Introduction

When I am dead, I hope it may be said
'His sins were scarlet, but his books were read'
(Hilaire Belloc)

The basic aim of this book is to present a simple and accessible account of some of the most basic ideas in the theory of non-uniformly hyperbolic diffeomorphisms, or more colloquially, 'Pesin theory'.

Part I consists of four chapters which contain basic material on the Oseledec theorem, the Ruelle-Pesin inequality, and the Pesin set. There is then a brief 'interlude' to mention some topical examples and to draw some motivation from the uniformly hyperbolic (or 'Axiom A') case. Then, Part II contains contains three chapters dealing with applications of this theory to periodic points, homoclinic points, and stable manifold theory.

In the course of the text I tried to bring out the following two themes

(i) *Generality*. Ultimately we want to arrive at a theory applicable to any smooth diffeomorphism of a compact surface (providing it has non-zero topological entropy);

(ii) *The rôle of measure theory*. In applying the theory it is remarkable how often invariant measures play a crucial role in situations where the hypothesis and conclusion are purely topological. In some sense, the Poincaré recurrence of invariant measures seems to compensate for the absence of the compactness often take for granted in uniformly hyperbolic systems.

This text is based on a short series of lectures I gave in the *Centro de Matematica do INIC na Universidade do Porto* between March and June 1989. These lectures were intended to give a basic introduction to some of the simpler and more accessible aspects of the theory (both for the benefit of the audience and myself). My choice of presentation was chiefly influenced by the more topologically oriented approaches in the work of Anatole Katok and Sheldon Newhouse. Because of the nature of the audience I was able to assume a solid background in undergraduate analysis, but not necessarily any specialist knowledge of either measure theory or ergodic theory.

In order to keep the exposition as breezy as possible I have unashamedly resorted to using two devices: (a) postponing the more tedious proofs to the ends of the relevant chapters (or indefinitely); and (b) restricting proofs to the case of surfaces if it provides a significant saving in effort.

Existing literature. There exist a number of very good accounts of certain parts of the theory, although these tend to be somewhat scattered in the literature, and for some topics the original articles still provided the best sources. At present the only textbook account of Pesin theory known to me is contained in the last chapter of *Teoria Ergodica*, by Ricardo Mañé [Mañé$_2$] (This book now has an English translation, but my references are to the original Portuguese language edition). Along with the preliminary sections of research articles by Katok [Katok] and Newhouse [Newhouse$_{1,2}$] this probably gives the clearest introduction to the foundations of the subject.

In addition, there exists a (currently) unfinished monograph by A. Katok and L. Mendoza which is already becoming a standard reference and promises to be a very useful introduction to the subject [Kat-Men]. I should also mention that there is a very comprehensive book by Katok and Strelycn [Kat-Str], but this goes far beyond the scope of an expository account.

These sources are nicely complemented by a clear account of the basic theory of stable manifolds contained in the survey article of Fathi, Herman and Yoccoz [Fa-He-Yo] (although in these notes I follow the last part of Mañé's article [Mañé$_1$] in order to preserve a slightly more topological flavor). In the last chapters we lightly touch upon the important, but harder, topics of absolute continuity, continuity of entropy and the C^∞ entropy conjecture. For these matters there now exist good accounts by Pugh-Shub [Pug-Shu] and Gromov [Gromov].

Finally, an interesting over-view of the theory (without any proofs) can also be found in a survey of Pesin and Sinai [Pes-Sin].

For the background material on ergodic theory there are a number of excellent modern references, of which my favorites are *An Introduction to Ergodic Theory* by Peter Walters [Walters] and *Topics in Ergodic Theory* by William Parry [Parry].

Acknowledgements. I would like to thank all the members of the audience of the original course, from which these notes are taken, for their very active participation. I am also very grateful to the anonymous referees for their very useful comments, to Sr[a] D. Zulmira Couto for typing the first draft from very rough notes, and to Luisa Magalhães and Viviane Baladi for their brave attempts to help me to reduce the number of errors and inconsistencies in the text. I would appreciate being notified of any further errors that readers may notice.

I am also grateful to David Ruelle for some encouraging comments on a preliminary version of this book, and to the authors of [Kat-Men] for making their notes available prior to publication.

Finally, at all stages of this work I was an investigador of the Instituto Nacional de Investigação Cientifica (INIC), an institution for which I have great respect and affection, and to which I would like to dedicate this book.

Chapter 1

Invariant measures and some ergodic theory

In this first chapter we shall describe some basic ideas in the ergodic theory of general measurable spaces. In later sections we shall specialize to diffeomorphisms of manifolds.

For the initiated we have included a more recent proof of the ergodic theorem, to help relieve the tedium. At the other extreme we have added an appendix (Appendix A) explaining some of the necessary background in measure theory.

1.1 Invariant measures.

In measure theory the basic objects are *measurable spaces* $(X,\mathfrak{B})$, where X is a set and $\mathfrak{B}$ is a sigma algebra (or σ-algebra). In ergodic theory the basic objects are measurable spaces $(X,\mathfrak{B})$ *and* a measurable transformation $T: X\to X$ (i.e. with $T^{-1}\mathfrak{B}\subset\mathfrak{B}$). Given such a transformation $T: X\to X$ we want to consider those probability measures $m: \mathfrak{B}\to\mathbb{R}^+$ which are 'appropriate' for T in the following sense.

Definition. A probability measure m is called *invariant* (or more informatively, *T-invariant*) if $m(T^{-1}B) = m(B)$ for all sets $B\in\mathfrak{B}$.

This definition just tells us that the sets B and T^{-1}B always have the same 'size' relative to the invariant measure m. It is easy to see that if the transformation T is a bijection and its inverse $T^{-1}: X\to X$ is again measurable then the above definition is equivalent to asking that $m(B) = m(TB)$ for all sets $B\in\mathfrak{B}$.

An alternative formulation of this definition is to ask that $T^*m=m$ where $T^*:\mathcal{M}\to\mathcal{M}$ is the map on the set $\mathcal{M}$ of all probability measures on X defined by $(T^*m)(B) = m(T^{-1}B)$, for all $B\in\mathfrak{B}$.

Notation. We shall denote the set of all T-invariant probability measures on X by $\mathcal{M}_{\text{inv}}$.

We shall now give a trivial lemma which just gives another reformulation of the definition.

Lemma 1.1 (Characterizing invariant measures). The following statements are equivalent:

(a) $m \in \mathcal{M}_{\text{inv}}$;
(b) $\int f \circ T \, dm = \int f \, dm$, for all $f \in L^1(X, \mathcal{B}, m)$.

Proof. (b)$\Rightarrow$(a). For any set $B \in \mathcal{B}$ let $f = \chi_B$ be the characteristic function defined by

$$\chi_B(x) = \begin{cases} 1 \text{ if } x \in B \\ 0 \quad \text{otherwise} \end{cases}$$

We then have that $m(B) = \int f \; dm = \int f \circ T \; dm = m(T^{-1}B)$, and by definition $m \in \mathcal{M}_{\text{inv}}$.

(a)$\Rightarrow$(b). Starting with a measure $m \in \mathcal{M}_{\text{inv}}$ we have that $\int \chi_B \circ T \, dm = m(B) = m(T^{-1}B) = \int \chi_B \, dm$, for any set $B \in \mathcal{B}$. Thus by approximating any function $f \in L^1(X, \mathcal{B}, m)$ (in the L^1 norm) by finite combinations of these characteristic functions the result follows. □

We shall now begin to build up a collection of stock examples of transformations and invariant measures.

Examples. (i) Let $X = [0,1)$ be the half open unit interval with the usual Borel σ-algebra. For our transformation we choose $T: X \to X$ to be the fractional part of x multiplied by ten: $Tx = 10x$ (modulo unity). This transformation has many invariant measures, from which we shall choose m to be the usual Lebesgue measure.

To see that m is invariant we first consider intervals of the form $[a,b)$ for which we have that $T^{-1}[a,b) = \bigcup_{i=0}^{9} [\frac{a+i}{10}, \frac{b+i}{10})$ and therefore

$$m(T^{-1}[a,b)) = \sum_{i=0}^{9} \left| \frac{b+i}{10} - \frac{a+i}{10} \right| = b - a = m([a,b)).$$

Once we have this result for intervals (and the algebra of all their finite unions) we can deduce the same for the the whole σ-algebra $\mathcal{B}$ by wheeling out the machinery of the extension theorem (which we describe in Appendix A). A similar approach can be applied to all of the following examples.

(ii) Let $X=(0,1)$ be the open unit interval with the usual Borel σ-algebra. We can choose our transformation $T: X\to X$ to be the *Gauss map* defined by $Tx = 1/x$ (modulo unity). An interesting choice of invariant measure is the Gauss measure defined by

$$m(B)=\int_B \frac{1}{\log 2\,(1+x)}\,dx,\ B\in\mathfrak{B}.$$

Here m is equivalent to the usual Lebesgue measure. (NB. To make T well defined we *should* throw out points $\{1/n \mid n>0\}$, etc. But since these are a set of zero measure (relative to m) we shall just adopt the convention of ignoring them.)

(iii) Let $X=\mathbb{R}^2/\mathbb{Z}^2$ be the standard flat torus and let $\mathfrak{B}$ be the usual Borel sigma algebra. We define the transformation $T: X\to X$ to be a rotation in each of the co-ordinates of the form $T(x_1,x_2)=(x_1+\alpha_1,x_2+\alpha_2)$ where $\alpha_1,\alpha_2\in\mathbb{R}$ (Figure 1). The usual Lebesgue-Haar measure on X is invariant.

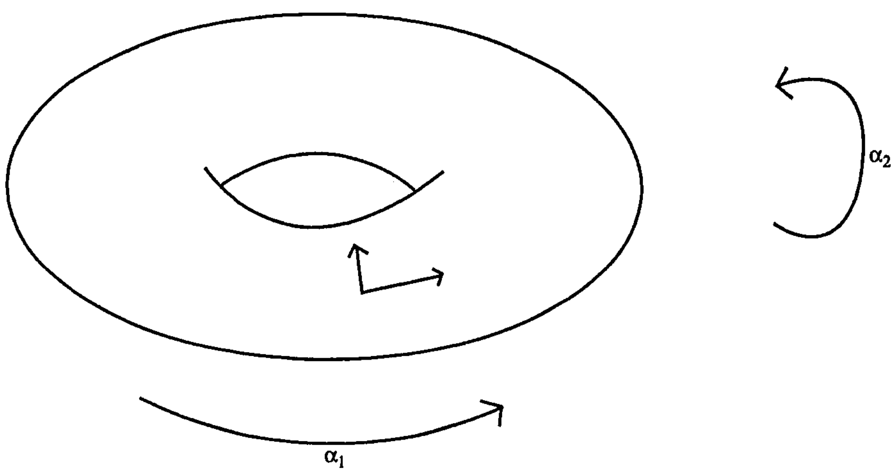

Figure 1: Irrational rotations on tori

(iv) Finally, we have the following trivial examples:

(a) Let $(X,\mathfrak{B})$ be any measurable space and let $T: X\to X$ be the identity transformation. Therefore, *all* probability measures are invariant and $\mathcal{M} = \mathcal{M}_{inv}$;

(b) Let X be any set and let $\mathfrak{B} = \{\emptyset, X\}$ be the trivial sigma algebra. There exists exactly one probability measure m defined by $m(X) = 1$, $m(\emptyset) = 0$ which will be invariant for any measurable transformation $T: X \to X$. Therefore, $\mathcal{M} = \mathcal{M}_{\text{inv}} = \{m\}$.

Standard convention. Whenever we state a result for *almost all* $x \in X$, with respect to the probability measure m (usually abbreviated to $a.a.(m)$ $x \in X$) it means that there exists $\Omega \in \mathfrak{B}$ with $m(\Omega) = 1$ for which the result holds for $x \in \Omega$. In most of our statements involving measures, this condition is implicit (even when not explicitly stated).

We now come to a rather basic question:

When does a transformation $T: X \to X$ actually have an invariant measure?

Fortunately, in all of the cases we will be interested in this turns out not to be a problem at all, by virtue of the following lemma (whose proof is both short *and* sweet).

Lemma 1.2. (Existence of invariant measures). If $T: X \to X$ is a homeomorphism of a compact topological space and $\mathfrak{B}$ is the usual Borel sigma algebra then there always exists at least one invariant measure (i.e. $\mathcal{M}_{\text{inv}} \neq \emptyset$).

Proof. The space $\mathcal{M}$ of all probability measures is a convex, compact, non-empty, topological space with respect to the weak* topology (a beautiful fact that we recall in Appendix A). Choose any measure $m \in \mathcal{M}$ and then for any $n \geq 1$ define a measure $m^{(n)}$ by the affine combination

$$m^{(n)} = \frac{m + T^* m + (T^*)^2 m + \ldots + (T^*)^{(n-1)} m}{n}$$

Since $\mathcal{M}$ is a convex space we see that $m^{(n)} \in \mathcal{M}$. Furthermore, since $\mathcal{M}$ is a compact space there must exist an accumulation point $\widetilde{m} \in \mathcal{M}$ for the sequence $\{m^{(n)}\}_{n=1}^{+\infty}$ i.e. $\widetilde{m} = \lim_{i \to +\infty} m^{(n_i)}$, for some subsequence $m^{(n_i)}$, $i \geq 1$. Finally, we can conclude that $\widetilde{m} \in \mathcal{M}_{\text{inv}}$ (and therefore $\mathcal{M}_{\text{inv}} \neq \emptyset$) since we can write

$$T^* \widetilde{m} = \lim_{i \to +\infty} T^* m^{(n_i)} = \lim_{i \to +\infty} \left(m^{(n_i)} + \frac{m - (T^*)^{n_i} m}{n_i} \right) = \lim_{i \to +\infty} m^{(n_i)} = \widetilde{m}$$

(NB. This proof is just a simple version of the Schauder fixed point theorem.) □

1.2 Poincaré Recurrence.

One of the most basic results, and probably also one of the most useful, in ergodic theory is the following.

Theorem 1.1 (Poincaré recurrence). Let $T: X \to X$ be a transformation on a measurable space $(X, \mathfrak{B})$ and let $m \in \mathcal{M}_{inv}$ be an invariant measure. For any set $B \in \mathfrak{B}$ almost all points $x \in B$ return to B under some iterate of T (i.e. $m(F)=0$ where $F = \{b \in B \mid T^n b \notin B, \forall n \geq 1\}$).

Proof. By definition, we have that $T^n F \cap B = \emptyset$ $\forall n \geq 1$ and so since $F \subset B$ it trivially follows that $T^n F \cap F = \emptyset$ for $n \geq 1$. By repeatedly applying the inverse of T we get that $T^{(n-k)} F \cap T^{-k} \subseteq T^{-k}(T^n F \cap F) = \emptyset$, for $n \geq k \geq 0$ and we can deduce that the sets $\{T^{-k} F\}_{k=0}^{+\infty}$ are all pairwise disjoint. Finally, we can write that

$$1 \geq m\left(\bigcup_{n=0}^{+\infty} T^{-n} F\right) = \sum_{n=0}^{+\infty} m(T^{-n} F) = \sum_{n=0}^{+\infty} m(F),$$

since by T invariance of the measure m we have $m(T^{-n} F) = m(F)$ for $n \geq 1$. Thus, we deduce that $m(F)=0$. □

1.3 Ergodic measures.

Now that we have mastered invariant measures we shall turn our attention to a particularly important type of invariant measure.

Definition. An invariant measure $m \in \mathcal{M}_{inv}$ is called *ergodic* if whenever $T^{-1} B = B$, for some $B \in \mathfrak{B}$, then either $m(B)=0$ or $m(B)=1$.

This definition just says that the only set B which also equals $T^{-1} B$ must have the same 'size' as the whole space X or be trivial, *relative to the invariant measure m*. It is easy to see that if the transformation T is a bijection and its inverse $T^{-1}: X \to X$ is again measurable then the above definition is equivalent to saying that whenever $B = TB$, for some set $B \in \mathfrak{B}$, then $m(B) = 0$ or $m(B) = 1$.

Notation. We shall denote the set of ergodic measures by $\mathcal{M}_{erg} = \{m \in \mathcal{M}_{inv} \mid m \text{ is ergodic}\}$.

We shall now give a trivial lemma which reformulates this definition.

Lemma 1.3 (Characterizing ergodic measures). The following statements are equivalent:
(a) $m \in \mathcal{M}_{erg}$;
(b) $f \circ T = f$, for some $f \in L^1(X, \mathfrak{B}, m) \Rightarrow f$ is constant a.e.(m).

Proof. (b)$\Rightarrow$(a). Given a set $B \in \mathfrak{B}$ satisfying $T^{-1}B = B$ we define $f = \chi_B$ to be the characteristic function of B. The condition on our set B implies that $f \circ T = f$ and so we deduce that f is constant (up to a set of measure zero). For a characteristic function this means the value is either 0 or 1, i.e. $m(B) = 0$ or $m(B) = 1$.

(a)$\Rightarrow$(b). If we could find a function $f \in L^1(X, \mathfrak{B}, m)$ which was *not* constant (up to a set of measure zero) then we could choose a real number c such that the set $B = \{x \in X \mid f(x) \geq c\}$ has measure $0 < m(B) < 1$. However the assumption $f \circ T = f$ implies that $T^{-1}B = B$, which leads us into a contradiction. □

A very reasonable question to ask at this stage is:

What makes ergodic measures particularly interesting?

We want to give two different answers to this question in the next two sections.

1.4 Ergodic decomposition.

Problems about arbitrary invariant measures in $\mathcal{M}_{inv}$ can frequently be reduced to (simpler) problems about ergodic measures in $\mathcal{M}_{erg}$ (at least when X is a topological space).

Given any compact, convex topological space C we define the *extremal points* to be the set

$$\text{Ext}(C) = \{x \in C \mid x = \alpha x_1 + (1-\alpha)x_2,\ 0 \leq \alpha \leq 1,\ x_1 \neq x_2 \Rightarrow \alpha = 0 \text{ or } \alpha = 1\}.$$

It is easy to see that if $C \neq \emptyset$ then $\text{Ext}(C) \neq \emptyset$ (Figure 2).

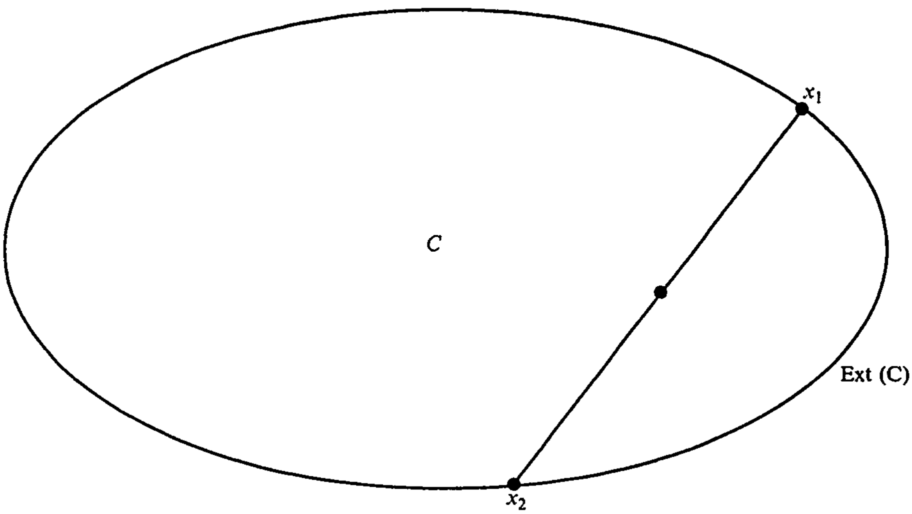

Figure 2: Convex sets and extremal points

With the choice $C = \mathcal{M}_{inv}$ we have the following useful result.

Lemma 1.4 (Ergodic decomposition). The extremal points of $\mathcal{M}_{inv}$ are precisely $\mathcal{M}_{erg}$ (i.e. $\text{Ext}(\mathcal{M}_{inv}) = \mathcal{M}_{erg}$). Furthermore, given any invariant measure $m \in \mathcal{M}_{inv}$ there exists a probability measure μ_m on the space $\mathcal{M}_{inv}$ such that:

(a) $\mu_m(\mathcal{M}_{erg}) = 1$; and

(b) $\int_X f \, dm = \int_{m' \in \mathcal{M}_{erg}} \left(\int_X f \, dm' \right) d\mu_m(m')$, for any $f \in L^1(X, \mathfrak{B}, m)$.

(i.e. the invariant measure m is an affine combination, weighted by μ_m, of *ergodic* measures m').

The first part of this lemma is an easy exercise (based on the observation that distinct ergodic measures are mutually singular, see