



105
232
THS

THE CAPACITY AND AMBIGUITY
OF A TRANSDUCER

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
WILLIAM MORGAN CONNER
1969



This is to certify that the

thesis entitled

THE CAPACITY AND AMBIGUITY OF A TRANSDUCER

presented by

William Morgan Conner

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Mathematics

John R Kenney
Major professor

Date Dec 2, 1969



ABSTRACT

THE CAPACITY AND AMBIGUITY OF A TRANSDUCER

By

William Morgan Conner

Up to now little work has been done on examining the correspondence between a discrete, noisy information channel with memory and the unit square. The correspondence is made by associating the infinite sequences of the channel with the expansions of points in the unit interval. We begin such an investigation in this paper, although not in this generality. We introduce the element of memory but not noise, and examine the following particular type of noiseless channel with memory.

Let $S = \{0, 1, \dots, b-1\}$ where $b \geq 2$ is an integer and let m be a positive integer. Let $\mathfrak{F}_m$ be the set of all functions f^* from $S^m = \underbrace{S \times S \times \dots \times S}_{m \text{ factors}}$ into S . Now let $f^* \in \mathfrak{F}_m$, let $x \in (0, 1]$ and let $\sum_{i=1}^{\infty} x_i b^{-i}$ be the (unique) nonterminating base b expansion of x . We define a function f from $(0, 1]$ into $[0, 1]$ by $f(x) = \sum_{i=1}^{\infty} y_i b^{-i}$ where $y_i = f^*(x_i \dots x_{i+m-1})$, $i = 1, 2, \dots$. We call f a transducer of memory m , and we call $\{x_i\}$ the input sequence and $\{y_i\}$ the received sequence.

We give two definitions for the capacity of the transducer and show that these two values for the capacity are equal. We then show that the Hausdorff dimension of the set of all received sequences is equal to the transducer capacity. This result

establishes a correspondence between a geometric property (dimension of the set of received sequences) and an information theoretic property (transducer capacity). It shows that the "size" (dimension) of the received set, which is an intuitive measure of the "capacity", is equal to the quantity formally defined as the capacity.

With geometric considerations as the motivation, we next define the ambiguity of the transducer. It is shown that the transducer has a homogeneity property by proving that the ambiguity is almost everywhere the same. The usual quantity used for ambiguity in information theory is a conditional entropy (called the equivocation by most authors) which represents the ambiguity averaged over all received sequences. Our definition is pointwise and gives results involving almost everywhere statements which are somewhat more satisfying than statements involving averages.

THE CAPACITY AND AMBIGUITY OF A TRANSDUCER

By

William Morgan Conner

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

1969

662772
7-1-78

ACKNOWLEDGMENTS

The author wishes to thank his advisor, Dr. John R. Kinney, for suggesting the topic of this thesis, and for his helpful suggestions, patience, and encouragement throughout the course of the work.

Thanks are due also to Dr. G.A. Hedlund of Yale University for sending copies of requested material to the author.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. HAUSDORFF DIMENSION AND ENTROPY	4
III. THE CAPACITY OF THE TRANSDUCER	7
IV. THE DIMENSION OF THE RECEIVED SET	13
V. THE AMBIGUITY OF THE TRANSDUCER	18
VI. CONCLUSION	29
BIBLIOGRAPHY	31

I. Introduction

Besicovitch [1], Eggleston [8, 9] and others have calculated the Hausdorff dimension (see Section II) of subsets of the unit interval defined by placing certain restrictions on the digits of expansions of numbers. For example, let $M(p)$, $0 \leq p \leq 1$, be the set of points in the unit interval containing 1 in their dyadic expansions in the proportion p , i.e., $x = .x_1x_2\ldots$ belongs to $M(p)$ if and only if $\lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n x_k = p$. Eggleston [8] has shown that the dimension of $M(p)$ is $p \log_2 p - (1-p) \log_2 (1-p)$. (A simplified proof is obtained by using a general theorem due to Billingsley [5], p. 142.)

Observing that this value for the dimension is the entropy of an information source, Kinney [18] and Billingsley [2,3,4] sought a connection between dimension theory and information theory by making the infinite sequences of symbols from the source correspond to the expansions of points in the unit interval. Theorem 1 of Kinney's paper asserts the existence of a set of measure one whose dimension is the entropy of a Markov source. Dym [7] recently extended this theorem to general stationary, ergodic sources. Theorem 2 of Kinney's paper is concerned with noiseless coding and shows that the dimension of a certain set corresponding to the coded messages is equal to the capacity of the noiseless channel. Smorodinsky [20] recently extended this theorem to very general noiseless channels.

As yet little work has been done on examining the correspondence between a discrete, noisy channel with memory and the unit square. Such an investigation is begun in this paper, although not in this generality. We introduce the element of memory but not noise, and examine the following particular type of noiseless channel with memory.

Let $S = \{0, 1, \dots, b-1\}$ where $b \geq 2$ is an integer and let m be a positive integer. Let $\mathfrak{F}_m$ be the set of all functions f^* from $S^m = \underbrace{S \times S \times \dots \times S}_{m \text{ factors}}$ into S . We note that $\text{crd } \mathfrak{F}_m = b^{b^m}$ where crd denotes the cardinal number of a set.

Let $f^* \in \mathfrak{F}_m$ and let n be a positive integer. We define a function f_n from S^{n+m-1} into S^n as follows. Let $x = x_1 \dots x_{n+m-1} \in S^{n+m-1}$ and let $y_i = f^*(x_i \dots x_{i+m-1})$, $i = 1, \dots, n$. Then the sequence $y = y_1 \dots y_n \in S^n$, and we define $f_n(x) = y$. Note that $f_1 = f^*$.

Now let $x \in (0, 1]$ and let $\sum_{i=1}^{\infty} x_i b^{-i}$ be the (unique) nonterminating base b expansion of x . We define a function f from $(0, 1]$ into $[0, 1]$ by $f(x) = \sum_{i=1}^{\infty} y_i b^{-i}$ where $y_i = f^*(x_i \dots x_{i+m-1})$, $i = 1, 2, \dots$. In the terminology of Shannon [19], the function f is an example of a transducer of memory m . In the terminology of Feinstein [10], Billingsley [5], Khinchin [16], and others f is a noiseless, discrete channel (or code) of memory m . We will occasionally call $\{x_i\}$ the input sequence and $\{y_i\}$ the output (or received) sequence; and, following Shannon, we will call f a transducer of memory m .

In Section II we define and give the basic properties of Hausdorff dimension and entropy. In Section III we give two definitions of the capacity of the transducer and show that they are equivalent. It is shown in Section IV that the dimension of the set of all received sequences is equal to the capacity. Finally, in Section V we define and examine the ambiguity of the transducer.

The pound sign (#) will be used throughout this paper to denote the end of a proof. Also, all logarithms in this paper are to the base b .

II. Hausdorff Dimension and Entropy

In this section we give the definitions and basic properties of Hausdorff dimension and entropy, two concepts we will be using later.

Let $x \in (0,1]$ and let $x = \sum_{i=1}^{\infty} x_i b^{-i}$ be the nonterminating base b expansion of x . Define $b_i(x) = x_i$ for all i , i.e., $b_i(x)$ is the i th digit of the base b expansion of x . A set of the form $\{x: b_i(x) = s_i, i = 1, \dots, n\}$, where $s_i \in S$, is denoted by $[s_1, \dots, s_n]$ and is called a cylinder of length b^{-n} . Note that $[s_1, \dots, s_n]$ is a half-open (open on the left) b -adic interval of length b^{-n} (i.e., an interval of the form $(\frac{j}{b^n}, \frac{j+1}{b^n}]$ for some j , $0 \leq j \leq b^n - 1$).

The following definition of dimension in the unit interval, which extends Hausdorff's original definition, is given by Billingsley [5]. Let $M \subset (0,1]$, let α and ρ be positive real numbers, and let μ be a probability measure on the Borel sets of $(0,1]$. Define

$$\mu_{\alpha}(M, \rho) = \inf \sum_i \mu(v_i)^{\alpha},$$

where the infimum is taken over all μ - ρ -coverings of M , a μ - ρ -covering being a covering by cylinders v_i with $\mu(v_i) < \rho$. It is clear that $\mu_{\alpha}(M, \rho) \leq \mu_{\alpha}(M, \rho')$ for $\rho' < \rho$, so the limit

$$\mu_{\alpha}(M) = \lim_{\rho \rightarrow 0} \mu_{\alpha}(M, \rho)$$

exists (but may be infinite). It can be shown that for fixed M there is an α_0 such that $\mu_\alpha(M) = \infty$ for $\alpha < \alpha_0$ and $\mu_\alpha(M) = 0$ for $\alpha > \alpha_0$. The number α_0 is called the (Hausdorff) dimension of M with respect to μ and is denoted by $\dim_\mu M$. We will denote Lebesgue measure by L and will write $\dim M$ instead of $\dim_L M$.

Some properties of dimension are:

$$(2.1) \quad \mu_\alpha(M) > 0 \text{ implies } \dim_\mu M \geq \alpha;$$

$$(2.2) \quad \dim_\mu M > \alpha \text{ implies } \mu_\alpha(M) = \infty;$$

$$(2.3) \quad \mu_\alpha(M) < \infty \text{ implies } \dim_\mu M \leq \alpha;$$

$$(2.4) \quad \dim_\mu M < \alpha \text{ implies } \mu_\alpha(M) = 0;$$

$$(2.5) \quad 0 \leq \dim_\mu M \leq 1;$$

$$(2.6) \quad \dim_\mu M = 0 \text{ if } M \text{ is countable; and}$$

$$(2.7) \quad \dim_\mu M = 1 \text{ if } M \text{ is a Borel set with } \mu(M) > 0.$$

Dimension is useful for indicating the "size" of a noncountable, measure zero set.

Entropy was introduced into information theory by Shannon [19]; Kolmogorov and Sinai have extended the notion of entropy to general measure preserving transformations (see Billingsley [5]). Discussions of entropy in information theory may be found in Feinstein [10] or Dym [7].

Let $\mathcal{B}$ be the Borel sets in $(0,1]$ and define a transformation T on $(0,1]$ by $Tx = \{bx\}$ for each $x \in (0,1]$, where $\{bx\}$ denotes the fractional part of bx . The transformation T

is a left shift on the digits of the base b expansion of x .

If $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$, T is said to be measure preserving. In such a case we will say that μ is stationary.

We now define the entropy of a stationary probability measure μ as follows. For each positive integer n define

$$H_n(\mu) = -\sum \mu([s_1, \dots, s_n]) \log \mu([s_1, \dots, s_n]),$$

where the summation is taken over all $s_1 \dots s_n \in S^n$, and where we take $0 \log 0$ to be zero. It can be shown that the limit

$$H(\mu) = \lim_{n \rightarrow \infty} \frac{H_n(\mu)}{n}$$

exists. We call $H(\mu)$ the entropy of μ .

Similar definitions can be given for the unit square. Define a transformation T_1 on $(0,1] \times (0,1]$ by $T_1(x,y) = (Tx, Ty)$ for each $(x,y) \in (0,1] \times (0,1]$. As before a probability measure P on the Borel sets of $(0,1] \times (0,1]$ will be called stationary if T_1 preserves P . For stationary P we define

$$H_n(P) = -\sum P([s_1, \dots, s_n, r_1, \dots, r_n]) \log P([s_1, \dots, s_n, r_1, \dots, r_n]),$$

where $[s_1, \dots, s_n, r_1, \dots, r_n] = \{(x,y) : b_i(x) = s_i, b_i(y) = r_i, 1 \leq i \leq n\}$, and where the summation is taken over all

$s_1 \dots s_n, r_1 \dots r_n \in S^n$. Again it can be shown that the limit

$$H(P) = \lim_{n \rightarrow \infty} \frac{H_n(P)}{n},$$

called the entropy of P , exists.

III. The Capacity of the Transducer

Let $f^* \in \mathfrak{F}_m$ and let f be the corresponding transducer of memory m . Let n be a positive integer and let

$$N(n) = \text{crd } f_n(S^{n+m-1}),$$

i.e., $N(n)$ is the number of distinct output sequences of length n which correspond to the b^{n+m-1} input sequences of length $n+m-1$. Following Shannon's terminology [19] we call

$$C = \lim_{n \rightarrow \infty} \frac{\log N(n)}{n}$$

the capacity of the transducer.

Theorem 3.1. The limit C exists and satisfies $0 \leq C \leq 1$.

Proof: Let k and n be positive integers. There are $N(k)$ different ways in which a received sequence of length $k+n$ may begin, and at most $N(n)$ different ways in which it may end. Hence

$$(3.2) \quad N(k+n) \leq N(k)N(n).$$

Also it is obvious that

$$(3.3) \quad N(k) \leq N(n) \quad \text{for } k \leq n.$$

We now follow a well-known procedure (see, for example, Feinstein [10], p. 85) to show that C exists. Let $a = \inf_n \frac{\log N(n)}{n}$ and let $\epsilon > 0$ be given. There exists an integer r such that

$\frac{\log N(r)}{r} \leq a + \epsilon$. For any integer $n \geq r$ define k_n by $(k_n - 1)r \leq n < k_n r$. By (3.3) and (3.2) we have

$$\log N(n) \leq \log N(k_n r) \leq k_n \log N(r),$$

and thus

$$\frac{\log N(n)}{n} \leq \frac{k_n r}{n} \frac{\log N(r)}{r} \leq \frac{k_n r}{(k_n - 1)r} (a + \epsilon) = \frac{k_n}{k_n - 1} (a + \epsilon).$$

As n approaches ∞ , k_n approaches ∞ and hence $\frac{k_n}{k_n - 1}$ approaches 1. It follows that $\limsup_n \frac{\log N(n)}{n} \leq a + \epsilon$. Since ϵ was arbitrary, we have $\limsup_n \frac{\log N(n)}{n} \leq a$, and since $\frac{\log N(n)}{n} \geq a$ for all n , we have $\liminf_n \frac{\log N(n)}{n} \geq a$. Thus C exists (and is equal to a).

Clearly $C \geq 0$ since $\frac{\log N(n)}{n} \geq 0$ for all n . As for $C \leq 1$, we note that $\frac{\log N(n)}{n} \leq \frac{\log b^n}{n} = 1$ for all n . #

We now apply another definition of capacity, introduced by Shannon [19] and expanded upon by others for noisy channels, to our transducer.

Let T be the transformation defined in Section II, and let μ be a probability measure on the Borel sets $\mathcal{B}$. Then T is called ergodic under μ if for each $B \in \mathcal{B}$ such that $T^{-1}B = B$, $\mu(B)$ is either zero or one. In such a case we will say that μ is ergodic, omitting reference to T . We will denote by $\mathcal{M}$ the set of all probability measures on $\mathcal{B}$ which are stationary (defined in Section II) and ergodic.

For each $x \in (0, 1]$ define a probability measure ν_x on $\mathcal{B}$ by letting ν_x assign unit mass to the point $f(x)$. Let $\mu \in \mathcal{M}$ and, for $M \in \mathcal{C}$, where $\mathcal{C}$ is the class of Borel sets in $(0, 1] \times (0, 1]$, define

$$(3.4) \quad P(M) = \int_{(0,1]} v_x(\{y: (x,y) \in M\}) \mu(dx).$$

To see that this integral is defined (i.e., that the integrand is measurable) we proceed as follows. Let B be a cylinder (b-adic interval). Then $f^{-1}(B)$ is the finite disjoint union of b-adic intervals and hence $f^{-1}(B) \in \mathcal{B}$. Since the class of b-adic intervals generates $\mathcal{B}$ (i.e., the smallest σ -algebra containing the class of b-adic intervals is $\mathcal{B}$), f is measurable with respect to $\mathcal{B}$.

Now let B be a fixed member of $\mathcal{B}$. We show that the function $g(x) = v_x(B)$ is measurable with respect to $\mathcal{B}$. The function g assumes only two values, namely zero on the set $D = \{x: f(x) \notin B\}$ and one on the set $E = \{x: f(x) \in B\}$. Thus g will be measurable if D and $E \in \mathcal{B}$. Since $E = f^{-1}(B)$ and f is measurable, $E \in \mathcal{B}$, and since D is the complement of E , $D \in \mathcal{B}$. Thus $v_x(B)$, as a function of x , is measurable.

We next show that the class $\mathcal{D}$ of all sets M in $\mathcal{C}$ for which the integrand in (3.4) is measurable is a monotone class.

Let $M_1 \subset M_2 \subset \dots$ be a sequence of sets in $\mathcal{D}$. Then

$$v_x(\{y: (x,y) \in \bigcup_i M_i\}) = v_x(\bigcup_i \{y: (x,y) \in M_i\}) = \lim_i v_x(\{y: (x,y) \in M_i\})$$

and hence $\bigcup_i M_i \in \mathcal{D}$ since $v_x(\{y: (x,y) \in \bigcup_i M_i\})$, being the limit of measurable functions, is measurable. Similarly $\bigcap_i M_i \in \mathcal{D}$ for a decreasing sequence $M_1 \supset M_2 \supset \dots$ of sets in $\mathcal{D}$. Thus $\mathcal{D}$ is a monotone class. Also $\mathcal{D}$ contains all rectangles $B \times C$, $B, C \in \mathcal{B}$, since the set $\{y: (x,y) \in B \times C\}$ is either empty or C , and we showed above that $v_x(C)$ is measurable. It follows easily that $\mathcal{D}$ contains the finite, disjoint unions of rectangles $B \times C$,

$B, C \in \mathcal{B}$. Thus by a well-known result (see, for example, Kingman and Taylor [17], p.18), we have $\mathcal{B} = \mathcal{C}$. Hence the integrand in (3.4) is measurable for all $M \in \mathcal{C}$.

It is easily verified that $P(M) = \mu(\text{proj}_x\{M \cap \text{graph of } f\})$, where proj_x denotes the projection on the x -axis, and that P is a probability measure. Also the set function λ defined, for $B \in \mathcal{B}$, by

$$(3.5) \quad \lambda(B) = P((0,1] \times B)$$

is easily seen to be a probability measure. We note that

$$\lambda(B) = \mu(\{x: f(x) \in B\}).$$

Theorem 3.6. P is stationary and $\lambda \in \mathcal{M}$.

Proof: We show first that for any $B \in \mathcal{B}$ we have

$$(3.7) \quad \{x: f(x) \in T^{-1}B\} = T^{-1}\{x: f(x) \in B\}.$$

Let A be the set on the left hand side of (3.7) and let A' be the set on the right hand side of (3.7). Let $x \in A$ and let $x_i = b_i(x)$, $i = 1, 2, \dots$, i.e., x_i is the i th digit of the non-terminating base b expansion of x . Then $f(x) \in T^{-1}B$ which implies that $f(x) = .y_1 y_2 \dots$ where $y \in S$ and the point $.y_1 y_2 \dots \in B$. Thus $f(Tx) = f(.x_2 x_3 \dots) = .y_1 y_2 \dots \in B$ which is the condition that $x \in A'$. Therefore

$$(3.8) \quad A \subset A'.$$

Now let $x \in A'$ and let $x_i = b_i(x)$, $i = 1, 2, \dots$. Then $f(Tx) \in B$, i.e., $f(.x_2 x_3 \dots) = .y_1 y_2 \dots$ where $.y_1 y_2 \dots \in B$. It follows that $f(.x_1 x_2 \dots) = .y_1 y_2 \dots$ for some $y \in S$, or

$f(x) \in T^{-1}B$. Thus $x \in A$ so

$$(3.9) \quad A' \subset A.$$

Inclusions (3.8) and (3.9) give (3.7).

Now for any $B \in \mathcal{B}$,

$$\begin{aligned} \lambda(T^{-1}B) &= \mu(\{x: f(x) \in T^{-1}B\}) \\ &= \mu(T^{-1}\{x: f(x) \in B\}) \\ &= \mu(\{x: f(x) \in B\}) \\ &= \lambda(B), \end{aligned}$$

where the next to the last equality follows by the stationarity of μ . Thus λ is stationary.

If $B \in \mathcal{B}$ is such that $T^{-1}B = B$, then by (3.7) we have

$$\begin{aligned} T^{-1}\{x: f(x) \in B\} &= \{x: f(x) \in T^{-1}B\} \\ &= \{x: f(x) \in B\}. \end{aligned}$$

It follows by the ergodicity of μ that $\lambda(B)$ (which is equal to $\mu(\{x: f(x) \in B\})$) is either zero or one. Thus λ is ergodic.

To see that P is stationary it suffices to prove that

$$P(T_1^{-1}(B \times C)) = P(B \times C)$$

for all $B, C \in \mathcal{B}$ (Billingsley [5], p. 4). We see that

$$\begin{aligned} P(T_1^{-1}(B \times C)) &= P(T^{-1}B \times T^{-1}C) \\ &= \mu(\text{proj}_x\{(T^{-1}B \times T^{-1}C) \cap \text{graph of } f\}) \\ &= \mu(\{x: f(x) \in T^{-1}C\} \cap T^{-1}B) \end{aligned}$$

$$\begin{aligned}
&= \mu(T^{-1}\{x: f(x) \in C\} \cap T^{-1}B) \\
&= \mu(T^{-1}(\{x: f(x) \in C\} \cap B)) \\
&= \mu(\{x: f(x) \in C\} \cap B) \\
&= P(B \times C). \quad \#
\end{aligned}$$

Now for $\mu \in \mathcal{M}$, all three measures μ , λ , and P are stationary so their entropies are defined. We let

$$(3.10) \quad R_{\mu} = H(\mu) + H(\lambda) - H(P)$$

and

$$(3.11) \quad C' = \sup_{\mu \in \mathcal{M}} R_{\mu}.$$

The number R_{μ} is called the rate of transmission of the transducer with respect to μ . See Billingsley [5] for an intuitive interpretation of R_{μ} . We now show that C' can be called the transducer capacity by proving the following theorem.

Theorem 3.12. $C' = C$.

Proof: For any integer $n \geq 1$, the triple (S^{n+m-1}, S^n, f_n) forms a discrete, noiseless, memoryless channel, where we think of a transmitted sequence $x \in S^{n+m-1}$ being received as the sequence $f_n(x) \in S^n$. The capacity C_n of this channel is easily computed to be $\log N(n)$ (see Feinstein [10] for the definition of capacity of a memoryless channel). Feinstein [11] has shown that

$$\lim_{n \rightarrow \infty} \frac{C_n}{n} = C'. \text{ But since } \lim_{n \rightarrow \infty} \frac{C_n}{n} = \lim_{n \rightarrow \infty} \frac{\log N(n)}{n} = C, \text{ we have } C = C'. \quad \#$$

IV. The Dimension of the Received Set

Let $f^* \in \mathfrak{F}_m$ and let f be the corresponding transducer of memory m . Let $Y = f((0,1])$ be the range of f (Y is the collection of all possible received sequences). In this section we show that the Hausdorff dimension of Y is C , the transducer capacity.

Lemma 4.1. The expression (3.10) for the rate of transmission R_μ of the transducer reduces to $R_\mu = H(\lambda)$.

Proof: Since by (3.10),

$$R_\mu = H(\mu) + H(\lambda) - H(P),$$

we must show that $H(\mu) - H(P) = 0$. We will make use of the following two facts: if p, q , and r are positive real numbers then

$$(4.2) \quad (p+q) \log(p+q) \geq p \log p + q \log q$$

and

$$(4.3) \quad (p+q) \log(p+q) + r \log r \leq (p+q+r) \log(p+q+r).$$

Both facts follow from the monotonicity of the logarithm.

Now

$$\begin{aligned} H_n(P) &= -\sum P([x_1, \dots, x_n, y_1, \dots, y_n]) \log P([x_1, \dots, x_n, y_1, \dots, y_n]) \\ (4.4) &= -\sum \mu([x_1, \dots, x_n] \cap \{x: f(x) \in [y_1, \dots, y_n]\}) \\ &\quad \log \mu([x_1, \dots, x_n] \cap \{x: f(x) \in [y_1, \dots, y_n]\}). \end{aligned}$$

The set $[x_1, \dots, x_n] \cap \{x: f(x) \in [y_1, \dots, y_n]\}$ is either empty or is the disjoint union of cylinders (b-adic intervals) of length $b^{-(n+m-1)}$. Then by using (4.2) and (4.3) on the summation (4.4) it is seen that

$$H_n(\mu) \leq H_n(P) \leq H_{n+m-1}(\mu).$$

Thus

$$\begin{aligned} H(\mu) &= \lim_{n \rightarrow \infty} \frac{H_n(\mu)}{n} \leq \lim_{n \rightarrow \infty} \frac{H_n(P)}{n} = H(P) \\ &\leq \lim_{n \rightarrow \infty} \frac{H_{n+m-1}(\mu)}{n} = \lim_{n \rightarrow \infty} \frac{H_{n+m-1}(\mu)}{n+m-1} = H(\mu). \end{aligned}$$

Hence $H(\mu) = H(P)$. #

Theorem 4.5. $\dim Y = C$.

Proof: Clearly, for each positive integer n , $N(n)$ b-adic intervals of length b^{-n} will cover Y . Let $\rho > 0$ and $\epsilon > 0$ be given, and choose a positive integer k such that $b^{-k} < \rho$ and $C + \epsilon > \frac{\log N(k)}{k}$. Then

$$\begin{aligned} L_{C+\epsilon}(Y, \rho) &\leq N(k)b^{-k(C+\epsilon)} \\ &< N(k)b^{-\log N(k)} = 1. \end{aligned}$$

Since ρ was arbitrary, it follows that $L_{C+\epsilon}(Y) \leq 1$, and thus $\dim Y \leq C + \epsilon$ by (2.3). Since ϵ was arbitrary, we have $\dim Y \leq C$.

We now show that $\dim Y \geq C$. Let $\epsilon > 0$ be given and choose $\mu \in \mathcal{M}$ so that $R_\mu > C' - \epsilon$. Then by Theorem 3.12 and Lemma 4.1 we have $H(\lambda) > C - \epsilon$. Now $\lambda \in \mathcal{M}$ by Theorem 3.6 and thus the

Shannon-McMillan-Breiman theorem (see Billingsley [5]) shows that

$$(4.6) \quad \lim_{n \rightarrow \infty} -\frac{1}{n} \log \lambda([b_1(y), \dots, b_n(y)]) = H(\lambda) \text{ a.e. } [\lambda].$$

Let M be the set of y 's for which (4.6) holds. Then $M \cap Y \subset M$ and by a general theorem due to Billingsley [5], p. 141, we have

$$\dim M \cap Y = H(\lambda) \dim_{\lambda} M \cap Y.$$

Now $\lambda(M) = 1$ and $\lambda(Y) = \mu(\{x: f(x) \in Y\}) = \mu((0,1]) = 1$ so $\lambda(M \cap Y) = 1$. Thus $\dim_{\lambda} M \cap Y = 1$ by (2.7), so

$$\dim M \cap Y = H(\lambda) > C - \epsilon.$$

Since ϵ was arbitrary, we have $\dim M \cap Y \geq C$, and since $M \cap Y \subset Y$, we have $\dim Y \geq \dim M \cap Y \geq C$. #

We remark that $\dim M \cap Y = H(\lambda)$ can also be shown by using a theorem of Dym [7, Theorem 2]. Dym provides a direct proof not involving Billingsley's general theorem.

We also note that both definitions of capacity given in Section III were used in proving Theorem 4.5. The C definition was used in showing $\dim Y \leq C$ and the C' definition was used in showing $\dim Y \geq C$.

We next construct a set Y' which has the same dimension as Y and which illustrates the structure of the received set. For each integer $n \geq 1$, let $I_1^n, I_2^n, \dots, I_{N(n)}^n$ be the $N(n)$ closed b -adic intervals of length b^{-n} which cover Y . Let

$$Y_n = I_1^n \cup I_2^n \cup \dots \cup I_{N(n)}^n,$$

and let

$$Y' = \bigcap_{n=1}^{\infty} Y_n.$$

It is clear that $Y' \supset Y$ since $Y_n \supset Y$ for all n , and since the dimension of a set is not changed by the addition of a countable number of points, the following theorem shows that $\dim Y' = \dim Y$.

Theorem 4.7. The set difference $Y' - Y$ is countable.

Proof: If $Y' - Y = \emptyset$ the theorem is true. Thus we assume

$Y' - Y \neq \emptyset$ and we let $y \in Y' - Y$. We assume for the moment that y is irrational; then the expansion $y = \sum_{i=1}^{\infty} y_i b^{-i}$ is unique.

For each integer $n \geq 1$ there exists an integer j_n satisfying $1 \leq j_n \leq N(n)$ such that $y \in I_{j_n}^n$. Let $E_n = f^{-1}(I_{j_n}^n)$; E_n is the union of those half-open (open on the left) b -adic intervals of length $b^{-(n+m-1)}$ which map under f into $I_{j_n}^n$. Since y is not an endpoint of any $I_{j_n}^n$, we have $I_{j_n}^n \supset I_{j_{n+1}}^{n+1}$, and thus $E_n \supset E_{n+1}$.

By the finite intersection property (see, for example, Hocking and Young [15], p. 19) $E = \bigcap_{n=1}^{\infty} \bar{E}_n \neq \emptyset$, where $\bar{E}_n$ is the closure of E_n . Since $y \notin Y$, $E' = \bigcap_{n=1}^{\infty} E_n = \emptyset$. Let $x \in E$; then $x \in \bar{E}_n$ for all n . Since $E' = \emptyset$, there exists a positive integer M such that $x \notin E_n$ for all $n \geq M$. Therefore, for all $n \geq M$, x is the left hand endpoint of one of the intervals comprising E_n . Thus x is a b -adic point.

Let $x = \sum_{i=1}^{\infty} x_i b^{-i}$ be the terminating expansion of x . Then since for any $n \geq M$, $x_1, \dots, x_{n+m-1}$ represent one of the intervals of E_n , we have $f_n(x_1 \dots x_{n+m-1}) = y_1 \dots y_n$. But since $x_i = 0$ for all i larger than some positive integer, we

must have $y_i = f^*(\underbrace{0\dots 0}_m \text{ zeros})$ for all i larger than some positive integer. This is a contradiction since the expansion of y , y being irrational, is nonrepeating. Therefore, since assuming that the set $Y' - Y$ contained an irrational point led to a contradiction, $Y' - Y$ must contain only rational points. #

That the set $Y' - Y$ may indeed be nonempty is shown by the following example.

Example 4.8. Let $b = 3$, $m = 1$, and let $f^* \in \mathfrak{F}_1$ be defined by $f^*(0) = 1$, $f^*(1) = 0$, and $f^*(2) = 2$. Then the ternary fraction $y = \sum_{i=1}^{\infty} y_i 3^{-i}$, $y_i = 1$ for all i , belongs to Y' since for each n the point $f(\underbrace{.0\dots 0}_{n \text{ zeros}} 111\dots)$ matches y in the first n places. However, $y \notin Y$ since the point 0 is not in the domain of f , and no other point can possibly map into y .

V. The Ambiguity of the Transducer

Let $f^* \in \mathfrak{F}_m$, $m \geq 1$, and let f be the corresponding transducer of memory m . Let $y \in (0,1]$ and for $n \geq 1$, let

$$M_n(y) = \text{crd } f_n^{-1}(b_1(y) \dots b_n(y)),$$

i.e., $M_n(y)$ is the number of input sequences of length $n+m-1$ which map into the output sequence $b_1(y) \dots b_n(y)$. In this section we examine the quantity $\lim_{n \rightarrow \infty} \frac{\log M_n(y)}{n}$, called the ambiguity of the transducer at the point y , and we examine the dimension of the set $M(y) = f^{-1}(y) = \{x: f(x) = y\}$. We will call $M(y)$ the ambiguity set of y .

When y is a b -adic point, the set $M(y)$ consists of two disjoint parts, namely,

$$(5.1) \quad A = \{x: f^*(b_i(x) \dots b_{i+m-1}(x)) \text{ is the } i\text{th digit of the non-terminating expansion of } y, \text{ for all } i \geq 1\}$$

and

$$(5.2) \quad B = \{x: f^*(b_i(x) \dots b_{i+m-1}(x)) \text{ is the } i\text{th digit of the terminating expansion of } y, \text{ for all } i \geq 1\}.$$

In this case we wish to consider only the set A so, for y a b -adic point, we redefine $M(y)$ to be the set A .

For the moment we restrict our attention to the case $m \geq 2$. We begin by defining a set of matrices and stating a theorem which allow computation of $M_n(y)$. These definitions and the theorem

appear in a report by Hedlund [13].

Let r and s be integers satisfying $0 \leq r \leq b^{m-1} - 1$, $0 \leq s \leq b^{m-1} - 1$, and let $r_1 r_2 \dots r_{m-1}$ and $s_1 s_2 \dots s_{m-1}$ be the b -adic representations of r and s respectively, i.e.,

$$r = r_1 b^{m-2} + \dots + r_{m-2} b + r_{m-1}$$

and

$$s = s_1 b^{m-2} + \dots + s_{m-2} b + s_{m-1}.$$

A sequence $x = x_1 x_2 \dots x_k \in S^k$ is said to begin with r and end with s provided the initial $(m-1)$ -sequence of x is $r_1 \dots r_{m-1}$ and the terminal $(m-1)$ -sequence of x is $s_1 \dots s_{m-1}$.

Let $A^i = (a_{rs}^i)$, $0 \leq i \leq b-1$, be the square matrices of order b^{m-1} defined as follows. For $0 \leq r, s \leq b^{m-1} - 1$, a_{rs}^i is the number of members of $f_1^{-1}(i) = (f^*)^{-1}(i)$ which begin with r and end with s .

Theorem 5.3. Let $y = y_1 \dots y_k \in S^k$ and let $W(y) = W(y_1 \dots y_k) = \begin{smallmatrix} y_1 & y_2 & \dots & y_k \\ A^1 & A^2 & \dots & A^k \end{smallmatrix}$. Then w_{rs} , the r, s entry of $W(y)$, $0 \leq r, s \leq b^{m-1} - 1$, is the number of members of $f_k^{-1}(y_1 \dots y_k)$ which begin with r and end with s .

Proof: The proof is by induction on the length k of the sequences. See Hedlund [13]. #

By the weight of a matrix A , denoted by $|A|$, we will mean the sum of all the entries of A .

Corollary 5.4. If $y \in (0, 1]$, then $M_n(y) = |W(b_1(y) \dots b_n(y))|$.

We now show that the ambiguity exists and has the same value for almost all y . We begin with a lemma.

Lemma 5.5. Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $n \times n$ matrices with nonnegative entries. Then $|AB| \leq |A||B|$.

Proof: We see that

$$(5.6) \quad |AB| = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ik} b_{kj}$$

and

$$(5.7) \quad |A||B| = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} \right) \left(\sum_{i=1}^n \sum_{j=1}^n b_{ij} \right).$$

It is clear that each term in the summation (5.6) appears on the right hand side of (5.7), and since all terms are nonnegative we have $|AB| \leq |A||B|$. #

We next state a theorem of Furstenberg and Kesten [12, Theorem 2] which is needed for the proof of Theorem 5.8.

Theorem (Furstenberg and Kesten). Let $X^1, X^2, \dots$ be a metrically transitive, stationary stochastic process with values in the set of $k \times k$ matrices. Define the norm of a matrix $A = (a_{ij})$ by $\|A\| = \max_{i,j} |a_{ij}|$, and for any positive real number t , let $\log^+ t = \max(\log t, 0)$. If $E(\log^+ \|X^1\|) < \infty$, then

$$\lim_{n \rightarrow \infty} \frac{\log \|X^n X^{n-1} \dots X^1\|}{n} = D$$

with probability one, where D is equal to

$$\lim_{n \rightarrow \infty} \frac{E(\log \|X^n X^{n-1} \dots X^1\|)}{n} \quad (\text{a limit which is shown to exist by}$$

Furstenberg and Kesten [12, Theorem 1]).

Theorem 5.8. Let f be a transducer of memory $m \geq 2$, let $\mu \in \mathcal{M}$, and let λ be defined by (3.5). Then $\lim_{n \rightarrow \infty} \frac{\log M_n(y)}{n}$ exists and has the same value for almost all $y [\lambda]$.

Proof: We define a stochastic process $Y_1, Y_2, \dots$ with domain $(0,1]$ and values in the set of $b^{m-1} \times b^{m-1}$ matrices as follows. For $y \in (0,1]$ define $Y_n(y) = A_T^{b_n(y)}$, where the subscript T denotes the transpose operation. By Theorem 3.6 we have $\lambda \in \mathcal{M}$. Then by the Furstenberg and Kesten theorem we have

$$(5.9) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \|Y_n(y) \dots Y_1(y)\|$$

exists and has the same value for almost all $y \in [\lambda]$. We may rewrite (5.9) as

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|(A^{b_1(y)} \dots A^{b_n(y)})_T\|$$

or

$$(5.10) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \|W_T(b_1(y) \dots b_n(y))\|.$$

The only property of the norm used by Furstenberg and Kesten in their proofs is that $\|AB\| \leq \|A\| \|B\|$ for any two matrices A and B . Since the matrices A^i are all nonnegative, we have the same inequality when the norm is replaced by the weight (Lemma 5.5). Hence we may use the weight instead of the norm in (5.10). Then, noting that

$$|W_T(b_1(y) \dots b_n(y))| = |W(b_1(y) \dots b_n(y))|,$$

and using Corollary 5.4, we have $\lim_{n \rightarrow \infty} \frac{\log M_n(y)}{n}$ exists and has the same value for almost all $y \in [\lambda]$. #

If we let D be the limit in Theorem 5.8, we have the following theorem.

Theorem 5.11. Let f be a transducer of memory $m \geq 2$, let $\mu \in \mathcal{M}$, and let λ be defined by (3.5). Then $\dim M(y) \leq D$ for almost all y $[\lambda]$.

Proof: Let $E = \{y: \lim_{n \rightarrow \infty} \frac{\log M_n(y)}{n} = D\}$. By Theorem 5.8, $\lambda(E) = 1$. Let $y \in E$. Clearly, for each positive integer n , $M_n(y)$ b -adic intervals of length $b^{-(n+m-1)}$ will cover $M(y)$. (It is here that we use the fact that for y a b -adic point, $M(y)$ is defined by the set (5.1). See Example 5.13). Now let $\rho > 0$ and $\epsilon > 0$ be given, and choose a positive integer k such that $b^{-(k+m-1)} < \rho$ and $D + \epsilon > \frac{\log M_k(y)}{k}$. Then

$$\begin{aligned} L_{D+\epsilon}(M(y), \rho) &\leq M_k(y) b^{-(k+m-1)(D+\epsilon)} \\ &< M_k(y) b^{-\log M_k(y)} b^{-(m-1)(D+\epsilon)} \leq 1. \end{aligned}$$

Since ρ was arbitrary, it follows that $L_{D+\epsilon}(M(y)) \leq 1$, and thus $\dim M(y) \leq D + \epsilon$ by (2.3). Since ϵ was arbitrary, we have

$$(5.12) \quad \dim M(y) \leq D.$$

Since (5.12) holds for all $y \in E$, the proof is completed. #

Example 5.13. We demonstrate in this example why for y a b -adic point, $M(y)$ is defined by the set (5.1). Let $b = 2$, $m = 2$, and let $f^* \in \mathfrak{Z}_2$ be defined by $f^*(00) = f^*(11) = 0$ and $f^*(01) = f^*(10) = 1$. Let y be the dyadic point $.0111\dots$, and note that $.1000\dots$ is also equal to y . Now $M_1(y) = \text{crd}(f^*)^{-1}(0) = \text{crd}\{00, 11\} = 2$. The two intervals $(0, 1/4]$ and $(3/4, 1]$ represented by the two sequences 00 and 11 do not cover $f^{-1}(y)$ since the point $x = .0111\dots$ belongs to $f^{-1}(y)$ but does not belong to either $(0, 1/4]$ or $(3/4, 1]$. However, if

$M(y)$ is defined to be the set (5.1) and not the set $f^{-1}(y)$, then it is clear that for all $n \geq 1$, the $M_n(y)$ intervals represented by the set of sequences $f_n^{-1}(b_1(y) \dots b_n(y))$ will cover $M(y)$, a property which is essential for the proof of Theorem 5.11.

The discussion so far in this section has been for transducers of memory $m \geq 2$, since the matrices A^i are not defined for $m = 1$. We now examine the case $m = 1$ separately.

Let $f^* \in \mathfrak{S}_1$ with f the corresponding transducer of memory 1. We represent f^* by a $b \times b$ matrix $D = (d_{ij})$, $0 \leq i, j \leq b-1$, as follows. The entry d_{ij} is one if $f^*(i) = j$ and is zero if $f^*(i) \neq j$. Thus each row of D contains a single entry of one and all other entries of the row are zero. Define $\ell_j = \sum_{i=0}^{b-1} d_{ij}$, $0 \leq j \leq b-1$; ℓ_j is the number of elements of S which map to j under f^* .

Let $\mu \in \mathcal{M}$ and define $p_i = \mu([i])$, $0 \leq i \leq b-1$ and $q_i = \sum_{j=0}^{b-1} d_{ji} p_j$, $0 \leq i \leq b-1$. We note that λ defined by (3.5) is such that $\lambda([i]) = q_i$, $0 \leq i \leq b-1$, since $\lambda([i]) = \mu(\{x: f(x) \in [i]\}) = \sum' \mu([j]) = q_i$, where $\sum'$ denotes the summation taken over those j for which $f^*(j) = i$.

We now prove the equivalent of Theorem 5.8 for the case $m = 1$. The proof is similar to that of Theorem 5.8, using ℓ_i instead of A^i and the ergodic theorem instead of the Furstenberg and Kesten theorem.

Theorem 5.14. Let f be a transducer of memory 1, let $\mu \in \mathcal{M}$, and let λ be defined by (3.5). Then the ambiguity at the point

$y, \lim_{n \rightarrow \infty} \frac{\log M_n(y)}{n}$, exists and has the same value (namely,

$q_0 \log \ell_0 + \dots + q_{b-1} \log \ell_{b-1}$) for almost all $y \in [\lambda]$.

Proof: Let $h_n^i(y)$ be the number of occurrences of i , $0 \leq i \leq b-1$, among the first n digits of the nonterminating base b expansion of y . By Theorem 3.6, $\lambda \in \mathcal{M}$ and we may apply the pointwise ergodic theorem (see Billingsley [5], p. 13) to conclude that for each i , $0 \leq i \leq b-1$,

$$(5.15) \quad \lim_{n \rightarrow \infty} \frac{h_n^i(y)}{n} = q_i$$

for almost all $y \in [\lambda]$. Let D_i , $0 \leq i \leq b-1$, be the set of y 's for which (5.15) holds. Then for $y \in F = \bigcap_{i=0}^{b-1} D_i$, $\lim_{n \rightarrow \infty} \frac{h_n^i(y)}{n} = q_i$ for all i , and $\lambda(F) = 1$ since $\lambda(D_i) = 1$, $0 \leq i \leq b-1$.

Now since for each $n \geq 1$ we obviously have

$$M_n(y) = \ell_0^{h_n^0(y)} \dots \ell_{b-1}^{h_n^{b-1}(y)},$$

then

$$\frac{\log M_n(y)}{n} = \frac{h_n^0(y)}{n} \log \ell_0 + \dots + \frac{h_n^{b-1}(y)}{n} \log \ell_{b-1},$$

where we take 0^0 to be one and $0 \log 0$ to be zero. Hence if $y \in F$, we have $\lim_{n \rightarrow \infty} \frac{\log M_n(y)}{n}$ exists and is equal to $q_0 \log \ell_0 + \dots + q_{b-1} \log \ell_{b-1}$. Since $\lambda(F) = 1$, the proof is completed. #

We now prove the equivalent of Theorem 5.11 for the case $m = 1$. In this case we are able to obtain an equality for the dimension of the ambiguity set rather than just an upper bound. We begin with a lemma.

Lemma 5.16. Let $y \in (0,1]$ and let $E_n = U'[x_1, \dots, x_n]$ where U' denotes the union over those $x_1, \dots, x_n$ such that

$f_n(x_1 \dots x_n) = b_1(y) \dots b_n(y)$ (there are clearly $M_n(y)$ such sequences $x_1, \dots, x_n$). Then

$$\bigcap_{n=1}^{\infty} E_n = M(y).$$

Proof: Let $x \in M(y)$. Then $f^*(b(x_i))$ equals the i th digit of the nonterminating expansion of y , i.e., $f^*(b(x_i)) = b_i(y)$.

Thus $f_n(b_1(x) \dots b_n(x)) = b_1(y) \dots b_n(y)$ for all n . Hence $x \in E_n$ so $M(y) \subset E_n$ for all n . Therefore

$$(5.17) \quad M(y) \subset \bigcap_{n=1}^{\infty} E_n.$$

Let $x \in \bigcap_{n=1}^{\infty} E_n$. Then $f_n(b_1(x) \dots b_n(x)) = b_1(y) \dots b_n(y)$ for all n so $f^*(b_n(x)) = b_n(y)$ for all n . Thus $x \in M(y)$ so

$$(5.18) \quad \bigcap_{n=1}^{\infty} E_n \subset M(y).$$

Inclusions (5.17) and (5.18) give the desired result. #

Theorem 5.19. Let f be a transducer of memory 1, let $\mu \in \mathcal{M}$, and let λ be defined by (3.5). Then

$$\dim M(y) = \sum_{i=0}^{b-1} q_i \log \ell_i$$

for almost all $y \in [\lambda]$.

Proof: Let $y \in (0,1]$ and let $y_i = b_i(y)$ for all $i \geq 1$. For each integer $k \geq 1$ and each sequence $x_1 \dots x_k \in S^k$, define a function p_k on S^k by

$$p_k(x_1 \dots x_k) = \begin{cases} 1/M_k(y) = 1/\ell_{y_1} \dots \ell_{y_k} & \text{if } f_k(x_1 \dots x_k) = y_1 \dots y_k \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that

$$(5.20) \quad p_k(x_1 \dots x_k) \geq 0 \quad \text{for all } k.$$

Also we have

$$(5.21) \quad \sum_{i \in S} p_1(i) = \ell_{y_1} / \ell_{y_1} = 1.$$

We show next that

$$(5.22) \quad \sum_{i \in S} p_{k+1}(x_1 \dots x_k i) = p_k(x_1 \dots x_k).$$

If $p_k(x_1 \dots x_k) = 0$ then $f_k(x_1 \dots x_k) \neq y_1 \dots y_k$ so $f_{k+1}(x_1 \dots x_k i) \neq y_1 \dots y_k y_{k+1}$ for all $i \in S$. Hence $p_{k+1}(x_1 \dots x_k i) = 0$ for all $i \in S$ so (5.22) is true in this case. If

$p_k(x_1 \dots x_k) = 1/\ell_{y_1} \dots \ell_{y_k}$ then $f_{k+1}(x_1 \dots x_k i) = y_1 \dots y_k y_{k+1}$ for the $\ell_{y_{k+1}}$ values of i for which $f^*(i) = y_{k+1}$. Hence

$p_{k+1}(x_1 \dots x_k i)$ is $1/\ell_{y_1} \dots \ell_{y_k} \ell_{y_{k+1}}$ for those $\ell_{y_{k+1}}$ values of i and is zero for the remaining values of i . Therefore

$$\sum_{i \in S} p_{k+1}(x_1 \dots x_k i) = \ell_{y_{k+1}} / \ell_{y_1} \dots \ell_{y_{k+1}} = p_k(x_1 \dots x_k)$$

so (5.22) is also true in this case.

Finally we see that for any sequence $x_1, \dots, x_k$ of elements of S , we have

$$(5.23) \quad \lim_{n \rightarrow \infty} p_{k+n}(x_1 \dots x_k \underbrace{0 \dots 0}_{n \text{ 0's}}) = 0.$$

It follows from (5.20), (5.21), (5.22) and (5.23) that there exists a probability measure ν_y on the unit interval such that

$$\nu_y([x_1, \dots, x_k]) = p_k(x_1 \dots x_k),$$

where $x_1 \dots x_k \in S^k$ (see Billingsley [5], p. 35).

Now let F be the set of y 's for which Theorem 5.14 holds, and let $y \in F$. If $x \in M(y)$ and if we set $x_i = b_i(x)$ and $y_i = b_i(y)$, then

$$\begin{aligned} \lim_{n \rightarrow \infty} - \frac{\log \nu_y([x_1, \dots, x_n])}{n} &= \lim_{n \rightarrow \infty} - \frac{\log(1/\ell_{y_1} \dots \ell_{y_n})}{n} \\ &= \lim_{n \rightarrow \infty} \frac{\log \ell_{y_1} \dots \ell_{y_n}}{n} = \lim_{n \rightarrow \infty} \frac{\log M_n(y)}{n} = \sum_{i=0}^{b-1} q_i \log \ell_i. \end{aligned}$$

The next to the last equality follows from the fact that we obviously have $M_n(y) = \ell_{y_1} \dots \ell_{y_n}$, and the last equality follows from Theorem 5.14. By a theorem of Billingsley [5], p. 141, we now have

$$(5.24) \quad \dim M(y) = \dim_{\nu_y} M(y) = \sum_{i=0}^{b-1} q_i \log \ell_i$$

for all $y \in F$.

If E_n is defined as in Lemma 5.15, it is seen that $\nu_y(E_n) = 1$. Since $\bigcap_{n=1}^{\infty} E_n = M(y)$ by Lemma 5.15 and since $\{E_n\}$ is a decreasing sequence, we have $\nu_y(M(y)) = 1$. Hence $\dim_{\nu_y} M(y) = 1$ by (2.7) and thus from (5.24) we have

$$\dim M(y) = \sum_{i=0}^{b-1} q_i \log \ell_i$$

for all $y \in F$. Since $\lambda(F) = 1$ by Theorem 5.14, the proof is completed. #

For the case $m = 1$ we were able to calculate the dimension of the ambiguity set $M(y)$ (Theorem 5.19), whereas in the case $m \geq 2$, we were able to obtain only an upper bound on $\dim M(y)$ (Theorem 5.11). Conjecture is that for the case $m \geq 2$, $\dim M(y)$ is actually equal to this upper bound. However, the method of proof of Theorem 5.19 cannot be used to prove this conjecture, since the consistency condition (5.22) may not hold for $m \geq 2$. The following example demonstrates this fact.

Example 5.25. Let $b = 2$, $m = 2$, and let $f^* \in \mathfrak{S}_2$ be defined by $f^*(00) = 0$ and $f^*(01) = f^*(10) = f^*(11) = 1$. The function f_2 is then as follows:

Domain Value	Functional Value
0 0 0	0 0
0 0 1	0 1
0 1 0	1 1
0 1 1	1 1
1 0 0	1 0
1 0 1	1 1
1 1 0	1 1
1 1 1	1 1

Let $y = .11$; we see that $M_1(y) = \text{crd} \{01, 10, 11\} = 3$ and $M_2(y) = \text{crd} \{010, 011, 101, 110, 111\} = 5$. If we define (as in the proof of Theorem 5.19) $p_2(00) = 0$, $p_2(01) = p_2(10) = p_2(11) = 1/3$ and $p_3(000) = p_3(001) = p_3(100) = 0$, $p(010) = p(011) = p(101) = p(110) = p(111) = 1/5$, then it is clear that the consistency condition (5.22) does not hold. Thus for any $y \in (3/4, 1]$, we cannot define a measure ν_y as we did in Theorem 5.19.

VI. Conclusion

This paper has begun an examination of the relationship between an information channel and the unit square. In particular, we have examined a special type of noiseless channel with memory called a transducer. This same transducer has been examined from a topological dynamics standpoint by Hedlund [14]. We now comment on our results and mention some difficulties encountered.

Theorem 4.5 shows that capacity is a reasonable term for the quantity C in the following sense. The "size" (dimension) of the set of all possible received sequences is a measure of the "capacity" of the transducer, since under no circumstances can the transducer be made to create new output sequences and hence increase the "size" of this set. This interpretation of "size" as "capacity", and the fact that the quantity C is equal to this "size" (Theorem 4.5), make capacity a reasonable term for C .

Attempts at calculating C explicitly by finding a difference equation satisfied by $N(n)$ were unsuccessful. During these attempts it was discovered that the difference equation method used by Shannon [19, Appendix 1] for his finite state sources calculates the number of state sequences of a given length rather than the number of output sequences as desired. Conant [6] is also aware of this problem.

Theorems 5.8, 5.11, 5.14, and 5.19 show that the transducer has a homogeneity property since almost all received sequences have the same ambiguity. The usual quantity used for ambiguity in information theory is a conditional entropy (called the equivocation by most authors) which represents the ambiguity averaged over all received sequences. Our pointwise ambiguity has a geometric interpretation and gives results involving almost everywhere statements which are somewhat more satisfying than statements involving averages. Yet to be resolved is the question of whether or not the inequality of Theorem 5.11 is actually an equality.

One direction in which our work may be extended is to consider more general channels by introducing the element of noise. Another direction is to consider how to make the transducer invertible, i.e., find a subset of the unit interval of measure one such that the transducer is a one-to-one function onto the set of all received sequences when restricted to this subset.

BIBLIOGRAPHY

1. Besicovitch, A.S., On the sum of digits of real numbers represented in the dyadic system, Math. Ann. 110(1935), 321-330.
2. Billingsley, P., Hausdorff dimension in probability theory, Ill. J. Math. 4(1960), 187-209.
3. _____, Hausdorff dimension in probability theory II, Ill. J. Math. 5(1961), 291-298.
4. _____, On the coding theorem for the noiseless channel, Ann. Math. Statist. 32(1961), 594-601.
5. _____, Ergodic Theory and Information, John Wiley and Sons, New York (1965).
6. Conant, R.C., Channel capacity of Moore automata, Information and Control 12(1968), 453-465.
7. Dym, H., On a class of monotone functions generated by ergodic sequences, Amer. Math. Monthly 75(1968), 594-601.
8. Eggleston, H.G., The fractional dimension of a set defined by decimal properties, Quart. J. Math. Oxford Ser. 20(1949), 31-36.
9. _____, Sets of fractional dimensions which occur in some problems of number theory, Proc. London Math. Soc. (2) 54 (1952), 42-93.
10. Feinstein, A., Foundations of Information Theory, McGraw-Hill, New York (1958).
11. _____, On the coding theorem and its converse for finite-memory channels, Information and Control 2(1959), 25-44.
12. Furstenberg, H. and H. Kesten, Products of random matrices, Ann. Math. Statist. 31(1960), 457-469.
13. Hedlund, G.A., Mappings on sequence spaces, Communications Research Division Technical Report No. 1, von Neumann Hall, Princeton, New Jersey, February 1961.

14. Hedlund, G.A., Transformations commuting with the shift, in Topological Dynamics, J. Auslander and W. Gottschalk, eds., W.A. Benjamin, New York (1968), 259-289.
15. Hocking, J. and G. Young, Topology, Addison-Wesley, Reading, Massachusetts (1961).
16. Khinchin, A.I., Mathematical Foundations of Information Theory, Dover, New York (1957).
17. Kingman, J.F.C. and S.J. Taylor, Introduction to Measure and Probability, Cambridge University Press, Cambridge (1966).
18. Kinney, J.R., Singular functions associated with Markov chains, Proc. Amer. Math. Soc. 9(1958), 603-608.
19. Shannon, C.E. and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, (1964). (reprinted from Bell System Tech. J. 27(1948), 379-423, 623-656.)
20. Smorodinsky, M., The capacity of a general noiseless channel and its connection with Hausdorff dimension, Proc. Amer. Math. Soc. 19(1968), 1247-1254.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03046 6787