# ENTROPY IN ERGODIC THEORY
# AND TOPOLOGICAL DYNAMICS

T.C. TREMAINE

A Thesis

in

the Department

of

Mathematics

Presented in Partial Fulfillment
of the Requirements for the degree of
Master of Science at
Concordia University
Montréal, Québec, Canada

February, 1983.

## ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. Bill Byers for his teaching, guidance and support throughout the preparation of this thesis and to Judith Hickey whose skillful typing is so evident in the pages that follow.

ABSTRACT

ENTROPY IN ERGODIC THEORY
AND TOPOLOGICAL DYNAMICS

T.C. TREMAINE

The concept of measure-theoretic entropy was originated by Kolmogoroff in 1958. Topological entropy began with the work of Adler, Konheim and McAndrew in 1965. A number of conputational theorems have been developed in the interim, including the Kolmogoroff-Sinai theorem and its topological analogue. Some of these are presented, together with proofs. A theorem by Goodman $(h(T) = \sup\{h_\mu(T) | \mu \in M(X,T)\})$ is presented and discussed. The work of M. Misiurewicz and W. Szlenk relating the growth number of piecewise monotone continuous maps to their topological entropy $(h(T) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{Var}(T^n))$ and the number of periodic points to the topological entropy $(h(T) \le \overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathrm{Card}\{x | T^n(x) = x\})$ is presented with proofs. An exposition of a paper by Block, Guckenheimer, Misiurewicz and Young is given. The main result (for a point of period $q \cdot 2^m$, $q$ (odd), $\frac{1}{2^m} \log \lambda_q \le h(T)$, where $\lambda_q = \max\{|z| : z \in \sigma_q\}$ and $\sigma_q$ is the set of roots of $x^q - 2x^{q-2} - 1 = 0$) is presented with a proof.

TABLE OF CONTENTS

## INTRODUCTION

In 1958 A.N. Kolmogoroff introduced a new invariant (entropy) into ergodic theory which he used to solve an outstanding problem concerning isomorphism among Bernoulli shifts. In Chapter II we present measure-theoretic entropy in the form in which it is currently used. Various computational devices are presented culminating in the Kolmogoroff-Sinai theorem which is proved and applied.

In 1965 R.A. Adler, A. Konheim, and M. McAndrew introduced an analogue of Kolmogoroff's invariant into topological dynamics. Most of the computational theorems for measure-theoretic entropy have their analogues for topological entropy. These are presented in Chapter III together with some examples of their application.

The variational principle for topological entropy is presented (without proof) and discussed in Cahpter III. This principle states the relationship between the two types of entropy and shows why the theory of the former type is so faithfully mirrored in its topological analogue.

In Chapter IV we discuss the work of M. Misiurewicz on the topological entropy of piecewise monotone continuous maps with special emphasis placed on the relatively simple linear Markov maps.

Chapters IV and V contain a discussion of the relationship between periodic points and topological entropy culminating in an exposition of the elegant results obtained by L. Block, J. Guckenheimer, M. Misiurewicz and L.S. Young. Their method involves modelling the dynamical behaviour on the interval by means of the dynamical behaviour of an oriented graph. Then results relating to the entropy of the original map can be derived by means of the graph and its structure.

Concerning the numbering system used throughout this work the following applies. Theorems, propositions and lemmas are together numbered sequentially. Sections and examples are numbered independently from the above and from each other.

# CHAPTER I

## PRELIMINARIES

### Section 1. Introduction

The purpose of this chapter is to introduce some of the fundamental concepts and definitions which will be used throughout.

### Section 2. Topological Spaces

A collection of sets $\tau \subseteq P(X)$ is called a topology on $X$ if

1) $\phi \in \tau$ and $X \in \tau$

2) If $A, B \in \tau$ then $A \cap B \in \tau$

3) If $\gamma \subseteq \tau$ then $\cup\{A | A \in \gamma\} \in \tau$.

The members of $\tau$ are referred to as open sets.

If $X' \subseteq X$ then the collection of sets $\tau' = \{A \cap X' | A \in \tau\}$ is called the subspace topology of $X'$ relative to $X$.

A collection of sets $\beta \subseteq P(X)$ is called a basis for $\tau$ if every member of $\tau$ is a union of members of $\beta$.

A collection of sets $\alpha \subseteq \tau$ are referred to as an open cover of $X' \subseteq X$ if $X' \subseteq \cup\{A | A \in \alpha\}$.

A collection of sets $\gamma \subseteq \tau$ is referred to as a finite open cover of $X' \subseteq X$ if $\gamma$ is an open cover of finite cordinality.

A set $X' \subseteq X$ is called compact if every open cover $\alpha$ of $X'$ has a subset $\alpha'$ which is a finite open cover of $X'$.

Suppose $f: X \to X$ is a function. The sequence $\langle x, f(x), f^2(x), \ldots \rangle$ shall be referred to as the orbit of x under f. $f^n(x)$ means $(\underbrace{f \circ f \circ \ldots \circ f}_{n \text{ times}})(x)$.

Many of the functions which will be considered will be defined on closed intervals of $\mathbb{R}$. The topology used will be the subspace topology relative to $\mathbb{R}$. The topology on $\mathbb{R}$ will be the usual topology which is formed by taking as a basis all sets of the form $(a,b); a,b \in \mathbb{R}$.

Section 3. Measure Spaces

Let $X$ be a set. A collection $\mathcal{A} \subseteq \mathcal{P}(X)$ is termed an algebra if

1) $E, F \in \mathcal{A}$ implies $E \cup F \in \mathcal{A}$

2) $F \in \mathcal{A}$ implies $F^c \in \mathcal{A}$.

A collection $\mathcal{B}$ of subsets of $X$ is called a $\sigma$-algebra if $\mathcal{B}$ is an algebra and if $B_n \in \mathcal{B}$ for $n \geq 1$ implies $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$.

It is clear that the intersection of any collection of algebras is also an algebra and the intersection of any collection of $\sigma$-algebras is also a $\sigma$-algebra.

Suppose we have an arbitrary collection $C$ of subsets of $X$. Let us refer to the intersection of all algebras containing $C$ as the algebra generated by $C$ and denoted it by $\mathcal{A}(C)$. Let us refer to the intersection of all $\sigma$-algebras containing $C$ as the $\sigma$-algebra generated by $C$ and denote it by $\mathcal{B}(C)$. $\mathcal{A}(C)$ and $\mathcal{B}(C)$ always exist since for any collection $C$ of subsets of $X$ there is always an algebra and $\sigma$-algebra containing $C$, namely the power set of $X$. If $\mathcal{B}$ is a $\sigma$-algebra of subsets of $X$ we shall refer to the pair $(X,\mathcal{B})$ as a measurable space.

A finite measure on the pair $(X,\mathcal{B})$ is a set function $\mu:\mathcal{B} \to [0,\infty)$ s.t.

1) $\mu(\phi) = 0$

2) $\mu(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n)$

for any collection $\{B_n\}_{n=1}^{\infty}$ of pairwise disjoint members of $\mathcal{B}$.

A <u>finite measure space</u> is a triple $(X,\mathcal{B},\mu)$ where $(X,\mathcal{B})$ is a measurable space and $\mu$ is a finite measure.

The triple $(X,\mathcal{B},\mu)$ is a <u>normalized measure space</u> (also referred to as a probability space) if $\mu(X) = 1$.

In order to establish various results pertaining to algebras and $\sigma$-algebras it is often useful to concentrate on a more restricted class of sets in a manner analogous to the use of sub-basic sets in topology.

A collection $\zeta$ of subsets of $X$ is called a <u>semi-algebra</u> if

1) $\phi \in \zeta$

2) $A,B \in \zeta$ implies $A \cap B \in \zeta$

3) If $A \in \zeta$ then

$A^c = \bigcup_{i=1}^{n} E_i$ where the finite collection $\{E_i\}_{i=1}^{n}$ consists of pairwise disjoint members of $\zeta$.

The following theorem shows how to construct the generated algebra $A(\zeta)$ from a semi-algebra $\zeta$.

1.1 <u>Theorem</u> ([21], p.19)

Suppose $\zeta$ is a semi-algebra of subsets of $X$. Then $A(\zeta)$ consists of those subsets of $X$ which are finite unions of pairwise disjoint members of $\zeta$.

Another useful concept for obtaining measure theoretic results is that of a monotone class.

A non empty class $M$ of subsets of $X$ is called <u>monotone</u> if

1) for every sequence $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots$ of sets in $M$, $\bigcup_{n=1}^{\infty} E_n \in M$.

2) for every sequence $E_1 \supseteq E_2 \supseteq E_3 \supseteq \ldots$ of sets in $M$, $\bigcap_{n=1}^{\infty} E_n \in M$.

It is clear that the intersection of any collection of monotone classes is also a monotone class. Furthermore, the power set of $X$ is a monotone class. Thus the following definition is meaningful.

If $C$ is any collection of subsets of $X$, let us refer to the intersection of all monotone classes containing $C$ as the <u>monotone class generated by C</u>, denoted $M(C)$.

The following theorem states the relationship between the monotone class generated by an algebra $A$ and the $\sigma$-algebra generated by $A$.

1.2 <u>Theorem</u> ([13], p.27)

Let $A$ be an algebra of subsets of $X$. Then $M(A) = B(A)$.

A substantial amount of emphasis will be placed on the use of partitions. Such objects are defined as follows. Suppose $(X,B,\mu)$ is a normalized measure space. A <u>partition</u> $\xi$ of $(X,B,\mu)$ is an at most countable subset of $B$ with the following properties:

1) If $A,B \in \xi$ then $A \cap B = \phi$ unless $A = B$.

2) $\sum_{A \in \xi} \mu(A) = 1$.

Let $(X,B,\mu)$ be a measure space. We may introduce the notion of conditional measure analogous to conditional probability.

Let $A,B \in B$. Then the <u>conditional measure of A given B</u>, denoted $\mu(A|B)$, is given by: $\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)}$. If $(X,B,\mu)$ is normalized then $\mu(A|B)$ is just the conditional probability of A given B.

Section 4. <u>Measure-Preserving Transformations</u>

In much of what follows we will be dealing with transformations which have the property of preserving the measure of sets under inverse images. They are in a sense analogous to continuous maps. Where continuous maps preserve the topological character of sets under inverse images a measure preserving map preserves their measure theoretic character.

Let $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ be two measure spaces. A transformation $T: X_1 \to X_2$ will be termed <u>measurable</u> if $T^{-1}\mathcal{B}_2 \subseteq \mathcal{B}_1$. If $T$ is measurable and $\mu_1(T^{-1}B) = \mu_2(B)$ for every $B \in \mathcal{B}_2$ then $T$ will be called <u>measure preserving</u> (abbreviated m.p.). If $T$ is bijective and both $T$ and $T^{-1}$ are measure preserving then $T$ will be referred to as an <u>invertible measure preserving</u> transformation.

The following theorem provides a means of checking whether a given map $T$ is m.p. by examining its behavior on a semi-algebra generating $\mathcal{B}_2$.

1.3 <u>Theorem</u> ([24], p.20)

Suppose $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ are normalized measure spaces and $T: X_1 \to X_2$ is a transformation. Let $\zeta$ be a semi-algebra generating $\mathcal{B}_2$. Then $T$ is m.p. if

1) $T^{-1}\zeta \subseteq \mathcal{B}_1$ and

2) $\mu_1(T^{-1}A) = \mu_2(A)$ for all $A \in \zeta$.

<u>Example 1</u>

Let $T: [0,1] \to [0,1]$ be defined by:

$$T(x) = \begin{cases} 2x & x \in [0, \frac{1}{2}] \\ 2-2x & x \in [\frac{1}{2}, 1] \end{cases}$$

Let $\mathcal{B}$ be the $\sigma$-algebra generated by the collection of all open subintervals of $[0,1]$. Let $m$ be Lebesgue measure. It can be readily shown that the collection $\zeta$ of all intervals (open, closed, half-open) together with $\phi$ is a semi algebra which generates $\mathcal{B}$. So, we need only show that

1) $T^{-1}\zeta \subseteq \mathcal{B}$ . and

2) $m(T^{-1}B) = m(B)$ for all $B \in \zeta$ .

Let $B \in \zeta$ . If $B = \phi$ then $T^{-1}\phi = \phi \in \mathcal{B}$ and $m(T^{-1}B) = m(B) = 0$ . If $B \neq \phi$ then $B$ is an interval. $T^{-1}B$ is a union of at most two intervals. Thus $T^{-1}B \in \mathcal{B}$ . Let $B = [a,b]$ . Then

$$T^{-1}B = [\frac{a}{2}, \frac{b}{2}] \cup [1-\frac{b}{2}, 1-\frac{a}{2}]$$

$$m(T^{-1}B) = (\frac{b}{2}-\frac{a}{2}) + (1-\frac{a}{2}-1+\frac{b}{2}) = b-a .$$

$$m(B) = b - a .$$

Thus $m(T^{-1}B) = m(B)$ . The same obviously holds for open and half-open intervals. Thus by Theorem 1.3 $T$ is m.p.

Suppose $(X,\tau)$ is a topological space. We may use the open sets of $X$ as a generating set for a $\sigma$-algebra on $X$ . Let the $\sigma$ algebra generated by the members of $\tau$ be referred to as the Borel class and be denoted by $\mathcal{B}(X)$ .

Let $M(X)$ be the collection of all normalized measures on the measurable space $(X,\mathcal{B}(X))$.

Let $M(X,T)$ be the set of all measures in $M(X)$ for which $T:X \to X$ is measure preserving. $M(X,T)$ will be referred to as the set of invariant measures with respect to $T$ .

The following theorem gives sufficient conditions for the existence of a measure in $M(X,T)$.

1.4 Theorem ([24], p.151)

Let $X$ be a compact metric space and $T:X \to X$ be continuous. Then $M(X,T)$ is non-empty.

Section 5. Sequence Spaces and Shift Maps

Let $X = \{0,1,2,\ldots,k-1\}$ .

The members of X shall be referred to as states or symbols and X shall be referred to as the state space.

$X^Z$ and $X^N$ will denote the infinite product spaces $\prod_{i=-\infty}^{\infty} X_i$ and $\prod_{i=0}^{\infty} X_i$ respectively where each is endowed with the product topology and $X_i = X$ for all $i$. The open sets of X are simply all subsets of X.

So, an element $x \in X^Z$ is an object of the form $(\ldots x_{-2}\, x_{-1}\, x_0 x_1 x_2 \ldots)$ and an element $x \in X^N$ is an object of the form $(x_0 x_1 x_2 \ldots)$ where $x_i \in X$ for every $i$. The spaces $X^Z$ and $X^N$ will be referred to as sequence spaces. For $i \in Z$ (or N) the term co-ordinate will be used. $x_i$ will be referred to as the value of the i-th co-ordinate of x. The notation $(x)_i = x_i$ will be used to indicate that the value of the i-th co-ordinate of x is being assessed.

The map $\sigma: X^Z \to X^Z$ defined by $(\sigma(x))_i = x_{i+1}$ will be called the shift on $X^Z$. Similarly the map $\sigma: X^N \to X^N$ defined by $(\sigma(x))_i = x_{i+1}$ will be referred to as the shift on $X^N$.

A finite sequence $(a_1, a_2, \ldots, a_N)$ of elements $a_j \in X$ shall be referred to as a block. N is called the length of the block. The block $A = (a_1, \ldots, a_N)$ is said to occur at the place m if $x_m = a_1$, $x_{m+1} = a_2, \ldots, x_{m+N-1} = a_N$.

For any m in Z (or N) and any block $A = (a_1, \ldots, a_N)$ let ${}_m[a_1, a_2, \ldots, a_N] = {}_m[A]$ denote the set of all $x \in X^Z$ (or $X^N$) such that $(a_1, a_2, \ldots, a_N)$ occurs in x at the place m. The set ${}_m[A]$ is called a cylinder of length N based on the block A at the place m.

Suppose there is a set function $\mu$ defined on the class of cylinders of $X^Z$ (or $X^N$) satisfying the properties:

1) $\sum_{a_o \in X} \mu({}_o[a_o]) = 1$ .

And, for any block $(a_o, a_1, \ldots, a_s)$ and any $n \in Z$ (or $N$)

2) $\mu({}_n[a_o, a_1, \ldots, a_s]) \geq 0$ .

3) $\mu({}_n[a_o, a_1, \ldots, a_s]) = \sum_{a_{s+1} \in X} \mu({}_n[a_o, a_1, \ldots, a_{s+1}])$

4) $\mu({}_n[a_o, a_1, \ldots, a_s]) = \sum_{a_{-1} \in X} \mu({}_n[a_{-1}, a_o, \ldots, a_s])$

The following theorem states that these four properties are sufficient to ensure that $\mu$ induces a unique measure $\bar{\mu}$ by extension and that $\bar{\mu}$ is in $M(X^Z, \sigma)$ (or $M(X^N, \sigma)$).

1.5 <u>Theorem</u> ([3], p.33-35)

Let $\mu$ be a function on the cylinder sets of $X^Z$ (or $X^N$) satisfying conditions (1) - (4). Then there exists a unique measure $\bar{\mu} \in M(X^Z, \sigma)$ (or $M(X^N, \sigma)$) s.t. $\bar{\mu}({}_n[A]) = \mu({}_n[A])$ for all cylinder sets ${}_n[A]$ in $X^Z$ (or $X^N$) .

Henceforth we shall refer to $\bar{\mu}$ as $\mu$ .

Two methods shall be utilized for producing measures on the sequence spaces $X^Z$ and $X^N$ .

A k-tuple $\pi = (p_0, p_1, \ldots, p_{k-1})$ with $p_i > 0$ and $\sum_{i=o}^{k-1} p_i = 1$ shall be referred to as a <u>probability vector</u>. A probability vector defines a set function $\mu_\pi$ on the cylinders of $X^Z$ (or $X^N$) by $\mu_\pi({}_n[a_o, a_1, \ldots, a_s]) = p_{a_o} p_{a_1} \ldots p_{a_s}$. Conditions (1) - (4) referred to previously are evidently satisfied. Thus, by Theorem 1.5 the set function $\mu_\pi$ has a unique extension, also denoted $\mu_\pi$ , which is invariant under the shift on both $X^Z$ and $X^N$ . $\mu_\pi$ is called a <u>Bernoulli measure</u>. The system $(X^Z, \mu_\pi, \sigma)$ is called the <u>two-sided Bernoulli $\pi$-shift</u>. The system $(X^N, \mu_\pi, \sigma)$ is called the <u>one-sided</u>

Bernoulli $\pi$-shift.

A second important method of introducing a measure on a sequence space is as follows.

Suppose we have a probability vector $\pi = (p_o,\ldots,p_{k-1})$ and a transition matrix (also called a stochastic matrix) which is defined as a $k \times k$ matrix $P = (p_{ij})_{i,j=o}^{k-1}$ with the properties

1) $p_{ij} \geq 0$ .
2) $\sum_{j=o}^{k-1} p_{ij} = 1$ .
3) $\sum_{i=o}^{k-1} p_i p_{ij} = p_j$ .

The pair $(\pi,P)$ defines a set function $\mu_{\pi p}$ on the cylinder sets of $X^Z$ (or $X^N$) by $\mu_{\pi P}({}_n[a_o,a_1,\ldots,a_k]) = P_{a_o} P_{a_o a_1} P_{a_1 a_2} \cdots P_{a_{k-1} a_k}$

Properties (1) - (4) referred to earlier are satisfied.

Thus, by Theorem 1.5 the set function $\mu_{\pi P}$ has a unique extension, also denoted $\mu_{\pi P}$ , which is invariant under the shift on both $X^Z$ and $X^N$ . $\mu_{\pi P}$ is called a Markov measure.

The system $(X^Z,\mu_{\pi P},\sigma)$ is called the two sided Markov $\pi$,P-shift. The system $(X^N,\mu_{\pi P},\sigma)$ is called the one sided Markov $\pi$,P-shift.

Section 6. Isomorphism

It is very useful to have a means of determining when two measure spaces are "equivalent". In topology two spaces are regarded as equivalent if there is a homeomorphism mapping one to the other. An analogous map between measure spaces would have to preserve the essential measure theoretic character of the space but would be able to "ignore" what is measure theoretically non-essential.

Suppose there are two probability spaces $(X_1, \mathcal{B}_1, \mu_1)$ and

$(X_2, \mathcal{B}_2, \mu_2)$. They are said to be isomorphic if there are sets $M_1 \in \mathcal{B}_1$ and $M_2 \in \mathcal{B}_2$ with $\mu_1(M_1) = \mu_2(M_2) = 1$ and a map $\phi: M_1 \to M_2$ which is invertible measure preserving with respect to the measure spaces $(M_1, \mathcal{B}_1', \mu_1)$ and $(M_2, \mathcal{B}_2', \mu_2)$. $\mathcal{B}_i'(i=1,2)$ is defined by $\mathcal{B}_i' = \{M_i \cap B \mid B \in \mathcal{B}_i\}$. $\phi$ is referred to as a measure theoretic isomorphism between $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$. The notation $(X_1, \mathcal{B}_1, \mu_1) \overset{\phi}{\rightleftarrows} (X_2, \mathcal{B}_2, \mu_2)$ will be used to indicate that $\phi$ is a m.t. isomorphism between the two given spaces. If $\phi: M_1 \to M_2$ is not invertible but only onto and measure preserving then $\phi$ is referred to as a measure theoretic homomorphism from $(X_1, \mathcal{B}_1, \mu_1)$ to $(X_2, \mathcal{B}_2, \mu_2)$. The notation $(X_1, \mathcal{B}_1, \mu_1) \overset{\phi}{\to} (X_2, \mathcal{B}_2, \mu_2)$ will be used in such a case.

We may extend the notion of equivalence to maps. This gives us a means of examining the behaviour of one map by studying another simpler or better understood map.

Suppose $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ are probability spaces and $T_1: X_1 \to X_1$ and $T_2: X_2 \to X_2$ are measure preserving transformations. We shall say that $T_1$ is isomorphic to $T_2$ if there is some map $\phi$ such that:

1) $(X_1, \mathcal{B}_1, \mu_1) \overset{\phi}{\rightleftarrows} (X_2, \mathcal{B}_2, \mu_2)$ and

2) $(\phi \circ T_1)(x) = (T_2 \circ \phi)(x)$

for all $x$ in the domain of $\phi$. We shall say that $T_1$ is homomorphic to $T_2$ or that $T_2$ is a factor of $T_1$ if there is a map $\psi$ such that

1) $(X_1, \mathcal{B}_1, \mu_1) \overset{\psi}{\to} (X_2, \mathcal{B}_2, \mu_2)$ and

2) $(\psi \circ T_1)(x) = (T_2 \circ \psi)(x)$

for all $x$ in the domain of $\psi$.

Example 2

Let the system $(T,I,\mathcal{B}(I),m)$ be defined as follows. $I=[0,1]$ with the usual subspace topology. $\mathcal{B}(I)$ is the Borel class of sets with Lebesgue measure $m$. $T:I\to I$ is given by $T(x)=2x \pmod 1$.

Let the system $(\sigma, X, \mathcal{B}(x),\mu_\pi)$ be defined as follows. $X=\{0,1\}^N$ where $\{0,1\}$ has the discrete topology and $\{0,1\}^N$ the product topology. $\mathcal{B}(X)$ is the Borel class of sets and $\mu_\pi$ the Bernoulli measure with $\pi=(\frac{1}{2},\frac{1}{2})$. $\sigma:X\to X$ is the one sided shift.

It can be shown that $T$ is isomorphic to $\sigma$.

It was previously stated (Sec. 5) that $\mu_\pi\in M(X,\sigma)$. We also observe that $T$ preserves Lebesgue measure $m$ since it splits each interval into two parts, the length of each being half the original. Since the collection of intervals forms a semi-algebra which generates $\mathcal{B}(I)$ the conclusion follows by 1.3.

We shall proceed to find a map $\phi$ such that $(I,\mathcal{B}(I),m)\overset{\phi}{\leftrightarrows}(X,\mathcal{B}(X),\mu_\pi)$. Let $D_1\subseteq I$ be the set of all dyadic rationals (those reals whose binary decimal expansion is eventually repeating ones or repeating zeros). Let $D_2\subseteq X$ be the set of all symbol sequences whose entries are eventually repeating ones or repeating zeros.

Both $D_1$ and $D_2$ have measure zero since they are both countable.

Let $M_1=I\setminus D_1$ and $M_2=X\setminus D_2$. Express each member of $M_1$ in binary decimal expansion and define the bijection $\phi:M_1\to M_2$ by $\phi(.x_0x_1x_2\ldots)=\langle x_0x_1x_2\ldots\rangle$.

Let $\zeta=\{{}_0[a_0,a_1,\ldots,a_n]\,|\,n\geq 0\}\cup\{\phi\}$. That is, $\zeta$ consists of all cylinder sets based at zero and the empty set. $\zeta$ is a semi algebra. It consists of open sets so $\mathcal{B}(\zeta)\subseteq\mathcal{B}(X)$. If we take finite unions of members of $\zeta$ we obtain all cylinder sets in $X$. Thus $A(\zeta)$ contains a sub-basis and thus a basis for $X$. Since this basis is countable ([8], p.174) it follows that every open set can be formed by countable unions of members of $A(\zeta)$. Thus $B(X) \subseteq B(\zeta)$. So, $B(\zeta) = B(X)$. Let $\zeta' = \{S \cap M_2 | S \in \zeta\}$ and $B'(X) = \{B \cap M_2 | B \in B(X)\}$. It is clear that $B(\zeta') = B'(X)$.

Let $\eta = \{(a,b) | a,b \text{ are dyadic rationals}\} \cup \{\phi\}$. Let $\eta' = \{H \cap M_1 | H \in \eta\}$ be the restriction of to $M_1$. Let $B'(I) = \{B \cap M_1 | B \subset B(I)\}$. $\eta'$ is a semi-algebra which generates $B'(I)$.

To show that $\phi$ is invertible measure preserving it is sufficient (by 1.3) to demonstrate that $\phi$ and $\phi^{-1}$ preserve the measures of members of $\zeta'$ and $\eta'$ resp. under inverse images.

Let ${}_0[a_0, a_1, \ldots, a_n]$ be a cylinder set im $M_2$. $\phi^{-1}({}_0[a_0, a_1, \ldots a_n])$ $= (.a_0a_1 \ldots a_n000\ldots\ ,\ .a_0a_1 \ldots a_n111\ldots)$. $\mu_\pi({}_0[a_0, a_1, \ldots, a_n]) = \frac{1}{2^{n+1}}$. And, $m(.a_0a_1 \ldots a_n000\ldots, .a_0a_1 \ldots a_n111\ldots) = \sum_{j=n+2}^{\infty} \frac{1}{2^j} = \frac{1}{2^{n+1}}$. Thus $\phi$ is measure preserving.

Let $J = (.a_0a_1 \ldots a_n000\ldots, .b_0b_1 \ldots b_n111\ldots)$ be a member of $\eta'$. $\phi(J)$ equals $\{{}_0[x_0, x_1, \ldots, x_n] | .a_0a_1 \ldots a_n \leq .x_0x_1 \ldots x_n \leq .b_0b_1 \ldots b_n\}$.

$$m(J) = \frac{1}{2^{n+1}} + \sum_{j=0}^{n} (b_j - a_j) 2^{-(j+1)}$$

$$\mu_\pi(\phi(J)) = \frac{1}{2^{n+1}} \left(\sum_{j=0}^{n} (b_j - a_j)^{n-j} + 1\right)$$

$$= \frac{1}{2^{n+1}} + \sum_{j=0}^{n} (b_j - a_j) 2^{-(j+1)} \ .$$

Thus $\phi^{-1}$ is measure preserving. Consequently, $(I, B(I), m) \overset{\phi}{\leftrightarrow} (X, B(X), \mu_\pi)$.

It is clear that $(\phi \circ T)(x) = (\sigma \circ \phi)(x)$ for every $x \in M_1$. Thus $T$ is isomorphic to $\sigma$.

Section 7. Vector Spaces

Some results from the theory of finite dimensional vector spaces will be required. For general definitions and theorems see [9] or [14].

Let M be the space of all $k \times k$ matrices with entries in C together with the norm $\|A\| = \sum_{i,j=0}^{k-1} |a_{ij}|$ where $A = (a_{ij})_{i,j=0}^{k-1}$. Now, for each $A \in M$ let $Sp(A) = \{\lambda \in C \mid Det(A - \lambda I) = 0\}$. Clearly, $Card(Sp(A)) \leq k$. Let $\rho(A) = \max\{|\lambda| : \lambda \in Sp(A)\}$. Sp(A) will be referred to as the spectrum of A and $\rho(A)$ the spectral radius of A.

1.6 Theorem (Spectral Radius Formula) ([22], p.235)

If $A \in M$ then $\lim_{n\to\infty} \|A^n\|^{1/n} = \rho(A)$.

Let A be a $k \times k$ matrix with non-negative real entries. Then A is called irreducible if for every pair i,j there is some n s.t. $a_{ij}^{(n)} > 0$ where $a_{ij}^{(n)}$ is the (i,j)-th entry in $A^n$.

1.7 Theorem (Perron-Frobenius) ([9], p.53)

If A is an irreducible $k \times k$ matrix then A has a positive eigenvalue r that is a simple root of $Det(A - xI) = 0$. The moduli of all other eigenvalues do not exceed r. Corresponding to r is an eigenvector with strictly positive coordinates.

# CHAPTER II

# MEASURE THEORETIC ENTROPY

## Section 1. Introduction

In 1958 Kolmogoroff introduced entropy into ergodic theory [15]. He was thereby able to solve an isomorphism problem relating to Bernoulli shifts.

Loosely speaking entropy is a numerical indicator of the degree of complexity in the orbit structure of a transformation. It is a dynamical invariant and thus can be used to partially clarify systems (those with unequal entropy are not equivalent).

The material of this chapter is derived primarily from Brown [6], Denker et al [7] and Walters [24].

## Section 2. The Entropy of a Partition

Let $(X, \mathcal{B}, \mu)$ be a normalized measure space. Recall that a partition $\xi$ of $(X, \mathcal{B}, \mu)$ is an at most countable subset of $\mathcal{B}$ with the two properties:

1) $A, B \in \xi$ $(A \neq B)$ implies $A \cap B = \phi$

2) $\sum_{A \in \xi} \mu(A) = 1$

**Definition**

Let $\xi, \eta$ be two partitions of $(X, \mathcal{B}, \mu)$. $\eta$ refines $\xi$, denoted $\xi \leq \eta$, if each member of $\xi$ is a union of members of $\eta$. $\eta$ refines $\xi$ (mod 0), denoted $\xi \overset{0}{\leq} \eta$, if for each $A \in \xi$ there is a subset $\eta' = \{B \in \eta \mid \mu(A \cap B) = \mu(B)\}$ of $\eta$ s.t. $\mu(A) = \sum_{B \in \eta'} \mu(B)$. ($\xi \overset{0}{\leq} \eta$ is equivalent to $\mu(A \cap B) = \mu(B)$ or $0$ for every $A \in \xi$ and $B \in \eta$). If $\xi \overset{0}{\leq} \eta$ and $\eta \overset{0}{\leq} \xi$ we will write $\xi \overset{0}{=} \eta$. It is clear that $\xi \leq \eta$ implies $\xi \overset{0}{\leq} \eta$.

Definition

Let $\xi, \eta$ be two partitions of $(X, \mathcal{B}, \mu)$. Their join, denoted $\xi \vee \eta$, is given by:

$$\xi \vee \eta = \{A \cap B \mid A \in \xi, \ B \in \eta\} .$$

($\xi \vee \eta$ is also a partition of $(X, \mathcal{B}, \mu)$).

Definition

Suppose $T:X \to X$ is measure preserving and $\xi$ is a partition of $(X, \mathcal{B}, \mu)$. Then $T^{-1}\xi = \{T^{-1}A \mid A \in \xi\}$.

The following proposition states some properties of inverse images of partitions.

2.1 Proposition

Let $\xi, \eta$ be two partitions of $(X, \mathcal{B}, \mu)$ and $T:X \to X$ be measure preserving. Then, for $n > 0$:

i) $T^{-n}(\xi \vee \eta) = T^{-n}\xi \vee T^{-n}\eta$ .

ii) $\xi \leq \eta \Rightarrow T^{-n}\xi \leq T^{-n}\eta$ .

iii) $\xi \overset{o}{\leq} \eta \Rightarrow T^{-n}\xi \overset{o}{\leq} T^{-n}\eta$ .

Proof:

i) $T^{-n}(\xi \vee \eta) = T^{-n}\{A \cap B \mid A \in \xi, \ B \in \eta\}$

$= \{T^{-n}(A \cap B) \mid A \in \xi, \ B \in \eta\}$

$= \{T^{-n}A \cap T^{-n}B \mid A \in \xi, \ B \in \eta\} = T^{-n}\xi \vee T^{-n}\eta$ .

ii) Let $T^{-n}A \in T^{-n}\xi$ . Since $\xi \leq \eta$ there is a collection $\gamma \subseteq \eta$ such that $A = \cup\{B \mid B \in \gamma\}$. Then,

$T^{-n}A = T^{-n}(\cup\{B \mid B \in \gamma\})$

$= \cup\{T^{-n}B \mid B \in \gamma\} = \cup\{T^{-n}B \mid T^{-n}B \in T^{-n}\gamma\}$

$\therefore \quad T^{-n}\xi \leq T^{-n}\eta$ .

iii) Let $T^{-n}A \in T^{-n}\xi$, $T^{-n}B \in T^{-n}\eta$. Since $\xi \overset{0}{\leq} \eta$, $\mu(A \cap B) = 0$ or $\mu(B)$.

Since $T$ is m.p.

$$\mu(T^{-n}A \cap T^{-n}B) = \mu(T^{-n}(A \cap B)) = \mu(A \cap B)$$

$$= 0 \quad \text{or} \quad \mu(B) .$$

But $\qquad \mu(B) = \mu(T^{-n}B)$

So $\mu(T^{-n}A \cap T^{-n}B) = 0$ or $\mu(T^{-n}B)$.

$\therefore \quad T^{-n}\xi \overset{0}{\leq} T^{-n}\eta$.

Definition

Let $\xi$ be a partition of $(X, \mathcal{B}, \mu)$. The entropy of $\xi$, denoted $H(\xi)$ is given by:

$$H(\xi) = -\sum_{A \in \xi} \mu(A) \log \mu(A) .$$

(Log to the base 2 will be used and define $0 \log 0 = 0$).

The more salient properties of the entropy map $H$ are clear from the following proposition.

2.2 Proposition

Let $p = (p_1, \ldots, p_k)$ be a probability vector and define $H^*(p) = -\sum_{i=1}^{k} p_i \log p_i$. Then:

i) $H^*(p) \geq 0$

ii) $H^*(p) \leq \log k$ with equality holding only when $p_i = \frac{1}{k}$ for every $i$.

And if $p' = (p_1, p_2, \ldots, p_\ell - \varepsilon, \ldots, p_m + \varepsilon, \ldots, p_k)$ where $0 < \varepsilon < p_\ell \leq \frac{1}{k} \leq p_m$

then:

iii) $H^*(p') < H^*(p)$

Proof:

i) Let $\phi(x) = \begin{cases} x \log x & x \in (0, \infty) \\ 0 & x = 0 \end{cases}$

$$\phi(x) \leq 0 \quad \text{for every} \quad x \in [0,1] .$$

Since, for every $i$, $0 \leq p_i \leq 1$ it follows that $H^*(p) \geq 0$.

ii) First we show that $\phi$ is convex. That is,

$\phi(\alpha x+\beta y) \leq \alpha\phi(x)+\beta\phi(y)$ for $x,y \in [0,\infty)$, $\alpha, \beta \geq 0$, $\alpha+\beta=1$.

Let $\alpha, \beta > 0$ be fixed with sum 1.

$$\phi'(x) = 1 + \log x \ .$$

$$\phi''(x) = \frac{1}{x} > 0 \quad \text{on} \quad (0,\infty)$$

Suppose $y > x$. Then $x < \alpha x + \beta y < y$. By the Mean Value Theorem there is some $z \in (\alpha x+\beta y, y)$ such that $\frac{\phi(y)-\phi(\alpha x+\beta y)}{y-(\alpha x+\beta y)} = \phi'(z)$. Now, $\alpha = 1-\beta$. So, $y-(\alpha x+\beta y) = y-\beta y-\alpha x = \alpha(y-x)$. Thus $\frac{\phi(y)-\phi(\alpha x+\beta y)}{\alpha(y-x)} = \phi'(z)$. Again by the Mean Value Theorem there is some $w \in (x, \alpha x+\beta y)$ such that $\frac{\phi(\alpha x+\beta y)-\phi(x)}{(\alpha x+\beta y)-x} = \phi'(w)$. Now, $\beta = 1-\alpha$. So, $\alpha x+\beta y-x = \beta(y-x)$. Thus, $\frac{\phi(\alpha x+\beta y)-\phi(x)}{\beta(y-x)} = \phi'(w)$. Consequently, $\phi(\alpha x+\beta y)-\phi(x) = \phi'(w)\,\beta(y-x)$. $\phi'' > 0$ on $(0,\infty)$, $w < z$, thus $\phi'(w) < \phi'(z)$. Therefore, $\frac{\phi(\alpha x+\beta y)-\phi(x)}{\beta(y-x)} < \frac{\phi(y)-\phi(\alpha x+\beta y)}{\alpha(y-x)}$. Since $y-x > 0$ we have $\alpha[\phi(\alpha x+\beta y)-\phi(x)] < \beta[\phi(y)-\phi(\alpha x+\beta y)]$ which gives $\alpha\phi(\alpha x+\beta y)+\beta\phi(\alpha x+\beta y) < \alpha\phi(x)+\beta\phi(y)$. We have $\alpha+\beta=1$. So, for $x,y > 0$, $\phi(\alpha x+\beta y) < \alpha\phi(x)+\beta\phi(y)$. This also holds for $x,y \geq 0$, $x \neq y$, since $\log\alpha < 0$ and $\phi(\alpha x) = \alpha x\log(\alpha x) = \alpha x\log\alpha + \alpha x\log x < \alpha\phi(x)$.

Set $\phi(\alpha x+\beta y) = \alpha\phi(x)+\beta\phi(y)$. To derive strict inequality we assumed that $x \neq y$ and $\alpha, \beta > 0$. So, one of these must be false. Therefore we have equality only when $x = y$ or $\alpha = 0$ or $\beta = 0$.

Now, we may extend this result by induction to give $\phi(\sum_{i=1}^{k}\alpha_i x_i) \leq \sum_{i=1}^{k}\alpha_i\phi(x_i)$ where $x_i \in [0,\infty)$, $\alpha_i \geq 0$, $\sum_{i=1}^{k}\alpha_i = 1$, and equality holds only when all $x_i$'s with non-zero coefficients are equal.

Suppose $\phi(\sum_{i=1}^{k-1}\beta_i x_i) \leq \sum_{i=1}^{k-1}\beta_i\phi(x_i)$ where $x_i \geq 0$, $\beta_i \geq 0$, $\sum_{i=1}^{k-1}\beta_i = 1$ and equality holds only when $\beta_i, \beta_j \neq 0$ implies $x_i = x_j$.

Now, consider the case where $\alpha_n > 0$, $\sum_{i=1}^{k} \alpha_i = 1$, and $x_i \geq 0$ for $i = 1,2,\ldots,k$. Let $\alpha = \sum_{i=1}^{k-1} \alpha_i$. Then $\alpha + \alpha_k = 1$ and $\sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha} = 1$. We may assume that $\alpha \neq 0$ and $\alpha_k \neq 0$ otherwise the situation is trivial.

Let $y = \frac{1}{\alpha} \sum_{i=1}^{k-1} \alpha_i x_i$. Then, $\phi(\sum_{i=1}^{k} \alpha_i x_i) = \phi(\alpha y + \alpha_k x_k)$. Now, since $\alpha + \alpha_k = 1$ we have by the previous result $\phi(\alpha y + \alpha_k x_k)$ $\leq \alpha\, \phi(y) + \alpha_k \phi(x_k) = \alpha\phi(\sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha} x_i) + \alpha_k \phi(x_k)$. By the induction hypothesis

$$\phi(\sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha} x_i) \leq \sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha} \phi(x_i) .$$

So, we have $\phi(\sum_{i=1}^{k} \alpha_i x_i) \leq \sum_{i=1}^{k} \alpha_i \phi(x_i)$. Now, set

$$\phi(\sum_{i=1}^{k} \alpha_i x_i) = \sum_{i=1}^{k} \alpha_i \phi(x_i). \text{ Then}$$

$$\phi(y) = \sum_{i=1}^{k} \frac{\alpha_i}{\alpha} \phi(x_i) \quad (*)$$

and $\phi(\alpha y + \alpha_k x_k) = \alpha\phi(y) + \alpha_k \phi(x_k)$. Since $\alpha \neq 0$ and $\alpha_k \neq 0$ it follows that $y = x_k$. Because of (*) and the induction hypothesis $\alpha_i, \alpha_j \neq 0$ implies $x_i = x_j$ for each pair $i,j \in \{1,2,\ldots,k-1\}$. Let $\bar{x} = x_j$ for any $j$ s.t. $\alpha_j \neq 0$. Then

$$x_k = y = \sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha} x_i = \bar{x} \sum_{i=1}^{k-1} \frac{\alpha_i}{\alpha} = \bar{x} .$$

So, the desired result follows.

Now, if we let $\alpha_i = \frac{1}{k}$ for all $i$ and $x_i = p_i$ for the probability vector $(p_1, p_2, \ldots, p_k)$ we get:

$$\phi(\frac{1}{k} \sum_{i=1}^{k} p_i) \leq \sum_{i=1}^{k} \frac{1}{k} \phi(p_i) .$$

$$\text{LHS} = \frac{1}{k}\left(\sum_{i=1}^{k} p_i\right)\log\left(\frac{1}{k}\sum_{i=1}^{k} p_i\right)$$

$$= \frac{1}{k}(\log 1 - \log k) = -\frac{\log k}{k}$$

$$\text{RHS} = \frac{1}{k}\sum_{i=1}^{k}\phi(p_i) = \frac{1}{k}\sum_{i=1}^{k} p_i \log p_i$$

$$= -\frac{1}{k} H^*(p_1, p_2, \ldots, p_k) .$$

Thus
$$-\frac{\log k}{k} \leq -\frac{1}{k} H^*(p_1, \ldots, p_k) .$$

$$\therefore \quad H^*(p_1, \ldots, p_k) \leq \log k$$

and $H^*(p_1,\ldots,p_k) = \log k$ only when $p_i = p_j$ for all pairs $i,j$ since $\alpha_i = \frac{1}{k} > 0$ for every every $i$. Thus $H^*(p_1,\ldots,p_k) = \log k$ only when $p_i = \frac{1}{k}$ for all $i$.

iii) Suppose $p_\ell \leq \frac{1}{k}$ and $p_m \geq \frac{1}{k}$. Such pairs always exist since $\sum_{i=1}^{k} p_i = 1$. Choose $\varepsilon > 0$ s.t. $p_\ell > \varepsilon > 0$. Then

$$-H^*(p_1,\ldots,p_k) = \sum_{i=1}^{k}\phi(p_i) \quad \text{and}$$

$$-H^*(p_1,\ldots,p_\ell - \varepsilon, p_m + \varepsilon, \ldots, p_k)$$

$$= \sum_{i=1}^{\ell-1}\phi(p_i) + \sum_{i=\ell+1}^{m-1}\phi(p_i) + \sum_{i=m+1}^{k}\phi(p_i)$$

$$+ \phi(p_\ell - \varepsilon) + \phi(p_m + \varepsilon) .$$

It is sufficient to show that $\phi(p_\ell) + \phi(p_m) < \phi(p_\ell - \varepsilon) + \phi(p_m + \varepsilon)$. By the Mean Value Theorem there is some $z \in (p_\ell - \varepsilon, p_\ell)$ such that

$$\frac{\phi(p_\ell) - \phi(p_\ell - \varepsilon)}{\varepsilon} = \phi'(z)$$

and there is some $w \in (p_m, p_m + \varepsilon)$ such that

$$\frac{\phi(p_m + \varepsilon) - \phi(p_m)}{\varepsilon} = \phi'(\omega) .$$

Now, $\phi''(x) = \frac{1}{x} > 0$ for $x \in (0,\infty)$. Since $p_\ell < p_m$ we have $z < w$. So, $\phi'(z) < \phi'(w)$. Thus $\phi(p_\ell) - \phi(p_\ell - \varepsilon) < \phi(p_m + \varepsilon) - \phi(p_m)$ and $\phi(p_\ell) + \phi(p_m) < \phi(p_\ell - \varepsilon) + \phi(p_m + \varepsilon)$.

$\therefore$ $H^*(p_1,\ldots,p_\ell - \varepsilon,\ldots,p_m + \varepsilon,\ldots,p_k) < H^*(p_1,\ldots,p_k)$.

It is clear from 2.2(iii) that for partitions of a given finite cardinality, the entropy map will give relatively lesser values to those partitions with relatively greater variation in the measure of their members.

Section 3. <u>The Entropy of a Measure Preserving Transformation</u>

We are now in a position to give a definition of the entropy of a map $T$ with respect to a partition $\xi$ and use this definition to define the entropy of $T$.

Denote the join product $\xi \vee T^{-1}\xi \vee \ldots \vee T^{-(n-1)}\xi$ by $\bigvee_{i=0}^{n-1} T^{-i}\xi$

<u>Definition</u>

Let $\xi$ be a partition of $(X, \mathcal{B}, \mu)$ such that $H(\xi) < \infty$ and $T:X\to X$ a measure preserving transformation. Then the <u>entropy of $T$ with respect to $\xi$</u>, denoted $h(T,\xi)$, is given by:

$$h(T,\xi) = \overline{\lim_{n\to\infty}} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$$

It will be shown subsequently that the limit superior is actually a limit and that $h(T,\xi)$ is always finite.

<u>Definition</u>

Let $T:X\to X$ be measure preserving. The <u>entropy</u> of $T$, denoted $h(T)$, is given by: $h(T) = \sup\{h(T,\xi) \mid H(\xi) < \infty\}$.

It will be shown in what follows that (1) entropy is invariant for isomorphism, (2) the supremum over all partitions of finite entropy equals the supremum over all finite partitions and (3) for some maps there is a finite partition $\xi$ s.t. $h(T) = h(T,\xi)$.

Firstly, however, it is necessary to develop further the theory of entropy on partitions.

## Section 4. Conditional Entropy

Let $\alpha,\beta$ be two partitions (not necessarily finite) of a probability space $(X, \mathcal{B}, \mu)$. For any $A,B \in \mathcal{B}$ with $\mu(B) > 0$ let $\frac{\mu(A \cap B)}{\mu(B)}$ be denoted by $\mu(A|B)$.

### Definition

The entropy of $\alpha$ given $\beta$ denoted $H(\alpha|\beta)$ is given by:

$$H(\alpha|\beta) = - \sum_{B \in \beta} \mu(B) \sum_{A \in \alpha} \mu(A|B) \log \mu(A|B) \ .$$

(members of $\beta$ with zero measure are excluded from the summation).

### 2.3 Proposition

Let $\alpha ; \gamma , \Delta$ be partitions of $(X, \mathcal{B}, \mu)$. Then:

i) $H(\alpha/\gamma) \geq 0$

ii) $\alpha \overset{0}{=} \Delta \Rightarrow H(\alpha/\gamma) = H(\Delta/\gamma)$

iii) $\gamma \overset{0}{=} \Delta \Rightarrow H(\alpha/\gamma) \quad H(\alpha/\Delta)$ .

Proof:

i) is obvious.

ii) Let $\alpha = \{A_i\}_{i \in I}$ (I is at most countable)

$\Delta = \{D_i\}_{i \in I}$ .

Since $\alpha \overset{0}{=} \Delta$ we may assume without loss of generality that $\mu(A_i \Delta D_i) = 0$ for each $i$ . Thus, for any $Q \in \mathcal{B}$ and any $i \in I$ $\mu(A_i \cap Q) = \mu(D_i \cap Q)$ .

Let $\gamma = \{C_j\}_{j \in J}$ (J is at most countable).

$$H(\alpha/\gamma) = -\sum_j \mu(C_j) \sum_i \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$$

$$= -\sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$$

$$= -\sum_{i,j} \mu(A_i \cap D_i \cap C_j) \log \frac{\mu(A_i \cap D_i \cap C_j)}{\mu(C_j)}$$

$$= -\sum_{i,j} \mu(D_i \cap C_i) \log \frac{\mu(D_i \cap C_j)}{\mu(C_j)} = H(\Delta/\gamma) \ .$$

iii) Let $\gamma = \{C_j\}_{j \in J}$

$\Delta = \{D_j\}_{j \in J}$

We may assume that $\mu(C_j \Delta D_j) = 0$ for each $j \in J$.

Let $\alpha = \{A_i\}_{i \in I}$.

$$H(\alpha/\gamma) = -\sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)}$$

$$= -\sum_{i,j} \mu(A_i \cap C_j \cap D_j) \log \frac{\mu(A_i \cap C_j \cap D_j)}{\mu(C_j \cap D_j)}$$

$$= -\sum_{i,j} \mu(A_i \cap D_j) \log \frac{\mu(A_i \cap D_j)}{\mu(D_j)}$$

$$= H(\alpha/\Delta) \ . \qquad \square$$

2.4 Proposition

Let $\alpha,\gamma,\Delta$ be partitions of $(X, \mathcal{B}, \mu)$ and $T:X\to X$ a measure preserving transformation. Then:

i) $H(\alpha \vee \gamma/\Delta) = H(\alpha/\Delta) + H(\gamma/\alpha \vee \Delta)$ .

ii) $H(\alpha \vee \gamma) = H(\alpha) + H(\gamma/\alpha)$

iii) $\alpha \overset{\circ}{\leq} \gamma \Rightarrow H(\alpha/\Delta) \leq H(\gamma/\Delta)$

iv) $\alpha \overset{\circ}{\leq} \gamma \Rightarrow H(\alpha) \leq H(\gamma)$

v) $\gamma \overset{\circ}{\leq} \Delta \Rightarrow H(\alpha/\gamma) \geq H(\alpha/\Delta)$

vi) $H(\alpha) \geq H(\alpha/\Delta)$

vii) $H(\alpha \vee \gamma/\Delta) \leq H(\alpha/\Delta) + H(\gamma/\Delta)$

viii) $H(\alpha \vee \gamma) \leq H(\alpha) + H(\gamma)$

ix) $H(T^{-1}\alpha/T^{-1}\gamma) = H(\alpha/\gamma)$

x) $H(T^{-1}\alpha) = H(\alpha)$

Proof:

Let $\alpha = \{A_i\}_{i \in I}$, $\gamma = \{C_j\}_{j \in J}$, $\Delta = \{D_k\}_{k \in K}$. (I,J,and K are at most countable).

i) $H(\alpha \vee \gamma/\Delta) = -\sum_{i,j,k} \mu(A_i \cap C_j \cap D_k) \log \dfrac{\mu(A_i \cap C_j \cap D_k)}{\mu(D_k)}$ and

$$\frac{\mu(A_i \cap C_j \cap D_k)}{\mu(D_k)} = \frac{\mu(A_i \cap C_j \cap D_k)}{\mu(A_i \cap D_k)} \; \frac{\mu(A_i \cap D_k)}{\mu(D_k)}$$

So, $\log \dfrac{\mu(A_i \cap C_j \cap D_k)}{\mu(D_k)} = \log \dfrac{\mu(A_i \cap C_j \cap D_k)}{\mu(A_i \cap D_k)} + \log \dfrac{\mu(A_i \cap D_k)}{\mu(D_k)}$ . Thus,

$$H(\alpha \vee \gamma/\Delta) = -\sum_{i,j,k} \mu(A_i \cap C_j \cap D_k)\left(\log \frac{\mu(A_i \cap C_j \cap D_k)}{\mu(A_i \cap D_k)} + \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)}\right)$$

$$= H(\gamma/\alpha \vee \Delta) - \sum_{i,j,k} \mu(A_i \cap C_j \cap D_k) \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)}$$

Summing over j gives:

$$H(\alpha \vee \gamma/\Delta) = H(\gamma/\alpha \vee \Delta) - \sum_{i,k} \mu(A_i \cap D_k) \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)}$$

$$= H(\gamma/\alpha \vee \Delta) + H(\alpha/\Delta)$$

ii) let $\eta = \{X\}$ and use (i). By (1) $H(\alpha \vee \gamma/\eta) = H(\alpha/\eta) + H(\gamma/\alpha \vee \eta)$

LHS: $H(\alpha \vee \gamma/\eta) = -\sum_{i,j} \mu(A_i \cap C_j \cap X) \log \dfrac{\mu(A_i \cap C_j \cap X)}{\mu(X)}$

$$= -\sum_{i,j} \mu(A_i \cap C_j) \log \mu(A_i \cap C_j) = H(\alpha \vee \gamma).$$

RHS: $H(\alpha/\eta) + H(\gamma/\alpha \vee \eta) = -\sum_i \mu(A_i \cap X) \log \frac{\mu(A_i \cap X)}{\mu(X)}$

$$-\sum_{i,j} \mu(C_j \cap A_i \cap X) \log \frac{\mu(C_j \cap A_i \cap X)}{\mu(A_i \cap X)}$$

$$= H(\alpha) + H(\gamma/\alpha)$$

$$\therefore \quad H(\alpha \vee \gamma) = H(\alpha) + H(\gamma/\alpha) .$$

iii) $\alpha \overset{0}{\leq} \gamma \Rightarrow \gamma \overset{0}{=} \alpha \vee \gamma$ . So, $H(\gamma/\Delta) = H(\alpha \vee \gamma/\Delta)$ by Prop. 2.3(ii).

By (i) $H(\alpha \vee \gamma/\Delta) = H(\alpha/\Delta) + H(\gamma/\alpha \vee \Delta)$ . Thus

$$H(\gamma/\Delta) = H(\alpha/\Delta) + H(\gamma/\alpha \vee \Delta) .$$

But $\quad H(\gamma/\alpha \vee \Delta) \geq 0$

So $\quad H(\alpha/\Delta) \leq H(\gamma/\Delta)$ .

iv) Let $\eta = \{X\}$. By (iii) $\alpha \overset{0}{\leq} \gamma \Rightarrow H(\alpha/\eta) \leq H(\gamma/\eta)$. But $H(\alpha/\eta) = H(\alpha)$ and $H(\gamma/\eta) = H(\gamma)$. $\therefore$ $H(\alpha) \leq H(\gamma)$ .

v) Fix $i,j$ and let

$$\alpha_k = \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \qquad x_k = \frac{\mu(A_i \cap D_k)}{\mu(D_k)}$$

Terms where $\mu(C_j) = 0$ or $\mu(D_k) = 0$ are considered to be zero since they are excluded in the summation for conditional entropy. Clearly

$$\sum_k \alpha_k = \frac{\mu(C_j)}{\mu(C_j)} = 1 \qquad \text{and}$$

$$\sum_k x_k = \sum_k \mu(A_i | D_k) = 1 \quad \text{since } (X, \mathcal{B}, \mu) \text{ is a probability space.}$$

Let $$\phi(x) = \begin{cases} x \log x & x \in (0,\infty) \\ 0 & x = 0 . \end{cases}$$

$$\phi\left(\sum_{k=1}^{n} \alpha_k x_k + \left(\sum_{k=n+1}^{\infty} \alpha_k\right)\left(\sum_{k=n+1}^{\infty} x_k\right)\right) \leq$$

$$\sum_{k=1}^{n} \alpha_k \phi(x_k) + \left(\sum_{k=n+1}^{\infty} \alpha_k\right) \phi\left(\sum_{k=n+1}^{\infty} x_k\right) \quad \text{for all } n \geq 1$$

(See proof of 2.2(ii)).

Since $\sum_{k=1}^{\infty} \alpha_k = \sum_{k=1}^{\infty} x_k = 1$, $(\sum_{k=n+1}^{\infty} \alpha_k) \to 0$ as $n \to \infty$ and $(\sum_{k=n+1}^{\infty} x_k) \to 0$ as $n \to \infty$. Therefore,

$$\sum_{k=1}^{n} \alpha_k x_k + (\sum_{k=n+1}^{\infty} \alpha_k)(\sum_{k=n+1}^{\infty} x_k) \to \sum_{k=1}^{\infty} \alpha_k x_k \quad \text{as } n \to \infty .$$

Since $\phi$ is continuous,

$$\lim_{n\to\infty} \phi\,(\sum_{k=1}^{\infty} \alpha_k x_k + (\sum_{k=n+1}^{\infty} \alpha_k)(\sum_{k=n+1}^{\infty} x_k)) = \phi(\sum_{k=1}^{\infty} \alpha_k x_k) .$$

And, clearly $\sum_{k=1}^{n} \alpha_k \phi(x_k) + (\sum_{k=n+1}^{\infty} \alpha_k)\phi(\sum_{k=n+1}^{\infty} x_k) \to \sum_{k=1}^{\infty} \alpha_k \phi(x_k)$ since

$$(\sum_{k=n+1}^{\infty} \alpha_k) \to 0 \quad \text{and} \quad \phi(\sum_{k=n+1}^{\infty} x_k) \to 0 \quad \text{as } n \to \infty .$$

Therefore $\phi(\sum_{k=1}^{\infty} \alpha_k x_k) \leq \sum_{k=1}^{\infty} \alpha_k \phi(x_k)$. Thus,

$$\phi(\sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \frac{\mu(A_i \cap D_k)}{\mu(D_k)}) \leq \sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \phi\left(\frac{\mu(A_i \cap D_k)}{\mu(D_k)}\right) . \text{ And,}$$

$$\gamma \overset{0}{\leq} \Delta \Rightarrow \sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \frac{\mu(A_i \cap D_k)}{\mu(D_k)} = \frac{\mu(A_i \cap C_j)}{\mu(C_j)} . \text{ So,}$$

$$\phi(\sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \frac{\mu(A_i \cap D_k)}{\mu(D_k)}) = \phi\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right) = \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} . \text{ Then}$$

$$\frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \leq \sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \frac{\mu(A_i \cap D_k)}{\mu(D_k)} \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)} . \text{ Then}$$

$$\mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \leq \sum_k \mu(D_k \cap C_j) \frac{\mu(A_i \cap D_k)}{\mu(D_k)} \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)}$$

$$= \sum_k \mu(A_i \cap D_k) \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)} , \quad \text{since} \quad \gamma \overset{0}{\leq} \Delta .$$

Summing over $i,j$ gives

$$\sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \leq \sum_{i,j} \mu(A_i \cap D_k) \log \frac{\mu(A_i \cap D_k)}{\mu(D_k)} .$$

Now, LHS $= -H(\alpha/\gamma)$ and RHS $= -H(\alpha/\Delta)$. Thus $-H(\alpha/\gamma) \leq -H(\alpha/\Delta)$

$\therefore$ $H(\alpha/\Delta) \leq H(\alpha/\gamma)$ .

vi) Let $\eta = \{X\}$ . Then $\eta \leq \Delta$ and using (v) gives:

$$H(\alpha/\Delta) \leq H(\alpha/\eta) = H(\alpha) \quad \therefore \quad H(\alpha/\Delta) \leq H(\alpha) .$$

vii) $H(\alpha \vee \gamma/\Delta) = H(\alpha/\Delta) + H(\gamma/\alpha \vee \Delta)$ by (i) . $\Delta \leq \alpha \vee \Delta$ so, by (v):

$$H(\gamma/\alpha \vee \Delta) \leq H(\gamma/\Delta) \quad \therefore \quad H(\alpha \vee \gamma/\Delta) \leq H(\alpha/\Delta) + H(\gamma/\Delta)$$

viii) Let $\eta = \{X\}$ . $H(\alpha \vee \gamma/\eta) \leq H(\alpha/\eta) + H(\gamma/\eta)$ by (vii).

LHS = $H(\alpha \vee \gamma)$ RHS = $H(\alpha) + H(\gamma)$ . $\therefore$ $H(\alpha \vee \gamma) \leq H(\alpha) + H(\gamma)$

ix) $$\mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} = \mu(T^{-1}(A_i \cap C_j)) \log \frac{\mu(T^{-1}(A_i \cap C_j))}{\mu(T^{-1}C_j)}$$

$$= \mu(T^{-1}A_i \cap T^{-1}C_j) \log \frac{\mu(T^{-1}A_i \cap T^{-1}C_j)}{\mu(T^{-1}C_j)}$$

$\therefore$ $H(\alpha/\gamma) = H(T^{-1}\alpha/T^{-1}\gamma)$ .

x) $\mu(A_i) \log \mu(A_i) = \mu(T^{-1}A_i) \log \mu(T^{-1}A_i)$ $\therefore$ $H(\alpha) = H(T^{-1}\alpha)$ .

We are now able to show that $\overline{\lim_{n\to\infty}} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$ is actually a limit and thus $h(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$ .

2.5 <u>Proposition</u>

If $\{a_n\}_{n=1}^{\infty}$ is a sequence in $\mathbb{R}$ such that $0 \leq a_{n+p} \leq a_n + a_p$ for all pairs $n,p \in \mathbb{N}$ then $\lim_{n\to\infty} \frac{a_n}{n}$ exists and equals $\inf\{\frac{a_n}{n} \mid n \geq 1\}$.

<u>Proof</u>:

Fix $p > 0$ . Each $n > 0$ can be written as $n = kp + i$ with $k \geq 0$ and $0 \leq i < p$ . Then

$$\frac{a_n}{n} = \frac{a_{i+kp}}{i+kp} \leq \frac{a_i}{kp} + \frac{a_{kp}}{kp} . \quad a_{kp} \leq ka_p \quad \text{so,}$$

$$\frac{a_n}{n} \leq \frac{a_i}{kp} + \frac{ka_p}{kp} = \frac{a_i}{kp} + \frac{a_p}{p}$$ . Let $n \to \infty$ . Then $k \to \infty$ since $n = kp + i$ and $p$ is fixed. Thus $\frac{a_i}{kp} \to 0$ . Therefore, $\overline{\lim_{n\to\infty}} \frac{a_n}{n} \leq \frac{a_p}{p}$ .

This is true for every $p > 0$, so

$$\overline{\lim_{n\to\infty}} \frac{a_n}{n} \leq \inf \{\frac{a_p}{p} \mid p \geq 1\} \quad . \qquad (1)$$

And the following must hold:

$$\inf \{\frac{a_n}{n} \mid n \geq 1\} \leq \underline{\lim_{n\to\infty}} \frac{a_n}{n} \quad . \qquad (2)$$

From (1) and (2) the proposition follows. □

Corollary

Let $T:X\to X$ be measure preserving and $\xi$ a partition of $(X, \mathcal{B}, \mu)$ with $H(\xi) < \infty$. Then $h(T,\xi) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$ .

Proof:

Set $\frac{a_n}{n} = \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$ and show that $a_{k+p} \leq a_k + a_p$ by means of 2.4(viii) and (x). Then apply 2.5. □

2.6 Proposition

If $\xi$ is a partition of $(X, \mathcal{B}, \mu)$ with $H(\xi) < \infty$ and $T:X\to X$ is measure preserving then:

$$h(T,\xi) \leq \frac{1}{n+1} H(\bigvee_{i=0}^{n} T^{-i}\xi) \leq \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$$

$$\leq H(\xi) \quad \text{for all} \quad n \geq 1 \ .$$

Thus $h(T,\xi)$ is finite.

Proof:

First we shall show by induction that for every $n > 0$

$$H(\bigvee_{i=0}^{n-1} T^{-i}\xi) = H(\xi) + \sum_{j=1}^{n-1} H(\xi / \bigvee_{i=1}^{j} T^{-i}\xi) \ .$$

For $n = 1$ it is clear. Assume

$$H(\bigvee_{i=0}^{p-1} T^{-i}\xi) = H(\xi) + \sum_{j=1}^{p-1} H(\xi / \bigvee_{i=1}^{j} T^{-i}\xi)$$

for $p \geq 1$ .

$$H(\bigvee_{i=0}^{p} T^{-i}\xi) = H(\bigvee_{i=1}^{p} T^{-i}\xi \vee \xi)$$

$$= H(\bigvee_{i=1}^{p} T^{-i}\xi) + H(\xi / \bigvee_{i=1}^{p} T^{-i}\xi) \qquad \text{by 2.4(ii).}$$

$$H(\bigvee_{i=1}^{p} T^{-i}\xi) = H(T^{-1}\bigvee_{i=0}^{p-1} T^{-i}\xi)$$

$$= H(\bigvee_{i=0}^{p-1} T^{-i}\xi) \qquad \text{by 2.4(x).}$$

$$H(\bigvee_{i=0}^{p-1} T^{-i}\xi) = H(\xi) + \sum_{j=1}^{p-1} H(\xi / \bigvee_{i=1}^{j} T^{-i}\xi)$$

by induction hypothesis. Thus,

$$H(\bigvee_{i=0}^{p} T^{-i}\xi) = H(\xi) + \sum_{j=1}^{p-1} H(\xi / \bigvee_{i=1}^{j} T^{-i}\xi)$$

$$+ H(\xi / \bigvee_{i=1}^{p} T^{-i}\xi)$$

$$= H(\xi) + \sum_{j=1}^{p} H(\xi / \bigvee_{i=1}^{j} T^{-i}\xi) \quad . \qquad \text{Therefore}$$

$$H(\bigvee_{i=0}^{n-1} T^{-i}\xi) = H(\xi) + \sum_{j=1}^{n-1} H(\xi / \bigvee_{i=1}^{j} T^{-i}\xi) \qquad \text{for all } n \geq 1 .$$

By 2.4(iii)

$$H(\xi / \bigvee_{i=0}^{n} T^{-i}\xi) \leq H(\xi / \bigvee_{i=1}^{j} T^{-i}\xi) \qquad \text{for } j \leq n$$

and $H(\xi / \bigvee_{i=0}^{n} T^{-i}\xi) \leq H(\xi)$ .

Thus $n\,H(\xi / \bigvee_{i=1}^{n} T^{-i}\xi) \leq H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$ . And we have,

$$H(\bigvee_{i=0}^{n} T^{-i}\xi) = H(\xi \vee \bigvee_{i=1}^{n} T^{-i}\xi)$$

$$= H(\bigvee_{i=1}^{n} T^{-i}\xi) + H(\xi / \bigvee_{i=1}^{n} T^{-i}\xi) \qquad \text{by 2.4(ii).}$$

$$= H(T^{-1}\bigvee_{i=0}^{n-1} T^{-i}\xi) + H(\xi / \bigvee_{i=1}^{n} T^{-i}\xi)$$

$$= H(\bigvee_{i=0}^{n-1} T^{-i}\xi) + H(\xi / \bigvee_{i=1}^{n} T^{-1}\xi) \qquad \text{by 2.4(x).}$$

Thus $\quad n\, H(\bigvee_{i=0}^{n} T^{-i}\xi) = n(H(\bigvee_{i=0}^{n-1} T^{-i}\xi) + H(\xi/\bigvee_{i=1}^{n} T^{-i}\xi))$

$$\leq nH(\bigvee_{i=0}^{n-1} T^{-i}\xi) + H(\bigvee_{i=0}^{n-1} T^{-i}\xi)$$

$$= (n+1)\, H(\bigvee_{i=0}^{n-1} T^{-i}\xi) \ .$$ Therefore,

$$\frac{1}{n+1} H(\bigvee_{i=0}^{n} T^{-i}\xi) \leq \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi) \quad \text{for all } n \geq 1 \ .$$

$$\therefore \quad h(T,\xi) \leq \frac{1}{n+1} H(\bigvee_{i=0}^{n} T^{-i}\xi)$$

$$\leq \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\xi) \leq H(\xi) < \infty \ . \qquad \square$$

Section 5. Properties of $h(T,\xi)$ and $h(T)$

In this section some basic properties of entropy will be established. This will include showing that entropy is invariant with respect to isomorphism.

2.7 Proposition

Suppose $\alpha$ and $\beta$ are partitions of $(X, \mathcal{B}, \mu)$ with $H(\alpha) < \infty$ and $H(\beta) < \infty$ . And suppose $T{:}X\to X$ is measure preserving. Then:

i) $h(T, \alpha \vee \beta) \leq h(T, \alpha) + h(T, \beta)$

ii) $\alpha \overset{0}{\leq} \beta \Rightarrow h(T, \alpha) \leq h(T, \beta)$

iii) $h(T, \alpha) \leq h(T, \beta) + H(\alpha/\beta)$

iv) $h(T, T^{-1}\alpha) = h(T, \alpha)$.

v) If $k \geq 1$ then $h(T, \alpha) = h(T, \bigvee_{i=0}^{k-1} T^{-i}\alpha)$

vi) If $T$ is invertible and $k \geq 1$ then

$$h(T, \alpha) = h(T, \bigvee_{i=-k}^{k} T^{-i}\alpha) \ .$$

Proof:

i) $H(\bigvee_{i=0}^{n-1} T^{-i}(\alpha \vee \beta)) = H(\bigvee_{i=0}^{n-1} (T^{-i}\alpha \vee T^{-i}\beta))$

$= H(\bigvee_{i=0}^{n-1} T^{-i}\alpha \vee \bigvee_{i=0}^{n-1} T^{-i}\beta)$

$\leq H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) + H(\bigvee_{i=0}^{n-1} T^{-i}\beta)$ by 2.4(viii).

Consequently,

$$\frac{1}{n}H(\bigvee_{i=0}^{n-1} T^{-i}(\alpha \vee \beta)) \leq \frac{1}{n}H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) + \frac{1}{n}H(\bigvee_{i=0}^{n-1} T^{-i}\beta)$$

$\therefore$ $h(T,\alpha \vee \beta) \leq h(T,\alpha) + h(T,\beta)$.

ii) If $\alpha \overset{o}{\leq} \beta$ then $T^{-n}\alpha \overset{o}{\leq} T^{-n}\beta$. Thus

$\bigvee_{i=0}^{n-1} T^{-i}\alpha \overset{o}{\leq} \bigvee_{i=0}^{n-1} T^{-i}\beta$. So,

$H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \leq H(\bigvee_{i=0}^{n-1} T^{-i}\beta)$ by 2.4(iv).

Then $\frac{1}{n}H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \leq \frac{1}{n}H(\bigvee_{i=0}^{n-1} T^{-i}\beta)$

$\therefore$ $h(T,\alpha) \leq h(T,\beta)$

iii) $H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \leq H((\bigvee_{i=0}^{n-1} T^{-i}\alpha) \vee (\bigvee_{i=0}^{n-1} T^{-i}\beta))$ by 2.4(iv).

$H(\bigvee_{i=0}^{n-1}T^{-i}\alpha \vee \bigvee_{i=0}^{n-1}T^{-i}\beta) = H(\bigvee_{i=0}^{n-1}T^{-i}\beta) + H(\bigvee_{i=0}^{n-1}T^{-i}\alpha / \bigvee_{i=0}^{n-1}T^{-i}\beta)$ by 2.4(ii).

$H(\bigvee_{i=0}^{n-1}T^{-i}\alpha / \bigvee_{i=0}^{n-1}T^{-i}\beta) \leq \sum_{i=0}^{n-1} H(T^{-i}\alpha / \bigvee_{j=0}^{n-1} T^{-j}\beta)$ by 2.4(vii).

$\sum_{i=0}^{n-1} H(T^{-i}\alpha / \bigvee_{i=0}^{n-1}T^{-j}\beta) \leq \sum_{i=0}^{n-1} H(T^{-i}\alpha / T^{-i}\beta)$ by 2.4(v).

$\sum_{i=0}^{n-1} H(T^{-i}\alpha / T^{-i}\beta) = n\,H(\alpha/\beta)$ by 2.4(ix).

Consequently,

$$\frac{1}{n}H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \leq \frac{1}{n}H(\bigvee_{i=0}^{n-1} T^{-i}\beta) + H(\alpha/\beta)$$

$\therefore$ $h(T,\alpha) \leq h(T,\beta) + H(\alpha/\beta)$

iv) $$h(T, T^{-1}\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=1}^{n} T^{-i}\alpha)$$

$$H(\bigvee_{i=1}^{n} T^{-i}\alpha) = H(T^{-1}\bigvee_{i=0}^{n-1} T^{-i}\alpha)$$

$$= H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \qquad \text{by } 2.4(x).$$

So, $\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=1}^{n} T^{-i}\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$

$\therefore \ h(T, T^{-1}\alpha) = h(T, \alpha)$ .

v) $$h(T, \bigvee_{i=0}^{k} T^{-i}\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{j=0}^{n-1} T^{-j} \bigvee_{i=0}^{k} T^{-i}\alpha)$$

$$= \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{k+n-1} T^{-i}\alpha)$$

$$= \lim_{n\to\infty} \left(\frac{k+n-1}{n}\right)\left(\frac{1}{k+n-1}\right) H(\bigvee_{i=0}^{k+n-1} T^{-i}\alpha) = h(T, \alpha).$$

vi) $h(T, \bigvee_{i=-k}^{k} T^{-i}\alpha) =$

$$h(T, T^{-k}\bigvee_{i=0}^{2k} T^{-i}\alpha) = h(T, \bigvee_{i=0}^{2k} T^{-i}\alpha) \qquad \text{by } 2.4(x).$$

$$h(T, \bigvee_{i=0}^{2k} T^{-i}\alpha) = h(T, \alpha) \qquad \text{by (v).}$$

2.8 <u>Proposition</u>

Let $T: X \to X$ be measure preserving. Then:

i) For $k = 1, 2, \ldots, h(T^k) = k\,h(T)$.

ii) If $T$ is invertible measure preserving then $h(T^k) = |k|\,h(T)$ for every $k \in Z$.

<u>Proof</u>:

i) Let $\xi$ be a partition of $(X, \mathcal{B}, \mu)$ with $H(\xi) < \infty$.

$$k\,h(T,\xi) = k \lim_{n\to\infty} \frac{1}{nk} H\Big(\bigvee_{i=0}^{nk-1} T^{-i}\xi\Big)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\Big(\bigvee_{j=0}^{n-1} T^{-kj}\big(\bigvee_{i=0}^{k-1} T^{-i}\xi\big)\Big)$$

$$= h\Big(T^k, \bigvee_{i=0}^{k-1} T^{-i}\xi\Big)\,. \qquad (1)$$

(1) implies $k\,h(T) \leq h(T^k)$ . By 2.7(ii)

$$h(T^k,\xi) \leq h\Big(T^k, \bigvee_{i=0}^{k-1} T^{-i}\xi\Big)\,. \qquad (2)$$

(1) and (2) imply $h(T^k,\xi) \leq k\,h(T,\xi)$ . Therefore $h(T^k) \leq k\,h(T)$ and part (i) follows.

ii) Let $\tau = T^{-1}$. $T:X\to X$ is i.m.p. so $\tau:X\to X$ is m.p. For $k > 0$ $h(\tau^k) = k\,h(\tau)$. $\therefore$ $h(T^{-k}) = k\,h(T^{-1})$. For $k < 0$ $h(\tau^k) = h(T^{|k|}) = |k|\,h(T)$ by (i) . So, to prove $h(T^{-k}) = k\,h(T)$ for $k > 0$ we must show that $h(T) = h(T^{-1})$ . Let $\xi$ be a finite partition of $(X, \mathcal{B}, \mu)$.

$$H\Big(\bigvee_{i=0}^{n-1} T^{i}\xi\Big) = H\Big(T^{-(n-1)} \bigvee_{i=0}^{n-1} T^{i}\xi\Big) \qquad \text{by 2.4(x).}$$

$$H\Big(T^{-(n-1)} \bigvee_{i=0}^{n-1} T^{i}\xi\Big) = H\Big(\bigvee_{i=0}^{n-1} T^{-i}\xi\Big)$$

$$\therefore\ \frac{1}{n} H\Big(\bigvee_{i=0}^{n-1} T^{i}\xi\Big) = \frac{1}{n} H\Big(\bigvee_{i=0}^{n-1} T^{-i}\xi\Big).$$

$$\therefore\ h(T^{-1}) = h(T) \qquad \square$$

The following proposition and its corollary give the invariance properties of entropy.

2.9 <u>Proposition</u>

Let $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ be two probability spaces. Let $T_1:X_1\to X_1$ and $T_2:X_2\to X_2$ be measure preserving transformations. If $T_1$ is homormorphic to $T_2$ then $h(T_1) \gtrsim h(T_2)$.

Proof:

Let $\xi$ be a partition of $(X_2, \mathcal{B}_2, \mu_2)$ with $H(\xi) < \infty$.

Since $T_1$ is homomorphic to $T_2$ there is a map $\phi$ such that $(X_1, \mathcal{B}_1, \mu_1) \xrightarrow{\phi} (X_2, \mathcal{B}_2, \mu_2)$ and $\phi \circ T_1 = T_2 \circ \phi$ on the domain of $\phi$. (See Chapter 1). Since $\phi$ is m.p., $\phi^{-1}\xi$ is a partition of $(X_1, \mathcal{B}_1, \mu_1)$ and $H(\phi^{-1}\xi) < \infty$.

Thus, $$h(T_1, \phi^{-1}\xi) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T_1^{-i}(\phi^{-1}\xi)\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} (T_1^{-i} \circ \phi^{-1})\xi\right).$$

Since $\phi \circ T_1 = T_2 \circ \phi$ we have,

$$H\left(\bigvee_{i=0}^{n-1}(T_1^{-i} \circ \phi^{-1})\xi\right) = H\left(\bigvee_{i=0}^{n-1}(\phi^{-1} \circ T_2^{-i})\xi\right)$$

$$= H\left(\bigvee_{i=0}^{n-1} \phi^{-1} \circ T_2^{-i}\,\xi\right) = H\left(\phi^{-1} \bigvee_{i=0}^{n-1} T_2^{-i}\,\xi\right)$$

$$= H\left(\bigvee_{i=0}^{n-1} T_2^{-i}\xi\right) \qquad \text{by 2.4(x).}$$

$\therefore \quad h(T_1, \phi^{-1}\xi) = h(T_2, \xi)$

$\therefore \sup \{h(T_1, \alpha) \mid H(\alpha) \leq \infty\} \geq \sup \{h(T_2, \beta) \mid H(\beta) < \infty\}$

$\therefore \quad h(T_1) \geq h(T_2)$ □

Corollary

If $T_1$ is isomorphic to $T_2$ then $h(T_1) = h(T_2)$.

Proof:

There is a map $\phi$ such that $(X_1, \mathcal{B}_1, \mu_1) \overset{\phi}{\leftrightarrow} (X_2, \mathcal{B}_2, \mu_2)$ and $\phi \circ T_1 = T_2 \circ \phi$. Therefore $T_1$ is homomorphic to $T_2$ and $T_2$ is homomorphic to $T_1$. $\therefore h(T_1) \geq h(T_2)$ and $h(T_2) \geq h(T_1)$.

$\therefore \quad h(T_1) = h(T_2)$. □

Thus entropy can sometimes be used to show that two maps are not isomorphic. For if $h(T_1) \neq h(T_2)$ then $T_1 \not\cong T_2$.

Section 6. The Computation of Entropy

In this section, a proof of the Kolmogoroff-Sinai theorem will be given. This theorem facilitates the computation of entropy for many maps and also shows that for some maps there is a finite partition $\xi$, such that $h(T) = h(T, \xi)$.

Definition

Two partitions $\alpha, \beta$ of $(X, \mathcal{B}, \mu)$ will be considered equivalent $(\alpha \sim \beta)$ iff $\alpha \overset{0}{=} \beta$.

Definition

For two partitions $\alpha, \beta$ of $(X, \mathcal{B}, \mu)$ we define the map $d(\alpha,\beta)$ by $d(\alpha,\beta) = H(\alpha/\beta) + H(\beta/\alpha)$.

To prove that $d$ is a metric requires the following proposition.

2.10 Proposition

Let $\alpha,\beta$ be two partitions of $(X, \mathcal{B}, \mu)$. Then:

$$H(\alpha/\beta) = 0 \text{ iff } \alpha \overset{0}{\leq} \beta$$

Proof:

Suppose $H(\alpha/\beta) = 0$

$$H(\alpha/\beta) = -\sum_{B\in\beta} \mu(B) \sum_{A\in\alpha} \frac{\mu(A\cap B)}{\mu(B)} \log \frac{\mu(A\cap B)}{\mu(B)}$$

If $H(\alpha/\beta) = 0$ then

$\frac{\mu(A\cap B)}{\mu(B)} \log \frac{\mu(A\cap B)}{\mu(B)} = 0$ for every $A \in \alpha$, $B \in \beta$. Then

$$\mu(A\cap B) = 0 \text{ or } \mu(A\cap B) = \mu(B)$$

$$\therefore \quad \alpha \overset{0}{\leq} \beta$$

Suppose $\alpha \overset{0}{\leq} \beta$. Then for every $A \in \alpha$, $B \in \beta$, $\mu(A\cap B) = 0$ or $\mu(A\cap B) = \mu(B)$. Then $\mu(A\cap B) \log \frac{\mu(A\cap B)}{\mu(B)} = 0$. $\therefore$ $H(\alpha/\beta) = 0$. □

2.11 <u>Proposition</u>

The map $d$ on the space of equivalence classes of partitions of $(X, \mathcal{B}, \mu)$ is a metric.

<u>Proof</u>:

1) $(d(\alpha,\beta) \geq 0)$ since $H(\alpha/\beta) \geq 0$ and $H(\beta/\alpha) \geq 0$ , $d(\alpha,\beta) \geq 0$ .

2) $(d(\alpha,\beta) = 0$ iff $\alpha \sim \beta)$. If $d(\alpha,\beta) = 0$ then $H(\alpha/\beta) = H(\beta/\alpha) = 0$. Then $\alpha \overset{0}{\leq} \beta$ and $\beta \overset{0}{\leq} \alpha$ $\therefore$ $\alpha \overset{0}{=} \beta$ $\therefore$ $\alpha \sim \beta$ . If $\alpha \sim \beta$ then $\alpha \overset{0}{=} \beta$ then $H(\alpha/\beta) = H(\beta/\alpha) = 0$ $\therefore$ $d(\alpha,\beta) = 0$ .

3) $(d(\alpha,\beta) = d(\beta,\alpha))$ Since

$$d(\alpha,\beta) = H(\alpha/\beta) + H(\beta/\alpha) = H(\beta/\alpha) + H(\alpha/\beta) = d(\beta,\alpha) .$$

4) $(d(\alpha,\gamma) \leq d(\alpha,\beta) + d(\beta,\gamma))$. Let $\gamma$ be a partition of $(X, \mathcal{B}, \mu)$.

$$H(\alpha/\gamma) = H(\alpha \vee \gamma) - H(\gamma) \qquad \text{by } 2.4(\text{ii}).$$

$$H(\alpha \vee \gamma) \leq H(\alpha \vee \beta \vee \gamma) \qquad \text{by } 2.4(\text{iv}).$$

Thus,
$$\begin{aligned} H(\alpha/\gamma) &\leq H(\alpha \vee \beta \vee \gamma) - H(\beta \vee \gamma) \\ &\quad + H(\beta \vee \gamma) - H(\gamma) \\ &= H(\alpha/\beta \vee \gamma) + H(\beta/\gamma) && \text{by } 2.4(\text{ii}). \\ &\leq H(\alpha/\beta) + H(\beta/\gamma) && \text{by } 2.4(\text{v}). \end{aligned}$$

Similarly, $H(\gamma/\alpha) \leq H(\gamma/\beta) + H(\beta/\alpha)$

Thus,
$$\begin{aligned} d(\alpha,\gamma) &= H(\alpha/\gamma) + H(\gamma/\alpha) \\ &\leq H(\alpha/\beta) + H(\beta/\gamma) + H(\gamma/\beta) + H(\beta/\alpha) \\ &= H(\alpha/\beta) + H(\beta/\alpha) + H(\beta/\gamma) + H(\gamma/\beta) \\ &= d(\alpha,\beta) + d(\beta,\gamma) . \end{aligned}$$

$$\therefore \quad d(\alpha,\gamma) \leq d(\alpha,\beta) + d(\beta,\gamma) . \qquad \square$$

If $T:X \to X$ is measure preserving, we may regard $h(T,\alpha)$ as a map which acts on finite partitions by the rule $\alpha \to h(T,\alpha)$. The following proposition states that this map is continuous, a fact which shall prove useful.

2.12 Proposition

The map $h(T,\cdot)$ on the space of finite parititons of $(X, \mathcal{B}, \mu)$ is continuous.

Proof:

Let $\alpha, \beta$ be finite partitions of $(X, \mathcal{B}, \mu)$. Then,

$h(T,\alpha) - h(T,\beta) \leq H(\alpha/\beta)$ and $h(T,\beta) - h(T,\alpha) \leq H(\beta/\alpha)$ by 2.7(iii).

Thus, $|h(T,\alpha) - h(T,\beta)| \leq \max \{H(\alpha/\beta), H(\beta/\alpha)\} \leq d(\alpha,\beta)$.

$\therefore$ $|h(T,\alpha) - h(T,\beta)| \leq d(\alpha,\beta)$. Continuity of $h(T,\cdot)$ follows. □

Definition

The set of equivalence classes of partitions of $(X, \mathcal{B}, \mu)$ with finite entropy together with the metric $d$ will be denoted by $\mathcal{F}$.

The following proposition shows that we can approximate (in the d-metric sense) any member of $\mathcal{F}$ by using finite partitions.

2.13 Propositions

The finite partitions are dense in $\mathcal{F}$.

Proof:

Let $\alpha = \{A_1, A_2, \ldots\} \in \mathcal{F}$. Write $\alpha_n = \{A_1, A_2, \ldots, A_{n-1}, B_n\}$ where $B_n = X - (A_1 \cup A_2 \cup \ldots \cup A_{n-1})$. Clearly $\alpha_n \leq \alpha$ and so, by 2.10 $H(\alpha_n/\alpha) = 0$. Thus,

$$d(\alpha,\alpha_n) = H(\alpha/\alpha_n) = H(\alpha \vee \alpha_n) - H(\alpha_n) \quad \text{by 2.4(ii)}$$

$$= H(\alpha) - H(\alpha_n) \quad \text{since } \alpha_n \leq \alpha .$$

$$- H(\alpha_n) = \sum_{i=1}^{n-1} \mu(A_i) \log \mu(A_i) + \mu(B_n) \log \mu(B_n)$$

$$+ H(\alpha) = \sum_{i=1}^{n-1} \mu(A_i) \log \mu(A_i) - \sum_{i=n}^{\infty} \mu(A_i) \log \mu(A_i)$$

So, $$d(\alpha,\alpha_n) = \mu(B_n) \log \mu(B_n) - \sum_{i=n}^{\infty} \mu(A_i) \log \mu(A_i)$$

Since $\alpha \in F$, $H(\alpha) < \infty$. Therefore $\sum_{i=n}^{\infty} \mu(A_i) \log \mu(A_i) \to 0$ as $n \to \infty$.

And, since $$\mu(B_n) = 1 - \sum_{i=1}^{n-1} \mu(A_i)$$

$$\mu(B_n) \to 0 \quad \text{as} \quad n \to \infty .$$

Thus $\mu(B_n) \log \mu(B_n) \to 0$ as $n \to \infty$.

$\therefore$ $$d(\alpha_n,\alpha) \to 0 \quad \text{as} \quad n \to \infty . \qquad \square$$

2.14 <u>Proposition</u> ([7])

Let $k \in N$ $\varepsilon > 0$. Then there is a $\delta = \delta(k,\varepsilon)$ such that for any two partitions

$$\alpha = \{A_1, A_2, \ldots, A_k\}$$

$$\beta = \{B_1, B_2, \ldots, B_k\}$$

with $\sum_{i=1}^{k} \mu(A_i \Delta B_i) < \delta$ we have the following:

i) $d(\alpha,\beta) < \varepsilon$

ii) $|H(\alpha) - H(\beta)| < \varepsilon/2$

iii) $|h(T,\alpha) - h(T,\beta)| < \varepsilon/2$

for any $T:X \to X$ m.p.

<u>Proof</u>:

i) Define the partition $\gamma$ by:

$$\gamma = \{A_i \cap B_j \mid i \neq j\} \cup \{\bigcup_{i=1}^{k} A_i \cap B_i\}$$

Then $$\alpha \vee \beta = \gamma \vee \beta$$

Now, $$\sum_{i \neq j} \mu(A_i \cap B_j) = \sum_{i=1}^{k} \left(\sum_{j=1}^{i-1} \mu(A_i \cap B_j) + \sum_{j=i+1}^{k} \mu(A_i \cap B_j)\right)$$

and $$\sum_{j=1}^{i-1} \mu(A_i \cap B_j) + \sum_{j=i+1}^{k} \mu(A_i \cap B_j)$$

$$\leq \mu(A_i - B_i) \leq \mu(A_i \Delta B_i) \quad \text{for each } i .$$

Thus $$\sum_{i \neq j} \mu(A_i \cap B_j) \leq \sum_{i=1}^{k} \mu(A_i \Delta B_i) < \delta$$

So, $\gamma$ has $k(k-1)$ sets of total measure less than $\delta$ and one set of measure greater than $1-\delta$. Thus

$$H(\gamma) \leq -(1-\delta)\log(1-\delta) - \delta \log \frac{\delta}{k(k-1)}$$

Thus if we choose $\delta$ small enough $H(\gamma) < \varepsilon/2$. Now, since

$$H(\beta) + H(\alpha/\beta) = H(\alpha \vee \beta) = H(\beta \vee \gamma)$$
$$\leq H(\beta) + H(\gamma) < H(\beta) + \varepsilon \quad \text{we have}$$

$$H(\beta) + H(\alpha/\beta) < H(\beta) + \varepsilon$$

Thus $\qquad H(\alpha/\beta) < \varepsilon/2.$

And, $\qquad H(\alpha) + H(\beta/\alpha) = H(\alpha \vee \beta) = H(\alpha \vee \gamma)$

$$\leq H(\alpha) + H(\gamma) \leq H(\alpha) + \varepsilon/2.$$

Thus $\qquad H(\beta/\alpha) < \varepsilon/2.$

Therefore $\qquad d(\alpha,\beta) < \varepsilon$

ii) $\qquad H(\alpha/\beta) = H(\alpha \vee \beta) - H(\beta).$

$\qquad H(\beta/\alpha) = H(\alpha \vee \beta) - H(\alpha) \qquad$ by 2.4(ii).

Thus $|H(\alpha/\beta) - H(\beta/\alpha)| = |H(\alpha) - H(\beta)|$

But $|H(\alpha/\beta) - H(\beta/\alpha)| < \varepsilon/2$

$\therefore \qquad |H(\alpha) - H(\beta)| < \varepsilon/2.$

iii) From the proof of 2.12,

$$|h(T,\alpha) - h(T,\beta)| \leq \max\{H(\alpha/\beta), H(\beta/\alpha)\} < \varepsilon/2. \quad \square$$

<u>Definition</u>

Let $\xi$ be a partition of $(X, \mathcal{B}, \mu)$. The algebra formed by taking arbitrary unions of members of $\xi$, denoted $A(\xi)$, will be referred to as the <u>algebra generated</u> by $\xi$.

<u>Definition</u>

Let $A, C$ be sub-algebras of $\mathcal{B}$. Their <u>join</u>, denoted $A \vee C$, is the smallest sub-algebra of $\mathcal{B}$ containing both $\mathcal{A}$ and $\mathcal{C}$. If $\{\mathcal{A}_i\}_{i\in I}$ is an indexed set of algebras then $\bigvee_{i\in I} \mathcal{A}_i$ is the smallest algebra containing every $\mathcal{A}_i$.

The following proposition is an existence theorem which states that finite partitions which are subsets of an algebra which generates $\mathcal{B}$ can be used to approximate (in the d-metric sense) any finite partition of $(X, \mathcal{B}, \mu)$.

2.15 <u>Proposition</u> ([24])

Let $\mathcal{B}_o \subseteq \mathcal{B}$ be an algebra such that $\mathcal{B}(\mathcal{B}_o) = \mathcal{B}$. If $\alpha = \{A_1, A_2, \ldots, A_k\}$ is a finite partition of $(X, \mathcal{B}, \mu)$ and $\varepsilon > 0$, then there is a finite partition $\Delta$ of $(X, \mathcal{B}, \mu)$, $\Delta = \{D_1, D_2, \ldots, D_k\} \subseteq \mathcal{B}_o$, such that $d(\alpha, \Delta) < \varepsilon$.

The following theorem serves as a computational device in its own right and is also useful in proving the main theorem of this section. The theorem states that if a refining sequence of partitions in $F$ generates $\mathcal{B}$ then the entropy of a map can be found by means of a limiting process. Part (2) of the proof shows that $h(T) = \sup\{h(T, \alpha) \mid \alpha \text{ finite}\}$.

2.16 <u>Theorem</u>

Let $\alpha_1 \overset{o}{\leq} \alpha_2 \overset{o}{\leq} \alpha_3 \leq \ldots$ be a sequence in $F$ which generates $\mathcal{B}$. That is, $\mathcal{B} = \mathcal{B}(\bigvee_{n=1}^{\infty} \mathcal{A}(\alpha_n))$. If $T: X \to X$ is measure preserving then $h(T) = \lim_{n\to\infty} h(T, \alpha_n)$.

<u>Proof</u>: ([7])

Let $\xi \in F$, $\varepsilon > 0$. Then by 2.13 there is a finite partition $\beta$ such that $d(\xi, \beta) < \varepsilon$. From proof of 2.12, $|h(T,\xi) - h(T,\beta)| < \varepsilon$. Thus, $h(T,\xi) < h(T,\beta) + \varepsilon$. Thus, $h(T,\xi) \leq \sup\{h(T,\beta) \mid \beta \text{ finite}\}$ since $\varepsilon$ is arbitrary. Thus $\sup\{h(T,\xi) \mid \xi \in F\} \leq \sup\{h(T,\beta) \mid \beta \text{ finite}\}$ (1)

since $\xi \in F$ is arbitrary.

Thus $$h(T) = \sup \{h(T,\xi) \mid \xi \in F\}$$
$$\leq \sup \{h(T,\beta) \mid \beta \text{ finite}\} .$$

But, since $\beta \in F$ for every $\beta$ finite we have

$$\sup \{h(T,\beta) \mid \beta \text{ finite}\} \leq \sup \{h(T,\xi) \mid \xi \in F\}$$

Thus $$h(T) = \sup \{h(T,\beta) \mid \beta \text{ finite}\} . \tag{2}$$

Since $\alpha_n \in F$ for all $n \geq 1$

$$h(T,\alpha_n) \leq \sup \{h(T,\xi) \mid \xi \in F\} = h(T)$$

Thus, $$h(T,\alpha_n) \leq h(T) \quad \text{for every} \quad n \geq 1 \tag{3}$$

Let $\varepsilon > 0$ be given and let $\alpha = \{A_1, A_2, \ldots, A_k\}$ be a finite partition of $(X, B, \mu)$ . Let $B_o = \bigvee_{n=1}^{\infty} A(\alpha_n)$ . By 2.15 there is a finite partition $\Delta = \{D_1, D_2, \ldots, D_k\} \subseteq B_o$ such that $d(\alpha,\Delta) < \varepsilon$ . Since $\Delta$ is finite there is some $n$ such that $\Delta \leq \alpha_n$ . Thus, by 2.7(ii) $h(T,\Delta) \leq h(T,\alpha_n)$. From proof of 2.12 $|h(T,\alpha) - h(T,\Delta)| < \varepsilon$ .

Consequently, $$h(T,\alpha_n) \geq h(T,\Delta) > h(T,\alpha) - \varepsilon. \tag{4}$$

Since $\alpha_1 \overset{o}{\leq} \alpha_2 \overset{o}{\leq} \alpha_3 \overset{o}{\leq} \ldots$, we have by 2.7(ii)

$$h(T,\alpha_1) \leq h(T,\alpha_2) \leq h(T,\alpha_3) \leq \ldots \quad .$$

Thus $$\sup \{h(T,\alpha_n) \mid n \geq 1\} = \lim_{n\to\infty} h(T,\alpha_n) \leq \infty \tag{5}$$

Since $\varepsilon$ is arbitrary we have by (4) $h(T,\alpha) \leq \sup\{h(T,\alpha_n) \mid n \geq 1\}$

But $\alpha$ is also arbitrary so,

$$\sup \{h(T,\alpha) \mid \alpha \text{ finite}\} \leq \sup \{h(T,\alpha_n) \mid n \geq 1\} .$$

By (2) $\quad h(T) \leq \sup \{h(T,\alpha_n) \mid n \geq 1\}$

By (3) $\quad \sup \{h(T,\alpha_n) \mid n \geq 1\} \leq h(T)$ .

By (5) $\quad \lim_{n\to\infty} h(T,\alpha_n) = h(T)$ □

The next theorem shows that for some maps there is a partition $\alpha$ such that $h(T) = h(T,\alpha)$ .

2.17 Theorem (Kolmogoroff-Sinai)

Let $T:X\to X$ be measure preserving. If there is a finite partition $\alpha$ such that $\mathcal{B} = \mathcal{B}(\bigvee_{i=0}^{\infty} T^{-i}\alpha)$ then $h(T) = h(T,\alpha)$ .

Proof:

Let $\alpha_n = \bigvee_{i=0}^{n-1} T^{-i}\alpha$ for $n \geq 1$ . Then $\alpha_n \leq \alpha_{n+1}$ for every $n$ . By 2.16 $h(T) = \lim_{n\to\infty} h(T,\alpha_n)$. By 2.7(v) $h(T,\alpha) = h(T, \bigvee_{i=0}^{k-1} T^{-i}\alpha)$ for $k \geq 1$ . Consequently, $h(T,\alpha) = h(T,\alpha_n)$ for $n \geq 1$ . Therefore

$$h(T,\alpha) = \lim_{n\to\infty} h(T,\alpha_n) = h(T) \qquad \square$$

We can use this theorem to compute the entropy for Bernoulli shifts, Markov shifts and maps on the unit internal of the real line.

Example 1

Let $T:[0,1)\to[0,1)$ be given by $T(x) = 2x \pmod 1$.

The collection of intervals of the form $[a,b)$ with dyadic rational end points generates the class of Borel sets $\mathcal{B}([0,1))$ since the dyadic rationals are dense on the unit interval.

Let $\alpha = \{[0,\tfrac{1}{2}), [\tfrac{1}{2},1)\}$ and let

$\alpha_n = \bigvee_{i=0}^{n-1} T^{-i}\alpha$ . Each $\alpha_n$ is of the form

$$\alpha_n = \{[0,\frac{1}{2^n}), [\frac{1}{2^n}, \frac{2}{2^n}), \ldots, [\frac{2^n-1}{2^n}, 1)\} .$$

Let $[a,b)$ be an arbitrary dyadic rational interval where

$$a = \frac{1}{2^{n_1}} + \frac{1}{2^{n_2}} + \ldots + \frac{1}{2^{n_l}}$$

and $$b = \frac{1}{2^{m_1}} + \frac{1}{2^{m_2}} + \ldots + \frac{1}{2^{m_k}} .$$

Since
$$a = \frac{2^{n_\ell - n_1} + 2^{n_\ell - n_2} + \ldots + 1}{2^{n_\ell}}$$

and
$$b = \frac{2^{m_k - m_1} + 2^{m_k - m_2} + \ldots + 1}{2^{m_k}}$$

$[a,b)$ is a union of members of $\alpha_n$ for $n = \max\{m_k, n_\ell\}$. Thus every open set in $[0,1)$ is an at most countable union of members of $\bigvee_{n=1}^{\infty} A(\alpha_n)$. Thus $\mathcal{B}([0,1)) \subseteq \mathcal{B}(\bigvee_{n=1}^{\infty} A(\alpha_n))$. Since every member of $\alpha_n$ is a Borel set $\mathcal{B}(\bigvee_{n=1}^{\infty} A(\alpha_n)) \subseteq \mathcal{B}([0,1))$. Therefore
$$\mathcal{B}([0,1)) = \mathcal{B}(\bigvee_{n=1}^{\infty} A(\alpha_n)).$$

Using 2.17 gives $h(T) = h(T,\alpha)$. With Lebesgue measure we have,
$$H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) = -2^n(\frac{1}{2^n}\log\frac{1}{2^n}) = n\log 2$$

Then
$$h(T,\alpha) = \lim_{n\to\infty}\frac{1}{n}(n\log 2) = 1 \quad .$$

We could have also used the fact that $\{\alpha_n\}_{n=1}^{\infty}$ is refining and generates $\mathcal{B}([0,1))$ and utilize 2.16.

The next two examples use the Kolmogoroff-Sinai Theorem to derive the entropy for Bernoulli and Markov shifts and are stated as propositions.

2.18 Proposition

Let $\pi = (p_0, p_1, \ldots, p_{k-1})$. The two-sided Bernoulli $\pi$-shift has entropy $-\sum_{i=0}^{k-1} p_i \log p_i$.

Proof:

Let $X = \{0,1,\ldots,k-1\}$ be the state space.

Let $\alpha = \{{}_0[0], {}_0[1], \ldots, {}_0[k-1]\}$ be a partition of the sequence space.

Let $\alpha_n = \bigvee_{i=-n}^{n} \sigma^i\alpha$ $\quad n = 0,1,2,\ldots$ and $\sigma$ is the shift on $\{0,1,\ldots,k-1\}^Z$. It is clear that $\alpha_0 \leq \alpha_1 \leq \alpha_2 \leq \ldots$. We need to show that this refining sequence generates $\mathcal{B}$.

$\alpha_n$ contains all cylinder sets of the form ${}_{-n}[a_{-n},\ldots,a_0,\ldots,a_n]$ where $a_i \in X$- for each $i$. $A(\alpha_n)$ contains all the unions of such sets.

Basic sets in the product topology are of the form

$$\ldots \times X \times X \times U_{-m} \times U_{-m+1} \times \ldots \times U_0 \times \ldots \times U_m \times X \times X \times \ldots$$

where $U_j \subseteq X$ (not necessarily proper) for each $j$. If $A$ is a basic set in the product topology then for some $n$, $A \in A(\alpha_n)$. Thus all basic sets are contained in $\bigvee_{n=0}^{\infty} A(\alpha_n)$. Thus all open sets are also contained in $\bigvee_{n=0}^{\infty} A(\alpha_n)$. Thus $\mathcal{B} = \mathcal{B}(\bigvee_{n=0}^{\infty} A(\alpha_n))$. Therefore, $h(\sigma) = h(\sigma,\alpha)$ by the Kolmogoroff-Sinai Theorem.

$$h(\sigma,\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha) .$$

Let ${}_o[i_0,i_1,\ldots,i_{n-1}]$ be a typical member of $\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha$.

$\mu({}_o[i_0,i_1,\ldots,i_{n-1}]) = p_{i_o}p_{i_1}\cdots p_{i_{n-1}}$. Thus

$$H(\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha) = -\sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} (p_{i_0}\cdots p_{i_{n-1}}) \log (p_{i_0}\cdots p_{i_{n-1}})$$

$$= -\sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} (p_{i_0}\cdots p_{i_{n-1}})[\log p_{i_0} + \ldots + \log p_{i_{n-1}}]$$

$$= -\sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0}\cdots p_{i_{n-1}} \log p_{i_0}$$

$$-\sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0}\cdots p_{i_{n-1}} \log p_{i_1}$$

$$-\ldots-\sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0}\cdots p_{i_{n-1}} \log p_{i_{n-1}}$$

$$= -\sum_{i_0=0}^{k-1} p_{i_0} \log p_{i_0} - \sum_{i_1=0}^{k-1} p_{i_1} \log p_{i_1}$$

$$-\dots-\sum_{i_{n-1}=0}^{k-1} p_{i_{n-1}} \log p_{i_{n-1}}$$

$$= -n \sum_{i=0}^{k-1} p_i \log p_i$$

Thus $\quad \frac{1}{n} H(\bigvee_{i=0}^{n-1} \sigma^{-i} \alpha) = -\sum_{i=0}^{k-1} p_i \log p_i$

Therefore $\quad h(\sigma,\alpha) = -\sum_{i=0}^{k-1} p_i \log p_i = h(\sigma)$ □

By replacing $Z$ by $N$ and $\bigvee_{i=-n}^{n} \sigma^{-i} \alpha$ by $\bigvee_{i=0}^{n} \sigma^{-i} \alpha$ we obtain the following:

<u>Corollary</u>

Let $\pi = (p_o, p_1, \dots, p_{k-1})$. The one-sided Bernoulli shift has entropy $-\sum_{i=0}^{k-1} p_i \log p_i$ .

2.19 <u>Proposition</u>

Let $\pi = (p_o, p_1, \dots, p_{k-1})$ and $P = (p_{ij})_{i,j=0}^{k-1}$

The two-sided Markov $\pi, P$ shift has entropy $-\sum_{i,j=0}^{k-1} p_i p_{ij} \log p_{ij}$

<u>Proof</u>:

We let $\alpha$ and $\alpha_n$ be as in proof of 2.18. Then $\mathcal{B} = \mathcal{B}(\bigvee_{n=0}^{\infty} A(\alpha_n))$. Consequently, by the Kolmogoroff-Sinai Theorem $h(\sigma) = h(\sigma,\alpha)$ . We have

$$h(\sigma,\alpha) = \lim_{\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} \sigma^{-i} \alpha) .$$

A typical element of $\bigvee_{i=0}^{n-1} \sigma^{-i} \alpha$ is the form ${}_o[i_o, i_1, \dots i_{n-1}]$ with measure $p_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}}$ . So, $H(\bigvee_{i=0}^{n-1} \sigma^{-i} \alpha) =$

$$-\sum_{i_0, i_1, \dots, i_{n-1}=0}^{k-1} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log (p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}})$$

$$= -\sum_{i_0, i_1, \dots, i_{n-1}=0}^{k-1} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} (\log p_{i_0} + \log p_{i_0 i_1} + \dots + \log p_{i_{n-2} i_{n-1}})$$

$$- \sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log p_{i_0 i_1}$$

$$- \ . \ . \ .$$

$$- \sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log p_{i_{n-2} i_{n-1}} \, .$$

Using $\sum_{j=0}^{k-1} p_{ij} = 1$ gives

$$- \sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log p_{i_0}$$

$= - \sum_{i=0}^{k-1} p_i \log p_i$ . Using $\sum_{j=0}^{k-1} p_{ij} = 1$ and $\sum_{i=0}^{k-1} p_i p_{ij} = p_j$ gives:

$$- \sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log p_{i_0 i_1}$$

$$= - \sum_{i_0,i_1=0}^{k-1} p_{i_0} p_{i_0 i_1} \log p_{i_0 i_1}$$

$$\vdots$$

$$- \sum_{i_0,i_1,\ldots,i_{n-1}=0}^{k-1} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log p_{i_{n-2} i_{n-1}}$$

$$= - \sum_{i_{n-2},i_{n-1}=0}^{k-1} p_{i_{n-2}} p_{i_{n-2} i_{n-1}} \log p_{i_{n-2} i_{n-1}}$$

Thus $\quad H(\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha) = - \bigvee_{i=0}^{n-1} p_i \log p_i - (n-1) \sum_{i,j=0}^{k-1} p_i p_{ij} \log p_{ij}$

Thus $\quad \frac{1}{n} H(\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha) = \frac{1}{n} \sum_{i=0}^{k-1} p_i \log p_i - (\frac{n-1}{n}) \sum_{i,j=0}^{k-1} p_i p_{ij} \log p_{ij}$ .

Therefore, $\quad h(\sigma,\alpha) = - \sum_{i,j=0}^{k-1} p_i p_{ij} \log p_{ij}$

$$= h(\sigma) \quad . \qquad \square$$

Corollary

The one sided $\pi, P$ Markov shoft has entropy $-\sum_{i,j=0}^{k-1} p_i p_{ij} \log p_{ij}$ .

We shall use the entropy formula for the Bernoulli shift to compute the entropy of the following map on $[0,1]$ .

Example 2

Let $(I, \mathcal{B}, m)$ be the measure space where $I = [0,1]$ , $\mathcal{B}$ is the Borel class, and $m$ is Lebesgue measure. Let $(X, \mathcal{B}(X), \mu_\pi)$ be the measure space where $X = \{0,1\}^N$ , $\mathcal{B}(X)$ is the Borel class, and $\mu_\pi$ is the Bernoulli measure with $\pi = (q, 1-q)$ for $0 < q < 1$ .

Let $T: I \to I$ be given by

$$T(x) = \begin{cases} \frac{x}{q} & x \in [0,q] \\ \frac{1-x}{1-q} & x \in [q,1] \end{cases}$$

and $\sigma: X \to X$ is the shift on $X$ .

It can be shown that $T$ is isomorphic to $\sigma$ and thus by 2.18 $h(T) = -q \log q - (1-q) \log (1-q)$ .

Let $I_0 = [0,q]$ and $I_1 = [q,1]$ . Define $\psi: X \to I$ by $\psi(<x_n>) = \bigcap_{n=0}^{\infty} T^{-n} I_{x_n}$ . It is clear that for each pair $i,j \in \{0,1\}$ $T(I_i) \supseteq I_j$ . Let $<x_n> \in X$ be arbitrary and $k \geq 1$ . Then

$I_{x_0} \cap T^{-1} I_{x_1} \cap \ldots \cap T^{-k} I_{x_k} \neq \phi$ since $T^k(I_{x_0}) \supseteq \ldots \supseteq T(I_{k_{k-1}}) \supseteq I_{x_k}$ . Consequently, the sequence $\{\bigcap_{n=0}^{k} T^{-n} I_{x_n}\}_{k=1}^{\infty}$ is a nested sequence of compact sets since $I_0$ and $I_1$ are closed and $T$ is continuous. Therefore, $\bigcap_{n=0}^{\infty} T^{-n} I_{x_n} \neq \phi$ . To show that $\psi(<x_n>)$ consists of exactly one point let $\lambda = \min\{\frac{1}{q}, \frac{1}{1-q}\}$ . $\lambda$ is the least value the slope of $T$ has on $[0,1]$ . Suppose $x,y \in \bigcap_{n=0}^{k} T^{-n} I_{x_n}$ . Then $|T^n(x) - T^n(y)| \geq \lambda^n |x - y|$ . Since $T^n(x)$ and $T^n(y)$ are in the same interval for $0 \leq n \leq k$ we have $|T^n(x) - T^n(y)| \leq \max\{q,1-q\} = \frac{1}{\lambda}$. Thus $|x - y| \leq \frac{1}{\lambda^{n+1}}$ for $0 \leq n \leq k$. Consequently, $|x - y| \leq \frac{1}{\lambda^{k+1}}$. If $x,y \in \bigcap_{n=0}^{\infty} T^{-n} I_{x_n}$ then $|x - y| \leq \frac{1}{\lambda^{k+1}}$ for every $k \geq 1$. Therefore $x = y$. It follows that $\psi$ is well defined.

$\psi$ is continuous. To show this let $d(<x_n>, <y_n>) = \sum_{n=0}^{\infty} \frac{|x_n - y_n|}{2^{n+1}}$ be a metric on $X$ ($d$ is compatible with the topology). For $d(<x_n>, <y_n>) < \frac{1}{2^{k+1}}$, $|\psi(<x_n>) - \psi(<y_n>)| < \frac{1}{2^k}$. Therefore $\psi$ is continuous.

Let $D = \bigcup_{i=0} T^{-i}\{q\}$. $m(D) = 0$ since $D$ is countable. $\psi^{-1}(D)$ consists of those symbol sequences terminating with 1000.... Since $\psi^{-1}(D)$ is countable, $\mu_\pi(\psi^{-1}(D)) = 0$.

Let $M_1 = X \setminus \psi^{-1}(D)$ and $M_2 = I \setminus D$. Suppose $<x_n>, <y_n> \in M_1$ and $<x_n> \neq <y_n>$. Let $x = \psi(<x_n>)$ and $y = \psi(<y_n>)$. Since $<x_n>, <y_n> \in M_1$, $T^n(x) \neq q$ and $T^n(y) \neq q$ for every $n$. Thus $T^n(x) \in I^o_{x_n}$ and $T^n(y) \in I^o_{y_n}$ for all $n$ where $I^o_0 = [0,q)$ and $I^o_1 = (q,1]$. Thus $x = \bigcap_{n=0}^{\infty} T^{-n} I^o_{x_n}$ and $y = \bigcap_{n=0}^{\infty} T^{-n} I^o_{y_n}$. Since $<x_n> \neq <y_n>$ there is some $j$ s.t. $x_j \neq y_j$. Then $T^j(x) \in I^o_{x_j}$, $T^j(y) \in I^o_{y_j}$ and $I^o_{x_j} \cap I^o_{y_j} = \phi$ since $x_j \neq y_j$. Therefore $x \neq y$ and $\psi|_{M_1}$ is 1 - 1. Let $x \in M_2$. Then there is some $<x_n> \in X$ s.t. $T^n(x) \in I^o_{x_n}$ for every $n$. Thus $x \in \bigcap_{n=0}^{\infty} T^{-n} I^o_{x_n} \subseteq \bigcap_{n=0}^{\infty} T^{-n} I_{x_n}$. Consequently, $x = \psi(<x_n>)$. Therefore, $\psi|_{M_1}: M_1 \to M_2$ is onto and thus a bijection.

To show $(\psi|_{M_1})^{-1}$ is continuous let $z \in M_2$. For some $<z_n> \in X$, $z = \psi(<z_n>)$. Since $z \in M_2$, $T^n(z) \in I^o_{z_n}$ for every $n$. Therefore $z = \bigcap_{n=0}^{\infty} T^{-n} I^o_{z_n}$. The restriction of $\bigcap_{n=0}^{k} T^{-n} I^o_{z_n}$ to $M_2$ is open in the supspace topology for every finite $k$. So, for each $k$ there is a $\delta_k > 0$ s.t. if $y \in M_2$ and $|y - z| < \delta_k$ then $y \in \bigcap_{n=0}^{k} T^{-n} I^o_{z_n}$. Thus $d((\psi|_{M_1})^{-1}(z), (\psi|_{M_1})^{-1}(y)) < \frac{1}{2^k}$. Therefore $(\psi|_{M_1})^{-1}$ is continuous. Consequently, $\psi|_{M_1} : M_1 \to M_2$ is a homeomorphism.

Since $\psi|_{M_1}$ is a homeomorphism it maps Borel sets to Borel sets. All cylinder sets ${}_o[a_0, a_1, \ldots, a_n]$ (restricted to $M_1$) together with $\phi$ form a semi-algebra which generates $\mathcal{B}'(X) = \{B \cap M_1 | B \in \mathcal{B}(X)\}$. The images of such cylinder sets under $\psi|_{M_1}$ form a semi-algebra which generates $\mathcal{B}'(I) = \{B \cap M_2 | B \in \mathcal{B}(I)\}$. Thus to show $\psi|_{M_1}$ is invertible measure preserving it is sufficient to show that $\mu_\pi({}_o[a_0, a_1, \ldots, a_n] = m(\psi|_{M_1}({}_o[a_0, a_1, \ldots, a_n]))$ for each cylinder set. Let ${}_o[a_0, a_1 \ldots, a_n]$ be arbitrary. $\mu_\pi({}_o[a_0, a_1, \ldots, a_n]) = p_{a_0} p_{a_1} \ldots p_{a_n}$. $\psi|_{M_1}({}_o[a_0, a_1, \ldots, a_n]) = I^o_{a_0} \cap T^{-1} I^o_{a_1} \cap \ldots \cap T^{-n} I^o_{a_n}$ which has Lebesgue measure $m(I^o_{a_0}) m(I^o_{a_1}) \ldots m(I^o_{a_n})$ due to the linear character of $T$. This equals $p_{a_0} p_{a_1} \ldots p_{a_n}$. Thus $\psi|_{M_1}$ is i. m. p. Consequently, $(X, \mathcal{B}(X), \mu_\pi) \overset{\psi}{\rightleftarrows} (I, \mathcal{B}, m)$.

It is clear that $(\psi \circ \sigma)(\langle x_n \rangle) = (T \circ \psi)(\langle x_n \rangle)$ for every $\langle x_n \rangle \in M_1$. Therefore $T$ is isomorphic to $\sigma$. Consequently,
$h(T) = h(\sigma) = -\sum_{i=0}^{1} p_i \log p_i = -q \log q - (1-q) \log (1-q)$.

The invariant property of entropy was used by Kolmogoroff ([15]) to show that the $(\frac{1}{2}, \frac{1}{2})$ Bernoulli shift is not isomorphic to the $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ Bernoulli shift since the former has entropy 1 and the latter entropy $\log_2 3$. He conjectured that entropy is a complete invariant for Bernoulli shifts. That is, two maps in this class would be isomorphic if they had the same entropy. This conjecture was proved by Donald Orstien [19] in 1969. His result will be stated here.

Definition

A probability space $(X, \mathcal{B}, \mu)$ is a Lebesgue space if it is isomorphic to a probability space which is the disjoint union of an at most countable collection of points $\{x_1, x_2, x_3, \ldots\}$ each of positive measure and the space $([0,s], L[0,s], m)$ where $L[0,s]$ is the $\sigma$-algebra of Lebesgue measurable subsets of $[0,s]$ and $m$ is Lebesgue measure. $s = 1 - \sum_{n=1}^{\infty} p_n$ where $p_n$ is the measure of $x_n$.

2.20 Theorem ([19], p.18-32)

Let $T_1, T_2$ be Bernoulli shifts whose state spaces are Lebesgue spaces. If $h(T_1) = h(T_2)$ then $T_1$ is isomorphic to $T_2$.

One of the many remarkable things about this theorem is that a Bernoulli shift with a countable state space can be isomorphic to a Bernoulli shift with finite state space.

# CHAPTER III

# TOPOLOGICAL ENTROPY

## Section 1. Introduction

The material in this chapter consists of an introduction to the notion of topological entropy of continuous maps on compact spaces. This is taken from Adler, Konheim and McAndrew [1] .

This is followed by a discussion of the variational principle for topological entropy and the existence of maximal measures. This is taken from Brown [6], Parry [20] and Walters [24] .

Following this, is a brief discussion concerning the classification of maps by means of topological entropy.

## Section 2. Topological Entropy of Maps on Compact Spaces

Let $X$ be a compact topological space. We can define the entropy of a continuous map $T$ in a manner analogous to that used in the measure theoretic context.

**Definition**

For any open cover $\alpha$ of $X$ let $N(\alpha)$ denote the number of sets in a subcover of minimum cardinality. A subcover with cardinality equal to $N(\alpha)$ is called minimal.

Since $X$ is compact $N(\alpha)$ is always finite and non-zero.

**Definition**

For any two open covers $\alpha,\beta$ of $X$, their join denoted $\alpha \vee \beta$ is given by: $\alpha \vee \beta = \{A \cap B \mid A \in \alpha , B \in \beta\}$ .

**Definition**

A cover $\beta$ is a refinement of a cover $\alpha$, denoted $\alpha \leq \beta$ , if every member of $\beta$ is a subset of some member of $\alpha$ .

3.1 <u>Proposition</u>

If $\alpha \leq \alpha'$ and $\beta \leq \beta'$ then $\alpha \vee \beta < \alpha' \vee \beta'$ .

<u>Definition</u>

Let the set of all open covers of $X$ be denoted by $\mathcal{U}(X)$.

3.2 <u>Proposition</u>

If $\alpha, \beta \in \mathcal{U}(X)$ then $\alpha \leq \alpha \vee \beta$ .

<u>Proof</u>:

$\alpha = \alpha \vee \{X\}$ and $\{X\} \leq \beta$ . Using Prop. 3.1: $\alpha = \alpha \vee \{X\} \leq \alpha \vee \beta$ . □

We define the topological entropy of a map $T$ analogously to the measure theoretic entropy. In the topological setting open covers take the place of partitions. Their entropy is defined as follows.

<u>Definition</u>

Let the <u>entropy</u> of an open cover $\alpha$ of $X$, denoted $H(\alpha)$, be given by: $H(\alpha) = \log N(\alpha)$ .

3.3 <u>Proposition</u>

Let $\alpha, \beta \in \mathcal{U}(X)$ . Then, $\alpha \leq \beta \Rightarrow N(\alpha) \leq N(\beta) \Rightarrow H(\alpha) \leq H(\beta)$ .

<u>Proof</u>:

Let $\{B_1, B_2, \ldots, B_{N(\beta)}\}$ be a minimal subcover of $\beta$ . Since $\alpha \leq \beta$ there is a subcover of $\alpha$, $\{A_1, A_2, \ldots, A_{N(\beta)}\}$ $\therefore$ $N(\alpha) \leq N(\beta)$ . The inequality $H(\alpha) \leq H(\beta)$ follows. □

3.4 <u>Proposition</u>

Let $\alpha, \beta \in \mathcal{U}(X)$. Then $\alpha \leq \beta \Rightarrow N(\alpha \vee \beta) = N(\beta)$

$$\Rightarrow H(\alpha \vee \beta) = H(\beta)$$

<u>Proof</u>:

$\beta = \{X\} \vee \beta \leq \alpha \vee \beta$ . And $\alpha \vee \beta \leq \beta$ since for each $B \in \beta$ there is some $A \in \alpha$ s.t. $B \subseteq A$ . Thus $B \subseteq A \cap B$ and consequently $\alpha \vee \beta \leq \beta$ .

By 3.3, $N(\alpha \vee \beta) \geq N(\beta)$ and $N(\alpha \vee \beta) \leq N(\beta)$. Thus $N(\alpha \vee \beta) = N(\beta)$ and therefore $H(\alpha \vee \beta) = H(\beta)$. □

3.5 <u>Proposition</u>

Let $\alpha, \beta \in \mathcal{U}(X)$. Then $N(\alpha \vee \beta) \leq N(\alpha) \cdot N(\beta)$ and $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.

<u>Proof</u>:

Let $\{A_1, A_2, \ldots, A_{N(\alpha)}\}$ and $\{B_1, B_2, \ldots, B_{N(\beta)}\}$ be minimal subcovers of $\alpha$ and $\beta$ respectively. Then $\{A_i \cap B_j | 1 \leq i \leq N(\alpha),\ 1 \leq j \leq N(\beta)\}$ is a subcover of $\alpha \vee \beta$ with cordinality less than or equal to $N(\alpha) \cdot N(\beta)$. Thus $N(\alpha \vee \beta) \leq N(\alpha)N(\beta)$. Therefore $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$ □

In the discussion of measure theoretic entropy the possibility of defining $H(\bigvee_{i=0}^{n-1} T^{-i} \alpha)$ was dependent on $T$ being measure preserving. This guaranteed that $\bigvee_{i=0}^{n-1} T^{-i} \alpha$ was a partition if $\alpha$ was a partition. Since we are defining topological entropy in a way analogous to measure theoretic entropy we shall use the wedge product $\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-(n-1)}\alpha$ and apply the entropy map $H$ to this open cover. We can only be sure that $\bigvee_{i=0}^{n-1} T^{-i} \alpha$ is an open cover if $T:X \to X$ is continuous. Consequently, for the remainder of this chapter let $T:X \to X$ be continuous. If $\alpha \in \mathcal{U}(X)$ then $T^{-1}\alpha = \{T^{-1} A | A \in \alpha\}$ is also in $\mathcal{U}(X)$.

3.6 <u>Proposition</u>

Let $\alpha, \beta \in \mathcal{U}(X)$. Then $\alpha \leq \beta$ implies $T^{-1}\alpha \leq T^{-1}\beta$.

<u>Proof</u>:

Let $T^{-1}B \in T^{-1}\beta$. Since $\alpha \leq \beta$ there is some $A \in \alpha$ s.t. $B \subseteq A$. Then $T^{-1}B \subseteq T^{-1}A$. But $T^{-1}A \in T^{-1}\alpha$. $\therefore$ $T^{-1}\alpha \leq T^{-1}\beta$. □

3.7 <u>Proposition</u>

Let $\alpha, \beta \in \mathcal{U}(X)$. Then: 1) $T^{-1}(\alpha \vee \beta) = T^{-1}\alpha \vee T^{-1}\beta$

2) $N(T^{-1}\alpha) \leq N(\alpha)$.

Equality holds when $T$ is onto.

Proof:

1) $T^{-1}(\alpha \vee \beta) = \{T^{-1}(A \cap B) \mid A \in \alpha, B \in \beta\}$

$$= \{T^{-1}A \cap T^{-1}B \mid A \in \alpha, B \in \beta\}$$

$$= T^{-1}\alpha \vee T^{-1}\beta .$$

2) Let $\{A_1, A_2, \ldots, A_{N(\alpha)}\}$ be a minimal subcover of $\alpha$. Then $\{T^{-1}A_1, T^{-1}A_2, \ldots, T^{-1}A_{N(\alpha)}\}$ is also a cover and thus a subcover of $T^{-1}\alpha$. Therefore $N(T^{-1}\alpha) \leq N(\alpha)$. If $T$ is onto $\{T^{-1}A_1, T^{-1}A_2, \ldots, T^{-1}A_{N(\alpha)}\}$ is minimal so equality holds in such a case. □

3.8 Proposition

Let $\alpha \in \mathcal{U}(X)$.

Let $\alpha \vee T^{-1}\alpha \vee \ldots \vee T^{-(n-1)}\alpha$ be denoted by $\bigvee_{i=0}^{n-1} T^{-i}\alpha$. Then $\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ exists and is finite.

Proof:

Let $a_n = H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$. Since $a_n \geq 0$ for every $n$ it is sufficient to show that $a_{m+n} \leq a_m + a_n$ (by 2.5)

$$a_{m+n} = H(\bigvee_{i=0}^{m+n-1} T^{-i}\alpha)$$

$$= H(\bigvee_{i=0}^{m-1} T^{-i}\alpha \vee \bigvee_{i=m}^{m+n-1} T^{-i}\alpha)$$

$$\leq H(\bigvee_{i=0}^{m-1} T^{-i}\alpha) + H(\bigvee_{i=m}^{m+n-1} T^{-i}\alpha) \qquad \text{(by 3.5)}$$

$$= a_m + H(T^{-m}\bigvee_{i=0}^{n-1} T^{-i}\alpha)$$

$$= a_m + \log N(T^{-m}\bigvee_{i=0}^{n-1} T^{-i}\alpha)$$

$$\leq a_m + \log N(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \qquad \text{(by 3.7)}$$

$$= a_m + H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) = a_m + a_n$$

Thus $\quad a_{m+n} \leq a_m + a_n$ . By 2.5 $\lim_{n\to\infty} \frac{a_n}{n} = \inf\{\frac{a_n}{n} | n \geq 1\}$ .

Therefore $\quad \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ exists and equals

$$\inf\{\frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) | n \geq 1\} \quad \square .$$

We are now in a position to define the topological entropy of a map with respect to an open cover $\alpha$ .

Definition

The entropy of $T:X\to X$ with respect to $\alpha \in \mathcal{U}(X)$ is denoted by $h(T,\alpha)$ and is given by: $h(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ .

The next proposition shows that the entropy of a map $T$ with respect to an open cover $\alpha$ is bounded above by the entropy of the cover.

3.9 Proposition

Let $\alpha \in \mathcal{U}(X)$ . Then $h(T,\alpha) \leq H(\alpha)$ .

Proof:

Let $a_n = H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$ as in proof of 3.8. Then $a_n \leq na_1$ . Thus, $\frac{a_n}{n} \leq a_1$ for every $n \geq 1$ . Since $h(T,\alpha) = \inf\{\frac{a_n}{n} | n \geq 1\}$ as in proof of 3.8 it follows that $h(T,\alpha) \leq a_1$. But $a_1 = H(\alpha)$ $\square$ .

3.10 Proposition

Let $\alpha,\beta \in \mathcal{U}(X)$ . If $\alpha \leq \beta$ then $h(T,\alpha) \leq h(T,\beta)$.

Proof:

By 3.6 $\quad T^{-n}\alpha \leq T^{-n}\beta \quad$ for every $n \geq 1$ .

By 3.1 $\quad \bigvee_{i=0}^{n-1} T^{-i}\alpha \leq \bigvee_{i=0}^{n-1} T^{-i}\beta \quad$ for every $n \geq 1$ .

By 3.3 $\quad H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) \leq H(\bigvee_{i=0}^{n-1} T^{-i}\beta) \quad$ for every $n \geq 1$ .

Thus

$$h(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha)$$

$$\leq \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\beta) = h(T,\beta) \quad \square$$

Just as measure theoretic entropy was defined by means of a supremum over all partitions with finite entropy so too topological entropy is defined by means of a supremum over all open covers.

Definition

The topological entropy of $T$, denoted $h(T)$, is given by:

$$h(T) = \sup\{h(T,\alpha) \mid \alpha \in \mathcal{U}(X)\}$$

If, in any given context, reference is made to both the topological and measure theoretic entropy of $T$ then the former will be denoted $h(T)$ and the latter $h_m(T)$ where $m$ denotes the measure being used.

In order to discuss the invariance properties of topological entropy it is necessary to introduce the notions of conjugacy and semi-conjugacy.

Definition

Let $X_1$ and $X_2$ be topological spaces and $T_i:X_i \to X_i$ continuous for $i = 1,2$. If $\phi:X_1 \to X_2$ is a continuous map with $\phi(X_1) = X_2$ and $\phi \circ T_1 = T_2 \circ \phi$ then $\phi$ is referred to as a semi-conjugacy from $T_1$ to $T_2$ and $T_2$ is said to be a factor of $T_1$. If $\phi$ is a homeomorphism then it is referred to as a conjugacy and $T_1$ and $T_2$ are said to be conjugate.

The invariance property of entropy is stated in the following proposition.

3.11 Proposition

Let $X_1$ and $X_2$ be compact spaces and $T_i:X_i \to X_i$ continuous for $i = 1,2$. If $T_2$ is a factor of $T_1$ then $h(T_1) \geq h(T_2)$. If $T_1$ and $T_2$ are conjugate then $h(T_1) = h(T_2)$.

Proof:

If $T_2$ is a factor of $T_1$ then there is a map $\phi:X_1 \to X_2$ which is continuous and onto s.t. $\phi \circ T_1 = T_2 \circ \phi$. Let $\alpha \in \mathcal{U}(X_2)$. Then

$$h(T_2,\alpha) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T_2^{-i}\alpha\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\phi^{-1} \bigvee_{i=0}^{n-1} T_2^{-i}\alpha\right) \qquad \text{by 3.7}$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} \phi^{-1} \circ T_2^{-i}\alpha\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T_1^{-i} \circ \phi^{-1}\alpha\right) = h(T_1,\phi^{-1}\alpha)$$

Since $\phi$ is onto and continuous the last term is well defined ($\phi^{-1}\alpha$ being a member of $\mathcal{U}(X_1)$). Thus $h(T_2) \le h(T_1)$. If $T_1$ and $T_2$ are conjugate then $h(T_1) \le h(T_2)$. $\therefore$ $h(T_1) = h(T_2)$ □

Let X be a compact topological space and $T:X\to X$ a continuous map. The following proposition states a useful relationship between the entropy of a map T and that of $T^k$ for $k \ge 1$.

3.12 Proposition

Let $k = 1,2,3,\ldots$. Then $h(T^k) = k\,h(T)$.

Proof:

Let $\alpha \in \mathcal{U}(X)$.

$$k\,h(T,\alpha) = k \lim_{n\to\infty} \frac{1}{nk} H\left(\bigvee_{i=0}^{nk-1} T^{-i}\alpha\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{nk-1} T^{-i}\alpha\right)$$

$$= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-ik} \bigvee_{j=0}^{k-1} T^{-j}\alpha\right)$$

$$= h\left(T^k, \bigvee_{j=0}^{n-1} T^{-j}\alpha\right) \qquad \therefore\ k\,h(T) \le h(T^k)$$

And, $\bigvee_{i=0}^{n-1} T^{-ik}\alpha \le \bigvee_{i=0}^{nk-1} T^{-i}\alpha$

So, $H\left(\bigvee_{i=0}^{n-1} T^{-ik}\alpha\right) \le H\left(\bigvee_{i=0}^{nk-1} T^{-i}\alpha\right)$ by 3.3

$$\frac{1}{k} h(T^k,\alpha) = \frac{1}{k} \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-ik}\alpha)$$

$$= \lim_{n\to\infty} \frac{1}{nk} H(\bigvee_{i=0}^{n-1} T^{-ik}\alpha)$$

$$\leq \lim_{n\to\infty} \frac{1}{nk} H(\bigvee_{i=0}^{nk-1} T^{-i}\alpha) = h(T,\alpha) .$$

So, $h(T^k,\alpha) \leq k\,h(T,\alpha)$ . $\therefore$ $h(T^k) \leq k\,h(T)$

$\therefore$ $h(T^k) = k\,h(T)$ □

The next theorem states an important result for maps which can be decomposed into two or more pieces $\{X_i\}$ s.t. $T|_{X_i}(X_i) \subseteq X_i$ . The proof requires the following lemma.

3.13 Lemma ([1])

Suppose $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ are two sequences of real numbers with $a_n \geq 1$ and $b_n \geq 1$ for every $n$ s.t. $\lim_{n\to\infty} \frac{\log a_n}{n} = a$ and $\lim_{n\to\infty} \frac{\log a_n}{n} = b$ . Then $\lim_{n\to\infty} \frac{\log(a_n + b_n)}{n} = \max\{a,b\}$ .

3.14 Theorem ([1])

Let $X_1$ and $X_2$ be two closed subsets of $X$ s.t. $X = X_1 \cup X_2$ and $T(X_1) \subseteq X_1$ and $T(X_2) \subseteq X_2$ . Then $h(T) = \max\{h(T_1), h(T_2)\}$ where $T_1 = T|_{X_1}$ and $T_2 = T|_{X_2}$ .

Proof:

Let $i = 1$ or $2$ .

For any $\alpha \in \mathcal{U}(X)$ the collection of sets $(\alpha)_i = \{A \cap X_i \mid A \in \alpha\}$ is an open cover of $X_i$ (in the subspace topology). $N_i$ will indicate the space whose cover is being counted. Thus $N_i((\alpha)_i) \leq N(\alpha)$. And for open covers $\alpha$ and $\beta$ of $X$ we also have $(\alpha \vee \beta)_i = \{A \cap B \cap X_i \mid A \in \alpha, B \in \beta\} = \{(A \cap X_i) \cap (B \cap X_i) \mid A \in \alpha, B \in \beta\} = (\alpha)_i \vee (\beta)_i$ . Furthermore, since $T(X_i) \subseteq X_i$

$$T_i^{-1}(\alpha)_i = \{T_i^{-1}(A \cap X_i) \mid A \in \alpha\}$$
$$= \{T^{-1}(A \cap X_i) \cap X_i \mid A \in \alpha\}$$
$$= (T^{-1}\alpha)_i$$

Now, let $\alpha_i$ be an arbitrary open cover of $X_i$ (with respect to the subspace topology). There is an open cover $\alpha$ of X s.t. $(\alpha)_i = \alpha_i$ where $\alpha = \{A \cup (X-X_i) \mid A \in \alpha_i\}$. Now, since $(\alpha)_i = \alpha_i$ we have $T_i^{-1}\alpha_i = (T^{-1}\alpha)_i$. Thus,

$$N_i(\bigvee_{j=0}^{n-1} T_i^{-j}\alpha_i) = N_i(\bigvee_{i=0}^{n-1} (T^{-j}\alpha)_i) .$$

And since, for any two covers $\beta, \gamma \in \mathcal{U}(X)$, $(\gamma \vee \beta)_i = (\gamma)_i \vee (\beta)_i$ it follows that $N_i(\bigvee_{j=0}^{n-1}(T^{-j}\alpha)_i) = N_i((\bigvee_{j=0}^{n-1} T^{-j}\alpha)_i)$. Distinct members of $(\bigvee_{j=0}^{n-1} T^{-i}\alpha)_i$ correspond to distinct members of $\bigvee_{j=0}^{n-1} T^{-j}\alpha$. Thus

$$N_i((\bigvee_{j=0}^{n-1} T^{-j}\alpha)_i) \leq N(\bigvee_{j=0}^{n-1} T^{-j}\alpha)$$ So, we have

$$N_i(\bigvee_{j=0}^{n-1} T_i^{-j}\alpha_i) \leq N(\bigvee_{j=0}^{n-1} T^{-j}\alpha)$$

$$\therefore \quad h(T_i, \alpha_i) \leq h(T, \alpha)$$

$$\therefore \quad h(T_i) \leq h(T) \qquad i = 1,2 .$$

Now, let $\alpha \in \mathcal{U}(X)$.

$$N(\bigvee_{j=0}^{n-1} T^{-j}\alpha) \leq N_1((\bigvee_{j=0}^{n-1} T^{-j}\alpha)_1) + N_2((\bigvee_{j=0}^{n-1} T^{-j}\alpha)_2)$$
$$= N_1(\bigvee_{j=0}^{n-1} (T^{-j}\alpha)_1) + N_2(\bigvee_{j=0}^{n-1} (T^{-j}\alpha)_2)$$
$$= N_1(\bigvee_{j=0}^{n-1} T_1^{-j}(\alpha)_1) + N_2(\bigvee_{j=0}^{n-1} T_2^{-j}(\alpha)_2) .$$ Thus,

$$\log N(\bigvee_{j=0}^{n-1} T^{-j}\alpha) \leq \log[N_1(\bigvee_{j=0}^{n-1} T_1^{-j}(\alpha)_1) + N_2(\bigvee_{j=0}^{n-1} T_2^{-j}(\alpha)_2)]$$

Now, let

$$a_n = N_1(\bigvee_{j=0}^{n-1} T_1^{-j}\ (\alpha)_1)$$

$$b_n = N_2(\bigvee_{j=0}^{n-1} T_2^{-j}\ (\alpha)_2)$$

Then $\lim_{n\to\infty} \frac{1}{n} \log a_n = h(T_1,(\alpha)_1)$

and $\lim_{n\to\infty} \frac{1}{n} \log b_n = h(T_2,(\alpha)_2)$

Applying the lemma gives

$$h(T,\alpha) = \lim_{n\to\infty} \frac{1}{n} \log N(\bigvee_{j=0}^{n-1} T^{-j}\ \alpha)$$

$$\leq \lim_{n\to\infty} \frac{1}{n} \log (a_n + b_n)$$

$$= \max\ \{h(T_1,(\alpha)_1), h(T_2,(\alpha)_2)\}$$

$$\therefore\quad h(T,\alpha) \leq \max\ \{h(T_1,(\alpha)_1), h(T_2,(\alpha)_2)\}$$

$$\therefore\quad h(T) \leq \max\ \{h(T_1), h(T_2)\}$$

$$\therefore\quad h(T) = \max\ \{h(T_1), h(T_2)\}$$ □

We can formulate a theorem analogous to one for measure theoretic entropy (i.e. 2.16) based on the concept of a refining sequence of covers.

Definition

A sequence $\{\alpha_n | n \geq 1\} \subseteq U(X)$ is refining if

1) $\alpha_n \leq \alpha_{n+1}$ $n \geq 1$

2) For every $\beta \in U(X)$ there is some n s.t. $\beta \leq \alpha_n$ .

3.15 Theorem

If $\{\alpha_n \mid n > 1\} \subseteq U(X)$ is refining then $h(T) = \lim_{n\to\infty} h(T,\alpha_n)$ .

Proof:

$$h(T) = \sup\ \{h(T,\beta) | \beta \in U(X)\} .$$

Let $\beta \in U(X)$ be arbitrary. Since $\{\alpha_n \mid n \geq 1\}$ is refining there is some n s.t. $\beta \leq \alpha_n$ . Thus, using 3.10, $h(T,\beta) \leq h(T,\alpha_n)$ .

Thus, $\sup\{h(T,\beta)\,|\,\beta\in \mathcal{U}(X)\} \le \sup\{h(T,\alpha_n)\,|\,n \ge 1\} = h(T)$. (1)

Since $\alpha_n \le \alpha_{n+1}$ we have, using 3.10, $h(T,\alpha_n) \le h(T,\alpha_{n+1})$. Therefore,

$$\lim_{n\to\infty} h(T,\alpha_n) = \sup\{h(T,\alpha_n)\,|\,n \ge 1\}. \quad (2)$$

(1) and (2) imply $\lim_{n\to\infty} h(T,\alpha_n) = h(T)$ □

Section 3. Topological Entropy on Compact Metric Spaces

Let $(X,d)$ be a compact metric space. Within the context of such a space we may prove a theorem analogous to the Kolmogoroff-Sinai theorem.

Definition

a) Let $A \subseteq X$. Then the diameter of $A$, denoted $d(A)$, is given by: $d(A) = \sup\{d(x,y)\,|\,x,y\in A\}$.

b) Let $\alpha\in \mathcal{U}(X)$. Then the diameter of $\alpha$, denoted $d(\alpha)$, is given by: $d(\alpha) = \sup\{d(A)\,|\,A\in\alpha\}$.

We shall make use of the following lemma in a revised form.

3.16 Lemma (Lebesgue Covering Lemma)

For every $\alpha\in \mathcal{U}(X)$ there is some $\varepsilon > 0$ s.t. $U \subseteq X$ and $d(U) < \varepsilon$ implies $U \subseteq A$ for some $A\in\alpha$.

Proof:

([24], p.18) □

Definition

Let $\alpha\in \mathcal{U}(X)$. The Lebesgue number for $\alpha$, denoted $\lambda(\alpha)$, is given by $\lambda(\alpha) = \sup\{\varepsilon > 0\,|\, U \subseteq X \text{ and } d(U) < \varepsilon \Rightarrow A \supseteq U \text{ for some } A\in\alpha\}$.

The Lebesgue number, roughly speaking, gives the size of the largest sets which are totally contained in some member of $\alpha$.

The Lebesgue Covering Lemma can be reformulated to suit our present purpose more adequately.

3.17 Lemma ([1], p. 314)

Let $\alpha,\beta \in \mathcal{U}(X)$. If $d(\beta) < \lambda(\alpha)$ then $\alpha \leq \beta$.

Corollary

If $\{\alpha_n\}_{n\in N}$ is a sequence in $\mathcal{U}(X)$ s.t. $\alpha_n \leq \alpha_{n+1}$ and $d(\alpha_n) \to 0$ as $n \to \infty$ then $\{\alpha_n\}_{n\in N}$ is refining.

The following example uses the corollary to 3.17 as well as 3.15.

Example 1

Let $T:[0,1] \to [0,1]$ be defined by:

$$T(x) = \begin{cases} 2x & x \in [0,\frac{1}{2}] \\ 2-2x & x \in [\frac{1}{2},1] \end{cases}$$

Let $\alpha_n \in \mathcal{U}([0,1])$ be given by:

$$\alpha_n = \{(\frac{k-1}{2^n}, \frac{k+1}{2^n}) \mid 2 \leq k \leq 2^n - 1\}$$

$$\cup\{[0,\frac{1}{2^{n-1}}),(1-\frac{1}{2^{n-1}}, 1]\}, \ n \geq 2 .$$

To show $\{\alpha_n\}$ is refining we first show that $\alpha_n \leq \alpha_{n+1}$. Let $2 < p \leq 2^{n+1} - 1$. If $(\frac{p-1}{2^{n+1}}, \frac{p+1}{2^{n+1}}) \subseteq (\frac{k-1}{2^n}, \frac{k+1}{2^n})$ for some $2 \leq k \leq 2^{n-1}$ then there must be a $k$ s.t.

$$\frac{k-1}{2^n} \leq \frac{p-1}{2^{n+1}} < \frac{p+1}{2^{n+1}} \leq \frac{k+1}{2^n}$$

that is, $2k \leq p+1 < p+3 \leq 2k+4$. If $p$ is even then let $2k = p$. Clearly, $p \leq p+1 < p+3 \leq p+4$. If $p$ is odd then let $2k = p+1$. Clearly, $p+1 \leq p+1 < p+3 \leq p+5$. Let $p = 2$ then $(\frac{1}{2^{n+1}}, \frac{3}{2^{n+1}}) \subseteq [0,\frac{1}{2^{n-1}}) \in \alpha_n$. For the sets $[0,\frac{1}{2^n}) \in \alpha_{n+1}$ and $(1-\frac{1}{2^n},1] \in \alpha_{n+1}$ we have $[0,\frac{1}{2^n}) \subseteq [0,\frac{1}{2^{n-1}}) \in \alpha_n$ and $(1-\frac{1}{2^n}, 1] \in \alpha_n$. Consequently, $\alpha_n \leq \alpha_{n+1}$ for every $n \geq 2$. Furthermore $d(\alpha_n) = \frac{1}{2^{n-1}}$. Thus $d(\alpha_n) \to 0$ as $n \to \infty$. $\therefore$ $\{\alpha_n\}$ is refining.

Let $S_k = N(\bigvee_{i=0}^{k-1} T^{-i} \alpha_n)$ for $n$ fixed. Then $S_{k+1} = 2S_k - 1$ and since $S_1 = N(\alpha_n)$ we have $S_{k+1} = 2^k N(\alpha_n) - 2^k + 1$ for $k \geq 1$. Thus

$$h(T,\alpha_n) = \lim_{k\to\infty} \frac{1}{k} \log S_k = \lim_{k\to\infty} \frac{1}{k+1} \log S_{k+1}$$

$$= \lim_{k\to\infty} \frac{1}{k+1} \log (2^k N(\alpha) - 2^k + 1) = 1$$

By use of 3.15, $h(T) = 1$.

We shall use the notion of a generator $\alpha$ for a homeomorphism $T$. Such generators, when they exist, facilitate the computation of $h(T)$.

Let $(X,d)$ be a compact metric space.

Definition

Let $T:X\to X$ be a homeomorphism. A finite $\alpha \in \mathcal{U}(X)$ is a generator for $T$ if for every sequence $\{A_n\}_{n=-\infty}^{\infty}$ of members of $\alpha$ the set $\bigcap_{n=-\infty}^{\infty} T^{-n} \bar{A}_n$ contains at most one point of $X$.

3.18 Proposition

If $T:X\to X$ is a homeomorphism and $\alpha \in \mathcal{U}(X)$ is a generator for $T$ then for every $\varepsilon > 0$ there is some $N > 0$ s.t. $d(\bigvee_{n=-N}^{N} T^{-n} \alpha) < \varepsilon$.

Proof: ([24], p.138)

By contradiction. Suppose there is some $\varepsilon > 0$ s.t. $d(\bigvee_{n=-N}^{N} T^{-n}\alpha) \geq \varepsilon$ for all $N = 1,2,3,\ldots$. Then for every $j = 1,2,3,\ldots$ there is some $x_j, y_j$ with $d(x_j,y_j) > \varepsilon$ and a sequence $\{A_{ij}\}_{i=-j}^{j}$ of members of $\alpha$ with $x_j, y_j \in \bigcap_{i=-j}^{j} T^{-i} A_{ij}$. There is a subsequence $\{j_k\}$ of natural numbers s.t. $x_{j_k} \to x$ and $y_{j_k} \to y$ since $X$ is compact. There is an $\varepsilon$ distance between all pairs $x_j, y_j$ thus $x \neq y$. Consider the sets $A_{oj_k}$. Since $\alpha$ is finite at least one member of $\alpha$ repeats infinitely often. Denote one such member by $A_o$. Thus $x_{j_k}, y_{j_k} \in A_o$ for infinitely many $k$ and thus $x,y \in \bar{A}_o$. Similarly, for each $n \in Z$ there is a member of $\alpha$, denoted $A_n$, s.t. $x,y \in T^{-n}\bar{A}_n$. Thus $x,y \in \bigcap_{n=-\infty}^{\infty} T^{-n}\bar{A}_n$. But $x \neq y$. Therefore $\alpha$ is not a generator. Contradiction. □

3.19 Theorem

If $T:X\to X$ is a homeomorphism and $\alpha$ is a generator for $T$ then $h(T,\alpha) = h(T)$.

Proof:

Let $\beta \in \mathcal{U}(X)$ with Lebesgue number $\lambda(\beta)$. By 3.13 we may choose $N > 0$ s.t. $d(\bigvee_{n=-N}^{N} T^{-n}\alpha) < \gamma(\beta)$. Then, by 3.17

$$\beta \leq \bigvee_{n=-N}^{N} T^{-n}\alpha. \text{ So,}$$

$$\begin{aligned} h(T,\beta) &\leq h(T, \bigvee_{n=-N}^{N} T^{-n}\alpha) \\ &= \lim_{k\to\infty} \frac{1}{k} H(\bigvee_{i=0}^{k-1} T^{-i} \bigvee_{n=-N}^{N} T^{-n}\alpha) \\ &= \lim_{k\to\infty} \frac{1}{k} H(\bigvee_{n=-N}^{N+k-1} T^{-n}\alpha) \\ &= \lim_{k\to\infty} \frac{1}{k} H(T^{-N} \bigvee_{n=0}^{2N+k-1} T^{-n}\alpha) \\ &= \lim_{k\to\infty} (\frac{2N+k-1}{k})(\frac{1}{2N+k-1}) H(\bigvee_{n=0}^{2N+k-1} T^{-n}\alpha) \\ &= h(T,\alpha). \end{aligned}$$

$$\therefore \quad h(T,\beta) \leq h(T,\alpha).$$

And since $\beta \in \mathcal{U}(X)$ is arbitrary it follows that $h(T) = h(T,\alpha)$. □

If we assume certain suitable conditions on a homeomorphism $T$ we can show that generators must exist.

Definition

A homeomorphism $T:X\to X$ is expansive if there is some $\delta > 0$ s.t. for every pair $x,y \in X$ $(x \neq y)$ there is some $n \in Z$ s.t. $d(T^n x, T^n y) > \delta$. The number $\delta$ is referred to as an expansive constant for T .

3.20 Proposition ([24], p.137-139)

If $T:X\to X$ is a homeomorphism then $T$ is expansive iff $T$ has a generator.

Proof:

Suppose $T$ is expanding and $\delta > 0$ is an expansive constant for $T$. Let $\alpha \in \mathcal{U}(X)$ be a finite cover by open balls of radius $\delta/2$. Suppose $x,y \in \bigcap_{n=-\infty}^{\infty} T^{-n}\bar{A}_n$ where $A_n \in \alpha$. Then $d(T^n x, T^n y) \le \delta$ for all $n \in Z$. Since $T$ is expanding $x = y$. $\therefore$ $\alpha$ is a generator.

Conversely, suppose $\alpha \in \mathcal{U}(X)$ is a generator for $T$. Let $0 < \delta < \lambda(\alpha)$ (see Defn. p. 62). If $d(T^n x, T^n y) \le \delta$ for every $n \in Z$ then every open interval with the pair $T^n x, T^n y$ as endpoints must be contained in some member of $\alpha$. Thus, for every $n \in Z$ there is some $A_n \in \alpha$ s.t. $T^n x, T^n y \in A_n$. Thus $x,y \in \bigcap_{n=-\infty}^{\infty} T^{-n} A_n$. Since $\bigcap_{n=-\infty}^{\infty} T^{-n}\bar{A}_n$ contains at most one point $x = y$. Thus $T$ is expanding. □

3.19 will be used presently to compute the entropy of shift maps.

3.21 Proposition

Let $Y = \{0,1,\dots,k-1\}$ have the discrete topology and $Y^Z$ the product topology. Define $\sigma: Y^Z \to Y^Z$ as in Chapter I Sec. 4. Then $h(\sigma) = \log k$.

Proof:

Let $A_j = {}_0[j]$ with $j = 0,1,\dots,k-1$. (see Ch.1. Sec.4). Then $\alpha = \{A_0, A_1, \dots, A_{k-1}\}$ is a generator for $\sigma$. By 3.19 $h(\sigma) = h(\sigma,\alpha)$.

$$h(\sigma,\alpha) = \lim_{n\to\infty} \frac{1}{n} H\Big(\bigvee_{i=0}^{n-1} \sigma^{-i}\alpha\Big) = \lim_{n\to\infty} \frac{1}{n} \log k^n = \log k \qquad \square$$

We shall define the topological Markov shift in a manner analogous to the usual Markov shift and then compute its entropy.

Definition

A $k \times k$ matrix $A$ with non-negative entries will be called irreducible if for each pair $i,j = 1,2,\ldots,k$ there is some $n$ s.t. $a_{ij}^{(n)} > 0$. $a_{ij}^{(n)}$ is the $(i,j)$-th entry of $A^n$.

Definition

Let $A = (a_{ij})_{i,j=0}^{k-1}$ be a $k \times k$ matrix with entries in $\{0,1\}$. Let $Y = \{0,1,2,\ldots,k-1\}$ have the discrete topology and $Y^Z$ the product topology. Let $X_A \subseteq Y^Z$ be given by $X_A = \{(y_n) \in Y^Z \mid a_{y_n y_{n+1}} = 1\}$. Let $\sigma_A$ be the shift on $X_A$. The system $(\sigma_A, X_A)$ is referred to as a two-sided Topological Markov shift. If we replace $Y^Z$ by $Y^N$ the system $(\sigma_A, X_A)$ is referred to as a one-sided Topological Markov shift.

Recall that a $k \times k$ matrix which is irreducible has by 1.7 a largest positive eigenvalue.

Let $A$ be a $k \times k$ matrix with non-negative entries.

3.22 Proposition

If $(\sigma_A, X_A)$ is a two-sided topological Markov shift and $A$ is irreducible then $h(\sigma_A) = \log \lambda$ where $\lambda$ is the largest positive eigenvalue of $A$.

Proof:

Let $\theta_n(X_A)$ be the number of cylinders $_0[i_0, i_1, \ldots, i_{n-1}]$ which are non-empty subsets of $X_A$. Then $\theta_n(X_A)$ is the number of n-tuples $(i_0, i_1, \ldots, i_{n-1})$ s.t. $a_{i_0 i_1} a_{i_1 i_2} \ldots a_{i_{n-2} i_{n-1}} = 1$. Thus,

$$\theta_n(X_A) = \sum_{i_0, i_1, \ldots, i_{n-1}=0}^{k-1} a_{i_0 i_1} a_{i_1 i_2} \ldots a_{i_{n-2} i_{n-1}}$$
$$= \sum_{i_0, i_{n-1}=0}^{k-1} a_{i_0 i_{n-1}}^{(n-1)} = \sum_{i,j=0}^{k-1} a_{ij}^{(n-1)} .$$

Define a norm on $k \times k$ matrices by:

$$\|(b_{ij})_{i,j=0}^{k}\| = \sum_{i,j=0}^{k-1} |b_{ij}| \quad . \qquad \text{Then,}$$

$$\theta_n(X_A) = \| A^{n-1} \|$$

for every $n$ since the entries of $A$ are non-negative.

The set $\alpha = \{{}_o[0], {}_o[1], \ldots, {}_o[k-1]\}$ is an open cover of $X_A$ in the subspace topology and is also a generator for $\sigma_A$. So, by 3.19

$$h(\sigma_A, \alpha) = h(\sigma_A) \ .$$

$$h(\sigma_A, \alpha) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} \sigma_A^{-i} \alpha) \quad \text{and}$$

$$N(\bigvee_{i=0}^{n-1} \sigma_A^{-i} \alpha) = \theta_n(X_A) \qquad \text{so,}$$

$$h(\sigma_A, \alpha) = \lim_{n\to\infty} \frac{1}{n} \log \theta_n(X_A) = \lim_{n\to\infty} \frac{1}{n} \log \|A^{n-1}\| \ .$$

Now, using the Spectral Radius formula (Chap.1 Sec.7) we have

$$\lim_{n\to\infty} \|A^n\|^{1/n} = \rho(A) = \lambda \ .$$

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^{n-1}\| = \lim_{n\to\infty} (\frac{n-1}{n}) \frac{1}{n-1} \log \|A^{n-1}\|$$

$$= \lim_{n\to\infty} \log \|A^{n-1}\|^{1/n-1} = \log \lambda \ . \qquad \square$$

We can establish results corresponding to 3.21 and 3.22 for the corresponding one-sided shifts. Suppose $Y = \{0,1,\ldots,k-1\}$ and $(T,X)$ is either $(\sigma, Y^N)$ or $(\sigma_A, X_A)$ where $A$ is an irreducible matrix. We observe that the cover $\alpha = \{{}_o[0], {}_o[1], \ldots, {}_o[k-1]\}$ has the property that for any sequence $\{A_n\}_{n=0}^{\infty}$ of members of $\alpha$ the set $\bigcap_{n=0}^{\infty} T^{-n} \overline{A}_n$ has at most one member in $X$. Utilizing the methods in the proofs of 3.13 and 3.19 we may conclude that $h(T) = h(T,\alpha)$. Thus we have:

3.23 <u>Proposition</u>

Let $Y = \{0,1,\dots,k-1\}$ have the power set topology and $Y^N$ the product topology. Then:

1) If $\sigma: Y^N \to Y^N$ is the one-sided shift on $Y^N$ then $h(\sigma) = \log k$

2) If $A$ is an irreducible $k \times k$ matrix and $\sigma_A: X_A \to X_A$ is the one-sided topological Markov shift then $h(\sigma_A) = \log \lambda$ where $\lambda$ is the largest positive eigenvalue of $A$.

We can use the entropy of the shift maps to approximate the entropy of certain maps on the unit interval. This will be demonstrated in Example 2.

Suppose $I$ is the closed unit interval, $\mathcal{B}(I)$ is the class of Borel sets and $\mu$ is any normalized measure defined on $(I, \mathcal{B}(I))$. If $\{x_n\}_{n=0}^{k} \subseteq I$ is a set of points s.t. $0 = x_0 < x_1 < \dots < x_k = 1$ then this set induces the partition $\{[0,x_1),[x_1,x_2),\dots,[x_{k-2},x_{k-1}),[x_{k-1},1)\}$ on $(I, \mathcal{B}(I), \mu)$. Where no ambiguity can arise the set $\{x_n\}_{n=0}^{k}$ will also be referred to as a partition (of I).

<u>Definition</u>

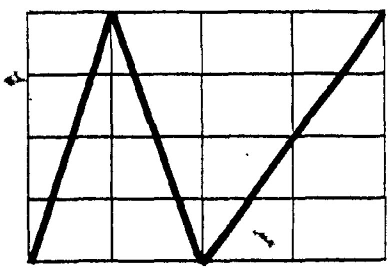
Suppose $T: I \to I$ is continuous and onto. If there is a partition $\{x_n\}_{n=0}^{k}$ of $I$ which contains all the extremal points of $T$ and $T(\{x_n\}) \subseteq \{x_n\}$ then $T$ shall be referred to as a <u>Markov map</u>. If $T$ is also linear on the intervals $[0,x_1],[x_1,x_2],\dots,[x_{k-1},1]$ then $T$ will be referred to as a <u>linear</u> Markov map.

<u>Remark</u>

Suppose $\{x_n\}_{n=0}^{k}$ is a partition of $I$ and $G$ is the set of ordered pairs $\{(x_0,y_0),(x_1,y_1),\dots,(x_k,y_k)\}$ where $y_i \in \{X_n\}$ for each $i$. Then $G$ uniquely determines a linear Markov map $T: I \to I$ where $G$ is a subset of the graph of $T$ .

Example 2

Let $T:[0,1]\to[0,1]$ be a linear Markov map defined on the partition $\{0,\frac{1}{4},\frac{1}{2},\frac{3}{4},1\}$ s.t. $\{(0,0),(\frac{1}{4},1),(\frac{1}{2},0),(\frac{3}{4},\frac{1}{2}),(1,1)\}$ is on the graph of $T$ . Then $T$ has a graph of the form



Let $I_0 = [0,\frac{1}{4}]$ $\quad$ $I_1 = [\frac{1}{4},\frac{1}{2}]$

$I_2 = [\frac{1}{2},\frac{3}{4}]$ $\quad$ $I_3 = [\frac{3}{4},1]$

Let $X = Y^N$ with $Y = \{0,1,2,3\}$ and $X_A \subseteq Y^N$ is all sequences allowed by

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

where $a_{ij} = 1$ if $T(I_i) \supseteq I_j$ and $a_{ij} = 0$ otherwise. Define the map $\psi:X_A\to I$ by $\psi(<x_n>) = \bigcap_{n=0}^{\infty} T^{-n} I_{x_n}$ . By the same argument as in Eg.3 Ch.II $\psi$ is well defined and continuous.

$\psi$ is clearly onto. If $x\in I_i$ then there is some $j$ s.t. $T(x)\in I_j$ and $a_{ij} = 1$ . So there is a sequence $<x_n>\in X_A$ s.t. $T^n(x)\in I_{x_n}$ . Then $\psi(<x_n>) = x$ .

Since $(T \circ \psi)(<x_n>) = (\psi \circ \sigma_A)(<x_n>)$ for every $<x_n> \in X_A$ it follows that $T$ is a factor of $\sigma_A$ . Thus $h(T) \leq h(\sigma_A)$.

Using the Spectral Radius formula gives the value 3 for the largest positive eigenvalue of A . Thus $h(T) \leq \log 3$ .

Using techniques which will be described in the next chapter we will show that $h(T)$ actually equals $\log 3$ .

Section 4. <u>The Variational Principle for Topological Entropy</u>

In 1965 Adler, Konheim and McAndrew [1] conjectured that $h(T) = \sup \{h_\mu(T) \mid \mu \in M(X,T)\}$ . This formula is referred to as the Variational Principle for topological entropy.

In 1969, Goodwyn [11] proved the inequality $h_\mu(T) \leq h(T)$ for every $\mu \in M(X,T)$ . In 1971 Goodman [10] proved the Variational Principle for continuous maps on compact metric spaces. This can be extended to uniformly continuous maps on general metric spaces.

The Variational principle can in fact be used as the definition of topological entropy. This has been done by Brown [6] for continuous maps on compact Hausdorff spaces.

3.24 <u>Theorem</u> ([23], [10])

If $(X,d)$ is a compact metric space and $T:X \to X$ is continuous then $h(T) = \sup \{h_\mu(T) \mid \mu \in M(X,T)\}$ .

Let $(X,d)$ be a compact metric space.

<u>Definition</u>

Suppose $T:X \to X$ is continuous. A point $x$ is referred to as <u>wandering</u> (w.r.t. T) if there is some open set $U$ s.t. $x \in U$ and $U \cap T^{-n} U = \phi$ for every $n \geq 1$ . The <u>non-wandering</u> set for $T$ , denoted $\Omega(T)$ , consists of all points of $X$ which are not wandering. The following is immediate:

3.25 <u>Proposition</u>

Let $T:X\to X$ be continuous. Then:

1) $\Omega(T)$ is closed.

2) $\Omega(T)$ contains all periodic points for $T$.

3.26 <u>Proposition</u>

Let $T:X\to X$ be continuous and $\mu \in M(X,T)$. Then $\mu(\Omega(T)) = 1$.

<u>Proof</u>: ([24], p.157)

$(X,d)$ is a compact metric space and thus has a countable basis for its topology. Let $\{U_n\}_{n=1}^{\infty}$ be such a basis. $\Omega(T)$ is closed so $X - \Omega(T)$ is open. Suppose $U_n \subseteq X - \Omega(T)$. Since $\mu \in M(X,T)$, $\mu(U_n) = \mu(T^{-1} U_n) = \ldots = \mu(T^{-k} U_n) = \ldots$. But $U_n \cap T^{-k} U_n = \phi$ for every $k$. Therefore $\mu(U_n) = 0$ for every $U_n \subseteq X - \Omega(T)$. Since there are at most a countable number of such sets it follows that $\mu(X - \Omega(T)) = 0$. $\therefore$ $\mu(\Omega(T)) = 1$ □

We can use the Variational Principle to show that all of the entropy is carried on the non-wandering set.

3.27 <u>Proposition</u>

Let $T:X\to X$ be continuous. Then, $h(T) = h(T|_{\Omega(T)})$.

<u>Proof</u>:

Since $\Omega(T) \subseteq X$ is a closed set with measure one $M(X,T) = M(\Omega(T), T|_{\Omega(T)})$. Thus for each $\mu \in M(X,T)$, $h_\mu(T) = h_\mu(T|_{\Omega(T)})$. Consequently, $h(T) = \sup\{h_\mu(T) \mid \mu \in M(X,T)\}$

$$= \sup\{h_\mu(T|_{\Omega(T)}) \mid \mu \in M(\Omega(T), T|_{\Omega(T)})\}$$

$$= h(T|_{\Omega(T)}) .$$ □

Using the notion of the non-wandering set we can formulate a weaker type of conjugacy than the usual which still preserves entropy.

Definition

Let $(X_1,d_1)$ and $(X_2,d_2)$ be compact metric spaces. Let $T_i:X_i\to X_i$ be continuous for $i = 1,2$. $T_1$ and $T_2$ are $\Omega$-conjugate if $T_1|_{\Omega(T_1)}$ and $T_2|_{\Omega(T_2)}$ are conjugate.

3.28 Proposition

If $T_1$ and $T_2$ are $\Omega$-conjugate then $h(T_1) = h(T_2)$ .

Proof:

Follows from 3.27. □

It is a question of interest whether and under what circumstances a measure exists which actually maximizes the entropy, i.e. where $h_\mu(T) = h(T)$.

Definition

Let $X$ be a topological space and $(X, \mathcal{B}(X))$ the corresponding measurable space. Let $T:X\to X$ be continuous. If $\mu\in M(X,T)$ satisfies $h_\mu(T) = h(T)$ then $\mu$ will be referred to as maximal. The set of all maximal measures will be denoted $M_{max}(X,T)$ .

Walters [24] and Grillenberger [12] have shown that there are homeomorphisms with $h(T) < \infty$ but $M_{max}(X,T) = \phi$ .

If $(X,d)$ is a compact metric space then $M(X,T)$ is non-empty (by 1.4). It can be shown that if $M(X)$ is given the right topology and $T$ is continuous then $M(X,T)$ is compact ([24], p.152). So, if we define the map $\mu \to h_\mu(T)$ on $M(X,T)$ and show that it is continuous (or even upper semi-continuous) then $\mu \to h_\mu(T)$ achieves a maximum. Thus $M_{max}(X,T) \neq \phi$. This method is used by Misiurewicz [16] to show that maximal measures exist for continuous maps of a certain type (piecewise montone). This will be discussed further in Chapter 4.

For the usual shift on $k$ symbols the maximal measure is the $(\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k})$ Bernoulli measure. (Compare 3.21 with 2.18 and 2.2(ii)).

Parry has shown that maximal measures exist for topological Markov shifts [20]. Let $Y = \{0,1,2,\ldots,k-1\}$ and let $X_A$ be the subset of $Y^Z$ (or $Y^N$) determined by the irreducible $k \times k$ matrix $A$ with entries in $\{0,1\}$. Let $\lambda$ be the largest positive eigenvalue of $A$, $\bar{u} = (u_0, u_1 \ldots u_{k-1})$ a strictly positive left eigenvector for $\lambda$, and

$$\bar{v} = \begin{pmatrix} v_0 \\ v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_{k-1} \end{pmatrix}$$

a strictly positive right eigenvector for $\lambda$ s.t. $\sum_{i=0}^{k-1} u_i v_i = 1$.

$\lambda$, $\bar{u}$ and $\bar{v}$ exist by the Perron-Frobenius Theorem ([9]).

Let $\pi = (u_0 v_0 \;\; u_1 v_1 \ldots u_{k-1} v_{k-1})$ and

$$P = (p_{ij})_{i,j=0}^{k-1}$$ where

$$p_{ij} = \frac{a_{ij} \; v_j}{\lambda \, v_i} .$$

Then the pair $(\pi, P)$ determines a Markov measure.

3.29 <u>Proposition</u> ([20], p.55 or [24] p.194)

Let $Y = \{0,1,2,\ldots,k-1\}$. Let $A$ be an irreducible $k \times k$ matrix with entries in $\{0,1\}$ and $\lambda$ its largest positive eigenvalue. Suppose $\bar{u}$ and $\bar{v}$ are strictly positive left and right eigenvectors of $A - \lambda I$ s.t. $\bar{u} \cdot \bar{v} = 1$. Let $\pi = (u_0 v_0, u_1 v_1, \ldots, u_{k-1} v_{k-1})$ and $P = (\frac{a_{ij} v_j}{\lambda \, v_i})_{i,j=0}^{k-1}$. Then the measure $\mu_{\pi P}$ is the unique maximal measure for $\sigma_A$.

Definition

The measure $\mu_{\pi P} \in M(X_A, \sigma_A)$ in 3.29 above will be referred to as the Parry measure.

Section 5. The Classification of Maps

Topological entropy, like its measure theoretic counterpart, is an invariant under isomorphism (i.e. conjugacy). The question of what additional assumptions must be introduced or what weaker forms of conjugacy might exist for which topological entropy is a complete invariant has only been answered in part.

It is clear that topological entropy is not a complete invariant for either conjugacy or $\Omega$-conjugacy of maps on the unit interval as the following example shows.

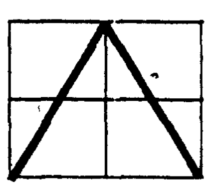
Example 3

Consider the two maps $T_1 : I \to I$ and $T_2 : I \to I$ given by:

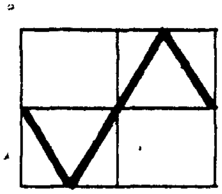
$$T_1(x) = \begin{cases} 2x & 0 \leq x \leq \frac{1}{2} \\ 2-2x & \frac{1}{2} \leq x \leq 1 \end{cases}$$

$$T_2(x) = \begin{cases} \frac{1}{2}-2x & 0 \leq x \leq \frac{1}{4} \\ 2x-\frac{1}{2} & \frac{1}{4} \leq x \leq \frac{3}{4} \\ \frac{3}{2}-2x & \frac{3}{4} \leq x \leq 1 \end{cases}$$

Their graphs are as follows:



$T_1$



$T_2$

From Eg.1 Ch.III $h(T_1) = 1$ and by means of 3.1 $h(T_2) = 1$ . $T_1$ has two fixed points but $T_2$ has three. Any conjugating homeomorphism must map fixed points to fixed points in a one-one manner. Thus $T_1$ and $T_2$ are neither conjugate nor $\Omega$-conjugate.

For topological Markov shifts Adler and Marcus [2] have shown that the entropy together with a quantity called the ergodic period is a complete set of invariants for a weakened form of conjugacy called almost topological conjugacy.

# CHAPTER IV

## ENTROPY OF PIECE-WISE MONOTONIC CONTINUOUS MAPS

### Section 1. Introduction

Many real valued functions with domain in the reals have the property of being monotone on a finite number of intervals whose union is the domain. When such maps are also continuous it is possible to readily compute good lower bounds for their entropy and in some cases (E.g. Linear Markov maps) to compute the entropy exactly.

The source material for this chapter is derived from two papers: Misiurewicz [16] and Misiurewicz and Szlenk [17].

### Section 2. T-mono Covers

We begin this section with some definitions from [16] and [17].

Let X be a compact Hausdorff space, $T:X\to X$ a continuous map and $\mathcal{U}(X)$ the set of all finite open covers of X. Let $\alpha,\beta$ be finite covers of X (not necessarily open). Let

$$\alpha^n = \bigvee_{i=0}^{n-1} T^{-i}\alpha = \{A_0 \cap T^{-1}A_1 \cap \dots \cap T^{-(n-1)}A_{n-1} \mid A_i \in \alpha\}$$

Definition

$$N(Y,\alpha) = \min\{\text{Card } C \mid C \subseteq \alpha,\ Y \subseteq \bigcup_{C \in C} C\}$$

$$N(\alpha\mid\beta) = \max_{B\in\beta} N(B,\alpha)$$

$$h(T,\alpha\mid\beta) = \lim_{n\to\infty}\frac{1}{n}\log N(\alpha^n\mid\beta^n)$$

$$H(T\mid\beta) = \sup\{h(T,\alpha\mid\beta)\mid\alpha\in\mathcal{U}(X)\}$$

$$h(T,\alpha) = h(T,\alpha\mid\beta) \quad \text{for} \quad \beta = \{X\}\ .$$

Note

If $\alpha \in \mathcal{U}(X)$ and $\beta = \{X\}$ then $N(\alpha|\beta) = N(\alpha)$ and $h(T,\alpha)$ has the same meaning as that assigned to it in the previous chapter.

Let $T:X \to X$ be continuous.

4.1 Proposition ([16]

Let $\alpha, \beta$ be finite covers of $X$.

Then: i) $h(T,\alpha|\beta) \leq \log N(\alpha|\beta)$

ii) $h(T,\alpha) \leq h(T,\beta) + h(T,\alpha|\beta)$

iii) $h(T) \leq h(T,\alpha) + h(T|\alpha)$.

Denote by $\mathcal{I}$ the set of all subintervals (open, closed, half-open, degenerate) of $I = [0,1]$. Let $T:I \to I$ be continuous.

Definition

A cover $\alpha$ of $X$ is called T-mono if $\alpha$ is finite, $\alpha \subseteq \mathcal{I}$, and for any $A \in \alpha$ the map $T|_A$ is monotone.

4.2 Proposition

If $T_1:I \to I$ and $T_2:I \to I$ are continuous, $\alpha$ is a $T_1$-mono cover and $\beta$ is a $T_2$-mono cover then $\alpha \vee T_1^{-1}\beta$ is a $(T_2 \circ T_1)$-mono cover.

Proof:

$\alpha \vee T_1^{-1}\beta = \{A \cap T_1^{-1}B \mid A \in \alpha, B \in \beta\}$ is finite. Let $A \in \alpha$ and $B \in \beta$. $(T_2 \circ T_1)|_{A \cap T_1^{-1}}$ is monotone since $A \cap T_1^{-1} B \subseteq A$, where $T_1$ is monotone, $T_1(A \cap T_1^{-1} B) \subseteq B$ where $T_2$ is monotone, and the composition of monotone maps is monotone. $A \cap T_1^{-1} B \in \mathcal{I}$ since $A \cap T_1^{-1} B = (T_1|_A)^{-1}(B)$, $T_1|_A$ is monotone and $B \in \mathcal{I}$. □

Corollary

If $\alpha$ is a T-mono cover then $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ is a $T^n$-mono cover.

4.3 Proposition

Let $T:I\to I$ be continuous and $\alpha \subseteq I$ a finite cover of $X$ . Then there is some $\beta \in \mathcal{U}(X)$ s.t. $h(T,\alpha|\beta) \leq \log 3$ .

Proof:

From 4.1(i) it is sufficient to find $\beta$ s.t. $N(\alpha|\beta) \leq 3$. Construct the cover $\beta$ by initially taking the interiors of the members of $\alpha$.

Let $\delta = \{d(A)|A$ a non-degenerate member of $\alpha\}$. Complete the cover by taking small open intervals with diameter less than $\delta/2$ . At most 3 members of $\alpha$ are required to cover any one member of $\beta$. Thus $N(\alpha/\beta) \leq 3$ . □

Definition

A map $T:I\to I$ is called piecewise monotone if there exists a T-mono cover of $X$ .

4.4 Proposition ([17], pp.47-48)

Let $T:I\to I$ be a piecewise monotone continuous map (p.m.c.); $\alpha$ a T-mono cover, and $\beta \subseteq I$ a finite cover. Then $h(T,\beta/\alpha) = 0$ .

Corollary

If $\alpha$ is a T-mono cover then $h(T|\alpha) = 0$ .

4.5 Theorem ([17])

If $T:I\ I$ is a p.m.c. map then $h(T) = \lim_{n\to\infty} \frac{1}{n} \log C_n$ where $C_n = \min\{\text{Card } \alpha | \alpha \text{ a } T^n\text{-mono cover}\}$.

Proof:

Let $\alpha_n$ be a $T^n$-mono cover of minimum cordinality $(n \geq 1)$. Let $m$ and $k$ be fixed positive integers. By 4.2 the cover $\alpha_k \vee T^{-k}\alpha_m$ is a $T^{m+k}$-mono cover. Thus:

$$C_{m+k} \leq \mathrm{Card}(\alpha_k \vee T^{-k}\alpha_m) \leq C_m C_k$$

and $\log C_{m+k} \leq \log C_m + \log C_k$ . By 2.5 $\lim_{n\to\infty} \frac{\log C_n}{n}$ exists.

$$h(T) = \frac{1}{n} h(T^n) \qquad \text{by } 3.12$$

$$h(T^n) \leq h(T^n, \alpha_n) + h(T^n | \alpha_n) \qquad \text{by } 4.1(iii)$$

$$h(T^n | \alpha_n) = 0 \qquad \text{by } 4.4 \text{ Cor.}$$

Thus, $h(T) = \frac{1}{n} h(T^n) \leq \frac{1}{n} h(T^n, \alpha_n)$ .

Since $h(T^n, \alpha_n) = \lim_{n\to\infty} \frac{1}{k} \log N(\alpha_n{}^k)$ and

$N(\alpha_n{}^k) \leq (N(\alpha_n))^k$ it follows that

$$h(T^n, \alpha_n) \leq \log N(\alpha_n) = \log \mathrm{Card}\ \alpha_n .$$

Therefore, $h(T) \leq \frac{1}{n} h(T^n, \alpha_n) \leq \frac{1}{n} \log \mathrm{Card}\ \alpha_n = \frac{1}{n} \log C_n$ .

$\therefore$ $h(T) \leq \lim_{n\to\infty} \frac{1}{n} \log C_n$ .

By 4.3 there is some $\beta_n \in \mathcal{U}(X)$ s.t. $h(T^n, \alpha_n | \beta_n) \leq \log 3$ for every $n$ .

$$C_{nk} \leq N\Big(\bigvee_{j=0}^{k-1} T^{-jn} \alpha_n\Big) \quad \text{since} \quad \bigvee_{j=0}^{k-1} T^{-jn} \alpha_n$$

is a $T^{nk}$–mono* cover.

$$h(T^n, \alpha_n) \leq h(T^n, \beta_n) + h(T^n, \alpha_n | \beta_n) \qquad \text{by } 4.1(iii).$$

Thus, $\lim_{k\to\infty} \frac{1}{k} \log C_k = \lim_{k\to\infty} \frac{1}{nk} \log C_{nk}$

$$\leq \lim_{k\to\infty} \frac{1}{nk} \log N\Big(\bigvee_{j=0}^{k-1} T^{-jn} \alpha_n\Big) = \frac{1}{n} h(T_n, \alpha_n)$$

$$\leq \frac{1}{n} h(T^n, \beta_n) + \frac{1}{n} h(T^n, \alpha_n | \beta_n) \leq \frac{1}{n} h(T^n) + \frac{1}{n} \log 3$$

$$= h(T) + \frac{1}{n}\log 3$$

$\therefore \quad \lim_{k\to\infty} \frac{1}{k} \log C_k \leq h(T)$

$\therefore \quad h(T) \leq \lim_{n\to\infty} \frac{1}{n} \log C_n \leq h(T)$ □

Corollary

If $\alpha$ is a T-mono cover then $h(T,\alpha) = h(T)$.

Proof:

$$h(T) = \lim_{k\to\infty} \frac{1}{nk} \log C_{nk} \leq \lim_{k\to\infty} \frac{1}{nk} \log N\left(\bigvee_{j=0}^{k-1} T^{-nj}\alpha^n\right)$$

$$= h(T,\alpha) = \frac{1}{n} h(T^n,\alpha^n) \leq h(T) + \frac{1}{n}\log 3$$ □

Definition

Let $\varepsilon > 0$ and $n \in N$. A set $E \subseteq I$ is said to be $(n,\varepsilon)$ separated (w.r.t. T) if for each pair $x,y \in E (x \neq y)$ $d(T^k(x),T^k(y)) > \varepsilon$ for some $k$ s.t. $0 \leq k \leq n$.

4.6 Theorem ([17], p.49)

Let $T: I \to I$ be p.m.c. and onto. Then

$$h(T) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{Var}(T^n).$$

Proof:

Since $T$ is p.m.c. and onto there is a monotone $g: I \to I$ s.t. $T \circ g = \mathrm{id}$. Then $\mathrm{Var}(T^n) = \mathrm{Var}(T^{n+1} \circ g)$ for $n \geq 0$. But $\mathrm{Var}(T^{n+1} \circ g) \leq \mathrm{Var}(T^{n+1})$ since $\mathrm{Range}(g) \subseteq I$. Thus $\mathrm{Var}(T^n) \leq \mathrm{Var}(T^{n+1})$ for every $n \geq 1$.

Let $\alpha \in \mathcal{U}(X)$ and $0 < 4\varepsilon < \lambda(\alpha)$. Choose a maximal $(n,\varepsilon)$-separated set $\{x_1, x_2, \ldots, x_s\}$ where $x_i \in I$ for each $i$ and $x_i < x_j$ for $i < j$. Since this set is $(n,\varepsilon)$-separated, for each $i \in \{1,2,\ldots,s-1\}$

there is some $k_i \in \{0,1,\dots,n-1\}$ s.t. $|T^{k_i}(x_i) - T^{k_i}(X_{i+1})| > \varepsilon$ .

Consequently,

$$(s-1)\varepsilon \leq \sum_{k=0}^{n-1} \mathrm{Var}(T^k) \leq n\,\mathrm{Var}(T^n) \ .$$

Consider the set

$$B_i(k,\varepsilon) = \{x \in I : |T^k(x) - T^k(x_i)| < 2\varepsilon\}$$

$$B_i = \bigcap_{k=0}^{n-1} B_i(k,\varepsilon)$$

The collection $\{B_i | i = 1,2,\dots,s\}$ forms a cover of $I$, otherwise $\{x_1,x_2,\dots,x_s\}$ is not maximal. Each set of the form $\{y : |y - T^k(x_i)| < 2\varepsilon\}$ has diameter strictly less than $4\varepsilon$ . So, there is some $A_k \in \alpha$ s.t. $\{y : |y - T^k(x_i)| < 2\varepsilon\} \subseteq A_k$ (By defin. of $\lambda(\alpha)$). And,

$$\begin{aligned} B_i(k,\varepsilon) &= \{x : |T^k(x) - T^k(x_i)| < 2\varepsilon\} \\ &= T^{-k}\{y : |y - T^k(x_i)| < 2\varepsilon\} \\ &\subseteq T^{-k} A_k \end{aligned}$$

Thus, $\bigcap_{k=0}^{n-1} B_i(k,\varepsilon) \subseteq A_0 \cap T^{-1} A_1 \cap \dots \cap T^{-(n-1)} A_{n-1}$ .

Therefore $N(\alpha^n) \leq s$ . Consequently,

$\frac{\varepsilon}{n}(N(\alpha^n) - 1) \leq \mathrm{Var}(T^n)$ And,

$\varliminf_{n\to\infty} \frac{1}{n} \log \frac{\varepsilon}{n}(N(\alpha^n) - 1) = h(T,\alpha)$. Thus

$$h(T,\alpha) \leq \varliminf_{n\to\infty} \frac{1}{n} \log \mathrm{Var}(T^n) \ .$$

Since $\alpha \in \mathcal{U}(X)$ is arbitrary, $h(T) \leq \varliminf_{n\to\infty} \frac{1}{n} \log \mathrm{Var}(T^n)$ . Now to establish the reverse inequality we note that $\mathrm{Var}(T^n) \leq C_n'$ with equality only when $T^n$ is surjective on each member of a $T^n$-mono cover of minimum cordinality. Using 4.5

$$\varliminf_{n\to\infty} \frac{1}{n} \log \mathrm{Var}(T^n) \leq \varliminf_{n\to\infty} \frac{1}{n} \log C_n = h(T) \qquad \square$$

4.7 Proposition

Suppose $T:I\to I$ is p.m.c. and $\alpha = \{[x_0,x_1],[x_1,x_2],\ldots,[x_{s-1},x_s]\}$ is a T-mono cover with $0 = x_0 \le x_1 \le \ldots \le x_s = 1$. Then

$$\mathrm{Var}(T) = \sum_{n=0}^{s-1} |T(x_{n+1}) - T(x_n)|$$

Proof:

Let $\{z_n\}_{n=0}^{k}$ be an arbitrary partition of $I$ with $0 = z_0 < z_1 < \ldots < z_{k-1} < z_k = 1$. Let $\{y_n\}_{n=0}^{\ell} = \{z_n\}_{n=0}^{k} \cup \{x_n\}_{n=0}^{s}$ where $0 = y_0 \le y_1 \le y_2 \le \cdots \le y_{\ell-1} \le y_\ell = 1$. $\{z_n\}_{n=0}^{k} \subseteq \{y_n\}_{n=0}^{\ell}$ so,

$$\sum_{n=0}^{k-1} |T(z_{n+1}) - T(z_n)| \le \sum_{n=0}^{\ell-1} |T(y_{n+1}) - T(y_n)| \,.$$ Since $\{x_n\}_{n=0}^{s} \subseteq \{y_n\}_{n=0}^{\ell}$ and $T$ is monotone on each interval $[x_n, x_{n+1}]$ we have

$$\sum_{n=0}^{s-1} |T(x_{n+1}) - T(x_n)| = \sum_{n=0}^{\ell-1} |T(y_{n+1}) - T(y_n)| \,.$$ Therefore

$$\sum_{n=0}^{k-1} |T(z_{n+1}) - T(z_n)| \le \sum_{n=0}^{s-1} |T(x_{n+1}) - T(x_n)|$$ □

4.6 and 4.7 enable us to compute the topological entropy of the following examples.

Example 1

Let $T:I\to I$ be given by

$$T(x) = \begin{cases} 2x & x \in [0,\frac{1}{2}] \\ 2-2x & x \in [\frac{1}{2},1] \end{cases}$$

$\alpha = \{[0,\frac{1}{2}],[\frac{1}{2},1]\}$ is a T-mono cover thus $\bigvee_{i=0}^{n-1} T^{-i}\alpha = \alpha^n$ is a $T^n$-mono cover. $\alpha^n$ consists of closed non-degenerate intervals. Each member of $\alpha^n$ is mapped homeomorphically onto $I$ by $T^n$. Thus $\mathrm{Var}(T^n) = N(\alpha^n) = 2^n$. Consequently, $h(T) = \lim_{n\to\infty} \frac{1}{n} \log 2^n = 1$.

Example 2

Let $T: I \to I$ be given by $T(x) = 4x(1-x)$. Let $\alpha$ be as in the previous example. In this case also each member of $\alpha^n$ is mapped homeomorphically onto $I$. Thus, $Var(T^n) = N(\alpha^n) = 2^n$. Thus $h(T) = 1$.

4.8 Proposition

Suppose $T: I \to I$ is p.m.c. and onto and $T'(x)$ exists at all but a finite number of points. If $|T'(x)| = k > 1$ then $h(T) = \log k$.

Proof:

Let $x_1 < x_2 < \ldots < x_s$ be the points where $T'(x)$ does not exist. Then, $\alpha = \{[0,x_1], [x_1,x_2], \ldots, [x_s,1]\}$ is a T-mono cover since $T$ is increasing or decreasing on each member. Let $\alpha^n$ be given by $\{[y_o,y_1], [y_1,y_2], \ldots, [y_{m-1},y_m]\}$ where $0 = y_0 \le y_1 \le y_2 \le \ldots \le y_m = 1$. Then $Var(T^n) = \sum_{j=0}^{m-1} |T^n(y_{j+1}) - T^n(y_j)|$. If $x \in (y_j, y_{j+1})$ then $(T^n)'(x)$ exists, since $x$ remains in the interiors of members of $\alpha$ under each iteration of $T$. Now, by the chain rule $|(T^n)'(x)| = k^n$. Thus, $|T^n(y_{j+1}) - T^n(y_j)| = k^n(y_{j+1} - y_j)$. So,

$$Var(T^n) = \sum_{j=0}^{m-1} k^n(y_{j+1} - y_j) = k^n . \quad \text{Consequently,}$$

$$h(T) = \lim_{n\to\infty} \frac{1}{n} \log k^n = \log k . \qquad \square$$

Example 3

Let $T: I \to I$ be given by:

$$T(x) = \begin{cases} \frac{1}{p} x & x \in [0,p] \quad 0<p<1 \\ 2 - \frac{1}{p} x & x \in [p,1] \end{cases}$$

Then $|T'(x)| = \frac{1}{p}$. $h(T) = \log \frac{1}{p} = -\log p$.

Section 3. Linear Markov Maps

This section will be devoted to applying the Corollary to 4.5 in order to compute the entropy of linear Markov maps.

4.9 Proposition

Let $T:I\to I$ be a linear Markov map defined s.t. $T$ is linear on each of the intervals $I_n = [x_n, x_{n+1}]$ where $0 = x_0 < x_1 < \ldots < x_k = 1$. Let $A = (a_{ij})_{i,j=0}^{k-1}$ be the $k \times k$ matrix defined by

$$a_{ij} = \begin{cases} 1 & \text{if } T(I_i) \supseteq I_j \\ 0 & \text{otherwise} \end{cases}.$$

If $A$ is irreducible then $h(T) = \log\lambda$ where $\lambda$ is the largest positive eigenvalue of $A$.

Proof:

$\alpha = \{I_0, I_1, \ldots, I_{k-1}\}$ is a T-mono cover so $h(T) = h(T,\alpha)$ by means of 4.5 Cor. Denote the sum of the entries of $A^n$ by $\Sigma A^n$. By the Spectral Radius formula (see Chap. 1. Sec. 6)

$$\lim_{n\to\infty} \frac{1}{n} \log \Sigma A^n = \log\lambda \quad \text{since } \Sigma A^n \text{ is a norm for } A^n.$$

So, it is sufficient to show $N(\alpha^n) = \Sigma A^n$.

Denote $I_{i_0} \cap T^{-1} I_{i_1} \cap \ldots \cap T^{-(n-1)} I_{i_{n-1}}$ by $(i_0, i_1, \ldots, i_{n-1})$ and $I^0_{i_0} \cap T^{-1} I^0_{i_1} \cap \ldots \cap T^{-(n-1)} I^0_{i_{n-1}}$ by $(i^0_0, i^0_1, \ldots, i^0_{n-1})$.

Suppose $(i_0, i_1, \ldots, i_{n-1}) \in \alpha^n$ is essential for $\alpha^n$ to cover $I$. Then there is some $x \in (i_0, i_1, \ldots, i_{n-1})$ which is not in any other member of $\alpha^n$. Thus $T^p(x) \in I^0_{i_p}$ for $0 \le p \le n-1$. An interior point of $I_i$ is mapped to an interior point of $I_j$ only if $a_{ij} = 1$. Thus $(i_0, i_1, \ldots, i_{n-1}) \in \alpha^n$ essential $\Rightarrow a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i_{n-2} i_{n-1}} = 1$.

Now, suppose $a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i_{n-2} i_{n-1}} = 1$. Since $T$ is finite-to-one there must be at least one interior point of $I_{i_0}$ which never hits an end point for any iteration of $T$ up to $n-1$. Thus $(i_0^0, i_1^0, \ldots, i_{n-1}^0) \neq \phi$ which implies $(i_0, i_1, \ldots, i_{n-1}) \in \alpha^n$ is essential. So, we have $(i_0, i_1, \ldots, i_{n-1}) \in \alpha^n$ is essential iff $a_{i_0 i_1} a_{i_1 i_2} \ldots a_{i_{n-2} i_{n-1}} = 1$. Therefore,

$$N(\alpha^n) = \text{Card}\,\{(i_0, i_1, \ldots, i_{n-1}) \mid a_{i_0 i_1}\, a_{i_1 i_2} \ldots a_{i_{n-2} i_{n-1}} = 1\}.$$

Now, $$\Sigma A^n = \sum_{i_0, i_1, \ldots, i_{n-1} = 0}^{k-1} a_{i_0 i_1}\, a_{i_1 i_2} \ldots a_{i_{n-2} i_{n-1}}$$

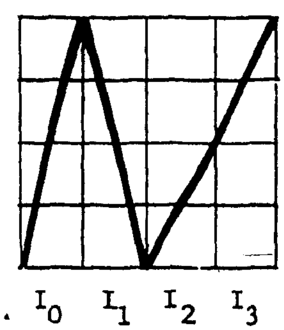
$$= \text{Card}\,\{(i_0, i_1, \ldots, i_{n-1}) \mid a_{i_0 i_1}\, a_{i_1 i_2} \ldots a_{i_{n-2} i_{n-1}} = 1\}.$$

Therefore $N(\alpha^n) = \Sigma A^n$. □

We now return to the map first introduced in Example 2 Ch. III.

<u>Example 3</u>

Let $T: I \to I$ be the linear Markov map defined in Example 2 Ch. III. $T$ has graph:




The matrix $A$ associated with $T$ is

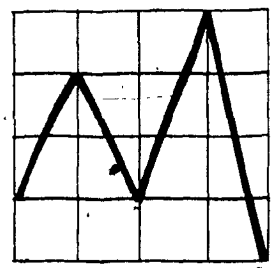
$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

which is irreducible. The Spectral Radius formula gives $\log \lambda = \log 3$.

Thus $h(T) = \log 3$ by **4.9**.

Example 4

Let $T:I\to I$ be the linear Markov map defined on $\{0,\frac{1}{4},\frac{1}{2},\frac{3}{4},1\}$ s.t. the set $\{(0,\frac{1}{4}),(\frac{1}{4},\frac{3}{4}),(\frac{1}{2},\frac{1}{4}),(\frac{3}{4},1),(1,0)\}$ lies on the graph of $T$. Thus $T$ has graph:



The matrix $A$ associated with $T$ is

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

which is irreducible. $x \approx 2.769$ is the largest positive eigenvalue of $A$. Thus $h(T) \approx \log 2.769$.

Section 4. Topological Entropy and Periodic Points

It was shown in the previous chapter that for $(X,d)$ a compact metric space and $T:X\to X$ continuous $h(T|_{\Omega(T)}) = h(T)$. We know that $\Omega(T)$ contains all periodic points so we might suspect a relation between periodic points and entropy. For p.m.c. maps the number of periodic points induces an upper bound for the entropy.

4.10 Theorem ([17], p.45)

If $T:I\to I$ is p.m.c. then $h(T) \le \overline{\lim_{n\to\infty}} \frac{1}{n} \log \mathrm{Card}\{x\in I \mid T^n(x) = x\}$

4.11 Proposition

If $T:I\to I$ is a polynomial of degree $k$ then $h(T) \le \log k$.

Proof:

If $T$ has degree $k$ then $T^n$ has degree $\leq k^n$. Thus $T^n(x) - x = 0$ has at most $k^n$ real roots. Thus

$$h(T) \leq \lim_{n\to\infty} \frac{1}{n} \log k^n = \log k \quad . \qquad \square$$

Section 5. P.M.C. Maps and Maximal Measures

Definition

Let $X$ be a compact Hausdorff space and $T:X\to X$ continuous. $T$ will be referred to as h-expansive if there is some $\alpha \in \mathcal{U}(X)$ s.t. $h(T|\alpha) = 0$.

4.12 Theorem ([16])

If $X$ is a compact Hausdorff space and $T$ is h-expansive then there is a measure $\mu \in M(X,T)$ s.t. $h_\mu(T) = h(T)$.

Corollary

If $T:I\to I$ is p.m.c. then $T$ has a maximal measure.

Proof:

By 4.4 Cor. $\qquad h(T|\alpha) = 0 \qquad \square$

We can use the Parry measure on topological Markov shifts to induce a measure on the unit interval which is maximal for linear Markov maps. The following definition and propositions will assist in establishing such a result.

Definition

Let $(X, \mathcal{B}, \mu)$ be a measure space and $T:X\to X$ measure preserving. $T$ is referred to as ergodic (w.r.t. $\mu$) if for every $B \in \mathcal{B}$, $T^{-1}B = B$ implies $\mu(B) = 0$ or $1$.

4.13 Proposition ([24])

If $\sigma$ is the one-sided or two-sided $(\pi, P)$ Markov shift then

$\sigma$ is ergodic (w.r.t.$\mu_{\pi P}$) iff $P$ is irreducible.

4.14 <u>Proposition</u> ([24])

If $\sigma$ is the one-sided or two-sided $(\pi,P)$ Markov shift and is ergodic then almost every sequence (w.r.t.$\mu_{\pi P}$) has a dense orbit under $\sigma$.

4.15 <u>Proposition</u>

Suppose $T:I\to I$ is a linear Markov map with $|T'(x)|>1$ defined with respect to the partition set $0 = x_0 < x_1 < x_2 < \ldots < x_{k-1} < x_k = 1$ and $A$ is the $k\times k$ matrix associated with $T$ as in 4.9. Suppose $A$ is irreducible and let $Y = \{0,1,2,\ldots,k-1\}$ and $X_A \subseteq Y^N$ be the subset of allowable sequences determined by $A$ (as in 4.9). Let $\mu_{\pi P}\in M(X_A,\sigma_A)$ be the Parry measure. Then there is a measure $\nu\in M(I,T)$ s.t. i) The systems $(\sigma_A, X_A, \mu_{\pi P})$ and $(T,I,\nu)$ are isomorphic.

ii) $\nu\in M_{max}(I,T)$.

<u>Proof</u>:

Let $I_0 = [x_0,x_1]$, $I_1 = [x_1,x_2],\ldots,I_{k-1} = [x_{k-1},x_k]$. Define the map $\psi:X_A\to I$ by $\psi(<x_n>) = \bigcap_{n=0}^{\infty} T^{-n} I_{x_n}$. As in Eg.3 Ch.II $\psi$ can be shown to be well defined and continuous.

Let $D = \bigcup_{n=0}^{\infty} T^{-n}(\{0,x_1,\ldots,x_{k-1},1\})$ and let $M_2 = I\setminus D$ and $M_1 = X_A\setminus\psi^{-1}D$. Then it can be shown as in Eg.3 Ch.II that $\psi|_{M_1}:M_1\to M_2$ is a homeomorphism.

No member of $D$ has a dense orbit in $I$ since each member is eventually fixed or periodic. It is clear that $\psi\circ\sigma_A = T\circ\psi$ for each $<x_n>\in M_1$. Since $\psi$ is continuous, no member of $\psi^{-1}(D)$ has a dense orbit in $X_A$. $A$ is irreducible so $P$ is also and thus by 4.13 $\sigma_A$ is ergodic w.r.t. $\mu_{\pi P}$. So, using 4.14, $\mu_{\pi P}(M_1) = 1$.

Since $\psi|_{M_1}$ is a homeomorphism it maps Borel sets to Borel sets.

Define the measure $\nu$ on $(I, \mathcal{B}(I))$ by $\nu(B) = \mu_{\pi P}(\psi^{-1}B)$ for each $B \in \mathcal{B}(I)$. With the induced measure $\nu$ defined on $(I, \mathcal{B}(I))$ the map $\psi|_{M_1}$ is a measure theoretic isomorphism between $(\sigma_A, X_A, \mu_{\pi P})$ and $(T, I, \nu)$. Thus (i) is proved.

By (i) $h_{\mu_{\pi P}}(\sigma_A) = h_\nu(T)$. Since $\psi$ is continuous and onto $h(\sigma_A) \geq h(T)$ by 3.11. Since $\mu_{\pi P}$ is a maximal measure for $\sigma_A$, $h(\sigma_A) = h_{\mu_{\pi P}}(\sigma_A)$. Thus $h_\nu(T) \geq h(T)$. So $\nu \in M_{max}(X, T)$ □

CHAPTER V

# PERIODIC POINTS AND ENTROPY OF CONTINUOUS MAPS

## Section 1. Introduction

We have seen in the previous chapter that the number of periodic points enables us to establish an upper bound for the entropy of p.m.c. maps. In this chapter it will be shown that periodic points allow us to find lower bounds for the entropy of general continuous maps.

The material in this chapter is derived primarily from the paper by Block, Guckenheimer, Misiurewicz and Young [4] as well as those by Misiurewicz and Sylenk [17] and Nitecki [18].

## Section 2. Maps and Graphs

Let $I$ be a closed interval in $\mathbb{R}$. Throughout this chapter a partition of $I$ will refer to a closed cover by subintervals of $I$ whose interiors are pairwise disjoint.

Let $\alpha = \{I_1, I_2, \ldots, I_s\}$ be a fixed partition of $I$ and $T:I \to I$ a fixed continuous map.

**Definition**

Suppose $H, J$ are subintervals of $I$. $H$ T-covers $J$ if there is some subinterval $K \subseteq H$ s.t. $f(K) = J$. $H$ T-covers $J$ n-times if there are $n$ subintervals $K_1, K_2, \ldots, K_n \subseteq H$ with pairwise disjoint interiors s.t. $T(K_i) = J$ for $i = 1, 2, \ldots, n$.

**5.1 Proposition**

Suppose $H$ and $J$ are subintervals of $I$ and $H$ T-covers $J$. Then $T(H) \supseteq J$.

**5.2 Proposition**

Suppose $H$ and $J$ are subintervals of $I$ with $J$ closed. If

$T(H) \supseteq J$ then $H$, T-covers $J$.

Proof:

Suppose $J$ is non-degenerate (otherwise the situation is trivial). Since $T(H) \supseteq J$ there are points $x_0, x_1 \in H$ $(x_0 \neq x_1)$ s.t. $T(x_0) \leq \min J$ and $T(x_1) \geq \max J$. Suppose $x_0 < x_1$. Define

$$\beta = \inf\{x \mid x_0 < x < x_1,\ T(x) \geq \max J\}$$

$$\alpha = \sup\{x \mid x_0 < x < \beta,\ T(x) \leq \min J\}\ .$$

By continuity of T: $\alpha \neq \beta$, $T(\alpha) = \min J$ and $T(\beta) = \max J$.

If $y \in J$ then by the Intermediate Value Theorem there is some $x \in [\alpha, \beta]$ s.t. $T(x) = y$. And, by definition of $\alpha$ and $\beta$, $x \in [\alpha, \beta]$ implies $T(x) \in J$. Therefore, $T([\alpha, \beta]) = J$.

The same result is obtained if $x_1 < x_0$ by defining $\alpha, \beta$ as follows.

Let
$$\beta = \inf\{x \mid x_1 < x < x_1,\ T(x) \leq \min J\}$$

$$\alpha = \sup\{x \mid x_1 < x < \beta,\ T(x) \geq \max J\}$$ □

Definition

The α-graph of T is an oriented graph with vertices $I_1, I_2, \ldots I_s$ s.t. if $I_i$ T-covers $I_j$ exactly n times then there are n arrows from $I_i$ to $I_j$.

Definition

Let the matrix $M_T = (m_{ij})_{i,j=1}^{s}$ be defined by:

$m_{ij}$ = number of arrows from $I_i$ to $I_j$ in the α-graph of T.

Definition

Let G denote the α-graph of T. The entropy of G, denoted h(G), is given by: $h(G) = \log \lambda$ where $\lambda$ is the spectral radius of $M_T$. (if $M_T$ is irreducible then $\lambda$ is the largest positive eigenvalue of $M_T$).

In [4] the authors present a lemma which states that $h(G) \leq h(T)$. A weaker form of this lemma will be proved subsequently.

The following fixed point theorem and its corollary will be used repeatedly.

5.3 Theorem

Let $J \subseteq I$ be a closed interval and $T: J \to I$ be continuous. If there are points $u, v \in J$ s.t. $T(u) \leq u$ and $T(v) \geq v$ then there is a fixed point for $T$ in $J$.

Proof:

If $T(u) = u$ or $T(v) = v$ then the existence is obvious so assume $T(u) < u$ and $T(v) > v$. Since $T$ is continuous on $J$ $T(x) - x$ is also. $T(x) - x < 0$ at $x = u$ and $T(x) - x > 0$ at $x = v$. By the Intermediate Value theorem there is an $x_0$ between $u$ and $v$ s.t. $T(x_0) - x_0 = 0$ □

Corollary

If $T: I \to I$ is continuous then $T$ has a fixed point in $I$.

Proof:

Recall that $I = [a,b]$ is a closed interval in $\mathbb{R}$. $T(a) \geq a$ and $T(b) \geq b$. □

The following lemma will be useful for proving the existence of periodic points of various orders by reference to the $\alpha$-graph of $T$.

5.4 Lemma

If $I_{a_0} \to I_{a_1} \to I_{a_2} \to \ldots \to I_{a_{n-1}} \to I_{a_n} = I_{a_0}$ is a loop in the $\alpha$-graph of $T$ then there is some $x \in I_{a_0}$ s.t. $T^n(x) = x$ and $T^i(x) \in I_{a_i}$ for $i = 0, 1, \ldots, n$.

Proof:

$I_{a_{n-1}} \to I_{a_n}$ so $I_{a_{n-1}}$ T-covers $I_{a_n}$. By definition there is a closed interval $K_1 \subseteq I_{a_{n-1}}$ s.t. $T(K_1) = I_{a_n}$. $I_{a_{n-2}} \to I_{a_{n-1}}$ so $I_{a_{n-2}}$ T-covers $I_{a_{n-1}}$. This implies $T(I_{a_{n-2}}) \supseteq I_{a_{n-1}} \supseteq K_1$ (by 5.1). By 5.2 there is a closed interval $K_2 \subseteq I_{a_{n-2}}$ s.t. $T(K_2) = K_1$. $I_{a_{n-3}} \to I_{a_{n-2}}$ so $I_{a_{n-3}}$ T-covers $I_{a_{n-2}}$. This implies $T(I_{a_{n-3}}) \supseteq I_{a_{n-2}} \supseteq K_2$. So, $I_{a_{n-3}}$ T-covers $K_2$. By 5.2 there is a closed interval $K_3 \subseteq I_{a_{n-3}}$ s.t. $T(K_3) = K_2$. Continuing this way gives $K_n \subseteq I_{a_0}$, $K_{n-1} \subseteq I_{a_1}, \ldots, K_2 \subseteq I_{a_{n-2}}$, $K_1 \subseteq I_{a_{n-1}}$ where $T(K_{n-j}) = K_{n-j-1}$ for $j<n-1$ and $T(K_1) = I_{a_0}$. Consequently,

$$\begin{aligned} T^n(K_n) &= T^{n-1}(K_{n-1}) = \ldots = T^2(K_2) \\ &= T(K_1) = I_{a_n} . \end{aligned}$$

But
$$K_n \subseteq I_{a_0} = I_{a_n} .$$

Therefore $T^n(K_n) \supseteq K_n$. There is a pair $u,v$ s.t. $T^n(u) \geq u$ and $T^n(v) \leq v$ otherwise $T^n(K_n) \not\supseteq K_n$. By 5.3 there is a $x_0 \in K_n$ s.t. $T^n(x_0) = x_0$. Since $x_0 \in K_n \subseteq I_{a_0}$ we have $T(x_0) \in K_{n-1} \subseteq I_{a_1}$, $T^2(x_0) \in K_{n-2} \subseteq I_{a_2}, \ldots, T^{n-1}(x_0) \in K_1 \subseteq I_{a_{n-1}}$. Therefore, $T^i(x_0) \in I_{a_i}$ $i = 0,1,2,\ldots,n$.

Definition

Let $p = \langle p_j \rangle_{j=0}^{k}$ be a finite sequence of elements of the set $\{1,2,\ldots,s\}$. Let $w(p) = \prod_{j=1}^{k} M_{p_{j-1}p_j}$ be referred to as the width of $p$. Call $p$ a path if $w(p) \neq 0$ and let $\ell(p) = k$ be the length of $p$.

Definition

Let $G$ be the $\alpha$-graph of $T$. $G'$ is a subgraph of $G$ denoted

$G' \leq G$, if $G'$ is an oriented graph whose vertices are members of $\alpha$ and where the number of arrows from $I_i$ to $I_j$ in $G'$ does not exceed the corresponding number in $G$.

Let $m'_{ij}$ denote the number of arrows from $I_i$ to $I_j$ in $G'$.

Let $H(G') = \log \lambda$ where $\lambda$ is the spectral radius of $M'_T$ where $M'_T = (M'_{ij})^s_{i,j=1}$.

5.5 Lemma

Let $G$ be the $\alpha$-graph of $T$. Let $G' \leq G$ be such that

$$m'_{ij} = \begin{cases} 1 & m_{ij} \neq 0 \\ 0 & m_{ij} = 0 \end{cases}.$$

Then $h(G') \leq h(T)$.

Proof:

$h(G') = \log \lambda$ where $\lambda$ is the spectral radius of $M'_T = (m'_{ij})^s_{i,j=1}$ (Some rows and columns of $M'_T$ might be zero). Let $\Sigma M'_T = \sum_{i,j=1}^{s} m'_{ij}$ then $h(G') = \lim_{n\to\infty} \frac{1}{n} \log \Sigma(M'_T)^n$ (see Chap. 1. Sec. 6). $\Sigma(M'_T)^n$ equals the number of distinct paths of length $n$ in $G'$.

Let $\langle j_0, j_1, \ldots, j_n \rangle$ be a path of length $n$ for $G'$. Then $I_{j_0} \to I_{j_1} \to \ldots \to I_{j_{n-1}} \to I_{j_n}$ forms a sequence of T-covers. $I_{j_{n-1}}$ T-covers $I_{j_n}$ so there is a closed interval $K_1 \subseteq I_{j_{n-1}}$ s.t. $T(K_1) = I_{j_n}$. Proceeding as in the proof of 5.4 we obtain closed intervals $K_n \subseteq I_{j_0}$, $K_{n-1} \subseteq I_{j_1}, \ldots, K_2 \subseteq I_{j_{n-2}}$, $K_1 \subseteq I_{j_{n-1}}$ s.t. $T(K_{n-i}) = K_{n-i-1}$ $i<n-1$ and $T(K_1) = I_{j_n}$. The $K$ sets can be chosen so that end points map to end points.

Now, $K_n \subseteq I_{j_0} \cap T^{-1} I_{j_1} \cap \ldots \cap T^{-n} I_{j_n}$. Furthermore, $T^n(K_n) = T^{n-1}(K_{n-1}) = \ldots = T^2(K_2) = T(K_1) = I_{j_n}$ where end points map to end points. $K^0_{n-i} \subseteq I^0_{j_i}$ for each $i<n-1$ since $K_{n-i} \subseteq I_{j_i}$.

Since end points map to end points it follows that

$T^n(K_n^o) = T^{n-1}(K_{n-1}^o) = \ldots = T(K_1^o) = I_{j_n}^o$. So, $K_n^o$ has disjoint intersection with every member of $\bigvee_{i=0}^{n} T^{-1}\alpha$ except $I_{j_0} \cap T^{-1} I_{j_1} \cap \ldots \cap T^{-n} I_{j_n}$. $K_n^o$ is non-empty since $T^n(K_n) = I_{j_n}$ (thus $K_n$ is non-degenerate). So the set $I_{j_0} \cap T^{-1} I_{j_1} \cap \ldots \cap T^{-n} I_{j_n}$ is essential for $\bigvee_{i=0}^{n} T^{-1}\alpha$ to cover $I$.

Therefore $\Sigma(M_T')^n \leq N(\bigvee_{i=0}^{n} T^{-1}\alpha)$ (The cardinality of closed covers is defined in Chapter 4 Sec. 1). Consequently, $h(G') \leq h(T,\alpha)$ upon taking limits.

Now, $h(T^n,\alpha^n) = n h(T,\alpha)$ since

$$\lim_{k\to\infty}\frac{1}{k}\log N(\bigvee_{i=0}^{k-1}T^{-in}\alpha^n) = \lim_{k\to\infty}\frac{1}{k}\log N(\bigvee_{i=0}^{nk-1}T^{-i}\alpha)$$

$$= n\lim_{k\to\infty}\frac{1}{nk}\log N(\bigvee_{i=0}^{nk-1}T^{-i}\alpha)$$

By 4.3 there is an open cover $\beta$ s.t. $h(T,\alpha|\beta) \leq \log 3$ and open covers $\beta_n$ s.t. $h(T^n,\alpha^n|\beta_n) \leq \log 3$ for each $n$. From 4.1(ii) $h(T^n,\alpha^n) \leq h(T^n,\beta_n) + h(T^n,\alpha^n|\beta_n)$. Thus $n h(T,\alpha) = h(T^n,\alpha^n) \leq h(T^n,\beta_n) + \log 3$ and $n h(T,\alpha) \leq h(T^n) + \log 3$ since $\beta_n$ is open for all $n$. Thus $h(T,\alpha) \leq \frac{1}{n} h(T^n) + \frac{1}{n}\log 3$. Since $n$ is abritrary and $H(T^n) = n h(T)$ (by 3.12) we have $h(G') \leq h(T,\cdot) \leq h(T)$. □

Definition

Let a set $R \subseteq \{1,2,\ldots,s\}$ be termed a rome if there is no path $p = \langle p_j \rangle_{j=0}^{k}$ with $p_0 = p_k$ s.t. $\{p_j\}_{j=0}^{k} \cap R = \phi$.

Note: A rome contains members of every closed loop.

Definition

Let a path $p = \langle p_j \rangle_{j=0}^{k}$ be termed simple w.r.t. a rome $R$ if

$\{p_0,p_k\} \subseteq R$ and $\{p_1,p_2,\ldots p_{k-1}\} \cap R = \phi$.

Note: Simple paths have only their initial and terminal points in the rome.

Definition

Let the matrix $A_T(R)$ be a function of a rome $R = \{r_1,r_2,\ldots,r_k\}$ $(i \neq j \Rightarrow r_i \neq r_j)$ defined as follows: $A_T(R) = (a_{ij})_{i,j=1}^{k}$ where $a_{ij}(\lambda) = \sum_p w(p)\, \lambda^{-\ell(p)}$ (the summation is over every simple path originating at $r_i$ and terminating at $r_j$).

5.6 Lemma

Let $R = \{r_1,r_2,\ldots,r_k\} \subseteq \{1,2,\ldots,s\}$ be a rome. Then $\mathrm{Det}(M_T - \lambda I) = (-1)^{s-k} \lambda^s\, \mathrm{Det}(A_T(R) - I)$.

Proof:

Suppose $R = \{1,2,\ldots,s\}$. Then the only simple paths are of the form $\langle r_i,r_j \rangle$ with length 1 and width $m_{ij}$. So, $a_{ij} = \lambda^{-1} m_{ij}$. Then $a_{ij} - \delta_{ij} = \lambda^{-1} m_{ij} - \delta_{ij} = \lambda^{-1}(m_{ij} - \lambda\, \delta_{ij})$ where

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

$$|A_T(R) - I| = \begin{vmatrix} \lambda^{-1}(m_{11}-\lambda) & \lambda^{-1}m_{12} \cdots\cdots \lambda^{-1}m_{1s} \\ \lambda^{-1}m_{21} & \lambda^{-1}(m_{22}-\lambda) \cdots \lambda^{-1}m_{2s} \\ \vdots & \\ \lambda^{-1}m_{s1} & \lambda^{-1}m_{s2} \cdots\cdots \lambda^{-1}(m_{ss}-\lambda) \end{vmatrix}$$

$$= \lambda^{-1} \begin{vmatrix} m_{11}-\lambda & \lambda^{-1}m_{12} \cdots\cdots \lambda^{-1}m_{1s} \\ m_{21} & \lambda^{-1}(m_{22}-\lambda) \cdots \lambda^{-1}m_{2s} \\ \vdots & \vdots \\ m_{s1} & \lambda^{-1}m_{s2} \cdots\cdots \lambda^{-1}(m_{ss}-\lambda) \end{vmatrix}$$

$$= \lambda^{-2} \begin{vmatrix} m_{11}-\lambda & m_{12} & \lambda^{-1}m_{13} \cdots\cdots & \lambda^{-1}m_{1s} \\ m_{21} & m_{22}-\lambda & \lambda^{-1}(m_{33}-\lambda) \cdots & \lambda^{-1}m_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ m_{s1} & m_{s2} & \lambda^{-1}m_{s3} \cdots\cdots & \lambda^{-1}(m_{ss}-\lambda) \end{vmatrix}$$

$$= \lambda^{-s} \begin{vmatrix} m_{11}-\lambda & m_{12} \cdots\cdots m_{1s} \\ m_{21} & m_{22}-\lambda \cdots\cdots m_{2s} \\ \vdots & \vdots \\ m_{s1} & m_{s2} \cdots\cdots m_{ss}-\lambda \end{vmatrix}$$

$$= \lambda^{-s} \; | M_T - \lambda I |$$

Consequently $|A_T(R) - I| = \lambda^{-s}|M_T - \lambda I|$

$$\therefore \qquad \lambda^{s}|A(R) - I| = |M - \lambda I| \; .$$

Now, if $S = \{s_1, s_2, \ldots, s_k\}$ is a rome and $s_0 \notin S$ then $J = \{s_0, s_1, s_2, \ldots, s_k\}$ is also a rome. Let

$$A_T(J) = (\tau_{ij})_{i,j=0}^{k}$$

$$A_T(S) = (\sigma_{ij})_{i,j=1}^{k}$$

$$\sigma_{ij} = \sum_{p} w(p)\lambda^{-\ell(p)}$$

where $p$ sums over all simple paths (w.r.t. S) beginning at $r_i$ and ending at $r_j$. Since $s_0 \notin S$, simple paths w.r.t. S may contain $s_0$ as a non-terminal point. So, $p$ sums over all simple paths (w.r.t. S) not passing through $s_0$ plus those that do. $\tau_{ij} = \sum_{p} w(p)\lambda^{-\ell(p)}$ where p sums over all simple paths (w.r.t. J) beginning at $r_i$ and ending at $r_j$. Since $s_0 \in J$ the summation for $\tau_{ij}$ excludes those paths passing through $s_0$. Thus $\sigma_{ij} = \tau_{ij} + \tau_{i0}\tau_{0j}$. Let $C_{ij} = \tau_{ij} - \delta_{ij}$. Then $A_T(J) - I = (C_{ij})_{i,j=0}^{k}$. So, $(\sigma_{ij} - \delta_{ij})_{i,j=1}^{k} = (\tau_{ij} - \delta_{ij} + \tau_{i0}\tau_{0j})_{i,j=1}^{k}$. Since $i \neq 0$, $j \neq 0$, $\tau_{i0} = C_{i0}$ $\tau_{0j} = C_{0j}$. So, $(\sigma_{ij} - \delta_{ij})_{i,j=1}^{k} = (C_{ij} + C_{i0}C_{0j})_{i,j=1}^{k}$. Consequently, $A_T(S) - I = (C_{ij} + C_{i0}C_{oj})_{i,j=1}^{k}$. S is a rome and $s_0 \notin S$ so $\tau_{00} = 0$ because there can be no path $<s_0, s_0>$. Therefore $C_{00} = -1$.

$$|A_T(J) - I| = \begin{vmatrix} C_{00} & C_{01} \cdots\cdots C_{0k} \\ C_{10} & C_{11} \cdots\cdots C_{1k} \\ \vdots & \vdots \qquad\qquad \vdots \\ C_{k0} & C_{k1} \cdots\cdots C_{kk} \end{vmatrix}$$

$$= \begin{vmatrix} C_{00} & C_{01} + C_{00}C_{01} & C_{02} \cdots\cdots C_{0k} \\ C_{10} & C_{11} + C_{10}C_{01} & C_{12} \cdots\cdots C_{1k} \\ \vdots & \vdots & \vdots \qquad\qquad \vdots \\ C_{k0} & C_{k1} + C_{k0}C_{01} & C_{k2} \cdots\cdots C_{kk} \end{vmatrix}$$

$$= \begin{vmatrix} C_{00} & C_{01} + C_{00}C_{01} & C_{02} + C_{00}C_{02} \cdots\cdots & C_{0k} \\ C_{10} & C_{11} + C_{10}C_{01} & C_{12} + C_{10}C_{02} & C_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ C_{k0} & C_{k1} + C_{k0}C_{01} & C_{k2} + C_{k0}C_{02} & C_{kk} \end{vmatrix} =$$

$$= \begin{vmatrix} c_{00} & c_{01}+c_{00}c_{01} & \cdots\cdots & c_{0k}+c_{00}c_{0k} \\ c_{10} & c_{11}+c_{10}c_{01} & \cdots\cdots & c_{1k}+c_{10}c_{0k} \\ \vdots & \vdots & & \vdots \\ c_{k0} & c_{k1}+c_{k0}c_{01} & \cdots\cdots & c_{kk}+c_{k0}c_{0k} \end{vmatrix}$$

$$= \begin{vmatrix} -1 & 0 & \cdots\cdots & 0 \\ c_{10} & c_{11}+c_{10}c_{01} & \cdots\cdots & c_{1k}+c_{10}c_{0k} \\ \vdots & \vdots & & \vdots \\ c_{k0} & c_{k1}+c_{k0}c_{01} & \cdots\cdots & c_{kk}+c_{k0}c_{0k} \end{vmatrix}$$

$$= -1 \begin{vmatrix} c_{11}+c_{10}c_{01} & \cdots\cdots & c_{1k}+c_{10}c_{0k} \\ \vdots & & \vdots \\ c_{k1}+c_{k0}c_{01} & \cdots\cdots & c_{kk}+c_{k0}c_{0k} \end{vmatrix}$$

$$= (-1) \; | A_T(S) - I |$$

$$\therefore \quad | A_T(J) - I | = (-1) \; | A_T(S) - I | .$$

So, by adding elements from $\{1,2,\ldots,s\}$ one at a time to the rome $S$ we eventually get the rome $R = \{1,2,\ldots,s\}$. For each element added the sign of the determinant changes once. So, if $\mathrm{card}(S) = k$ then $s - k$ elements must be added to make $\{1,2,\ldots,s\}$ and the sign of the determinant changes $s - k$ times. Therefore

$(-1)^{s-k} | A_T(S) - I| = |A_T(R) - I|$. Since

$\lambda^s | A_T(R) - I| = |M_T - \lambda I|$. It follows that

$(-1)^{s-k} \lambda^s | A_T(S) - I| = |M_T - \lambda I|$. □

Section 3. Periodic Points

Let $I$ be a closed interval in $\mathbb{R}$ and $T:I\to\mathbb{R}$ a fixed continuous map. Recall that $\text{Orb } x = \langle x, T(x), T^2(x), \ldots \rangle$. Orb $x$ can be either finite or infinite. In this section we will consider partitions which are induced by finite orbits (i.e. periodic cycles).

Suppose $\text{Orb } x = \langle x, T(x), \ldots, T^n(x) \rangle$. Then $\{x, T(x), T^2(x), \ldots, T^n(x), \min I, \max I\}$ induces a partition on $I$. This partition results from ordering the above set by the usual ordering $\mathbb{R}$ and then forming closed intervals.

Definition

A point $Z \in I$ separates Orb $x$ if for each $y \in \text{Orb } x$ $T(y) < Z \Rightarrow y < Z$.

5.7 Proposition

Suppose Orb $x$ is of finite cardinality and $\alpha$ is the partition induced by Orb $x$ as described above. Let $[a,b] \in \alpha$ be defined by:

$a = \max\{y \in \text{Orb } x \mid T(y) > y\}$ and $b = \min\{y \in \text{Orb } x \mid a < x\}$.

Then either there is some $K \in \alpha$ unequal to $[a,b]$ s.t. $K$ T-covers $[a,b]$ or every point of $[a,b]$ separates Orb $x$.

Proof:

Denote the points of Orb $x$ in increasing order on the real line by $p_0, p_1, \ldots, p_{n-1}$.

If there is no $K \in \alpha$ s.t. $K$ T-covers $[a,b]$ then for every pair $p_j, p_{j+1} \le a$, $T(p_j), T(p_{j+1}) \ge b$ and for every pair $p_i, p_{i+1} \ge b$ $T(p_i), T(p_{i+1}) \le a$. Thus each point of $[a,b]$ separates Orb $x$.

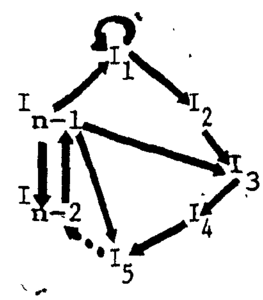
If $[a,b]$ does not separate Orb $x$ then there is at least one point $p_i$ s.t. either 1) $p_i < a$ and $T(p_i) < a$

or 2) $p_i > b$ and $T(p_i) > b$.

(The inequalities are strict because of the definition of $a$ and $b$).
If 1) then there must also be a point $p_j < a$ s.t. $T(p_j) \geq b$.
If 2) then there must also be a point $p_j > b$ s.t. $T(p_j) \leq a$.
This is because at least one point on either side of $[a,b]$ must change sides under the action of $T$. So there is a $K \in \alpha$ $(K \neq [a,b])$ s.t. $T(K) \supseteq [a,b]$. Thus $K$ T-covers $[a,b]$ by 5.2. □

5.8 <u>Theorem</u>

Suppose $T: I \to \mathbb{R}$ has a point $x$ of odd period $n > 1$ and no points of odd period less than $n$ (apart from 1). Let $\alpha$ be the partition of $I$ induced by the set $\{x, T(x), \ldots, T^{n-1}(x), \min I, \max I\}$ and let $\beta$ be the partition of ~~$J =$~~ $[\min \text{Orb } x, \max \text{Orb } x]$ by the set $\{x, T(x), \ldots, T^{n-1}(x)\}$. Then the $\alpha$-graph of $T$ contains a subgraph of the form:




(There are arrows from $I_{n-1}$ to all <u>odd</u> vertices).

The vertices of the subgraph are all members of $\beta$.

<u>Proof</u>:

Let $[a,b]$ be defined as in 5.6. Denote $[a,b]$ by $I_1$. $T(b) \leq a$ and $T(a) \geq b$ so, by the I.V. Theorem $T(I_1) \supseteq I_1$. Thus by 5.2 $I_1$ T-covers $I_1$. By 5.3 Cor. there is a $z \in I_1$ s.t. $T(z) = z$.

Denote the points of Orb $x$ in their order on the real line by $p_0, p_1, \ldots, p_{n-1}$. For some $k$, $T^k(a) = p_{n-1}$ and for some $j$, $T^j(a) = p_0$. Consequently, $T^k([a,z]) \supseteq [z, p_{n-1}]$ and $T^j([a,z]) \supseteq [p_0, z]$. The union $[p_0, z] \cup [z, p_{n-1}]$ equals $J$ which includes every member of $\beta$.

Consequently, there is a sequence of T-covers from $I_1$ to each other member of $\beta$.

Now, $n$ is odd so Orb $x$ has an odd number of elements. So, there is no point in $I_1$ which separates Orb $x$. Using 5.6 there is a $K \in \beta$ s.t. $K$ T-covers $I_1$.

Consequently there is a loop from $I_1$ to $I_1$ different from $I_1$.

Let $I_1 \to I_2 \to I_3 \to \dots \to K \to I_1$ be the shortest loop from $I_1$ to $I_1$ other than $I_1$.

Claim: This loop has length equal to $n$.

Suppose the contrary. Then there is a loop $I_1 \to I_2 \to I_3 \to \dots \to I_\ell \to I_1$ with $\ell < n-1$. (Any loop of length $\geq n$ cannot be shortest).
If $\ell$ is even then $I_1 \to I_2 \to \dots \to I_\ell \to I_1 \to I_1$ gives (using 5.4) a point of odd period $\ell + 1 < n$. This contradicts the assumption that $n$ is the least odd period. If $\ell$ is odd then $I_1 \to I_2 \to \dots \to I_\ell \to I_1$ gives (using 5.4) a point of odd period $\ell < n$. A contradiction.
$\therefore$ $I_1 \to I_2 \to \dots \to I_{n-1} \to I_1$ is the shortest loop from $I_1$ to $I_1$ other than $I_1$.

Claim: $I_2$ is adjacent to $I_1$.

Either $T([a,z]) \supseteq I_2$ or

$T([a,b]) \supseteq I_2$.

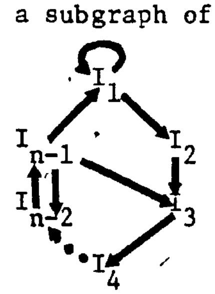
If $I_2$ is not adjacent to $I_1$ then $I_1$ T-covers some other member of $\beta$ other than $I_1$ and $I_2$. If that is the case then, $I_1 \to I_2 \to \dots \to I_{n-1} \to I_1$ is not the shortest loop. A contradiction of what was previously proven. $\therefore$ $I_2$ is adjacent to $I_1$.

The sets $I_2, I_3, \dots, I_{n-2}$ T-cover only one other member of $\beta$ otherwise a shorter loop than $I_1 \to I_2 \to \dots \to I_{n-1} \to I_1$ would be possible.

Consequently, $I_2$ T-covers $I_3$ and no other interval. Suppose $I_2$ is to the left of $I_1$ (relative to ordering on real line). Then $I_3$ is to the immediate right of $I_1$ since $x = a$ is the right end-point of $I_2$, $T(a) \geq b$, and $I_2$ T-covers $I_3$ only. Now $x = b$ is the left end point of $I_3$ and $T(b)$ is the left end point of $I_2$. Since $I_3$ T-covers $I_4$ only $I_4$ must be to the immediate left of $I_2$.

Continuing this way gives the following order on the real line: $I_{n-1} \ldots I_4\ I_2\ I_1\ I_3\ I_5 \ldots I_{n-2}$. (If we suppose $I_2$ is to the right of $I_1$ we get the reverse ordering $I_{n-2} \ldots I_3\ I_1\ I_2 \ldots I_{n-1}$).

Now, since $I_{n-3}$ T-covers $I_{n-2}$ the right end point of $I_{n-1}$ is mapped to the right end point of $I_{n-2}$. Furthermore, $I_{n-1}$ T-covers $I_1$ so the left end point of $I_{n-1}$ must be mapped to $x = a$. Therefore, $I_{n-1}$ T-covers $I_1, I_3, I_5, \ldots, I_{n-2}$. Consequently, the $\alpha$-graph of $T$ contains a subgraph of the form




with arrows from $I_{n-1}$ to all <u>odd</u> vertices.

□

**5.9 <u>Lemma</u>**

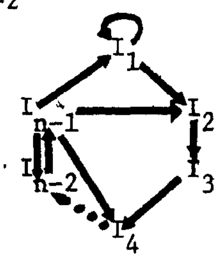
If $T$ has a point of period $n$ which is even then it has a point of period 2.

<u>Proof</u>:

In the proof of the previous theorem the fact that $n$ was odd was used in only one place. It was shown that for $n$ odd there must be some $K \in \beta$ s.t. $K$ T-covers $I_1$. If there is no such $K \in \beta$ then by 5.6 each point of $I_1$ separates Orb $x$. Consequently,

$[p_0,a]$ T-covers $[b,p_{n-1}]$ and $[b,p_{n-1}]$ T-covers $[p_0,a]$. Thus $[p_0,a]\to[b,p_{n-1}]\to[p_0,a]$ gives a fixed point for $T^2$ by 5.4. So there is a point of period 2 for $T$.

If there is a $K\in\beta$ s.t. $K$ T-covers $I_1$ then $I_1\to I_2\to\ldots\to K\to I_1$ is the shortest loop with $K = I_{n-1}$. This induces the ordering $I_{n-1}\ldots I_3\ I_1\ I_2\ldots I_{n-2}$ or $I_{n-2}\ldots I_2\ I_1\ I_3\ldots I_{n-1}$. Then $I_{n-1}$ T-covers $I_1,I_2,\ldots,I_{n-2}$. Consequently we have the subgraph




with arrows from $I_{n-1}$ to all even vertices.

Then the loop $I_{n-1}\to I_{n-2}\to I_{n-1}$ gives a fixed point for $T^2$ (by 5.4). Since $T(I_{n-2}) = I_{n-1}$ and the end points of $I_{n-2}$ have period $n$, the fixed point for $T^2$ is a point of least period 2 for $T$. □

5.10 Theorem (Sarkovskii, ([23])

Let $T:I\to\mathbb{R}$ be continuous. Order $\mathbb{N}$ as follows: $3\vdash 5\vdash 7\vdash 9\vdash\ldots\vdash 2\cdot 3\vdash 2\cdot 5\vdash\ldots\vdash 2^2\cdot 3\vdash 2^2\cdot 5\vdash\ldots\vdash 2^2\vdash 2\vdash 1$. If $n\vdash k$ and $T$ has a point of least period $n$ then $T$ has a point of least period $k$.

Proof:

There are four cases:

1) $n = 2^m,\ k = 2^\ell,\ \ell < m$

2) $n = p\cdot 2^m$ p(odd), $k = q\cdot 2^m$ q(even)

3) $n = p\cdot 2^m$ p(odd), $k = q\cdot 2^m$ q(odd) $> p$

4) $n = p\cdot 2^m$ p(odd), $k = 2^\ell,\ m \le \ell$

Case 1)

Suppose $\ell = 0$. There is a point $x$ with least period $n$ so arrange the members of $\text{Orb } x = \{x, T(x), \ldots, T^n(x)\}$ in their usual order in $\mathbb{R}$. Choose $p_j, p_{j+1} \in \text{Orb } x$ s.t. $p_j < p_{j+1}$ and $T(p_{j+1}) \leq p_j < T(p_j)$. (This is possible since we may let $p_j$ be the greatest member of Orb $x$ which moves to the right under the action of $T$). Then $[p_j, p_{j+1}]$ T-covers $[p_j, p_{j+1}]$ which ensures the existence of a fixed point for $T$ ( using 5.4).

If $\ell > 0$ let $\psi = T^{k/2}$. Then $\psi = T^{2^{\ell-1}}$ and $(T^{2^{\ell-1}})^{2^{m+1-\ell}} = T^{2^m}$. Since $T$ has a point of period $2^m$, $\psi$ has a point of period $2^{m+1-\ell}$. Thus ( using 5.8) $\psi$ has a point of least period 2. But $\psi^2 = T^{2^\ell}$ so $T$ has a point of least period $2^\ell$.

Case 2)

$T^{2^m}$ has a point $x$ of period $p$. Let $\beta$ be the partition of $J = [\min \text{Orb } x, \max \text{Orb } x]$ by the members of $\text{Orb } x = \{x, T^{2^m}, T^{2 \cdot 2^m}(x), \ldots, T^{(p-1)2^m}(x)\}$.

Suppose $q < p$. Since $p$ is odd and $q$ is even there is by Theorem 5.7 a loop (in the $\beta$ graph of $T^{2^m}$) of the form:

$I_{p-1} \to I_{p-q} \to I_{p-q+1} \to \ldots \to I_{p-1}$. This implies ( using 5.4) the existence of a point of period $q$ for $T^{2^m}$.

Suppose $q > p$. Since $p$ is odd there is ( using 5.7) a loop of the form: $I_1 \to I_2 \to \ldots \to I_{p-1} \to I_1 \to I_1$. Since $I_1$ T-covers $I_1$ this loop may be extended to $I_1 \to I_2 \to \ldots \to I_{p-1} \to I_1 \to I_1 \to \underbrace{I_1 \to \ldots \to I_1}_{q-p \text{ times}}$

By 5.4 a point of period $q$ exists for $T^{2^m}$.

Because of the way the members of $\beta$ are arranged on the interval $J$ ( using 5.7) $q$ must be a least period in both of the above instances. So for even $q$ there is a point of least period $q$ for $T^{2^m}$. To show that $q\cdot 2^m$ is also a least period for $T$ we proceed by supposing the contrary and deriving the contradiction that $q$ is not a least period for $T^{2^m}$.

(i) $T^{q\cdot 2^m}(x) = x$. If $T^{r-2^m}(x) = x$ where $r < q$ then $x$ is not a point of least period $q$ for $T^{2^m}$.

(ii) $T^{q\cdot 2^m}(x) = x$. If $T^{r-2^\ell}(x) = x$ where $r < q$ and $\ell < m$ then $(T^{r-2^\ell})^{m-\ell}(x) = x$ also. So $T^{r-2^m}(x) = x$ and consequently $x$ is not a point of least period $q$ for $T^{2^m}$.

(iii) $T^{q\cdot 2^m}(x) = x$. If $T^{q\cdot 2^\ell}(x) = x$ where $\ell < m$ then $T^{(q/2)\cdot 2^{\ell+1}}(x) = x$ (since $q$ is even). This reduces to (i) or (ii).

Case 3)

As for $q(\text{even}) > p$ we show that for $q(\text{odd}) > p$ there is a point of period $q$ for $T^{2^m}$ by means of a loop

$$I_1 \to I_2 \to \dots \to I_{p-1} \to I_1 \to I_1 \to \underbrace{I_1 \to \dots \to I_1}_{q-p \text{ times}} .$$

To show that such a point is also a point of least period $q\cdot 2^m$ for $T$ we proceed as in (i) and (ii) of case 2). If $T^{q\cdot 2^\ell}(x) = x$ where $\ell < m$ let $n = q\cdot 2^\ell$. Then $k = (q\cdot 2^{m-\ell})2^\ell$ which fits case 2).

Case 4)

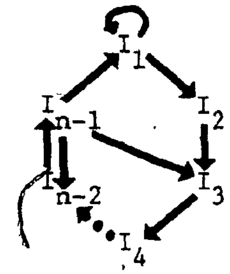
Let $q\cdot 2^m = 2\cdot 2^m$. Then $q(\text{even}) < p$. By case 2) there is a point of least period $2^{m+1}$ for $T$. $\ell \leq m$ so by case 1) there is a point of least period $k = 2^\ell$ for $T$. □

Section 4. Entropy and Periodic Points

In this section a lower bound for the entropy of a map $T$ will be established using the existence of periodic points for $T$.

5.11 Proposition

Let $G$ denote the $\alpha$-graph




(where $n$ is odd and there are arrows from each $I_j$ to $I_{j+1}$ as well as from $I_{n-1}$ to every odd vertex). Let $S_n$ denote all roots (real and complex) of $x^n - 2x^{n-2} - 1$. Let $\sigma_n = \max\{|z| \mid z \in S_n\}$. Then $h(G) = \log \sigma_n$.

Proof:

The set $\{1, n-1\}$ is a rome. $h(G) = \log \rho(M)$ where $\rho$ denotes the spectral radius and $M$ is the matrix associated with $G$ (as in the definition of $h(G)$).

By 5.5, $|M - \lambda I| = (-1)^{s-k} \lambda^s \, |A(R) - I|$ where $s = n-1$, $k = 2$ and

$$A(R) = \begin{bmatrix} \lambda^{-1} & \lambda^{-(n-2)} \\ \lambda^{-1} & \lambda^{-2} + \lambda^{-4} + \ldots + \lambda^{-(n-3)} \end{bmatrix}.$$

So, $$|A(R) - I| = \frac{\lambda}{1+\lambda}(\lambda^{-(n-3)} - 1) - \lambda^{-1} - \lambda^{-(n-1)} + 1$$

Consequently,

$$\lambda^{n-1}|A(R) - I| = \frac{\lambda^n - 2\lambda^{n-2} - 1}{1+\lambda}.$$ Therefore

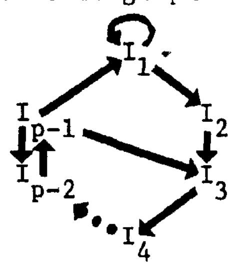
$$\rho(M) = \max\{|\lambda| : |M - \lambda I| = 0\}$$
$$= \sigma_n$$ □

5.12 Theorem

Let $T: I \to \mathbb{R}$ be continuous. Suppose $T$ has a point of least period $p \cdot 2^m$ with $p$(odd) $> 1$. Then $\frac{1}{2^m} \log \sigma_p \leq h(T)$ where $\sigma_n = \max \{|z| : z \in S_n\}$ and $S_n$ is all roots (real and complex) of $x^p - 2x^{p-2} - 1$.

<u>Proof</u>:

Let $x$ be a point s.t. $T^{p \cdot 2^m}(x) = x$. Let Orb $x = \langle x, T^{2^m}(x), T^{2 \cdot 2^m}(x), \ldots, T^{(p-1)2^m}(x) \rangle$. Let $\alpha$ be the partition of $I$ by the members of Orb $x$. Let $G$ denote the $\alpha$-graph of $T^{2^m}$. By 5.8, $G$ has a subgraph $G'$ of the form:




(there is an arrow from $I_{p-1}$ to each odd vertex).

For any pair of vertices $I_i, I_j$ in $G'$ there is at most one arrow from $I_i$ to $I_j$. So, by 5.5 $h(G') \leq h(T^{2^m})$. By 3.12 $h(G') \leq 2^m h(T)$. By 5.11, $h(G'') = \log \sigma_n$. Therefore, $\frac{1}{2^m} \log \sigma_n \leq h(T)$ □

## BIBLIOGRAPHY


1. Adler, R.A., Konheim, A. and McAndrew, M., Topological entropy, Trans. Amer. Math. Soc., 114 (1965), p. 309-319.

2. Adler, R.A. and Marcus, B., Topological entropy and equivalence of dynamical systems, Memoirs Amer. Math. Soc., 20, no. 219, (1979).

3. Billingsley, Patrick, Ergodic Theory and Information, Wiley, New York, 1965.

4. Block, L., Guckheimer, J., Misiurewicz, M. and Young, L.S., Periodic points and topological entropy of one dimensional maps, in Global Theory of Dynamical Systems, Lecture Notes in Mathematics, No. 819, Springer-Verlag, New York, 1980.

5. Bowen, R., Entropy for group endomorphisms and homogenous spaces, Trans. Amer. Math. Soc., 153 (1970), p. 401-414.

6. Brown, J.R., Ergodic Theory and Topological Dynamics, Academic Press, New York, 1976.

7. Denker, M., Grillenberger, C. and Sigmund, K., Ergodic Theory on Compact Spaces, Lecture Notes in Mathématics, No. 527, Springer-Verlag, Berlin, 1976.

8. Dugundji, J., Topology, Allyn and Bacon, Inc., Boston, 1966.

9. Gantmacher, F.R., The Theory of Matrices, Vol. II, Chelsea, New York, 1959.

10. Goodman, T.N.T., Relating topological entropy and measure entropy, Bull. London Math. Soc., 3 (1971), p. 176-180.

11. Goodwyn, L.W., Topological entropy bounds measure-theoretic entropy, Proc. Amer. Math. Soc., 23 (1969), p. 679-688.

12. Grillenberger, C., Ensembles minimaux sans mesure d'entropie maximale, Monatsh. Math., 32 (1976), p. 275-285.

13. Halmos, Paul R., Finite-Dimensional Vector Spaces, Springer-Verlag, New York, 1974.

14. Halmos, Paul R., Measure Theory, Springer-Verlag, New York, 1974.

15. Kolmogoroff, A.N., Entropy per unit time as a metric invariant of automorphisms (Russian), Dokl. Adad. Nauk. SSSR, 124 (1959), p. 754-755.

16. Misiurewicz, M., Topological conditional entropy, Stud. Math., 55 (1976).

17. Misiurewicz, M. and Szlenk, W., Entropy of piecewise monotone mappings, Stud. Math., 67 (1980).

18. Nitecki, Z., Topological dynamics on the interval, Preprint.

19. Ornstein, D., Bernoulli shifts with the same entropy are isomorphic, Adv. in Math. 4 (1970), p. 337-352.

20. Parry, W., Intrinsic Markov chains, Trans. Amer. Math. Soc., 112 (1964), p. 55-66.

21. Parthasarathy, K.R., Introduction to Probability and Measure, MacMillan, London, 1977.

22. Rudin, W., Functional Analysis, McGraw-Hill, New York, 1973.

23. Sarkovskii, A.N., Coexistence of cycles of a continuous map of a line into itself, Ukr. Math. J., 16 (1964), p. 61-71.

24. Walters, P., An Introduction to Ergodic Theory, Springer-Verlag, New York, 1982.