mixing levels. We find that if any tower transformation (including the primitive transformation) generates Carathéodory's definition then any other tower transformation on the same primitive transformation (including the primitive transformation) also generates Carathéodory's definition. This means we choose any tower extension we like in terms of finding mixing properties. However, as we have mentioned, the most important development in investigating towers is the further investigation of splinters.

We first define what we will mean by tower extensions, both of finite and infinite level. Throughout this chapter we will work with a generic measure space which we will denote by $(X_0, B_0, \mu_0)$ from which we will derive tower
extensions. The space \((X_0, \mathcal{B}_0, \mu_0)\) that is the base of our tower extensions, as we have mentioned, is also known as the \textit{primitive space}. Together with the primitive space we will consider an arbitrary invertible transformation \(T_0\) defined on \(X_0\), which will be known as the \textit{primitive transformation}. We can now define what we mean by tower extensions in the general sense. It is the obvious extension of the usual two level tower extension that we have already seen. Since we will have difficulty expressing \(n\) overlines in an appropriate form, we will need to alter our notation.

\textbf{Definitions 5.2}

To extend \((X_0, \mathcal{B}_0, \mu_0, T_0)\) to the usual two level tower extension we take some \(A_0 \in \mathcal{B}_0\) and define a bijection \(\tau_0 : A_0 \to B_0\) where \(B_0\) is a set chosen such that \(X_0 \cap B_0 = \emptyset\). Then

\[ X_1 = X_0 \cup B_0 \]
\[ \mathcal{B}_1 = \{ D \subset X_1 : D \cap X_0 \in \mathcal{B}_0, \tau_0^{-1}(D \cap B_0) \in \mathcal{B}_0 \} \]
\[ \mu_1(B) = \mu_0(B \cap X_0) + \mu_0(\tau_0^{-1}(B \cap B_0)) \quad \text{for all } B \in \mathcal{B}_1 \]

and

\[ T_1(x) = \begin{cases} T_0(x), & x \in X_0 - A_0, \\ \tau_0(x), & x \in A_0, \\ T_0(\tau_0^{-1}(x)), & x \in B_0. \end{cases} \]

We then define the \textbf{usual \(n\)-level tower construction} \((X_n, \mathcal{B}_n, \mu_n)\) with \(T_n\) inductively as follows. Let \(A_{n-1}\) be a measurable subset of \(B_{n-2}\) and \(B_{n-1}\) be a set such that \(X_{n-1} \cap B_{n-1} = \emptyset\) and \(\tau_{n-1} : A_{n-1} \to B_{n-1}\) be the bijection between \(A_{n-1}\) and \(B_{n-1}\) in the previous inductive step. Then

\[ X_n = X_{n-1} \cup B_{n-1}, \]
\[ \mathcal{B}_n = \{ D \subset X_n : D \cap X_{n-1} \in \mathcal{B}_{n-1}, \tau_{n-1}^{-1}(D \cap B_{n-1}) \in \mathcal{B}_{n-1} \}, \]
\[ \mu_n(B) = \mu_{n-1}(B \cap X_{n-1}) + \mu_{n-1}(\tau_{n-1}^{-1}(B \cap B_{n-1})), \quad \text{for all } B \in \mathcal{B}_n \]
and

\[ T_n(x) = \begin{cases} \tau_0(x), & x \in X \cap A_0, \\ \tau_k(x), & x \in A_k, \quad 0 \leq k \leq n-1, \\ T_0(\tau_0^{-1} \circ \tau_1^{-1} \circ \ldots \circ \tau_k^{-1}(x)), & x \in B_k - A_{k+1} \cup A_n, \quad 0 \leq k \leq n-1. \end{cases} \]

We extend this process to define the usual \( \infty \)-level tower extension, \((X^*, B^*, \mu^*)\) as follows.

\[ X^* = X_0 \cup \bigcup_{n=0}^{\infty} B_n, \]
\[ B^* = \bigcup_{n=0}^{\infty} B_n, \]
\[ T^*(x) = T_{n+1}(x) \quad x \in X_n \]

and

\[ \mu^*(B) = \mu_0(B \cap X_0) + \sum_{n=1}^{\infty} \mu_n(B \cap B_{n-1}). \]

As in this chapter we will be frequently using compositions of sequences of transformations we define the following notation.

**Definition 5.3**

Let \( \{f_n\}_{n=1}^k \) be a sequence of functions such that \( f_n : X_n \to X_{n+1} \) for each \( 1 \leq n \leq k \), then we define

\[ \circ_{n=0}^{k-1} f_{k-n} : X_1 \to X_{k+1} \]

by

\[ \circ_{n=0}^{k-1} f_{k-n}(x) = f_k \circ f_{k-1} \circ \ldots \circ f_1(x) \]

for each \( x \in X_1 \).

In this construction we require that \( \mu^*(X^*) < \infty \). Also let \( \mu_{n*} \) and \( \mu^* \) be the associated outer measures for \( \mu_n \) and \( \mu^* \) respectively.
Note especially that it can be shown that $\mu_n$ and $\mu^*$ are measures.

We also need to show that $T_n$ and $T^*$ are measure preserving if and only if $T$ is. To do this we need to introduce further new notation. We need to look at an alternative method of defining the tower transformations. That is, instead of seeing the tower as a sequence of sets of decreasing size, one resting on top of the previous one, we can picture the tower as a sequence of disjoint stacks of sets of the same size. Each stack is a finite collection of ordered sets, each placed on top of the one before it. The set at the bottom of each stack is called the base of that stack, denoted $S_{i0}$, that is the base or ground (0-th) floor of the $i$-th stack. The set at the top of each stack is called the ceiling and is denoted by $S_{i|n|^{-1}}$, the ceiling or top ($(n_i - 1)$-th) floor of the $i$-th stack where $n_i$ is the number of piled sets in the $i$th stack. We also denote by $S_{ik}$ the $k$-th level of the stack (where we are counting up from the bottom with the base being counted as 0). To join the two ideas, for a usual tower extension we consider stack $i$ as the unique stack that has a base

$$S_{i0} = \{x \in X : \bar{T}_0^n(x) \in X, \bar{T}_m(x) \not\in X, \ 1 \leq m < n_i\}.$$ 

We can then order the sets so that the $i-$th stack has $i + 1$ sets piled in the stack, that is $n_i = i + 1$. In this case clearly $S_{ii} = \bigcup_{j=0}^{i} \bar{T}_{i-j}(S_{i0})$ and

$$B_i = \bigcup_{n=0}^{\alpha} \bigcup_{j=0}^{i} \bar{T}_{i-j}(S_{n0})$$

for each level $B_i$ in our original method of constructing towers, where $\alpha$ is the number of levels in the tower, which may be finite or countably infinite.

**Theorem 5.4**

Let $(X, B, \mu)$ be a measure space and $(\bar{X}, \bar{B}, \bar{\mu})$ be a tower extension on $(X, B, \mu)$ (we do not distinguish whether it is finite or infinite). We let $T$
be a transformation on $X$ and $\breve{T}$ be the tower transformation on $\breve{X}$. In this case $T$ is $\mu$-measure preserving if and only if $\breve{T}$ is $\breve{\mu}$-measure preserving.

Proof:
Suppose that $T$ is $\mu$-measure preserving. Note also that the definition of $\breve{\mu}$ ensures that $\tau_n$ is measure preserving for each $n \in \mathbb{N}$. Now let $D \subset \breve{X}$ so that

$$D = \bigcup_{n=1}^{\alpha} \bigcup_{i=0}^{n-1} (D \cap S_n)$$

and

$$\breve{T}^{-1}(D) = T^{-1}D \cap S_{10} \cup \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} \tau_i(T^{-1}D \cap S_{n_0}) \right) \cup \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} \tau_i^{-1}(D \cap S_{n+1}) \right)$$

Hence, since $S_{nm} \cap S_{ij} = \emptyset$ whenever $n \neq i$ or $m \neq j$, we have

$$\breve{\mu}(\breve{T}^{-1}D) = \breve{\mu}\left(T^{-1}(D \cap S_{10}) \cup \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} \tau_i(T^{-1}(D \cap S_{n_0})) \right) \cup \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} \tau_i^{-1}(D \cap S_{n+1}) \right) \right)$$

$$= \breve{\mu}(T^{-1}(D \cap S_{10})) + \sum_{n=2}^{\alpha} \breve{\mu}(\bigcup_{i=0}^{n-2} \tau_i(T^{-1}(D \cap S_{n_0})))$$

$$+ \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \breve{\mu}(\tau_i^{-1}(D \cap S_{n+1}))$$

$$= \breve{\mu}(T^{-1}(D \cap S_{10})) + \sum_{n=2}^{\alpha} \breve{\mu}(T^{-1}(D \cap S_{n_0}))$$

$$+ \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \breve{\mu}((D \cap S_{n+1}))$$

$$= \breve{\mu}(D \cap S_{10}) + \sum_{n=2}^{\alpha} \breve{\mu}(D \cap S_{n_0}) + \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \breve{\mu}((D \cap S_{n+1}))$$

$$= \breve{\mu}(D \cap X) + \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \breve{\mu}((D \cap S_{n+1}))$$

$$= \breve{\mu}\left(\bigcup_{n=1}^{\alpha} (D \cap S_{n_0})\right) + \breve{\mu}\left(\bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} (D \cap S_{n+1})\right)$$

163
Thus $\tilde{T}$ is $\tilde{\mu}$-measure preserving.

Conversely, suppose that $\tilde{T}$ is $\tilde{\mu}$-measure preserving. Then we need to show that $T$ is $\mu$-measure preserving. Let $D \subset X$. We again note that $\tilde{T}^{-1}D$ is a subset of the ceiling of the tower. That is

$$\tilde{T}^{-1}D \subset \bigcup_{n=0}^{\alpha-1} S_{n_n},$$

so that

$$\tilde{T}^{-1}D = \bigcup_{n=1}^{\alpha} (D \cap S_{n_n}).$$

Note that we could also have used $S_{n_{n_n-1}}$. However, by our earlier construction of the $S_n$ we know that $i_i = i + 1$ for each $i = 0, 1, ..., \alpha - 1$. We now have that

$$\mu(T^{-1}(D)) = \tilde{\mu}(T^{-1}(D))$$

$$= \tilde{\mu}\left(\bigcup_{n=1}^{\alpha} (T^{-1}D \cap S_{n_0})\right)$$

$$= \sum_{n=1}^{\alpha} \tilde{\mu}(D \cap S_{n_0})$$

$$= \sum_{n=1}^{\alpha} \tilde{\mu}(\alpha_{i=0}^{n-1} \tau_i(D \cap S_{n_0}))$$

$$= \sum_{n=1}^{\alpha} \tilde{\mu}(D \cap S_{n_n})$$

$$= \tilde{\mu}\left(\bigcup_{n=1}^{\alpha} (D \cap S_{n_n})\right)$$
\[
\begin{align*}
&= \mu(T^{-1}D) \\
&= \mu(D) \\
&= \mu(D).
\end{align*}
\]

Which, by definition gives us that \( T \) is \( \mu \)-measure preserving which completes the proof. 

We have previously mentioned induced and skyscraper transformations. When we consider the \( B_n \)'s used in the above definitions as levels or floors in a tower, one placed above the previous, we can describe the difference between a derived tower transformation and an induced skyscraper transformation. The derived tower transformation creates new external sets to the original set and correspondingly extends the transformation that was defined on the original space. The induced skyscraper construction requires that the transformation on the original space be of a special type. The process of constructing the induced transformation involves breaking the original set up into some sequence of measurable parts and then placing these parts one atop the other. The resulting skyscraper construction is then expected to behave in the same way as a tower transformation. Thus the transformation must already be in a form capable of supporting this concept.

One example of an induced skyscraper transformation is the first of Kaku-tani's transformations that we considered. However, we postpone constructing an example until later in the chapter where we give an example of a weakly mixing induced skyscraper transformation on \([0,1)\).

The aim is now to consider tower transformations. To do so, however, we need to use the concept of 'splinters' and so we will now give a more thorough explanation, and a formal definition, of what they are.
Under iterations (or inverse iterations) of a transformation the image of a set will usually be broken up in some sense, this may entail being broken up from being topologically connected to not being topologically connected or being broken up as a subset of one level of a tower transformation to having non empty intersection with multiple levels of the tower transformation. We have so far been concerned with the connected sets, intervals. Further when investigating their orbits we have usually only been interested in the elements of the orbits of these intervals that are also (essentially) intervals, that is, are also connected or in the same form as the original set. We would now like to be able to allow the sets we consider to break up, in some understandable way into parts that will be called splinters, and to able to be able to sufficiently understand the behaviour of the orbit of the original set in terms of the orbits of the splinters. For this chapter, we will not be concerned with topological connectedness, we will be concerned only with the breaking up of a set from one into multiple levels of a tower transformation.

**Definition 5.4**

Let $(X, \mathcal{B}, \mu)$ be a primitive space with primitive transformation $T$. Let $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu}, \tilde{T})$ be a usual tower extension of either finite or infinite levels. Let $n \in \mathbb{N}$ and $B \subseteq \tilde{X}$. Then a subset $V$ of $B$ is an $n$-splinter of $B$ if for each $0 \leq p \leq n$, either $\tilde{T}^{-p}(V) \subseteq X_0$ or $\tilde{T}^{-p}(V) \subseteq B_k$ for some tower level $B_k$. That is, $V$ is an $n$-splinter of $B$ if for each $0 \leq p \leq n$, $\tilde{T}^{-p}(V)$ is a subset of some level of the tower construction.

We now are in a position to investigate some of the properties of the usual $n$ level tower constuctions for finite $n$ with respect to generating the Carathéodory definition of measurable sets. We present the properties in essentially the
same structure as we did for a usual two level tower extension for irrational rotations. That is we show that a $T_n$-invariant set of positive measure must have an intersection with $X_0$ of positive measure. We then show that $T_n$ invariance of a set $B$ implies $T_0$-invariance of $B \cap X_0$. These two results are presented as Lemmas 5.6 and 5.7. We then present the first important result for finite level tower extensions in Theorem 5.4.

**Lemma 5.6**

Let $(X_0, \mathcal{B}_0, \mu_0)$ be a measure space and $T_0$ be a measure preserving bijection. Let $(X_n, \mathcal{B}_n, \mu_n, T_n)$ be the usual $n$-level tower extension of $(X_0, \mathcal{B}_0, \mu_0, T_0)$. If $V \subset X_n$ is $T_n$-invariant and $\mu_n(V) > 0$, then $\mu_0(V \cap X_0) > 0$.

**Proof:**

After an iteration of $T_n$, $V$ may be broken up into as many as $n$-disjoint $n$-splinters, one per level. After a second iteration, the same may happen to each of these $n$-splinters so that there may be as many as $n \times n = n^2$ disjoint $n$-splinters of $V$. Continuing in this manner, we see that after $n$ iterations of $T_n$ there will be some finite number $k \leq n^n$ of disjoint $n$-splinters of $V$. Let these $n$-splinters be listed as $V_1, V_2, ..., V_k$. Then

$$0 < \mu_n(V) = \mu_n \left( \bigcup_{n=1}^{k} V_n \right) \leq \sum_{n=1}^{k} \mu_n(V_n),$$

so that there must be some $j \in \{1, 2, ..., k\}$ such that $\mu_n(V_j) > 0$.

Also, by the construction of $T_n$ there must be an $m \in \{0, 1, 2, 3, ..., n-1\}$ such that $T_n^m(V_j) \subset X_0$ and as $V$ is $T_n$-invariant we have that $T_n^m(V_j) \subset V \cap X_0$. Also as $T_n$ is a measure preserving bijection $\mu_n(T_n^m(V_j)) = \mu_n(V_j) > 0$. Thus

$$\mu_0(V \cap X_0) = \mu_n(V \cap X_0) \geq \mu_n(T_n^m(V_j))$$

167
Lemma 5.7

Let \((X_0, B_0, \mu_0)\) be a finite measure space and \(T_0\) be a measure preserving bijective transformation on \(X\). Let \((X_n, B_n, \mu_n)\) and \(T_n\) be the usual \(n\)-level tower extension. If \(V \subset X_n\) is \(T_n\)-invariant, then \(V \cap X_0\) is \(T_0\)-invariant.

Proof:

Let \(V\) be a \(T_n\)-invariant subset of \(X_n\). Now let \(x \in T_0^{-1}(V \cap X_0)\) so that \(T_0(x) \in V \cap X_0\). We therefore have that \(x \in X_0\).

If \(x \in V \cap X_0\), then

\[
T_n^{-1}(x) = \circ_{i=0}^m \tau_{m-i}(T_0^{-1}(x))
\]

for some \(0 \leq m \leq n - 1\) and hence

\[
T_n^{-m-1}(x) = \circ_{i=0}^m \tau_i^{-1}(\circ_{i=0}^m \tau_{m-i}(T_0^{-1}(x))) = T_0^{-1}(x).
\]

as \(V\) is \(T_n\)-invariant we have \(T_0^{-1}(x) \subset V\). Also, clearly \(T_0^{-1}(x) \subset X_0\) and hence \(T_0^{-1}(x) \subset V \cap X_0\). As this is true for all \(x \in V \cap X_0\) it follows that \(T_0^{-1}(V \cap X_0) \subset V \cap X_0\).

Conversely, suppose \(x \in V \cap X_0\), then \(x \in X_0\) and thus \(T_0(x) \in X_0\).

Then, if \(x \in X_0 \cap V \cap A_0^n\), then \(T_0(x) = T_n(x) \in V\). Also, if \(x \in A_0\) then there must exist some \(m \in \{0, 1, 2, ..., n - 1\}\) such that \(T_n^m(x) = T_0(x)\), that is \(T_0(x) = T_n^m(x) \in V\). Therefore

\[
x \in V \cap X_0 \quad \Rightarrow \quad T_0(x) \in V \cap X_0
\]

\[
\Rightarrow \quad x \in T_0^{-1}(V \cap X_0).
\]
Thus $V \cap X_0 \subset T_0^{-1}(V \cap X_0)$ and therefore we now have that

$$V \cap X_0 = T_0^{-1}(V \cap X_0)$$

and hence the result.

Our main theorem concerning finite level tower extensions of transformations clearly concerns generating Carathéodory's definition of measurable sets. With our current definition of a transformations ability to generate the Carathéodory definition of measurable sets we need to have a concept of what an interval of density is. As we are assuming in Theorem 5.4 that the primitive space generates Carathéodory's definition we are, at this stage, implying that the primitive space must have a concept of intervals and that the process of finding intervals of density must work. That is, we must be able to find intervals of density of any sufficiently small measure for any set of positive outer measure in the primitive space. The fact that the concept of intervals of density must make sense in the primitive space is actually sufficient. However, we will need to give a definition of a subinterval in the more general situation of the tower extensions of a space in which intervals and intervals of density make sense.

While we can do this directly, we will in later chapters be using a more general tool than intervals of density but we would like to continue to use some the results that we prove later in this chapter. For this reason we give the following definitions.
Definition 5.5

Let \((X, B, \mu)\) be a measure space and \(B\) be a subset of \(X\). Then a set \(J \in B\) is called a set of density to within \(\varepsilon\) for \(B\) if

\[
\mu_*(B \cap J) > (1 - \varepsilon)\mu(J).
\]

Definition 5.6

Let \((X, B, \mu)\) be a measure space. A collection \(\mathcal{J} \subset B\) will be called a collection of basic sets if for each pair of subsets \(A\) and \(B\) of \(X\) and each \(\varepsilon > 0\) there exist two sets \(J_1\) and \(J_2\) in \(\mathcal{J}\) of equal measure such that \(J_1\) is a set of density to within \(\varepsilon\) for \(A\) and \(J_2\) is a set of density to within \(\varepsilon\) for \(B\).

These collections of basic sets will eventually be redefined in Chapter 6 as a collection of standard sets. We make the redefinition in Chapter 6 when the reason for the name is given discussion. We make this corresponding definition now to allow us to prove certain results that we would like to have hold for collections of basic (standard) sets as well as for intervals as we have been using in this chapter.

Note in particular that subintervals or \(b\)-adic subintervals (for some \(b \in \mathbb{N}\)) are collections of basic sets in \([0,1)\), hence for the duration of this chapter a collection of basic sets can simply be thought of as the set of subintervals of \([0,1)\). While the same definition of a collection of basic sets will work for the measure space that results from a tower extension of a measure space, we wish to give this case special consideration so that we can also consider the collection of basic sets for the tower extension as being an extension of the collection of basic sets for the primitive space. That is, we need to make the following definition.
Definition 5.7
For each collection of basic sets, \( J \) on a measure space \((X_0, \mathcal{B}_0, \mu_0)\) we define a tower collection of basic sets, \( \tilde{J} \) for a tower extension \((\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})\) corresponding to \( J \) by defining an element of \( \tilde{J} \) to be either an element of \( J \) or a subset, \( V \), of \( \mathcal{B}_k \) for some level of the tower extension, \( \mathcal{B}_k \) such that \( \omega_{n=1}^{n-1}(V) \in J \).

We point out that in particular, we have that a subinterval of a tower extension \((\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})\) of a measure space \((X_0, \mathcal{B}_0, \mu_0)\) for which intervals make sense is either a subinterval of \( X_0 \) or a subset \( V \), of \( \mathcal{B}_k \), for some level of the tower extension \( \mathcal{B}_k \), such that \( \omega_{n=1}^{n-1}(V) \) is a subinterval of \( X_0 \).

With these definitions we will prove results about the generation of Carathéodory's definition, we will prove the generalised case but note as corollaries the particular cases that were developed at this point of the research that are results in terms of the intervals with which we are already familiar.

Theorem 5.5
Let \((X_n, \mathcal{B}_n, \mu_n, T_n)\) be the usual \( n \)-level tower extension of a primitive dynamical system \((X_0, \mathcal{B}_0, \mu_0, T_0)\), let \( J \) be a chosen collection of basic sets for \((X, \mathcal{B}, \mu)\) and let \( J_n \) be the tower collection of basic sets for \( X_n \). Let each \( T \)-invariant subset \( A \) of \( X_0 \) for which there is a \( \theta = \theta(A) < 2 \) such that for all \( J \in J \)

\[
\mu_0(A \cap J) + \mu_0(A^c \cap J) \leq \theta \mu_0(J),
\]

be such that either \( A \) or \( A^c \) is a set of outer measure zero. Now, let \( B \) be a \( T_n \)-invariant subset of \( X_n \). Suppose that there is a \( \theta < 2 \) such that for all
$V \in \mathcal{J}_n$

$$\mu_n(B \cap V) + \mu_n(B^c \cap V) \leq \theta \mu_n(V).$$

Then, either $B$ or $B^c$ is a set of measure zero.

**Proof:**

As $B$ is $T_n$-invariant, Lemma 5.7 gives us that $B \cap X_0$ is $T_0$-invariant. Also as

$$\mu_n(B \cap J) + \mu_n(B^c \cap J) > \theta \mu_n(J)$$

for each $J \in \mathcal{J}_n$ we know that

$$\mu_0((B \cap X_0) \cap K) + \mu_0((B^c \cap X_0) \cap K) = \mu_n(B \cap K) + \mu_n(B^c \cap K)$$

$$> \theta \mu_n(K)$$

$$= \theta \mu_0(K)$$

for each subinterval $K \in \mathcal{J}$. It follows, from the conditions stated in the theorem that either $\mu_0(B \cap X_0) = 0$ or $\mu_0(B^c \cap X_0) = 0$.

Now suppose that both $B$ and $B^c$ have strictly positive $\mu_n$ outer measure. Then by Lemma 5.6 we would have that both $\mu_0(B \cap X_0)$ and $\mu_0(B^c \cap X_0)$ are strictly positive. This contradiction implies that either $B$ or $B^c$ must have zero $\mu_n$ outer measure.

\(\diamondsuit\)

Note that by choosing $\mathcal{J}$ to be subintervals of a space $X_0$ for which intervals make sense, the conditions on each $T$-invariant set $A$ imply, by the definition of a transformation that generates Carathéodory's definition, that $T$ generates Carathéodory's definition. That is, we have the following Corollary.

**Corollary 5.1**

Let $(X_n, B_n, \mu_n, T_n)$ be the usual $n$-level tower extension of a primitive dynamical system $(X_0, B_0, \mu_0, T_0)$ that generates the Carathéodory definition of
measurable sets. Now, let $B$ be a $T_n$-invariant subset of $X_n$. Suppose that there is a $\theta < 2$ such that for all subintervals $V$ of $X_n$

$$\mu_{n^*}(B \cap V) + \mu_{n^*}(B^c \cap V) \leq \theta \mu_n(V).$$

Then, either $B$ or $B^c$ is a set of measure zero.

We now show that we can prove almost identical results for usual $\infty$ level tower transformations. Lemmas 5.8 and 5.9 give the $\infty$ level tower extension results corresponding to Lemmas 5.6 and 5.7. Similarly Theorem 5.5 gives the corresponding result for $\infty$ level tower extensions to Theorem 5.4.

**Lemma 5.8**

Let $(X_0, B_0, \mu_0)$ be a measure space and $T_0$ be an invertible measure preserving transformation. Let $(X^*, B^*, \mu^*, T^*)$ be the usual $\infty$-level tower extension of $(T_0, B_0, \mu_0, T_0)$. Let $V \subseteq X^*$ be $T^*$-invariant, then if $\mu^*(V) > 0$, $\mu_{0^*}(V \cap X_0) = \mu^*(V \cap X_0) > 0$.

**Proof:**

Let $\mathcal{J}$ be a collection of basic sets for $(X_0, B_0, \mu_0)$ and let $\mathcal{J}^*$ be the tower collection of basic sets for $(X^* B^*, \mu^*)$.

As the levels of the tower are measurable, we have

$$0 < \mu^*(V)$$

$$= \mu^*(V \cap (X_0 \cup \bigcup_{n=0}^{\infty} B_n))$$

$$= \mu_{0^*}(V \cap X_0) + \sum_{n=1}^{\infty} \mu_{n^*}(V \cap B_{n-1}).$$

Thus either $\mu_{0^*}(V \cap X_0) > 0$ or there is an $m \in \mathbb{N}$ such that $\mu_{m^*}(V \cap B_{m-1}) > 0$. In the first case there is nothing more to show.
For the second case, note that since the \( \tau_i \) defined as the bijections between the levels of the tower are measure preserving bijections we know that for any \( C \in B_0 \), that is \( C \in B_{m-1}, C \subset X_0 \), we have that

\[
\mu \left( \circ_{j=0}^{m-1} \tau_{m-1-j}(C) \right) = \mu(C).
\]

Now, set some \( \varepsilon > 0 \). Then, by the definition of \( \mu_m \), there is a sequence of sets \( \{C_i\}_{i=0}^{\infty} \subset B_0 \) such that

\[
\circ_{j=0}^{m-1} \tau_j^{-1}(V \cap B_{m-1}) \subseteq \bigcup_{i=1}^{\infty} C_i
\]

and

\[
\mu_m \left( \circ_{j=0}^{m-1} \tau_j^{-1}(V \cap B_{m-1}) \right) > \left( \sum_{i=1}^{\infty} \mu_m(C_i) \right) - \varepsilon.
\]

Note that as

\[
\circ_{j=0}^{m-1} \tau_j^{-1}(V \cap B_{m-1}) \subseteq \bigcup_{i=1}^{\infty} C_i,
\]

\[
V \cap B_{m-1} \subseteq \bigcup_{i=1}^{\infty} \circ_{j=0}^{m-1} \tau_{m-1-j}(C_i)
\]

and hence

\[
\mu_m(V \cap B_{m-1}) \leq \sum_{i=1}^{\infty} \mu_m \left( \circ_{j=0}^{m-1} \tau_{m-1-j}(C_i) \right)
\]

so that we have

\[
\mu_m(V \cap X_0) \geq \mu_m \left( \circ_{j=0}^{m-1} \tau_j^{-1}(V \cap B_{m-1}) \right)
\]

\[
> \left( \sum_{i=1}^{\infty} \mu_m(C_i) \right) - \varepsilon
\]

\[
= \left( \sum_{i=1}^{\infty} \mu_m \left( \circ_{j=0}^{m-1} \tau_{m-1-j}(C_i) \right) \right) - \varepsilon
\]

\[
\geq \mu_m(V \cap B_{m-1}) - \varepsilon.
\]

(19)

As this is true for any \( \varepsilon > 0 \) we have

\[
\mu_0(V \cap X_0) = \mu_*(V \cap X_0)
\]
Lemma 5.9

Let \((X_0, B_0, \mu_0)\) be a measure space, and \(T_0\) be an invertible measure preserving transformation on \(X_0\). Let \((X^*, B^*, \mu^*, T^*)\) be the usual \(\infty\)-level tower extension of this space. Let

\[ B = \bigcap_{n=0}^{\infty} \{ x \in X_0 : T^{*n}(x) \in B_n \} \]

and

\[ \Psi = \bigcup_{n \in \mathbb{Z}} T^{*n}(B). \]

If \(V \subset X^*\) is \(T^*\)-invariant, then \((V \cap X_0 \cap \Psi^c)\) is \(T_0\)-invariant.

Proof:

Let \(V\) be a \(T^*\)-invariant subset of \(X^*\).

Now, let \(x \in T_0^{-1}(V \cap X_0 \cap \Psi^c)\) so that \(T_0(x) \in V \cap X_0 \cap \Psi^c\). In this case we have \(x \in X_0\) and must consider the two cases of \(x \in A_0 \cap \Psi^c\) and \(x \in A_0^c \cap \Psi^c\). In the second case

\[ T_0(x) = T^*(x) \in V \cap X_0 \cap \Psi^c \subset V \cap \Psi^c. \]

This implies that

\[ x \in (V \cap X_0 \cap \Psi^c). \]

(Where we have used that \(V\) is \(T^*\)-invariant and that \(T^*(x) \in \Psi^c \iff x \in \Psi^c\).)

In the case where \(x \in A_0\), we have that \(x \in X_0\) and as \(x \notin \Psi\) there must
exist some minimum \( k \in \mathbb{N} \) such that \( T^{*k}(x) \notin B_{k-1} \). For this \( k \)

\[
T_0(x) = T^{*k}(x) \in V \cap X_0 \cap \Psi^c
\]

and in a similar way to the case just discussed this implies that

\[
x \in V \cap X_0 \cap \Psi^c.
\]

We now have that \( x \in T^{-1}(V \cap X_0 \cap \Psi^c) \Rightarrow x \in V \cap X_0 \cap \Psi^c \). That is, 

\[
T^{-1}(V \cap X_0 \cap \Psi^c) \subset V \cap X_0 \cap \Psi^c.
\]

Conversely, let \( x \in V \cap X_0 \cap \Psi^c \). Clearly \( x \in X_0 \) and hence \( T_0(x) \in X_0 \). Also, as \( x \in V \cap \Psi^c \), if \( x \in A_0^c \) then \( T_0(x) = T^*(x) \in V \cap \Psi^c \). Otherwise \( x \in A_0 \) and in this case as \( x \notin \Psi \) there is a \( k \in \mathbb{N} \) such that \( T_0(x) = T^{*k}(x) \in V \cap \Psi^c \) and hence \( T_0(x) \in V \cap X_0 \cap \Psi^c \). Therefore, we now have

\[
x \in V \cap X_0 \cap \Psi^c \Rightarrow T_0(x) \in V \cap X_0 \cap \Psi^c \Rightarrow x \in T_0^{-1}(V \cap X_0 \cap \Psi^c).
\]

Meaning that \( V \cap X_0 \cap \Psi^c \subset T_0^{-1}[V \cap X_0 \cap \Psi^c] \) and therefore

\[
V \cap X_0 \cap \Psi^c = T_0^{-1}(V \cap X_0 \cap \Psi^c).
\]

\[\diamondsuit\]

**Theorem 5.6**

Let \((X^*, B^*, \mu^*, T^*)\) be the usual \( n \)-level tower extension of a primitive dynamical system \((X_0, B_0, \mu_0, T_0)\), let \( J \) be a chosen collection of basic sets for \((X, B, \mu)\) and let \( J^* \) be the tower collection of basic sets for \( X^* \). Let each \( T \)-invariant subset \( A \) of \( X_0 \) for which there is a \( \theta = \theta(A) < 2 \) such that for all \( J \in J \)

\[
\mu_0^*(A \cap J) + \mu_0^*(A^c \cap J) \leq \theta \mu_0(J),
\]

be such that either \( A \) or \( A^c \) is a set of outer measure zero. Now, let \( D \) be a \( T^* \)-invariant subset of \( X^* \). Suppose that there is a \( \theta < 2 \) such that for all \( V \in J^* \)

\[
\mu^*(D \cap V) + \mu^*(D^c \cap V) \leq \theta \mu^*(V).
\]
Then, either $D$ or $D^c$ is a set of measure zero.

**Proof:**

As in Lemma 5.9, let

$$B = \bigcap_{n=0}^{\infty} \{ x \in X_0 : T^{*n}(x) \in B_n \}$$

and

$$\Psi = \bigcup_{n \in \mathbb{Z}} T^{*n}(B).$$

As

$$\mu^*(X^*) < +\infty$$

and

$$\mu^*(X^*) = \mu^*(X_0) + \sum_{n=0}^{\infty} \mu^*(B_n),$$

we know that for all $\varepsilon > 0$ there is an $n_0 \in \mathbb{N}$ such that

$$\sum_{n=n_0}^{\infty} \mu^*(B_n) < \varepsilon.$$

Since

$$\mu^*(B) = \mu^* \left( \bigcap_{n=0}^{\infty} \{ x \in X_0 : T^{*n}(x) \in B_n \} \right)$$

$$= \lim_{j \to \infty} \sum_{n=j}^{\infty} \mu^*(o_{i=0}^{n-1} \tau^{-1}_n(B_n))$$

$$= \lim_{j \to \infty} \sum_{n=j}^{\infty} \mu^*(B_n),$$

it follows that for all $\varepsilon > 0$, $\mu^*(B) < \varepsilon$ and thus $\mu^*(B) = 0$. Then as for all $n \in \mathbb{Z}$, $\mu^*(T^{*n}(B)) = \mu^*(B)$ we have that

$$\mu^*(\Psi) = \mu^* \left( \bigcup_{n \in \mathbb{Z}} T^{*n}(B) \right) \leq \sum_{n \in \mathbb{Z}} \mu^*(T^{*n}(B)) = 0.$$

Now, if $\mu^*(D) > 0$ and $\mu^*(D^c) > 0$, then by Lemma 5.8 $\mu^*(D \cap X_0) > 0$ and $\mu^*(D^c \cap X_0) > 0$ Also, from Lemma 5.9, we have that $D \cap X_0 \cap \Psi^c$ is
$T_0$-invariant and as $\mu^*(\Psi) = 0$, $\mu^*(D \cap X_0 \cap \Psi^c) = \mu^*(D \cap X_0) > 0$. It follows that $(D \cap X_0 \cap \Psi^c)^c$ is $T_0$-invariant and $\mu^*((D \cap X_0 \cap \Psi^c)^c) > \mu^*(D^c \cap X_0) > 0$. By the stated conditions $T_0$-invariant sets, we cannot have $\mu^*(D^c \cap X_0) > 0$ and $\mu^*(D \cap X_0) > 0$, therefore either $D$ or $D^c$ has $\mu^*$ outer measure zero.

Again, we note that should we chose subintervals of a space on which intervals makes sense as our collection of basic sets, we would have that $T_0$ generates the Carathéodory definition of measurable sets. Hence we can state the following Corollary.

**Corollary 5.2**

Let $(X^*, \mathcal{B}^*, \mu^*, T^*)$ be the usual $\infty$-level tower extension of the primitive dynamical system $(X_0, \mathcal{B}_0, \mu_0, T_0)$ which generates the Carathéodory definition of measurable sets. Now, let $D$ be a $T^*$-invariant subset of $X^*$. Suppose that there is a $\theta < 2$ such that for all subintervals $V$ of $X^*$

$$\mu^*(D \cap V) + \mu^*(D^c \cap V) \leq \theta \mu^*(V).$$

Then, either $D$ or $D^c$ is a set of outer measure zero.

Note that the concept of splinters has not been used very regularly, however, the use was very important. By way of splinters we are able to allow the sets (at this stage we're still only actually considering intervals) to break up into finitely many pieces under iterations by a transformation and to ‘arrive’ in the desired subset of the space in their own time. Due to such information, when we consider a class of transformations defined by possessing certain characteristics, we end up with a large collection of transformations. The results of this investigation are discussed in the next chapter.

We have also seen that towers are essentially impotent when it comes to
generating Carathéodory's definition in that the generation has no effect on whether a transformation will generate the Carathéodory definition or not. We can therefore choose any extension on a primitive space that satisfies the mixing properties we would like. However, an even stronger result is true. That is, if a tower transformation generates Carathéodory's definition, then so too does the primitive space. This gives us the further comfort of being able to choose the tower extension for which proving the generation of Carathéodory's definition is easiest when demonstrating that a given type of transformation generates the definition in the first place.

This equivalence result is proved below.

**Theorem 5.7**

Suppose that \((X_0, \mathcal{B}_0, \mu_0)\) is a finite measure space, that \(T_0\) is an invertible measure preserving transformation on \(X_0\) and that \(\mathcal{J}\) is a collection of basic sets for \(X_0\). Let \((X^*, \mathcal{B}^*, \mu^*, T^*)\) be a usual tower extension of either a finite or infinite type and \(\mathcal{J}^*\) be a tower collection of basic sets for \(X^*\). Then the following two properties are equivalent.

(i) \(T_0\) has the property that if, whenever \(A\) is a \(T_0\)-invariant subset of \(X\) for which there is a \(\theta = \theta(A) < 2\) such that for all \(J \in \mathcal{J}\)

\[
\mu_0^*(A \cap J) + \mu_0^*(A^c \cap J) \leq \theta \mu_0(J),
\]

either \(A\) or \(A^c\) is a set of outer measure zero.

(ii) \(T^*\) has the property that if, whenever \(A\) is a \(T^*\)-invariant subset of \(X\) for which there is a \(\theta = \theta(A) < 2\) such that for all \(J \in \mathcal{J}^*\)

\[
\mu^*(A \cap J) + \mu^*(A^c \cap J) \leq \theta \mu^*(J),
\]

179
either A or Ac is a set of outer measure zero

Proof:
If $T_0$ has property (i), then depending on whether $(X^*, B^*, \mu^*, T^*)$ is a finite or infinite extension Theorem 5.5 or Theorem 5.5 gives the required result. We therefore need only show that if $T^*$ has property (ii) then $T_0$ has property (i).

Now suppose that if $A$ is a $T^*$-invariant subset of $X^*$ such that there is a $\theta < 2$ so that for all $J \in \mathcal{J}^*$

$$\mu^*(A \cap J) + \mu^*(A^c \cap J) \leq \theta \mu^*(J)$$

then $\mu^*(A) = 0$ or $\mu^*(A^c) = 0$. Next, let $A_0$ be a $T_0$-invariant subset of $X_0$ and suppose that there is a $\theta < 2$ such that for all $J_0 \in \mathcal{J}$

$$\mu_0^*(A_0 \cap J_0) + \mu_0^*(A^c_0 \cap J_0) \leq \theta \mu_0(J_0).$$

We now need to show that either $\mu_0(A_0) = 0$ or $\mu_0(A^c_0) = 0$.

Let $A_n = \bigcup_{i \in \mathbb{Z}} (T^*)^i(A_0)$ so that $A_n$ is $T^*$-invariant and note that $A_n \cap X_0 = A_0$.

Suppose there is a $Z \in \mathcal{J}^*$ such that

$$\mu^*(A_n \cap Z) + \mu^*(A^c_n \cap Z) > \theta \mu^*(Z).$$

Now, $Z \subset B_k$ for some $k \in \mathbb{N}$ and so $(T^*)^{-k-1}(Z) \subset X_0$, and we will denote $(T^*)^{-k-1}(Z)$ by $Z_0$. We then note that $(T^*)^{-k-1}(A_n \cap Z) \subset A_0 \cap Z_0$ and that $(T^*)^{-k-1}(A^c_n \cap Z) \subset A^c_0 \cap Z_0$. Thus using that $A_n$ is $T^*$-invariant and that $T^*$ is measure preserving we have

$$\mu_0^*(A_0 \cap Z_0) + \mu_0^*(A^c_0 \cap Z_0) \geq \mu_0^*((T^*)^{-k-1}(A_n \cap Z)) + \mu_0^*((T^*)^{-k-1}(A^c_n \cap Z))$$

$$= \mu^*((T^*)^{-k-1}(A_n \cap Z)) + \mu^*((T^*)^{-k-1}(A^c_n \cap Z))$$

$$= \mu^*(A_n \cap Z) + \mu^*(A^c_n \cap Z)$$

$$> \theta \mu^*(Z)$$

$$= \theta \mu^*(Z_0)$$

180
which contradicts the conditions on $A_0$ in $X_0$ and thus such a $Z$ does not exist. We now have that for the $T^*$-invariant set $A_n$ for all $Z \subset X_n$, $Z \in \mathcal{J}^*$

$$\mu_*(A \cap Z) + \mu_*(A^c \cap Z) \leq \theta \mu^*(Z).$$

Hence either $\mu^*(A_n) = 0$ and thus $\mu_0*(A_0) = \mu_*(A_0) \leq \mu_*(A_n) = 0$ or $\mu^*(A^c_n) = 0$ and thus $\mu_0(A^c_0) = \mu_*(A^c_0) \leq \mu_*(A^c_n) = 0$. That is either $\mu_0*(A_0) = 0$ or $\mu_0*(A^c_0) = 0$. This is the required result and thus the proof is complete.

As with the previous two theorems, choosing our collection of basic sets to be subintervals of a set $X_0$ on which the concept of intervals makes sense we arrive at a Corollary of particular interest to us at this time.

**Corollary 5.3**

*Suppose that $(X_0, B_0, \mu_0)$ is a finite measure space and that $T_0$ is an invertible measure preserving transformation on $X_0$. Let $(X^*, B^*, \mu^*, T^*)$ be a usual tower extension of either a finite or infinite type. Then $(X_0, B_0, \mu_0, T_0)$ generates the Carathéodory definition of measurable sets if and only if $(X^*, B^*, \mu^*, \psi^*)$ does.*

One of the aims of considering tower transformations was to find constructions of transformations that generate Carathéodory's definition of measurable sets that would perhaps have stronger mixing properties than weak mixing. We do not know what the mixing properties are of all of the transformations that can be constructed by tower extensions of transformations that generate Carathéodory's definition. However, we can show that increasing the number of tower levels doesn't necessarily strengthen the mixing properties. We do this below by providing the proof that there is a weak but not strong mixing transformation for any finite level tower construction.
We also show that there is an infinite level weak mixing construction by constructing an infinite level induced skyscraper transformation.

In the following results we continue to use the established notation for tower constructions, that is we use $B_i$ to denote the $i+1$th level of the tower, $\tau_i$ to denote the bijection between the $i$th and the $i+1$th level of the tower and so forth.

**Theorem 5.8**

*For each $n > 1$ there is a usual $n$ level tower construction with a Kakutani I type primitive transformation that is weak but not strong mixing.*

**Proof:**

Let $P$ be a Kakutani I type transformation defined so that $P$-orbits of $b$-adic intervals are essentially the set of all $b$-adic intervals of the same order. Let $n \in \mathbb{N}$. We construct the weak mixing but not strong mixing usual $n$ level tower transformation as follows. Choose any

$$A_0 = \bigcup_{n=1}^{\infty} C_n$$

where each $C_n$ is a $b$-adic interval of order $2n-1$, \{\$C_n$\} is a disjoint sequence and the $C_n$ satisfy the property that for each $n \in \mathbb{N}$ there is a $b$-adic interval of order $n$, $J$, such that

$$\bigcup_{m=n+1}^{\infty} C_m \subset J.$$  

An example of such a sequence of $C_n$'s is the sequence $I_n$ in Kakutani's original definition. We then construct the remaining levels by using $A_i$'s defined by

$$A_i = \tau_{i-1} \left( A_{i-1} - o^{2i-2}_{j=0} \tau_j(C_i) \right),$$

for $i \in \{1, 2, ..., n-1\}$ where for each $i \in \{0, 1, ..., n-2\} A_{i+1}$ is any set such that $(X \cup \bigcup_{j=0}^{i} A_j) \cap A_{i+1} = \emptyset$ and $\tau_i$ is any bijection mapping $A_i$ to $A_{i+1}.$
We denote the resulting tower extension by \((X_n, B_n, \mu_n, P_n)\) and show that this transformation is weak but not strong mixing.

In order to show that \(\psi_n\) is weak mixing on \((X_n, B_n, \mu_n)\) assume that there is an eigenvalue \(e^{2\pi i \lambda}\) such that \(\lambda \not\equiv 0 (\text{mod} 1)\) with an associated eigenfunction \(f \in L^2(\mu_n)\) so that

\[ f(P_n(x)) = e^{2\pi i \lambda} f(x) \quad \text{for all } x \in X_n. \]

By Lemma 4.9 we know that for any such eigenfunction \(|f|\) is a constant and so we can assume that \(|f| = 1\). Following the same style of proof as for Kakutani’s original Kakutani I type function we define a function \(u\) to give us the number of iterations of \(P_n\) required to have the same effect as \(P_0\). That is for each \(x \in [0, 1)\)

\[ u(x) = 1 + \sum_{i=0}^{n-1} \chi_{A_i}(P_n^i(x)) \]

and we can then define

\[ u_m = \sum_{i=0}^{m-1} u(P_0^i). \]

The inclusion of the greater number of characteristic functions in the definition of \(u\) follows as there are (potentially) a greater number of levels for \(x\) to pass through before returning to the primitive space. Then, analogously to Kakutani we find that

\[ u_{bm} \in \{W_m, W_m + 1, ..., W_m + (n - 1)\} \]

where

\[ W_m = b^n + \sum_{j=1}^{n} (n - j + 1) \sum_{i=1}^{k_j} b^{n-(2(j-1))-(2i-1)} \]

183
and $k_j$ is the number of terms in the sequence

$$\{x = n - (2(j - 1)) - (2i - 1) : i \in \mathbb{N}, x \geq 0\}$$

The motivation of this $W_m$ comes from the fact that each element of $X_0$ is an element of a $b^m$ adic interval and as the images of the $A_j$'s into $X_0$ under $\sigma_{i=0}^{j-1}(A_i)$ are $b$-adic intervals of some order. For this reason $u_n$ will be constant on all $b^m$ adic intervals except on the interval, $I$, in which $A \cap I$ consists of $b$-adic intervals of higher order than $m$. Note that in the definition of the $A_j$'s we required that there would be only one such $b$-adic interval of order $m$. Then the $u_{b^m}$ value will be the sum of all these constant values (as the orbit of $\psi_0$ takes the $x$ through each $b$-adic interval of order $m$) plus whatever value it takes when it is in the non-constant valued interval.

Note also that we must have that $\mu_n \{x \in X_0 : u_{b^m}(x) = W_m + i\} \neq 0$ for each $i \in \{0, 1, ..., n - 1\}$ since $\tilde{\mu}(A_i) < \tilde{\mu}(A_j)$ whenever $i > j$. Now

$$\int_{X_0} |f(P_n^{b^m}(x)) - f(x)|^2 d\mu(x)$$

$$= \int_{X_0} |f(P_0^{u_{b^m}}(x))(x)) - f(x)|^2 d\mu(x)$$

$$= \int_{X_0} |e^{2\pi i u_{b^m}(x)\lambda} - 1|^2 d\mu(x)$$

$$= \sum_{j=0}^{n-1} |e^{2\pi i (W_m + j)} - 1| \mu(\{x \in X_0 : u_{b^m}(x) = W_m + j\})$$

We also know, by Lemma 4.8 that this integral approaches zero with $m$ so that

$$\lambda(W_m + j) \equiv 0 (mod 1)$$

for all

$$j \in \{0, 1, 2, ..., n - 1\}$$

so that

$$\lambda \equiv 0 (mod 1)$$

184
which is a contradiction and thus \( P_n \) has a continuous spectrum and thus \( P_n \) is weakly mixing.

To show that \( P_n \) is not strong mixing we note that \( P_n^{um}(x) = P_0^{m}(x) \) for all \( x \in X_0 \). Thus for any \( b \)-adic interval, \( B \), of order less than or equal to \( m \),

\[
P_n^{um}(B) \subseteq \bigcup_{j=0}^{n-1} P_n^{W_m+j}(B)
\]

\[
= \bigcup_{j=0}^{n-1} P_n^j(P_0^{b^m}(B))
\]

\[
= \bigcup_{j=0}^{n-1} P_n^j(B)
\]

Then select a \( m_0 \in \mathbb{N} \) so that \( b^{m_0} > n \). Now set \( A = \bigcup_{j=0}^{n-1} P_n^j(B) \) and \( B = X_n - \bigcup_{j=0}^{n-1} P_n^j(B) \). \( \mu_n(A), \mu_n(B) > 0 \) and thus \( \mu_n(A)\mu_n(B) > 0 \), however for all \( m > m_0 \), \( \mu_n(P_n^{um}(A) \cap B) = 0 \), and so \( P_n \) is not strongly mixing. \( \diamond \)

We now construct an infinite tower (an induced skyscraper) that is weak mixing but not strong mixing. This example is due to Chacon [8].

We start with the usual Lebesgue measure space, \((X, B, \mu)\) on \( X = [0, 1) \). The general idea is to split the interval \([0, 1)\) up into disjoint measurable parts (in fact subintervals of equal length), which will be labelled as the sequence \( \{I_n\}_{n=1}^{\infty} \). We then allow the transformation to map \( I_k \) linearly onto \( I_{k+1} \) for each \( k \). When these disjoint intervals are chosen appropriately the transformation has the desired properties. As it is somewhat impossible to choose a countably infinite set of subintervals of equal length, the intervals are generated by an iterative process. Due to the definition of the transformation's dependence on intervals of equal length the idea behind how we will show that the transformation generates Carathéodory's definition is imme-
•••clear.

The iterative generation of the transformation is described below.

STEP 1:
We divide the unit interval into three subintervals \([0, 1/3), [1/3, 2/3), [2/3, 1)\) and denote these intervals by \(R, I_1, I_2^1\). We then define the transformation \(\tau_1\) on \(I_1\) to be a linear measure-preserving mapping of positive slope (that is \(x \mapsto x + 1/3\)) onto \(I_2^1\) and \(\tau_1 = 0\) elsewhere.

STEP n+1:
We suppose that \([0, 1)\) is partitioned into \(p(n) + 1\) intervals \(\{R_n, I_{n,1}^1, ..., I_{n,p(n)}^n\}\) of equal length and such that \(\mu(R_n) + \sum_{i=1}^{p(n)} \mu(I_n^i) = 1\). Also we assume that \(\tau_n\) has so far been defined on \(\bigcup_{i=1}^{p(n)-1} I_n^i\) as a linear measure preserving mapping of positive slope of \(I_n^k\) onto \(I_{n+1}^{k+1}\) for each \(k \in \{1, 2, ..., p(n) - 1\}\).

We then form the partition \(\{R_{n+1}, I_{n+1,1}^1, I_{n+1,2}^2, ..., I_{n+1,p(n)+1}^{n+1}\}\) as follows. We write each interval as the union of two consecutive intervals of equal length \(I_{n,1}^i = I_{n,1}^i + I_{n,2}^i\) and similarly we write \(R_n\) as the union of two consecutive intervals \(R_n = A_{n+1} + B_{n+1}\) where \(\mu(I_{n,1}^1) = \mu(B_{n+1})\). Then

\[R_{n+1} = A_{n+1},\]
\[I_{n+1}^k = I_{n,1}^k \quad \text{for each } k \in \{1, 2, ..., p(n)\},\]
\[I_{n+1}^{p(n)+k} = I_{n,2}^k \quad \text{for each } k \in \{1, 2, ..., p(n)\},\]

and
\[I_{n+1}^{p(n)+1} = I_{n+1}^{2p(n)+1} = B_{n+1}.\]

We have that \(\tau_n\) is a linear measure-preserving map of positive slope from \(I_{n+1}^k\) onto \(I_{n+1}^{k+1}\) for each \(k \in \{1, 2, ..., p(n)-1, p(n)+1, p(n)+1, ..., 2p(n)-1\}\).
and we now define \( r_{n+1} \) to equal \( r_n \) on \( \bigcup_{i=1}^{p(n)-1} I_i \), and on \( I_{n,1}^{p(n)} \) and \( I_{n,2}^{p(n)} \) we define \( r_n \) as a linear measure-preserving map onto \( I_{n,1}^1 \) and \( B_{n+1} \) respectively and \( r_{n+1} = 0 \) elsewhere.

Note that as we start with 3 intervals and at each stage we double the number of intervals, we must have that \( p(n) = 3 \cdot 2^n - 1 \) for each \( n \in \mathbb{N} \). Also note that \( r_n \) is defined linearly on all but the 2 intervals coming from \( R_{n-1} \) at each stage. That is, \( r_n \) is defined linearly on \( p(n) - 1 \) of the \( p(n) + 1 \) intervals that we divide \([0, 1)\) into at each stage.

We now define our weakly but not strongly mixing transformation as

\[
\xi = \lim_{n \to \infty} r_n.
\]

As the \( r_n \) are defined on an increasing sequence of sets whose union is \( X \) and as the \( r_n \) agree wherever their domains intersect the limit, \( \xi \) is well defined and is defined for all \( x \in X \).

Before proving the desired properties of \( \xi \) we prove a convenient lemma.

**Definition 5.8**

A \((2, 3)\)-adic interval of order \( n \) is a subinterval of \([0, 1)\) of the form

\[
\left[ \frac{k}{3 \cdot 2^n}, \frac{k+1}{3 \cdot 2^n} \right)
\]

with \( k \in \{0, 1, 2, \ldots, 3 \cdot 2^n \} \).

**Lemma 5.10**

Let \((X, B, \mu)\) be the usual Lebesgue measure space on \( X = [0, 1] \) and \( B \subseteq X \) be such that \( \mu_*(B) > 0 \) where \( \mu_* \) is the usual outer measure associated with \( \mu \). Then for any \( \varepsilon > 0 \) and all sufficiently large \( n \) there is a \((2, 3)\)-adic interval
\[ \frac{[k/(3.2^n), (k + 1)/(3.2^n)], k \neq 0, \text{ that is an interval of density for } B \]

**Proof:**

Let \( \varepsilon > 0 \). If \( \mu_*(B) > 0 \) then there is an \( n_0 \in \mathbb{N} \) such that \( 2^{-n_0} < \mu_*(B) \) so that \( \mu_*(B \cap (X - [0, 3 \cdot 2^{n_0}])) > 0 \). Then by Lemma 4.7 for all sufficiently large \( n \) there is a dyadic interval of density of order \( n \) \( J_0 \) to within \( \varepsilon \) for \( B \cap (X - [0, 3 \cdot 2^{n_0}]) \) and hence also for \( B \).

\( J_0 \) can be split into three \((2,3)\)-adic intervals of order \( n \) \( J_1, J_2, J_3 \). Should none of these be an interval of density to within \( \varepsilon \) for \( B \) then

\[
\mu_*(B \cap J_0) \leq \sum_{i=1}^{3} \mu_*(B \cap J_i) < \sum_{i=1}^{3} (1 - \varepsilon) \mu(J_i) = (1 - \varepsilon) \mu(J_0).
\]

Therefore there is a \((2,3)\)-adic interval of density of order \( n \), \( J \subseteq J_0 \), to within \( \varepsilon \) for \( B \). As \( J \subseteq J_0 \) is an interval of density for \( B \cap (X - [0, 3.2^{n_0}]) \), clearly \( J_0 \cap [0, 3.2^{n_0}) = \emptyset \) and hence \( J \neq [0, 3.2^{n_0}) \). \( \diamond \)

We now prove directly that \( \xi \) generates Carathéodory's definition of measurable sets.

**Theorem 5.9**

*Chacon's \( \xi \) generates the Carathéodory definition of measurable sets.*

**Proof:**

We note that as \( \xi \) agrees with \( \tau_n \) on \( X - R_n \) for each \( n \in \mathbb{N} \) \( \xi \) is a linear measure preserving function on intervals of the form

\[
\begin{bmatrix}
\frac{p}{3.2^n} & \frac{p + 1}{3.2^n}
\end{bmatrix}
\]

188
on $X - R_n$. That is, on
\[
\left[ \frac{p}{3 \cdot 2^n}, \frac{p + 1}{3 \cdot 2^n} \right]
\]
for $p \in \{1, 2, \ldots, 3 \cdot 2^n - 1\}$. Note also that each such interval is the interval $I_{n+1}^k$ for some $k$ and that
\[
\{I_{n+1}^k, I_{n+1}^{k-1}, \ldots, I_{n+1}^1\} \subset O_\xi(I_{n+1}^k).
\tag{20}
\]

Now we wish to prove that if $B$ is a $\xi$-invariant set such that there is a $\theta < 2$ with the property that
\[
\mu_*(B \cap J) + \mu_*(B^c \cap J) \leq 2\mu(J)
\]
for each interval $J$ then either $B$ or $B^c$ has outer measure zero.

Suppose that $\mu_*(B) > 0$ and $\mu_*(B^c) > 0$, then by Lemma 5.10 for any sufficiently large $n$ there is a $(2, 3)$-adic interval of density of order $n$, $J$, to within $\varepsilon$ for $B$. Similarly there is a $(2, 3)$-adic interval of density of order $n$, $K$, to within $\varepsilon$ for $B^c$. Note also that, again by Lemma 5.10, $J$ and $K$ can be chosen so that $[0, (3 \cdot 2^n)^{-1}] \not\subset \{J, K\}$.

Then by (20) either $K \in O_\xi(J)$ or $J \in O_\xi(K)$. Without loss of generality suppose that $K \in O_\xi(J)$, that is, suppose that $K = \xi^{-p}J$ for some $p \in \mathbb{N}$. As $\xi$ is measure preserving and as $B$ is $\xi$-invariant, we now have
\[
\mu_*(B^c \cap J) = \mu_*(\xi^{-p}(B^c \cap K)) = \mu_*(B^c \cap K) > (1 - \varepsilon)\mu(K) = (1 - \varepsilon)\mu(J),
\]
so that $J$ is an interval of density for $B^c$. Thus
\[
\mu_*(B \cap J) + \mu_*(B^c \cap J) > 2(1 - \varepsilon)\mu(J).
\]

189
Hence for an appropriate choice of $\varepsilon$ we have contradicted that

$$\mu_*(B \cap J) + \mu_*(B^c \cap J) \leq 2\mu(J)$$

for each interval $J$. Therefore either $\mu_*(B) = 0$ or $\mu_*(B^c) = 0$.

Following a technical lemma describing a property of $\xi$, we prove that $\xi$ is weak mixing but not strong mixing.

**Lemma 5.11**

*For each $n \in \mathbb{N}$ and each $k \in \{1, 2, \ldots, p(n)\}$*

$$\mu(I_n^k \cap \xi^{p(n)}(I_n^k)) = \frac{1}{2}\mu(I_n^k)$$

and

$$\mu(I_n^k \cap \xi^{p(n)+1}(I_n^k)) = \frac{1}{4}\mu(I_n^k).$$

**Proof:**

By the definition of $\xi$ we know that $\xi(I_n^k) = I_n^{k+1}$ (unless $k = p(n)$) and hence that $\mu(\xi(I_n^k)) = \mu(I_n^k)$. Inductively we get (or immediately if $k \in \{p(n), p(n) - 1\}$) that $\xi^{p(n)-k}(I_n^k) = I_n^{p(n)}$ and hence that $\mu(\xi^{p(n)-k}(I_n^k)) = \mu(I_n^{k})$. Now note that $\xi$ maps half of $I_n^{p(n)}$ linearly onto half of $I_n^1$ and half of $I_n^{p(n)}$ linearly onto half of $R_n$. In the latter case we know that $\xi$ maps half of $\xi(I_n^{p(n)}) \cap R_n$ linearly onto one quarter of $I_n^1$ and the other half onto $R_{n+1}$. Also the half of $\xi^{p(n)-k+1}(I_n^k)$ that is in $I_n^1$ is mapped by $\xi$ into $I_n^2$. Hence we have accounted for all of $\xi^{p(n)-k+1}(I_n^k)$ and $\xi^{p(n)-k+2}(I_n^k)$ which gives us that

$$\mu(\xi(I_n^{p(n)-k+1}) \cap I_n^1) = \frac{1}{2}\mu(I_n^k)$$

and

$$\mu(\xi(I_n^{p(n)-k+2}) \cap I_n^1) = \frac{1}{4}\mu(I_n^k).$$
Now we use that $\xi$ maps $I_n^1$ linearly onto $I_n^2$ to get that

$$
\mu(\xi(I_n^{p(n)-k+2}) \cap I_n^2) = \mu(\xi(I_n^{p(n)-k+1}) \cap I_n^1) = \frac{1}{2}\mu(I_n^k)
$$

and

$$
\mu(\xi(I_n^{p(n)-k+3}) \cap I_n^2) = \mu(\xi(I_n^{p(n)-k+2}) \cap I_n^1) = \frac{1}{4}\mu(I_n^k).
$$

Finally, since $\xi$ maps $I_n^j$ linearly onto $I_n^{j+1}$ for each $j \in \{1, 2, ..., p(n) - 1\}$, and since $k \in \{1, 2, ..., p(n)\}$ we can repeat the previous step $k - 1$ times to get that

$$
\mu(\xi(I_n^{p(n)}) \cap I_n^k) = \mu(\xi(I_n^{p(n)-1}) \cap I_n^{k-1})
$$

= ...

= $\mu(\xi(I_n^{p(n)-k+1}) \cap I_n^1)$

= $\frac{1}{2}\mu(I_n^k)$

and

$$
\mu(\xi(I_n^{p(n)+1}) \cap I_n^k) = \mu(\xi(I_n^{p(n)}) \cap I_n^{k-1})
$$

= ...

= $\mu(\xi(I_n^{p(n)-k+2}) \cap I_n^1)$

= $\frac{1}{4}\mu(I_n^k)$.

\[\diamondsuit\]

**Theorem 5.10**

*Chacon's $\xi$ is weak but not strong mixing.*

**Proof:**

We show that $\xi$ is weak mixing via Theorem 3.10. That is, we show that $\xi$ has a continuous spectrum. We therefore need to show that the only eigenvalue for $\xi$ is 1 and the only eigenfunctions are constants.
Since we can write any \( f \in L^2(X, \mathcal{B}, \mu) \) as the limit of a sequence of simple functions and since any set in \( \mathcal{B} \) can (up to a set of measure zero) be constructed by countable unions and intersections of \((2,3)\)-adic intervals we can write any \( f \in L^2(X, \mathcal{B}, \mu) \) as the limit of a sequence of simple functions \( \{f_n\} \) such that \( f_n \) is constant on \((2,3)\)-adic intervals of order \( n \).

It follows that for any \( \varepsilon > 0 \) there is an \( n \in \mathbb{N} \), an \( \alpha \in \mathbb{R} \) and a \((2,3)\)-adic interval of order \( n \), \( I_n^k \) (for some \( k \)), such that

\[
|f(x) - \alpha| < \varepsilon
\]

for all \( x \in I_n^k \) apart for perhaps a set of measure less than \( \varepsilon \mu(I_n^k) \).

Let \( \lambda \) be an eigenvalue for \( \xi \) with eigenfunction \( f \in L^2(X, \mathcal{B}, \mu) \), so that

\[
f(\xi(x)) = e^{2\pi i \lambda} f(x) \quad \text{for all } x \in X.
\]

Let \( 0 < \varepsilon < 1/4 \) and let \( I_n^k \) and \( \alpha \) be chosen such that \( |f(x) - \alpha| < \varepsilon \) for all \( x \in I_n^k \) apart from at most a set of measure less than \( \varepsilon \). Then, by Lemma 5.11 there is an \( x \in I_n^k \) for which we can find

\[
\delta_1 = f(x) - \alpha \quad \text{and}
\]

\[
\delta_2 = f(\xi^{p(n)}(x)) - \alpha,
\]

such that \( |\delta_1|, |\delta_2| < \varepsilon \). Similarly there must be an \( x \in I_n^k \) for which we can find

\[
\delta_3 = f(x) - \alpha \quad \text{and}
\]

\[
\delta_4 = f(\xi^{p(n)+1}(x)) - \alpha,
\]

such that \( |\delta_3|, |\delta_4| < \varepsilon \). Then as

\[
f(\xi^{p(n)}(x)) = e^{2\pi i \lambda p(n)} f(x)
\]
and
\[ f(\xi^{p(n)+1}(x)) = e^{2\pi i \lambda (p(n)+1)} f(x) \]
we have
\[ \alpha + \delta_2 = e^{2\pi i \lambda p(n)}(\alpha + \delta_1) \quad \text{and} \]
\[ \alpha + \delta_4 = e^{2\pi i \lambda (p(n)+1)}(\alpha + \delta_3) \]
so that
\[ e^{2\pi i \lambda} = \frac{(\alpha + \delta_1)(\alpha + \delta_4)}{(\alpha + \delta_3)(\alpha + \delta_2)}. \]
As this true for each 0 < \varepsilon < 1/4 and hence each 0 < |\delta_1|, |\delta_2|, |\delta_3|, |\delta_4| < 1/4 we have
\[ e^{2\pi i \lambda} = 1. \]

In order to show that \( \xi \) is weak mixing we now need only show that any eigenfunction for the eigenvalue 1 is a constant function \( \mu \) almost everywhere.

Suppose there is an eigenfunction for the eigenvalue 1 that is not constant \( \mu \text{-a.e.} \) Then there is a \( \beta \in \mathbb{R} \) such that
\[ B = \{x : f(x) < \beta\} \in \mathcal{B}, \]
\[ \mu(B) > 0, \]
\[ B^c = \{x : f(x) \geq \beta\} \in \mathcal{B} \quad \text{and} \]
\[ \mu(B^c) > 0. \]

Then by Lemma 5.10 there exist (2,3)-adic intervals of density of order \( n \), \( I_{n+1}^k \) and \( I_{n+1}^j \) to within 1/2 for \( B \) and \( B^c \) respectively. Further,
\[ [0, (3 \cdot 2n)^{-1}) \notin \{I_{n+1}^k, I_{n+1}^j\}. \]
Assume, without loss of generality that \( k > j \). Then as for each \( x \in X \)
\[ f(\xi^p(x)) = (e^{2\pi i \lambda})^p f(x) = f(x) \]
for any $p \in \mathbb{N}$ we have

\[
\mu(B^c \cap I_{n+1}^k) = \mu(\{x \in I_{n+1}^k : f(x) \geq \beta\})
\]

\[
= \mu(\{x \in I_{n+1}^j(x) : f(\xi^{k-j}(x)) \geq \beta\})
\]

\[
= \mu(\{x \in I_{n+1}^j(x) : f(x) \geq \beta\})
\]

\[
= \mu(B \cap I_{n+1}^j)
\]

\[
> \frac{1}{2} \mu(I_{n+q}).
\]

Hence

\[
\mu(B \cap I_{n+1}^k) + \mu(B^c \cap I_{n+1}^k) > \mu(I_{n+1}^k)
\]

which is impossible since $B$ is measurable. Therefore $f$ is constant $\mu$-a.e.

To show that $\xi$ is not strong mixing we use the fact that for each $A \in \mathcal{B}$

\[
\mu(A) = \inf\{\sum_{i=1}^{\infty} \mu(V_n) : A \subseteq \bigcup_{i=1}^{\infty} V_n, V_n \text{ are disjoint intervals}\}
\]

\[
= \inf\{\sum_{i=1}^{j} \mu(I_n^i) : A \subseteq \bigcup_{i=1}^{j} I_n^i, I_n^i \text{ are (2,3)-adic intervals of order } n - 1\}
\]

Using this we know that for any $A \in \mathcal{B}$ and any $\varepsilon > 0$ we can find a collection of (2,3)-adic intervals of order $n$ for some $n \in \mathbb{N}$ such that

\[
\mu\left(\bigcup_{i=1}^{j} I_n^i - A\right) < \varepsilon.
\]

Then by Lemma 5.11 we know that

\[
\mu\left(\xi^{p(n)}\left(\bigcup_{i=1}^{j} I_n^i\right) \cap \bigcup_{i=1}^{j} I_n^i\right) = \frac{1}{2} \mu\left(\bigcup_{i=1}^{j} I_n^i\right).
\]

We can then consider a subset $I$ of $\bigcup_{i=1}^{j} I_n^i$ for which $\mu(I) = 1/2 \mu(\bigcup_{i=1}^{j} I_n^i)$ and for which $\xi^{p(n)}(I) \subseteq \bigcup_{i=1}^{j} I_n^i$. Further we know that

\[
\xi^{p(n)}(\bigcup_{i=1}^{j} I_n^i - I) \cap \bigcup_{i=1}^{j} I_n^i = \emptyset.
\]
We now know that
\[
\xi^{p(n)}(A \cap I) \subset \bigcup_{i=1}^{j} I_n^i \\
\mu(A \cap I) \geq \mu(I) - \varepsilon
\]
and hence, as \( \xi^{-1} \) is measure preserving
\[
\mu(\xi^{p(n)}(A \cap I)) \geq \mu(I) - \varepsilon
\]

Therefore
\[
\mu(\xi^{p(n)}(A) \cap A) \geq \mu\left(\xi^{p(n)}(A) \cap \bigcup_{i=1}^{j} I_n^i\right) - \varepsilon \\
\geq \mu(I) - 2\varepsilon \\
= \frac{1}{2} \mu\left(\bigcup_{i=1}^{j} I_n^i\right) - 2\varepsilon \\
\geq \frac{1}{2} \mu(A) - 2\varepsilon.
\]

As this is true for any \( \varepsilon > 0 \) and any \( A \in \mathcal{B} \) we know that for each \( n \in \mathbb{N} \)
\[
\mu(\xi^{p(n)}([0, 0.1])) \geq \frac{1}{2} \mu([0, 0.1]) > (\mu([0, 0.1]))^2
\]
which means that
\[
\lim_{n \to \infty} \mu(\xi^n([0, 0.1]) \cap [0, 0.1]) \neq (\mu([0, 0.1]))^2.
\]
Therefore \( \xi \) cannot be strong mixing. \( \diamond \)

The other obvious type of extension to investigate is that of product spaces. The property of generating Carathéodory's definition does not translate as well back and forth through product spaces and projections as it does through tower transformations. However, the results we do get in a limited investigation are interesting when compared with Theorem 3.7.
We conclude this chapter by demonstrating the results of our investigation into product spaces and discussing the mentioned comparison with Theorem 3.7.

We must start by defining what we mean by a product space generating Carathéodory’s definition of measurable sets. We will only consider products of spaces on which the definition of generating Carathéodory’s definition makes sense. That is, on spaces for which the concept of intervals makes sense.

**Definition 5.9**

Let $(X \times Y, \mathcal{B} \times \mathcal{D}, \mu \times \nu, \psi \times \phi)$ be the cartesian product of two dynamical systems of finite measure where $X$ and $Y$ are both subintervals of $\mathbb{R}$. Then if, whenever $A$ is a $\psi \times \phi$-invariant subset of subset of $X \times Y$ for which there is a $\theta = \theta(A) < 2$ such that for all rectangles $J_1 \times J_2$, with $J_1$ a subinterval of $X$ and $J_2$ a subinterval of $Y$,

$$(\mu_* \times \nu_*)(A \cap J_1 \times J_2) + (\mu_* \times \nu_*)(A^c \cap J_1 \times J_2) \leq \theta(\mu \times \nu)(J_1 \times J_2),$$

either $(\mu_* \times \nu_*)(A) = 0$ or $(\mu_* \times \nu_*)(A^c) = 0$, $\psi \times \phi$ is said to generate the Carathéodory definition of measurable sets.

The property of product spaces with respect to generating Carathéodory’s definition is proven below.

**Theorem 5.11**

If $\psi \times \phi$ generates Carathéodory’s definition of measurable sets then so too does both $\psi$ and $\phi$.

**Proof:**

Suppose that $\psi \times \phi$ generates Carathéodory’s definition. Let $B$ be a
\( \phi \)-invariant subset of \( Y \) and suppose there is a \( \theta < 2 \) such that for all intervals \( J \),

\[
\nu_*(B \cap J) + \nu_*(B^c \cap J) \leq \theta \nu(J).
\]

Note that \( X \times B \) is \( \psi \times \phi \)-invariant and that for all intervals \( J \), \( \mu(X \cap J) = \mu(J) \) and \( \mu(X^c \cap J) = 0 \). Then for any rectangle \( J_1 \times J_2 \)

\[
\mu_* \times \nu_*(X \times B \cap J_1 \times J_2) + \mu_* \times \nu_*((X \times B)^c \cap J_1 \times J_2) \\
= \mu_*(X \times B \cap J_1 \times J_2) + \mu_*(X \times B^c \cap J_1 \times J_2) \\
= \mu(X \cap J_1)\nu_*(B \cap J_2) + \mu(X \cap J_1)\nu_*(B^c \cap J_2) \\
= \mu(J_1)(\nu_*(B \cap J_2) + \nu_*(B^c \cap J_2)) \\
\leq \mu(J_1)(\theta \nu(J_2)) \\
= \theta \mu(J_1)\nu(J_2) \\
= \theta \mu \times \nu(J_1 \times J_2).
\]

Thus by the condition that leads to \( \psi \times \phi \) generating Carathéodory's definition, either \( \mu_* \times \nu_*(X \times B) = 0 \) or \( \mu_* \times \nu_*(X \times B^c) = \mu_*(X \times B^c) = 0 \). That is, either \( \nu_*(B) = \mu(X)\nu_*(B) = \mu_* \times \nu_*(X \times B) = 0 \) or (in a similar manner) \( \nu_*(B^c) = 0 \). Thus \( \phi \) generates Carathéodory's definition. The fact that \( \psi \) also generates Carathéodory's definition follows in a similar way.

An obvious corollary of Theorem 5.10, which is actually covered by Theorem 5.10 and thus requires no proof is stated below.

**Corollary 5.3**

If \( \psi \times \psi \) generates Carathéodory's definition then so does \( \psi \).

To prove that the Corollary 5.1 and thus Theorem 5.10 fails in the other direction we construct a counter example.
Theorem 5.12
\[ \psi \text{ generating Carathéodory's definition of measurable sets is insufficient to ensure that } \psi \times \psi \text{ generates Carathéodory's definition.} \]

Proof:
We prove the Theorem by way of a counter example. It is known that the original Kakutani I type transformation, \( \psi \), generates Carathéodory's definition. We construct our counter example based on this \( \psi \). We look at the orbit

\[ O_{\psi \times \psi}([0, 1/4] \times [0, 1/4]) \]

which is essentially

\[ \{[0, 1/4] \times [0, 1/4], [1/2, 3/4] \times [1/2, 3/4], [1/4, 1/2] \times [1/4, 1/2], [3/4, 1] \times [3/4, 1]\}. \]

Clearly \( K = \bigcup \{J : J \in O_{\psi \times \psi}([0, 1/4] \times [0, 1/4])\} \) is \( \psi \times \psi \)-invariant. Also as \( K \) and \( J_1 \times J_2 \) are measurable for all rectangles \( J_1 \times J_2 \)

\[ \mu \times \mu(K \cap J_1 \times J_2) + \mu \times \mu(K^c \cap J_1 \times J_2) = \mu \times \mu(J_1 \times J_2) \leq \theta \mu \times \mu(J_1 \times J_2) \]

where \( \theta = 1 < 2 \). Now note that

\[ \mu \times \mu(K) = \sum \{\mu(J) : J \in O_{\psi \times \psi}([0, 1/4] \times [0, 1/4])\} = \frac{1}{4^2} = 1/4 \]

and that as \( K \)-is measurable \( \mu \times \mu(X \times X - K) = 1 - 1/4 = 3/4 \). Hence the required conditions do not hold for \( \psi \times \psi \) to generate Carathéodory's definition.

The comparison to Theorem 3.7 is comparing what strength of mixing the above Theorems suggest that the property of generating Carathéodory's definition should be related to. From Theorem 3.7 we gather that a transformation \( T \) is weak mixing if and only if \( T \times T \) is. On the other hand if \( T \times T \)
is ergodic then by Theorem 3.7 $T$ is weak mixing and thus ergodic, however, if $T$ is ergodic we do not necessarily have that $T \times T$ is ergodic as otherwise $T$ being ergodic would imply that $T$ is weak mixing which is obviously false.

We have just seen that the one way implication is the behaviour displayed by the property of generating Carathéodory’s definition. Hence we now aim to show that only ergodicity is required to ensure that a transformation generates the Carathéodory definition of measurable sets. While it is not true that only ergodicity is required, we get close. We discuss the relationship between ergodicity and the property of generating Carathéodory’s definition of measurable sets in the next chapter.

5.3 Notes

The second example of a weak but not strong mixing transformation is given by Kakutani in [21]. The result of Theorem 5.2 was known to Kakutani who gave an outline of the proof of which we give a completion in [22]. Chacon’s weak but not strong mixing transformation was first presented by Chacon in [8]. Theorem 5.10 is due to Chacon [8]. Lemma 5.3 is an extension of Lemma 4.8 and Lemma 5.5 is a generalisation of Lemma 4.7. The remaining results, Lemmas 5.1, 5.2, 5.3, 5.6, 5.7, 5.8, 5.9 and 5.10, Theorems 5.1, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.11 and 5.12 and Corollaries 5.1, 5.2 and 5.3 are original. Kakutani knew of the tower representation of transformations similar to his transformation that we present in Chapter 4 as he discusses in [22]. An excellent discussion of ergodic theory highlighting the use of tower transformations is given by Friedman [13]. Liouville numbers were mentioned in this chapter. For a discussion on Liouville numbers and their place as a subset of the real numbers a good starting point is Niven [33].
6 An Ergodic Theorem for Measurable Sets: The Carathéodory definition generated

The results so far have been presented in a chronological order to show the progression of ideas as we approach the most general results. We have now looked at all of the work leading up to the final results. All of the tools that will be used in the general results that we prove in this chapter have been motivated, though there is still some generalisation needed. We now use the generalisations of these tools that have repetitively appeared to find what general types of transformations will generate the Carathéodory definition of measurable sets.

In the previous chapters we have found several specific examples of transformations that generate the Carathéodory definition of measurable sets. We are now looking to provide general results which will describe the specific properties that are necessary and or sufficient for a transformation to generate the Carathéodory definition. As mentioned in previous chapters, we would like these properties to be as general and simple as possible, so that the definition will then have been shown to arise from a large and natural class of transformations. As noted at the end of the previous chapter, the property of generating Carathéodory's definition behaves like the property of ergodicity (as opposed to the property of weak mixing) in some cases. It therefore seemed reasonable to try to use only the ergodic property. To look at more general spaces we also need, as mentioned, to generalise some of our tools to move away from the specific spaces for which the tools have been previously used.

Our line of thinking, in terms of which properties to aim for, was justified in that we have successfully proved that any ergodic outer measure preserving
transformation (with restrictions on the type of measure yet to be discussed) will generate the Carathéodory definition. This is the main result with respect to the generation of Carathéodory’s definition. In this chapter we will first prove a theorem that will allow general generation of Carathéodory’s definition. The main results are then presented in two theorems. The appropriate generalisations of the property of generating Carathéodory’s definition, splinters and intervals of density, will be given as they become necessary.

Of the two theorems, the first (Theorem 6.2), in using and following from the concepts developed so far, will show the importance and relevance of the ideas so far developed. Then a generalisation of Theorem 6.2, Theorem 6.3, will give our most general and powerful result. Theorem 6.3 actually covers the result of Theorem 6.2, however, the proof of Theorem 6.3 is more a manipulation of the tools used in Theorem 6.2 and so Theorem 6.2 remains to help demonstrate how the previous ideas lead to the final result.

In order to prove these general results we generalise the mathematical tools that we will use. So far, as we have noted quite thoroughly, every proof of the ability to generate the definition by way of analogy to Theorem 5.1 in [32] in the sense of Definition 2.5 has used two properties. Firstly the existence of intervals of density and secondly a rotation of some form mapping an interval or set (at least approximately) onto another one of the same measure. Our generalisation of an ‘interval of density’ is ‘sets of density’ as the concept of an interval will not necessarily be available in a general measure space.

We have actually already given a definition of sets of density in Chapter 5 to allow for some results to proved in more generality, but we repeat the definition here for convenience.
Definition 5.5

If \((X, \mathcal{B}, \mu)\) is a measure space, \(\mathcal{B} \subseteq X\) has positive outer measure and \(\varepsilon > 0\), then \(J \in \mathcal{B}\) is called a set of density to within \(\varepsilon\) for \(\mathcal{B}\) if

\[
\mu^*(B \cap J) > (1 - \varepsilon)\mu(J).
\]

We can note that in any non-atomic measure space we will in fact be able to find a set in \(\mathcal{B}\), say \(J\) for which

\[
\mu^*(B \cap J) = \mu(J).
\]

However, we will not be using this fact as in general, we will not allow the choice of an arbitrary set in \(\mathcal{B}\) to use as a set of density but rather the choice must come from a specified subset of \(\mathcal{B}\). The reason for this is the use of intervals up to this point. An important aspect of the generation of Carathéodory's definition has been the ability to use intervals as a type of set of comparison for the arbitrary sets that were considered. As we have generalised intervals of density and may no longer have intervals, we need to choose our sets of density to form a subset of \(\mathcal{B}\) that allows for the preservation of this tool. We shall therefore establish the properties of intervals that we need to retain and develop a collection of sets that will take their place. As this class of sets will be, as the intervals were, used, in a sense, as a collection of 'standard' sets for arbitrarily chosen sets to be compared to, this class of sets will be referred to as the collection of standard sets and an element of this collection will be called a standard set. This class of sets will be denoted by \(\mathcal{J}\). The following definition identifies the characteristics that identify the standard sets.
Definition 6.1

Let \((X, \mathcal{B}, \mu)\) be a measure space. A collection \(\mathcal{J} \subset \mathcal{B}\) will be called a collection of standard sets if for each pair of subsets \(A\) and \(B\) of \(X\) and each \(\varepsilon > 0\) there exist two sets \(J_1\) and \(J_2\) in \(\mathcal{J}\) of equal measure such that \(J_1\) is a set of density to within \(\varepsilon\) for \(A\) and \(J_2\) is a set of density to within \(\varepsilon\) for \(B\).

Note that every dynamical system \((X, \mathcal{B}, \mu, T)\) will have a collection of standard sets as \(\mathcal{B}\) will satisfy the conditions and hence, itself, be a collection of standard sets. Note also that the collection of intervals with Lebesgue measure on some subset of \(\mathbb{R}\) (the collection which has been the subject, or equivalent to the subject, of all of our discussions so far) satisfies these conditions and so is a collection of standard sets. We also mention that the smaller we can make \(\mathcal{J}\) the more general (and hence stronger) our results become. To have been satisfied with simply the \(\sigma\)-algebra would have weakened the results in this section and in fact would have meant that the previous results would not have been covered. Other examples of collections of standard sets include the set of all dyadic intervals in \(\mathbb{R}\) with the Borel \(\sigma\)-algebra and the collection of open balls on an \(n\)-sphere with Lebesgue measure.

With these tools in hand we can give a formal definition of what it means for a transformation in a general ergodic space to generate the Carathéodory definition of measurable sets.

Definition 6.2

Let \(T\) be a transformation on a measure space \((X, \mathcal{B}, \mu)\) and \(\mathcal{J}\) be a chosen collection of standard sets for \((X, \mathcal{B}, \mu)\). Then if, whenever \(A\) is a \(T\)-invariant subset of \(X\) for which there is a \(\theta = \theta(A) < 2\) such that for all \(J \in \mathcal{J}\)

\[
\mu_*(A \cap J) + \mu_*(A^c \cap J) \leq \theta \mu(J),
\]
either $A$ or $A^c$ is a set of outer measure zero, $T$ is said to generate Carathéodory’s definition of measurable sets or simply to generate Carathéodory’s definition.

Note that under this definition of generating Carathéodory’s definition, Theorem 5.7 gives us that any transformation will generate Carathéodory’s definition if and only if each usual tower extension of the transformation does.

In this chapter’s first theorem we prove that with the above definitions we need not actually go so far as proving that all of the requirements of Definition 6.2 are met but rather that Carathéodory’s definition will definitely be generated from an earlier stage in the argument.

**Theorem 6.1**

Suppose that $(X, \mathcal{B}, \mu, T)$ is a dynamical system on a measurable space and let $\mathcal{J}$ be a prechosen collection of standard sets. Suppose that for any $T$-invariant $B \subset X$ and for any $J \in \mathcal{J}$, $\mu_*(B \cap K) \geq \mu_*(B \cap J)$ for all $K \in \mathcal{J}$ with $\mu(K) = \mu(J)$. Then $T$ generates the Carathéodory definition of measurable sets.

**Note:** Clearly for any $K \in \mathcal{J}$ we can use $K$ as our original set to get $\mu_*(B \cap J) \geq \mu_*(B \cap K)$ and thus have $\mu_*(B \cap K) = \mu_*(B \cap J)$. However, while using ‘$=$’ or ‘$\geq$’ has little effect on the proof of this theorem, the phrasing allows simpler proofs of our main theorems later on.

**Proof of Theorem 6.1:**

For $T$ to generate the Carathéodory definition we need to show that if $B$ is
a $T$-invariant set for which there is a $\theta = \theta(B) > 0$ such that

$$\mu_*(B \cap J) + \mu_*(B^c \cap J) > (1 - \epsilon)\mu(J)$$

for each $J \in \mathcal{J}$ then either $\mu_*(B) = 0$ or $\mu_*(B^c) = 0$.

Suppose that neither $B$ nor $B^c$ have zero outer measure. Then for $\epsilon = (2 - \theta)/2$ we can find sets of density $J_1$ and $J_2$ of equal length for $B$ and $B^c$ respectively. Thus

$$\mu_*(B \cap J_1) > (1 - \epsilon)\mu(J_1) \text{ and } \mu_*(B^c \cap J_2) > (1 - \epsilon)\mu(J_2).$$

Then by the hypothesis

$$\mu_*(B \cap J_2) \geq \mu_*(B \cap J_1) > (1 - \epsilon)\mu(J_1) = (1 - \epsilon)\mu(J_2).$$

Thus

$$\mu_*(B \cap J_2) + \mu_*(B^c \cap J_2) > 2(1 - \epsilon)\mu(J_2)$$

$$= 2(1 - (2 - \theta)/2)\mu(J_2)$$

$$= (2 - (2 - \theta))\mu(J_2)$$

$$= \theta\mu(J_2).$$

This contradiction gives the result.

While this result is useful in this section, for a given transformation, if trying to prove the property of generation without Theorem 6.3, it may often be easier to prove that the transformation generates Carathéodory's definition directly, rather than via Theorem 6.1.

The next theorem, Theorem 6.2 is the first result presented that shows a general class or type of transformation that generates the Carathéodory definition. While the theorem directly following Theorem 6.2, Theorem 6.3,
immediately supercedes the generality of Theorem 6.2, as mentioned earlier, Theorem 6.2 remains as it draws together the ideas of the previous work and provides what would otherwise have been a missing link between the previous work and the inspiration for the ultimate result presented here, Theorem 6.3.

Theorem 6.2 shows that in an ergodic dynamical system that is almost bijective, the transformation must generate Carathéodory's definition. First we define what is meant by almost bijective.

**Definition 6.3**

Let $(X, \mathcal{B}, \mu, T)$ be a dynamical system. The transformation $T$ is called almost onto or onto $\mu$-a.e. if there is a set $Z$ with $\mu(Z) = 0$ such that $T(X) = X - Z$.

An almost bijective transformation is a transformation that is one to one and almost onto.

The idea behind the proof, given our previous work is relatively simple, though we must first generalise the concept of splinters that was introduced in Chapter 5 by Definition 5.4. Here we will not have any tower levels on which to define the splinters, instead we will talk about splinters of one set with respect to another. This proves to be a generalisation of the original definition when considering the induced tower transformation on the space. In our definition we use $|B|$ to denote the cardinality of the set $B$.

**Definition 6.4**

Let $A$ and $B$ be two elements of the $\sigma$-algebra of the dynamical system $(X, \mathcal{B}, \mu, T)$. Then an $n$-splinter of $A$ with respect to $B$ is either a
\[ \{ x \in A : \min \{ p : p \in \mathbb{N}; T^{-p}x \in B \} = m; |N \cap \{ s; T^{-s}x \in B \}| = q \} \]

for some \( 1 \leq m, q \leq n \), or a set

\[ \{ x \in A : T^{-i}x \notin B \text{ for all } 1 \leq i \leq n \}. \]

A splinter of \( A \) with respect to \( B \) is either a set

\[ \{ x \in A : \min \{ p : T^{-p}x \in B - \cup_{i=1}^{p-1} T^{-i}A \} = m \} \] (21)

for some \( m \in \mathbb{N} \), or a set

\[ \{ x \in A : T^{-i}x \notin B \text{ for all } i \in \mathbb{N} \}. \]

It can be easily seen that in terms of Definition 6.4, Definition 5.4 says that an \( n \)-splinter of \( V \) is an \( n \)-splinter of \( V \) with respect to \( X_0 \). The reason for the subtraction of \( \cup_{i=1}^{p-1} T^{-i}A \) in (21) shall be seen shortly.

We can now consider the idea behind the proof of Theorem 6.2. We discuss the idea behind the proof extensively to make abundantly clear the relation of the previous results to this chapter. The theorem is proved if we can show that the conditions for Theorem 6.1 are satisfied and hence that is our aim. We therefore take some arbitrary subset \( B \) of some ergodic dynamical system with a collection of standard sets. The idea of Theorem 6.2 to use splinters when rotating our standard sets comes immediately when one considers that the elements of \( J \) have no requirement to be of the same 'shape' (in the sense that all intervals are a connected line in \( \mathbb{R}^1 \) and hence the same shape) and so we must find another method of transporting one standard set to another. For this reason we can abandon the idea of rotating the element \( J_1 \) of \( J \) so that it retains its shape and 'arrives' at \( J_2 \) all on the same iteration.
Instead we break $J_1$ up into its splinters with respect to $J_2$ and let the splinters 'arrive' in $J_2$ in their own time. Provided that $\mu$-almost all of $J_1$ does indeed arrive in $J_2$ and that each of the splinters has its own unique 'arrival point' (so as to distribute $B$ with the same density in $J_2$ as it had in $J_1$) the result should (and does) follow easily. To guarantee that each splinter that does arrive has its own arrival point, we need that the transformation be onto almost everywhere and one-to-one. Without this, $T^{-1}$ will not be measure preserving, also inverse images of points will be sets, thus making the tracking of splinters movements impossible. Additionally we note that once we have the required invertibility properties, the definition of a splinter, particularly the subtraction of $\bigcup_{i=1}^{p-1} T^{-i}A$ in (21) at the crucial point of the definition commented on earlier ensures the uniqueness of arrival position for the splinters. While the splinters may not initially arrive in a unique part, each splinter will arrive in a unique part of $J_2$ without all of the sections of $J_2$ that have already been used as arrival sets for other splinters of $J_1$. That $\mu$-almost all of $J_1$ does arrive in an allotted part of $J_2$ is then given by the ergodic property of the transformation. This last point is quite clearly a vital one, and in proving it we make use of Theorem 3.1.

We will also be needing the following technical lemma that gives us countable additivity on certain types of sequences of not necessarily measurable sets.

**Lemma 6.1**

Suppose that $(X, B, \mu)$ is a measure space, $B \subset X$ and $\{A_n\}$ is a sequence of disjoint sets in $B$ such that

$$\lim_{n \to \infty} \mu \left( \bigcup_{i=n}^{\infty} (A_i) \right) = 0.$$
Then
\[ \mu_\ast \left( \bigcup_{i=1}^\infty (A_i \cap B) \right) = \sum_{i=1}^\infty \mu_\ast (A_i \cap B). \]

**Proof:**

By countable subadditivity we immediately have
\[ \mu_\ast \left( \bigcup_{i=1}^\infty (A_i \cap B) \right) \leq \sum_{i=1}^\infty \mu_\ast (A_i \cap B). \]

For the converse, first note that as the \( A_i \)'s are measurable we have
\[ \mu_\ast \left( \bigcup_{i=1}^\infty (A_i \cap B) \right) = \mu_\ast (A_1 \cap B) + \mu_\ast \left( \bigcup_{i=2}^\infty (A_i \cap B) \right) = \ldots = \sum_{i=1}^m \mu_\ast (A_i \cap B) + \mu_\ast \left( \bigcup_{i=m+1}^\infty (A_i \cap B) \right), \]
for any finite \( m \).

Now let \( \varepsilon > 0 \). Then as \( \lim_{n \to \infty} \mu(\bigcup_{i=n}^\infty (A_i)) = 0 \) there is an \( n_0 \in \mathbb{N} \) such that
\[ \mu(\bigcup_{i=n_0}^\infty A_i) < \varepsilon \]
and we therefore have
\[ \sum_{i=1}^\infty \mu_\ast (A_i \cap B) = \sum_{i=1}^{n_0-1} \mu_\ast (A_i \cap B) + \sum_{i=n_0}^{\infty} \mu_\ast (A_i \cap B) \]
\[ \leq \sum_{i=1}^{n_0-1} \mu_\ast (A_i \cap B) + \sum_{i=n_0}^{\infty} \mu_\ast (A_i \cap B) \]
\[ < \varepsilon + \mu_\ast \left( \bigcup_{i=1}^{\infty} (A_i \cap B) \right) - \mu_\ast \left( \bigcup_{i=n_0}^{\infty} (A_i \cap B) \right) \]
\[ \leq \varepsilon + \mu_\ast \left( \bigcup_{i=1}^{\infty} (A_i \cap B) \right). \]

We therefore have that
\[ \mu_\ast \left( \bigcup_{i=1}^\infty (A_i \cap B) \right) > \sum_{i=1}^\infty \mu_\ast (A_i \cap B) - \varepsilon \]
for each $\varepsilon > 0$ and hence
\[
\mu_* \left( \bigcup_{i=1}^{\infty} (A_i \cap B) \right) \geq \sum_{i=1}^{\infty} \mu_*(A_i \cap B)
\] \hfill \diamond

**Theorem 6.2**

*Suppose that $T$ is a measure-preserving ergodic transformation, onto $\mu$-a.e. and one-to-one, on the measure space $(X, \mathcal{B}, \mu)$. Then for any collection of standard sets $J$ $T$ generates the Carathéodory definition.*

**Proof:**

Let $J$ be a collection of standard sets for $(X, \mathcal{B}, \mu)$.

By Theorem 6.1 it is sufficient to prove that if $B$ is $T$-invariant and $J \in J$ then $\mu_*(B \cap K) = \mu_*(B \cap J)$ for all $K \in J$ such that $\mu(K) = \mu(J)$.

Now, let $J_1$ and $J_2$ be fixed but arbitrary elements of $J$ of equal measure. Then, let

\[ A_1 = T^{-1}(J_2) \cap J_1 \in \mathcal{B} \]

and then, inductively, let

\[ A_k = (J_1 - \bigcup_{i=1}^{k-1} A_i) \cap (J_2 - \bigcup_{i=1}^{k-1} T^{-i} A_i). \]

These $A_k$ are an indexation of the set of splinters of $J_1$ with respect to $J_2$. Clearly $\{A_i\}_{i=1}^{\infty}$ is a disjoint sequence of measurable sets, as is $\{T^{-i} A_i\}_{i=1}^{\infty}$.

Thus as $\mu(\bigcup_{i=1}^{\infty} A_i) \leq \mu(J_1) < \infty$ and $\mu(\bigcup_{i=1}^{\infty} T^{-i} A_i) \leq \mu(J_1) < \infty$

\[
\mu(\bigcup_{i=1}^{\infty} T^{-i} A_i) = \sum_{i=1}^{\infty} \mu(T^{-i} A_i)
\]
\[= \sum_{i=1}^{\infty} \mu(A_i)
\]
\[= \mu(\bigcup_{i=1}^{\infty} A_i).\]
Note that $J_1 - \cup_{i=1}^{\infty} A_i$, $J_2 - \cup_{i=1}^{\infty} T^{-i} A_i \in B$.

It is clear that $\mu(\cup_{i=1}^{\infty} A_i) \leq \mu(J_1)$. Now suppose that $\mu(\cup_{i=1}^{\infty} A_i) \neq \mu(J_1)$, then $\mu(\cup_{i=1}^{\infty} A_i) < \mu(J_1)$, $\mu(J_1 - \cup_{i=1}^{\infty} A_i) > 0$ and $\mu(J_2 - \cup_{i=1}^{\infty} T^{-i} A_i) > 0$. As $T$ is ergodic, by Theorem 3.1 there is an $n \in \mathbb{N}$ such that

$$\mu(T^{-n} (J_1 - \cup_{i=1}^{\infty} A_i) \cap (J_2 - \cup_{i=1}^{\infty} T^{-i} A_i)) > 0.$$ 

But then, as

$$(J_1 - \cup_{i=1}^{\infty} A_i) \subseteq (J_1 - \cup_{i=1}^{n-1} A_i)$$

and

$$(J_2 - \cup_{i=1}^{\infty} T^{-i} A_i) \subseteq (J_2 - \cup_{i=1}^{n-1} T^{-i} A_i)$$

we would then have

$$\emptyset \neq (T^{-n} (J_1 - \cup_{i=1}^{\infty} A_i) \cap (J_2 - \cup_{i=1}^{\infty} T^{-i} A_i))$$

$$\subseteq (T^{-n} (J_1 - \cup_{i=1}^{n-1} A_i) \cap (J_2 - \cup_{i=1}^{n-1} T^{-i} A_i))$$

$$= T^{-n}(\{x \in (J_1 - \cup_{i=1}^{n-1} A_i) : T^{-n}x \in (J_2 - \cup_{i=1}^{n-1} T^{-i} A_i)\})$$

$$= T^{-n}A_n.$$ 

Hence there is $x \in X$ such that $T^n x \notin A_n$ and $T^n x \in A_n$, a contradiction which gives us that

$$\mu(J_1 - \cup_{i=1}^{\infty} A_i) = 0 = \mu(J_2 - \cup_{i=1}^{\infty} T^{-i} A_i).$$

That is, we have

$$\mu(J_1 \triangle \cup_{i=1}^{\infty} A_i) = 0 \text{ and } \mu(J_2 \triangle \cup_{i=1}^{\infty} T^{-i} A_i) = 0.$$ 

Now, using Lemma 6.1, and the fact that $B$ is $T$-invariant we get

$$\mu_*(B \cap J_1) = \mu_*(B \cap \cup_{i=1}^{\infty} A_i) + \mu_*(B \cap (J_1 - \cup_{i=1}^{\infty} A_i))$$

$$= \mu_*(B \cap \cup_{i=1}^{\infty} A_i)$$

211
\begin{align*}
&= \mu_*(\bigcup_{i=1}^{\infty} (B \cap A_i)) \\
&= \sum_{i=1}^{\infty} \mu_*(B \cap A_i) \\
&= \sum_{i=1}^{\infty} \mu_*(T^{-i}(B \cap A_i)) \\
&= \mu_*(\bigcup_{i=1}^{\infty} T^{-i}(B \cap A_i)) \\
&= \mu_*(\bigcup_{i=1}^{\infty} T^{-i}B \cap T^{-i}A_i) \\
&= \mu_*(\bigcup_{i=1}^{\infty} B \cap T^{-i}A_i) \\
&= \mu_*(B \cap \bigcup_{i=1}^{\infty} T^{-i}A_i) + 0 \\
&= \mu_*(B \cap \bigcup_{i=1}^{\infty} T^{-i}A_i) + \mu_*(B \cap (J_2 - \bigcup_{i=1}^{\infty} T^{-i}A_i)) \\
&= \mu_*(B \cap J_2).
\end{align*}

Hence the conditions for Theorem 6.1 are satisfied and the proof is complete.

\diamondsuit

Before moving on to Theorem 6.3 we note that one useful consequence of
Theorem 6.2 is the existence of cross product spaces that have the ability to
generate Carathéodory's definition, which was an open question at the end of
Chapter 5. In the proof, we use the result that $T \times T$ is measure preserving
with respect to $(X \times X, B \times B, \mu \times \mu)$ if $T$ is measure preserving with respect
to $(X, B, \mu)$. A proof of this can be found in Walters [45].

**Corollary 6.1**

*The cross product of any bijective weak mixing transformation satisfying the
conditions of Theorem 6.2 generates the Carathéodory definition of measurable sets.*

**Proof:**

We know such weakly mixing transformations exist, and Kakutani's trans-
formation discussed in Chapter 4 is one example. Thus this Corollary does
indeed prove that product spaces that generate the Carathéodory definition of measurable sets do indeed exist.

Now suppose that \( T \) is a weakly mixing bijective measure-preserving transformation. We know that \( T \) being weakly mixing implies that \( T \times T \) is ergodic. Also, as mentioned above, \( T \times T \) is \( \mu \times \mu \)-measure preserving as \( T \) is \( \mu \)-measure preserving. Then if \( T \times T(x, y) = T \times T(u, v) \) we have \( (T(x), T(y)) = (T(u), T(v)) \) so that \( T(x) = T(u) \) and \( T(y) = T(v) \). \( T \) is bijective, and hence \( x = u, y = v \) and thus \( (x, y) = (u, v) \) so that \( T \) is one to one. Finally, if \( (x, y) \in X \times X \) then \( x, y \in X \) and as \( T \) is onto there are \( u, v \in X \) such that \( T(u) = x \) and \( T(v) = y \). That is, \( T \times T(u, v) = (T(u), T(v)) = (x, y) \) so that \( T \times T \) is onto and hence we now have that \( T \times T \) is bijective. As \( T \times T \) satisfies the conditions of Theorem 6.2 we know that \( T \times T \) generates Carathéodory's definition. \( \diamond \)

6.1 \( T \) generates if \( T \) is outer measure preserving and ergodic

As has been mentioned, we will further generalise the result presented in Theorem 6.2. We prove that if \( T \) is an outer measure preserving ergodic transformation on a measure space then \( T \) must generate the Carathéodory definition. The idea used is similar to that used in Theorem 6.2 though we need to be more careful in choosing and defining the splinters we use. In fact, the way we move \( J_1 \) onto \( J_2 \) does not technically use splinters at all. As has been discussed, Theorem 6.3 is more an evolution of Theorem 6.2 directly rather than from the previous chapters. Instead of splitting the \( J_1 \) set into countably many splinters simultaneously, and finding them unique 'arrival' sets in \( J_2 \), rotating them individually, and then reassembling, a method compatible with avoiding the difficulties presented by noninvertible transfor-
mations must be found. This method involves ‘rotating’ the entire set $J_1$ and simply removing the bits of the iterated inverse images that intersect new parts of $J_2$. Although some new technical difficulties then present themselves in the wake of this change, including the necessity for a new use of ergodicity to show that $J_2$ is $\mu$ almost everywhere covered, the logical structure of the Theorem is the same. The necessity for the outer measure preserving property comes when we need to preserve the outer measure of a not necessarily measurable set as we ‘rotate’ it through $X$. This problem has so far not been considered in great detail as we have so far only considered (at worst) almost bijective transformations. As any almost bijective transformation (proven in Proposition 6.1) is outer measure preserving, the issue is not necessary to specifically dwell on as we mentioned in Chapter 2. Clearly the requirement for transformations to be outer measure preserving is an unusual one and hence we must consider in some way what sort of transformation is outer measure preserving. This issue is given some consideration in the next section.

Theorem 6.3 only proves that outer measure preserving ergodicity is sufficient for the generation of the Carathéodory definition of measurable sets. To prove that ergodicity is necessary for $T$ to generate the Carathéodory definition is much simpler and we prove the ‘equivalence’ in Theorem 6.4.

**Theorem 6.3**

*Suppose that $T$ is an outer measure preserving ergodic transformation on the space $(X, \mathcal{B}, \mu)$. Then for any collection of standard sets $\mathcal{J}$ $T$ generates the Carathéodory definition.*

**Proof:**

Let $\mathcal{J}$ be a collection of standard sets for $(X, \mathcal{B}, \mu)$. 

214
As in Theorem 6.2 it is sufficient to prove that the conditions of Theorem 6.1 hold.

Now, let $J_1$ and $J_2$ be fixed but arbitrary elements of equal measure in $\mathcal{J}$, and we construct our sets as follows. Let

$$A_1 = T^{-1}J_1 \cap J_2 \in \mathcal{B}$$

and

$$B_1 = T^{-1}J_1 - A_1 = T^{-1}J_1 - J_2 \in \mathcal{B}.$$ 

Then, we continue the definitions of these sets inductively as

$$A_n = T^{-1}B_{n-1} \cap (J_2 - \bigcup_{i=1}^{n-1} A_i) \in \mathcal{B},$$

and

$$B_n = T^{-1}B_{n-1} - A_n = T^{-1}B_{n-1} - \left( J_2 - \bigcup_{i=1}^{n-1} A_i \right) \in \mathcal{B}.$$  \hspace{1cm} (22)

Note that the sequence of sets $\{A_i\}_{i=1}^\infty$ is necessarily measurable and disjoint and hence for each $n \in \mathbb{N}$

$$A_{n+1} \cap \bigcup_{i=1}^{n} A_i = \emptyset$$ \hspace{1cm} (23)

We also claim that

$$\mu(B_n) = \mu(J_2 - \bigcup_{i=1}^{n} A_i) \text{ for each } n \in \mathbb{N}. \hspace{1cm} (24)$$

The claim is obviously true for $n = 1$ as using the measurability of $J_1$, $J_2$ and the fact that $T^{-1}(\mathcal{B}) \subset \mathcal{B}$

$$\mu(J_2 - A_1) = \mu(J_2) - \mu(A_1)$$

$$= \mu(T^{-1}J_1) - \mu(A_1)$$

$$= \mu(T^{-1}J_1) - \mu(T^{-1}J_1 \cap J_2)$$

$$= \mu(T^{-1}J_1 \cap J_2^c)$$

$$= \mu(B_1).$$

215
Now assuming \( \mu(B_n) = \mu(J_2 - \cup_{i=1}^{n} A_i) \) for some \( n \in \mathbb{N} \), we have, using the measurability of \( J_2, A_{n+1} \), that \( T \) is measure preserving and (23)

\[
\begin{align*}
\mu(J_2 - \cup_{i=1}^{n+1} A_i) & = \mu((J_2 - \cup_{i=1}^{n} A_i) - A_{n+1}) \\
& = \mu((J_2 - \cup_{i=1}^{n} A_i)) - \mu(A_{n+1}) \\
& = \mu(B_n) - \mu(A_{n+1}) \\
& = \mu(T^{-1}B_n) - \mu(A_{n+1}) \\
& = \mu(T^{-1}B_n \cap A_{n+1}^c) \quad \text{(by \( A_{n+1} \)'s measurability)} \\
& = \mu(T^{-1}B_n - A_{n+1}) \quad \text{(putting into the appropriate form)} \\
& = \mu(B_{n+1}) \text{(by (22))},
\end{align*}
\]

so that by mathematical induction the claim, (24), is true for all \( n \in \mathbb{N} \) as stated. Additionally since \( \mu(J_2) < \infty \) and \( \{J_2 - \cup_{i=1}^{n} A_i\}_n^\infty \) is a decreasing sequence,

\[
\lim_{n \to \infty} \mu(J_2 - \cup_{i=1}^{n} A_i)
\]
exists and hence so does \( \lim_{n \to \infty} \mu(B_n) \).

We now wish to prove that \( \mu(J_2 - \cup_{i=1}^{\infty} A_i) = 0 \). To do this we firstly note that for each \( m, n \in \mathbb{N} \)

\[
\mu(T^{-m}B_n \cap (J_2 - \cup_{i=1}^{\infty} A_i))
\]

\[
= \mu((B_{n+m} \cup \cup_{i=n+1}^{n+m} T^{2-m+n-i} A_i) \cap (J_2 - \cup_{i=1}^{\infty} A_i))
\]

\[
= \mu((B_{n+m} \cap (J_2 - \cup_{i=1}^{\infty} A_i) \cup \cup_{i=n+1}^{n+m} T^{2-m+n-i} A_i \cap (J_2 - \cup_{i=1}^{\infty} A_i))
\]

\[
\leq \mu((B_{n+m} \cap (J_2 - \cup_{i=1}^{\infty} A_i) + \mu(\cup_{i=n+1}^{n+m} T^{2-m+n-i} A_i \cap (J_2 - \cup_{i=1}^{\infty} A_i))
\]

\[
= 0 + \mu(\cup_{i=n+1}^{n+m} T^{2-m+n-i} A_i \cap (J_2 - \cup_{i=1}^{\infty} A_i))
\]

\[
\leq \mu(\cup_{i=n+1}^{\infty} T^{2-m+n-i} A_i \cap (J_2 - \cup_{i=1}^{\infty} A_i))
\]

\[
\leq \sum_{i=n+1}^{\infty} \mu(T^{2-m+n-i} A_i \cap (J_2 - \cup_{i=1}^{\infty} A_i))
\]
\[ \sum_{i=n+1}^{\infty} \mu(T^{2-m+n-i} A_i) \]
\[ = \sum_{i=n+1}^{\infty} \mu(A_i) \]
\[ = \mu(\bigcup_{i=n+1}^{\infty} A_i); \tag{25} \]

where the fourth relation follows from the fact that for any \( n \in \mathbb{N} \)

\[ B_n \cap J_2 \subset \bigcup_{i=1}^{n-1} A_i. \]

Now suppose that \( \mu(J_2 - \bigcup_{i=1}^{\infty} A_i) = c \neq 0 \) so that \( \lim_{n \to \infty} \mu(B_n) = c. \) Thus we can find an \( n_0 \in \mathbb{N} \) such that

\[ \mu(B_{n_0}) < c + \frac{c^2}{2}, \]

which, by (24), also means that

\[ \mu(J_2 - \bigcup_{i=1}^{n_0} A_i) < c + \frac{c^2}{2} \]

and hence

\[ \mu(\bigcup_{i=n_0+1}^{\infty} A_i) = \mu(J_2 - \bigcup_{i=1}^{n_0} A_i) - \mu(J_2 - \bigcup_{i=1}^{\infty} A_i) < c + \frac{c^2}{2} - c = \frac{c^2}{2}. \]

Then by (25) we have that for each \( m \in \mathbb{N} \)

\[ \mu(T^{-m} B_{n_0} \cap (J_2 - \bigcup_{i=1}^{\infty} A_i)) \leq \mu(\bigcup_{i=n_0+1}^{\infty} A_i) < \frac{c^2}{2}. \]

Thus, for each \( m \in \mathbb{N} , \)

\[ \frac{1}{m} \sum_{j=1}^{m} \mu(T^{-j} B_{n_0} \cap (J - \bigcup_{i=1}^{\infty} A_i)) < \frac{1}{m} \sum_{i=1}^{m} \frac{c^2}{2} \]
\[ = \frac{c^2}{2}. \]

Hence,

\[ \limsup_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} \mu(T^{-j} B_{n_0} \cap (J - \bigcup_{i=1}^{\infty} A_i)) \leq \frac{c^2}{2} \]
\[ \leq \frac{\mu(B_{n_0}) \mu(J_2 - \bigcup_{i=1}^{\infty} A_i)}{2} \]
\[ < \mu(B_{n_0}) \mu(J_2 - \bigcup_{i=1}^{\infty} A_i). \]
Thus even when the limit exists

$$\lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \mu(T^{-m}B_{n_0} \cap (J - \cup_{i=1}^{\infty} A_i)) \neq \mu(B_{n_0})\mu(J_2 - \cup_{i=1}^{\infty} A_i).$$

So that, whether or not a limit exists, Theorem 3.3 would imply that $T$ is not ergodic and therefore

$$\mu(J_2 - \cup_{i=1}^{\infty} A_i) = 0. \quad (26)$$

That is, $\mu(J_2 \Delta \cup_{i=1}^{\infty} A_i) = 0$. (Note that the limit will exist, but it is not necessary to show that it does here.)

So far in this proof we have shown that in a similar sense to Theorem 6.2, $J_2$ can be made up of pieces that come from $J_1$. However, we do not know what these pieces looked like when they were in $J_1$. Hence we have yet to prove that any $T$-invariant set $B$ intersected with $J_1$ will be represented appropriately by these pieces. We complete the proof by showing that they are.

Let a $T$-invariant set $B$ be given. Then, again using the measurability of the $B_n$s, the $A_n$s and the outer measure preserving property of $T$, we note that

$$\mu_*(J_1 \cap B) = \mu_*(T^{-1}(J_1 \cap B))$$

$$\leq \mu_*(J_2 \cap T^{-1}(J_1 \cap B)) + \mu_*(J_2 \cap T^{-1}(J_1 \cap B))$$

$$= \mu_*(J_2 \cap T^{-1}(J_1 \cap B)) + \mu_*(B_1 \cap T^{-1}(J_1 \cap B))$$

$$= \mu_*(J_2 \cap T^{-1}(J_1) \cap T^{-1}B) + \mu_*(B_1 \cap T^{-1}(J_1) \cap T^{-1}B)$$

$$= \mu_*(A_1 \cap T^{-1}B) + \mu_*(B_1 \cap T^{-1}B)$$

$$= \mu_*(A_1 \cap B) + \mu_*(B_1 \cap B)$$

and in a similar manner we can continue and note that

$$\mu_*(J_1 \cap B) \leq \mu_*(B_n \cap B) + \sum_{i=1}^{n} \mu_*(A_i \cap B) \quad (27)$$

218
for each $n \in \mathbb{N}$. We also have, using the measurability of $\bigcup_{i=1}^{\infty} A_i$ and (26), that

\[
\mu_*(J_2 \cap B) = \mu_*(\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cup (J_2 - \bigcup_{i=1}^{\infty} A_i)\right) \cap B) \\
= \mu_*(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B) \cup (J_2 - \bigcup_{i=1}^{\infty} A_i) \cap B) \\
= \mu_*(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B) + \mu_*(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B) \\
= \mu_*(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B).
\]  

(28)

Now we also have by Lemma 6.2

\[
\sum_{i=1}^{\infty} \mu_*(A_i \cap B) = \mu_*(\bigcup_{i=1}^{\infty} A_i \cap B),
\]

Hence, using (28)

\[
\mu_*(J_2 \cap B) = \sum_{i=1}^{\infty} \mu_*(A_i \cap B).
\]

Finally, since

\[
\lim_{n \to \infty} \mu(B_n) = \mu(J_2 - \bigcup_{i=1}^{\infty} A_i) = 0,
\]

for any $\varepsilon > 0$ there is an $n_0 \in \mathbb{N}$ such that $\mu(B_{n_0}) < \varepsilon$. Hence for any $\varepsilon > 0$ we have, by (27), that

\[
\mu_*(J_1 \cap B) \leq \mu_*(B_{n_0} \cap B) + \sum_{i=1}^{n_0} \mu_*(A_i \cap B) \\
< \varepsilon + \sum_{i=1}^{n_0} \mu_*(A_i \cap B) \\
\leq \varepsilon + \sum_{i=1}^{\infty} \mu_*(A_i \cap B) \\
= \varepsilon + \mu_*(J_2 \cap B).
\]

Since this is true for any $\varepsilon > 0$ we have that

\[
\mu_*(J_1 \cap B) \leq \mu_*(J_2 \cap B)
\]
which satisfies the requirements of Theorem 6.1.

\[ \text{Theorem 6.4} \]

Let \( T \) be an outer-measure-preserving transformation \( T \) on a measure space \((X, \mathcal{B}, \mu)\). Then for any collection of standard sets \( J \) \( T \) generates the Carathéodory definition if and only if it is ergodic.

**Proof:**

Let \( J \) be a collection of standard sets for \((X, \mathcal{B}, \mu)\).

Theorem 6.3 gives us that if \( T \) is ergodic then it will generate the Carathéodory definition. Now suppose that \( T \) is not ergodic, then there is a \( T \)-invariant set \( B \in \mathcal{B} \) such that \( 0 < \mu(B) < 1 \). As \( B \) is measurable

\[
\mu(B \cap J) + \mu(B^c \cap J) = \mu(J) \leq \theta \mu(J)
\]

with \( \theta = 1 \) for all \( J \in J \).

However \( \mu(B) \neq 0 \) and \( \mu(B^c) = \mu(X) - \mu(B) = 1 - \mu(B) \neq 0 \). Thus, by Definition 6.3, \( T \) does not generate the Carathéodory definition of measurable sets.

\[ \text{\diamond} \]

Before considering what outer measure preserving transformations are we give generalisations of Theorem 6.2 and Theorem 6.4 by demonstrating two results that describe some families of transformations that generate Carathéodory’s definition of measurable sets.

We first extend the definitions necessary to describe what a family of transformations that generate Carathéodory’s definition is.

We extend the definitions of measure preserving, orbits, invariance, ergodic-
ity and generating Carathéodory's definition.

Definition 6.5
Let $\Psi$ be a family of transformations on a measure space $(X, \mathcal{B}, \mu)$. $\Psi$ is said to be measure preserving if for each $B \in \mathcal{B}$, $\psi^{-1}(B) \in \mathcal{B}$ and $\mu(B) = \mu(\psi^{-1}(B))$ for each $\psi \in \Psi$.

Definition 6.6
Let $X$ be a set and $\Psi$ be a family of transformations. For any subset $D$ of $X$ the $\Psi$-orbit of $D$ is

$$O_\Psi(D) = \{\psi_1^{s_1} \circ \psi_2^{s_2} \circ \cdots \circ \psi_r^{s_r}(D) : r \in \mathbb{N}, s_1, \ldots, s_r \in \mathbb{Z}, \psi_1, \ldots, \psi_r \in \Psi\},$$

or, if $\Psi$ is understood, $O_\Psi(D)$ is simply the orbit of $D$. A similar definition is made for the orbit of a point $x \in X$. We define the $\Psi$-orbit of $x$ or simply the orbit of $x$, if $\Psi$ is understood, to be

$$O_\Psi(x) = \{y = \psi_1^{s_1} \circ \psi_2^{s_2} \circ \cdots \circ \psi_r^{s_r}(x) : r \in \mathbb{N}, s_1, \ldots, s_r \in \mathbb{Z}, \psi_1, \ldots, \psi_r \in \Psi\},$$

For the following definitions and the next two theorems we will also need the following definition relating to orbits.

Definition 6.7
Let $\Psi$ be a family of transformations defined on a set $X$ and $D \subset X$. The negative half $\Psi$-orbit of $D$ is

$$O_{\Psi^{-}}(D) = \{\psi_1^{-s_1} \circ \psi_2^{-s_2} \circ \cdots \circ \psi_r^{-s_r}(D) : r \in \mathbb{N}, s_1, \ldots, s_r \in \mathbb{N} \cup \{0\}, \psi_1, \ldots, \psi_r \in \Psi\},$$

or if $\Psi$ is understood, $O_{\Psi^{-}}(D)$ is called simply the negative half orbit of $D$.

This last definition is given to provide us with some information about the
sets in an orbit (the negative half orbit) of a set. That is, if we know that $\Psi$ is measure preserving on a measure space $(X, \mathcal{B}, \mu)$ and that $D \in \mathcal{B}$ then we know that for all $B \in O_\Psi(D), B \in \mathcal{B}$ and $\mu(B) = \mu(D)$.

**Definition 6.8**

Let $\Psi$ be a family of transformations on a measure space $(X, \mathcal{B}, \mu)$ and let $A \subset X$. A is said to be $\Psi$-invariant if for each $\psi \in \Psi$, $\psi^{-1}(A) = A$.

**Definition 6.9** A family of measure preserving transformations, $\Psi$ on a measure space $(X, \mathcal{B}, \mu)$ is said to be **ergodic** if for each $\Psi$ invariant $A \in \mathcal{B}$, $\mu(A) = 0$ or $\mu(A) = \mu(X)$.

**Definition 6.10**

Let $\Psi$ be a family of transformations on a measure space $(X, \mathcal{B}, \mu)$ and $\mathcal{J}$ be a chosen collection of standard sets for $(X, \mathcal{B}, \mu)$. Then if, whenever $A$ is a $\Psi$-invariant subset of $X$ for which there is a $\theta = \theta(A) < 2$ such that for all $J \in \mathcal{J}$

$$\mu(A \cap J) + \mu(A^c \cap J) \leq \theta \mu(J),$$

either $A$ or $A^c$ is a set of outer measure zero, $\Psi$ is said to generate Carathéodory’s definition of measurable sets or simply to generate Carathéodory’s definition.

The generalisations are interesting in that they show that for a family of transformations to generate the Carathéodory definition of measurable sets, the family of transformations must behave as a whole in a similar way to the individual transformations. These generalisation became further interesting later on when an example of a family of transformations that generates the Carathéodory’s definition was found for which no element of the family of
transformations generated Carathéodory's definition as an individual transformation.

This example of a family of transformations is important in Chapter 7 and is defined and investigated in Chapter 7. The proof of the fact that the family (but none of the elements) generates Carathéodory's definition follows from Theorem 6.5 but is given in Theorem 7.4 following further necessary preliminary Lemmas.

We now consider the generalisations of Theorems 6.2 and 6.4 respectively.

First we note that in the proof of Theorem 6.1 there is no use of any property of individual transformations that a family does not have and we can therefore state the following corollary.

**Corollary 6.2**

Suppose that \((X, \mathcal{B}, \mu)\) is a measure space and that \(\Psi\) is a family of transformations defined on \(X\). Let \(\mathcal{J}\) be a prechosen collection of standard sets. Suppose that for any \(\Psi\)-invariant \(B \subset X\) and for any \(J \in \mathcal{J}\), \(\mu_* (B \cap K) \geq \mu_* (B \cap J)\) for all \(K \in \mathcal{J}\) with \(\mu(K) = \mu(J)\). Then \(\Psi\) generates the Carathéodory definition of measurable sets.

We will make use of Corollary 6.2 in the generalisations. To generalise Theorem 6.2 we also need the following Lemma.
Lemma 6.2

Let $\Phi$ be a countable ergodic family of transformations on a measure space $(X, \mathcal{B}, \mu)$, then if $A \in \mathcal{B}$ has positive measure and $B \in \mathcal{B}$ is a $\Phi$-invariant set of positive measure there exists $C \in O^-_\Phi(A)$ such that $\mu(C \cap B) > 0$.

Proof:

As $\mathbb{N} \times \mathbb{N}$ is countable, \{${\phi^{-s} : \phi \in \Phi, s \in \mathbb{N}}$\} is also countable. Similarly, we also have that since $\prod_{i=1}^{\infty} \mathbb{N}$ is countable for each $\tau \in \mathbb{N}$

$\Phi_\tau(A) = \{\phi_{-s_1} \circ ... \circ \phi_{-s_\tau}(A) : s_1, ..., s_\tau \in \mathbb{N}, \phi_1, ..., \phi_\tau \in \Phi\}$

is countable. Hence, as

$$O^-_\Phi(A) = \bigcup_{n=1}^{\infty} \Phi_n(A)$$

$O^-_\Phi(A)$ is countable. We can thus order the elements of $O^-_\Phi(A)$ as $A_1, A_2, ....$

Suppose that for each $C \in O^-_\Phi(A)$, $\mu(C \cap B) = 0$. Then for each $n \in \mathbb{N}$, $\mu(A_n \cap B) = 0$. Giving

\[
0 \leq \mu \left( \left( \bigcup_{n=1}^{\infty} A_n \right) \cap B \right) = \mu \left( \bigcup_{n=1}^{\infty} (A_n \cap B) \right) \leq \sum_{n=1}^{\infty} \mu(A_n \cap B) = 0 \leq \mu(B).
\]

Hence

\[
0 < \mu(A) \leq \mu \left( \bigcup_{n=1}^{\infty} A_n \right) = \mu \left( \left( \bigcup_{n=1}^{\infty} A_n \right) \cap B \right) + \mu \left( \left( \bigcup_{n=1}^{\infty} A_n \right) \cap B^c \right)
\]

224
Thus $0 < \mu(B) < \mu(X)$ which contradicts $\Phi$ being ergodic as $B$ is $\Phi$-invariant.

With this Lemma we can now prove the first generalisation of our main theorems.

**Theorem 6.5**

Let $\Psi$ be a family of transformations on a measure space $(X, B, \mu)$ and suppose that there is an ergodic countable subfamily $\Phi$ of invertible transformations. Then for any collection of standard sets $J$, $\Psi$ generates Carathéodory's definition.

**Proof:**

Let $J$ be a collection of standard sets for $(X, B, \mu)$.

Let $J_1$ and $J_2$ be elements of equal measure from a standard collection of sets, $J$ for $(X, B, \mu)$. Also let $B$ be a $\Phi$-invariant subset of $X$. To prove the theorem we must satisfy the conditions of Corollary 6.2.

As was seen in Lemma 6.2 $O_\Phi^-$ is countable and so we can order the set $O_\Phi^-(J_1)$ as $J_{1,1}, J_{1,2}, \ldots$. Note that for each $n \in \mathbb{N}$

$$J_{1,n} = \phi_1^{-s_1} \circ \phi_2^{-s_2} \circ \cdots \circ \phi_r^{-s_r}(J_1)$$

for some $r \in \mathbb{N}$, $s_1, s_2, \ldots, s_r \in \mathbb{N}$ and $\phi_1, \ldots, \phi_r \in \Phi$. For each $n \in \mathbb{N}$ denote
the $\phi_1^{-s_1} \circ \phi_2^{-s_2} \circ \ldots \circ \phi_r^{-s_r}$ associated with $J_{1,n}$ be $\Phi^{-n}$ (so that $\Phi^{-n}(J_1) = J_{1,n}$.

We note that since each $\phi \in \Phi$ is a measure preserving and invertible transformation, $\Phi^{-n}(B) \in B$ whenever $B \in B$ and $\mu(\Phi^{-i}(B)) = \mu(\Phi^{-j}(B))$ for each $B \in B$ and each pair $i, j \in \mathbb{N}$.

The rest of the proof is essentially identical to that for Theorem 6.2. While we would like to be able to just refer to Theorem 6.2 we cannot as while $\Phi^{-n}$ behave in all the necessary ways like the inverse images of the transformation $T$ used in Theorem 6.2. $\Phi$ is not actually a transformation.

Now, let

$$A_1 = \Phi^{-1}(J_2) \cap J_1 \in B$$

and then, inductively, let

$$A_k = (J_1 - \bigcup_{i=1}^{k-1} A_i) \cap (J_2 - \bigcup_{i=1}^{k-1} \Phi^{-i} A_i)).$$

These $A_k$ are an indexation of a set of splinters of $J_1$ with respect to $J_2$.

Clearly $\{A_i\}_{i=1}^{\infty}$ is a disjoint sequence of measurable sets, as is $\{\Phi^{-i} A_i\}_{i=1}^{\infty}$.

Thus as $\mu(\bigcup_{i=1}^{\infty} A_i) \leq \mu(J_1) < \infty$ and $\mu(\bigcup_{i=1}^{\infty} T^{-i} A_i) \leq \mu(J_1) < \infty$

$$\mu(\bigcup_{i=1}^{\infty} \Phi^{-i} A_i) = \sum_{i=1}^{\infty} \mu(\Phi^{-i} A_i) = \sum_{i=1}^{\infty} \mu(A_i) = \mu(\bigcup_{i=1}^{\infty} A_i).$$

Note that $J_1 - \bigcup_{i=1}^{\infty} A_i, J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i \in B$.

It is clear that $\mu(\bigcup_{i=1}^{\infty} A_i) \leq \mu(J_1)$. Now suppose that $\mu(\bigcup_{i=1}^{\infty} A_i) \neq \mu(J_1)$, then $\mu(\bigcup_{i=1}^{\infty} A_i) < \mu(J_1), \mu(J_1 - \bigcup_{i=1}^{\infty} A_i) > 0$ and $\mu(J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i) > 0$. As $\Phi$ is ergodic, by Lemma 6.2 there is an $n \in \mathbb{N}$ such that

$$\mu(\Phi^{-n}(J_1 - \bigcup_{i=1}^{\infty} A_i) \cap (J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i)) > 0.$$ 

But then, as

$$(J_1 - \bigcup_{i=1}^{\infty} A_i) \subseteq (J_1 - \bigcup_{i=1}^{n-1} A_i)$$
and

\[(J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i) \subseteq (J_2 - \bigcup_{i=1}^{n-1} \Phi^{-i} A_i)\]

we would then have

\[\emptyset \neq (\Phi^{-n} (J_1 - \bigcup_{i=1}^{\infty} A_i) \cap (J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i))\]

Hence

\[\mu(J_1 - \bigcup_{i=1}^{\infty} A_i) = 0 = \mu(J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i).\]

Now, using Lemma 6.1, and the fact that \(B\) is \(\Phi\)-invariant we get

\[
\mu_*(B \cap J_1) = \mu_*(B \cap \bigcup_{i=1}^{\infty} A_i) + \mu_*(B \cap (J_1 - \bigcup_{i=1}^{\infty} A_i)) \\
= \mu_*(B \cap \bigcup_{i=1}^{\infty} A_i) \\
= \mu_*(\bigcup_{i=1}^{\infty} (B \cap A_i)) \\
= \sum_{i=1}^{\infty} \mu_*(B \cap A_i) \\
= \sum_{i=1}^{\infty} \mu_*(\Phi^{-i}(B \cap A_i)) \\
= \mu_*(\bigcup_{i=1}^{\infty} \Phi^{-i}(B \cap A_i)) \\
= \mu_*(\bigcup_{i=1}^{\infty} \Phi^{-i} B \cap \Phi^{-i} A_i) \\
= \mu_*(\bigcup_{i=1}^{\infty} B \cap \Phi^{-i} A_i) \\
= \mu_*(B \cap \bigcup_{i=1}^{\infty} \Phi^{-i} A_i) \\
= \mu_*(B \cap (J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i)) + \mu_*(B \cap (J_2 - \bigcup_{i=1}^{\infty} \Phi^{-i} A_i)) \\
= \mu_*(B \cap J_2).
\]

Hence the conditions for Corollary 6.2 are satisfied for \(\Psi\) so that \(\Psi\) generates Carathéodory's definition. Now suppose that \(B\) is a \(\Psi\)-invariant set for which there is a \(\theta < 2\) that allows \(B\) to satisfy

\[\mu_*(B \cap J) + \mu_*(B^c \cap J) \leq \theta \mu(J)\]

227
for each $J \in \mathcal{J}$ where $\mu_*$ is the usual outer measure associated with $\mu$. As $B$ is $\Psi$-invariant it must also be $\Phi$-invariant. Then, as $\Phi$ generates Carathéodory's definition we know that either $B$ or $B^c$ has outer measure zero. Thus $\Psi$ generates Carathéodory's definition.

Due to the inability to preserve measure of forward iterations of transformations that are not invertible, we are not able to achieve as significant a generalisation for Theorem 6.4 as we are for Theorem 6.2. The generalisation, in fact follows, essentially, straight away.

**Theorem 6.6**

Let $\Psi$ be a family of measure preserving transformations on a measure space $(X, \mathcal{B}, \mu)$. If there is a $\psi \in \Psi$ that is an outer measure preserving ergodic transformation then $\Psi$ generates the Carathéodory definition.

**Proof:**

Let $\mathcal{J}$ be a collection of standard sets for $X$. Suppose that $B$ is a $\Psi$-invariant set for which there is a $\theta < 2$ that allows $B$ to satisfy

$$\mu_*(B \cap J) + \mu_*(B^c \cap J) \leq \theta \mu(J)$$

for each $J \in \mathcal{J}$ where $\mu_*$ is the usual outer measure associated with $\mu$. As $B$ is $\Psi$-invariant it must also be $\psi$-invariant. Then, as $\psi$ generates Carathéodory's definition we know that either $B$ or $B^c$ has outer measure zero. Thus $\Psi$ generates Carathéodory's definition.

We now proceed to investigating the properties of outer measure preserving transformations.
6.2 Outer measure preserving transformations

After demonstrating the relatively simple result that an invertible measure preserving transformation must be outer measure preserving we begin this investigation by showing that a transformation, $T$, is outer measure preserving if for each member, $B$, of the $\sigma$-algebra, $T(B)$ is measurable. This transfers our discussion to investigating which transformations have measurable images of Borel sets. We establish some results showing that there are a lot of transformations that do have measurable images of Borel sets, and also provide an easy method of demonstrating that all of the specific examples of transformations that we have considered so far (irrational rotations, Kakutani I type transformations etc.) are outer measure preserving. We also show that there are measure preserving transformations that are not outer measure preserving by providing a counter example. This counter example also highlights the importance of the measurability of images of Borel sets as the images of Borel sets in the provided example are not always measurable. We conclude the chapter by showing that the present chapter, and in particular Theorem 6.4 includes all of the extra cases of transformations that generate Carathéodory's definition provided by Chapter 5. This is done by showing that if the primitive transformation is ergodic and outer measure preserving, then so is any derived tower transformation.

**Proposition 6.1**

*An invertible measure preserving transformation is measure preserving.*

**Proof:**

As $T$ is invertible and outer measure preserving, for all $B \in \mathcal{B}$ we know that $\mu(B) = \mu(T^{-1}(B))$ and that

$$\mu(T(B)) = \mu(T^{-1}(T(B)))$$
Now let $B \subseteq X$, Then

$$\mu^*(B) = \inf \{\mu(A) : B \subseteq A, A \in \mathcal{B}\}$$

and

$$\mu^*(T^{-1}(B)) = \inf \{\mu(A) : T^{-1}(B) \subseteq A, A \in \mathcal{B}\}.$$

Thus we have that

$$\mu^*(T^{-1}(B)) = \inf \{\mu(A) : T^{-1}(B) \subseteq A, A \in \mathcal{B}\} \leq \inf \{\mu(T^{-1}(A)) : B \subseteq A, A \in \mathcal{B}\}$$

$$= \inf \{\mu(A) : T^{-1}(B) \subseteq A, A \in \mathcal{B}\}$$

$$= \mu^*(B).$$

Similarly we also have that

$$\mu^*(B) = \inf \{\mu(A) : (B) \subseteq A, A \in \mathcal{B}\}$$

$$\leq \inf \{\mu(T(A)) : T^{-1}(B) \subseteq A, A \in \mathcal{B}\}$$

$$= \inf \{\mu(A) : T^{-1}(B) \subseteq A, A \in \mathcal{B}\}$$

$$= \mu^*(T^{-1}(B)).$$

Hence $\mu^*(B) \leq \mu^*(T^{-1}(B)) \leq \mu^*(B)$ and thus $\mu^*(B) = \mu^*(T^{-1}(B))$. \hfill \diamondsuit

**Lemma 6.3**

*If $T$ is a measure preserving transformation on $(X, \mathcal{B}, \mu)$ such that $T(A)$ is measurable for any $A \in \mathcal{B}$, then for any $B \subseteq X$

$$\mu^*(T^{-1}B) = \mu^*(B).$$
Proof:

Let $B \subseteq X$, then

$$\mu_*(B) = \inf \{ \mu(A) : B \subseteq A, A \in \mathcal{B} \}$$

and

$$\mu_*(T^{-1}B) = \inf \{ \mu(A) : T^{-1}B \subseteq A, A \in \mathcal{B} \}.$$\n
Note that as for all $A \in \mathcal{B}$ such that $B \subseteq A$, $T^{-1}(A) \in \mathcal{B}$, $T^{-1}B \subseteq T^{-1}A$ and $\mu(T^{-1}A) = \mu(A)$ so that $\mu_*(T^{-1}(B)) \leq \mu_*(B)$.

Now suppose $\mu_*(B) \neq \mu_*(T^{-1}B)$. Choose $A \in \mathcal{B}$ such that $B \subseteq A$ and $C \in \mathcal{B}$ such that $T^{-1}(B) \subseteq C$ and $\mu(C) < \mu_*(B)$. Now

$$\mu(T^{-1}A) = \mu(A).$$

$$T^{-1}B \subseteq T^{-1}(A) \cap C,$$

and hence

$$\mu(T^{-1}(A) \cap C) \leq \mu(C) < \mu_*(B).$$

Note that

$$T^{-1}(A) \cap C \subseteq T^{-1}(T(T^{-1}(A) \cap C)) \in \mathcal{B}$$

so that

$$\mu(T^{-1}(A) \cap C) \leq \mu(T^{-1}(T(T^{-1}(A) \cap C))) = \mu(T(T^{-1}(A) \cap C))$$

and hence

$$\mu(T(T^{-1}A - C)) \geq \mu(T^{-1}A - C)$$

$$= \mu(T^{-1}A) - \mu(C)$$

$$> \mu(A) - \mu_*(B).$$
We also have that $T^{-1}B \cap (T^{-1}A - C) = \emptyset$ so that $B \cap (T(T^{-1}A - C)) = \emptyset$. That is $B \subseteq A - T(T^{-1}A - C)$ and hence

$$
\begin{align*}
\mu_*(B) & \leq \mu(A - T(T^{-1}A - C)) \\
& = \mu(A) - \mu(T(T^{-1}A - C)) \\
& < \mu(A) - (\mu(A) - \mu_*(B)) \\
& = \mu_*(B).
\end{align*}
$$

This contradiction gives the result. \hfill \diamond

At one stage, the investigation of the problem of describing functions under which the image of Borel sets is measurable was an active area of research. The pursuit of the solution to the following theorem, Theorem 6.7, in fact lead to the creation of Suslin sets. Theorem 6.7, like the spectral theorem, is a deep result that requires substantial work to motivate and prove. Therefore, like the spectral theorem, we shall state the Theorem without proof. In the theorem the denotation of a measure space will be of the form $(X, \mathcal{B}, \mu)$ where $X$ is the set, $\mathcal{B}$ is the Borel $\sigma$-algebra and $\mu$ is a measure defined on $\mathcal{B}$.

**Theorem 6.7**

*Let $(X, \mathcal{B}, \nu)$ and $(Y, \mathcal{D}, \mu)$ be two measure spaces. Let $X$ be a complete separable metric space and $Y$ be a Hausdorff space in which every closed set is $\mu$-measurable. Then if $f : X \to Y$ is continuous the $f$ image of Borel sets is $\mu$-measurable.*

We will use Theorem 6.7 (generally in the case where $(Y, \mathcal{D}, \mu) = (X, \mathcal{B}, \nu)$ so that we need only concern ourselves with one measure) to show that each of the transformations and types of transformations that we have considered so far are outer measure preserving. As some of the transformations have
been neither invertible nor continuous we need to extend Theorem 6.7 to use it to show that each of the specific examples so far considered are in fact outer measure preserving. We were able to show that Theorem 6.7 can be extended by using the fact that the set of discontinuities in our examples is small in the sense that they are countable and have at most one limit point. In fact we show that, whenever the set of discontinuities of a transformation, $T$, are small in the sense that they are nowhere dense (a classification that will be defined shortly), Theorem 6.7 can be extended to apply to $T$.

Before demonstrating the extension possible on Theorem 6.7 we need to introduce some definitions and a couple of preliminary propositions. First note that

1. a subspace of a complete metric space with the inherited metric is complete if and only if it is closed, and

2. single points in a Hausdorff space are closed.

We next need to define two classifications of sets.

**Definition 6.11**

A discrete subset of a topological space $(X, T)$ is a subset $D$ of $X$ for which the topology $T_D$ on $D$ inherited from $T$ is discrete. That is, individual points are open.

For us, the important aspect of this definition is that for each element, $d$, of a discrete set $D$ in a topological space there is an open set whose intersection with $D$ is $d$.  

233
Definition 6.12

A nowhere dense subset of a topological space is a set whose closure has no interior.

We now define a sequence of collections of sets that are of importance to this discussion. It will be shown that if a transformation, $T$, from a complete separable metric measure space into a hausdorff measure space whose closed sets are measurable has a set of discontinuities that is an element of one of these collections then $T$ images of Borel sets will be measurable.

Definition 6.13

Let $(X, T)$ be a topological space. Let $D_1$ denote the discrete sets. Suppose, now that the collection $D_n$ is defined for some $n \in \mathbb{N}$ then $D \in D_{n+1}$ if there is a set $D_1 \in D_1$ such that for each open set, $F_{D_1}$, $D_1 \subset F_{D_1}$,

$$D \cap F_{D_1}^c \in D_n.$$

The collection $D_n$ will be called the collection of sets discrete at the $n$th level or simply the $n$-discrete sets. Similarly an element $D \in D_n$ will be known as a set discrete at the $n$th level or simply a $n$-discrete set.

To provide insight into this definition, we give an example of a $D_2$ set that is not a $D_1$ set and an example of a $D_3$ that is not a $D_2$ set. Note that the sequence $N = \{1/n : n \in \mathbb{N}\}$ is a discrete and hence $D_1$ set. However, $N \cup \{0\}$ is not discrete. However, $\{0\}$ is a discrete and hence $D_1$ set. Further for any open set $F_0$ around 0, $N_1 = ((N \cup \{0\}) - F_0) \subset N$ and hence $N_1 \in D_1$. Therefore $N \cup \{0\}$ is a $D_2$ set but not a $D_1$ set. In a similar way it can be shown that $\{(1/n, 1/m) : n, m \in \mathbb{Z}, m, n \neq 0\} \subset \mathbb{R}^2$ is a $D_3$ set but not a $D_2$ set.
Finally before demonstrating the extension to Theorem 6.7 we prove two propositions that are used not only in the following theorem but also in some subsequent Theorems exploring the properties \( n \)-discrete sets.

**Proposition 6.2**

For each discrete subset, \( D \), of a metric space there exists a family of balls \( \{ F_d : d \in D \} \) such that \( d \in F_d \) and \( F_d \cap F_g = \emptyset \) whenever \( d \neq g \).

**Proof:**

As \( D \) is discrete, for each \( d \in D \) there exists an open set \( G_d \) containing \( d \) and no other points in \( D \). \( G_d \) contains a ball \( B(d, r_d) \) that is open, is centred on \( d \) (and thus contains \( d \)) and no other points in \( D \). Then the collection required is \( \{ F_d \} \) defined by \( F_d = B(d, r_d / 3) \).

**Proposition 6.3**

A discrete subset of a separable metric space is countable.

**Proof:**

Let \((X, d)\) be a separable metric space and \( D \) be a discrete subset. Then, by Proposition 6.2 there is a collection of disjoint open balls \( \{ F_d \}_{d \in D} \) such that \( D \cap F_d = d \) for each \( d \in D \). As \( X \) is separable there is a dense countable subset of \( X \), say \( S \), which must have a non-empty intersection with each \( F_d \). Should \( D \) not be countable, then \( \{ F_d \}_{d \in D} \) would not countable and thus \( S \) could not be countable. This contradiction gives the result.

**Theorem 6.8**

Let \((X, \mathcal{B}, \nu)\) and \((Y, \mathcal{D}, \mu)\) be two measure spaces. Let \( X \) be a complete separable metric space and \( Y \) be a Hausdorff space in which every closed set is \( \mu \)-measurable. Then if \( f : X \to Y \) has a set of discontinuities, \( D \in \mathcal{D}_n \), for some \( n \in \mathbb{N} \) then the \( f \) image of Borel sets are \( \mu \)-measurable.
Proof:
We prove the result by induction. That is, we first prove the case where the
set of discontinuities of \( f \) is a \( \mathcal{D}_1 \) set. Let \( D \) be the set of discontinuities.
We know, by Propositions 6.2 and 6.3 that there is a countable collection of
disjoint balls \( \mathcal{B} = \{ \overline{B(d, r_d)} : d \in D \} \).

For any ball \( \overline{B(d, r_d)} \in \mathcal{B} \) let
\[
C_{d,n} = \overline{B(d, r_d)} - B(d, r_d/n).
\]
For each \( n \in \mathbb{N} \), \( (C_{d,n}, B|_{C_{d,n}}, \mu|_{C_{d,n}}) \) is a measure space on a closed (and thus
complete) separable metric space, \( C_{d,n} \). Further \( f \) restricted to \( C_{d,n} \) maps \( C \)
continuously into the Hausdorff space \( (Y, \mathcal{D}, \mu) \) for which \( \mu \) measures closed
sets.

Therefore, by Theorem 6.7, for all \( B \in \mathcal{B}|_{C_{d,n}} \), \( f(B) \) is \( \mu \) measurable. Further
\( f(d) \) is a point in \( Y \), which is therefore closed and thus measurable.

Therefore for any \( B \in \mathcal{B} \),
\[
 f(B \cap \overline{B(d, r_d)}) = f \left( \bigcup_{n=1}^{\infty} (B \cap C_{d,n}) \cup B \cap \{d\} \right) \\
= \bigcup_{n=1}^{\infty} f(B \cap C_{d,n}) \cup f(B \cap \{d\})
\]
which is a countable union of measurable sets and thus is measurable.

Next we consider
\[
X_1 = X - \bigcup_{d \in D} B(d, r_d).
\]
\( (X_1, \mathcal{B}|_{X_1}, \mu|_{X_1}) \) is measure space on a complete separable metric space over
which \( f \) is a continuous mapping into \( Y \). Thus for any \( B \in \mathcal{B} \), \( f(B \cap X_1) \) is
μ measurable.

Hence for any $B \in \mathcal{B}$

$$f(B) = f \left( (B \cap X_1) \cup \bigcup_{d \in D} (B \cap B(d, r_d)) \right)$$

which is a countable union of measurable sets and thus measurable.

We now suppose that the result is true for a $\mathcal{D}_n$ set and prove that it must then be true for a $\mathcal{D}_{n+1}$ set. The proof is the same as the above but we use the inductive hypothesis instead of using the thesis’ Theorem 6.7.

We are now supposing that $D \in \mathcal{D}_{n+1}$ is the set of discontinuities of $f$. We know that as $D \in \mathcal{D}_n$ there is a $D_1 \in \mathcal{D}_1$ such that for each collection of open sets $\{F_d : d \in D_1, F_d \cap D_1 = d\}$, $D \cap (X - \cup_{d \in D_1} F_d) \in \mathcal{D}_n$. By propositions 6.2 and 6.3 we know that we can choose $\{F_d\}_{d \in D_1}$ to be a countable collection of open balls with radii $r_d$ whose closures are disjoint. We then set

$$F_{d,m} = F_d - B \left( d, \frac{r_d}{m} \right)$$

Each of the spaces $(F_{d,m}, \mathcal{B}|_{F_{d,m}}, \mu|_{F_{d,m}})$ are measure spaces defines on a complete separable metric space over which $f$ is continuous except for a $n$-discrete set of discontinuities. Thus for each $B \in \mathcal{B}$, $f(B \cap F_{d,m})$ is measurable. Also $f(d)$ is a single point and thus $\mu$ measurable. Therefore for each $B \in \mathcal{B}$

$$f(B \cap F_{d}) = f \left( (B \cap \{d\}) \cup \bigcup_{m=1}^{\infty} (B \cap F_{d,m}) \right)$$

$$= f(B \cap \{d\}) \cup \bigcup_{m=1}^{\infty} f(B \cap F_{d,m})$$

237
which is a countable union of measurable sets and thus measurable. Also, letting

\[ X_2 = X - \bigcup_{d \in D_1} F_d, \]

we note that \((X_2, \mathcal{B}|_{F_d}, \mu|_{F_d})\) is a measure space defined on a complete separable metric space such that \(f\) maps \(X_2\) continuously into \(Y\) except for a \(n\)-discrete set of discontinuities. By the earlier arguments in this proof we know that this implies that for any \(B \in \mathcal{B}\), \(f(B \cap X_2)\) is measurable. Therefore we now have that if \(B \in \mathcal{B}\)

\[
f(B) = f \left( (B \cap X_2) \cup \bigcup_{d \in D_1} (B \cap \overline{F}_d) \right)
= f(B \cap X_2) \cup \bigcup_{d \in D_1} f(B \cap \overline{F}_d),
\]

which is a countable union of measurable sets and thus is measurable. That is, under a function \(f\) mapping \(X\) into \(Y\) that has a \((n + 1)\)-discrete set of discontinuities, the image of Borel sets is \(\mu\) measurable. By mathematical induction we now know that the result is true for any \(n \in \mathbb{N}\) which proves the Theorem.

We now provide some properties of the \(n\)-discrete sets.

**Theorem 6.9**

*Let \(X\) be a separable metric space. Then for any \(n \in \mathbb{N}\), \(\mathcal{D}_n\) is countable.*

**Proof:**

Assume that for some \(n \in \mathbb{N}\) any \(\mathcal{D}_n\) set is countable. Let \(D \in \mathcal{D}_{n+1}\) Then there is a countable set \(D_1 \in \mathcal{D}_1\) and a corresponding countable set of disjoint balls \(\{F_d\}_{d \in D_1}\) of radius \(r_d\) such that \(F_d \cap D_1 = d\) for each \(d \in D_1\). Further
$D \cap (X - \cup_{d \in D_1} F_d) \in D_n$. We therefore know that

$$G_n = D \cap \left( X - \bigcup_{d \in D_1} B \left( d, \frac{r_d}{n} \right) \right) \in D_n$$

and thus is countable. Since

$$D \subseteq D_1 \cup \bigcup_{n \in \mathbb{N}} G_n,$$

which must be countable, it follows that $D$ is countable and thus as $D$ is an arbitrary set in $D_{n+1}$ any set in $D_{n+1}$ is countable. Combined with Proposition 6.3, which gives us that $D_1$ is countable, this inductive step gives the result.

We wish to prove one other property of $n$-discrete sets. That is that, at least in a metric space, they are nowhere dense. We use $Int(D)$ to denote the interior of a set $D$.

**Theorem 6.10**

Let $X$ be a separable metric space that has non isolated points. Then for any $n \in \mathbb{N}$, $D_n$ is nowhere dense.

**Proof:**

We again use mathematical induction. Suppose that $D \in D_1$. Let $x \in Int(D)$. Then there is an open ball $B(x, r) \subset Int(D)$ around $x$ which must contain a point $d \in D$. Then there must be an open set $F_d \subset B(x, r)$ containing $d$ but no other point in $D$. Let $y \in F_d - \{d\}$ so that $y \in Int(D)$. The ball $B(y, |d - y|/2)$ then contains no points of $D$ contradicting $y \in D$. Thus $Int(D) = \emptyset$ and hence we know that any $D_1$ set is nowhere dense.

Now suppose that any $D_n$ set is nowhere dense and let $D \in D_{n+1}$. Then there is a $D_1 \in D_1$ such that for any collection of balls (which will be count-
able) \( \{F_d : d \in D_1\} \), \( D \cap (X - \cup_{d \in D_1} F_d) \in \mathcal{D}_n \). Hence for any \( m \in \mathbb{N} \)

\[
D \cap \left( X - \bigcup_{d \in D_1} B \left( d, \frac{1}{m} \right) \right)
\]

has empty interior. Hence for any \( m \in \mathbb{N} \)

\[
\text{Int}(\overline{D}) \subseteq \bigcup_{d \in D_1} B \left( d, \frac{1}{m} \right) \subset \bigcup_{d \in D} B \left( d, \frac{1}{m - 1} \right)
\]

and therefore

\[
\text{Int}(\overline{D}) \subseteq \bigcap_{m \in \mathbb{N}} \bigcup_{d \in D_1} B \left( d, \frac{1}{m} \right) = D_1.
\]

As \( D_1 \) is countable it cannot contain an open set in a metric space and therefore \( \text{Int}(\overline{D}) = \emptyset \).

Combining the above results we can show that each of the specific transformations considered so far are outer measure preserving.

**Corollary 6.3**

*Irrational rotations, Kakutani I type transformations, Chacon's \( \xi \) and*

\[
T : [0, 1] \rightarrow [0, 1]
\]

*defined by*

\[
T(z) = 2(z)(\text{mod} \ 1)
\]

*are outer measure preserving*

**Proof:**

Irrational rotations are continuous functions on complete separable metric spaces and hence by Theorem 6.7 and Lemma 6.3 must be outer measure preserving. \( T \) is a transformation on a complete separable metric space with only one discontinuity. The set of discontinuities for \( T \) are therefore nowhere dense and Theorem 6.8 gives us that \( T \) is outer measure preserving.
The Kakutani I type transformations and Chacon's $\xi$ are shown to be outer measure preserving in the same way. Let $P$ be an arbitrary transformation selected from Chacon's $\xi$ and the Kakutani I type transformations. $P$ is defined either on $(2, 3)$-adic or $b$-adic intervals in $([0, 1], B, \mu)$. Where $B$ and $\mu$ are the usual Lebesgue $\sigma$-algebra and measure associated with $[0, 1)$. Then $P$ is continuous on all but one $(2, 3)$-adic or $b$-adic interval of order $n$ for each $n$. Further let the sequence $\{J_n\}$ be given by defining $J_n$ to be the $(2, 3)$-adic or $b$-adic interval of order $n$ over which $P$ is not continuous for each $n$. Then by the definition of Chacon's $\xi$ and Kakutani I type transformations we know that $\{J_n\}$ is a decreasing sequence and that

$$\bigcap_{n=1}^{\infty} J_n$$

is a single point. Thus this single point is the only possible limiting point for the points at which $P$ is discontinuous. As a single point is discrete we know that the set of discontinuities for $P$ must be a subset of a 2-discrete set. The only problem is that $[0, 1)$ is not complete. We get around this problem as follows. We define $P_1$ on $[0, 1]$ (with the usual Lebesgue $\sigma$-algebra, $B_1$ and measure, $\mu_1$ associated with it) as

$$P_1(x) = \begin{cases} 
P(x), & x \in [0, 1), \\
1, & x = 1.
\end{cases}$$

Then $P_1$ is a transformation defined on a complete metric space with a 2-discrete set of discontinuities. Thus for all $B \in B$, $B \in B_1$ and thus by Theorem 6.8 $P(B) = P_1(B)$ is $\mu_1$ measurable. To show that $P(B)$ is $\mu$ measurable we take $\mu_*$ and $\mu_{1,*}$ as the usual outer measures associated with $\mu$ and $\mu_1$ respectively and take $C \subset [0, 1) \subset [0, 1]$. Then using the fact that $X \in B_1$ and thus is $\mu_1$-measurable and that $\mu_* = \mu_{1,*}$ on $[0, 1)$ we have

$$\mu_*(C) = \mu_{1,*}(C)$$

241
\[= \mu_1,*(C \cap P(B)) + \mu_1,*(C \cap P(B)^c_{[0,1]})\]
\[= \mu_1,*(C \cap P(B)) + \mu_1,*(C \cap P(B)^c_{[0,1]} \cap [0,1]) + \mu_1,*(C \cap P(B)^c_{[0,1]} \cap [0,1]^c)\]
\[= \mu_1,*(C \cap P(B)) + \mu_1,*(C \cap P(B)^c_{[0,1]} \cap [0,1]) + 0\]
\[= \mu_1,*(C \cap P(B)) + \mu_1,*(C \cap P(B)^c_{[0,1]} \cap [0,1])\]
\[= \mu_1,*(C \cap P(B)) + \mu_1,*(C \cap P(B)^c_{[0,1]} \cap [0,1])\]

where the superscripts \(c_{[0,1]}\) and \(c_{[0,1]}\) denote complementation in \([0,1]\) and \([0,1]\) respectively. Hence for all \(B \in \mathcal{B}, P(B)\) is measurable and hence outer measure preserving.

We conclude this section by defining an measure preserving transformation that is not outer measure preserving. Since the closure of the rationals in \([0,1]\) is \([0,1]\), Theorem 6.8 does not include a transformation with the rationals as a set of discontinuities. Such a transformation is in fact one of the types of transformation for which we have no results concerning their outer measure preserving properties. To guarantee that a transformation be outer measure preserving (at least for a complete separable metric space) we know that the closure of the points of discontinuity must have no interior. But we don't know that having such a set of discontinuities is sufficient. On the other hand, by the transformation that we construct below, we know that not all transformations with a set of discontinuities with positive outer measure can be outer measure preserving. This is because the constructed example has a set of discontinuities of outer measure zero. However the set of discontinuities is uncountable so this example is still not instructive in the case for transformations with the rationals as a set of discontinuities.

We now construct the example of a transformation of a measure space onto itself that is measure preserving but not outer measure preserving. We also
note that in this case the image of Borel sets is not necessarily measurable, so the condition of requiring that images of Borel sets be measurable may still be both necessary and sufficient.

Let \((X, \mathcal{B}, \mu)\) be the usual Lebesgue measure space on \([0,1]\). Let \(C\) be the Cantor set, \(C_L = C \cap [0,1/2)\) and \(C_U = C \cap [1/2,1]\). We know that \(C\) is equal to the set of elements of \([0,1]\) whose ternary expansion does not contain the digit '1'. We therefore know that there is a bijection between \(C\) and \([0,1]\), \(c_1 : C \rightarrow [0,1]\) that takes each element of \(C\), and changes each '2' in its ternary expansion to a '1' and gives \(x\) the value whose binary expansion is the resulting expression. Also for each \(x \in [0,1]\), let \(x = 0.x_1x_2x_3...\) be \(x\)'s ternary expansion. Define \(c_2\) as \(c_2(x) = 0.x_2x_3...\) In this manner \(c_2\) can be considered as a bijection either between \(C_U\) and \(C\) or \(C_L\) and \(C\).

Next let \(G\) be a non-measurable subset of \([0,1]\) with positive outer measure. The \(\mathcal{P}\)-invariant non-measurable set described in section 7.3 is sufficient as an example of such a set. In fact, we may assume that

\[
\mu_*(G \cap B) = \mu_*(G^c \cap B) = \mu(B)
\]

for any \(B \in \mathcal{B}\). We define the transformation \(T : X \rightarrow X\) by

\[
T(x) = \begin{cases} 
\chi_G(c_1 \circ c_2(x)) \cdot c_1 \circ c_2(x), & x \in C_L, \\
c_2(x), & x \in C_U, \\
x, & \text{otherwise}.
\end{cases}
\]

We now prove that \(T\) has the desired properties.

**Theorem 6.11**

*The transformation \(T\) as constructed above is a measure preserving transformation that is not outer measure preserving and for which the image of borel sets is not always \(\mu\)-measurable.*

243
Proof:
Firstly, as $\mu_*(C_L) = 0$, $C \in \mathcal{B}$ and $\mu(C_L) = 0$. Also $T(C_L) = G \cup \{0\}$ and so is not measurable. Also $G \subset X$ has positive outer measure but $T^{-1}(G) \subset C$ and hence, as $C \in \mathcal{B}$, $\mu_*(T^{-1}(G)) = 0$. Hence all that remains is to show that $T$ is measure preserving.

Let $B \in \mathcal{B}$. Then

$$T^{-1}(B) = (B - C) \cup T^{-1}((G \cup C) \cap B) = (B - C) \cup C_1 \quad (29)$$

where $C_1 = T^{-1}((G \cup C) \cap B) - (B - C) \subset C$. As $C_1 \subset C$, $\mu_*(C_1) = 0$ and hence $C_1 \in \mathcal{B}$ so that by (29) $T^{-1}(B)$ is measurable. Further, noting that $C$ is measurable and hence that

$$\mu(B - C) = \mu(B) - \mu(C)$$

we have that

$$\mu(B) = \mu(B - C)$$

$$\leq \mu((B - C) \cup C_1)$$

$$= \mu(T^{-1}(B))$$

$$= \mu((B - C) \cup C_1)$$

$$\leq \mu(B - C) + \mu(C_1)$$

$$= \mu(B) + 0$$

$$= \mu(B).$$

Thus $T$ is measure preserving and hence $T$ has all of the desired properties.

$\Diamond$

244
6.3 Towers and ergodic outer measure preserving transformations

With the results shown already in this chapter and the results presented in Chapter 5, it is clear that a transformation on an appropriate measure space is ergodic if and only if any derived tower transformation is. This result follows immediately from Theorem 5.7 and Theorem 6.4. However, in looking for transformations that generate the Carathéodory definition, the proof of equivalence of generation of spaces and their derived towers needs its own separate result. To remedy this problem and ensure that all of our results are covered by Theorem 6.4 we show directly that the property of ergodicity and the outer measure preserving property (or lack thereof) is also necessarily shared by dynamical systems and their derived tower systems.

Theorem 6.12 is another equivalent definition of ergodicity that we have proved that makes Theorem 6.13, the result demonstrating that the property of ergodicity is shared by primitive dynamical systems and their derived systems, easier to prove. We then show in Theorem 6.14 that primitive and derived theorems also share the property of preserving outer measure. These Theorems are the last results that we will present concerning the generation of Carathéodory's definition. The chapters that follow will then discuss the transformation invariant subsets of a generating dynamical system that do not satisfy the requirements to be a set considered in defining a transformation that generates Carathéodory's definition. As we have discussed in Chapters 1 and 2 the counterexamples are vital in showing the significance and validity of our method of generating Carathéodory's definition.

Theorem 6.12

A transformation $T$ on the measure space $(X, \mathcal{B}, \mu)$ of finite positive mea-
sure is ergodic if and only if for all $T$-invariant sets $B$ in $B$, $\mu(B) = 0$ or $\mu(B \cap C) = \mu(C)$ for all sets $C \in B$

Proof:
Suppose $T$ is ergodic and $B \in B$ is $T$-invariant, then $\mu(B) = 0$ or $\mu(B) = \mu(X)$.
Let $C \in B$. If $\mu(C) = 0$, then $0 \leq \mu(B \cap C) \leq \mu(C) = 0$ so that $\mu(C) = 0 = \mu(B \cap C)$.

Now consider $\mu(C) > 0$. We know $\mu(B \cap C) \leq \mu(C)$. Now suppose that $\mu(B \cap C) \neq \mu(C)$. Then

$$
\mu(B) = \mu(B \cap C) + \mu(B \cap C^c) < \mu(C) + \mu(C^c) = \mu(X),
$$

which is impossible as $T$ is ergodic. Therefore $\mu(B \cap C) = \mu(C)$.

On the other hand suppose that for all $T$-invariant sets $B \in B$, $\mu(B \cap C) = \mu(C)$ for all $C \in B$ or $\mu(B) = 0$. We then have $\mu(B) = 0$ or $\mu(B) = \mu(B \cap X) = \mu(X)$ so that $T$ is ergodic. \hfill \diamond

Theorem 6.13
A tower transformation is ergodic if and only if the base transformation is.

Proof:
Let $((X^*, B^*, \mu^*), T^*)$ be the tower space extension of $((X, B, \mu), T)$ of either an infinite or finite number of levels. Let the tower be made up of the base space $X = X_0$ and as usual extended with $X_1, X_2, \ldots$ as $\tau_0, \tau_1, \ldots$ images of $A_0, A_1, \ldots$ which are measurable subsets of $X_0, X_1, \ldots$

Suppose that $(X, B, \mu, T)$ is ergodic and that $B$ is $T^*$-invariant, $\mu^*(B) > 0$
and that $B \in B^*$. Then by Lemma’s 5.6 and 5.7 $B$ is $T$-invariant and $\mu(B) > 0$ which implies that $\mu(B \cap X) = \mu(X)$ and in fact by Theorem 6.5, $\mu(B \cap C) = \mu(C)$ for all $C \in B$.

Then, if $D \in B^*$, let $D_0 = D \cap X_0$, $D_1 = D \cap X_1$, ..., $D_n = D \cap X_n$, .... Then $\{D_n\}_n$ is a disjoint measurable sequence and therefore

$$
\mu^*(B \cap D) = \mu^* \left( \bigcup_{i=0}^{\alpha} (B \cap D_i) \right)
$$

$$
= \sum_{i=0}^{\alpha} \mu^*(B \cap D_i)
$$

where $\alpha$ is the number of levels in $X^*$ and may be $\infty$. Also for each $D_i$, $T^{-k}D_i = \tau_{k-1}^{-1} \beta_k^{-1} \in B$ so that $T^{-k}(B \cap D_i) \in B$ and using Theorem 6.5

$$
\mu(T^{-k}(B \cap D_i)) = \mu(T^{-k}B \cap T^{-k}D_i)
$$

$$
= \mu(B \cap T^{-k}D_i)
$$

$$
= \mu(T^{-k}D_i)
$$

$$
= \mu(D_i).
$$

Hence

$$
\mu^*(B \cap D) = \sum_{i=0}^{\alpha} \mu^*(B \cap D_i)
$$

$$
= \sum_{i=1}^{\alpha} \mu^*(D_i)
$$

$$
= \mu^*(D).
$$

Theorem 6.5 then gives us that $T^*$ must be ergodic.

Conversely, suppose that $T^*$ is ergodic. Then as $B \subseteq B^*$, if $B \subseteq X$ is $T$-invariant, clearly for all $x \in X$, $(X \cap O_{T^*}(x)) \subseteq O_T(x)$ and hence

$$
\left( X \cap \bigcup_{x \in B} O_{T^*}(x) \right) \subseteq \bigcup_{x \in B} O_T(x)
$$
and thus

\[(X \cap \bigcup \{D : D \in O_{T^*}(B)\}) \subseteq \bigcup \{D : D \in O_T(B)\}.\]

Therefore using Theorem 6.5 and the fact that \(\bigcup \{D : D \in O_{T^*}(B)\}\) is \(T^*\)-invariant

\[
\mu(\bigcup \{D : D \in O_T(B)\}) \geq \mu^*(\bigcup \{D : D \in O_T(B)\}) \\
\geq \mu^*(\bigcup \{D : D \in O_{T^*}(B)\} \cap X) \\
= \mu^*(X) \\
= \mu(X).
\]

Then by Theorem 3.1, \(T\) is ergodic. \(\diamondsuit\)

Specifically considering the effect that towers may have on the property of preserving outer measure, one consequence that may possibly have arisen from the outer measure preserving requirement is that we may have run into the event of finding that tower extensions or restrictions do not preserve the property of being outer measure preserving. Since we have only been able to prove that \(T\) generates the Carathéodory definition if it is outer measure preserving rather than also showing that the outer measure preserving property is required, this would have demonstrated that our method of proof was not adequate for this problem. Therefore, it is of interest to show that \(T\) is outer measure preserving if and only if \(\tilde{T}\) is.

To prove this important property we reintroduce the notion of looking at tower transformations as stacks. Recall that this means that, instead of seeing the tower as a sequence of sets of decreasing size, one resting on top of the previous one, we can picture the tower as a sequence of disjoint stacks of sets of the same size. Each stack is a finite collection of ordered sets, each
placed on top of the one before it. The set at the bottom of each stack, the base of that stack, denoted \( S_{i_0} \), is the base or ground (0-th) floor of the \( i \)-th stack. The set at the top of each stack, the ceiling, is denoted by \( S_{i_{n_i - 1}} \), is the ceiling or top \(((n_i - 1)\)-th) floor of the \( i \)-th stack where \( n_i \) is the number of piled sets in the \( i \)th stack. Recall also that we denote by \( S_{i_k} \) the \( k \)-th level of the stack (where we are counting up from the bottom with the base being counted as 0). Recall that we join the two ideas as follows. For a usual tower extension we consider stack \( i \) as the unique stack that has a base

\[
S_{i_0} = \{ x \in X : \bar{T}^{i+1}(x) \in X, \bar{T}^m(x) \not\in X, \ 1 \leq m < i + 1 \}.
\]

We can then order the sets so that the \( i \)-th stack has \( i + 1 \) sets piled in the stack, that is \( n_i = i + 1 \). In this case clearly \( S_{i_{i-1}} = \bigcup_{j=0}^{i} \tau_{i-j}(S_{i_0}) \) and

\[
B_i = \bigcup_{n=i}^{\alpha} \bigcup_{j=0}^{i} \tau_{i-j}(S_{n_0})
\]

for each level \( B_i \) in our original method of constructing towers, where \( \alpha \) is the number of levels in the tower, which may be finite or countably infinite.

We are now in a position to prove the final theorem for this chapter.

**Theorem 6.14**

Let \((X, \mathcal{B}, \mu)\) be a measure space and \((\bar{X}, \bar{\mathcal{B}}, \bar{\mu})\) be a tower extension on \((X, \mathcal{B}, \mu)\) (we do not distinguish whether it is finite or infinite). We let \( T \) be a transformation on \( X \) and \( \bar{T} \) be an associated transformation on \( \bar{X} \). In this case \( T \) is \( \mu_* \)-measure preserving if and only if \( \bar{T} \) is \( \bar{\mu}_* \)-measure preserving.

**Proof:**

Suppose that \( T \) is \( \mu_* \)-measure preserving. Note also that since \( \tau_n \) is a bijection for each \( n \in \mathbb{N} \), \( \tau_n \) and \( \tau_n^{-1} \) are outer measure preserving for each \( n \in \mathbb{N} \).
Now let $D \subset \tilde{X}$ so that

$$D = \bigcup_{n=1}^{\alpha} \bigcup_{i=0}^{n-1} (D \cap S_{n_i})$$

and

$$\tilde{T}^{-1}(D) = T^{-1}D \cap S_{10} \cup \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} \tau_i(T^{-1}(D \cap S_{n_0})) \right) \cup \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} \tau_i^{-1}(D \cap S_{n_{i+1}}) \right)$$

Hence, since $S_{nm} \cap S_{ij} = \emptyset$ whenever $n \neq i$ or $m \neq j$, we can use Lemma 6.1 to get

$$\tilde{\mu}_*(\tilde{T}^{-1}D) = \tilde{\mu}_*(T^{-1}(D \cap S_{10})) + \sum_{n=2}^{\alpha} \tilde{\mu}_*(\bigcup_{i=0}^{n-2} \tau_i(T^{-1}(D \cap S_{n_0})))$$

$$+ \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \tilde{\mu}_*(\tau_i^{-1}(D \cap S_{n_{i+1}}))$$

$$= \tilde{\mu}_*(T^{-1}(D \cap S_{10})) + \sum_{n=2}^{\alpha} \tilde{\mu}_*(T^{-1}(D \cap S_{n_0}))$$

$$+ \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \tilde{\mu}_*((D \cap S_{n_{i+1}}))$$

$$= \tilde{\mu}_*(D \cap S_{10}) + \sum_{n=2}^{\alpha} \tilde{\mu}_*(D \cap S_{n_0}) + \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \tilde{\mu}_*((D \cap S_{n_{i+1}}))$$

$$= \tilde{\mu}_*(D \cap X) + \sum_{n=2}^{\alpha} \sum_{i=0}^{n-2} \tilde{\mu}_*((D \cap S_{n_{i+1}}))$$

$$= \tilde{\mu}_* \left( \bigcup_{n=1}^{\alpha} (D \cap S_{n_0}) \right) + \tilde{\mu}_* \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} (D \cap S_{n_{i+1}}) \right)$$

$$= \tilde{\mu}_* \left( \bigcup_{n=1}^{\alpha} (D \cap S_{n_0}) \right) \cup \left( \bigcup_{n=2}^{\alpha} \bigcup_{i=0}^{n-2} (D \cap S_{n_{i+1}}) \right)$$
\[
\begin{align*}
&= \tilde{\mu}_* \left( \bigcup_{n=1}^{\alpha} \bigcup_{i=0}^{n-1} (D \cap S_{n_i}) \right) \\
&= \tilde{\mu}_*(D).
\end{align*}
\]

Thus \( \tilde{T} \) is \( \tilde{\mu}_* \)-measure preserving.

Conversely, suppose that \( \tilde{T} \) is \( \tilde{\mu}_* \)-measure preserving. Then we need to show that \( T \) is \( \mu_* \)-measure preserving. Let \( D \subset X \). We again note that \( \tilde{T}^{-1}D \) is a subset of the ceiling of the tower. That is

\[
\tilde{T}^{-1}D \subset \bigcup_{n=0}^{\alpha-1} S_{n_n},
\]

so that

\[
\tilde{T}^{-1}D = \bigcup_{n=1}^{\alpha} (D \cap S_{n_n}).
\]

Note that we could also have used \( S_{n_n-1} \). However, by our earlier construction of the \( S_n \) we know that \( i_1 = i + 1 \) for each \( i = 0, 1, ..., \alpha - 1 \). We now have, again using Lemma 6.1, that

\[
\begin{align*}
\mu_*(T^{-1}(D)) &= \tilde{\mu}_*(T^{-1}(D)) \\
&= \tilde{\mu}_* \left( \bigcup_{n=1}^{\alpha} (T^{-1}D \cap S_{n_0}) \right) \\
&= \sum_{n=1}^{\alpha} \tilde{\mu}_*(D \cap S_{n_0}) \\
&= \sum_{n=1}^{\alpha} \tilde{\mu}_*(\varrho_{i=0}^{n-1} \tau_i(D \cap S_{n_0})) \\
&= \sum_{n=1}^{\alpha} \tilde{\mu}_*(D \cap S_{n_n}) \\
&= \tilde{\mu}_* \left( \bigcup_{n=1}^{\alpha} (D \cap S_{n_n}) \right) \\
&= \tilde{\mu}_*(\tilde{T}^{-1}D) \\
&= \tilde{\mu}_*(D)
\end{align*}
\]
= \mu_*(D).

Which, by definition gives us that \( T \) is \( \mu_* \)-measure preserving which completes the proof.

\[\square\]

### 6.4 Notes

Apart from Lemma 6.1 and Theorem 6.7, the results presented in this chapter are all original. Lemma 6.1 is due to Nillsen [31]. An extended discussion and proof of Theorem 6.7 can be found in Federer [11]. Further discussion on images of Borel sets can be found in Freiwald [12], Holick’y [19] and Yoshio [48]. A good source for further reading on nowhere dense sets, and category in measure in general is Oxtoby [34].
7 Non-measurable Sets

The previous three chapters have discussed quite extensively transformations that generate Carathéodory’s definition in the sense that was introduced in the discussion of Chapter 2. However, we also noted in Chapter 2 that in order to really be able to be considered as generating Carathéodory’s definition a transformation must do more than simply show that measurable sets must satisfy Carathéodory’s condition. That is, we noted that we must be able to show that for any transformation we wish to consider as generating Carathéodory’s definition, the transformation must also distinguish between the ‘well behaved’ and ‘poorly behaved’ sets. That is, we should be able to distinguish between what are afterwards defined as measurable and non-measurable sets.

In order to demonstrate that a transformation, \( T \), does indeed make such a distinction, we must, as noted in Chapter 2, find a non-measurable set, \( A \), that is \( T \)-invariant and is such that for each ‘subinterval’ (or as defined later, any element of a standard collection of sets), \( J \), of the measure space, \( (X, \mathcal{B}, \mu) \), on which \( T \) is defined

\[
\mu_+(A \cap J) + \mu_+(A^c \cap J) = 2\mu(J).
\]  

(30)

Although we have not been able to show that each transformation known to generate Carathéodory’s definition has a non-measurable set with the required properties associated with it, we do have some fairly general results about non-measurable transformation invariant sets. We present these at the beginning of this chapter. We then go on to show that each of the specific one to one examples considered, (irrational rotations, Kakutani type I transformations and Chacon’s \( \xi \)) do have non-measurable sets with the required properties associated with them.
Further, we show that, as with generating Carathéodory's definition, any
tower extension of a primitive system will have a non-measurable set with
the required properties if and only if the primitive transformation does.

The discussion in this chapter, complementing the discussion on generating
Carathéodory's definition in Chapters 4, 5 and 6 about the non-measurable
sets necessary to make any justification of the statement that transforma-
tions generate and motivate Carathéodory's definition concludes the thesis.
Except when otherwise mentioned, the results in this chapter are original,
although some of the work (as specified in the notes) is incorporated into
[27]. However, the work as a whole represents new results in this area.

7.1 General results on non-measurable sets

Just as with the generation of Carathéodory's definition, we find that the
concept of orbits is vital in considering transformation invariant sets. For
this reason we take the time to recall the definitions of orbits that will be
required in this chapter.

Firstly, suppose that $T$ is a transformation defined on the measure space
$(X, B, \mu)$, and then recall from Definition 4.6 that the $T$-orbit of a set $D \subseteq X$
is defined as

$$O_T(D) = \{T^n(D) : n \in \mathbb{Z}\}.$$ 

Also, recall that if there is a $j \in \mathbb{Z}$ such that $T^j(x)$ is not defined for some
$x \in X$ then the $T$-orbit of $x$ is

$$O_T(x) = \{T^n(x) : n \in \mathbb{Z}, n > j\}$$
but that if such a \( j \) does not exist, the \( T \)-orbit of a point \( x \in X \) is defined as

\[
O_T(x) = \{T^n(x) : n \in \mathbb{Z}\}.
\]

Secondly, suppose that \( \Psi \) is a family of transformations defined on \((X, B, \mu)\).
Then, from Definition 6.7 we have the \( \Psi \)-orbit of \( D \subseteq X \) defined by

\[
O_\Psi(D) = \{\psi_1^{s_1} \circ \psi_2^{s_2} \circ \ldots \circ \psi_r^{s_r}(D) : r \in \mathbb{N}, s_1, \ldots, s_r \in \mathbb{Z}, \psi_1, \ldots, \psi_r \in \Psi\}.
\]

We also have the \( \Psi \)-orbit of a point \( x \in X \) defined by

\[
O_\Psi(x) = \{\psi_1^{s_1} \circ \psi_2^{s_2} \circ \ldots \circ \psi_r^{s_r}(x) : r \in \mathbb{N}, s_1, \ldots, s_r \in \mathbb{Z}, \psi_1, \ldots, \psi_r \in \Psi, \psi_1^{s_1} \circ \psi_2^{s_2} \circ \ldots \circ \psi_r^{s_r}(x) \text{ is defined}\},
\]

We now need to consider some further properties of orbits and how different orbits fit together in \( X \). Firstly, it is clear that if \( \Psi \) is a group of transformations then for any \( r \in \mathbb{N} \), any selection \( \psi_1, \psi_2, \ldots, \psi_r \in \Psi \) and any selection \( s_1, s_2, \ldots, s_r \in \mathbb{Z} \)

\[
\psi_1^{s_1} \circ \psi_2^{s_2} \circ \ldots \circ \psi_r^{s_r} \in \Psi
\]

so that

\[
O_\Psi(x) = \bigcup_{\psi \in \Psi} \{\ldots, \psi^{-1}(x), x, \psi(x), \psi^2(x), \ldots\} = \bigcup_{\psi \in \Psi} O_\psi(x) = \Psi(x).
\]

Secondly, recalling Definition 6.8 of \( \Psi \)-invariant sets, we show that \( O_\psi(x) \) is the smallest \( \Psi \)-invariant set containing \( x \) and similarly \( \bigcup_{C \in O_\psi(D)} C \) is the smallest \( \Psi \)-invariant set containing \( D \). In order to do this we must first show an implication of Definition 6.8

**Proposition 7.1**

Let \( \Psi \) be a family of transformations on a set \( X \). Then, a subset \( D \) of \( X \) is \( \Psi \)-invariant if and only if \( D \) is \( \psi \)-invariant for each \( \psi \in \Psi \).
Proof:
Suppose that $D$ is $\Psi$-invariant. Then for each $\psi \in \Psi$, $\psi^{-1}(D) \in O_\Psi^{-1}(D)$ and hence as $D$ is $\Psi$-invariant $\psi^{-1}(D) = D$ so that $D$ is $\psi$-invariant.

Conversely, suppose that $D$ is $\psi$-invariant for each $\psi \in \Psi$. Let $B \in O_\Psi^{-1}(D)$ so that there are $r \in \mathbb{N}, s_1, s_2, \ldots, s_r \in \mathbb{N}$ and $\psi_1, \psi_2, \ldots, \psi_r \in \Psi$ such that

$$B = \psi_1^{-s_1} \circ \psi_2^{-s_2} \circ \cdots \circ \psi_r^{-s_r}(D).$$

Repetitively using the fact that for each $\psi \in \{\psi_1, \psi_2, \ldots, \psi_r\}$, $D$ is $\psi$-invariant we have that

$$B = \psi_1^{-s_1} \circ \psi_2^{-s_2} \circ \cdots \circ \psi_r^{-s_r}(D)$$
$$= \psi_1^{-s_1} \circ \psi_2^{-s_2} \circ \cdots \circ \psi_r^{-s_r+1}(D)$$
$$\vdots$$
$$= \psi_1^{-1}(D)$$
$$= D.$$

From Definition 6.8 it then follows that $D$ is $\Psi$-invariant. \(\boxdot\)

Proposition 7.2

Let $X$ be a set and $\Psi$ be a family of transformations defined on $X$. Then $O_\Psi(x)$ is the smallest $\Psi$-invariant set containing $x$ and $\bigcup_{C \in O_\Psi(D)} C$ is the smallest $\Psi$-invariant set containing $D$.

Proof:

Suppose that $A$ is a $\Psi$-invariant set containing $x$. Now, clearly for each $\psi \in \Psi$, $\psi^{-1}(x) \in A$ as $A$ is $\Psi$ and hence $\psi$-invariant. Also, if $\psi(x) \notin A$, $x \notin \psi^{-1}(A)$ which would contradict $A$ being $\psi$-invariant. By using this argument repetitively it follows that if $r \in \mathbb{N}$, $\psi_1, \psi_2, \ldots, \psi_r \in \Psi$ and $s_1, s_2, \ldots, s_r \in \mathbb{Z}$,

$$\psi_1^{s_1} \circ \psi_2^{s_2} \circ \cdots \circ \psi_r^{s_r}(x) \in A$$
so that each element of \(O_\Psi(x)\) is in \(A\) and hence \(O_\Psi(x) \subseteq A\). Hence any \(\Psi\)-invariant set containing \(x\) contains \(O_\Psi(x)\).

We now need to show that \(O_\Psi(x)\) is \(\Psi\)-invariant. As it is clear that for each element of \(O_\Psi(x)\), say \(y = \psi_1^{s_1} \circ \psi_2^{s_2} \circ \ldots \circ \psi_r^{s_r}(x)\), and any \(\psi \in \Psi\), \(\psi^{-1} y \in O_\Psi(x)\) so that \(\psi^{-1}(O_\Psi(x)) \subseteq O_\Psi(x)\) and similarly for any \(y \in O_\Psi(x)\), \(\psi(y) \in O_\Psi(x)\) and hence \(y = \psi^{-1}(\psi(y)) \in \psi^{-1}(O_\Psi(x))\) so that \(O_\Psi(x) \subseteq \psi^{-1}(O_\Psi(x))\). Therefore for any \(\psi \in \Psi\), \(\psi^{-1}(O_\Psi(x)) = O_\Psi(x)\) and hence \(O_\Psi(x)\) is \(\Psi\)-invariant.

Now suppose that \(A\) is a \(\Psi\)-invariant set containing \(D\). Clearly for any \(\psi \in \Psi\), \(\psi^{-1}(D) \subseteq \psi^{-1}(A) = A\). Also if \(\psi(D) - A \neq \emptyset\) then there is an \(x \in D \subseteq A\) such that \(\psi(x) \notin A\) which implies that \(x \notin \psi^{-1}(A)\) contradicting the \(\Psi\)-invariance of \(A\). Therefore \(\psi(D) \subseteq A\). By using this argument repetitively, we know that for any \(r \in \mathbb{N}\), \(\psi_1, \psi_2, \ldots, \psi_r\), and \(s_1, s_2, \ldots, s_r \in \mathbb{Z}\)

\[
\psi_1^{s_1} \circ \psi_2^{s_2} \circ \ldots \circ \psi_r^{s_r}(D) \subseteq A
\]

so that any \(\Psi\)-invariant set containing \(D\) also contains each element of \(O_\Psi(D)\) and hence

\[
\bigcup_{C \in O_\Psi(D)} C.
\]

Finally we need now only show that \(\bigcup_{C \in O_\Psi(D)} C\) is \(\Psi\)-invariant. Clearly, for any \(x \in \bigcup_{C \in O_\Psi(D)} C\), \(x \in B\) for some \(B \in O_\Psi(D)\). Also, by the definition of \(O_\Psi(D)\), it is clear that for any \(\psi \in \Psi\), \(\psi^{-1}(B) \in O_\Psi(D)\) and \(\psi(B) \in O_\Psi(D)\). Therefore

\[
\psi^{-1}(x) \subseteq \psi^{-1}(B) \subseteq \bigcup_{C \in O_\Psi(D)} C.
\]

Also \(\psi(x) \in \psi(B) \subseteq \bigcup_{C \in O_\Psi(D)} C\) and hence

\[
x \in \psi^{-1}\left(\bigcup_{C \in O_\Psi(D)} C\right).
\]
It follows that
\[
\psi^{-1}\left( \bigcup_{C \in O_{\Psi}(D)} C \right) \subseteq \bigcup_{C \in O_{\Psi}(D)} C
\]
and that
\[
\bigcup_{C \in O_{\Psi}(D)} C \subseteq \psi^{-1}\left( \bigcup_{C \in O_{\Psi}(D)} C \right)
\]
and hence
\[
\psi^{-1}\left( \bigcup_{C \in O_{\Psi}(D)} C \right) = \bigcup_{C \in O_{\Psi}(D)} C.
\]
As this is true for any \( \psi \in \Psi \) we have that \( \bigcup_{C \in O_{\Psi}(D)} C \) is \( \Psi \)-invariant. 

Note that by considering \( \Psi = \{\psi\} \) Proposition 7.2 also gives us that the smallest \( \psi \)-invariant sets \( A \) and \( B \) containing an element \( x \in X \) and a subset \( D \) of \( X \) respectively are \( O_{\psi}(x) \) and \( O_{\psi}(D) \) respectively.

The final property specifically concerning orbits that we consider is an important technical lemma to be used in the construction of non-measurable sets.

**Lemma 7.1**

Let \( \Psi \) be a family of one-to-one transformations defined on a set \( X \). Then the following hold.

1. Two \( \Psi \)-orbits in \( X \) are either identical or disjoint.

2. A subset \( A \) of \( X \) is \( \Psi \)-invariant if and only if \( A \) is a union of \( \Psi \)-orbits.
   
   That is, \( A \) is \( \Psi \)-invariant if and only if there is a subset \( W \) of \( X \) such that
   \[
   A = \bigcup_{w \in W} O_{\Psi}(w).
   \]
Proof:
To prove (1) we show that if two \( \Psi \)-orbits have an intersection then they are identical. Suppose that \( x, y \in X \) and that

\[
O_\Psi(x) \cap O_\Psi(y) \neq \emptyset,
\]
so that there is an element, \( z \) of \( X \) such that \( z \in O_\Psi(x) \cap O_\Psi(y) \). We therefore have that there are \( n, m \in \mathbb{N} \), collections \( s_1, s_2, ..., s_n \) and \( t_1, t_2, ..., t_m \) in \( \mathbb{Z} \) as well as collections \( \psi_1, \psi_2, ..., \psi_n \) and \( \phi_1, \phi_2, ..., \phi_m \) in \( \Psi \) such that

\[
z = \psi_1^{s_1} \circ \psi_2^{s_2} \circ \cdots \circ \psi_n^{s_n}(x),
\]

\[
z = \phi_1^{t_1} \circ \phi_2^{t_2} \circ \cdots \circ \phi_m^{t_m}(y)
\]

and hence that

\[
y = \phi_m^{-t_m} \circ \cdots \circ \phi_1^{-t_1} \circ \psi_1^{s_1} \circ \cdots \circ \psi_n^{s_n}(x)
\]

and

\[
x = \psi_n^{-s_n} \circ \cdots \circ \psi_1^{-s_1} \circ \phi_1^{s_1} \circ \cdots \circ \phi_m^{-t_m}(y).
\]

Thus \( y \in O_\Psi(x) \) and \( x \in O_\Psi(y) \).

Now, let \( w \in O_\Psi(x) \) so that there is a \( r \in \mathbb{N} \) a collection \( u_1, u_2, ..., u_r \in \mathbb{Z} \) and a collection \( \rho_1, \rho_2, ..., \rho_r \in \Psi \) such that

\[
w = \rho_1^{u_1} \circ \rho_2^{u_2} \circ \cdots \circ \rho_r^{u_r}(x),
\]

and hence

\[
y = \phi_m^{-t_m} \circ \cdots \circ \phi_1^{-t_1} \circ \psi_1^{s_1} \circ \cdots \circ \psi_n^{s_n} \circ \rho_r^{-u_r} \circ \cdots \circ \rho_1^{-u_1}(w)
\]

so that \( w \in O_\Psi(y) \). Therefore \( O_\Psi(x) \subseteq O_\Psi(y) \). Similarly we can show that \( O_\Psi(y) \subseteq O_\Psi(x) \) and therefore \( O_\Psi(x) = O_\Psi(y) \) which proves (1).
To prove (2), firstly, if \( A = \bigcup_{w \in W} O_\Psi(w) \) for some \( W \subset X \) and \( \psi \in \Psi \), we have

\[
\psi^{-1}(A) = \psi^{-1} \left( \bigcup_{w \in W} O_\Psi(w) \right) = \bigcup_{w \in W} \psi^{-1}(O_\Psi(w)) = \bigcup_{w \in W} O_\Psi(w) = A.
\]

Thus \( A \) is \( \psi \)-invariant for any \( \psi \in \Psi \) and thus by Proposition 7.1 \( A \) is \( \Psi \)-invariant.

Conversely suppose that \( A \) is \( \Psi \)-invariant, Then, using Proposition 7.2 we know that for each \( a \in A \), \( O_\Psi(a) \subset A \) and hence that

\[
\bigcup_{a \in A} O_\Psi(a) \subset A.
\]

Clearly

\[
A = \bigcup_{a \in A} a \subset \bigcup_{a \in A} O_\Psi(a)
\]

and hence

\[
A = \bigcup_{a \in A} O_\Psi(a),
\]

which proves (2) and thus completes the proof of the theorem.

We now define and consider some of the properties of transversals which are instrumental in the construction of non-measurable sets.

**Note:** For this discussion, unless otherwise specified, we will use \( \iota \) to denote the identity transformation of a group of transformations.
Definition 7.1
Let $G$ be a group of transformations on a set $X$. A subset $A$ of $X$ is called a partial $G$-transversal if $x, y \in A$ and $G(x) = G(y)$ implies that $x = y$. That is, any two distinct points of $A$ belong to distinct $G$-orbits.
A subset $A$ of $X$ is called a (complete) $G$-transversal if every $G$-orbit contains a necessarily unique point in $A$. That is, a subset $A$ of $X$ is a $G$-transversal if it is a partial $G$-transversal such that

$$X = \bigcup_{a \in A} G(a).$$

The properties of transversals that we are interested in are proven in the following result. Note especially the third point in the result which actually, assuming the axiom of choice, provides the existence of the type of non-measurable set that we will considering. The actual approach used is based on a well known method for constructing non-measurable sets but is more general.

Proposition 7.3
Let $(X, \mathcal{B}, \mu)$ be a measure space such that $0 < \mu(X) < \infty$. Let $G$ be a group of $\mu$-measure preserving transformations on $X$ such that if $g \in G$ and if $x \in X$ are such that $g(x) = x$ (that is $g$ has a fixed point) then $g = \iota$. Then the following hold

1. If $A$ is a $G$-transversal and $g \in G$, $g(A)$ is a $G$-transversal. Moreover, if $g, h \in G$ and $g \neq h$, then

$$g(A) \cap h(A) = \emptyset.$$ 

Thus $g(A) \mapsto g$ is a one-to-one function from $G$ into the set of all $G$-transversals, and $G$ acts as a group of permutations on the set

$$\{ h(A) : h \in G \}.$$
2. If \( A \) is a \( G \)-transversal,
\[
X = \bigcup_{g \in G} g(A).
\]

3. If \( G \) is countable and infinite and \( A \) is a \( G \)-transversal, then \( A \not\in B \).

Proof:

(1)
Let \( A \) be a \( G \)-transversal and \( g \in G \). Suppose that \( x, y \in g(A) \) are in the same \( G \)-orbit. Then there are \( a, b \in A \) and \( h \in G \) such that \( g(a) = x \), \( g(b) = y \) and \( h(x) = y \). It follows that
\[
hg(a) = h(x) = y = g(b),
\]
which gives \( g^{-1}hg(a) = b \) showing that \( a \) and \( b \) are in the same \( G \)-orbit. As \( A \) is a \( G \)-transversal, it follows that \( a = b \) and hence that
\[
x = g(a) = g(b) = y,
\]
so that \( g(A) \) is a partial \( G \)-transversal. To show that \( g(A) \) is a \( G \)-transversal we note that for any \( z \in X \) there is an \( a \in A \cap G(z) \) and hence \( g(a) \in G(z) \). Thus every \( G \)-orbit contains a point in \( g(A) \) so that \( g(A) \) is a \( G \)-transversal.

Now note that if \( g, h \in G \) are such that \( g(A) \cap h(A) \neq \emptyset \) there are \( a, b \in A \) such that \( g(a) = h(b) \). However, this implies that \( a = g^{-1}h(b) \) so that \( a, b \in G(b) \). As \( g(A) \) is a \( G \)-transversal, this implies that \( a = b \) and hence \( g^{-1}h(b) = b \). From the assumptions on the elements of \( G \) this implies that \( g^{-1}h = \iota \) and hence that \( h = g \). Therefore we have
\[
g, h \in G, g \neq h \Rightarrow g(A) \cap h(A) = \emptyset.
\]

It follows that \( g(A) \rightarrow g \) for each \( g \in G \) is one-to-one. Since this mapping is one-to-one and \( G \) is a group it also follows that \( G \) acts as a group of permutations on the set \( \{g(A) : a \in G\} \).
(2)

As \( A \) is a \( G \)-transversal we have

\[
X = \bigcup_{a \in A} G(a)
\]

\[
= \bigcup_{a \in A} \left( \bigcup_{g \in G} g(a) \right)
\]

\[
= \bigcup_{g \in G} \left( \bigcup_{a \in A} g(a) \right)
\]

\[
= \bigcup_{g \in G} g(A),
\]

which proves point (2).

(3)

Assume that \( G \) is countably infinite and that \( A \) is a \( G \)-transversal with \( A \in \mathcal{B} \). Then \( g(A) \in \mathcal{B} \) for all \( g \in G \). As \( \mu \) is a measure on \( \mathcal{B} \) it is \( \sigma \)-additive on \( \mathcal{B} \) and it follows that

\[
\mu(X) = \mu \left( \bigcup_{g \in G} g(A) \right)
\]

\[
= \sum_{g \in G} \mu(g(A))
\]

\[
= \sum_{g \in G} \mu(A).
\]

If \( \mu(A) = 0 \) we would then have that \( \mu(X) = 0 \), a contradiction. Alternatively if \( \mu(A) > 0 \) we would have that \( \mu(X) = \infty \) which is also a contradiction. It therefore follows that \( A \not\in \mathcal{B} \) which proves point (3). \( \diamond \)

An important point following from the above is that if \( A \) is a \( G \)-transversal we then have that

\[
g \in G, g \neq \iota \Rightarrow g(A) \cap A = \emptyset.
\]
In this way a $G$-transversal can be considered to be 'anti-invariant' under the action of elements of $G$.

The following theorem looks at some general conditions on a family of transformations, $\Psi$ that will ensure that we can find $\Psi$-invariant non-measurable sets. We later look at further conditions necessary to ensure that such non-measurable sets satisfy all of the conditions we would like them to satisfy as discussed during the preamble to this Chapter.

It is in this theorem that the axiom of choice is used, or rather, Zorn's Lemma, which is equivalent. We shall therefore, for convenience, state Zorn's Lemma for reference.

**Lemma 7.2 (Zorn's Lemma)**

*If $S$ is any nonempty partially ordered set in which every chain has an upper bound, then $S$ has a maximal element. That is, should $x$ be the maximal element, $y \in S$ and $y \geq x$ implies $x \geq y$.***

**Theorem 7.1**

Let $(X, \mathcal{B}, \mu)$ be a measure space such that $0 < \mu(X) < \infty$. Let $G$ be a group of $\mu$-measure preserving transformations on $X$ with the property that if $g \in G$ is such that there is an $x \in X$ for which $g(x) = x$ then $g = \iota$. Let $\Psi$ be a family of one-to-one $\mu$-measure preserving transformations on $X$. Consider the following conditions.

(a) For each $x \in X$, The $\Psi$-orbit $O_\Psi(x)$ is a partial $G$-transversal.
(b) If $g \in G$ and $x \in X$, there is $h \in G$ such that $g(O_\Psi(x)) \subseteq O_\Psi(h(x))$.
(c) $G$ is infinite and countable.
(d) For all \( g \in G \), \( g^{-1} \Psi g \subseteq \Psi \).

Then if (a) and (b) hold there is a \( \Psi \)-invariant \( G \)-transversal. If (a), (b) and (c) hold, there is a \( \Psi \)-invariant \( G \)-transversal which is not in \( \mathcal{B} \), so that in this case there is a non-measurable \( \Psi \)-invariant subset of \( X \). If all conditions hold then \( X \) is the disjoint union of a countably infinite family of non-measurable \( \Psi \)-invariant \( G \)-transversals.

**Proof:**

Assume that (a) and (b) hold. Let \( \mathcal{Y} \) denote the family of all partial \( G \)-transversals in \( X \) which are \( \Psi \)-invariant. As \( \emptyset \in \mathcal{Y} \) we know that \( \mathcal{Y} \neq \emptyset \).

Now, let \( \{X_\alpha\}_{\alpha \in I} \) be a chain in \( \mathcal{Y} \) ordered by inclusion. That is, for \( \alpha, \beta \in \mathcal{Y} \), either \( X_\alpha \subseteq X_\beta \) or \( X_\beta \subseteq X_\alpha \). Let

\[
Z = \bigcup_{\alpha \in I} X_\alpha.
\]

Then, using the fact that each \( X_\alpha \) is \( \Psi \)-invariant, for \( \psi \in \Psi \)

\[
\psi^{-1}(Z) = \psi^{-1} \left( \bigcup_{\alpha \in I} X_\alpha \right)
= \bigcup_{\alpha \in I} \psi^{-1}(X_\alpha)
= \bigcup_{\alpha \in I} X_\alpha
= Z.
\]

Thus \( Z \) is \( \Psi \)-invariant by Proposition 7.1 and we now proceed to show that \( Z \) is also a partial \( G \)-transversal.

To this end, let \( x, y \in Z \) be such that \( x, y \) are in the same \( G \)-orbit. Then, as \( \{X_\alpha\} \) is a chain there is \( \beta \in I \) such that \( x, y \in X_\beta \). We must then have \( x = y \) as they are in the same \( G \)-orbit, and because \( X_\beta \) is a partial \( G \)-transversal.
Thus $Z$ is a $\Psi$-invariant $G$-transversal and thus in $\mathcal{Y}$. This shows that Zorn’s Lemma may be applied to $\mathcal{Y}$ and thus we deduce that $\mathcal{Y}$ has a maximal element, say $Y$.

In order to apply the previous results on non-measurability we need $Y$ to be a $G$-transversal, we therefore prove that $Y$ is a $G$-transversal.

Now, $Y \in \mathcal{Y}$ and so is a $\Psi$-invariant partial $G$-transversal. Assume that $Y$ is not a $G$-transversal. Then there is a $z \in X$ such that

$$z \notin \bigcup_{y \in Y} G(y). \quad (32)$$

Put

$$Y' = Y \cup O_\Psi(z).$$

We show that $Y' \in \mathcal{Y}$, which will then give us the result by contradicting $Y$ being a maximal element of $\mathcal{Y}$, as it is clear that $z \in Y'$, $z \notin Y$ and $Y \subseteq Y'$.

If $\psi \in \Psi$ we have that

$$\psi^{-1}(Y') = \psi^{-1}(Y \cup O_\Psi(z))$$

$$= \psi^{-1}(Y) \cup \psi^{-1}(O_\Psi(z))$$

$$= Y \cup O_\Psi(z)$$

$$= Y',$$ \quad (33)

so that $Y'$ is $\Psi$-invariant.

In order to show that $Y'$ is a partial $G$-transversal, let $u, v \in Y'$ and $g \in G$ be such that $g(u) = v$. We need to consider four cases.
CASE I.

This is when \( u, v \in Y \). In this case \( u = v \) as \( Y \) is a partial \( G \)-transversal.

CASE II.

This is when \( u, v \in O_\psi(z) \). Then, by assumption (a), \( O_\psi(z) \) is a partial \( G \)-transversal and hence \( u = v \).

CASE III.

This is when \( u \in Y \) and \( v \in O_\psi(z) \). By assumption (b), there is \( h \in G \) such that \( g^{-1}(O_\psi(z)) \subseteq O_\psi(h(z)) \). Then, as \( g(u) = v \), we have

\[
u = g^{-1}(v) \in g^{-1}(O_\psi(z)) \subseteq O_\psi(h(z)).
\]

It then follows from Lemma 7.1 that \( h(z) \in O_\psi(u) \subseteq Y \) and hence that

\[
z \in h^{-1}(Y) \subseteq G(Y) = \bigcup_{g \in G} g(Y),
\]

which contradicts (32). Thus CASE III cannot occur.

CASE IV.

This is when \( u \in O_\psi(z) \) and \( v \in Y \). This is the same case as CASE III. with the roles of \( u \) and \( v \) reversed and so cannot occur by the same argument as that for CASE III.

From these cases we deduce that \( u = v \) in Cases I and II and that Cases III and IV cannot occur so that \( u = v \) and hence \( Y' \) is a \( \Psi \)-invariant partial \( G \)-transversal and hence \( Y' \in \mathcal{Y} \). As \( Y \) is a proper subset of \( Y' \) this contradicts the choice of \( Y \) as a maximal element in \( \mathcal{Y} \) and hence \( Y \) must be a \( \Psi \)-invariant \( G \)-transversal.
If we further assume that \( (c) \) holds then by Proposition 7.3 (3) we know that \( Y \) cannot be in \( B \).

Now suppose that all of the assumptions hold and let \( Y \) be a \( \Psi \)-invariant \( G \)-transversal which, necessarily, is not in \( B \). By Proposition 7.3 (1) we also know that \( g(Y) \) is a \( G \)-transversal for any \( g \in G \). We now show that for any \( g \in G \), \( g(Y) \) is \( \Psi \)-invariant.

If \( \psi \in \Psi \) and \( g \in G \), by assumption \( (d) \) there is a \( \phi \in \Psi \) such that

\[
g^{-1}\psi g = \phi,
\]

so that

\[
\psi = g\phi g^{-1},
\]

and hence

\[
\psi^{-1} = g\phi^{-1}g^{-1}.
\]

Then, using the fact that \( Y \) is \( \Psi \)-invariant,

\[
\psi^{-1}g(Y) = g\phi^{-1}g^{-1}g(Y)
= g\phi^{-1}(Y)
= g(Y),
\]

so that \( g(Y) \) is \( \Psi \)-invariant. Since

\[
X = \bigcup_{g \in G} g(Y)
\]

it now follows from Proposition 7.3 (2) that this union expresses \( X \) as the disjoint union of a countably infinite set of \( \Psi \)-invariant \( G \)-transversals.  

\[\diamondsuit\]

The above theorem shows that the construction of a non-measurable set
involves identifying a set that is $\Psi$-invariant but, in the sense discussed following Proposition 7.3, $G$-'anti-invariant'. Another point that should be considered arises when we allow for the condition that elements of $\Psi$ and $G$ commute. In this case assumption $(d)$ in Theorem 7.1 that $g^{-1}\Psi g \subseteq \Psi$ for each $g \in G$ automatically follows. We also note that the condition that for each $g \in G$ and each $x \in X$ there is a $h \in G$ such that $g(O_x(x)) \subseteq O_x(h(x))$ will imply that $g(O_x(x)) = O_x(h(x))$.

As the elements in the groups $G$ and the families $\Psi$ chosen to demonstrate the existence of non-measurable $T$-invariant sets with the desired properties for the specific transformations $T$ considered earlier in the thesis do commute, we consider the effects of these simplifications in the following proposition. This proposition concludes this section and we then consider the outer measure of the $\Psi$-invariant non-measurable sets that we have constructed.

**Proposition 7.4**

Let $(X, \mathcal{B})$ be a measurable space. Let $G$ be a group of $\mathcal{B}$-measurable transformations on $X$ with the property that if $g \in G$ is such that there is an $x \in X$ with $g(x) = x$, then $g = \iota$. Let $\Psi$ be a family of one to one transformations on $X$. Assume that whenever $g \in G$ and $x \in X$, there is a $h \in G$ such that $g(O_x(x)) \subseteq O_x(h(x))$. Then the following conditions hold.

(a) For any $g \in G$ and $x \in X$, there is a $h \in G$ such that

$$g(O_x(x)) = O_x(h(x)),$$

and $G$ acts as a group of permutations on the set of all $\Psi$-orbits of single points.

(b) Assume that for each $x \in X$, the $\Psi$-orbit $O_x(x)$ is a partial
G - transversal. Then if \( g(O_{\Psi}(x)) = O_{\Psi}(x) \), \( g = \iota \). In this case \( G \) acts as a group of permutations on the set of all \( \Psi \)-orbits but with no non-trivial fixed points.

**Proof:**

Let \( g \in G \) and \( x \in X \). Then by our assumptions there are \( h,k \in G \) such that

\[
g(O_{\Psi}(x)) \subseteq O_{\Psi}(h(x)) \text{ and } g^{-1}(O_{\Psi}(h(x))) \subseteq O_{\Psi}(k(x)).
\]

So we have

\[
O_{\Psi}(x) \subseteq g^{-1}(O_{\Psi}(h(x))) \subseteq O_{\Psi}(k(x)).
\]

However, by Lemma 7.1, we know that two \( \Psi \)-orbits are either identical or disjoint. It follows that

\[
O_{\Psi}(x) = O_{\Psi}(k(x))
\]

and hence that

\[
O_{\Psi}(x) = g^{-1}(O_{\Psi}(h(x))).
\]

This gives that \( g(O_{\Psi}(x)) = O_{\Psi}(h(x)) \). From this it follows immediately that \( G \) acts as a group of permutations on the set of \( \Psi \)-orbits under the action of \( G \) given by

\[
O_{\Psi}(x) \mapsto g(O_{\Psi}(x)),
\]

for each \( g \in G \).

Now let \( g \in G \) and \( x \in X \) be such that \( g(O_{\Psi}(x)) = O_{\Psi}(x) \). Then there is a \( y \in O_{\Psi}(x) \) such that \( g(y) = x \). It follows that \( y = g^{-1}(x) \) and clearly \( x \in G(x) \). Now since \( x,y \in G(x) \) and we know that \( O_{\Psi}(x) \) is a \( G \)-transversal it follows that \( x = y \). This implies that \( g(x) = x \) and from the assumptions it follows that \( g = \iota \). Hence fixed points of the mapping \( O_{\Psi}(x) \mapsto g(O_{\Psi}(x)) \) can only arise from the identity transformation. \( \diamond \)
7.2 The outer measure of sets invariant under transformations that generate the Carathéodory definition of measurable sets

We now consider the outer measures of the \( \Psi \)-invariant non-measurable sets that we have shown exist in the above theorems. As discussed at the beginning of this Chapter, in order to support the earlier chapters we will need to be able to find (especially when considering \( \Psi = \{T\} \) for some transformation, \( T \), of particular interest) a \( \Psi \)-invariant non-measurable set such that for any element of a collection of standard sets, \( J \),

\[
\mu_*(A \cap J) + \mu_*(A^c \cap J) = 2\mu(J).
\]

We show in this section that, when \( \Psi \) satisfies some additional requirements, the \( \Psi \)-invariant non-measurable sets constructed in the previous section satisfy this condition. We start by proving that with the additional requirements

\[
\mu_*(A \cap X) + \mu_*(A^c \cap X) = 2\mu(X).
\]

This is followed by a second result allowing this condition on outer measure to be satisfied for any element of the \( \sigma \)-algebra of the measure space on which the family \( \Psi \) is defined. This second result clearly covers all elements of a collection of standard sets as, by definition, such sets must all be members of the \( \sigma \)-algebra.

**Theorem 7.2**

Let \((X, B, \mu)\) be a probability space where \( \mu \) is a complete measure. Let \( \Psi \) be a countable family of one-to-one transformations on \( X \) such that for each \( \psi \in \Psi \) and \( Y \in B \), \( \psi(Y), \psi^{-1}(Y) \in B \) and assume that \( \Psi \) is \( \mu \)-measure preserving and ergodic. Let \( A \) be a \( \Psi \)-invariant subset of \( X \) with \( A \notin B \). Then \( \mu_*(A) = 1 \).
Proof:
Let $A$ be a $\Psi$-invariant subset of $X$ with $A \not\subseteq B$. As $A \not\subseteq B$, $\mu_*(A) > 0$. We suppose that $\mu_*(A) < 1$. Let $\varepsilon > 0$ be such that $\mu_*(A) + \varepsilon < 1$. Then there is a $B \in \mathcal{B}$ such that $A \subseteq B$ and $\mu_*(A) \leq \mu(B) < \mu_*(A) + \varepsilon < 1$. Now, for each $r \in \mathbb{N}, s_1, s_2, \ldots, s_r \in \mathbb{Z}$ and $\psi_1, \psi_2, \ldots, \psi_r \in \Psi$ put

$$A(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r) = \{\psi_1^{s_1} \circ \cdots \circ \psi_r^{s_r}(x) : x \in A\}$$

$$B(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r) = \{\psi_1^{s_1} \circ \cdots \circ \psi_r^{s_r}(x) : x \in B\}$$

and

$$X(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r) = \{\psi_1^{s_1} \circ \cdots \circ \psi_r^{s_r}(x) : x \in X\}.$$

Observe that $B(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r)$ and $X(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r)$ are in $B$. Now let

$$y \in A \cap A(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r)^c.$$

If we knew that $s_i \leq 0$ for each $s_i \in \{s_1, \ldots, s_r\}$ then we would know that such a $y$ would not exist as $A$ is $\Psi$-invariant. However, this is not the case so we must take a different approach.

If $y \in X(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r)$ we would have $y = \psi_1^{s_1} \circ \cdots \circ \psi_r^{s_r}(x)$ for some $x \in X$ and we must then have $x = \psi_r^{-s_r} \circ \cdots \circ \psi_1^{-s_1}(y)$ which in turn must then be in $A$ by the invariance of $A$. Therefore

$$y = \psi_1^{s_1} \circ \cdots \circ \psi_r^{s_r}(x) \in A(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r),$$

which is a contradiction. It follows that

$$A \cap A(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r)^c \subseteq X \cap X(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r)^c. \quad (34)$$

As each $\psi \in \Psi$ is one-to-one and $\mu$-invariant we have that for each $D \in \mathcal{B}$,

$$\mu(\psi^{-1}(D)) = \mu(D) \text{ and } \mu(\psi(D)) = \mu(\psi^{-1}(\psi(D))) = \mu(D).$$

272
So, for all \( r \in \mathbb{N}, s_1, s_2, \ldots, s_r \in \mathbb{Z} \) and \( \psi_1, \psi_2, \ldots, \psi_r \in \Psi \) we have

\[
0 \geq \mu(X \cap X(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r)^c) \\
= \mu(X) - \mu(X \cap X(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r)) \\
\geq \mu(X) - \mu(X(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r)) \\
= \mu(X) - \mu(X) \\
= 0.
\]

Consequently we have, from (34), that

\[
\mu_*(A \cap A(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r)^c) = 0,
\]

and hence we can deduce that

\[
\mu_* \left( A \cap \left( \bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} A(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r) \right)^c \right) = 0. \tag{35}
\]

Now put

\[
C = \bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} B(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r). \tag{36}
\]

Note that \( C \subset B \) and that \( C \in \mathcal{B} \). Further, note that for any \( \phi \in \Phi \) we have

\[
\phi^{-1}(C) = \phi^{-1} \left( \bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} B(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r) \right) \\
= \bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} \phi^{-1} \left( B(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r) \right) \\
= \bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} B(s_1, \ldots, s_r, -1, \psi_1, \ldots, \psi_r, \psi) \\
= C.
\]

The last step of the above step follows by noting that for each \( \psi \in \Psi \) the set

\[
\{ \psi_1^{s_1} \circ \cdots \circ \psi_r^{s_r}(x) : r \in \mathbb{N}, s_1, \ldots, s_r \in \mathbb{Z}, \psi_1, \ldots, \psi_r \in \Psi \}
\]

is \( \psi \)-invariant. We have therefore established that \( C \) is \( \Psi \)-invariant and as \( \Psi \) is ergodic we know that \( \mu(C) \in \{0, 1\} \).
As $A \subseteq B$ it follows that $A(s_1, \ldots, s_r, \psi_1, \ldots, \psi_r) \subseteq B(s_1, \ldots, s_2, \psi_1, \ldots, \psi_r)$ for any selection of $r$, $s_i$'s and $\psi_i$'s. Thus

$$\bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} \bigcap_{\phi_1, \ldots, \phi_r \in \Phi} A(s_1, \ldots, s_r, -1, \psi_1, \ldots, \psi_r, \psi)$$

is a subset of

$$\bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} \bigcap_{\phi_1, \ldots, \phi_r \in \Phi} B(s_1, \ldots, s_r, -1, \psi_1, \ldots, \psi_r, \psi)$$

and hence from (36), $C^c$ is a subset of

$$\left( \bigcap_{r \in \mathbb{N}} \bigcap_{s_1, \ldots, s_r \in \mathbb{Z}} \bigcap_{\phi_1, \ldots, \phi_r \in \Phi} A(s_1, \ldots, s_2, -1, \psi_1, \ldots, \psi_r, \psi) \right)^c,$$

so that from (35) it follows that

$$\mu_*(A \cap C^c) = 0.$$

We now have

$$0 < \mu_*(A) \leq \mu_*(A \cap C) + \mu_*(A \cap C^c) = \mu_*(A \cap C) \leq \mu(C) \leq \mu(B) < 1,$$

which contradicts $\mu(C) \in \{0, 1\}$. Therefore $\mu_*(A) = 1$. \hfill \diamondsuit
Proposition 7.5

Let \((X, \mathcal{B}, \mu)\) be a probability space with a complete measure. Let \(G\) be a countably infinite group of \(\mu\)-measure preserving transformations on \(X\) with the property that if \(g \in G\) is such that there is an \(x \in X\) with \(g(x) = x\), then \(g = i\). Let \(\Psi\) be a family of one-to-one transformations on \(X\) such that for each \(\psi \in \Psi\) and \(Y \in \mathcal{B}\), \(\psi(Y), \psi^{-1}(Y) \in \mathcal{B}\), and assume that \(\Psi\) is \(\mu\)-measure preserving and ergodic. Assume further that for each \(x \in X\) and \(g \in G\), the \(\Psi\)-orbit \(O_\psi(x)\) is a partial \(G\) transversal, there is \(h \in G\) such that \(g(O_\psi(x)) = O_\psi(h(x))\) and \(g^{-1} \Psi g \subseteq \Psi\). Then the following hold.

(a) There is a disjoint sequence \(\{J_n\}\) of \(\Psi\)-invariant subsets of \(X\) such that

\[
X = \bigcup_{n=1}^\infty J_n, J_n \notin \mathcal{B} \text{ and } \mu_*(J_n) = 1 \text{ for all } n.
\]

In this case, for any \(M \in \mathcal{B}\) with \(\mu(M) > 0\),

\[
\sum_{n=1}^\infty \mu_*(M \cap J_n) = \infty. \quad (37)
\]

(b) If \(n \in \mathbb{N}, n \geq 2\), there are \(\Psi\)-invariant subsets \(K_1, K_2, \ldots, K_n\) of \(X\) such that

\[
X = \bigcup_{i=1}^n K_i, K_i \notin \mathcal{B} \text{ and } \mu_*(K_i) = 1 \text{ for all } i = 1, 2, \ldots, n.
\]

In this case for any \(M \in \mathcal{B}\),

\[
\sum_{i=1}^n \mu_*(M \cap K_i) = n\mu(M). \quad (38)
\]

Proof:

To prove (a), Theorem 7.1 gives us that \(X\) can be expressed as the disjoint union of a sequence of non-measurable sets which we can denote by \(\{J_n\}\). Also, each \(J_n\) has outer measure 1 as proven in Proposition 7.4. This proves

275
the first part of (a).

Next, if \( n \in \mathbb{N}, \ n \geq 2 \) we can take the sets \( K_1, \ldots, K_{n-1} \) to be defined as \( K_i = J_i \) for each \( i \in \{1, 2, \ldots, n-1\} \) and define \( K_n = \bigcup_{i=n}^{\infty} J_n \). From this definition we immediately get that \( K_i = J_i \not\in \mathcal{B} \) and that \( \mu_*(K_i) = \mu_*(J_i) = 1 \) for each \( i \in \{1, 2, \ldots, n-1\} \). For \( K_n \) note that \( 1 \geq \mu_*(K_n) \geq \mu_*(J_n) = 1 \) so that \( \mu_*(K_n) = 1 \). Since \( \mu_*(K_n^c) \geq \mu_*(K_1) = 1 \) we must also have that \( K_n \not\in \mathcal{B} \). Hence for this selection of \( \{J_n\}, K_1, K_2, \ldots, K_n \) we have proven the result apart form the two equations (37) and (38).

Now, let \( M \in \mathcal{B} \) and note that for each \( i \in \{1, 2, \ldots, n\} \) we have

\[
\mu_*(M \cap K_i) + \mu_*(M^c \cap K_i) = \mu_*(K_i),
\]

so that adding these terms we get

\[
\sum_{i=1}^{n} \mu_*(M \cap K_i) + \mu_*(M^c \cap K_i) = \sum_{i=1}^{n} \mu_*(K_i) = \sum_{i=1}^{n} 1 = n.
\]

Thus, if \( \sum_{i=1}^{n} \mu_*(M \cap K_i) < n \mu(M) \) we have that

\[
n = \sum_{i=1}^{n} \mu_*(M \cap K_i) + \mu_*(M^c \cap K_i)
< n \mu(M) + \sum_{i=1}^{n} \mu_*(M^c \cap K_i)
\leq n \mu(M) + n \mu(M^c)
= n,
\]

a contradiction. So, \( \sum_{i=1}^{n} \mu_*(M \cap K_i) \geq n \mu(M) \). Also

\[
\sum_{i=1}^{n} \mu_*(M \cap K_i) \leq \sum_{i=1}^{n} \mu(M) = n \mu(M)
\]
so that we have
\[ \sum_{i=1}^{n} \mu_*(M \cap K_i) = n\mu(M) \]
which proves (38).

We prove (37) through (38) as follows. Let \( M \in \mathcal{B} \) with \( \mu(M) > 0 \). Let \( N \in \mathbb{R} \). Then there is an \( n \in \mathbb{N} \) such that \( n\mu(M) > N \). We then have, using (38) that
\[
\sum_{i=1}^{\infty} \mu_*(M \cap J_i) \geq \sum_{i=1}^{n} \mu_*(M \cap K_i) = n\mu(M) > N.
\]
As this is true for any \( N \in \mathbb{R} \) we have that
\[
\sum_{i=1}^{\infty} \mu_*(M \cap J_i) = \infty.
\]

\[ \Diamond \]

7.3 Examples of non-measurable sets

We have now established some results concerning non-measurable sets, especially non-measurable sets that are \( \Psi \)-invariant non-measurable sets for some family \( \Psi \) of transformations on a measure space. We now turn to using these results to show that there are transformation invariant non-measurable sets satisfying the condition (30) for the specific examples of transformations that generate the Carathéodory definition that we have considered throughout the thesis. We do this in order to confirm that they not only generate Carathéodory’s definition but also distinguish between the ‘well behaved’ and
'poorly behaved' sets as discussed earlier in the chapter.

We begin by considering the irrational rotations. In this case we can actually prove that there is the required type of non-measurable set for any family of irrational rotations, provided that they are independent. Recall that when considering rotations we use the measure space of the unit circle $T$ with the usual associated Lebesgue measure space.

The results in this section are not included in any other work, although Theorem 7.8 is simply a generalisation of a result that has been included in [25] and [27].

**Definition 7.2**

A family $\{\rho_\beta : \beta \in B\}$ of rotations on the unit circle $T$ is called independent if whenever $\beta_1, \beta_2, \ldots, \beta_n$ are distinct elements in $B$ and $j_1, j_2, \ldots, j_n \in \mathbb{Z}$ are such that

$$\rho_{\beta_1}^{j_1} \circ \rho_{\beta_2}^{j_2} \circ \cdots \circ \rho_{\beta_n}^{j_n} = \iota,$$

then $j_1 = j_2 = \ldots = j_n = 0$.

In order to prove that there are irrational rotation invariant non-measurable sets satisfying (30), we first introduce some notation regarding the calculation of images of rotations. For each $x \in T$, we can express $x$ as $x = e^{2\pi i \theta(x)}$ for some $\theta \in [0, 1)$. Using this notation we note that if $g$ is a rotation on $T$ then there is some $\gamma \in [0, 1)$ such that $g(x) = e^{2\pi i (\theta(x)+\gamma)}$ for each $x \in T$. We now note that

$$g(x) = e^{2\pi i (\theta(x)+\gamma)}$$

$$= e^{2\pi i (0+\gamma)} e^{2\pi i \theta(x)}$$

278
\[ = g(1)x. \]

To simplify the messy notation in performing calculations in the following Theorem we use this notation. (That is, we write \( g(x) \) as the product of \( g(1) \) and \( x \).) Also note that under this notation if \( g \) is a rotation on \( \mathbb{T} \) then \( g^{-1}(x) = (g(1))^{-1}x \).

**Theorem 7.3**

Let \( \mathcal{R} \) be an independent family of irrational rotations on \( \mathbb{T} \). Then there is an \( \mathcal{R} \)-invariant \( \mu \)-non-measurable set, \( A \), such that (30) is satisfied. That is, for any selected collection of standard sets, \( \mathcal{J} \), for \((\mathbb{T}, \mathcal{B}, \mu)\),

\[ \mu_*(A \cap J) + \mu_*(A^c \cap J) = 2\mu(J). \]

for each \( J \in \mathcal{J} \).

**Proof:**

Let \( \mathcal{J} \) be a collection of standard sets for \((\mathbb{T}, \mathcal{B}, \mu)\).

To show that there exist \( \mathcal{R} \)-invariant non-measurable sets we need to show that the conditions for Theorem 7.1 are satisfied. To show that these \( \mathcal{R} \)-invariant non-measurable sets satisfy (30) we show that the conditions for Proposition 7.5 are satisfied. This will show that (30) is true for all elements of \( \mathcal{B} \) and therefore is also true for all elements of \( \mathcal{J} \).

To show that the conditions for Theorem 7.1 are satisfied, let \( G \) be the group of rational rotations which is infinite and countable and for which the only transformation with a fixed point is the identity. Let \( \Psi = \mathcal{R} \) which is a family of one-to-one \( \mu \)-measure preserving transformations on \( X \). We then need to show that the following conditions hold.
(a) For each $x \in X$, the $\Psi$-orbit $O_\Psi(x)$ is a partial $G$-transversal.
(b) If $g \in G$ and $x \in X$, there is $h \in G$ such that $g(O_\Psi(x)) \subseteq O_\Psi(h(x))$.
(c) $G$ is infinite and countable.
(d) For all $g \in G$, $g^{-1}\Psi g \subseteq \Psi$.

We have already shown (c).

To show (b) let $g \in G$, $x \in X$ and set $h = g$. Then for any $y \in O_\Psi(x)$ there are $r \in \mathbb{N}$, $s_1, s_2, ..., s_r \in \mathbb{Z}$ and $\psi_1, \psi_2, ..., \psi_r \in \Psi$ such that

$$y = \psi_1^{s_1} \circ \psi_2^{s_2} \circ \cdots \circ \psi_r^{s_r}(x) = \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) x.$$

Also, $z \in O_\Psi(h(x))$ if and only if it is of the form

$$z = \phi_1^{t_1} \circ \phi_2^{t_2} \circ \cdots \circ \phi_n^{t_n}(h(x)) = \left( \prod_{j=1}^{n} \phi_j^{t_j}(1) \right) h(1)x,$$

for some collection of elements $t_1, t_2, ..., t_n \in \mathbb{Z}$ and $\phi_1, \phi_2, ..., \phi_n \in \Psi$. Now

$$g(y) = g(1)y = g(1) \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) x = \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) g(1)x = \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) h(1)x \in O_\Psi(h(x)).$$

Therefore $g(O_\Psi(x)) \subseteq O_\Psi(h(x))$ and (b) is satisfied.

To show (d), let $g \in G$ and $\psi \in \Psi$. Then

$$g^{-1}\psi g(x) = (g(1))^{-1}\psi(1)g(1)x = (g(1))^{-1}g(1)\psi(1)x = \psi(1)x = \psi(x).$$
for each $x \in X$ so that $g^{-1}\psi g = \psi$ and as this is true for each $\psi \in \Psi$ we have that $g^{-1}\Psi g \subseteq \Psi$ which satisfies (d).

To show that (a) is satisfied, suppose that $z, y \in O_\psi(x)$ and that $z$ and $y$ are in the same $G$-orbit. Then there is a $g \in G$ such that $z = g(y) = g(1)y$. Also there is are $u \in \mathbb{N}, m_1, m_2, ..., m_u \in \mathbb{Z}$ and $\rho_1, \rho_2, ..., \rho_u \in \Psi$ such that

$$y = \rho_1^{m_1} \circ \rho_2^{m_2} \circ \cdots \circ \rho_u^{m_u}(x) = \left( \prod_{j=1}^{u} \rho_j^{m_j}(1) \right) x,$$

and $t \in \mathbb{N}, n_1, n_2, ..., n_t \in \mathbb{Z}$ and $\phi_1, \phi_2, ..., \phi_t \in \Psi$ such that

$$z = \phi_1^{n_1} \circ \phi_2^{n_2} \circ \cdots \circ \phi_t^{n_t}(x) = \left( \prod_{j=1}^{t} \phi_j^{n_j}(1) \right) x,$$

and thus we can write

$$y = \psi_1^{s_1} \circ \psi_2^{s_2} \circ \cdots \circ \psi_u^{s_u}(x) = \left( \prod_{j=1}^{u} \psi_j^{s_j}(1) \right) x,$$

for some $r \in \mathbb{N}, s_1, s_2, ..., s_r \in \mathbb{Z}$ and $\psi_1, \psi_2, ..., \psi_r \in \Psi$. Hence

$$z = g(1) \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) z.$$

This implies that

$$g(1) \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) = 1$$

and hence that

$$\left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) = (g(1))^{-1}$$

As $g \in G$ there is a $k \in \mathbb{N}$ such that $g^k = 1$ and hence that $(g^{-1})^k = 1$. thus

$$\left( \prod_{j=1}^{r} \psi_j^{k_s_j}(1) \right) = \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right)^k = 1.$$
Since \( \Psi = \mathcal{R} \) is independent this implies that \( k s_1 = k s_2 = \ldots = k s_r = 0 \) and hence that \( s_1 = s_2 = \ldots = s_r = 0 \). We now have
\[
y = \left( \prod_{j=1}^{r} \psi_j^0(1) \right) z
= \left( \prod_{j=1}^{r} 1 \right) z
= z
\]
so that \( O_\Psi(x) \) is a partial \( G \)-transversal.

Therefore all the conditions of Theorem 7.1 are satisfied and hence there is an \( \mathcal{R} \)-invariant non-measurable subset of \( \mathbb{T} \). Further we know that \( \mathbb{T} \) can be expressed as the disjoint union of countably many \( \mathcal{R} \)-invariant non-measurable sets.

To satisfy the conditions of Proposition 7.5 we need to satisfy the same conditions as those for Theorem 7.1 with three additional requirements. Allowing the same allocations of transformations to take the places of \( G \) and \( \Psi \) these additional requirements are as follows.

\( (e) \) \( \Psi \) needs to be ergodic.

\( (f) \) For all \( M \in \mathcal{B}, \psi \in \Psi, \psi(M), \psi^{-1}(M) \in \mathcal{B} \).

\( (g) \) For all \( g \in G \) and \( x \in X \) there is a \( h \in G \) such that
\[
g(O_\Psi(x)) = O_\Psi(h(x)).
\]

Theorem 4.2 gives us that any irrational rotation is ergodic. Hence if \( A \in \mathcal{B} \) is \( \Psi \)-invariant then \( A \) is \( \psi \)-invariant for some \( \psi \in \Psi \) (it is actually invariant for any \( \psi \in \Psi \) but this stronger condition is unnecessary) and hence, as \( \psi \) is ergodic, \( \mu(A) \in \{0, 1\} \) so that \( \Psi \) must be ergodic. Therefore \( (e) \) is satisfied.
For each $\psi \in \Psi$, $\psi, \psi^{-1}$ are continuous transformations on a complete separable metric (and hence Hausdorff) space in which all closed sets are measurable. Thus Theorem 6.7 gives us that condition (f) is satisfied.

Finally, we know that by selecting $h = g$ we have that $g(O_{\Psi}(x)) \subseteq O_{\Psi}(h(x))$. To show the converse, suppose that $y \in O_{\Psi}(h(x)) = O_{\Psi}(g(x))$. Then there are $r \in \mathbb{N}, s_1, s_2, ..., s_r \in \mathbb{Z}$ and $\psi_1, \psi_2, ..., \psi_r \in \Psi$ such that

$$y = \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) g(1)x.$$ 

Thus

$$y = \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) g(1)x = g(1) \left( \prod_{j=1}^{r} \psi_j^{s_j}(1) \right) x \in g(O_{\Psi}(x)).$$

Therefore $g(O_{\Psi}(x)) = O_{\Psi}(h(x))$ and hence (g) is satisfied which proves the Theorem.

We prove that Kakutani Type I transformations and Chacon’s $\xi$ have the required associated non-measurable sets in a less direct manner. As all of the relevant primitive transformations are defined on the same space we will, for the remainder of the section allow $(X, \mathcal{B}, \mu)$ to be the Lebesgue measure space on the interval $[0, 1) \subset \mathbb{R}$.

For simplicity of notation we will define $\oplus$ to be the addition operation on $X$ defined by

$$a \oplus b = (a + b)(\text{mod } 1).$$
for all $a, b \in X$. We also note that $(X, \oplus)$ is a group with this definition of $\oplus$.

Now, for any $p \in \mathbb{N}$, denote by $\mathbb{Q}_p$ the set of $p$-adic rationals and define the family of transformations on $X$

$$\Phi_p = \{ \phi_d : d \in \mathbb{Q}_p \},$$

where for each $d \in \mathbb{Q}_p$, $\phi_d$ is defined for all $x \in X$ by

$$\phi_d(x) = x \oplus d.$$

Note that $(\mathbb{Q}_p, \oplus)$ is a subgroup of $(X, \oplus)$ for each $p \in \mathbb{N}$, and that the operation $\oplus$ is commutative on $X$. Note also that as each $\phi_d \in \Phi_p$ is a translation on $X$ it is measure preserving, and as each such $\phi_d$ is also invertible it is also outer measure preserving. We prove that for any $p \in \mathbb{N}$, $p > 1$, there is a $\Phi_p$-invariant set which is non-measurable in the manner we require for the Kakutani type I transformations and for Chacon's $\xi$. We then show that for appropriate choices of $p$ such a set is the desired non-measurable set for either Chacon's $\xi$ or a Kakutani type I transformation. Note that in general, when we state that a set is non-measurable in the manner that we require, we mean that there is a set that satisfies (30).

We first give two results about $\Phi_p$-invariance that shall be required. Also, we show, as stated in Chapter 6, that there is a family of transformations that generates Carathéodory's definition of measurable sets, but that no individual element of the family does. In fact, for any $p \in \mathbb{N}$, $\Phi_p$ constitutes such a family.

**Proposition 7.6**

*For any $x \in X$ and any $p \in \mathbb{N}$, $O_{\Phi_p}(x) = x \oplus \mathbb{Q}_p$*
Proof:
For any \( x \in X, d \in \mathbb{Q}_p \) there is a \( \phi_d \in \Phi_p \) such that \( \phi_d(x) = x \oplus d \) and hence \( x \oplus d \in O_{\Phi_p}(x) \) so that \( x \oplus \mathbb{Q}_p \subseteq O_{\Phi_p}(x) \).

Conversely, suppose that \( y \in O_{\Phi_p}(x) \) then there are \( r \in \mathbb{N}, s_1, s_2, ..., s_r \in \mathbb{Z}, d_1, d_2, ..., d_r \in \mathbb{Q}_p \) and \( \phi_{d_1}, \phi_{d_2}, ..., \phi_{d_r} \in \Phi_p \) such that as \( \mathbb{Q}_p \) is a subgroup

\[
y = \phi_{s_1}^{d_1} \circ \phi_{s_2}^{d_2} \circ ... \circ \phi_{s_r}^{d_r}(x) = x \oplus s_1 d_1 \oplus s_2 d_2 \oplus ... \oplus s_r d_r \in x \oplus \mathbb{Q}_p
\]

so that \( O_{\Phi_p}(x) \subseteq x \oplus \mathbb{Q}_p \) and thus \( O_{\Phi_p}(x) = x \oplus \mathbb{Q}_p \).

Proposition 7.7
For each \( p \in \mathbb{N}, \Phi_p \) is ergodic.

Proof:
Suppose that \( B \in B \) is \( \Phi_p \)-invariant such that \( 0 < \mu(B) < 1 \). Then \( \mu(B^c) > 0 \). It then follows, by Lemma 5.5, that there are \( p \)-adic intervals of density to within \( 1/2 \), \( J_1 \) and \( J_2 \), of the same measure for \( B \) and \( B^c \) respectively.

Then we note that the lower end points, \( d_1 \) and \( d_2 \), of the intervals \( J_1 \) and \( J_2 \) must be \( \mathbb{Q}_p \)-adic rationals and thus \( d_3 = (d_2 - d_1)(\text{mod } 1) \) is also in \( \mathbb{Q}_p \). Therefore \( \phi_{d_3}(J_1) = J_2 \) and hence (as \( B \) in \( \Phi_p \)-invariant)

\[
\mu(B \cap J_2) = \mu(\phi_{d_3}^{-1}(B \cap J_2)) = \mu(\phi_{d_3}^{-1}(B) \cap \phi_{d_3}^{-1}(J_2)) = \mu(B \cap J_1) > \frac{1}{2} \mu(J_1)
\]
\[
\frac{1}{2} \mu(J_2).
\]

Therefore, as \(B \in \mathcal{B}\) we have

\[
\mu(J_2) = \mu(B \cap J_2) + \mu(B^c \cap J_2) > 2 \frac{1}{2} \mu(J_2) = \mu(J_2).
\]

This contradiction gives us that \(\mu(B) \in \{0, 1\}\) and hence \(\Psi_p\) is ergodic. \(\diamondsuit\)

**Theorem 7.4**

*For any \(p \in \mathbb{N}\) and any collection of standard sets \(\mathcal{J}\), \(\Phi_p\) generates the Carathéodory definition of measurable sets. Conversely, for each \(\phi \in \Phi_p\), \(\phi\) does not generate the Carathéodory definition of measurable sets.*

**Proof:**

Let \(\mathcal{J}\) be a standard collection of sets for \((X, \mathcal{B}, \mu)\), the usual Lebesgue measure space on \([0, 1)\).

Let \(p \in \mathbb{N}\). By proposition 7.7, \(\Phi_p\) is ergodic. As each \(\phi \in \Phi_p\) is a translation, that is, there is a \(p\)-adic rational, \(b\) such that for each \(x \in X\), \(\phi(x) = x \oplus b\), we have that \(\phi\) is a bijection and is therefore invertible. Clearly, it is also measure preserving. Also, as \(\Phi_p\) is in one-to-one correspondence with the elements of \(\mathbb{Q}_p\) we know that \(\Phi_p\) is countable. It therefore follows from Theorem 6.5 that \(\Phi_p\) generates the Carathéodory definition of measurable sets.

Now, let \(\phi \in \Phi_p\) so that there is a \(p\)-adic rational, \(b\), such that \(\phi(x) = x \oplus b\) for each \(x \in X\). It follows, since \(b\) is a \(p\)-adic rational and thus \(pb \in \mathbb{Z}\), that

\[
O_\phi(x) = \{x \oplus nb : n \in \mathbb{Z}\} = \{x \oplus nb : n \in \{0, 1, \ldots, p - 1\}\}.
\]

Let \(A = [0, 1/2p)\). Then

\[
O_\phi(A) = \bigcup_{n \in \mathbb{Z}} \phi^n(A)
\]

286
Let \( B = \bigcup_{C \in \mathcal{O}(A)} C \) so that \( B \) is \( \phi \)-invariant. Then as

\[
\{ [n/2p, (n+1)/2p) : n \in \{0, 1, \ldots, p-1\} \}
\]

is a disjoint collection of sets we have

\[
\mu(B) = \mu \left( \bigcup_{n=0}^{p-1} \left[ \frac{n}{2p}, \frac{n+1}{2p} \right) \right) = \sum_{n=0}^{p-1} \mu \left( \left[ \frac{n}{2p}, \frac{n+1}{2p} \right) \right) = \sum_{n=0}^{p-1} \frac{1}{2p} = \frac{1}{2}.
\]

Hence \( 0 < \mu(B) < \mu(X) \) and \( B \) is \( \phi \)-invariant and therefore \( \phi \) is not ergodic. Therefore, by Theorem 6.4, \( \phi \) cannot generate the Carathéodory definition of measurable sets. \( \diamond \)

Now, with Propositions 7.6 and 7.7 we can prove that for each \( p \in \mathbb{Q}_p \) there is a \( \Psi_p \)-invariant non-measurable set satisfying the properties desired for the transformations that generate Carathéodory’s definition.

**Theorem 7.5**

Let \( \mathcal{J} \) be a collection of standard sets for \((X, \mathcal{B}, \mu)\). Then for any \( p \in \mathbb{N} \) there is a \( \Phi_p \)-invariant \( \mu \)-non-measurable set, \( A \), such that (30) is satisfied. That
is, for each \( J \in \mathcal{J} \),

\[
\mu_*(A \cap J) + \mu_*(A^c \cap J) = 2\mu(J).
\]

**Proof:**

As in Theorem 7.3, we need only to satisfy the requirements of Theorem 3.3 and Proposition 7.5. Let \( p \in \mathbb{N} \) and let \( b \) be relatively prime to \( p \). Set

\[
G = \Phi_b
\]

and

\[
\Psi = \Phi_p.
\]

We therefore have that \( G \) and \( \Psi \) are both countably infinite families (in fact groups) of \( \mu \)-measure preserving invertible transformations. We firstly need to satisfy

(a) For each \( x \in X \), The \( \Psi \)-orbit \( O_\psi(x) \) is a partial \( G \)-transversal.

(b) If \( g \in G \) and \( x \in X \), there is \( h \in G \) such that \( g(O_\psi(x)) \subseteq O_\psi(h(x)) \).

(c) \( G \) is infinite and countable.

(d) For all \( g \in G \), \( g^{-1}\Psi g \subseteq \Psi \).

(c) is immediate.

To show (a), let \( y, z \in O_\psi(x) \) such that \( y \) and \( z \) are both in the same \( G \)-transversal. We then have that for some \( p \)-adic rationals \( p_1 \) and \( p_2 \)

\[
y = x \oplus p_1 \quad \text{and} \quad z = x \oplus p_2
\]

and hence

\[
y = z \oplus p_3
\]
where \( p_3 = (p_1 - p_2) \pmod{1} \) is a \( p \)-adic rational. Also, as \( y \) and \( z \) are in the same \( G \)-orbit, we know that there is a \( g \in G \) such that \( y = g(z) = z \oplus b_1 \) where \( b_1 \) is a \( b \)-adic rational. Therefore

\[
z \oplus b_1 = z \oplus p_3
\]

and hence

\[
b_1 \equiv p_3 (\pmod{1})
\]

which can only occur if \( b_1 = p_3 = 0 \) as \( p \) and \( b \) are relatively prime. Hence \( y = z \oplus p_3 = z \oplus 0 = z \) so that \( O_\Psi(x) \) is a \( G \)-transversal.

To show (b) set \( h = g \) so that there is a \( b \)-adic rational \( b_1 \) such that for all \( x \in X \), \( h(x) = g(x) = x \oplus b_1 \). Let \( y \in g(O_\Psi(x)) \) so that for some \( p \)-adic rational \( p_1 \) we have

\[
y = g(x \oplus p_1) \\
= x \oplus p_1 \oplus b_1 \\
= x \oplus b_1 \oplus p_1 \\
= h(x) \oplus p_1 \\
\in O_\Psi(h(x)),
\]

where the final inclusion is ensured by Proposition 7.6. Therefore \( g(O_\Psi(x)) \subseteq O_\Psi(h(x)) \).

To show (d), let \( g \in G \) and \( \psi \in \Psi \) so that there is a \( b \)-adic rational \( b_1 \) and a \( p \)-adic rational \( p_1 \) such that for all \( x \in X \), \( g(x) = x \oplus b_1 \), \( \psi(x) = x \oplus p_1 \) and \( g^{-1}(x) = x \oplus (1 - b_1) \).

\[
g^{-1}\psi g(x) = g^{-1}\psi(x \oplus b_1) \\
= g^{-1}(x \oplus b_1 \oplus p_1) \\
= x \oplus b_1 \oplus p_1 \oplus (1 - b_1)
\]
As this is true for each $x \in X$, $g^{-1}\psi g \in \Psi$, and as this is true for each $\psi \in \Psi$ we have that $g^{-1}\Psi g \subseteq \Psi$.

We therefore know, by Theorem 7.1, that there is a $\Psi$-invariant non-measurable set and in fact that $X$ can be represented by a countable union of disjoint $\Psi$-invariant non-measurable sets. To prove that there are sets with the appropriate outer measure properties we must further show the following.

(e) $\Psi$ is ergodic.

(f) For all $M \in \mathcal{B}$, $\psi \in \Psi$, $\psi(M), \psi^{-1}(M) \in \mathcal{B}$.

(g) For all $g \in G$ and $x \in X$ there is a $h \in G$ such that

$$g(O_\psi(x)) = O_\psi(h(x)).$$

(e) is proven in Proposition 7.7. (f) follows from the fact that each $\psi \in \Psi$ is a measurable invertible transformation.

For (g) we already know that setting $h = g$ gives us that $g(O_\psi(x)) \subseteq O_\psi(h(x))$. Again noting that there is a $b$-adic rational $b_1$ such that for each $x \in X$, $g(x) = h(x) = x \oplus b_1$, let $y \in O_\psi(h(x))$ so that there is a $\psi \in \Psi$ and associated $p$-adic rational $p_1$ for which

\[
y = \psi(h(x)) = \psi(x \oplus b_1) = x \oplus b_1 \oplus p_1.
\]
Thus, using Proposition 7.6, we have

\[ y = x \oplus b_1 \oplus p_1 \]
\[ = x \oplus p_1 \oplus b_1 \]
\[ \in O_\psi(x) \oplus b_1 \]
\[ = g(O_\psi(x)). \]

Therefore \( O_\psi(h(x)) \subseteq g(O_\psi(x)) \) and hence \( g(O_\psi(x)) = O_\psi(h(x)) \). \hfill \Box

The way we utilize Theorem 7.5 to give us results for Kakutani type I transformations and for Chacon's \( \xi \) is via the following the following proposition. In the proposition we consider the sets

\[ \bigcup_{\psi \in \Psi \psi(x)^n : n \in \mathbb{Z}, \psi^n(x) \text{ is defined}} \]

for some family of transformations \( \Psi \). This is similar to \( O_\psi(x) \) except that for a given \( x \) these sets allow for integers, \( n \) such that \( \psi^n(x) \) is a set of more than one point.

**Proposition 7.8**

Let \( \Psi \) be a family of transformations on \( (X, \mathcal{B}, \mu) \). Suppose that for some \( p \in \mathbb{N} \)

\[ \bigcup_{\psi \in \Psi \psi^n(x) : n \in \mathbb{Z}, \psi^n(x) \text{ is defined}} \subseteq x \oplus \mathbb{Q}_p \]

for each \( x \in X \). Then, if \( A \) is \( \Phi_p \)-invariant it is \( \Psi \)-invariant.

**Proof:**

Let \( p \in \mathbb{N} \) be such that

\[ \bigcup_{\psi \in \Psi \psi^n(x) : n \in \mathbb{Z}, \psi^n(x) \text{ is defined}} \subseteq x \oplus \mathbb{Q}_p \]
for each \( x \in X \), let \( A \) be \( \Phi_p \)-invariant and let \( \psi \in \Psi \). Then, if \( a \in A \),

\[
\psi^{-1}(a) \subseteq a \oplus \mathbb{Q}_p \\
= O_{\Phi_p}(a) \\
\subseteq \bigcup_{a \in A} O_{\Phi_p}(a) \\
= \bigcup_{a \in A} \bigcup_{\phi \in \Phi_p} \phi(a) \\
= \bigcup_{\phi \in \Phi_p} \phi(A) \\
= \bigcup_{\phi \in \Phi_p} A \\
= A
\]

and hence

\[
\phi^{-1}(A) \subseteq A. \quad (40)
\]

Note also that \( O_{\Phi_p}(a) \subseteq A \). Then, conversely, if \( a \in A \)

\[
\psi(a) \in O_{\Phi}(a) \\
\subseteq a \oplus \mathbb{Q}_p \\
= O_{\Phi_p}(a) \\
\subseteq A,
\]

hence there exists a \( a_1 \in A \) such that \( \psi(a) = a_1 \). Therefore \( a \in \psi^{-1}(a_1) \subseteq \psi^{-1}(A) \).

Hence \( A \subseteq \psi^{-1}(A) \). Combining with (40) we have that \( \psi^{-1}(A) = A \) and hence, since \( \psi \) was arbitrarily chosen, we have that \( A \) is \( \Psi \)-invariant. \( \diamond \)

A particular case of Proposition 7.8 is when the family, \( \Psi \), of transformations is one to one. In this case the inverse image of any point has at most one point and hence in this case

\[
\bigcup_{\psi \in \Psi} \{ \psi^n(x) : n \in \mathbb{Z}, \psi^n(x) \text{ is defined} \} \subseteq O_{\Phi}(a),
\]

292
so that in the statement of Proposition 7.8 it is sufficient to require that there be a \( p \in \mathbb{N} \) such that \( O_\Psi(x) \subseteq x \oplus \mathbb{Q}_p \) for each \( x \in X \). This notationally neater case will also be of use to us and so we state the obvious corollary that comes from this observation.

**Corollary 7.1**

Let \( \Psi \) be a family of one to one transformations. Suppose that for some \( p \in \mathbb{N} \)

\[
O_\Psi(x) \subseteq x \oplus \mathbb{Q}_p
\]

for each \( x \in X \). Then, if \( A \) is \( \Phi_p \)-invariant it is \( \Psi \)-invariant.

From Proposition 7.8 it follows that if \( \Psi \) is a family of transformations on \( (X, \mathcal{B}, \mu) \) for which there is a \( p \in \mathbb{N} \) such that for each \( x \in X \),

\[
\bigcup_{\psi \in \Psi} \{ \psi^n(x) : n \in \mathbb{Z}, \psi^n(x) \text{ is defined} \} \subseteq x \oplus \mathbb{Q}_p
\]

then there is a \( \Phi_p \) and hence \( \Psi \)-invariant non-measurable set with all the properties that we would like for a transformation generating Carathéodory’s definition. Therefore, showing that, in view of the comments preceding Proposition 7.8, for each Kakutani type I transformation, \( T \), or for Chacon’s \( \xi \) there is a \( p \in \mathbb{N} \) for which

\[
\{ T^n(x) : n \in \mathbb{Z}, T^n(x) \text{ is defined} \} \subseteq x \oplus \mathbb{Q}_p
\]

or

\[
\{ \xi^n(x) : n \in \mathbb{Z}, \xi^n(x) \text{ is defined} \} \subseteq x \oplus \mathbb{Q}_p
\]

for each \( x \in X \) is sufficient to show that there are the desired non-measurable sets for these transformations. After showing that this is the case, we complete the section by showing that for some of the Kakutani type I transformations, say for example \( T \), there is in fact a \( p \in \mathbb{N} \) such that \( O_T(x) = x \oplus \mathbb{Q}_p \).
for each $x \in X$.

**Theorem 7.6**

If $\xi$ is Chacon's transformation there is a $\xi$-invariant $\mu$-non-measurable set, $A$, such that (30) is satisfied. That is, for any collection of standard sets $J$ for $(X, \mathcal{B}, \mu)$,

$$\mu_*(A \cap J) + \mu_*(A^c \cap J) = 2\mu(J)$$

for each $J \in \mathcal{J}$.

**Proof:**

Let $\mathcal{J}$ be a standard collection of sets for $\mathcal{J}$ We prove the theorem by showing that for each $x \in X$,

$$\{\xi^n(x) : n \in \mathbb{Z}, \xi^n(x) \text{ is defined} \} \subseteq x \oplus \mathbb{Q}_6.$$  

We use the same notation in discussing Chacon's $\xi$ as we did in constructing it in section 5.2. Let $x \in X$, then $x \in I_n^k$ for some $(2, 3)$-adic interval, $I_n^k$ of order $n$. As $\xi$ maps $I_n^k$ linearly and with positive slope onto another $(2, 3)$-adic interval $I_n^i$ of the same length, we must have that on $I_n^k$, $\xi(x) = x + c$ for some constant $c$. As the lower endpoint $i_n^k$ of $I_n^k$ must be mapped to the lower endpoint $i_n^i$ of $I_n^i$ and both of these points are $(2, 3)$-adic rationals we must have that $c = i_n^i - i_n^k$ is a $(2, 3)$-adic rational. As every $(2, 3)$-adic rational is of the form $q/(3 \cdot 2^n) = l/6^n$ for some $q \in \{0, 1, \ldots, 3 \cdot 2^n - 1\}$ and some $l \in \{0, 1, \ldots, 6^n - 1\}$ we have that $\xi(x) \in x \oplus \mathbb{Q}_6$. Hence for each $x \in X$, $\xi(x) \in x \oplus \mathbb{Q}$. It follows that for each $x \in X$ there is a 6-adic rational $s(x)$ such that $\xi(x) = x \oplus s(x)$.

Now suppose that $\xi^n(x) \in x \oplus \mathbb{Q}_6$ so that there is a 6-adic rational $t$ such that $\xi^n(x) = x \oplus t$. Then

$$\xi^{n+1}(x) = \xi(\xi^n(x))$$
= \xi(x \oplus t)
= x \oplus t \oplus s(x + t)
= x \oplus t_1

where \( t_1 = t + s(x + t) \) which is a 6-adic rational. By induction we now have that \( \xi(x) \in x \oplus \mathbb{Q}_6 \) for each \( n \in \mathbb{N} \). Now let

\[
\xi^n(x) = x \oplus s(x)
\tag{41}
\]

for some 6-adic rational \( s(x) \) and some \( n \in \mathbb{N} \cup \{0\} \). Suppose that for some \( y \in \xi^{n-1}(x) \),

\[
y = x \oplus r
\]

for some \( r \notin \mathbb{Q}_6 \). In this case, as \( y \in X \), we have that

\[
\xi^n(x) = x \oplus r \oplus s(y).
\]

By (41) we then have that \( s(x) = r \oplus s(y) \). As \( r \) is not a 6-adic rational this is impossible. Therefore, for each \( y \in \xi^{n-1}(x) \), \( y = x \oplus s \) for some 6-adic rational \( s \). Thus for each \( n \in \mathbb{Z} \) and each \( x \in X \) we have that \( \xi^n(x) \subseteq x \oplus \mathbb{Q}_6 \) and hence that

\[
\{ \xi^n(x) : n \in \mathbb{Z}, \xi^n(x) \text{ is defined} \} \subseteq x \oplus \mathbb{Q}_6.
\]

By Proposition 7.8 and Theorem 7.5 this proves the theorem. \( \diamond \)

**Theorem 7.7**

*If \( T \) is a Kakutani type I transformation there is a \( T \)-invariant \( \mu \)-non-measurable set, \( A \), such that (30) is satisfied. That is, for any collection of standard sets, \( \mathcal{J} \), for \( (X, \mathcal{B}, \mu) \)

\[
\mu_*(A \cap J) + \mu_*(A^c \cap J) = 2\mu(J)
\]

for each \( J \in \mathcal{J} \).*
Proof:
Let \( J \) be a collection of standard sets for \((X, \mathcal{B}, \mu)\).

Let \( T \) be a Kakutani type I transformation. \( T \) is thus a piecewise linear
transformation with \( \frac{dT}{dx} = 1 \) on each piece. Further there is a \( b \in \mathbb{N} \) such that
for each \( n \in \mathbb{N} \)
\[
O_T([0, b^{-n})) = \left\{ \left[ \frac{k}{b^n}, \frac{k+1}{b^n} \right) : k \in \{0, 1, \ldots, b^n - 1\} \right\},
\]
which in turn implies that for each \( n \in \mathbb{N} \)
\[
O_T\left( \left[ \frac{j}{b^n}, \frac{j+1}{b^n} \right) \right) = \left\{ \left[ \frac{k}{b^n}, \frac{k+1}{b^n} \right) : k \in \{0, 1, \ldots, b^n - 1\} \right\}, \tag{42}
\]
for each \( j \in \{0, 1, \ldots, b^n - 1\} \).

We prove the theorem by showing that for each \( x \in X, \)
\[
\{ T^n(x) : n \in \mathbb{Z}, T^n(x) \text{ is defined} \} \subseteq x \oplus \mathbb{Q}_b.
\]
Let \( x \in X \). Then as \( T \) is piecewise linear \( x \) is in a piece, say \( P \), and further
there must be a \( b \)-adic subinterval, say \( J \) of \( P \). As \( \frac{dT}{dx} = 1 \) on \( P \) we must have
that \( T(y) = y \oplus c \) for some \( c \in \mathbb{R} \) for each \( y \in P \) (and hence each \( y \in J \)). Now
suppose that \( c \notin \mathbb{Q}_b \), Then \( T(J) = J \oplus c \) is not a \( b \)-adic interval and infact
\( \mu(T(J) \triangle J) > 0 \) which contradicts (42). Therefore \( c \in \mathbb{Q}_b \). We therefore have
that \( T(x) = x \oplus b \) for some \( b \in \mathbb{Q}_b \) and hence for each \( x \in X \) \( T(x) \in x \oplus \mathbb{Q}_b \)
and thus there is a \( b \)-adic rational \( b(x) \) such that \( T(x) = x \oplus b(x) \).

Now suppose that \( T^n(x) \in \mathbb{Q}_b \) and hence there is a \( b \)-adic rational \( b \) such that
\( T^n(X) = x \oplus b \), then
\[
T^{n+1}(x) = T(T^n(x)) = T(x \oplus b)
\]
296
where \( b_1 = (b \oplus b(x \oplus b))(\text{mod } 1) \in \mathbb{Q}_b. \) Therefore, by mathematical induction we have that for each \( x \in X \) and each \( n \in \mathbb{N} \) there is a \( b \in \mathbb{Q}_b \) such that \( T^n(x) = x \oplus b \in x \oplus \mathbb{Q}_b. \)

Also, suppose that \( T^n(x) = x \oplus b \in x \oplus \mathbb{Q}_b \) and \( y \in T^{n-1}(x) \). Then we have \( T(y) = y \oplus b(y) \) and \( T(y) = T^n(x) = x \oplus b. \) Thus \( y \oplus b(y) = x \oplus b \) and hence \( y = x \oplus b \oplus (1 - b(y)) \in x \oplus \mathbb{Q}_b. \) Hence by mathematical induction we have that for each \( n \in \mathbb{Z} \) and each \( x \in X, \) \( T^n(x) \subseteq x \oplus \mathbb{Q}_b \) and therefore we have

\[
\{T^n(x) : n \in \mathbb{Z}, T^n(x) \text{ is defined}\} \subseteq x \oplus \mathbb{Q}_b.
\]

\[\Diamond\]

A result that comes out of the proof of Theorem 7.7 is now noted as a Corollary so that we can refer to it.

**Corollary 7.2**

Let \( T \) be a Kakutani type I transformation such that

\[
O_T([0, b^{-n})) = \left\{ \left[ \frac{k}{b^n}, \frac{k+1}{b^n} \right) : k \in \{0, 1, \ldots, b^n - 1\} \right\}.
\]

Then for each \( x \in X \)

\[
\{T^n(x) : n \in \mathbb{Z}, T^n(x) \text{ is defined}\} \subseteq x \oplus \mathbb{Q}_b.
\]

This implies, in particular, that for each \( x \in X \) there is a \( b \)-adic rational \( \beta(x) \) such that

\[
T(x) = x \oplus \beta(x).
\]
We now show that for some Kakutani type I transformations we have that there is a $b \in \mathbb{N}$ such that for each $x \in X$

$$O_T(x) = \mathbb{Q}_b.$$ 

In order to do this we must define the type of Kakutani type I transformations we will be considering. Let $T$ be a Kakutani type I transformation so that $T$ is piecewise linear with $\frac{dT}{dx} = 1$ on each piece, and such that there is a $b \in \mathbb{N}$ with the property that for each $n \in \mathbb{N}$

$$O_T([0, b^{-n})) \equiv \left\{ \left[ \frac{k}{b^n}, \frac{k + 1}{b^n} \right] : k \in \{0, 1, ..., b^n - 1\} \right\}.$$ 

The additional condition that we need for the transformations is that for each $n \in \mathbb{N}$, we have that for each $k \in \{0, 1, ..., b^n - 2\}$

$$\left[ \frac{k}{b^n}, \frac{k + 1}{b^n} \right]$$

is a piece over which $T$ is linear and for each $j \in \{0, 1, ..., b^n - 1\}$

$$T\left( \left[ \frac{j}{b^n}, \frac{j + 1}{b^n} \right] \right)$$

is another $b$-adic interval of order $n$. Further

$$T^j \left( \left[ \frac{0}{b^n}, \frac{1}{b^n} \right] \right) = [1 - b^{-n}, 1)$$

if and only if $j = b^n - 1$. We know that such transformations exist as Kakutani’s original transformation satisfies these requirements as proven in Lemma 4.4. Such a transformation will be called a *Kakutani type II transformation*. An important property of these new transformations is

**Lemma 7.3**

Let $T$ be a Kakutani type II transformation and $b \in \mathbb{N}$ be such that

$$O_T((0, b^{-n})) \equiv \left\{ \left( \frac{k}{b^n}, \frac{k + 1}{b^n} \right) : k \in \{0, 1, ..., b^n - 1\} \right\}$$

298
Then for each \( m \in \{0, 1, \ldots, b^n - 1\} \) there is a \( k \in \{0, 1, \ldots, b^n - 1\} \) such that

\[
T^k([0, b^{-n})) = \left[ \frac{m}{b^n}, \frac{m + 1}{b^n} \right).
\]

Proof:

By the definition of Kakutani type II transformations we know that

\[
\{T^k([0, b^{-n})) : k \in \{0, 1, \ldots, b^n - 1\} \} \subseteq \left\{ \left[ \frac{m}{b^n}, \frac{m + 1}{b^n} \right) : m \in \{0, 1, \ldots, b^n - 1\} \right\}.
\]

Suppose that the two sets are not equal, then there must be \( k_1, k_2 \) such that

\[
T^{k_1}([0, b^{-n})) = T^{k_2}([0, b^{-n})).
\]

However, in this case we would have that for each \( n \in \mathbb{Z} \)

\[
T^n((0, b^{-n})) = T^{k_1 + ((n - k_1) \mod (k_2 - k_1))}((0, b^{-n})).
\]

Thus there is a set \( G = \{T^k([0, b^{-n})) : k_1 \leq k < k_2\} \) such that

\[
O_T([0, b^{-n})) \subseteq G \\
\neq \left\{ \left[ \frac{k}{b^n}, \frac{k + 1}{b^n} \right) : k \in \{0, 1, \ldots, b^n - 1\} \right\}.
\]

This contradiction proves the result.

\( \diamond \)

Theorem 7.8

Suppose that \( T \) is a Kakutani type II transformation and that \( b \in \mathbb{N} \) is such that

\[
O_T([0, b^{-n})) \equiv \left\{ \left[ \frac{k}{b^n}, \frac{k + 1}{b^n} \right) : k \in \{0, 1, \ldots, b^n - 1\} \right\}
\]

Then the following statements hold.

(i) Let \( n \in \mathbb{N} \) and let \( x \in [0, b^{-n}) \). Then for all \( k \in \{0, 1, \ldots, b^n - 1\} \),

\[
T^k(x) \equiv x \mod b^{-n}.
\]

(ii) If \( x \in X \), \( O_T(x) = x \oplus \mathbb{Q}_b \).

299
Proof:

(i) Let \( n \in \mathbb{N} \) be given. Then, if \( k = 0 \), the result holds. Consider when \( k = 1 \). Then as \( x \in [0, b^{-n}) \) we know, by Corollary 7.2, that \( T(x) = x \oplus \beta(x) \) for some \( b \)-adic \( \beta(x) \). We also know that on \([0, b^{-n})\), \( T \) is a linear transformation with slope 1 and hence is defined on \([0, b^{-n})\) by \( T(y) = y + \beta(x) \). Also as \( T \) is a Kakutani type II transformation, \( T([0, b^{-n})) \) is a \( b \)-adic interval of order \( n \). This interval has a lower endpoint of the form \( k/b^n \) and hence

\[
T(0) = 0 + \beta(x) = 0 + \frac{k}{b^n},
\]

so that \( \beta(x) = k/b^n \). Therefore

\[
T(x) = x + \frac{k}{b^n} \equiv x \pmod{b^{-n}}.
\]

Now assume that \( k \in \{0, 1, 2, \ldots, b^n - 2\} \) and that

\[
T^k(x) \equiv x \pmod{b^{-n}}.
\]

so that

\[
T^k(x) = x + \frac{m}{b^n}
\]

for some \( m \in \mathbb{N} \). Then, as \( k \neq b^n - 1 \),

\[
T^k([0, b^{-n})) = \left[ \frac{l}{b^n}, \frac{l + 1}{b^n} \right),
\]

for some \( l \in \{0, 1, \ldots, b^n - 2\} \), and hence

\[
T^k(x) \in \left[ \frac{l}{b^n}, \frac{l + 1}{b^n} \right).
\]

Also

\[
T^{k+1}([0, b^{-n})) = \left[ \frac{j}{b^n}, \frac{j + 1}{b^n} \right)
\]

for some \( j \in \{0, 1, \ldots, b^n - 1\} \). Now \( T \) is linear over \([l/b^n, (l + 1)/b^n)\) with \( T(l/b^n) = j/b^n \) and therefore \( T \) is defined by \( T(y) = y \oplus (j - 1)/b^n \) for all
\[ y \in [l/b^n, (l + 1)/b^n). \] Therefore

\[
T^{k+1}(x) = T(T^k(x)) \\
= T \left( x + \frac{m}{b^n} \right) \\
= \left( x + \frac{m}{b^n} \right) \oplus \frac{j - l}{b^n} \\
= x (\text{mod } b^{-n}).
\]

Thus (i) follows by induction.

(ii) We know from Corollary 7.2 that \( O_T(x) \subseteq x \oplus \mathbb{Q}_b \) and hence we need to prove that \( x \oplus \mathbb{Q}_b \subseteq O_T(x) \).

Let \( x \in X \) and \( d \in \mathbb{Q}_b \). Then there is an \( n \in \mathbb{N} \) and \( p \in \{0, 1, ..., b^n - 1\} \) such that \( d = p/b^n \). Also there exist \( q \in \{0, 1, ..., b^n - 1\} \) and \( r \in [0, b^{-n}) \) such that

\[ x = \frac{q}{b^n} + r. \]

Then,

\[ x \oplus d = \frac{m}{b^n} + r, \]

where \( m = p + q \) if \( x + d \in X \), and \( m = p + q - b^n \) if \( x + d \notin X \). Note that \( m \in \{0, 1, ..., b^n - 1\} \). By Lemma 7.3 there is a \( k \in \{0, 1, ..., b^n - 1\} \) such that

\[ T^k([0, b^{-n})) = \left[ \frac{m}{b^n}, \frac{m + 1}{b^n} \right). \]

Hence, there is \( s \in [0, b^{-n}) \) such that

\[ T^k(r) = \frac{m}{b^n} + s. \]

But as \( r \in [0, b^{-n}) \), part (i), already proved, shows that there is an \( l \in \mathbb{Z} \) such that

\[ T^k(r) = \frac{l}{b^n} + r. \]
Hence

\[ 0 = T^k(r) - T^k(r) \]
\[ = \frac{m}{b^n} + s - \frac{l}{b^n} - r \]
\[ = \frac{m - l}{b^n} + s - r, \]

so that

\[ \frac{m - l}{b^n} = r - s \in (-b^{-n}, b^{-n}). \]

But as \( m - l \in \mathbb{Z} \), we deduce that \( m - l = r - s = 0 \) so that \( r = s \). Thus,

\[ T^k(r) = \frac{m}{b^n} + r = x \oplus d. \]

Also, again by Lemma 7.3, there is a \( k_1 \in \{0, 1, \ldots, b^n - 1\} \) such that

\[ T^{k_1}([0, b^{-n})) = \left[ \frac{q}{b^n}, \frac{q + 1}{b^n} \right). \]

Hence, there is \( s_1 \in [0, b^{-n}) \) such that

\[ T^{k_1}(r) = \frac{q}{b^n} + s_1. \]

Also, by part (i) already proved, there is an \( l_1 \in \mathbb{Z} \) such that

\[ T^{k_1}(r) = \frac{l_1}{b^n} + r. \]

As before, we can deduce that \( r = s_1 \) and so we have

\[ T^{k_1}(r) = \frac{q}{b^n} + r = x. \]

Thus, we have

\[ T^k(r) = x \oplus d \text{ and } T^{k_1}(r) = x. \]

Hence,

\[ T^{k-k_1}(x) = x \oplus d, \]

and therefore \( x \oplus d \in O_T(x) \). Thus \( x \oplus \mathbb{Q}_b \subseteq O_T(x). \) \( \diamond \)
7.4 Transformation invariant non-measurable sets and tower extensions

In the previous section we showed that there were transformation invariant non-measurable sets with desirable properties for certain transformations. Each of these transformations was a 'primitive' transformation of the transformations that have been considered in this thesis. To show that appropriate non-measurable sets exist in association with each of the tower extensions of the transformations that we have considered, we need to show that each of the tower extensions of these primitive transformations also has associated non-measurable sets. We show in this section that the existence of an appropriate non-measurable set for a primitive transformation is equivalent to the existence of one for any tower extension on the primitive transformation. This result concludes the chapter and therefore the Thesis.

Definition 7.3

Let \((X_0, \mathcal{B}_0, \mu_0)\) be a measure space of finite measure, \(\mathcal{J}_0\) be a collection of standard sets for \((X_0, \mathcal{B}_0, \mu_0)\), \(\psi_0\) be a (primitive) transformation and let \((X_m, \mathcal{B}_m, \mu_m, \psi_m)\) be the usual tower extension of either a finite or infinite type. Then the usual extension of \(\mathcal{J}_0\), \(\mathcal{J}_m\), is

\[
\mathcal{J}_m = \left\{ J = J_0 \cup \bigcup_{i=1}^{\alpha} \bigcup_{j=1}^{\alpha_i} \tau_i^{-j}(J_i) : J_0, \ldots, J_\alpha \in \mathcal{J}_0 \right\}
\]

where \(\alpha\) is the number of levels in the tower.

Theorem 7.9

Let \((X_0, \mathcal{B}_0, \mu_0)\) be a measure space of finite measure, \(\mathcal{J}_0\) be a collection of standard sets for \((X_0, \mathcal{B}_0, \mu_0)\) and \(\psi_0\) be an outer measure preserving bijective transformation on \(X_0\). Let \((X_m, \mathcal{B}_m, \mu_m, \psi_m)\) be the usual tower extension of either a finite or infinite type and let \(\mathcal{J}_m\) be the usual extension of \(\mathcal{J}_0\)
If \( V \) is an example of the required type of non measurable set in \( X_0 \), then
\[
V_m = \bigcup \{ \psi_m^n(V) : n \in \mathbb{Z} \}
\]
is an example of a non-measurable set of the required type in \( X_m \).

Conversely, if \( V \) is a non measurable set of the required type in \( X_m \) then
\( V \cap X_0 \) is a non-measurable set of the required form in \( X_0 \).

**Proof:**

Let \( \alpha \) denote the number of levels in the tower and note that we use the usual notation in the construction of the tower (that is \( X_i = X_0 \cup \bigcup_{j=1}^{i} B_j \) etc) and let \( B_0 = X_0 \). Suppose \( V \) is a non measurable set of the required type in \( X_0 \), so that \( V \subset X_0 \), \( V \) is \( \psi_0 \)-invariant and for each \( J \in \mathcal{J}_0 \),

\[
\mu_0^*(V \cap J) + \mu_0^*(V^c \cap J) = 2\mu_0(J).
\]

Let

\[
V_m = \bigcup_{n \in \mathbb{Z}} \psi_m^n(V),
\]

so that \( V_m \) is \( \psi_m \)-invariant. Now let \( J \in \mathcal{J}_m \), for each \( i \in \{1, ..., \alpha\} \) note that

\[
o_j^{-1}(J \cap B_i) = J_i \in \mathcal{J}_0.
\]

Then, using the fact that

\[
\mu_0^*(V \cap J) + \mu_0^*(V^c \cap J) = 2\mu_0(J),
\]

for each \( J \in \mathcal{J} \), we have

\[
\mu_*(J \cap B_i \cap V) + \mu_*(J \cap B_i \cap V^c)
\]

\[
= \mu_*(\omega_{j=0}^{i-1}(J \cap B_i \cap V)) + \mu_*(\omega_{j=0}^{i-1}(J \cap B_i \cap V^c))
\]

\[
= \mu_*(J_{B_i} \cap V) + \mu_*(J_{B_i} \cap V^c)
\]

\[
= 2\mu(J_{B_i}).
\]

Of course \( J \cap X_0 = J \cap B_0 \in \mathcal{J}_0 \) and hence

\[
\mu_*(J \cap B_0 \cap V) + \mu_*(J \cap B_0 \cap V^c) = 2\mu(J_{B_0}).
\]
Then, since $X_0, B_1, \ldots, B_\alpha$ is a disjoint countable sequence of measurable sets, we can use Lemma 6.1 to obtain

$$
\mu_\ast(J \cap V) + \mu_\ast(J \cap V^c) = \mu_\ast\left(\bigcup_{i=0}^{\alpha} (J \cap B_i \cap V)\right) + \mu_\ast\left(\bigcup_{i=0}^{\alpha} (J \cap B_i \cap V^c)\right)
$$

$$
= \sum_{i=0}^{\alpha} \mu_\ast(J \cap B_i \cap V) + \sum_{i=0}^{\alpha} \mu_\ast(J \cap B_i \cap V^c)
$$

$$
= \sum_{i=0}^{\alpha} 2\mu(J_B)
$$

$$
= \sum_{i=0}^{\alpha} 2\mu(J \cap B_i)
$$

$$
= \mu\left(\bigcup_{i=0}^{\alpha} (J \cap B_i)\right)
$$

$$
= 2\mu(J).
$$

which proves the first part of the proof.

For the converse suppose that $V \subset X_m$ is a set of the required non measurable form so that $V$ is $\psi_m$-invariant and for each $J \in \mathcal{J}_m$

$$
\mu_\ast(V \cap J) + \mu_\ast(V^c \cap J) = 2\mu_m(J).
$$

By either Lemma 5.7 or Lemma 5.9, depending on whether the tower extension is finite or infinite, $V \cap X_0$ is $\psi_0$-invariant.

As each $J \in \mathcal{J}_0$ is also in $\mathcal{J}_m$ it follows immediately that for each $J \in \mathcal{J}_0$

$$
\mu_\ast(V \cap J) + \mu_\ast(V^c \cap J) = 2\mu_m(J).
$$

Hence $V \cap X_0$ is a non-measurable set of the required form and so the theorem is proved. \hfill \diamondsuit
7.5 Notes

This chapter is essentially original work although much of it follows the work that is presented in [27]. A version of each of Propositions 7.1, 7.2, 7.3, 7.4 and 7.5, Lemma 7.1 and Theorems 7.1, 7.2, and 7.8 is presented in [27]. Zorn’s Lemma, equivalent to the Axiom of Choice in presented in Zorn [49]. The remaining results are original to this work. Further reading on non-measurable sets can be found in Eoin [10], Federer [11], Halperin [16], Kakutani and Oxtoby [23] and Wagon [45]. In particular, Itzkowitz and Shahkmatov [20] give a discussion on similar topics to those that we present here as a representation of those in [27]. Also, Pierański and Wojciechowski [36] present a similar result to irrational rotation invariant sets but for Tori.
8 Bibliography


[28] Kronecker, L.: Vorlesungen über Zahlen theorie (Reprint), Springer-
Verlag, 1978.


