DISSERTATION

A RATIO ERGODIC THEOREM ON BOREL ACTIONS OF $\mathbb{Z}^d$ AND $\mathbb{R}^d$

Submitted by

Eric Norman Holt

Department of Mathematics

In partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2009

COLORADO STATE UNIVERSITY

June 17, 2009

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ERIC NORMAN HOLT ENTITLED "A RATIO ERGODIC THEOREM ON BOREL ACTIONS OF $\mathbb{Z}^d$ AND $\mathbb{R}^d$" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Dr. Gerhard Dangelmayr

Dr. Donald Estep

Dr. Tim Penttila

Dr. Hoanan Wang

Adviser: Dr. Daniel Rudolph

Department Head: Dr. Simon Tavener

ABSTRACT OF DISSERTATION

A RATIO ERGODIC THEOREM ON BOREL ACTIONS OF $\mathbb{Z}^d$ AND $\mathbb{R}^d$

We prove a ratio ergodic theorem for free Borel actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$ on a standard Borel probability space. The proof employs an extension of the Besicovitch Covering Lemma, as well as a notion of coarse dimension that originates in an upcoming paper of Hochman. Due to possible singularity of the measure, we cannot use functional analytic arguments and therefore diffuse the measure onto the orbits of the action. This diffused measure is denoted μ_x, and our averages are of the form $\frac{1}{\mu_x(B_n)}\int_{B_n} f \circ T^{-v}(x)d\mu_x$. A Følner condition on the orbits of the action is shown, which is the main tool used in the proof of the ergodic theorem. Also, an extension of a known example of divergence of a ratio average is presented for which the action is both conservative and free.

Eric Norman Holt
Department of Mathematics
Colorado State University
Fort Collins, Colorado 80523
Summer 2009

ACKNOWLEDGEMENTS

All that I have is a gift from someone else. I here try to express gratitude to all of those that have given a major contribution to my educational achcivements, though I will undoubtedly leave some out.

I extend a warm thank you to my parents Norman and Joyce for their constant support of my education. I am very grateful to the excellent educators at Shoals and Mitchell schools in Indiana and especially thank Mr. Michael Dorsett for modeling the type of challenging, enjoyable mathematics classroom that I hope to be able to create for my future students. I thank Dr. James Lesko and Dr. Elizabeth Brown for their insightful career advice. I would like to thank my graduate advisor, Dr. Daniel Rudolph, who invested uncountable hours training me and guiding me in this project, and I am certainly a better mathematician because of him. I cannot express enough appreciation to my wife, Jennifer, for her encouragement and availability to help out in any way possible. I would like to thank my Heavenly Father for teaching me both patience with myself and perseverance. Finally, thank you, whoever you are, for taking the time to look at the ideas I have laid out in this dissertation!

TABLE OF CONTENTS

LIST OF SYMBOLS

1. $F = \mathbb{Z}^d$ or $\mathbb{R}^d$.
2. $||v|| =$ the Euclidean norm of v in F.
3. $B_\rho = \{v \in F : ||v|| < \rho\}$.
4. $B_\rho(w) = \{v \in F : ||v - w|| < \rho\}$.
5. $\partial_r B_\rho = B_\rho \setminus B_{\rho - r}$.
6. $\partial_r^* B_\rho = B_{\rho + r} \setminus B_{\rho - r}$.
7. $\mathrm{cdim}_{R_0} Y$: see Definition 2.2.7.
8. $(X, \mathcal{F}, \mu) =$ a standard Borel probability space.
9. $T =$ a Borel action of F on $(X, \mathcal{F}, \mu)$.
10. $I =$ the function that takes $X \times F$ to $X \times F$ by $I(x, v) = (T^v(x), v)$.
11. $m =$ Haar measure on F.
12. $\hat{\mu}_N = \frac{1}{m(B_N)} I^*(\mu \times m)|_{X \times B_N}$.
13. $\mathcal{H}_N = \{A \times B_N : A \in \mathcal{F}\}$.
14. $\hat{\mu}_{x,N}(E) =$ a version of $E_{\hat{\mu}_N}(X \times E|\mathcal{H}_N)$, where $E \subset B_N$ is Borel.
15. $N_x =$ the smallest N such that $\hat{\mu}_{x,N}(B_N) > 0$.
16. $\mu_x = \lim_{N \to \infty} \dfrac{\hat{\mu}_{x,N}}{\hat{\mu}_{x,N}(B_{N_x})}$.
17. $X_{0,N} = \{x \in X : \hat{\mu}_{x,N}(B_N) > 0\}$.
18. $X_0 = \{x \in X : \mu_x(F) > 0\}$.
19. $A_0 = \{x \in X_0 : \lim_{R \to \infty} \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} = 0$ for all $r > 0\}$.
20. $[0,1)^* = [0,1) \setminus \{\frac{k}{2^n} : k \in \mathbb{N} \cup \{0\}, n \in \mathbb{N}\}$.
21. $\mathcal{L}^* = \{A \cap [0,1)^* : A \in \mathcal{L}\}$.

22. $f =$ an $L^1(\mu)$ function, which is usually assumed to be nonnegative.

23. $\theta =$ a measure on $(X, \mathcal{F}, \mu)$ defined by $\theta(A) = \int_A f d\mu$.

24. $\theta_x =$ the measure θ diffused on the orbits of T.

25. $Y_0' = \{x \in X : \theta_x(F) > 0\}$.

26. $Y_0 = Y_0' \cup \{x : f(T^v(x)) = 0$ for all v outside some set of m measure zero$\}$.

27. $A_n(f, x) = \frac{1}{\mu_x(B_N)} \int_{B_N} f \circ T^v(x) d\mu_x$.

28. $\nu_x =$ a measure equivalent to μ_x for which $A_n(f, x) = \frac{\theta_x(B_N)}{\nu_x(B_N)}$.

29. $A_{\alpha,\beta} = \{x : \liminf_{n \to \infty} A_n(f, x) < \alpha < \beta < \limsup_{ntoo} A_n(f, x)\}$.

30. $A^*_{\alpha,\beta} =$ a subset of $A_{\alpha,\beta}$ that is invariant and has the same μ measure as $A_{\alpha,\beta}$.

Chapter 1

Introduction

The first milestones in ergodic theory come in the early 1930's when von Neumann [25] and Birkhoff [4] each publish an ergodic theorem. Both results assume a measure-preserving, invertible transformation on a σ-finite measure space. A few years later Hopf extends Birkhoff's result for L^1 functions by proving a ratio ergodic theorem [9]. Hopf's theorem is presented in terms of a weighted average, but can also be seen as generalizing the measure-preserving requirement to nonsingular transformations. In 1944, Hurewicz gives an even further generalization by proving an ergodic theorem that allows for singularity of the system [10]. The results of both Hopf and Hurewicz assume conservativity. It is five decades until another major pointwise ergodic theorem of this form is presented. In 2007, Feldman [6] uses a maximal inequality proven by Lindenstrass and Rudolph [14] to show an ergodic theorem for non-singular actions of $\mathbb{Z}^d$. The multidimensional group $\mathbb{Z}^d$ requires him to average over hypercubes $[-n, n]^d$, which are centered at the origin, rather than the most natural extension of the earlier results, which is to average over $[0, n]^d$. This is necessary because an example of divergence is known for the latter type of average with $d > 1$ [13]. Hochman recently extended Feldman's Ratio Ergodic Theorem by removing the assumption of conservativity on both the action and the components [8].

The climax of this dissertation is the proof of a ratio ergodic theorem for Borel actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$ which assumes neither nonsingularity nor conservativity. This first, introductory chapter gives a foundation to the ergodic theorem by describing the ergodicity condition and reviewing the classical ergodic theorems that pertain to pointwise convergence. Chapter 2 presents the Besicovitch Covering Lemma, as well as a recent extension of the Besicovitch Covering Lemma due to Hochman. In Chapter 3, the measure of a standard Borel probability space is diffused onto the orbits of a free Borel $\mathbb{Z}^d$ or $\mathbb{R}^d$ action. Such diffusion of a measure has a long history in the setting of the leaves of a foliation, but was introduced by Lindenstauss and Rudolph in the case of a Borel action [14]. Also in this chapter, a Følner condition of the diffused measure on the orbits of such an action is proven, which is the first original work presented. The second such work is an extension of the Krengel and Brunel example of divergence of ratio averages so that the system is free and conservative, which is in Chapter 4. Also, the ratio ergodic theorems of Feldman and Hochman are reviewed here. Chapter 5 includes the statement and proof of the main result, as well as a few examples. Finally, Chapter 6 describes questions which arise from the results presented in this dissertation.

1.1 Ergodicity

The study of dynamical systems, at its most basic level, is the branch of mathematics that deals with values which change in time. A state space and a discrete or continuous transformation are used to quantify this idea. Physicists introduced the ergodic hypothesis, which is the notion that statistical properties of the system over time in a single experiment will be the same as the statistical properties across the state space [16]. This probably came about by observations that the statistics are indeed often the same. One basic statistical analysis one can do is an average, or expectation. Ergodic theory began as the study of such averages, though naturally the field has grown to include subjects that extend beyond the scope of computing average values. Ergodic theorems, in turn, are results concerning the existence and value of a time average under certain conditions.

Consider a mass attached to a spring in a frictionless system. For the state space, one must consider velocity or momentum as well as the position of the mass. This is necessary so that the system is deterministic: knowing the value at one time enables us to compute the value at any other time. The state space, then, is $I \times \mathbb{R}$ for some open interval I (for our purposes, what happens at the endpoints of this interval is not important). We may use either a discrete time transformation, representing measurement at subsequent intervals of a constant amount of time, or a continuous time transformation. An orbit is a circle for the continuous time case and is an at most countable subset of a circle for the discrete time case. Suppose that $f(x)$ is the total energy of the system at state x. By conservation of energy we have that f remains constant under the transformation. Thus, the average value of f over time in a particular experiment is clearly not the same as the expected value of f across the state space (the former is a number whereas the latter is infinite). On the other hand, if the state space is restricted to a certain energy level, then the average value of f over time in one experiment is the same as the expected value of f over the state space.

This example demonstrates the notion of ergodicity, which is the condition that time averages equal space averages. Also seen in this example is the ergodic decomposition: any system can be broken up into ergodic components. To make this notion of ergodicity more precise, suppose we have a state space X. The state space is assumed to come with a natural probability measure μ on the σ-algebra $\mathcal{F}$ of X. This allows us to write down a space average:

$$\int_X f d\mu.$$

We assume a discrete time system and let $T : X \to X$ be an $\mathcal{F}$ measurable time evolution map that gives evolution over one unit of time. We then formulate a time average:

$$A_n(f) := \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k.$$

This is known as the Cezáro average, and is only one of many time averages that can be considered. Ergodicity, in this case, is the property that $A_n(f)$ converges to $\int f$. The convergence may be pointwise, L^1, L^p, or uniform, for example.

In application of the results, we may wonder where the probability measure μ comes from. Indeed, one natural way to find μ is just to run the experiment and let $\mu(A)$ be the proportion of times that the state is in A. This, however, is just computing a time average, so the space average equals the time average for free (assuming that the average converges). Another way to approach this issue is to search for a measure that is invariant under T. Suppose the state space is compact and the map T is continuous. We can start with *any* probability measure ν and look at the sequence of measures

$$\nu_n := \frac{1}{n}\sum_{k=0}^{n-1} \nu \circ T^k.$$

We don't know if these measures converge, but we do know that a convergent subsequence exists. If we start with a system in $\mathbb{R}^n$ and the Lebesgue measure, then the subsequential limits of $\{\nu_n\}$ are called Krylov-Bogolubov measures. Any weak* subsequential limit of ν_n is an invariant measure and therefore a reasonable choice for μ [5].

The above formulation of the condition of ergodicity is a property of the set $\{f, (X, \mathcal{F}, \mu), T\}$. However, the usual definition of ergodicity is a property of the set $\{(X, \mathcal{F}, \mu), T\}$ and looks quite different.

Definition 1.1.1. *The system (X, T) is said to be* **ergodic** *if $T^{-1}(A) = A$ implies $\mu(A) = 0$ or 1 for $A \in \mathcal{F}$.*

This definition appears to have nothing to do with space averages and time averages. How can the two notions of ergodicity be reconciled? Suppose that for all $f \in L^1$, $A_n(f) \to \int f d\mu$ almost surely. Further, suppose $A \in \mathcal{F}$ is T invariant

$(T^{-1}(A) = A)$ and $\mu(A) > 0$. Letting $f = 1_A$ gives

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_A\left(T^k(x)\right) = \mu(A) \tag{1.1.1}$$

for a.e. x. Choose $x_0 \in A$ for which (1.1.1) holds. Now $1_A(T^k(x_0)) = 1$ for all $k \in \{0, 1, 2, ...\}$ and we see $\mu(A) = 1$. Therefore, T-invariant sets must have measure zero or one, and the system is ergodic. So, we see how space averages equalling time averages implies Definition 1.1.1. In the next section, we will see the converse: Definition 1.1.1 implies space averages equal time averages.

1.2 Early Ergodic Theorems

We now look at some early ergodic theorems.

The most basic ergodic theorem is the Law of Large Numbers, which says that the average value of a Bernoulli random variable converges to the expected value as the number of trials goes to infinity. Although the law is stated in the context of random variables rather than dynamical systems, the language and result are easily transferrable. The function f is the random variable and the state space is all infinite sequences of samples. We endow the state space with a Bernoulli measure, which is based on the original measure of the sample space. For example, suppose a fair die is rolled repeatedly. The Law of Large Numbers says us that, almost surely, the ratio of fours rolled to total number of rolls converges to $\frac{1}{6}$, since $\frac{1}{6}$ is the expected value, or space average, of the random variable (which takes the value 1 when a four is rolled and 0 otherwise) [11, 3].

Next we turn our attention to the von Neumann Ergodic Theorem and the Birkhoff Ergodic Theorem. As with the various laws of large numbers, the difference between these two theorems is the type of convergence that is asserted. In this dissertation, we are concerned with ergodic theorems that give pointwise convergence. The von Neumann theorem gives L^2 convergence, but is included because it is in this setting that the identity of the function which the time averages converge to is the clearest.

Theorem 1.2.1. *(von Neumann Ergodic Theorem) [25] Suppose $(X, \mathcal{F}, \mu)$ is a σ-finite measure space and T is a measurable, measure-preserving transformation. Then for $f \in L^2(\mu)$, there is an $\bar{f} \in L^2(\mu)$ for which*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to \bar{f}$$

in $L^2(\mu)$.

We notice a special property of $\bar{f}$: it is T-invariant. To see this, we compare the time averages of f and $f \circ T$ in L^2:

$$\begin{aligned}\int \left| \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k - \frac{1}{n}\sum_{k=0}^{n-1} (f \circ T) \circ T^k \right|^2 d\mu &= \int \left| \frac{1}{n}(f - f \circ T^n) \right|^2 d\mu \\ &\leq \frac{2}{n^2}||f||_2,\end{aligned}$$

and the right hand side goes to zero as n approaches infinity. The von Neumann Ergodic Theorem, then, gives that

$$\frac{1}{n}\sum_{k=0}^{n-1} ((f \circ T) \circ T^k) \to \bar{f}$$

in $L^2(\mu)$. In other words, $\overline{f \circ T} = \bar{f}$.

This property is notable because it is the main tool used in characterizing $\bar{f}$. The proof of the von Neumann Ergodic Theorem is constructive. It shows that $\bar{f}$ is the projection of f onto the L^2-subspace of T-invariant functions.

The Birkhoff Ergodic Theorem came soon after the von Neumann theorem (even though the publication dates imply the opposite).

Theorem 1.2.2. *(Birkhoff Ergodic Theorem) [4] Suppose $(X, \mathcal{F}, \mu)$ is a probability space, T is a measurable and measure-preserving transformation on X, and $f \in L^1(\mu)$. Then*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \to \bar{f} \tag{1.2.1}$$

both a.e. and in $L^1(\mu)$, and $\bar{f}$ is the conditional expectation of f given the σ-algebra of T-invariant sets.

We can use the Birkhoff Ergodic Theorem to show the other direction in relating the two notions of ergodicity given above, having already seen that time averages equalling space averages implies invariant sets have measure zero or one. Now suppose that the pair $\{(X, \mathcal{F}, \mu), T\}$ is ergodic (Definition 1.1.1) and $f \in L^1(\mu)$. The conditional expectation of f given the σ-algebra of T-invariant sets is just the expected value of f in this case. Therefore, the Birkhoff Ergodic Theorem says that time averages converge to the expected value of the function almost surely.

The Birkhoff theorem is often stated on a σ-finite measure space, although the characterization of the limit does not hold in this case since the conditional expectation is not defined for infinite measures. If there is a T-invariant subset of positive, finite measure, then one can restrict the space to this set and apply the characterization given in Theorem 1.2.2. If no such set exists, then $\bar{f}$ is 0 a.e [20].

We will need a new definition to continue our review of the early ergodic theorems.

Definition 1.2.3. *The system $\{(X, \mathcal{F}, \mu), T\}$ is* **conservative** *if $\mu(T^{-k}(A) \cap A) = \emptyset$ and for all $k \in \mathbb{N}$ and $A \in \mathcal{F}$ implies $\mu(A) = 0$.*

Hopf extended Birkhoff's theorem, which was in turn extended by Hurewicz. We now review these two results. Again, our statement will assume a probability space, although convergence holds in both cases for σ-finite measure spaces. (The σ-finite case adds an assumption of conservativity for Hopf's theorem, whereas conservativity is automatic for the measure-preserving transformation on a probability space, which is a condition of Theorem 1.2.4.)

Theorem 1.2.4. *(Hopf Ergodic Theorem) [9] Suppose $(X, \mathcal{F}, \mu)$ is a probability space, T is a measurable, measure-preserving transformation of X, $f \in L^1(\mu)$, and*

$g : X \longrightarrow \mathbb{R}$ is measurable with respect to the Lebesgue σ-algebra on $\mathbb{R}$ and positive almost everywhere. Then

$$\frac{\sum_{k=0}^{n} f \circ T^k}{\sum_{k=0}^{n} g \circ T^k} \underset{n\to\infty}{\longrightarrow} \frac{E(f|\mathcal{I})}{E(g|\mathcal{I})} \tag{1.2.2}$$

almost everywhere, where $\mathcal{I}$ is the σ-algebra of T invariant sets.

We notice that this is a generalization of the Birkhoff Ergodic Theorem by taking g to be the constant function 1. Since the Hopf theorem involves a ratio of sums, it is known as a ratio ergodic theorem.

A little over a decade after Birkhoff's Ergodic Theorem and seven years after Hopf's Ergodic Theorem, Hurewicz proved an even more general result removing the assumption that the system be measure-preserving. In fact, Hurewicz allowed the system to be singular: we may have $A \in \mathcal{F}$ with $\mu(A) = 0$ and $\mu(T^{-1}(A)) > 0$.

Rather than assuming a function $f \in L^1(\mu)$ as Birkhoff and Hopf do, Hurewicz starts with a countably additive set function F on $\mathcal{F}$ which is absolutely continuous with respect to μ. This countably additive set function is just a signed measure on $\mathcal{F}$. Taking $f = \frac{dF}{d\mu}$ (see [17] for the Radon-Nikodym Theorem on signed measures) and $g = \frac{d\mu\circ T}{d\mu}$, this is the Hopf theorem.

Theorem 1.2.5. *(Hurewicz Ergodic Theorem) [10] Suppose $(X, \mathcal{F}, \mu)$ is a probability space and T is a measurable and measurably invertible transformation of X. Let F be a finite, countably additive set function on $\mathcal{F}$ which is absolutely continuous with respect to μ, and consider the point densities*

$$f_n := \frac{d(\sum_{k=0}^{n} F \circ T^k)}{d(\sum_{k=0}^{n} \mu \circ T^k)}. \tag{1.2.3}$$

If the system is conservative, then f_n converges a.e.

Recent work on the ratio ergodic theorem has considered an action of $\mathbb{Z}^d$ or $\mathbb{R}^d$ instead of a transformation T. Nevertheless, the Birkhoff, Hopf, and Hurewicz ergodic theorems provide a foundation for the more recent ratio ergodic theorems. The arguments used on the actions of higher dimensional groups are very similar to those used in these early ergodic theorems. We return to the ratio ergodic theorem in Chapter 4, but first build a few tools for our proof in the next few chapters.

Chapter 2

Covering Lemmas

Covering lemmas play a crucial role in various arguments in Ergodic Theory. The most basic such lemma is that if I_1, I_2, and I_3 are intervals and $I_1 \cap I_2 \cap I_3$ is nonempty, then one of the intervals may be discarded so that the remaining two intervals cover the same set as the original three. This is essentially the Besicovitch Covering Lemma for $\mathbb{R}$. In this chapter, we prove the Besicovitch Covering Lemma and then move on to an extension by Hochman. All results in this chapter are presented in $\mathbb{R}^d$, but the results immediately follow for $\mathbb{Z}^d$ as well (using Haar instead of Lebesgue measure). In both settings, the Euclidean metric is used. A ball in $\mathbb{R}^d$ with radius ρ and center c, denoted $B_\rho(c)$, here means $\{x \in \mathbb{R}^d : d(x, c) < \rho\}$. If the center is not specified, then it is assumed to be the origin.

2.1 The Besicovitch Covering Lemma

Theorem 2.1.1. *(The Besicovitch Covering Lemma) [2, 24] For $\mathbb{R}^d$, there is a natural number C such that the following holds. If $E \subset \mathbb{R}^d$ is bounded, and for all $v \in E$ we have a ball $B_{\rho(v)}(v)$ with $\rho(v) > 0$, then there exist subsets $E_1, ..., E_C \subset E$ such that $v_1, v_2 \in E_k$, $v_1 \neq v_2$ implies $B_{\rho(v_1)}(v_1) \cap B_{\rho(v_2)}(v_2) = \emptyset$, and*

$$E \subset \bigcup_{\substack{v \in E_k \\ 1 \leq k \leq C}} B_{\rho(v)}(v).$$

We will employ the following two lemmas in the proof of the Besicovitch Covering Lemma.

Lemma 2.1.2. *For all real numbers r and R, let K be the maximum number of disjoint balls of radius $\frac{r}{2}$ in $\mathbb{R}^d$ that can be placed inside a ball of radius $3R$ (notice K depends only on d and the ratio $\frac{R}{r}$). Then for any collection of $K+1$ vectors $v_0, v_1, v_2, ..., v_K \in \mathbb{R}^d$ and function $\rho : \{v_1, ..., v_K\} \to [r, R]$ with $d(v_i, v_j) \geq \rho(v_j)$ for all $0 \leq i < j \leq K$, there is a $1 \leq k \leq K$ such that $d(v_0, v_k) > \rho(v_0) + \rho(v_k)$.*

Proof. Fix r, R, and d. Let K be the maximum number of disjoint balls of radius $\frac{r}{2}$ that can be placed inside a ball of radius $3R$. Assume v_j and ρ are as stated. For the sake of contradiction, suppose that for all $0 \leq j \leq K$, $d(v_0, v_j) \leq \rho(v_0) + \rho(v_j)$. This implies that the collection of balls $\{B_{\frac{r}{2}}(v_j) : 0 \leq j \leq K\}$ is pairwise disjoint and contained in $B_{3R}(v_0)$, which contradicts the definition of K.

□

The next lemma is rather technical.

Lemma 2.1.3. *Suppose $d \geq 2$ and $a, b, c \in \mathbb{R}^d$ are not collinear. Let $d_1 := d(a,b), d_2 := d(b,c), d_3 := d(a,c)$, and $\psi := \angle abc$ (see Figure 2.1). Also, suppose positive real numbers $\rho(a), \rho(b), \rho(c)$ and the following constraints:*

1. *$d_1 \in [\rho(a), \rho(a) + \rho(b))$ and $d_2 \in [\rho(c), \rho(c) + \rho(b))$,*

2. *$d_3 \geq \rho(a)$,*

3. *$\rho(a) \geq \frac{49}{50}\rho(c)$, and*

4. *$10\rho(b) \leq \rho(c)$.*

Then $m(\psi) \geq \frac{\pi}{8}$.

Proof. Suppose, for the sake of contradiction, that the assumptions hold and $m(\psi) < \frac{\pi}{8}$. Thus, $\cos(\psi) \geq \frac{9}{10}$, and we apply the law of cosines to $\triangle abc$ and estimate the

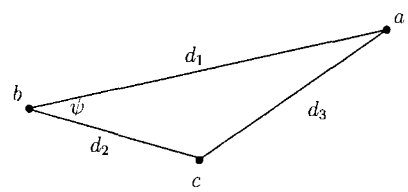

Figure 2.1: The setup for Lemma 2.1.3.

quantities involved:

$$\begin{aligned} (d_3)^2 &= (d_1)^2 + (d_2)^2 - 2d_1 d_2 \cos\psi \\ (\rho(a))^2 &< \left(\frac{54}{49}\rho(a)\right)^2 + \left(\frac{11}{10}\rho(c)\right)^2 - 2\rho(a)\rho(c)\frac{9}{10}. \end{aligned}$$

We combine like terms, move the last term to the left hand side, and divide by $\rho(a)\rho(c)$ to get

$$\frac{9}{5} < \frac{515}{2,401} \cdot \frac{\rho(a)}{\rho(c)} + \frac{121}{100} \cdot \frac{\rho(c)}{\rho(a)}.$$

which implies $\rho(a) > 7\rho(c)$ by assumption 3. Applying assumptions 1 and 4 gives $d_3 > \frac{63}{10}d_2$. Concavity of sin on $[0, \frac{\pi}{2}]$ and the law of sines gives $\frac{63}{10} < \frac{\sin(\psi)}{\sin(\angle bac)}$. Since $\frac{x}{y} - \frac{\sin(x)}{\sin(y)} > 0$ for $x \in (0, \frac{\pi}{8}), y \in (0, x)$, we have $\frac{63}{10} < \frac{m(\psi)}{m(\angle bac)}$.

Let e be a new point which is colinear with a and c such that $d(a, e) = d(a, b)$. Then $d(c, e) < \rho(b)$, while $d(b, e) > d_2 - \rho(b) > d(c, e)$ and $d(b, c) = d_2 > d(c, e)$. Thus, $\overline{ce}$ is the shortest leg of $\triangle bce$, so $m(\angle cbe) < \frac{\pi}{3}$. Since $\triangle abe$ is isosceles,

$$\begin{aligned} m(\psi) + m(\angle cbe) &= \frac{\pi - m(\angle bac)}{2} \\ &\geq \frac{\pi}{2} - \frac{10}{126}m(\psi), \end{aligned}$$

which implies $m(\psi) \geq \frac{63}{136}\pi - \frac{42\pi}{136}$. This contradicts $m(\psi) < \frac{\pi}{8}$. □

Proof. Now we prove Theorem 2.1.1. Suppose E and $\rho(v)$ are as stated.

It is not difficult to see that Theorem 2.1.1 holds for $d = 1$ by letting $C = 2$. We may assume, then, that $d \geq 2$. Let $s = \text{diam}(E)$ and K be the maximum number of balls of radius one that can be placed inside a ball of radius $\frac{300}{49}$. Also, let M be the maximum number of non-origin points that can be placed in $\mathbb{R}^d$ such that the measure of any angle with the origin as the vertex and one of these points on each leg is at least $\frac{\pi}{8}$. Let $C = 115K + M + 1$, which depends only on d. If there is a $v \in E$ with $\rho(v) > s$, then we let $E_1 = \{v\}$ and $E_j = \emptyset$ for $1 < j \leq C$, and the result holds. Assume, then, that $\rho(v) \leq s$ for all $v \in E$.

For each natural number i, let

$$G_i := \left\{ v \in E : \left(\frac{49}{50}\right)^i s < \rho(v) \leq \left(\frac{49}{50}\right)^{i-1} s \right\}.$$

Notice that $G_1, G_2, ...$ are pairwise disjoint and cover E.

We build the sets E_k simultaneously, moving through the sets G_i one at a time. For step one, choose $v_1 \in G_1$ (if G_i is empty, then skip to step two) and let $v_1 \in E_1$. Suppose $v_1, ..., v_{j-1}$ have been chosen and placed in an appropriate set E_k. Choose $v_j \in G_1 \setminus \cup_{1 \leq i < j} B_{\rho(i)}(v_i)$, skipping to step two if this set is empty. Let n_j be the smallest n such that for each $1 \leq k < n$, there is a $v \in E_k$ with $d(v_j, v) < \rho(v_j) + \rho(v)$ and place $v_j \in E_{n_j}$. Notice that $\{B_{\rho(v)}\}_{v \in E_k}$ is a pairwise disjoint collection of balls for each k. Also, the balls $\{B_{\frac{r_1}{2}}(v_j)\}$ are pairwise disjoint, where $r_1 = \frac{49}{50}s$. Since the set G_1 is bounded, this process terminates.

Now suppose steps one through $l - 1$ have been completed. Let k_l be such that v_{k_l} has been defined, but v_{k_l+1} has not (or, if no v_j has been defined, let $k_l = 0$). Also, let

$$G_l' := G_l \setminus \bigcup_{1 \leq j \leq k_l} B_{\rho(v_j)}(v_j).$$

Choose $v_{k_l+1} \in G_l'$, or skip to step $l + 1$ if G_l' is empty. Let n_{k_l+1} be the smallest n such that for each $1 \leq k < n$, there is a $v \in E_k$ with $d(v_{k_l+1}, v) < \rho(v_{k_l+1}) + \rho(v)$, and place $v_{k_l+1} \in E_{n_{k_l+1}}$. Now suppose v_i has been chosen and placed in an appropriate E_k for $k_l + 1 \leq i < k_l + j$. Choose $v_{k_l+j} \in G_l' \setminus \bigcup_{k_l+1 \leq i < k_l+j} B_{\rho(v_i)}(v_i)$, or skip to the next step if this set is empty. Let n_{k_l+j} be the smallest n such that for each $1 \leq k < n$, there is a $v \in E_k$ with $d(v_{k_l+j}, v) < \rho(v_{k_l+j}) + \rho(v)$ and place $v_{k_l+j} \in E_{n_{k_l+j}}$. Again, the balls $\{B_{\frac{r_l}{2}}(v_j)\}$ are pairwise disjoint, where $r_l = (\frac{49}{50})^l s$. Since G_l is bounded, step l terminates.

If we show that n_j is never larger than C, then the proof is complete. Suppose, for sake of contradiction, that the above procedure is carried out and there is some j at step l for which $n_j = C + 1$. This gives increasing natural numbers $i_1, i_2, ..., i_C$ such that $d(v_j, v_{i_k}) < \rho(v_j) + \rho(v_{i_k})$ for all $1 \leq k \leq C$.

Lemma 2.1.2 implies that at most K of the vectors v_{i_k} can be from a given G_m. Thus, $i_{C-115K} \leq k_{l-115} + 1$. Also, for all $1 \leq k \leq C - 115K$, $\rho(v_{i_k}) \geq 10\rho(v_j)$, since $(\frac{50}{49})^{114} > 10$ and $(\frac{50}{49})^{114}$ is the ratio of the lower bound for ρ on G_{l-115} to the upper bound for $\rho(v_j)$. Since $C - 115K > M$, we may choose distinct vectors w_1 and w_2 from $v_{i_1}, ..., v_{i_{C-115K}}$ such that $m(\angle w_1 v_j w_2) < \frac{\pi}{8}$. Without loss of generality, assume w_1 comes before w_2 in the list $v_{i_1}, ..., v_{i_{C-115K}}$. Apply Lemma 2.1.3 with $a = w_1, b = v_j$, and $c = w_2$ to get that $m(\angle w_1 v_j w_2) \geq \frac{\pi}{8}$. This is a contradiction, so n_j is never larger than C, and the proof is complete. $\square$

Given a set of points in $\mathbb{R}^d$ and a ball centered at each point, it would be helpful to reduce to a disjoint sub-collection of these balls that still cover the set. The Vitali Covering Lemma comes close to providing this, but the sub-collection given covers all but some positive fraction of the Lebesgue mass of the set. Since we will be working with measures that are not Lebesgue, this is not good enough for us. Instead, we use the Besicovitch Covering Lemma, which shows that there must be C collections of balls such that each collection has pairwise disjoint balls and the balls in the C collections together union to the entire set. This allows us to cover a fraction $\frac{1}{C}$ of the mass of the set, with respect the measure we use. Much work has been done to determine the value of C [7]. For our purposes, however, it will be sufficient that C is finite.

2.2 Hochman's Lemma

In this section, we follow a line of reasoning that extends the commonly used Besicovitch covering lemma on $\mathbb{R}^d$ so that, in each subcollection, two distinct balls are not only disjoint, but are no less than a certain positive distance apart. This is used to prove a statement that says balls with heavy boundaries cannot be too common if the measure on $\mathbb{R}^d$ is finite. The results and arguments in this section are due to Hochman [8]. It should be noted that Hochman's treatment includes more general metrics on $\mathbb{R}^d$.

First we start with some terminology.

Definition 2.2.1. *A collection of subsets of* $\mathbb{R}^d$ *has* **multiplicity** M *if every element of* $\mathbb{R}^d$ *is contained in at most* M *elements of the collection.*

For example, if $d = 1$, then the collection of sets $\{\{1, 4, 5\}, \{0, 1, 2\}, B_2(4), \mathbb{R}\}$ has multiplicity three since $1, 4$, and 5 each lie in three sets in the collection. Notice that this collection also has multiplicity M for any integer $M \geq 3$.

Definition 2.2.2. *A collection of balls* $\mathcal{U}$ *in* $\mathbb{R}^d$ *is* **well-separated** *if the distance between any two balls in* $\mathcal{U}$ *is at least* $\text{rmin}\mathcal{U} = \min_{B \in U}\{\text{radius}(B)\}$.

A collection being well-separated allows us to extend the radii of the balls in the collection by up to $\frac{rmin\mathcal{U}}{2}$ and still have a disjoint collection.

Suppose m is Lebesgue measure on $\mathbb{R}^d$. We have the following improvement of the Besicovitch Covering Lemma.

Lemma 2.2.3. *Suppose* $E \subset \mathbb{R}^d$ *is bounded and to each* $v \in E$ *there corresponds a ball* $B_{\rho(v)}(v)$ *with* $\rho(v) > 0$. *Then there exists a number* χ, *which depends only on* $\mathbb{R}^d$, *and a partition* $\{E_1, E_2, ..., E_\chi\}$ *of* E *such that each* E_j *is countable,* $\{B_{\rho(v)}(v) : v \in E_j\}$ *is well-separated for each* j, *and*

$$E \subset \bigcup_{\substack{v \in E_j \\ 1 \leq j \leq \chi}} B_{\rho(v)}(v).$$

Proof. Let C be the constant from the Besicovitch Covering Lemma. Also, let D be the doubling constant for $\mathbb{R}^d$: for any ball $B_r(v)$ in $\mathbb{R}^d$, $m(B_{2r}(v)) \leq D \cdot m(B_r(v))$.

Let $v \in \mathbb{R}^d$ and $R > 0$. Suppose $\{B_1, B_2, ..., B_n\}$ is a collection of balls with radii $\geq R$, centers in $B_{3R}(v)$, and multiplicity C. Shrinking each ball to radius R gives

$$n \cdot m(B_R(v)) \leq C \cdot m(B_{4R}(v)) \leq D^2 \cdot C \cdot m(B_R(v)),$$

and we see $n \leq CD^2$. Let $\chi = CD^2 + 1$.

Suppose there is a $v \in E$ with $\rho(v) > \operatorname{diam}(E)$. We let $E_1 = \{v\}$ and $E_2 = ... = E_\chi = \emptyset$, and the result holds. We may assume, then, that $\rho(v) \leq \operatorname{diam}(E)$ for each $v \in E$.

Let $v_1, v_2, ...$ be a sequence of vectors from E such that the ordered collection of balls $\{B_{\rho(v_1)}(v_1), B_{\rho(v_2)}(v_2), ...\}$ covers E, has nonincreasing radii, multiplicity C, and $v_i \notin \cup_{j<i} B_{\rho(v_j)}(v_j)$. The fact that such a collection of balls may be chosen follows from the Besicovitch Covering Lemma. Color the balls inductively, starting with $B_{\rho(v_1)}(v_1)$, with $CD^2 + 1$ colors so that each color gives a well-separated collection. By the work in the previous paragraph, $B_{\rho(v_j)}(v_j)$ cannot be within $\rho(v_j)$ of more than CD^2 of the balls which are already covered, so an appropriate coloring is always available. □

Fix a value $\chi = \chi(\mathbb{R}^d)$ that satisfies Lemma 2.2.3.

Corollary 2.2.4. *Suppose ν is a measure on $\mathbb{R}^d$, $E \subset \mathbb{R}^d$ is bounded, $0 < \nu(E) < \infty$, and each $v \in E$ corresponds to a ball $B_{\rho(v)}(v)$. Then there exist $v_1, v_2, ... \in E$ such that $\{B_{\rho(v_i)}(v_i)\}_{i \in \mathbb{N}}$ is well-separated and covers at least $\frac{1}{\chi}$ of the ν-mass of E.*

Proof. By Lemma 2.2.3, let $\{E_1, E_2, ..., E_\chi\}$ be a partition of E such that $\{B_{\rho(v)}(v) : v \in E_j\}$ is well-separated for each j and $\cup_j \{B_{\rho(v)}(v) : v \in E_j\}$ covers E. By finite additivity,

$$\nu(E) = \sum_{1 \leq j \leq \chi} \nu\left(\bigcup_{v \in E_j} B_{\rho(v)}(v)\right).$$

The result follows. $\square$

We now build some notation. Suppose $r, \rho > 0$ and $v \in \mathbb{R}^d$. We let

$$\partial_r B_\rho(v) := B_\rho(v) \setminus B_{\rho-r}(v)$$

and

$$\partial_r^* B_\rho(v) := B_{\rho+r}(v) \setminus B_{\rho-r}(v).$$

Also, for a collection $\mathcal{U}$ of balls in $\mathbb{R}^d$, let $\partial\mathcal{U} = \{\partial B : B \in \mathcal{U}\}$, where ∂B is the usual topological boundary. We may extend the definition of well-separated to this setting.

Definition 2.2.5. *A collection of balls $\mathcal{V}$ in $\mathbb{R}^d$ has that $\partial\mathcal{V}$ is* **well-separated** *if* $\mathrm{rmin}\mathcal{V} \leq \inf\{d(\partial B_1, \partial B_2) : B_1, B_2 \in \mathcal{V}\}$.

Note that a collection of balls $\mathcal{V}$ may have the property that $\partial\mathcal{V}$ is well-separated even though $\mathcal{V}$ is not well-separated. Consider, for example, the collection $\mathcal{V} = \{B_N : N \in \mathbb{N}\}$.

Next, we have a lemma that allows us to capture mass of a set inside thick boundaries $\partial_R^* B$ when each vector of the set itself is the center of many balls with heavy thick boundaries.

Lemma 2.2.6. *Suppose that ν is a finite measure on $\mathbb{R}^d$, $0 < \epsilon, \delta < 1$, $p = \lceil \frac{2\chi}{\epsilon\delta} \rceil + 1$, $r > 0$,*

1. *$E \subset \mathbb{R}^d$ is bounded and $\nu(E) > \delta\nu(\mathbb{R}^d)$,*

2. *$\max(11, r) \leq r_1 < R_1 < r_2 < \ldots < r_p < R_p$, and*

3. *for each $1 \leq i \leq p$, we have a function $\rho_i : E \to [r_i, R_i]$ and $\nu(\partial_r B_{\rho_i(v)}(v)) > \epsilon\nu(B_{\rho_i(v)}(v))$ for all $v \in E$.*

Then there is an integer $1 \leq k < p$ and $\mathcal{V} \subset \cup_{i>k}\{B_{\rho_i(v)}(v) : v \in E\}$ such that $\partial\mathcal{V}$ is well separated and

$$\nu((\bigcup_{B\in\mathcal{V}} \partial^*_{2R_k} B) \cap E) > \frac{1}{2}\nu(E).$$

Proof. We recursively define $\mathcal{V}_j \subset \cup_{i>p-j}\{B_{\rho_i(v)}(v) : v \in E\}$ such that $\mathcal{V}_j$ is well separated and $\nu(\cup_{B\in\mathcal{V}_j}\partial_r B) \geq j\frac{\epsilon}{2\chi}\nu(E)$ until one of the $\mathcal{V}_j$'s satisfy the conclusions. We begin by letting $\mathcal{V}_0 = \emptyset$.

Suppose $\mathcal{V}_{j-1}$ has been constructed and satisfies the above mentioned conditions. If $\nu((\cup_{B\in\mathcal{V}_{j-1}}\partial^*_{2R_{p-j}}B) \cap E) > \frac{1}{2}\nu(E)$, then let $\mathcal{V} = \mathcal{V}_{j-1}$ and $k = p - j$ to satisfy the conclusions. Otherwise, we build $\mathcal{V}_j$. Notice $\nu(E \setminus \cup_{B\in\mathcal{V}_{j-1}}\partial^*_{2R_{p-j}}B) \geq \frac{1}{2}\nu(E)$. We apply Corollary 2.2.4 to select $E_j \subset (E \setminus \cup_{B\in\mathcal{V}_{j-1}}\partial^*_{2R_{p-j}}B)$ such that $V_j := \{B_{\rho_j(v)}(v) : v \in E_j\}$ is well-separated and $\nu(E \cap \cup_{B\in V_j}B) \geq \frac{1}{2\chi}\nu(E)$. Let $\mathcal{V}_j = \mathcal{V}_{j+1} \cup V_j$. Since each element of E_j is of distance at least $2R_k$ from the boundary of any ball in $\mathcal{V}_{j-1}$, $\partial\mathcal{V}_j$ is well-separated. Also,

$$\begin{aligned}
\nu\left(\bigcup_{B\in\mathcal{V}_j}\partial_r B\right) &= \nu\left(\bigcup_{B\in\mathcal{V}_{j-1}}\partial_r B\right) + \nu\left(\bigcup_{B\in V_j}\partial_r B\right) \\
&\geq (j-1)\frac{\epsilon}{2\chi}\nu(E) + \frac{\epsilon}{2\chi}\nu(E) \\
&= j\frac{\epsilon}{2\chi}\nu(E).
\end{aligned}$$

Notice that $\nu(\cup_{B\in\mathcal{V}_j}\partial_r B)$ would be larger than $\nu(\mathbb{R}^d)$ for $j = p$, so the process must terminate before building $\mathcal{V}_p$. □

We would like to use the fact that the boundary of a ball has lower dimension than the ball itself. Since we are dealing with thick boundaries, though, we need to define a different type of dimension.

Definition 2.2.7. *If Y is a metric space and $R_0 > 1$, then $\mathrm{cdim}_{R_0}Y = k$ (read* **Y has coarse dimension k at scales $\geq \mathbf{R_0}$***) is defined by recursion on k:*

1. $\mathrm{cdim}_{R_0}\emptyset = -1$ *for any R_0,*

2. $\mathrm{cdim}_{R_0}Y$ *is the minimum integer* k *for which* $\mathrm{cdim}_{tR_0}\partial_t^* B_\rho(v) \leq k-1$ *for any* $t \geq 1, \rho \geq tR_0$, *and* $v \in Y$.

For example, $\mathrm{cdim}_{R_0}E = 0$ if and only if E is nonempty and $\mathrm{diam}E < R_0 - 1$. Also, if E has $\mathrm{diam}E \geq R_0 - 1$ and E is composed of two nonempty, disjoint sets, each of diameter less than $R_0 - 1$, then $\mathrm{cdim}_{R_0}E = 1$.

We show that $\mathbb{R}^d$ has finite coarse dimension for scales ≥ 11.

Lemma 2.2.8. *There exists* $k \in \mathbb{N}$ *such that* $\rho(1) > \rho(2) > ... > \rho(k) \geq 11$ *and* $v_1, v_2, ..., v_k \in \mathbb{R}^d$ *with* $v_i \notin \cup_{j<i} B_{\rho(j)-1}(v_j)$ *implies*

$$\bigcap_{i=1}^{k} \partial_1^* B_{\rho(i)}(v_i) = \emptyset.$$

Proof. Let k be larger than the most points that can be placed around v in $\mathbb{R}^d$ so that the angle formed by any pair of these points on the legs and v as the vertex is at least $\arccos(\frac{187}{198})$. Suppose the assumptions hold and

$$v \in \bigcap_{i=1}^{k} \partial_1^* B_{\rho(i)}(v_i).$$

Choose $1 \leq j < i \leq k$ and note $\rho(j) > \rho(i)$. Let $\alpha := d(v_i, v_j), \beta := d(v, v_i), \gamma := d(v, v_j)$, and $\phi := \angle v_i v v_j$, and we have the following three estimates:

$$\begin{aligned} \rho(j) - 1 &< \alpha, \\ |\rho(i) - \beta| &\leq 1, \text{ and} \\ |\rho(j) - \gamma| &\leq 1. \end{aligned}$$

The law of cosines states that

$$\alpha^2 = \beta^2 + \gamma^2 - 2\beta\gamma\cos\phi.$$

We use this to estimate $\cos\phi$:

$$\begin{aligned}\cos\phi &= \frac{\beta^2+\gamma^2-\alpha^2}{2\beta\gamma}\\ &< \frac{(\rho(i)+1)^2+(\rho(j)+1)^2-(\rho(j)-1)^2}{2(\rho(i)-1)(\rho(j)-1)}\\ &= \frac{\frac{\rho(i)}{\rho(j)}+\frac{2}{\rho(j)}+\frac{4}{\rho(i)}+\frac{1}{\rho(i)\rho(j)}}{2-\frac{2}{\rho(j)}-\frac{2}{\rho(i)}+\frac{2}{\rho(i)\rho(j)}}\\ &\leq \frac{187}{198}.\end{aligned}$$

Thus, $\phi < \arccos(\frac{187}{198})$ for all $1 \leq j < i \leq k$, which contradicts the definition of k. $\square$

Proposition 2.2.9. $\mathbb{R}^d$ *has finite coarse dimension at scales* ≥ 11.

Proof. Suppose $\mathrm{cdim}_{11}\mathbb{R}^d = l$. We work backwards through the inductive definition of coarse dimension. For any $t(1) \geq 1$, $\rho(1) \geq 11t(1)$, and $v \in \mathbb{R}^d$, we have

$$\mathrm{cdim}_{t(1)11}\partial^*_{t(1)}B_{\rho(1)}(v) = l-1.$$

This means that for any $t(1), t(2) \geq 1$, $\rho(1) \geq 11t(1)$, $\rho(2) \geq 11t(1)t(2)$, $v_1 \in \mathbb{R}^d$, and $v_2 \in \partial^*_{t(1)}B_{\rho(1)}(v_1)$, we have that

$$\mathrm{cdim}_{t(1)t(2)11}\left(\partial^*_{t(2)}B_{\rho(2)}(v_2) \cap \partial^*_{t(1)}B_{\rho(1)}(v_1)\right) \leq l-2.$$

How do we see that l is finite? To say that $\mathbb{R}^d$ has finite coarse dimension no larger than $k-1$ is to say that for any $t(1), t(2), ..., t(k) \geq 1$, $\rho(i)$ for $1 \leq i \leq k$ such that $\rho(i) \geq t(1)t(2)\cdots t(i)11$, and $v_i \in \mathbb{R}^d$ for $1 \leq i \leq k$ with $v_i \in \partial^*_{t(j)}B_{\rho(j)}(v_j)$ for all $1 \leq j < i$, we have

$$\mathrm{cdim}_{t(1)\cdots t(k)11}\bigcap_{i=1}^{k}\partial^*_{t(i)}B_{\rho(i)}(v_i) = -1,$$

i.e.,$\cap_{i=1}^{k}\partial^*_{t(i)}B_{\rho(i)}(v_i) = \emptyset$.

Let k' be the k from Proposition 2.2.8 and k'' be the largest $k \in \mathbb{N}$ such that there exist $v_1, ...v_k \in B_{\frac{12}{11}}$ with $d(v_i, v_j) \geq 1 - \frac{1}{11}$ for all $1 \leq i < j \leq k$. Let $k = k'k'' + 1$. Suppose we are given

1. $t(1), t(2), ..., t(k) \geq 1$,

2. $\rho(1), \rho(2), ..., \rho(k)$ with $\rho(i) \geq t(1)t(2)\cdots t(i)11$, and

3. $v_1, v_2, ..., v_k \in \mathbb{R}^d$ such that $v_i \in \partial^*_{t(j)} B_{\rho(j)}(v_j)$ for $j < i$.

We would like to show $\cap_{i=1}^k \partial^*_{t(i)} B_{\rho(i)}(v_i) = \emptyset$.

Claim: we can obtain a subsequence of length k' such that $\{\rho(j_i)\}_{i=1}^{k'}$ is decreasing. To see this, we show there is some $2 \leq j \leq k'' + 1$ with $\rho(j) < \rho(1)$, and the argument may be repeated for $\rho(j)$, and so on. Suppose, for the sake of contradiction, that $\rho(j) \geq \rho(1)$ for $2 \leq j \leq k'' + 1$. First, notice

$$v_j \in B_{\rho(1)+t(1)}(v_1) \subset B_{\frac{12}{11}\rho(1)}(v_1)$$

for all $2 \leq j \leq k'' + 1$. Second, if $2 \leq j < i \leq k'' + 1$, then

$$d(v_i, v_j) \geq \rho(j) - t(j) \geq \rho(j)(1 - \frac{1}{11}) \geq \rho(1)(1 - \frac{1}{11}).$$

Scaling the norm by a factor of $\frac{1}{\rho(1)}$ shows that we have contradicted the definition of k''.

Having reduced to a subsequence of length k' with decreasing radii, let $t = \max_i t(i)$ and scale the norm by a factor of $\frac{1}{t}$. Lemma 2.2.8 shows $\cap_{i=1}^k \partial^*_{t(i)} B_{\rho(i)}(v_i) = \emptyset$, so $\mathrm{cdim}_{11}\mathbb{R}^d \leq k - 1$. $\square$

We now have access to the main result of this section, which will allow us to show a Følner condition on the orbits of a $\mathbb{Z}^d$ or $\mathbb{R}^d$ action. Suppose F is a subset of $\mathbb{R}^d$.

Theorem 2.2.10. *Fix $\epsilon, \delta \in (0, 1)$ and $r > 0$. Suppose $cdim_{11}F = k$, and let q be an integer no smaller than $(\frac{2^{4k}\chi}{\epsilon\delta} + 2)^k \cdot (\frac{2^{7k}\chi}{\epsilon\delta^2})^k$. Also, suppose ν is a finite measure on F,*

1. *$E \subset F$ is bounded,*

2. r_i and R_i are numbers for $1 \leq i \leq q$ such that $\max(11, r) \leq r_1 < R_1$, $R_{i+1} > r_{i+1} > 11R_i$, and

3. $\rho_i : E \longrightarrow [r_i, R_i]$ for $1 \leq i \leq q$ such that $\nu(\partial_r B_{\rho_i(v)}(v)) > \epsilon \nu(B_{\rho_i(v)}(v))$ for each $v \in E$.

Then $\nu(E) \leq \delta \nu(F)$.

Proof. We proceed by induction on k. For the base case, suppose $k = -1$, which is trivial since this implies $F = \emptyset$.

Suppose, then, that the result holds for $\text{cdim}_{11} F = k - 1$. For the sake of contradiction, assume $\nu(E) > \delta \nu(F)$. For each $1 \leq i \leq q$, let

$$\mathcal{U}_i := \{B_{\rho_i(v)}(v) : v \in E\}.$$

Let $N = \lceil \frac{q}{\lceil \frac{2\chi}{\epsilon\delta} \rceil + 1} \rceil$ and notice $N \geq (\frac{2^{4k}\chi}{\epsilon\delta} + 2)^{k-1} \cdot (\frac{2^{7k}\chi}{\epsilon\delta^2})^k$. We apply Lemma 2.2.6 to every N^{th} r_i, R_i, and ρ_i to get a collection of balls $\mathcal{V} \subset \cup_{k' < i < \frac{q}{N}} \mathcal{U}_{iN}$ such that $\partial\mathcal{V}$ is well-separated and $\nu(F_0 \cap E) > \frac{1}{2}\nu(E)$, where $F_0 = \cup_{B \in \mathcal{V}} \partial^*_{2R_{k'N}} B$. Let $p = \lceil (\frac{2^{4k}\chi}{\epsilon\delta} + 1)^{k-1} \cdot (\frac{2^{7k}\chi}{\epsilon\delta^2})^{k-1} \rceil$, $M = \frac{N}{p}$, and $m(j) = k'N + jp$ for $0 \leq j < M$. The constant p corresponds to q for parameters $k-1, \chi, \frac{\epsilon}{2}$, and $\frac{\delta}{8}$. Notice $M > \frac{64\chi}{\epsilon\delta^2}$. We consider the following M collections of covers of E:

$$\mathcal{U}_{m(0)+1}, \mathcal{U}_{m(0)+2}, \dots, \mathcal{U}_{m(0)+p}$$

$$\mathcal{U}_{m(1)+1}, \mathcal{U}_{m(1)+2}, \dots, \mathcal{U}_{m(1)+p}$$

$$\vdots$$

$$\mathcal{U}_{m(M-1)+1}, \mathcal{U}_{m(M-1)+2}, \dots, \mathcal{U}_{m(M-1)+p}$$

For $0 \leq j < M$, let $F_j := \cup_{B \in \mathcal{V}} \partial^*_{2R_{m(j)}} B$. Notice $F_j \subset F_{j+1}$, F_j is a disjoint union since $\partial\mathcal{V}$ is well-separated and $\text{rmin}(\mathcal{V}) > 2R_{m(j)}$, and $\nu(F_j \cap E) > \frac{1}{2}\nu(E)$ for all $0 \leq j < M$. Fix $0 \leq j < M - 1$ and for $B \in \mathcal{V}$ let $F_B = \partial^*_{2R_{m(j)}} B$. Also, let $\mathcal{V}_j := \{B \in \mathcal{V} : \nu(F_B \cap E) > \frac{\delta}{4}\nu(F_B)\}$. Fix $B \in \mathcal{V}_j$.

We would like to apply the induction hypothesis to the set F_B. However, we do not know that all of the balls with heavy boundaries still have heavy boundaries when restricted to F_B. So, we have to throw away some points. In fact, the induction hypothesis will give us a lower bound for the ν mass of the points we have to throw away. For $m(j) < i \leq m(j) + p$, let

$$E_B^i := \left\{ v \in E \cap F_B : \nu(F_B \cap \partial_r B_{\rho_i(v)}(v)) > \frac{\epsilon}{2}\nu(F_B \cap B_{\rho_i(v)}(v)) \right\},$$

$E_B := \cap_{m(j)<i\leq m(j)+p} E_B^i$, and $E_B' := F_B \cap E \setminus E_B$. We apply the induction hypothesis to the space F_B with bounded subset E_B and parameters $k-1, \frac{\epsilon}{2}, \frac{\delta}{8}, \nu|_{F_B}$ and p instead of parameters k, ϵ, δ, ν, and q, respectively. This gives that $\nu(E_B) \leq \frac{\delta}{8}\nu(F_B)$. In turn,

$$\nu(E_B') > \frac{\delta}{8}\nu(F_B),$$

since $B \in \mathcal{V}_j$. For $v \in E_B'$, let $m(j) < i_v \leq m(j) + p$ such that $v \notin E_B^{i_v}$. Thus,

$$\nu(F_B^c \cap \partial_r B_{\rho_{i_v}(v)}(v)) > \frac{\epsilon}{2}\nu(B_{\rho_{i_v}(v)}(v))$$

for all $v \in E_B'$. Now $\{B_{\rho_{i_v}(v)}(v) : v \in E_B'\}$ is a cover of E_B', and we apply Corollary 2.2.4 to get a well-separated subcollection $\mathcal{C}$ covering $\frac{1}{\chi}$ of the ν mass of E_B'. We have the following estimate:

$$\begin{aligned} \nu(\bigcup_{B' \in \mathcal{C}} \partial_r B' \cap (F \setminus F_B)) &> \frac{\epsilon}{2}\nu(\bigcup_{B' \in \mathcal{C}} B') \\ &> \frac{\epsilon}{2\chi}\nu(E_B') \\ &> \frac{\epsilon\delta}{16\chi}\nu(F_B). \end{aligned}$$

Notice $\bigcup_{B' \in \mathcal{C}} \partial_r B' \cap (F \setminus F_B)$ is contained in $F_{j+1} \setminus F_j$.

We un-fix B and have

$$\begin{aligned}
\frac{\delta}{2}\nu(F) &< \nu(F_j \cap E) \\
&= \nu((\bigcup_{B\in\mathcal{V}_j} \partial^*_{2R_{m(j)}} B) \cap E) + \nu((\bigcup_{B\in\mathcal{V}\setminus\mathcal{V}_j} \partial^*_{2R_{m(j)}} B) \cap E) \\
&\leq \nu(\bigcup_{B\in\mathcal{V}_j} F_B \cap E) + \frac{\delta}{4}\nu(\bigcup_{B\in\mathcal{V}\setminus\mathcal{V}_j} \partial^*_{2R_{m(j)}} B) \\
&\leq \nu(\bigcup_{B\in\mathcal{V}_j} F_B) + \frac{\delta}{4}\nu(F).
\end{aligned}$$

This gives that $\frac{\delta}{4}\nu(F) < \nu(\bigcup_{B\in\mathcal{V}_j} F_B)$, which is a disjoint union and allows us to estimate the ν mass between F_{j+1} and F_j:

$$\begin{aligned}
\nu(F_{j+1} \setminus F_j) &\geq \sum_{B\in\mathcal{V}_j} \nu(\partial^*_{2R_{m(j)+p}} B \setminus \partial^*_{2R_{m(j)}} B) \\
&> \sum_{B\in\mathcal{V}_j} \frac{\epsilon\delta}{16\chi}\nu(F_B) \\
&= \frac{\epsilon\delta}{16\chi}\nu(\bigcup_{B\in\mathcal{V}_j} F_B) \\
&> \frac{\epsilon\delta^2}{64\chi}\nu(F)
\end{aligned}$$

We see that this is true for all $0 \leq j < M-1$, which gives $\nu(\cup_j F_j) > \nu(F)$, a contradiction. □

Chapter 3

Diffusion of Measure

Let F be $\mathbb{Z}^d$ or $\mathbb{R}^d$. In the first section of this chapter we construct measures μ_x on F for μ-a.e. x such that for any $A \in \mathcal{F}$ and $N > 0$,

$$\mu(A) = \int \frac{\mu_x(A_x \cap B_N)}{\mu_x(B_N)} d\mu_N^*,$$

where $A_x := \{v \in F : T^v(x) \in A\}$, T is a free Borel action of F on a standard Borel probability space $(X, \mathcal{F}, \mu)$, and μ_N^* is a measure on X. Such diffusion of the measure is common in the setting of the leaves of a foliation but was only recently applied to the general context of Borel actions on Polish spaces. In the second section we prove a Følner condition for F on the diffused measures μ_x.

3.1 Construction of the Diffused Measure

Definition 3.1.1. *The measure spaces* $(X, \mathcal{F}, \mu)$ *and* $(Y, \mathcal{G}, \nu)$ *are* **measurably isomorphic** *if there exist*

1. $X_0 \subset X$ *and* $Y_0 \subset Y$ *with* $\mu(X \setminus X_0) = 0$, $\nu(Y \setminus Y_0) = 0$ *and*

2. *a measurable bijection* $\phi : X_0 \to Y_0$ *such that* $\nu = \mu \circ \phi^{-1}$.

The map ϕ *is called an* **isomorphism**.

Definition 3.1.2. *A* **standard probability space** *(or* **Lebesgue probability space***) is a probability space which is measurably isomorphic to an interval with the Borel sets and Lebesgue measure joined with at most countably many atoms.*

A standard Borel probability space will be the setting in which we work. We can look at the preimage of the dyadic intervals under such an isomorphism to get a sequence of partitions $\mathcal{P}_n$ that refines to points almost surely on a standard probability space. Each partition includes the singletons of the atoms.

To build the diffused measures, we construct a "twisted" measure on $X \times F$ and then define a measure on F using the Rohklin decomposition. Let m be Haar measure (Lebesgue if $F = \mathbb{R}^d$, counting if $F = \mathbb{Z}^d$) on the Borel sets (denoted $\mathcal{B}$) of F. We use the function $I : X \times F \to X \times F$, defined by $I(x, v) := (T^v x, v)$, to twist the space $X \times F$. Let $N \in \mathbb{N}$ and $\hat{\mu}_N := \frac{1}{m(B_N)} I^*(\mu \times m)|_{X \times B_N}$, a Borel probability measure on $X \times B_N$. Notice that for $M > N$, $\hat{\mu}_M|_{X \times B_N}$ is equivalent to $\hat{\mu}_N$.

The Rohklin decomposition can be used to pull the measure $\hat{\mu}_N$ down to orbits, so a description of this decomposition of a measure will be useful. Suppose $(X, \mathcal{F}, \mu)$ is a standard Borel probability space. Let $(Y, \mathcal{G}, \hat{\lambda})$ be an interval of length λ_0 together with the points $1, 2, \ldots$, which each are assigned mass $\lambda_1, \lambda_2, \ldots$, such that $(X, \mathcal{F}, \mu)$ is isomorphic to $(Y, \mathcal{G}, \hat{\lambda})$. Then there exists a measurable isomorphism ψ between $(X, \mathcal{F}, \mu)$ and $(Y \times Y, \mathcal{G} \times \mathcal{G}, \nu)$ for some measure ν (see Figure (3.1)). There is also a measurable isomorphism ϕ between $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \hat{\lambda})$. The Rohklin Theorem says that for any sub σ-algebra $\mathcal{H}$ of $\mathcal{F}$, ψ and ν may be chosen so that the σ-algebra $\{A \times Y : A \in \mathcal{G}\}$ in $Y \times Y$ is the image of $\mathcal{H}$ under ψ almost surely. The pullbacks of the vertical slices $\{a\} \times Y$ are called fibres. Notice, then, that a fibre is a maximal set of points in X which are indistinguishable under $\mathcal{H}$. Let $P_n(x)$ be the element of the partition $\mathcal{P}_n$ that contains x and π_2 be projection onto the second coordinate in $Y \times Y$. A fibre almost surely can be given a measure μ_x by

$$\mu_x(A) := \lim_{n \to \infty} \frac{\mu(\psi^{-1}(\psi(P_n(x)) \times \pi_2(\psi(A))))}{\mu(\psi^{-1}((\psi(P_n(x)) \times Y)))}. \tag{3.1.1}$$

Integrating these measures over μ gives back μ, i.e., for any $f \in L^1(\nu)$,

$$\int f d\mu = \int \int f d\mu_x d\mu. \tag{3.1.2}$$

For $x \in X$, the measure μ_x is a version of the conditional expectation of μ given $\mathcal{H}$, which is denoted $E_\mu(A|\mathcal{H})$ [20, 18].

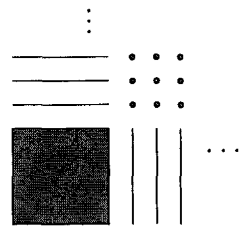

Figure 3.1: The space $Y \times Y$.

Let $\mathcal{H}_N := \{A \times B_N : A \in \mathcal{F}\}$, and apply the Rohklin Theorem to $X \times B_N$ using the sub σ-algebra $\mathcal{H}_N$. The fibres correspond to sets $\{x\} \times B_N$, since these are precisely the sets of undistinguishable points under $\mathcal{H}_N$. The fibre measures, which we call $\hat{\mu}_{x,N}$, are measures on Borel subsets of $\{x\} \times B_N$. To keep our notation as clean as possible, we consider the measures $\hat{\mu}_{x,N}$ as measures on the Borel sets of B_N. We let $f(x, v) = 1_E(v)$ for a Borel $E \subset B_N$ in (3.1.2) to obtain

$$\int_{X \times B_N} 1_E(v) d\hat{\mu}_N = \int_{X \times B_N} \hat{\mu}_{x,N}(E) d\hat{\mu}_N. \tag{3.1.3}$$

Equation (3.1.3) is enough to uniquely determine measures $\hat{\mu}_{x,N}$ (up to a set of measure 0). To construct the measures, we use (3.1.1), which in our case is

$$\hat{\mu}_{x,N}(E) = \lim_{n \to \infty} \frac{\hat{\mu}_N(A_n \times E)}{\hat{\mu}_N(A_n \times B_N)}, \tag{3.1.4}$$

where $A_n = P_n(x)$ for some choice of ϕ. The set of points x for which $\hat{\mu}_{x,N}$ is defined under the construction, which we denote as $X_{0,N}$, is precisely the set of points for which there are A_n that refine to x and have $\hat{\mu}_{x,N}(A_n(x) \times B_N) > 0$ for each n. There may be many measurable isomorphisms between $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \hat{\lambda})$, and $\hat{\mu}_{x,N}$ is defined if there exists a measurable isomorphism between these two spaces such that $P_n(x) \times B_N$ has positive $\hat{\mu}_N$ measure for each n. Let $X_0 := \cup X_{0,N}$. We will see that X_0 is a set which is invariant under T and of full μ measure.

Letting $E = B_N$ in (3.1.3), we see that for each N, $X_{0,N}$ has $\hat{\mu}_N(X_{0,N} \times B_N) = 1$. Also, fixing an $x \in X_{0,N}$ results in measures that extend each other up to rescaling:

$$\hat{\mu}_{x,N} = \frac{\hat{\mu}_{x,N+1}|_{B_N}}{\hat{\mu}_{x,N+1}(B_N)}.$$

For each $x \in X_0$, we let N_x be the smallest natural number N for which $\hat{\mu}_{x,N}(B_N) > 0$ and define a Borel measure on all of F by

$$\mu_x = \lim_{N \to \infty} \frac{\hat{\mu}_{x,N}}{\hat{\mu}_{x,N}(B_{N_x})}.$$

We have constructed measures μ_x on F for $x \in X_0$ that pull the measure μ down to the orbits. In what ways do the measures μ_x represent the structure of the system? We address this question in two fashions. First, we show in Proposition 3.1.4 that composing the measures μ_x with a shift on F is equivalent to shifting the base point x. This shows that these orbit measures behave correctly when the base point is moved to somewhere else on the same orbit. Also, as a corollary, this shows that X_0 is T invariant and of full μ measure. Second, in Proposition 3.1.6, we prove a statement that allows us to interpret the measure μ in terms of the orbit measures μ_x.

Lemma 3.1.3. *Suppose $N > 0$, $w \in F$, and $M \geq N + ||w||$. Then for any $A \in \mathcal{F}$,*

$$\hat{\mu}_M(A \times B_N) = \hat{\mu}_M(T^w(A) \times B_N(-w)).$$

Proof. Let N, w, and M be as stated and choose $A \in \mathcal{F}$. By the definition of $\hat{\mu}_M$,

$$\hat{\mu}_M(T^w(A) \times B_N(-w)) = \frac{1}{m(B_M)} \cdot (\mu \times m)(I^{-1}(T^w(A) \times B_N(-w))).$$

Now we can apply Fubini's Theorem to the right hand side and use the fact that m is shift-invariant:

$$\begin{aligned} \int\int 1_{I^{-1}(T^{-w}(A) \times B_N(-w))} d\mu dm &= \int_{B_N(-w)} \int 1_{T^{-w}(A)}(T^v(x)) d\mu dm \\ &= \int_{B_N} \int 1_A(T^v(x)) d\mu dm. \end{aligned}$$

Thus, $(\mu \times m)(I^{-1}(T^{-w}(A) \times B_N(-w))) = (\mu \times m)(I^{-1}(A \times B_N))$, and the result follows by dividing by $m(B_N)$. ∎

Let $\tau_w(v) := v + w$ for $v, w \in F$.

Proposition 3.1.4. *For $x \in X_0$, μ_x is equivalent to $\mu_{T^w(x)} \circ \tau_{-w}$ for each $w \in F$.*

Proof. Suppose $x \in X_0$ and $\{A_n\}$ is a sequence of $\mathcal{F}$ measurable sets that decrease to $\{x\}$ and have $\hat{\mu}_{N_x}(A_n \times B_{N_x}) > 0$ for each $n \in \mathbb{N}$. Choose $w \in F$. We first use Lemma 3.1.3 to show that $\frac{\mu_x(B_N)}{\mu_{T^w(x)}(B_N(-w))}$ does not depend on N for $N \geq N_x$, assuming $\mu_{T^w(x)}$ exists. Suppose $N_2 > N_1 \geq N_x$, and let $M = N_2 + ||w||$. We have

$$\frac{\left(\frac{\mu_x(B_{N_1})}{\mu_{T^w(x)}(B_{N_1}(-w))}\right)}{\left(\frac{\mu_x(B_{N_2})}{\mu_{T^w(x)}(B_{N_2}(-w))}\right)} = \frac{\mu_x(B_{N_1})\mu_{T^w(x)}(B_{N_2}(-w))}{\mu_x(B_{N_2})\mu_{T^w(x)}(B_{N_1}(-w))}.$$

Now, since $B_{N_1}, B_{N_2}(-w) \subset B_M$, we use (3.1.1) to see

$$\frac{\left(\frac{\mu_x(B_{N_1})}{\mu_{T^w(x)}(B_{N_1}(-w))}\right)}{\left(\frac{\mu_x(B_{N_2})}{\mu_{T^w(x)}(B_{N_2}(-w))}\right)} = \frac{\lim_{n\to\infty} \hat{\mu}_M(A_n \times B_{N_1})\hat{\mu}_M(T^w(A_N) \times B_{N_2}(-w))}{\lim_{n\to\infty} \hat{\mu}_M(A_n \times B_{N_2})\hat{\mu}_M(T^w(A_n) \times B_{N_1}(-w))}.$$

By Lemma 3.1.3, the right hand side is 1. Let $k_{x,w} = \frac{\mu_x(B_N)}{\mu_{T^w(x)}(B_N(-w))}$ for $N > N_x$.

Suppose x, w, and A_n are as above. Also, suppose $E \subset B_N$ is Borel, $N \geq N_x$, $M = N + ||w||$. We again use (3.1.1) and find

$$\begin{aligned}
\mu_x(E) &= \mu_x(B_M) \lim_{n\to\infty} \frac{\hat{\mu}_M(A_n \times E)}{\hat{\mu}_M(A_n \times B_M)} \\
&= \mu_x(B_M) \lim_{n\to\infty} \frac{\hat{\mu}_M(A_n \times B_N)}{\hat{\mu}_M(A_n \times B_M)} \cdot \lim_{n\to\infty} \frac{\hat{\mu}_M(A_n \times E)}{\hat{\mu}_M(A_n \times B_N)}.
\end{aligned}$$

The first limit is just $\frac{\mu_x(B_N)}{\mu_x(B_M)}$. We apply Lemma 3.1.3 to the second limit gives

$$\mu_x(E) = \mu_x(B_N) \lim_{n\to\infty} \frac{\hat{\mu}_M(T^w(A_n) \times E - w)}{\hat{\mu}_M(T^w(A_n) \times B_N(-w))}.$$

We evaluate the limit to find that $\hat{\mu}_{T^w(x)}$ does exist and the right hand side is exactly $k_{x,w} \cdot \mu_{T^w(x)} \circ \tau_{-w}(E)$. The result follows by continuity from below on σ-finite measures. □

Corollary 3.1.5. *The set X_0 is T invariant and $\mu(X_0) = 1$.*

Proof. Proposition 3.1.4 implies that $\mu_{T^v(x)}(F) > 0$ for every $v \in F$ whenever $\mu_x(F) > 0$. This is exactly invariance of X_0.

It was noted earlier that $\hat{\mu}_N(X_{0,N} \times B_N) = 1$ for any N. By continuity from below, this implies

$$\begin{aligned} 1 &= \hat{\mu}_N(X_0 \times B_N) \\ &= \frac{1}{m(B_N)}(\mu \times m)(I^{-1}(X_0 \times B_N)) \\ &= \mu(X_0). \end{aligned}$$

□

As the final result in this section, we have an equation that allows us to interpret the measure μ in terms of the orbit measures μ_x. Recall $A_x = \{v \in F : T^v(x) \in A\}$.

To use μ_x, we need another measure: the projection of $\hat{\mu}_N$ onto the first coordinate. This projection, which we denote as μ_N^*, is a measure on X and for $A \in \mathcal{F}$ is equal to $\hat{\mu}_N(A \times B_N)$. Writing this as an integral and untwisting the measure gives

$$\mu_N^* = \frac{1}{m(B_N)} \int_{B_N} (T^v)^* \mu dm.$$

If $A \in \mathcal{F}$ is an invariant set, then $\mu_N^*(A) = \mu(A)$. For example, $\mu_N^*(X_0) = \mu(X_0) = 1$ for any N. The measure μ_N^* allows us to use Fubini's Theorem on $\hat{\mu}_N$. For any $g \in L^1(\hat{\mu}_N)$,

$$\int_{X \times B_N} g d\hat{\mu}_N = \int_X \int_{B_N} g d\hat{\mu}_{x,N} d\mu_N^*.$$

Proposition 3.1.6. *For $A \in \mathcal{F}$ and $N > 0$,*

$$\mu(A) = \int \frac{\mu_x(A_x \cap B_N)}{\mu_x(B_N)} d\mu_N^*.$$

Proof. Let $A \in \mathcal{F}$ and $N > 0$. We can integrate over B_N, normalize, and twist the inside integral to get

$$\mu(A) = \frac{1}{m(B_N)} \int_{B_N} \left(\int 1_{T^v(A)} d(T^v)^* \mu \right) dm.$$

Since $\frac{1}{m(B_N)} d(T^v)^* \mu dm = d\hat{\mu}_N$, we apply 3.1.3 and see

$$\mu(A) = \int \hat{\mu}_{x,N}(A_x \cap B_N) d\hat{\mu}_N.$$

Recall that μ_x is defined so that $\hat{\mu}_{x,N} = \frac{\mu_x|_{B_N}}{\mu_x(B_N)}$, so

$$\begin{aligned} \mu(A) &= \int \int_{B_N} \frac{\mu_x(A_x \cap B_N)}{\mu_x(B_N)} d\hat{\mu}_{x,N} d\mu_N^* \\ &= \int \frac{\mu_x(A_x \cap B_N)}{\mu_x(B_N)} d\mu_N^*. \end{aligned}$$

□

3.2 A Følner Condition on Orbits

We can now prove a Følner condition on orbits via the diffused measures. A Følner condition states that, for any $r > 0$, the ratio of the measure of $\partial_r B$ to the measure of B goes to zero as the radius of the ball goes to infinity.

Proposition 3.2.1. *For any $r, R > 0$ and $x \in X_0$,*

$$\int \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} d\mu_R^* = \frac{m(\partial_r B_R)}{m(B_R)}.$$

Proof. Let $r, R > 0$. We use that μ_R^* is the projection of $\hat{\mu}_R$ onto the first coordinate:

$$\int \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} d\mu_R^* = \int_{X \times B_R} \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} d\hat{\mu}_R.$$

Recall that $\frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} = \hat{\mu}_{x,R}(\partial_r B_R)$. So, we can apply (3.1.3) and get

$$\int \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} d\mu_R^* = \int_{X \times B_R} 1_{\partial_r B_R}(v) d\hat{\mu}_R.$$

Now we untwist the measure $\hat{\mu}_R$ to get

$$\int \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} d\mu_R^* = \frac{m(\partial_r B_R)}{m(B_R)},$$

which completes the proof. □

We prove the Følner Condition with respect to the sequence of measures μ_N^*.

Corollary 3.2.2. *Let $r > 0$. Then*

$$\lim_{R\to\infty} \int \left| \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} \right| d\mu_R^* = 0.$$

Proof. Let $r > 0$. By Proposition 3.2.1,

$$\lim_{R\to\infty} \int \left| \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} \right| d\mu_R^* = \lim_{R\to\infty} \frac{m(\partial_r B_R)}{m(B_R)} = 0.$$

□

We now prove the Følner Condition pointwise almost surely.

Theorem 3.2.3. *(Følner Condition) For μ-a.e. $x \in X_0$,*

$$\lim_{R\to\infty} \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} = 0$$

for all $r > 0$.

Proof. Let $x \in X_0$, a set of full μ measure. Let $r > 0$ be fixed and for any $\epsilon > 0$,

$$A_\epsilon := \left\{ x \in X_0 : \limsup_{R\to\infty} \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} > \epsilon \right\}.$$

We want to show that $\mu(A_\epsilon) = 0$ for all $\epsilon > 0$. Suppose, for the sake of contradiction, that $\epsilon > 0$ and $\mu(A_\epsilon) = 2a > 0$. We inductively reduce this set so that we have a structure on the radii of the boundary-heavy balls that will allow us to use Theorem 2.2.10. Reduce to $A_1 \subset A_\epsilon$ with $\mu(A_1) \geq \frac{3}{2}a$ and $\max(10, r) \leq r_1 < R_1$ such that $x \in A_1$ implies there is a $\rho_1(x)$ with $r_1 \leq \rho_1(x) \leq R_1$ and $\mu_x(\partial_r B_{\rho_1(x)}) > \epsilon\mu_x(B_{\rho_1(x)})$. Having defined A_{i-1}, reduce to $A_i \subset A_{i-1}$ with $\mu(A_i) \geq (1 + \frac{1}{2^i})a$ and $11R_{i-1} < r_i < R_i$ such that $x \in A_i$ implies there is a $\rho_i(x)$ with $r_i \leq \rho_i(x) \leq R_i$ and $\mu_x(\partial_r B_{\rho_i(x)}) > \epsilon\mu_x(B_{\rho_i(x)})$. Let $A := \cap_i A_i$. By continuity from above, $\mu(A) \geq a > 0$.

Let q be an integer no less than $(\frac{2^{4k}\chi}{\epsilon\delta} + 2)^k \cdot (\frac{2^{7k}\chi}{\epsilon\delta^2})^k$ and $r := R_q$. Suppose $R > 0$. By Proposition 3.1.6,

$$a \leq \int \frac{\mu_x(A_x \cap B_R)}{\mu_x(B_R)} d\mu_R^*.$$

We let $D_R := \{x \in X_0 : \frac{\mu_x(A_x \cap B_R)}{\mu_x(B_R)} \geq a^2\}$, and it follows that $\mu^*_R(D_R) \geq \frac{a}{a+1}$. Suppose $x \in D_R$. We apply Theorem 2.2.10 with $E = A_x \cap B_{R-r}$, $\delta = \frac{a^2}{2}$, and $\nu = \mu_x|_{B_R}$ to find that

$$\mu_x(A_x \cap B_{R-r}) < \frac{a^2}{2}\mu_x(B_R).$$

But $x \in D_R$, so $\mu_x(A_x \cap B_R) \geq a^2\mu_x(B_R)$. Thus,

$$\mu_x(A_x \cap \partial_r B_R) \geq \frac{a^2}{2}\mu_x(B_R)$$

for any $x \in D_R, R > 0$.

Using Corollary 3.2.2, we choose $\hat{R} > 0$ such that

$$\int \frac{\mu_x(\partial_r B_{\hat{R}})}{\mu_x(B_{\hat{R}})} d\mu^*_{\hat{R}} \leq \frac{a^3}{2a+4}.$$

We see that $X' = \{x \in X_0 : \frac{\mu_x(\partial_r B_{\hat{R}})}{\mu_x(B_{\hat{R}})} \geq \frac{a^2}{2}\}$ has $\frac{a^3}{2a+4} \geq \frac{a^2}{2}\mu^*_{\hat{R}}(X')$, i.e., $\mu^*_{\hat{R}}(X') \leq \frac{a}{a+2}$. But $D_{\hat{R}} \subset X'$, and $\mu^*_{\hat{R}}(D_{\hat{R}}) \geq \frac{a}{a+1}$. This is a contradiction and shows that $\mu(A_\epsilon) = 0$. For $r > 0$, let

$$A_{0,r} := \left\{x \in X_0 : \lim_{R \to \infty} \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} = 0\right\}.$$

We have $A_{0,s} \subset A_{0,r}$ and $\mu(A_{0,r}) = 1$ for any $s > r > 0$. Let $A_0 = \cap_{r>0} A_{0,r}$. By continuity from above, $\mu(A_0) = 1$. This completes our proof. □

Finally, we show that the set of points which satisfy the Følner condition is invariant under the action T. Notice

$$A_0 = \{x \in X_0 : \lim_{R \to \infty} \frac{\mu_x(\partial_r B_R)}{\mu_x(B_R)} = 0 \text{ for all } r > 0\}.$$

Theorem 3.2.4. *The set A_0 of points which satisfy the Følner condition is T invariant and $\mu(A_0) = 1$.*

Proof. That $\mu(A_0) = 1$ was already shown in Theorem 3.2.3. So, we need to show invariance of A_0.

Suppose $x \in A_0 \subset X_0$. Let $r > 0$, $v \in F$, $r' = r + ||v||$, and $\epsilon > 0$. Choose N_0 such that $R_0 \geq N_0$ implies $\frac{\mu_x(\partial_{r'} B_{R_0})}{\mu_x(B_{R_0})} < \epsilon_0$, where $\epsilon_0 = \min(\sqrt{\epsilon}, 1 - \sqrt{\epsilon})$. Let

$N = N_0 + ||v||$ and suppose $R \geq N$. Notice $B_{R-||v||} \subset B_R(-v)$. We apply Proposition 3.1.4 to get

$$\begin{aligned} \mu_x(B_R) &< (1-\epsilon_0)\mu_x(B_{R-||v||}) \\ &\leq k_{v,x}(1-\epsilon_0)\mu_{T^v(x)}(B_R). \end{aligned}$$

Since $\partial_r B_R - v \subset \partial_{r'} B_R$, we may again apply Proposition 3.1.4 to obtain

$$\begin{aligned} \mu_{T^v(x)}(\partial_r B_R) &\leq \frac{1}{k_{v,x}}\mu_x(\partial_{r'} B_R) \\ &< \frac{1}{k_{v,x}}\epsilon_0 \mu_x(B_R) \\ &< \epsilon \mu_{T^v(x)}(B_R). \end{aligned}$$

The suppositions imply that $\mu_{T^v(x)}(B_R)$ is positive for large R, so the proof is complete. $\square$

Chapter 4

Ergodic Theorems on Actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$

Recent explorations of the ergodic theorem have involved the generalization of ergodic theorems beyond a transformation on a probability space to a group action on a probability space. The groups that are most often considered are $\mathbb{Z}^d$ and $\mathbb{R}^d$. Since the average for the ergodic theorems in Section 1.2 was taken over intervals $[0, n]$, the most natural extension is to average over $[0, n]^d$. In the first section of this chapter, we examine an example of a measure preserving, conservative action of $\mathbb{Z}^2$ on $[0, \infty)$ for which the ratio averages over the hypercubes $[0, n]^2$ diverge on a set of positive measure. This shows that the most natural extension of the Birkhoff, Hopf, and Hurewicz ergodic theorems to actions of $\mathbb{Z}^d$ does not hold. Nevertheless, versions of these theorems do exist for actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$, when the average is taken over hypercubes of the form $[-n, n]^d$. These results are reviewed in the second section of this chapter.

4.1 An Example of Divergence

We describe an example of a measure preserving, conservative action T of $\mathbb{Z}^2$ on all but a null set of $[0, \infty)$ such that the ratio averages

$$\frac{\sum_{v \in [0,n]^d} f \circ T^v}{\sum_{v \in [0,n]^d} g \circ T^v} \tag{4.1.1}$$

fail to converge a.e. as $n \to \infty$ for certain $f, g \in L^1([0,\infty))$. The action T^v has components $T^{(0,1)} = T^{(1,0)} = S$, where S is a measure-preserving, invertible transformation on all but a null set of $([0,1), \mathcal{L}, \lambda)$. Both the transformation S and the functions f and g are defined by a recursive procedure. A step in the recursion process extends the definitions of S, f, and g to a larger portion of the space. (Note: this example is original, but it is both inspired by and closely related to a construction given by Krengel and Brunel [13].) Implementation of the cutting and stacking procedure is used to gain the conservativity in our example, and we use a product space formulation to make the action free. This system is an improvement on that given by Krengel and Brunel because their example is neither conservative nor free. For a description of the cutting and stacking method, see [20, 22].

To begin the construction, let S take $[0,1)$ to $[1,2)$ by addition of one. We are able to describe the transformation S by a cutting and stacking procedure, so let the interval $[1,2)$ be stacked above $[0,1)$ and the transformation given by moving one step vertically (see Figure 4.1). Now, for $x \in [0,2)$, let

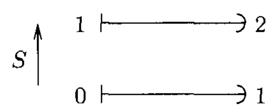

Figure 4.1: The first stage of the construction.

$$f(x) = 0 \text{ and } g(x) = \begin{cases} 1 \text{ if } 0 \leq x < 1, \\ 0 \text{ if } 1 \leq x < 2. \end{cases}$$

This completes the first stage of the construction.

We now perform the second stage of the construction. We cut the stack in half by a vertical slice and place the left half beneath the right half (see Figure 4.2). This means S remains the same on $[0,1)$, but now $[1,\frac{3}{2})$ is taken to $[\frac{1}{2},1)$ by subtraction of $\frac{1}{2}$. Stack 76 more intervals, each of length $\frac{1}{2}$, above $[\frac{3}{2},2)$. Thus, S maps $[\frac{3}{2},2)$ to $[2,\frac{5}{2})$, $[2,\frac{5}{2})$ to $[\frac{5}{2},3)$, ..., and $[39,\frac{79}{2})$ to $[\frac{79}{2},40)$ by addition of $\frac{1}{2}$ (see Figure 4.3). For $x \in [2,40)$, let

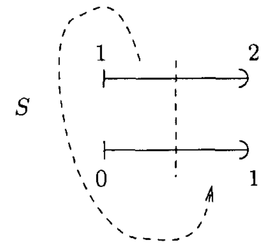

Figure 4.2: The second stage of the construction.

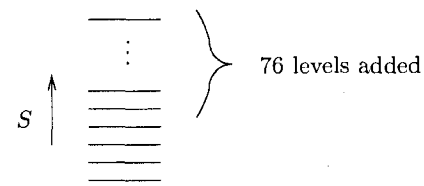

Figure 4.3: S at the second stage of the construction.

$$g(x) = 0 \text{ and } f(x) = \begin{cases} \frac{1}{6} \text{ if } \frac{39}{2} \leq x < 20), \\ 0 \text{ otherwise.} \end{cases}$$

This completes the second stage of the construction.

Suppose that stage $2n$ has been completed. Let i be the largest integer such that $S^i(0)$ is defined and let j be the largest number such that f and g are defined on $[0, j)$. The number i, then, is one less than the number of levels in our stack at the end of the previous stage. We make a vertical slice down the middle of the stack and put the left half underneath the right half. This defines $S^k(0)$ for $0 \leq k \leq 2i+1$ by moving vertically one level in the stack. Let N be an integer such that $\sqrt{N} \geq \sum_{k=0}^{2i+1}(k+1)(f \circ S^k)(0)$. Place $2N + 2i + 2$ new levels, all of the same length as those levels already in the stack, on top of the stack, with the new levels being taken consecutively from $[j, \infty)$ (see Figure 4.4). Let the $N + 2i + 2^{\text{th}}$ level be representing $[a, b)$ and the top level be representing $[c, d)$. For $x \in [j, d)$, let

$$f(x) = 0 \text{ and } g(x) = \begin{cases} \frac{1}{\sqrt{N}} \text{ if } a \leq x < b, \\ 0 \text{ otherwise.} \end{cases}$$

This completes the $2n + 1^{\text{th}}$ stage of the construction. Since $2n + 1$ is odd, we extended the locations for which g is positive on $[0, \infty)$. Choose an $x \in [0, 1]$ and

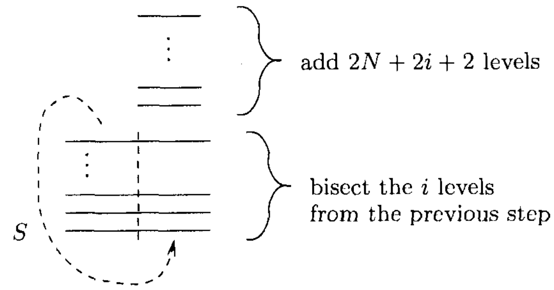

Figure 4.4: Stage $2n+1$ of the construction.

let N_0 be such that $S^{N_0}(x) \in [a,b)$. It can be seen in the construction above that $N_0 > N$. Thus,

$$\begin{aligned}
\frac{\displaystyle\sum_{v\in[0,N_0]^2}(f\circ T^v)(x)}{\displaystyle\sum_{v\in[0,N_0]^2}(g\circ T^v)(x)} &= \frac{\displaystyle\sum_{k=0}^{N_0}(k+1)(f\circ S^k)(x)}{\displaystyle\sum_{k=0}^{N_0}(k+1)(g\circ S^k)(x)} \\
&\leq \frac{\sqrt{N}}{(N_0+1)\frac{1}{\sqrt{N}}} \\
&< 1.
\end{aligned}$$

For stage $2n+2$, let $l = 2N+4i+4$, which is the largest integer for which $S^l(0)$ is defined. As before, make a vertical slice through the middle of the stack and place the left half underneath the right half. Let M be an integer such that $\sqrt{M} \geq 2\sum_{k=0}^{2l+1}(k+1)(g\circ S^k)(0)$ and place $2M+2l+2$ new levels of the same length on top of the stack, taken consecutively from $[d,\infty)$. Let the $M+2l+2^{\text{th}}$ level be representing $[r,s)$ and the top level be representing $[t,u)$. For $x\in[t,u)$, let

$$g(x) = 0 \text{ and } f(x) = \begin{cases} \frac{1}{\sqrt{M}} \text{ if } r \leq x < s, \\ 0 \text{ otherwise.} \end{cases}$$

This completes step $2n+2$ of the construction. Since $2n+2$ is even, we extended the locations for which f is positive on $[0,\infty)$. Now choose $x\in[0,1]$ and let M_0 be such that $S^{M_0}(x) \in [r,s)$. This implies $M_0 > M$ and

$$\begin{aligned}\frac{\sum\limits_{v\in[0,M_0]^2}(f\circ T^v)(x)}{\sum\limits_{v\in[0,M_0]^2}(g\circ T^v)(x)} &= \frac{\sum\limits_{k=0}^{M_0}(k+1)(f\circ S^k)(x)}{\sum\limits_{k=0}^{M_0}(k+1)(g\circ S^k)(x)}\\ &\geq \frac{(M_0+1)\frac{1}{\sqrt{M}}}{\frac{\sqrt{M}}{2}}\\ &> 2.\end{aligned}$$

This procedure is carried out to define a transformation S on $([0,\infty),\mathcal{L},\lambda)$ and functions $f, g \in L^1([0,\infty))$. A representation of the construction on the real number line is given in Figure 4.5.

For $A \subset [0,\infty)$, let

$$A^* := A \setminus \left\{\frac{k}{2^n} : k \in \{0,1,..\}, n \in \mathbb{N}\right\}.$$

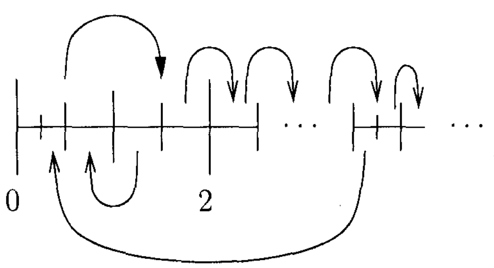

Figure 4.5: S is a measure preserving transformation $[0,\infty)$ that takes dyadic intervals to dyadic intervals linearly.

It may seem odd that we used an infinite measure space when up to this point we have only been working with probability spaces. This was done because the natural extension of the pointwise ergodic theorem is true for measure-preserving actions of $\mathbb{Z}^d$ or $\mathbb{R}^d$ on a probability space. What remains, then, is the non-measure-preserving case. It is easier to describe this construction as a measure-preserving action on an infinite space than as a nonsingular action on a probability space. Nevertheless, we can modify the system so that the action of $\mathbb{Z}^2$ is on a probability space. To do so, simply put a probability measure on $[0,\infty)$ in the following way: for a Lebesgue measurable $E \subset [0,\infty)$, let

$$\nu(E) = \sum_{i=1}^{\infty} \frac{\lambda(E \cap [i-1,i))}{2^i}.$$

Also let $\hat{f}(x) = f(x) \cdot 2^{\lceil x \rceil}$ and $\hat{g}(x) = g(x) \cdot 2^{\lceil x \rceil}$ for $x \in [0,\infty)$. We have $\hat{f}, \hat{g} \in L^1([0,\infty)^*, \mathcal{L}^*, \nu))$, and the ratio averages

$$\frac{\displaystyle\sum_{v \in [0,n]^d} \frac{d(\nu \circ T^v)}{d\nu}(\hat{f} \circ T^v)}{\displaystyle\sum_{v \in [0,n]^d} \frac{d(\nu \circ T^v)}{d\nu}(\hat{g} \circ T^v)}$$

diverge on $[0,1)^*$, since they are identical to the averages in (4.1.1).

In the ratio ergodic theorem of Hopf, the function g was assumed to be positive almost everywhere, whereas our function g is definitely not so. However, g may be added to an extremely small constant function on the probability space $([0,\infty), \mathcal{L}, \nu)$. The constant function can be made small enough so that the lim sup and lim inf still do not match, and the ratio average still diverges.

We now show that the action T is conservative. Since $\nu << \lambda$ and $\lambda << \nu$, conservativity will apply to the measure ν if proven for λ. Lebesgue measurable sets can be approximated from above arbitrarily well by open sets, so we first show recurrence for open sets. We use special types of recurrence that are stricter than conservativity [12].

Definition 4.1.1. *A transformation S is* **rigid** *if there exists a sequence of natural numbers $n_1, n_2, \dots$ such that for any measurable set A of finite measure,*

$$\liminf_{i \to \infty} \lambda(S^{n_i}(A) \cap A) = \lambda(A).$$

Definition 4.1.2. *A transformation S is* **partially rigid with factor r** *if there exists a sequence of natural numbers $n_1, n_2, \dots$ such that for any measurable set A of finite measure,*

$$\liminf_{i \to \infty} \lambda(S^{n_i}(A) \cap A) \geq r\lambda(A).$$

Rigidity implies partial rigidity, and partial rigidity implies conservativity. The reverse directions, however, do not hold. To be partially rigid, a transformation must have a sequence of the return times in which each set recurs to a certain fraction of its mass. To be rigid is to have a sequence of return times in which each set recurs to almost all of its mass. These two definitions have analogs for actions of $\mathbb{Z}^2$, for which the sequence of natural numbers n_i is replaced by a sequence of vectors in $\mathbb{Z}^2$.

We show the transformation S built earlier in this section is partially rigid. For each natural number i, let n_i be the height of the stack in stage i of the construction of S. Notice that each of the n_i intervals in the stack at stage i is of length $\frac{1}{2^{i-1}}$.

Lemma 4.1.3. *Any interval (a, b) has*

$$\liminf_{i \to \infty} \lambda(S^{n_i}((a,b)) \cap (a,b)) \geq \frac{1}{2}\lambda((a,b)).$$

Proof. Fix $\epsilon > 0$ and suppose (a, b) is nonempty (the result is trivial for an empty interval). Choose l such that $S((a,b))$ is defined no later than stage l. Suppose $k \geq l$, a_0, and b_0 are natural numbers such that $[\frac{a_0}{2^k}, \frac{b_0}{2^k}) \subset (a,b)$ and

$$\lambda\left((a,b) \setminus [\frac{a_0}{2^k}, \frac{b_0}{2^k})\right) < \epsilon\lambda((a,b)).$$

At stage k, then, all but less than a fraction ϵ of the mass of (a, b) is a union of levels of the stack at stage k. Looking forward to stage $k + 1$, each of these levels (which has now been cut into two pieces) has half of its mass return to the level when S is applied n_{k+1} times. Further, looking forward to stage $k + 2$, each level has half of its mass return to the level when S is applied n_{k+2} times. More generally, for any $i \geq k$,

$$\lambda\left(S^{n_i}([\frac{a_0}{2^k}, \frac{b_0}{2^k})) \cap [\frac{a_0}{2^k}, \frac{b_0}{2^k})\right) = \frac{1}{2}\lambda\left([\frac{a_0}{2^k}, \frac{b_0}{2^k})\right).$$

Since $[\frac{a_0}{2^k}, \frac{b_0}{2^k}) \subset (a, b)$, for any $i \geq k$,

$$\begin{aligned}
\lambda\left(S^{n_i}((a,b)) \cap (a,b)\right) &\geq \frac{1}{2}\lambda\left([\frac{a_0}{2^k}, \frac{b_0}{2^k})\right) \\
&\geq \frac{1}{2}(1-\epsilon)\lambda((a,b)).
\end{aligned}$$

Thus,

$$\liminf_{i\to\infty} \lambda\left(S^{n_i}((a,b)) \cap (a,b)\right) \geq \frac{1}{2}(1-\epsilon)\lambda((a,b))$$

for any $\epsilon > 0$ and the result follows. $\square$

Lemma 4.1.4. *The action S on $([0,\infty), \mathcal{L}, \lambda)$ is partially rigid with factor $\frac{1}{2}$.*

Proof. Suppose $A \in \mathcal{L}$ with $\lambda(A) < \infty$. Let $\epsilon > 0$ and U be an open set in $[0,\infty)$ such that $A \subset U$ and $\lambda(U \setminus A) < \epsilon$. Let $I_1, I_2, ..., I_k$ be pairwise disjoint open intervals contained in U such that $\lambda(U \setminus \cup_{j=1}^{k} I_j) < \epsilon$. Let l be a natural number such that for any $i \geq l$ and $1 \leq j \leq k$,

$$\lambda(S^{n_i}(I_j) \cap I_j) \geq (\frac{1}{2} - \epsilon)\lambda(I_j). \tag{4.1.2}$$

The existence of such an l follows from Lemma 4.1.3. For any $i \geq l$,

$$\lambda(S^{n_i}(A) \cap A) > \sum_{j=1}^{k} \lambda(S^{n_i}(I_j) \cap I_j) - 2\epsilon,$$

since $\lambda(\cup_{j=1}^{k} I_j \setminus A) < \epsilon$. We then use (4.1.2) and sum over j to get

$$\begin{aligned}\lambda(S^{n_i}(A) \cap A) &> (\frac{1}{2} - \epsilon)(1-\epsilon)\lambda(U) - 2\epsilon \\ &\geq (\frac{1}{2} - \epsilon)(1-\epsilon)\lambda(A) - 2\epsilon\end{aligned}$$

for any $i \geq l$. For any ϵ we can choose such an l, so

$$\liminf_{i\to\infty} \lambda(S^{n_i}(A) \cap A) \geq \frac{1}{2}\lambda(A).$$

$\square$

Corollary 4.1.5. *T is partially rigid with factor $\frac{1}{2}$.*

Proof. For $i \in \mathbb{N}$, let $v_i = (0, n_i)$, where n_i is the height of the stack in the ith stage of the construction of S. For any $A \in \mathcal{L}^*$,

$$\liminf_{i\to\infty} \lambda(T^{v_i}(A) \cap A) \geq \frac{1}{2}\lambda(A)$$

by Lemma 4.1.4. $\square$

We now modify the action T so that it is a free action. Let $(S^1, \mathcal{F}, \mu)$ be the unit circle in $\mathbb{R}^2$ with the Lebesgue σ-algebra and the Lebesgue probability measure, α_1 and α_2 be irrational numbers for which $\frac{\alpha_1}{\alpha_2}$ and $\alpha_1 - \alpha_2$ are also irrational, and π be the action of $\mathbb{Z}^2$ on S^1 whose component actions are rotation by α_1 and α_2. It is well known that an irrational rotation on the unit circle is rigid, is an isometry, and has the property that every orbit is dense with respect to the Euclidean metric from $\mathbb{R}/\mathbb{Z}$, in which the length of an interval on S^1 is equal to its measure. Now let $T \times \pi$ be the action of $\mathbb{Z}^2$ on the product space $([0,\infty)^* \times S^1, \mathcal{L}^* \times \mathcal{F}, \lambda \times \mu)$ in which T defines the action on the first coordinate and π gives the action on the second coordinate.

First, we note that $T \times \pi$ is a free action. This is true since π is free. Second, we note that the action $T \times \pi$ is partially rigid with factor $\frac{1}{2}$.

To see the partial rigidity, first consider a cylinder set $A \times B$ of $[0,\infty)^* \times S^1$ in which A and B are intervals and let $\epsilon > 0$. Let m_i be a sequence in which rotation of T_1 by $m_i(\alpha_1 - \alpha_2)$ is within $\frac{1}{i}$ of rotation by $n_i\alpha_1$, and notice this sequence is independent of B. Such m_i can be chosen because rotation by $\alpha_1 - \alpha_2$ is an isometry and has that every orbit is dense. Let k_1 be a natural number such that $i \geq k_1$ implies $\lambda(T^v(A) \cap A) \geq \frac{1}{2}(1-\epsilon)\lambda(A)$ for any $v \in \mathbb{Z}^2$ with $||v|| = n_i$. Notice that rotation by $\alpha_1 - \alpha_2$ is the same as $\pi^{(1,-1)}$, and let k_2 be a natural number so that $i \geq k_2$ implies $\mu(\pi^{(n_i+m_i,-m_i)}(B) \cap B) \geq (1-\epsilon)\nu(B)$. For $i \geq \max(k_1, k_2)$,

$$(\lambda \times \mu)((T \times \pi)^{(n_i+m_i,-m_i)}(A \times B) \cap A \times B) \geq \frac{1}{2}(1-\epsilon)^2(\lambda \times \mu)(A \times B).$$

This gives partial rigidity of $T \times \pi$ with factor $\frac{1}{2}$ on cylinder sets. To prove partial rigidity with factor $\frac{1}{2}$ of general $\mathcal{L} \times \mathcal{F}$ sets, we can approximate from above by cylinder sets as we did in the proof of Corollary 4.1.4.

We may take functions f and g in $L^1(\lambda \times \mu)$ that only depend on the first coordinate for which the ratio average over $[0,n]^2$ diverges on a set of positive measure, using the construction given above. We then see that the most natural extension of the ratio ergodic theorem for actions of $\mathbb{Z}$ does not hold for conservative, free actions of $\mathbb{Z}^2$.

Why do we get divergence of the ratio averages in higher dimensions while it can be proven that they converge in one dimension? One explanation goes back to the Besicovitch Covering Lemma. A sequence of hypercubes $\{[0, n_k]^d\}_{k=1}^{\infty}$ share a common corner rather than sharing a center. The Besicovitch Covering Lemma does not hold, however, on these types of sets. To display what is meant by this, consider the collection of sets $\{[x, 1]^d : 0 \leq x < 1\}$. If the Besicovitch Covering Lemma held, then we could find some finite number of subcollections such that each subcollection is disjoint and the union of sets in all subcollections covers $\{(x, x, ..., x) \in \mathbb{R}^d : x \in (0, 1]\}$. However, any two of these sets intersect nontrivially and no finite subcollection will cover the diagonal in $(0, 1]^d$, so the Besicovitch Covering Lemma does not hold. (This reasoning also applies to the $d = 1$ case, but the Besicovitch Covering Lemma is not used in the proof of ergodic theorems in this setting. It is, however, a standard tool in proving ergodic theorems on actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$.)

4.2 Ratio Ergodic Theorems on Actions of $\mathbb{Z}^d$

While versions of the Birkhoff, Hopf, and Hurewicz ergodic theorems do not hold for averages over $[0, n]^d$, Feldman and Hochman have shown that the theorems do have analogs for actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$ when averaging over hypercubes centered at the origin ($[-n, n]^d$).

Definition 4.2.1. *A measure space $(X, \mathcal{F}, \mu)$ is called a* **Polish space** *if there exists a metric on X such that the Borel sets generate $\mathcal{F}$ and the metric is complete and separable.*

A Polish probability space is also a standard probability space [16], so the Polish space assumed by Feldman is more restrictive than the standard Borel probability space we assume.

Theorem 4.2.2. *(Feldman) [6] (2007) Suppose T is a measurable, invertible, non-singular, conservative action of $\mathbb{Z}^d$ on the Polish probability space $(X, \mathcal{F}, \mu)$ such that the component actions $T_i := T^{e_i}$ of $\mathbb{Z}$ are also conservative. Then for any $f, g \in L^1(\mu)$ with $E(g|\mathcal{I}) > 0$ a.e., the ratio averages*

$$\frac{\displaystyle\sum_{v \in B_n} f \circ T^{-v} \cdot \frac{d(T^v)^*\mu}{d\mu}}{\displaystyle\sum_{v \in B_n} g \circ T^{-v} \cdot \frac{d(T^v)^*\mu}{d\mu}} \tag{4.2.1}$$

converge to $\frac{E(f|\mathcal{I})}{E(g|\mathcal{I})}$ almost everywhere, where $\mathcal{I}$ is the σ-algebra of sets which are invariant under the action T.

The ratio averages in (4.2.1) look like somewhat of a compromise between those of Hopf (1.2.2) and Hurewicz (1.2.3). Without loss of generality, the function g can be assumed to be one (see Corollary 5.1.4). Hochman extended Feldman's result by proving the same averages converge a.e. without the assumption of directional conservativity or conservativity of the action.

The ratio ergodic theorem stated and proven in Section 5 improves the above results by allowing for singularity of the dynamical system. The action T is assumed to be Borel and free, but there is no connection, beyond being Borel, that is assumed between this action and the measure μ. This can be seen as a version of the Hurewicz ergodic theorem for actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$ (or, rather, of the extension of Hurewicz by Oxtoby that assumes neither non-singularity nor conservativity [15]).

Every ergodic theorem mentioned thus far is proven by a maximal inequality. With a maximal inequality in hand, convergence of the averages is reduced to finding a dense family in L^1 for which convergence can be shown. This is typically taken to be the set of coboundaries, $\{f - f \circ T^v : f \in L^1(\mu), v \in F\}$. Feldman and Hochman use a maximal inequality that was proven by Lindenstrauss and Rudolph [14]. The first step in this method (proving a maximal inequality) has been completed for the singular case by Rudolph [21]. However, our proof bypasses the usual maximal inequality, instead using the Følner condition found in Section 3.2. With the Følner condition in hand, convergence of the ratio average is then proven for all $f \in L^1$ instead of using the set of coboundaries or some other dense family.

Chapter 5

An Ergodic Theorem for Borel Actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$

We can now state and prove an ergodic theorem for Borel actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$.

Theorem 5.0.3. *If T is a free Borel action of $F(=\mathbb{Z}^d$ or $\mathbb{R}^d)$ on the standard Borel probability space $(X, \mathcal{F}, \mu)$ and $f \in L^1(\mu)$, then*

$$\lim_{R\to\infty} \frac{1}{\mu_x(B_R)} \int_{B_R} f \circ T^{-v}(x) d\mu_x$$

converges for μ-a.e. x. Furthermore, denoting the a.e. pointwise limit as $\hat{f}(x)$, the averages converge to $\hat{f}$ in $L^1(\mu)$ and $\hat{f} = E(f|\mathcal{I})$, where $\mathcal{I}$ is the σ-algebra of sets which are invariant under the action T.

Before proving Theorem 5.0.3, we review a similar result. Suppose $\mathcal{H}^n$ is an n dimensional real hyperbolic space. There are $n-1$ dimensional spheres, called horospheres, which are perpendicular to the geodesics. These spheres are all tangent to $\partial(\mathcal{H}^n)$, and the collection of horospheres covers $\mathcal{H}^n$. In 1982, Rudolph used a Følner condition to show a mean ergodic theorem on these horospheres. This implied that the geodesic flow, when equipped with a natural measure [23], is isomorphic to a Bernoulli flow [19].

5.1 Proof of the Ratio Ergodic Theorem on Borel Actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$

Suppose $f \in L^1(\mu)$ is a nonnegative function and for $A \in \mathcal{F}$ let

$$\theta(A) := \int_A f d\mu.$$

We construct the diffused measure on θ as described in Section 3 and get Borel probability measures θ_x on F for $x \in Y_0'$, an invariant set of full θ measure. Let

$$Y_0 := Y_0' \cup \{x \in X \setminus Y_0' : f(T^v(x)) = 0 \text{ for } m\text{-a.e. } v \in F\}.$$

For $x \in Y_0 \setminus Y_0'$, let $\theta_x = 0$, the trivial measure on F.

Lemma 5.1.1. *Y_0 is T-invariant and of full μ measure.*

Proof. Invariance of Y_0 is obvious, so we need to show that $\mu(Y_0) = 1$. Let $Y^* = X \setminus Y_0$ and suppose, for the sake of contradiction, that $\mu(Y^*) > 0$. For each $x \in Y^*$, let $E_x = \{v \in F : f(T^v(x)) > 0\}$. Notice $m(E_x) > 0$ for each $x \in Y^*$. Using continuity from below, choose $N^* > 0$ and a subset $Y_{N^*} \subset Y^*$ such that $\mu(Y_{N^*}) > 0$ and for each $x \in Y_{N^*}$, $m(E_x \cap B_{N^*}) > 0$. We have

$$\int_{B_{N^*}} \int_{Y^*} f d\mu dm = 0,$$

since $\theta(Y^*) = 0$ implies the inside integral is zero. We twist the integral (notice Y^* is invariant) and switch the order of integration to get

$$\int_{Y^*} \int_{B_{N^*}} f \circ T^v(x) dm d\mu = 0.$$

This implies $\mu(Y^*) = 0$, since the inside integral is now positive for every $x \in Y_{N^*}$, which is a contradiction and completes the proof. □

Lemma 5.1.2. *For $x \in X_0 \cap Y_0$, there is a real number k_x such that*

$$\frac{d\theta_x}{d\mu_x}(v) = k_x \cdot f \circ T^{-v}(x) \tag{5.1.1}$$

for μ_x a.e. v.

Proof. Let $x \in X_0 \cap Y_0$. If $x \notin Y_0'$, then let $k_x = 0$ and (5.1.1) holds. Suppose $x \in Y_0'$ and let M_x be the minimal natural number M such that $\theta_x(B_M) > 0$. Since $\theta << \mu$ and $\theta_x << \mu_x$, we have $M_x \geq N_x$.

First, we calculate $\frac{d\hat{\theta}_N}{d\hat{\mu}_N}$ for $N > 0$. Choose $N > 0$. For any measurable $S \subset X \times B_N$, we untwist the measure $\hat{\theta}_N$ to get

$$\hat{\theta}_N(S) = \frac{1}{m(B_N)} \int_{I(S)} d\theta \times m.$$

We now use the definition of θ and retwist the measure:

$$\hat{\theta}_N(S) = \frac{1}{m(B_N)} \int_{I(S)} f(x) d\mu \times m = \int_S f \circ T^{-v}(x) d\hat{\mu}_N.$$

This shows that $\frac{d\hat{\theta}_N}{d\hat{\mu}_N} = f \circ T^{-v}$.

Second, we show that $\frac{\theta_x(B_N)}{\int_{B_N} f \circ T^{-v} d\mu_x}$ is constant in N for $N \geq M_x$. Suppose $N_2 > N_1 \geq M_x$. We apply (3.1.4) to get

$$\begin{aligned}
\frac{\left(\frac{\theta_x(B_{N_1})}{\int_{B_{N_1}} f \circ T^{-v} d\mu_x}\right)}{\left(\frac{\theta_x(B_{N_2})}{\int_{B_{N_2}} f \circ T^{-v} d\mu_x}\right)} &= \lim_{n \to \infty} \frac{\hat{\theta}_{N_2}(A_n \times B_{N_1})}{\hat{\theta}_{N_2}(A_n \times B_{N_2})} \cdot \lim_{n \to \infty} \frac{\int_{A_n \times B_{N_2}} f \circ T^{-v} d\hat{\mu}_{N_2}}{\int_{A_n \times B_{N_1}} f \circ T^{-v} d\hat{\mu}_{N_2}} \\
&= \lim_{n \to \infty} \frac{\int_{A_n \times B_{N_1}} f \circ T^{-v} d\hat{\mu}_{N_2}}{\int_{A_n \times B_{N_2}} f \circ T^{-v} d\hat{\mu}_{N_2}} \cdot \lim_{n \to \infty} \frac{\int_{A_n \times B_{N_2}} f \circ T^{-v} d\hat{\mu}_{N_2}}{\int_{A_n \times B_{N_1}} f \circ T^{-v} d\hat{\mu}_{N_2}},
\end{aligned}$$

which is one. Let $k_x = \frac{\theta_x(B_N)}{\int_{B_N} f \circ T^{-v} d\mu_x}$ for $N \geq M_x$.

Finally, we characterize $\frac{d\theta_x}{d\mu_x}$. The Radon-Nikodym derivative $\frac{d\theta_x}{d\mu_x}$ is the unique Borel function of F such that $\theta_x(E) = \int_E \frac{d\theta_x}{d\mu_x} d\mu_x$ for any Borel set $E \subset F$. Notice that $k_x \cdot f \circ T^v(x)$, as a function of v, is Borel. Let $E \subset B_N$ be Borel and $N \geq M_x$. Again, we use the sets A_n to identify the diffused measure:

$$\begin{aligned}
\theta_x(E) &= \theta_x(B_N) \lim_{n \to \infty} \frac{\hat{\theta}_N(A_n \times E)}{\hat{\theta}_N(A_n \times B_N)} \\
&= \theta_x(B_N) \lim_{n \to \infty} \frac{\int_{A_n \times E} f \circ T^{-v} d\hat{\mu}_N}{\int_{A_n \times B_N} f \circ T^{-v} d\hat{\mu}_N} \\
&= \theta_x(B_N) \frac{\int_E f \circ T^{-v} d\mu_x}{\int_{B_N} f \circ T^{-v} d\mu_x} \\
&= \int_E k_x \cdot f \circ T^{-v} d\mu_x.
\end{aligned}$$

The set E was assumed to be bounded. Continuity from below on θ_x and μ_x completes the result. $\square$

We now define some new notation:

$$A_n(f,x) := \frac{1}{\mu_x(B_n)} \int_{B_n} f \circ T^{-v}(x) d\mu_x.$$

Also, let $\nu_x = k_x \cdot \mu_x$ for $x \in Y_0$ and $\nu_x = \mu_x$ for $x \in X_0 \setminus Y_0$, where k_x is that from Lemma 5.1.2. Thus, $A_n(f,x) = \frac{\theta_x(B_n)}{\nu_x(B_n)}$ for $x \in X_0$. Recall that we want to show pointwise convergence of $A_n(f,x)$ outside a set of measure zero. Let

$$A_{\alpha,\beta} := \left\{ x : \liminf_{n\to\infty} A_n(f,x) < \alpha < \beta < \limsup_{n\to\infty} A_n(f,x) \right\}.$$

To prove a.e. convergence in Theorem 5.0.3, we need to show $\mu(A_{\alpha,\beta}) = 0$ for any $\alpha < \beta$.

Lemma 5.1.3. *Let $\alpha < \beta$ be given. Then there is a subset $A^*_{\alpha,\beta}$ of $A_{\alpha,\beta}$ of the same measure which is T-invariant.*

Proof. For a.e. $x \in A_{\alpha,\beta}$, $\lim_{R\to\infty} \frac{\nu_x(\partial_r B_R)}{\nu_x(B_R)} = 0$ and $\lim_{R\to\infty} \frac{\theta_x(\partial_r B_R)}{\theta_x(B_R)} = 0$ for all $r > 0$ by Theorem 3.2.3. Let $A^*_{\alpha,\beta}$ be the set of such x, and we see $A^*_{\alpha,\beta}$ has the same measure as $A_{\alpha,\beta}$. Let $x \in A^*_{\alpha,\beta}$. Fix $w \in F$. We would like to know that $T^w(x) \in A^*_{\alpha,\beta}$. Theorem 3.2.4 says that the Følner condition holds for $T^w(x)$, so we only need to show $T^w(x) \in A_{\alpha,\beta}$. Let $b = \limsup_{n\to\infty} A_n(f,x)$. Let $n > M_x$ and $w \in F$. We can write $A_n(f, T^w(x))$ in terms of θ_x and ν_x:

$$\begin{aligned} A_n(f,T^w(x)) &= \frac{1}{\mu_x(B_n(-w))} \int_{B_n(-w)} f \circ T^{-v}(x) d\mu_x \\ &= \frac{\theta_x(B_n(-w))}{\nu_x(B_n(-w))} \\ &= \frac{\theta_x(B_n)}{\nu_x(B_n)} \left(\frac{\theta_x(B_n(-w))}{\theta_x(B_n)} \cdot \frac{\nu_x(B_n)}{\nu_x(B_n(-w))} \right). \end{aligned}$$

By the Følner condition, choose N such that $n \geq N$ implies

$$\frac{\theta_x(B_n(-w))}{\theta_x(B_n)} \cdot \frac{\nu_x(B_n)}{\nu_x(B_n(-w))} \geq \frac{2b + 2\beta}{3b + \beta}.$$

Now for every $n \geq N$ with $A_n(f,x) > \frac{3b+\beta}{4}$ (notice there are an infinite number of such n),

$$A_n(f, T^w(x)) \geq \frac{3b+\beta}{4} \cdot \frac{2b+2\beta}{3b+\beta} = \frac{b+\beta}{2} > \beta.$$

Thus, $\limsup_{n\to\infty} A_n(f, T^w(x)) > \beta$. An analogous argument shows

$$\liminf_{n\to\infty} A_n(f, T^w(x)) < \alpha,$$

so $T^w(x) \in A^*_{\alpha,\beta}$. □

Proof. We now prove Theorem 5.0.3. First, we show a.e. convergence of $A_n(f,x)$. For the sake of contradiction, suppose $\mu(A_{\alpha,\beta}) > 0$. Thus, $\mu(A^*_{\alpha,\beta}) > 0$. Notice $A^*_{\alpha,\beta} \subset Y_0$. We groom the set $A^*_{\alpha,\beta}$ to obtain a structure on the pairs n, x which have $A_n(f,x) > \beta$. Let $r_1 = 0$. Choose $A_1 \subset A^*_{\alpha,\beta}, R_1 > r_1$, and $\rho_1 : A_1 \to [r_1, R_1]$ such that $\mu(A_1) > (\frac{\alpha}{\beta})^{2^{-5}} \mu(A^*_{\alpha,\beta})$ and each $x \in A_1$ has $A_{\rho_1(x)}(f,x) > \beta$. We now inductively define measurable sets $A_k \subset A^*_{\alpha,\beta}$, positive numbers r_k and R_k, and functions $\rho_k : A_k \to [r_k, R_k]$ for all $k \in \mathbb{N}$.

Suppose $A_{k-1}, r_{k-1}, R_{k-1}$, and ρ_{k-1} have been defined. We let $A^*_{k-1} \subset A_{k-1}$ and $r_k > R_{k-1}$ such that $\mu(A^*_{k-1}) > (\frac{\alpha}{\beta})^{2^{-4-(k-1)}} \mu(A_{k-1})$ and $x \in A^*_{k-1}$ implies $\frac{\theta_x(\partial_{R_{k-1}} B_n)}{\theta_x(B_n)} < 1 - \sqrt[3]{\frac{\alpha}{\beta}}$ for all $n \geq r_k$. Now choose $A_k \subset A^*_{k-1}, R_k$, and $\rho_k : A_k \to [r_k, R_k]$ such that $x \in A_k$ implies $A_{\rho_k(x)}(f,x) > \beta$ and $\mu(A_k) > (\frac{\alpha}{\beta})^{2^{-4-k}} \mu(A^*_{k-1})$. Notice these properties also hold for the base case $k = 1$. Now let $A = \cap_k A_k$ and we have $\mu(A) \geq \sqrt[8]{\frac{\alpha}{\beta}} \cdot \mu(A^*_{\alpha,\beta})$ by continuity from above on finite measures.

We proceed with a Besicovitch covering argument. Let C be the Besicovitch constant for F, and let K be such that $1 - (\frac{C-1}{C})^K \geq \sqrt[3]{\frac{\alpha}{\beta}}$. Also, choose $y \in A$ and $N > R_K$ such that

$$\begin{aligned} \frac{\mu_y(\partial_{R_K} B_N)}{\mu_y(B_N)} &< \sqrt[4]{\frac{\alpha}{\beta}} - \sqrt[3]{\frac{\alpha}{\beta}}, && (5.1.2) \\ \frac{\mu_y(A_y \cap B_N)}{\mu_y(B_N)} &> \sqrt[4]{\frac{\alpha}{\beta}}, \text{ and} && (5.1.3) \\ A_N(f,y) &< \alpha, && (5.1.4) \end{aligned}$$

where the second estimate uses Lemma 5.1.3. Let $\hat{A} = A_y \cap B_{N-R_K}$. Apply the Besicovitch Covering Lemma to the set $\hat{A}$ with the ball $B_{\rho_K(T^v(y))}(v)$ corresponding to each $v \in \hat{A}$ to get a subset $\hat{A}_K \subset \hat{A}$ with $\mu_y(\hat{A}_K) \geq \frac{\mu_y(\hat{A})}{C}$ such that $\{B_{\rho_K(T^v(y))}(v)\}_{v \in \hat{A}_K}$ is a pairwise disjoint collection of balls in F.

Suppose that $\hat{A}_{K-j}$ has been chosen for $0 \leq j < K-1$. We inductively define subsets $\hat{A}_{K-j}$ for $0 \leq j < K-1$. Apply the Besicovitch Covering Lemma to the set $\hat{A} \setminus \cup_{i=0}^{j} \hat{A}_i$ with corresponding balls $B_{\rho_{K-j-1}(T^v(y))}(v)$ to get

$$\hat{A}_{K-j-1} \subset \hat{A} \setminus \bigcup_{i=0}^{j} \hat{A}_i$$

such that $\{B_{\rho_{K-j-1}(T^v(y))}(v)\}_{v \in \hat{A}_{K-j-1}}$ is a pairwise disjoint collection of balls in F and $\mu_y(\hat{A}_{K-j-1}) \geq \frac{1}{C}\mu_y(\hat{A} \setminus \cup_{i=0}^{j} \hat{A}_i)$.

This procedure terminates after defining $\hat{A}_1$, and

$$\mu_y\left(\hat{A} \setminus \bigcup_{i=1}^{K} \hat{A}_i\right) \leq \left(\frac{C-1}{C}\right)^K \mu_y(\hat{A}). \tag{5.1.5}$$

We now estimate $A_N(f,y)$ by using the balls $\{B_{\rho_j(T^v(y))}(v)\}_{v \in \hat{A}_j, 1 \leq j \leq K}$ to cover most of the μ_y mass of $\hat{A}$. For $1 \leq j \leq K$, let

$$\mathcal{C}_j := \{B_{\rho_j(T^v(y))}(v)\}_{v \in \hat{A}_j}$$

and

$$\mathcal{C} := \bigcup_{B \in \cup \mathcal{C}_j} B \setminus \partial_{R_{j-1}} B,$$

where $R_0 = 0$. Note that $\mathcal{C}$ is a disjoint union. Also, every $B \in \mathcal{C}_j$ has $\frac{\theta_y(B)}{\nu_y(B)} > \beta$ and $\frac{\theta_y(\partial_{R_{j-1}} B)}{\theta_y(B)} < 1 - \sqrt[3]{\frac{\alpha}{\beta}}$, so

$$\frac{\theta_y(B \setminus \partial_{R_{j-1}} B)}{\nu_y(B)} > \sqrt[3]{\frac{\alpha}{\beta}}\beta.$$

Let $\mathcal{D} = \bigcup_{B \in \cup \mathcal{C}_j} B$. This implies $\frac{\theta_y(\mathcal{C})}{\nu_y(\mathcal{D})} > \sqrt[3]{\frac{\alpha}{\beta}}\beta$, and since $\mathcal{C} \subset B_N$,

$$\frac{\theta_y(B_N)}{\nu_y(\mathcal{D})} > \sqrt[3]{\frac{\alpha}{\beta}}\beta. \tag{5.1.6}$$

Further, $\bigcup_{i=1}^{K} \hat{A}_i$ is covered by $\mathcal{D}$, so (5.1.5) gives

$$\mu_y(\mathcal{D}) \geq \left(1 - \left(\frac{C-1}{C}\right)^K\right) \mu_y(\hat{A}) \geq \sqrt[3]{\frac{\alpha}{\beta}} \mu_y(\hat{A}). \tag{5.1.7}$$

We have

$$\begin{aligned} \mu_y(\hat{A}) &\geq \mu_y(A_y \cap B_N) - \mu_y(\partial_{R_K} B_N) \\ \mu_y(\hat{A}) &> \left(\sqrt[4]{\frac{\alpha}{\beta}} - \left(\sqrt[4]{\frac{\alpha}{\beta}} - \sqrt[3]{\frac{\alpha}{\beta}}\right)\right) \mu_y(B_N) \\ \mu_y(\hat{A}) &> \sqrt[3]{\frac{\alpha}{\beta}} \beta \mu_y(B_N) \end{aligned} \tag{5.1.8}$$

by (5.1.2) and (5.1.3). Recall that ν_y and μ_y are equivalent, and

$$\begin{aligned} A_N(f, y) &= \frac{\theta_x(B_N)}{\nu_y(\mathcal{D})} \cdot \frac{\nu_y(\mathcal{D})}{\nu_y(\hat{A})} \cdot \frac{\nu_y(\hat{A})}{\nu_y(B_N)} \\ &= \frac{\theta_x(B_N)}{\nu_y(\mathcal{D})} \cdot \frac{\mu_y(\mathcal{D})}{\mu_y(\hat{A})} \cdot \frac{\mu_y(\hat{A})}{\mu_y(B_N)} \\ &> \sqrt[3]{\frac{\alpha}{\beta}} \sqrt[3]{\frac{\alpha}{\beta}} \sqrt[3]{\frac{\alpha}{\beta}} \beta \\ &= \alpha, \end{aligned}$$

which contradicts (5.1.4). So, $\mu(A_{\alpha,\beta}) = 0$ for all $\alpha < \beta$ and $A_n(f)$ converges μ a.e. to a function $\hat{f}$.

The rest of the proof is standard. We next show that this convergence is also in $L^1(\mu)$. Fubini's Theorem can be applied to show that $\int A_n(f) d\mu = \int f d\mu$ for any $n > 0$. Thus,

$$\begin{aligned} \int \hat{f} d\mu &\leq \int \lim_{n \to \infty} A_n(f) d\mu \\ &\leq \liminf_{n \to \infty} \int A_n(f) d\mu \\ &= \int f d\mu, \end{aligned}$$

which gives that

$$||\hat{f}||_1 \leq ||f||_1. \tag{5.1.9}$$

Now suppose $g \in L^1(\mu)$ is a nonnegative, bounded function. This implies $A_n(g)$ has the same bound as g, and $A_n(g) \to \hat{g}$ in $L^1(\mu)$ by the Lebesgue Dominated Convergence Theorem. Also,

$$||A_n(f) - \hat{f}||_1 \leq ||A_n(f) - A_n(g)||_1 + ||A_n(g) - \hat{g}||_1 + ||\hat{g} - \hat{f}||_1.$$

We know $||A_n(f) - A_n(g)||_1 = ||f - g||_1$ and $||\hat{g} - \hat{f}||_1 \leq ||f - g||_1$ by (5.1.9), so for large enough n, $||A_n(g) - \hat{g}||_1 < ||f - g||_1$. Additionally,

$$||A_n(f) - \hat{f}||_1 \leq 3||f - g||_1$$

for large n. Since the bounded functions are dense in L^1, we have that $A_n(f) \to \hat{f}$ in L^1.

Finally, we need $\hat{f} = E(f|\mathcal{I})$. It is enough to show that $\hat{f}$ is invariant under the action T and $\int_A \hat{f} d\mu = \int_A f d\mu$ for any invariant set A. By Proposition 3.1.4,

$$A_n(f, T^w(x)) = \frac{\theta_{T^w(x)}(B_n)}{\nu_{T^w(x)}(B_n)} = \frac{\theta_x(B_n(-w))}{\nu_x(B_n(-w))},$$

and T invariance follows from the Følner condition. Suppose $T^{-w}(A) = A$ for all $w \in F$ and $A \in \mathcal{F}$. By L^1 convergence and Fubini's Theorem,

$$\begin{aligned} \int_A \hat{f} d\mu &= \lim_{n \to \infty} \int_A A_n(f) d\mu \\ &= \int_A f d\mu. \end{aligned}$$

We have proven Theorem 5.0.3 for nonnegative $f \in L^1(\mu)$. For $f \in L^1(\mu)$, write $f = f^+ - f^-$ such that f^+, f^- are nonnegative $L^1(\mu)$ functions, and the result follows. $\square$

A ratio ergodic theorem is a theorem about the convergence of a weighted average. In what sense is Theorem 5.0.3 a weighted average? First, because the measure μ can be taken to be any standard Borel probability measure, the measure μ may be altered to adjust the weighting of the average. Changing the measure

μ, however, may change which functions are L^1 and therefore change the functions to which the theorem applies. Alternatively, we have the following result that is a more traditional notion of a weighted average.

Corollary 5.1.4. *Suppose T is a free Borel action of F ($= \mathbb{Z}^d$ or $\mathbb{R}^d$) on the standard Borel probability space $(X, \mathcal{F}, \mu)$. Then for any $f, g \in L^1(\mu)$ with $E(g|\mathcal{I}) > 0$ μ-a.e.,*

$$\lim_{n\to\infty} \frac{\int_{B_n} f \circ T^{-v}(x) d\mu_x}{\int_{B_n} g \circ T^{-v}(x) d\mu_x} = \frac{E(f|\mathcal{I})(x)}{E(g|\mathcal{I})(x)}$$

for μ-a.e. x.

Proof. Notice

$$\frac{\int_{B_n} f \circ T^{-v}(x) d\mu_x}{\int_{B_n} g \circ T^{-v}(x) d\mu_x} = \frac{A_n(f,x)}{A_n(g,x)}.$$

Apply Theorem 5.0.3. □

5.2 Examples

We now look at some examples.

1. Suppose T is a measure-preserving, free action of F on the standard Borel probability space $(X, \mathcal{F}, \mu)$. In this case, the twisting of the measure actually does not change the measure at all, and the diffused measures μ_x are equivalent to m. Thus,

$$\begin{aligned} A_n(f,x) &= \frac{1}{\mu_x(B_n)} \int_{B_n} f \circ T^{-v}(x) d\mu_x(v) \\ &= \frac{1}{m(B_n)} \int_{B_n} f \circ T^{-v}(x) dm, \end{aligned}$$

which is just the average of f at x over the ball B_n. Theorem 5.0.3 implies that this average converges a.e. to $E(f|\mathcal{I})$ as $n \to \infty$. In particular, if T is ergodic, then the average converges a.e. to the expectation of f.

2. Suppose T is a free, nonsingular action of $\mathbb{Z}^d$ or $\mathbb{R}^d$ on the standard Borel probability space $(X, \mathcal{F}, \mu)$. For $f \in L^1(\mu)$, we show that the averages $A_n(f,x)$ are the same as those considered by Feldman and Hochman.

Suppose $B \in \mathcal{B}$, $m(B) = 0$, and $x \in X_0$. We have $I^*(\mu \times m)(X \times B) = 0$, which implies $\hat{\mu}_N(X \times B) = 0$ for any N. So, $\mu_x(B) = 0$ and we see $\mu_x << m$. Thus, there exists a Radon-Nikodym derivative $\frac{d\mu_x}{dm}$. We will see that $\frac{d\mu_x}{dm}(v) = c_x \cdot \frac{d(T^v)^*\mu}{d\mu}(x)$ for some constant c_x.

Suppose $E \subset B_N$ is Borel and $\mu_x(E) > 0$. We use sets A_n that decrease to x to calculate $\mu_x(E)$:

$$\begin{aligned}
\mu_x(E) &= \mu_x(B_N) \lim_{n\to\infty} \frac{\hat{\mu}_N(A_n \times E)}{\hat{\mu}_N(A_n \times B_N)} \\
&= \mu_x(B_N) \lim_{n\to\infty} \frac{\int_E \int_{A_n} d(T^v)^*\mu dm}{\int_{B_N} \int_{A_n} d(T^v)^*\mu dm} \\
&= \mu_x(B_N) \lim_{n\to\infty} \frac{\int_E \left(\int_{A_n} \frac{d(T^v)^*\mu}{d\mu} d\mu\right) dm}{\int_{B_N} \left(\int_{A_n} \frac{d(T^v)^*\mu}{d\mu} d\mu\right) dm}. \qquad (5.2.1)
\end{aligned}$$

Since $(X, \mathcal{F}, \mu)$ is a standard Borel probability space, it is measurably isomorphic to an interval with the Borel sets and Lebesgue measure, along with at most countably many atoms. It is known that the limit as $n \to \infty$ of the inner integrals in (5.2.1) is equal to $\frac{d(T^v)^*\mu}{d\mu}(x)$ when the space is an interval, and it is not hard to see that the same is true for a point mass x. We can use the isomorphism which corresponds to the sets A_n to take the limit. This gives

$$\mu_x(E) = \mu_x(B_N) \frac{\int_E \frac{d(T^v)^*\mu}{d\mu}(x) dm}{\int_{B_N} \frac{d(T^v)^*\mu}{d\mu}(x) dm}.$$

Thus, for any $F \subset B_N$ which also has $\mu_x(F) > 0$,

$$\frac{\mu_x(E)}{\mu_x(F)} = \frac{\int_E \frac{d(T^v)^*\mu}{d\mu}(x) d\mu}{\int_F \frac{d(T^v)^*\mu}{d\mu}(x) d\mu}.$$

This is true for any B_N and sets $E, F \subset B_N$ of positive μ_x measure, so $\frac{d\mu_x}{dm}(v) = c_x \frac{d(T^v)^*\mu}{d\mu}(x)$ for some constant c_x.

We now have that for $n > 0, f \in L^1(\mu)$, and a.e. $x \in X$,

$$\begin{aligned} A_n(f,x) &= \frac{1}{\mu_x(B_N)} \int_{B_N} f \circ T^{-v} d\mu_x \\ &= \frac{\int_{B_N} f \circ T^{-v}(x) \cdot c_x \cdot \frac{d(T^v)^*\mu}{d\mu}(x) dm}{\int_{B_N} c_x \cdot \frac{d(T^v)^*\mu}{d\mu}(x) dm} \\ &= \frac{\int_{B_N} f \circ T^{-v}(x) \cdot \frac{d(T^v)^*\mu}{d\mu}(x) dm}{\int_{B_N} \frac{d(T^v)^*\mu}{d\mu}(x) dm}. \end{aligned} \tag{5.2.2}$$

The expression in (5.2.2) is exactly the ratio averages considered by Feldman and Hochman. This shows that Theorem 5.0.3 is an extension of the ratio ergodic theorems of Feldman and Hochman, which we reviewed in section 4.2.

3. Consider $X = \mathbb{R}^d$ with the Borel σ-algebra and let T be the action of $\mathbb{Z}^d$ given by translation. For any nontrivial standard Borel measure μ, this system is not conservative, since $[0,1)^2$ is a wandering set and $\cup_{v \in \mathbb{Z}^d} T^v([0,1)^2) = X$. For any $f, g \in L^1(X, \mu)$ with $E(g|\mathcal{I}) > 0$ a.e., the ratio averages

$$\frac{\int_{B_n} f \circ T^{-v}(x) d\mu_x}{\int_{B_n} g \circ T^{-v}(x) d\mu_x}$$

converge μ-a.e. to $\frac{E(f|\mathcal{I})}{E(g|\mathcal{I})}$.

Chapter 6

Conclusion

Ergodic theorems are at the foundation of measurable dynamics. They begin the classification of dynamical systems that proceeds through entropy and orbit equivalence, and this has been explored in-depth in the case of a measure-preserving transformation. However, such theory has not been constructed for dynamical systems that are not measure-preserving. Can a parallel theory be built for this case? This question is yet to be addressed, but the establishment of the ratio ergodic theorem gives a starting point for such theory.

In this dissertation we used several tools to prove an ergodic theorem. First, we used an extension of the Besicovitch Covering Lemma due to Hochman to get a statement about the frequency of boundary-heavy balls in $\mathbb{Z}^d$ or $\mathbb{R}^d$. This line of reasoning uses the Besicovitch Covering Lemma, Doubling Condition, and the notion that the boundary of a ball is of lower dimension than the ball itself. Second, we diffused the measure of a probability space onto the orbits using a conditional expectation. This was necessary to even state what the ratio theorem should look like for possibly singular transformations. The statement from Hochman was then used to show a Følner condition on these diffused measures. Finally, we saw that the Følner condition and the Besicovitch Covering Lemma could be used to prove the ratio ergodic theorem. This result improves previous work by dropping the assumption of nonsingularity, but it also is a new method of proving ergodic theorems that bypasses the usual maximal inequality and dense family.

Finally, we describe a few questions that arise from Theorem 5.0.3. The theorem shows that convergence of the ratio average is intrinsic to the machinery of Borel actions of $\mathbb{Z}^d$ and $\mathbb{R}^d$ on standard Borel probability spaces and does not need a connection between the measure and the action. What can one say about the rate of convergence? This is probably a difficult question to address due to the few assumptions made. Also, does such a general ratio ergodic theorem hold on actions of groups other than $\mathbb{Z}^d$ and $\mathbb{R}^d$? Can convergence be shown on a more general class of averaging sets? One may be able to follow the same line of reasoning by using a Følner Condition and the Besicovitch Covering Lemma to positively answer either of these questions.

Bibliography

[1] M. Becker. A ratio ergodic theorem for groups of measure-preserving transformations. *Ill. J. Math.*, 27:562–570, 1983.

[2] A. S. Besicovitch. A general form of the covering principle and relative differentiation of additive functions. *Proc. Cambridge Philos. Soc.*, 41:103–110, 1945.

[3] P. Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, 1979.

[4] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. USA*, 17:656–60, 1931.

[5] M. Brin and G. Stuck. *Introduction to dynamical systems.* Cambridge University Press, 2002.

[6] J. Feldman. A ratio ergodic theorem for commuting, convservative, invertible transformations with quasi-invariant measure summed over symmetric hypercubes. *Ergodic Theory Dynam. Systems*, 27(4):1135–1142, 2007.

[7] Z. Fredi and P. A. Loeb. On the best constant for the Besicovitch Covering Theorem. *Proc. of the Amer. Math. Soc.*, 121(4):1063–1073, 1994.

[8] M. Hochman. A ratio ergodic theorem for non-singular multiparameter actions. *To appear in the Journal of the European Mathematical Society.*

[9] E. Hopf. *Ergodentheorie.* J. Springer, Berlin, 1937.

[10] W. Hurewicz. Ergodic theorem without invariant measure. *Ann. of Math.*, 45(2):192–206, 1993.

[11] M. Keane. The essence of the law of large numbers. *Algorithms, Fractals, and Dynamics*, pages 125–129, 1995.

[12] J. King and B. Weiss. A group rotation factor of a nonrigid rank-1 map. *Dynamical Systems*, 1342:425–439, 1988.

[13] U. Krengel. *Ergodic theorems*, volume 6 of *de Gruyter Studies in Mathematics.* Walter de Gruyter & Co., Berlin, 1985. With a supplement by Antoine Brunel.

[14] E. Lindenstrauss. Invariant measures and arithmetic quantum unique ergodicity. *Ann. of Math. (2)*, 163(1):165–219, 2006.

[15] J. Oxtoby. On the ergodic theorem of Hurewicz. *Annals of Mathematics*, 49(4):872–884, 1948.

[16] K. Petersen. *Ergodic theory.* Cambridge University Press, 1983.

[17] I. K. Rana. *An introduction to measure and integration*, volume 45 of *Graduate Studies in Mathematics.* Narose Publishing House, second edition, 2002.

[18] V. A. Rohlin. Selected topics in the metric theory of dynamical systems. *American Mathematical Society Translations*, 49(Series 2):171–240, 1966.

[19] D. Rudolph. Ergodic behaviour of Sullivan's geometric measure on a geometrically finite hyperbolic manifold. *Ergodic Theory Dynam. Systems*, 2:491–512, 1982.

[20] D. Rudolph. *Fundamentals of measurable dynamics.* Clarendon Press, 1990.

[21] D. Rudolph. Ergodic theory on Borel foliations by $\mathbb{R}^n$ and $\mathbb{Z}^n$. *Contemporary Mathematics*, 444:89–113, 2007.

[22] P. Shields. *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics.* American Mathematical Society, 1996.

[23] D. Sullivan. The density at infinity of a discrete group of hyperbolic motions. *Inst. Hautes Études Sci. Publ. Math.*, 50:171–202, 1979.

[24] M. Taylor. *Measure theory and integration*, volume 76 of *Graduate Studies in Mathematics.* American Mathematical Society, 2006.

[25] J. Von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Nat. Acad. Sci. USA*, 18:70–82, 1932.