INTRODUCTION TO ERGODIC THEORY WITH APPLICATIONS TO PHYSICS

KAMERON MEHLING

ABSTRACT. This paper explores the basics of Ergodic Theory, motivated largely by the rich and paramount history physics has had with this field of dynamical systems. This pursuit begins with an introduction to measure theory, enabling proof of both the Poincaré recurrence theorem and Birkhoff ergodic theorem. Following these results we undergo a more advanced approach grounded upon smooth manifolds to achieve the culminating result of this paper, a proof of Liouville's theorem.

CONTENTS

1. Introduction 1
2. Introduction Measure Theory 1
3. Poincaré Recurrence Theorem 3
4. Birkhoff Ergodic Theorem 5
5. Introduction to Riesz Representation Theorem 9
6. Introduction to Manifolds 12
7. Liouville's Theorem 14
Acknowledgments 19
References 19

1. INTRODUCTION

Ergodic Theory is a branch of dynamical systems developed to solve problems in statistical mechanics. Initially rooted in one field of physics, ergodic theory has since blossomed into a vital component of various fields of mathematics and physics today. The focus of ergodic theory is measure-preserving transformations, a concept that will be explained following some preliminary results of measure theory and dynamics.

2. INTRODUCTION MEASURE THEORY

Abstractly, a dynamical system is a system that models the evolution of an object's state within a geometrical space. The vagueness of this definition is indicative of the wide variety of mathematical fields and problems that dynamical systems may represent, but for our purposes one should think of the 'object'as a particle and its 'state' as the particle's position. The particle's evolution in time is generally represented by a transformation T, however we have yet to construct a space on which T operates.

A *measure space* is a tuple (X, Σ, μ) where:

- X is any non-empty set
- Σ is a σ-algebra on the set X
- μ is a measure on the space (X, Σ)

All transformations T will be defined on a measure space (X, Σ, μ), but it remains to define σ-algebra and measure μ.

Definition 2.1. A σ-algebra on a set X is any collection of subsets such that:

- $X \in \Sigma$
- $\forall\, S \in \Sigma,\ X \backslash S \in \Sigma$
- Σ is closed under *countable* unions, i.e. $S_1, S_2, ...S_n \in \Sigma$ for n ≥ 1 and $S = S_1 \cup S_2 \cup ... \cup S_n$ then $S \in \Sigma$.

It should be noted that the pair (X, Σ) is called a *measurable space* and can be any collection of sets that satisfy the $\sigma-$algebra conditions.

Example 2.2. *Let*$X = \{x, y, z\}$. We may construct $\Sigma = \big\{\emptyset, \{x, y\}, \{z\}, \{x, y, z\}\big\}$ and see (X, Σ) is a measurable space. Furthermore there may exist more than one σ-algebra over a set since $\psi = \big\{\emptyset, \{x, z\}, \{y\}, \{x, y, z\}\big\}$ also creates a measurable space (X, ψ)

Remark 2.3. All measurable spaces must have $\emptyset \in \Sigma$, and for finite or countably infinite sets X then the powerset $\mathscr{P}(X)$ will form a measurable space $(X, \mathscr{P}(X))$.

A measurable space (X, Σ) coupled with measure μ produces a measure space (X, Σ, μ). The measure μ provides a way to compare sets within Σ, however all measures must obey the following restrictions:

Definition 2.4. Given a nonempty set X with a σ-algebra on X, we define $\mu : \Sigma \to \mathbb{R}$ to be a measure if:

- $\forall S \in \Sigma, \mu(S) \geq 0,$
- $\mu(\emptyset) = 0,$
- For all countable collections $\{S_i\}_{i=1}^{\infty}$, of pairwise disjoint sets in Σ : $\mu(\bigcup\limits_{i=1}^{\infty} S_i) = \sum\limits_{i=1}^{\infty} S_i.$

Example 2.5. The most famous measure is the Lebesgue measure which conveys our normal understanding of length and volume. In 1-dimension, if $b > a$, $\mu([a, b]) = b - a$. Generalizing this to higher dimensions, given a box in $\mathbb{R}^n$ bounded by $\mathscr{B} = \big([a_1, b_1], [a_2, b_2], ...[a_n, b_n]\big)$ then $\mu(\mathscr{B}) = \prod\limits_{i=1}^{n} (b_i - a_i)$.

Example 2.6. Another interesting measure is the Dirac measure $\delta_a, a \in X$, where for any set $S \in \Sigma$ we denote $\delta_a(S) = 1$ if $a \in S$ and $\delta_a(S) = 0$ for $a \notin S$.

These two examples merely scratch the surface of the multitude of measures that can be endowed to create measure spaces. One final restriction remains in order to setup the parameters for the spaces we will study, that the geometrical space our transformation T operates over is a special kind of measure space known as a *probability space.*

Definition 2.7. A probability space is a measure space (X, Σ, μ) that also satisfies the condition that $\mu(X) = 1$. Additionally, this means all probability spaces are also finite-measure spaces, a subset of measure spaces in which $\mu(X) < \infty$.

Remark 2.8. Utilizing the above examples of measures, one should be convinced that a measure space utilizing the Dirac measure is always a probability space. As for a measure space with the Lebesgue-measure, it would only be a probability space if $a_i = r, b_i = (r+1)\ \forall r \in \mathbb{R}^n$ and for $1 \leq i \leq n$.

By now, the reader has learned all requisite tools and concepts to begin our study of ergodic theory. All the previous definitions and examples have been applicable to generic dynamical systems and the foundations of measure theory, however ergodic theory is concerned with a special set of evolutions modeled by transformations T called *measure-preserving.*

3. Poincaré Recurrence Theorem

Definition 3.1. A *measure-preserving dynamical system* is a probability space (X, Σ, μ) coupled with a transformation $T : X \to X$ such that

$$\forall A \in \Sigma,\ \mu(A) = \mu(T(A)) \tag{3.2}$$

Further elaborating upon this transformation T, since our domain and range are both X, we know for all $x \in X$ one can take numerous iterations such that we have a total trajectory $T^{-n}x, ...T^{-1}x, x, T^1x, ...T^nx \in X$ and it can be thought of each T as evolving the system by a unit of time.

Remark 3.3. Note that due to the notion of the transformation T as advancing 'time' we have $T^s \circ T^k = T^{s+k}$,

Definition 3.4. For T over (X, Σ, μ) and $E \in \Sigma$, any point $x \in E$ such that $T^nx \in E$ for some $n > 0$ is called a *recurrence point* in the set E.

Theorem 3.5 (Poincaré Recurrence Theorem)**.** *For any measure-preserving T and set $E \in \Sigma$, almost all $x \in E$ are recurrence points.*

The Poincaré Recurrence Theorem (PRT) is practical for physicists since it concludes that given an initial system state, eventually the system will evolve in time to produce the same configuration. Historically, there have been attempts to apply this result to particle systems which are central to statistical physics. By applying the PRT to a set of particles, one seemingly gains information about trajectories and eventual configurations, however this also would appear to contradict the 2nd Law of Thermodynamics (explained below following the proof).

Proof. Construct the set $N = \{x \in E : T^n x \notin E \ \forall \ n > 0\}$. Furthermore, we know we can rewrite this as

$$N = E \cap \left(\bigcap_{n=1}^{\infty} T^{-n}(X \backslash E) \right).$$

Using our first construction of N we know if $x \in N$, then

$$Tx, T^2x, ..., T^nx \notin E \tag{3.6}$$

and since

$$N \subseteq E, \quad T^n x \notin N \ \forall \ n. \tag{3.7}$$

This result implies $x \notin T^{-n}N$ so $N \cap T^{-n}N = \emptyset$. We therefore now see that each set $T^{-1}N, T^{-2}N, ...T^{-n}N$ is pairwise disjoint. This being said, by properties of dynamical transformations (Remark 1.3) for $n_1, n_2 \in \mathbb{N}, n_1 < n_2$ we see

$$T^{-n_1}N \cap T^{-n_2}N = T^{-n_1}\left(N \cap T^{-(n_2-n_1)}N\right) = \emptyset.$$

We therefore now see each set $N, T^{-1}N, T^{-2}N, ...T^{-n}N$ are disjoint allowing (Defintion 1.2)

$$1 \geq \mu\left(\bigcup_{n=0}^{\infty} N\right) = \sum_{n=0}^{\infty} \mu(T^{-n}N) = \sum_{n=0}^{\infty} \mu(N)$$

The above inequality clearly can only hold if $\mu(N) = 0$. □

The 2nd Law of Thermodynamics states that entropy, interpreted as the level of disorganization within a system, always increases with time. This Law was originally introduced as H-theorem by Ludwig Boltzmann, and it should be clear how these two theorems, Poincaré and H, are contradictory. Given an initial configuration of particles, PRT states after a long period of time they will return to the exact same positions. This would contradict the conclusion that entropy always increases since the entropy of the system would be equal at two separate times; when the particle configuration returned to that of a previous time.

As to which theorem prevails, this is up to debate for physicists to this day. Mathematician Ernst Zermelo was a staunch critic of H-theorem and some physicists follow his acceptance of Poincaré recurrence, while others build theoretical constructs that disregard recurrence and are grounded upon thermodynamics. Some even believe there is no contradiction, as the 2nd Law only holds as the limit of particles in a system goes to infinity which implies the time for a recurrence of particle configuration would also be infinite. Finally, others even still think that physical conditions such as gravity have an impact that makes both theorems applicable but blurs their compatibility.

4. Birkhoff Ergodic Theorem

Beyond knowing that a system will return to a prior state, ergodic theory can also enlighten physicists as to the average state a system will obtain. This is of monumental importance as the following theorem will prove a result that had been assumed by physicists for decades before its proof. First we need to define some terminology regarding functions.

Definition 4.1. An L^p space is a space of functions over a measure space (X, Σ, μ) such that if $f \in L^p, ||f||_p < \infty$, where we define

$$||f||_p = \left(\int\limits_X |f|^p d\mu \right)^{\frac{1}{p}} . \tag{4.2}$$

For our purposes, we shall be concerned with $f \in L^p(X, \Sigma, \mu)$. Before tackling the Birkhoff ergodic theorem, we will prove an imperative lemma often referred to as the *Maximal Ergodic Theorem.*

Lemma 4.3. *If T is measure-preserving over a measure space (X, Σ, μ), and given $f \in L^p(X, \mu)$, then for the set $H = \{x \in X : \sup\limits_{n \geq 0} \sum\limits_{k=0}^{n} f(T^k x) \geq 0\}$ we have $\int\limits_H f d\mu \geq 0$.*

Proof. Given $x \in X$, we introduce some notation.

$$s_n(x, f) = \sum_{k=0}^{n-1} f(T^k x), \quad s_0(x) = 0. \tag{4.4}$$

Using this notation, we obtain the relation

$$s_l(Tx) = s_{l+1}(x) - f(x) \tag{4.5}$$

For $k \geq 0$ this is a important result, which will be explicitly shown below, but first some facts.

$$T : X \to X, \quad \therefore Tx \in X$$
$$\bar{x} := Tx$$

with this notation we can rewrite our expansions as follows:

$$s_l(Tx) = s_l(\bar{x}) = 0 + f(\bar{x}) + f(T\bar{x}) + ... + f(T^{l-1}\bar{x})$$

$$s_{l+1}(x) = 0 + f(x) + f(Tx) + ... + f(T^l x) = f(x) + f(\bar{x}) + f(T\bar{x}) + ... + f(T^{l-1}\bar{x}).$$

All subsequent terms agree with the first expansion, hence comparing them we recover (4.5). Continuing we introduce

$$\Phi_n(x) = \max(0, s_1(x), ..., s_n(x)),$$
$$\Phi_n^*(x) = \max(s_1(x), ..., s_n(x)).$$

One may convince themselves that by extension, for $n \geq 0$

$$\Phi_{n+1}^*(x) - f(x) = \Phi_n(Tx). \tag{4.6}$$

Rearranging our formula we obtain

$$f(x) = \Phi_{n+1}^*(x) - \Phi_n(Tx) \geq \Phi_n^*(x) - \Phi_n(Tx), \quad \text{and}$$

$$\int_H f(x)d\mu \geq \int_H \Phi_n^*(x)d\mu - \int_H \Phi_n(Tx)d\mu.$$

But for any $x \in H$ we know $\Phi_n^*(x) = \Phi_n(x)$ by construction and for $x \notin H$ we have $\Phi_n(x) = 0$. This culminates to prove

$$\int_H \Phi_n^*(x)d\mu = \int_H \Phi_n(x)d\mu = \int_X \Phi_n(x)d\mu,$$

and since Φ_n is non-negative over X we have

$$\int_H \Phi_n(Tx)d\mu \leq \int_X \Phi_n(Tx)d\mu.$$

This enables us to obtain the desired result that

$$\int_H f(x)d\mu \geq \int_X \Phi_n(x)d\mu - \int_X \Phi_n(Tx)d\mu = 0.$$

□

With the lemma thus proved, we begin the larger challenge posed by the Birkhoff Ergodic Theorem.

Theorem 4.7 (Birkhoff ergodic theorem). *Let (X, Σ, μ) be a finite measure space equipped with a measure-preserving transformation T, and take $f \in L^1(X, \Sigma, \mu)$. Then for almost all $x \in X$ the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) := \bar{f}(x)$$

exists. Furthermore, $\overline{f} \in L^1(X, \Sigma, \mu)$ and

$$\int_X \bar{f}(x)d\mu = \int_X f(x)d\mu$$

The following proof, along with a similar result proven by Von Neumann around the same time, gave confirmation to physicists throughout many fields who had been treating the time and space average as identical for years, known today as the Ergodic Hypothesis. This theorem gave physics a verifiable way to justify this conclusion, however it wasn't long before exceptions were found. The ergodic proofs were published in 1931-1932, but by 1953 researchers had discovered what is known today as the Fermi–Pasta–Ulam-Tsingou problem. This was a paradox in which highly complicated physical systems from chaos theory displayed behavior that wasn't ergodic in nature, but periodic. This has lead to realizations that physically, there may be restrictions when applying the ergodic theorems, however, their setup and results have been vital to the development of physics since.

Proof. First, our aim is to show that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \lim_{n\to\infty} \frac{1}{n} s_n(x)$ (Recall from 4.4) exists. It suffices to show that for any $a, b \in \mathbb{Q}$, the following set

$$E = \left\{ x \in X : \lim_{n\to\infty^-} \frac{1}{n} s_n(x) < a < b < \lim_{n\to\infty^+} \frac{1}{n} s_n(x) \right\} \tag{4.8}$$

has $\mu(E) = 0$. Consider the function

$$g(x) = \begin{cases} f(x) - b & \text{for } x \in E, \\ 0 & \text{for } x \notin E. \end{cases} \tag{4.9}$$

Applying Lemma 4.3 to the set $H(g) = \left\{ x \in X : \sup_{n\geq 0} s_n(x, g) > 0 \right\}$ we obtain

$$\int_{H(g)} g(x) d\mu \geq 0. \tag{4.10}$$

For convenience we will introduce another set

$$\bar{H}(g) = \left\{ x \in X : \sup_{n\geq 1} \frac{1}{n} s_n(x, g) > 0 \right\} \tag{4.11}$$

and using the fact $s_0(x, g) = 0$ and $n \geq 1$ it is simple to show that $H(g) = \bar{H}(g)$. This result allows us to see

$$H(g) = \left\{ x \in X : \sup_{n\geq 1} \frac{1}{n} s_n(x, g) > 0 \right\} = \left\{ x \in X : \sup_{n\geq 1} \frac{1}{n} s_n(x, f) > b \right\}. \tag{4.12}$$

By construction we know

$$b < \lim_{n\to\infty^+} \frac{1}{n} s_n(x), \tag{4.13}$$

hence $E \subseteq H(g)$. However we know $g(x) = 0 \; \forall x \notin E$, so this implies $s_n(x, g) = 0, x \notin H(g)$. Therefore $H(g) \subseteq E \implies H(g) = E$. Therefore we have

$$\int_E g(x) d\mu \geq 0 \quad \text{or} \quad \int_E f(x) d\mu \geq b\mu(E), \tag{4.14}$$

We look to obtain a similar result by introducing

$$g'(x) = \begin{cases} a - f(x) & \text{for } x \in E, \\ 0 & \text{for } x \notin E. \end{cases} \tag{4.15}$$

and using same procedure as above one would obtain

$$\int\limits_E f(x) d\mu \leq a\mu(E). \tag{4.16}$$

Combining the two prior inequalities we have

$$\int\limits_E f(x) d\mu \leq a\mu(E) \leq b\mu(E) \leq \int\limits_E f(x) d\mu. \tag{4.17}$$

This inequality can only exist if $\mu(E) = 0$, so we have shown for almost all $x \in X$, hence $\lim\limits_{n\to\infty^+} \frac{1}{n} s_n(x) = \bar{f}(x)$ exists. We continue

$$\int\limits_X \bar{f}(x) = \int\limits_X \frac{1}{n} s_n(x) d\mu \leq \frac{1}{n} \int\limits_X \sum_{k=0}^{n-1} |f(T^k x)| d\mu \leq \frac{1}{n} \int\limits_X n|f(x)| d\mu = \int\limits_X |f(x)| < \infty \tag{4.18}$$

which therefore proves the first part of our theorem, that $\bar{f}(x) \in L^1(X, \Sigma, \mu)$.

Remark 4.19. The above inequality can be obtained due to the measure-preserving property of T and the fact $f \in L^1(X, \Sigma, \mu)$.

we continue by introducing for $l \in \mathbb{Z}$ and $n \in \mathbb{N}$

$$C = \{x \in X : \frac{l}{n} < \bar{f}(x) < \frac{l+1}{n}\}. \tag{4.20}$$

We will still be utilizing Lemma 4.3, by working with the function

$$f^* := \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$$

But by our previous construction we know f^* converges to $\bar{f}$ meaning that for almost all $x \in X$ and any $\epsilon > 0$ then

$$|\bar{f}(x) - f^* x| < \epsilon,$$

meaning

$$C = \{x \in X : \frac{l}{n} - \epsilon \leq f^*(x) < \frac{l+1}{n} + \epsilon\} \tag{4.21}$$

so one may recognize that with a little algebra the first inequality in the set C greatly resembles H (4.3) and one reaches the conclusion that

$$(\frac{l}{n} - \epsilon)\mu(C) \leq \int\limits_C f d\mu, \tag{4.22}$$

and since the ϵ is arbitrary we may revert back

$$\frac{l}{n}\mu(C) \leq \int\limits_C f d\mu. \tag{4.23}$$

A more immediate result also follows that

$$\int_C \bar{f} d\mu \leq \frac{l+1}{n}\mu(C). \tag{4.24}$$

Hence combining our above two inequalities we obtain

$$\left| \int_C \bar{f} d\mu - \int_C f d\mu \right| \leq (\frac{1}{n})\mu(C). \tag{4.25}$$

Adding our final fact that by taking all $l \in \mathbb{Z}$ we sum over all possible values so

$$X = \bigcup_{l \in \mathbb{Z}} C, \tag{4.26}$$

and since this is a collection of disjoint unions we can state

$$\left| \int_X \bar{f} d\mu - \int_X f d\mu \right| \leq \sum_{l=-\infty}^{l=\infty} \left| \int_{C_{\frac{l}{n}, \frac{l+1}{n}}} \bar{f} d\mu - \int_{C_{\frac{l}{n}, \frac{l+1}{n}}} f d\mu \right| \leq \sum_{p=-\infty}^{p=\infty} \frac{1}{n}\mu(C) = \frac{1}{n}. \tag{4.27}$$

Taking $n \to \infty$ we obtain $\int_X \bar{f} d\mu = \int_X f d\mu$.

□

The Birkhoff ergodic theorem has simple applications to physical situations. Imagine a billiard on a pool table, where once the ball is struck it continues forever by bouncing along the walls. It can be shown that if the pool table is rectangularly shaped, then the ball will follow a periodic trajectory and hence be a non-ergodic system. However, if the table is polygon-shaped then in all most all cases the system is now ergodic, except when the angle of any sides are rational multiples of π. There are many similar reformulations of these problems, with proofs, so we encourage the reader to pursue further reading [1] if intrigued by these situations that provide a physical background for the important concept proved above.

5. Introduction to Riesz Representation Theorem

We shall further develop our background in measure theory by introducing the necessary Riesz Representation Theorem. This section is pivotal along with our ensuing discussion of manifolds to prove Liouville's theorem .

Definition 5.1. A *Borel set* is any set within a topological space that can be constructed by the operations of countable union, countable intersection, and relative complement applied to open sets.

Remark 5.2. The collection of all Borel sets over a topological space (such as a manifold M) creates a σ-algebra (Def 2.1) known as the *Borel algebra*.

Definition 5.3. A *Borel measure* is any measure defined over all open sets (hence operates over all Borel sets).

Definition 5.4. A Borel measure on a locally compact metric space is called *regular* if:

- $\mu(S) < \infty$ for all compact S,
- For every Borel set $E \subseteq M$, $\mu(E) = \inf\{\mu(G) : E \subseteq G, G \text{ open}\}$, and
- For every Borel set $E \subseteq M$, $\mu(E) = \sup\{\mu(L) : L \subseteq E, L \text{ compact}\}$ if $\mu(E) < \infty$ or instead E is open.

All the above formalities allow us introduce the Riesz representation theorem, which we shall present without proof due to its complexity.

Theorem 5.5. *Given a locally compact metric space, M, any positive linear functional $l \in C_c(M)$ (continuous complex functions over M) has one corresponding regular Borel measure on M such that*

$$l(f) = \int_M f(x) d\mu(x). \tag{5.6}$$

With the above theorem, we are enabled to prove a more pivotal result for our purposes in ergodic theory. Some more vocabulary remains to be defined.

Definition 5.7. A measurable function g is *invariant* with respect to the transformation T if

$$g(T^t(x)) = g(x) \quad \forall\, t \in \mathbb{R}^1. \tag{5.8}$$

Remark 5.9. We have used the terminology prior that a measure space (X, Σ, μ) is a probability space if $\mu(X) = 1$. Such a space is naturally endowed with a probability measure μ which we shall also refer to as a *normalized* measure.

Utilizing all the definitions throughout this section, we are capable of introducing and tackling an existence proof fundamental in the pursuit of Liouville's theorem.

Lemma 5.10. *Given a compact metric space M, then the space of all probability measures over M, denoted $P(M)$, is weakly compact.*

Proof. First we introduce another metric space $C(M, \mathbb{R}) = \{f : M \to \mathbb{R} | f \text{ is continuous}\}$. We will also define a metric on $C(M, \mathbb{R})$,

$$||f||_\infty = \sup_{x \in M} |f(x)|, \tag{5.11}$$

and further establish

$$\rho(f, g) = ||f - g||_\infty. \tag{5.12}$$

Remark 5.13. We shall proceed to use the following notation $\mu(f)$. This may be confusing as one hasn't defined how to use a measure on a function, but instead the above denotes $\mu(f) = \int_M f d\mu$.

One additional property of $C(M, \mathbb{R})$ is the fact it is *separable*, meaning that it contains countable dense subsets. Due to this property we choose a subset

$$\{f_i\}_{i=1}^{\infty} \subset C(X) \tag{5.14}$$

which is countable and dense. Given a sequence $\mu_n \in P(M)$ we see

$$|\mu_n(f_1)| \leq ||f_1||_\infty \quad \forall\ n \in \mathbb{N}. \tag{5.15}$$

Since $|\mu_n(f_1)| \in \mathbb{R}$ and is bounded we know there exists a convergent subsequence we shall denote $\mu_{n_1}(f_1)$. Use this subsequence as above on the function f_2, we obtain:

$$|\mu_{n_1}(f_2)| \leq ||f_2||_\infty. \tag{5.16}$$

We have another bounded sequence of real numbers, so it must have a convergent subsequence we shall denote $\mu_{n_2}(f_2)$.

Continuing in this way we have a collection of nested sequences such that

$$\{\mu_{n_i}\} \subset \{\mu_{n_{i-1}}\} \quad i \geq 1, \tag{5.17}$$

and each $\mu_{n_i}(f_j)$ converges for $1 \leq j \leq i$. If we examine the diagonal sequence we know that for $i \leq n$ that μ_{n_n} is a subsequence of μ_{n_i}, thus $\mu_{n_n}(f_i)$ converges for $1 \leq i \leq n$.

Since we began with the criterion that $\{f_i\}$ be dense, we know $\forall\ f \in C(M)$ that $\forall\ \epsilon > 0, \exists\ f_i \in C(M)$ s.t. $||f - f_i||_\infty < \epsilon$. Also by the convergence of $\mu_{n_n}(f_i)$, we have for $N \in \mathbb{N}$, $i \leq N \leq n, m$, and $\forall\ \epsilon > 0$,

$$|\mu_{n_n}(f_i) - \mu_{m_m}(f_i)| \leq \epsilon. \tag{5.18}$$

Thus by the triangle inequality applied to the above inequalities, we obtain:

$$|\mu_{n_n}(f) - \mu_{m_m}(f)| \leq |\mu_{n_n}(f) - \mu_{n_n}(f_i)| + |\mu_{n_n}(f_i) - \mu_{m_m}(f_i)| + |\mu_{m_m}(f_i) - \mu_{m_m}(f)| < 3\epsilon.$$

Hence we see $\mu_{n_n}(f)$ converges, and define the limit $w(f) = \lim_{n\to\infty} \mu_{n_n}(f) = \lim_{n\to\infty} \int f d\mu_{n_n}$. We deduce the following facts about w:

- $w(\alpha f + g) = \lim_{n\to\infty} (\alpha \int f d\mu_{n_n} + \int g d\mu_{n_n}) = \alpha w(f) + w(g)$,
- $|w(f)| \leq ||f||_\infty$,
- $\forall\ f \geq 0, w(f) = \lim_{n\to\infty} \int f d\mu_{n_n} \geq 0$.

Hence these facts show that $w(f)$ is a positive linear functional with $f \in C_c(M)$ over a compact metric space so we may apply Riesz representation theorem (5.5). Therefore $\exists\, \mu \in P(M)$ such that $w(f) = \int f d\mu = \lim_{n\to\infty} \int f d\mu_{n_n}$. Hence we have proved all sequences $\mu_n \in P(M)$ have weakly convergent subsequences, which is equivalent to $P(M)$ being weakly compact. □

We conclude the section with the following theorem on the existence of invariant measures.

Theorem 5.19. *For any continuous map T of the compact metric space M into itself, there exists a normalized Borel measure μ invariant with respect to T.*

Proof. Take an arbitrary normalized Borel measure μ^0 over M and consider the sequence

$$\mu_n(f) = \frac{1}{n}\sum_{k=0}^{n-1}\mu^0(f(T^k x)), \quad f \in C(M). \tag{5.20}$$

We know by Lemma 5.10 that for

$$\mu_{n_s}(f) \subset \mu_n(f), \quad \mu_{n_s} \to \mu \text{ and} \tag{5.21}$$

$$\int_M f(x)d\mu = \lim_{n_s\to\infty}\int_M f(x)d\mu_{n_s} \tag{5.22}$$

for any function $f \in C(M)$. Now examine

$$\left|\int_M f(Tx)d\mu_{n_s} - \int_M f(x)d\mu_{n_s}\right| = \tag{5.23}$$

$$\left|\frac{1}{n_s}\sum_{k=0}^{n_s-1}\int_M f(T^{k+1}x)d\mu^0 - \frac{1}{n_s}\sum_{k=0}^{n_s-1}\int_M f(T^k x)d\mu^0\right| = \tag{5.24}$$

$$\frac{1}{n_s}\left|\int_M f(T^{n_s}x)d\mu^0 - \int_M f(x)d\mu^0\right| \leq \tag{5.25}$$

$$\frac{2}{n_s}\max|f| \to 0 \tag{5.26}$$

when $n_s \to \infty$. But applying this with equation (5.22) we have

$$\lim_{n_s\to\infty}\left|\int_M f(Tx)d\mu_{n_s} - \int_M f(x)d\mu_{n_s}\right| = \left|\int_M f(Tx)d\mu - \int_M f(x)d\mu\right| = 0. \tag{5.27}$$

Hence we have shown there exists a Borel measure μ which is invariant with respect to T. □

6. Introduction to Manifolds

To further espouse the utility of ergodic theory for physics, we must introduce the reader to a basic background in the smooth manifolds.

Definition 6.1. An n-dimensional manifold is any topological space M that for each $x \in M$, there is a neighborhood that is *homeomorphic* to E, an n-dimensional Euclidean space.

Definition 6.2. A function $f : U \to V$ is homeomorphic over two topological spaces (U, M), (V, E) if:

- f is a bijection,
- f is continuous, and

- f has a continuous inverse, f^{-1}.

Definition 6.3. If given $U_\alpha \subset M$ such that we have a homeomorphism $f_\alpha : U_\alpha \to E$ with the property that $\bigcup_{\alpha=1}^{n} U_\alpha = M$ then the collection of *charts* (U_α, f_α) form the *atlas* denoted (U, f).

Remark 6.4. A *smooth* manifold is an object which satisfies definition 6.1 and possesses charts which are all infinitely differentiable.

Remark 6.5. A *compact* manifold has a a finite collections of (U_α, f_α) that can form an atlas.

Definition 6.6. A diffeomorphism is any differentiable map $G : M \Rightarrow N$ between two manifolds M, N such that:

- $G : M \to N$ is a bijection,
- $G^{-1} : N \to M$ is differentiable, and
- $G \in C^r$ if both function are r times differentiable.

Proposition 6.7. *For a fixed altas (U, f) over a n-dimensional manifold M, with T a homeomorphic map $T : M \to M$, if T is also a diffeomorphism of class C^r, for there to exists a smooth invariant measure μ, then it must be* $\det \left\| \frac{\partial \phi_i}{\partial x_j} \right\| = 1$ *where ϕ_i are the component functions of T, $1 \le i \le n$.*

Proof. For $x \in M$ we denote

$$y = T(x), \quad y_i = \phi_i(x_1, ..., x_m) \quad 1 \le i \le m. \tag{6.8}$$

By Theorem 5.19 we know there exists an invariant measure with respect to T, but since T is a diffeomorphism the measure must be compatible with this smooth structure. Therefore we introduce a *smooth invariant measure*

$$\mu(A) = \int_A p(x) dx_1 ... dx_m \tag{6.9}$$

for $A \subset U_i$ and $p(x)$ a measurable function. This measure is smooth, meaning for a given atlas the measure is defined by a density function that weights the value based on the chart being evaluated (here our $p(x)$), and by the above theorems is invariant so

$$\mu(A) = \mu(TA) \tag{6.10}$$

$$\int_A p(x) dx_1 ... dx_m = \int_{TA} p(y) dy_1 ... dy_m. \tag{6.11}$$

But using integration by substitution, we find

$$\int_{TA} p(y) dy_1 ... dy_m = \int_A p(Tx) \det \left\| \frac{\partial \phi_i}{\partial x_j} \right\| dx_1 ... dx_m, \tag{6.12}$$

hence

$$\int_A p(x) dx_1 ... dx_m = \int_A p(Tx) \det \left\| \frac{\partial \phi_i}{\partial x_j} \right\| dx_1 ... dx_m. \tag{6.13}$$

Since A was an arbitrary subset of U,

$$p(x) = p(Tx)\det\left\|\frac{\partial \phi_i}{\partial x_j}\right\|. \tag{6.14}$$

If we assume the function $p(x)$ is continuous then the above equality holds everywhere (including overlapping charts U_i, U_j). Furthermore, we know the invariance of our measure implies

$$p(x) = p(T^n x)\det\left\|\frac{\partial \phi_i}{\partial x_j}\right\| \quad \forall\ n \in \mathbb{N}. \tag{6.15}$$

Since M is compact, there exists a finite number of $N \in \mathbb{N}$ such that $\bigcup_{i=0}^{N} U_i = U$ so at least one U_k contains an infinite subsequence

$$p(T^{n_s} x)\det\left\|\frac{\partial \phi_i}{\partial x_j}\right\| \in U_k \quad 1 \leq k \leq N. \tag{6.16}$$

Now one could apply Poincaré Recurrence to prove some $T^{n_s}x = x$, showing in order for there to exist a smooth measure μ invariant to T, then

$$\det\left\|\frac{\partial \phi_i^{n_s}}{\partial x_j}\right\| = 1 \tag{6.17}$$

where $\phi_i^{n_s}$ are the component functions of T.

□

Example 6.18. If x_0 is a fixed point for the transformation T, then for the existence of a smooth invariant measure it is required that $\det||\frac{\partial \phi_i}{\partial x_j}|| = 1$. In the case that x_0 were an attractive fixed point, hence $\det||\frac{\partial f_i}{\partial x_j}|| < 1$, then a smooth invariant measure is impossible.

7. Liouville's Theorem

We now have all the suppository information and proof to tackle Liouville's Theorem. Given M an n-dimensional compact orientable manifold of class $C^r, r \geq 2$ we introduce the concept of a *tangent vector*. Given $p \in M$, we call v a tangent vector if there exists a C^r function

$$\gamma : \mathbb{R} \to M \text{ s.t}$$
$$\gamma(0) = p, \quad \dot{\gamma}(0) = v$$

Since there is no restriction γ be unique, each $p \in M$ has an accompanying *tangent space* defined as:

$$T_pM := \left\{\dot{\gamma}(0) \,|\, \gamma : \mathbb{R} \to M \text{ a } C^r \text{ function}, \gamma(0) = p\right\}. \tag{7.1}$$

But the above construction is only point-specific, so the collection of all tangent spaces creates a *tangent bundle*:

$$TM = \bigcup_{p \in M} T_pM \tag{7.2}$$

The tangent bundle on M gives rise to a *vector field* on M, a construction which for every $p \in M$ assigns a vector $v \in T_pM$. This can be thought of as a collection of one element from each tangent space or a section of the tangent bundle, but in any case it should be apparent due to the smoothness of the tangent spaces that the vector field is smooth as well. We denote this vector field

$$X : M \to \mathbb{R}^m, \quad X(p) = \big(X_1(p_1, ...p_m)...X_m(p_1, ...p_m)\big) \; \forall \; p \in M. \tag{7.3}$$

By characteristics of the smooth vector field structure we know their must be a differential operator over M, which using the relation

$$D_v f(x) = (\nabla f(x)) \cdot v \tag{7.4}$$

we state for all $f \in C^1(M)$

$$D_X f = \sum_{k=1}^{m} X_k \cdot \frac{\partial f}{\partial x_k}. \tag{7.5}$$

Furthermore, for our smooth manifold M endowed with a vector field X, for a given interval $I \subset \mathbb{R}$ we call $\gamma : I \to M$ an *Integral Curve* if:

$$\dot{\gamma}(t) = X(\gamma(t)) \quad \forall \; t \in I \tag{7.6}$$

Without deviating from our study of differential manifolds, we introduce a pivotal result in ordinary differential equations known as the Picard-Lindelöf Theorem.

Theorem 7.7 (Picard-Lindelöf theorem)**.** *Given X a continuous function which also obeys the Lipschitz condition, then for $I := [a, b] \subset \mathbb{R}$, $X : I \times \mathbb{R}^n \to \mathbb{R}^n$, and $\dot{\gamma}(t) = X(\gamma(t), t)$ then the ODE possesses a unique solution on $[a, a + \epsilon]$ for fixed choice of $\gamma(0) = p$.*

Remark 7.8. The Lipschitz condition requires $|X(x, t) - X(y, t)| \leq L|x - y|$ for any $x, y \in \mathbb{R}^n$ and for some $L \in \mathbb{R}$. Now all continuously differentiable functions are Lipschitz continuous, hence by smoothness of X the condition is satisfied.

It should be apparent that this theorem directly applies to our problem and more plainly means that for a given $t_0 \in \mathbb{R}$ the system of equations

$$\frac{dx_k}{dt} = X_k(x_1, ..., x_m) \tag{7.9}$$

has a unique solution in a small region around t_0. We now wish to extend our integral curves such that for all $p \in M$,

$$\gamma : \mathbb{R} \to M \quad \text{and} \tag{7.10}$$

$$\dot{\gamma}(t) = X(\gamma(t)), \quad \gamma(0) = p. \tag{7.11}$$

The above condition is a result of X being a *complete* vector field, meaning all flow curves exist for any time. Now we present without proof

Lemma 7.12. *For any compact manifold $M \subset \mathbb{R}^m$, then all vector fields X over M are complete.*

The proof is grounded in differential geometry and can be viewed in (*Introduction to Differential Geometry*, Robbin & Salamon).

This concludes all relevant background and concepts necessary to introduce and prove Liouville's Theorem.

Theorem 7.13. *(*Liouville Invariant Measure Theorem*) In order for a measure μ (finite or σ-finite) to be invariant with respect to $\{T^t\}$, it is necessary and sufficient to have the relation*

$$\sum_{k=1}^{m} \frac{\partial}{\partial x_k}(pX_k) = 0. \tag{7.14}$$

Proof. Suppose we have M, a compact m-dimensional closed and orientable manifold of class C^∞. Endow M with a smooth vector field X where a vector in the field is represented by

$$X(x_1, ..., x_m) = (X_1(x_1, ...x_m), ..., X_m(x_1, ...x_m)). \tag{7.15}$$

Since X is smooth we know each X_k is C^∞ for $1 \leq k \leq m$. Therefore if we consider the system of equations

$$\begin{cases} \frac{dx_1}{dt} = X_1(x_1, ...x_m) \\ \quad \vdots \\ \frac{dx_m}{dt} = X_m(x_1, ...x_m) \end{cases} \tag{7.16}$$

we can apply the above existence and uniqueness theorems to establish a one parameter group $\{T^t\}$ of diffeomorphisms satisfying the above system of equations.

Now we choose a differential m-form ω of class C^∞ on M. Without specifying the details, this enables us to integrate over our manifold M without identifying specific coordinates, thus preventing issues when integrating over charts of M with different local coordinates. This differential form creates a continuous linear functional over $C(M)$ by

$$\omega(f) = \int_M f(x)\omega(dx), \quad f \in C(M) \tag{7.17}$$

If we only consider positive functionals such that $\omega(f) > 0$ if $f > 0$ then we can define a measure $\mu_\omega(f) = \omega(f)$

. Now by definition, a manifold is orientable if the Jacobian determinant of all transition maps are positive

$$\det||\phi_i o \phi_j^{-1}|| > 0. \tag{7.18}$$

Since this allows us to choose an oriented atlas which remains positive along transitions and each individual chart, this means our form w can be replaced by a nonnegative density function of C^∞ given by $p(x)$. This enables us to review Theorem 4.20 and 5.17 to find an invariant measure.

Since we know $\{T^t\}, t \in \mathbb{R}$ provides an integral curve for a given t then we know for all $x \in M$ there exists $k \in \mathbb{R}$ where the integral curve has traversed the manifold and returned such that $x = T^k x$. Now if we introduce $\mu_{\omega_t}(f) = \omega_t(f) =$

$\omega(f(T^t x))$, then we know for $t = k$

$$\mu_\omega(f) = \int_M f(x)p(x)dx \text{ and} \tag{7.19}$$

$$\mu_{\omega_k}(f) = \mu_\omega(f(T^k)) = \int_M f(T^k x)p(x)dx. \tag{7.20}$$

For $\mu_\omega(f) = \mu_{\omega_k}(f)$ we would need μ_ω to be a T invariant measure, which can only be true if

$$\det||\frac{\partial \phi_i^k}{\partial x_j}|| = 1, \tag{7.21}$$

where ϕ_i^k are the component functions of T^k. However we know due to the periodicity that

$$x_n = \phi_n^k(x_1, ...x_n...x_m) \quad \forall\ 1 \leq n \leq m, \tag{7.22}$$

hence we have

$$\det||\frac{\partial \phi_i^k}{\partial x_j}|| = \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_1}{\partial x_2} & \cdots & \frac{\partial x_1}{\partial x_m} \\ \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_m}{\partial x_1} & \frac{\partial x_m}{\partial x_2} & \cdots & \frac{\partial x_m}{\partial x_m} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{7.23}$$

So obviously our condition is satisfied, so we know there exists a smooth invariant measure for a continuous density such that

$$\mu_\omega(f) = \int_M f(x)p(x)dx = \int_M f(T^k(x)p(x)dx = \mu_{\omega_k}(f). \tag{7.24}$$

This has been defined for a single $k \in \mathbb{R}$, however this proof requires the existence of an invariant measure for $\{T^t\}$. Let us denote $d\mu = p(x)dx_1, ...dx_m$ and we begin our final construction.

Consider the function $f \in C^\infty(M)$ which is concentrated in one neighborhood U_i i.e.:

$$B = \{x \in M : f(x) \neq 0\}, \quad \bar{B} \subseteq U_i \tag{7.25}$$

Remark 7.26. $\bar{B}$ denotes the closure of the set B.

Next we can choose a t_0 such that for all $|t| < t_0$ then $f(T^t(x))$ is also concentrated in U_i. This result is based on the continuity of the functions f, T. To prove the invariance then all that remains is to show

$$\int_M f(x)pdx_1, ...dx_m = \int_M f(T^t x)pdx_1, ...dx_m \tag{7.27}$$

for all f, t_0 as outlined above.

One notices that the right side of this equality is continuously differentiable while the left side is a constant C. Therefore we have

$$\int_M f(T^t x) p dx_1, ...dx_m|_{t=0} = \int_M f(x) p dx_1, ...dx_m = C. \tag{7.28}$$

But differentiating both sides we find

$$\frac{d}{dt}\int_M f(T^t x) p dx_1, ...dx_m|_{t=0} = 0. \tag{7.29}$$

Since $f(x) = 0$ for $x \notin U_i$ we change

$$0 = \frac{d}{dt}\int_M f(T^t x) p dx_1, ...dx_m|_{t=0} = \frac{d}{dt}\int_{U_i} f(T^t x) p dx_1, ...dx_m|_{t=0}. \tag{7.30}$$

Now applying this total integral to a function $f(t, x_1, ...x_m)$ one obtains

$$\frac{d}{dt}\int_{U_i} f(T^t x) d\mu|_{t=0} = \frac{\partial}{\partial t}\int_{U_i} f(T^t x) d\mu|_{t=0} \quad + \int_{U_i} \sum_k^m \frac{\partial f(T^t x)}{\partial x_k}\frac{dx_k}{dt} d\mu|_{t=0}. \tag{7.31}$$

The first term on the right side is equivalently zero in either order you evaluate the operations, and the second term becomes

$$\frac{d}{dt}\int_{U_i} f(T^t x) d\mu|_{t=0} = \int_{U_i} \sum_k^m X_k(x)\frac{\partial f(x)}{\partial x_k} d\mu. \tag{7.32}$$

Now using integration by parts we obtain

$$\int_{U_i} \sum_k^m X_k(x)\frac{\partial f(x)}{\partial x_k} d\mu = \sum_k^m X_k(x) f(x)|_{U_i} \quad - \int_{U_i} \sum_k^m \frac{\partial X_k}{\partial x_k} f(x) d\mu. \tag{7.33}$$

Since $\bar{B} \subset U_i$ there exists $x_0 \in U_i$ such that $f(x_0) = 0$ making the first term disappear. Therefore

$$0 = \int_{U_i} \sum_k^m \frac{\partial X_k}{\partial x_k} f(x) d\mu = \int_{U_i} \sum_k^m \frac{\partial (pX_k)}{\partial x_k} f(x) dx_1, ...dx_m, \tag{7.34}$$

and expanding this we obtain for any f satisfying the conditions above

$$\int_{U_i} \sum_k^m \frac{\partial (pX_k)}{\partial x_k} f(x) dx_1, ...dx_m = \int_M \sum_k^m \frac{\partial (pX_k)}{\partial x_k} f(x) dx_1, ...dx_m = 0, \tag{7.35}$$

which can only be true if

$$\sum_{k=1}^m \frac{\partial}{\partial x_k}(pX_k) = 0. \tag{7.36}$$

□

This result seems inconsequential, however, it enables the application of the Hamiltonian formalism, a significant simplification in the field of mechanics. Although the result is not immediate from the prior proof, it is a corollary. The impact this theorem has on physics cannot be understated, as it has enabled a Hamiltonian mechanics, a simplification that has made many problems solveable that would previously have been incredibly tedious or impossible. For a more robust explanation of the impact Liouville's theorem has on physics one should read *Mathematical Methods of Classical Mechanics*, Arnold.

Acknowledgments. It is a pleasure to thank my mentor, Owen Barrett. He has played an instrumental role in providing guidance and assisting in my personal development as a mathematician. This paper has been made possible by his support, of which I am incredibly grateful. I would also like to thank Peter May, for running a fantastic Uchicago Math REU which enables young mathematicians to push bounds and pursue dreams.

References

[1] Sun Woo Park. An Introduction to Dynamical Billiards. https://math.uchicago.edu/ may/REU2014/REUPapers/Park.pdf, 2014

[2] Dietmar Salamon & Joel Robbins. Introduction to Differential Geometry. Prepress, 2018

[3] I.P. Cornfield, S.V. Fomin, & Ya. G. Sinai. Ergodic Theory. A Series of Comprehensive Studies in Mathematics, 1982

[4] Charles Walkden. Ergodic Theory. Lecture Series, 2018

[5] Bercovici, Hari, Brown, Arlen, Pearcy, Carl. Measure and Integration. Spring, 2016

[6] P.D. Gujrati. Poincare Recurrence, Zermelos Second Law Paradox, and Probabilistic Origin in Statistical Mechanics. https://arxiv.org/pdf/0803.0983.pdf, 2018

[7] Ruifeng Dong & Dejan Stojkovic. Gravity can significantly modify classical and quantum Poincare recurrence theorems. https://arxiv.org/pdf/1611.00792.pdf, 2017

[8] Wassim M. Haddad, VijaySekhar Chellaboina, and Sergey G. Nersesov. Time-Reversal Symmetry, Poincare Recurrence, Irreversibility, and the Entropic Arrow of Time: From Mechanics to System Thermodynamics. http://folk.ntnu.no/skoge/prost/proceedings/cdc-ecc05/pdffiles/papers/1251.pdf, 2005

[9] V.S. Sunder. The Riesz representation theorem. https://www.imsc.res.in/ sunder/rrt1.pdf, 2008

[10] Calvin Moore. Ergodic theorem, ergodic theory, and statistical mechanics. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4343160/, 2015