N68-32351

THE RATE DISTORTION FUNCTION FOR A CLASS OF SOURCES

David J. Sakrison

Technical Memo

NASA Grant NGR-05-003-143

Distribution of this document is unlimited.

Series No. 9, Issue No. 35
July 22, 1968

UNIVERSITY OF CALIFORNIA, BERKELEY

THE RATE DISTORTION FUNCTION FOR A CLASS OF SOURCES

David J. Sakrison
Department of Electrical Engineering and Computer Sciences
University of California
Berkeley, California 94720

Technical Memo on NASA Grant
NGR-05-003-143

Space Sciences Laboratory
Series No. 9, Issue No. 35

This work was supported by the National Aeronautics and Space Administration under Grant NGR 05-003-143

LIST OF SYMBOLS

U	random source output symbol
$\tilde{U}$	random receiver output symbol
α	index on source distribution
γ	index on transition distribution
$\mathcal{A}$	class of source distributions
$P_{\alpha\gamma}(u,\tilde{u})$	$\alpha\gamma$ probability distribution
$E_{\alpha\gamma}(\cdot)$	expectation with respect to $\alpha\gamma$ distribution
$I_{\alpha\gamma}(U;\tilde{U})$	average mutual information with respect to $\alpha\gamma$ distribution
$I_{\alpha\gamma}$ or $I_{\alpha\gamma}(u;\tilde{u})$	random variable corresponding to above
$R_\alpha(d)$	rate distortion function for source α
$R_{\mathcal{A}}(d)$	rate distortion function of the class $\mathcal{A}$

Proposed Running Head "Rate of a Class of Sources"

number of pages 41

number of tables 0

number of figures 1

Abstract

Shannon's rate distortion theory applies to a source with a known probability distribution. In many instances an encoder must be designed when only vague a priori knowledge is available concerning the source distribution. We define the rate distortion function for a class of sources, prove a corresponding coding theorem, and present several examples.

1. INTRODUCTION

In this paper we consider the source encoding problem shown in Figure 1. Present results have considered the case in which the source is characterized by a single probability distribution which is known to the designer of the encoding system. This is in distinct contrast to many physical situations, e.g., space experimentation, in which either only vague a priori knowledge of the source is available or only limited data such as estimates of second moments are available to the designer of the encoder. The objective of this paper is to consider what is the best that can be achieved by a single fixed encoding system that must be designed to work with any source from among a class of possible sources. In particular we show that if the rate distortion function is suitably redefined then the positive and negative sides of Shannon's encoding theorem (Shannon, 1948, 1960; Gallager, 1968) still remain true. The treatment involved is to a certain extent the dual of the encoding for a class of noisy channels considered by Blackwell, Breiman, and Thomasian (1959) and our proof of the positive side of the encoding theorem is based on a lemma of Gallager (1968) and the methods used by Blackwell, Breiman and Thomasian (1959).

We denote the random output of the source by U and the random replica of U generated by the decoder at the receiver by $\tilde{U}$. We consider sources whose outputs are sequences of independent identically distributed random quantities U. The random source outputs U take on values u which are points in a separable metric space $\mathcal{U}$ whose metric we denote by $\rho(\cdot,\cdot)$. A source is then characterized by its probability distribution. We will index source distributions by α and denote a class of sources by $\mathcal{A}$. The open sets of $\mathcal{U}$ determined by ρ will be measurable sets for all source distributions considered.

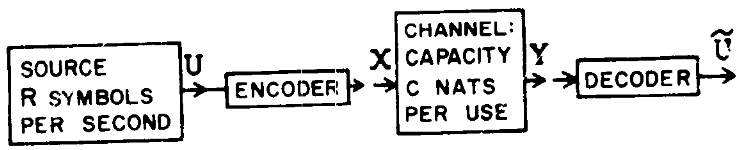

FIG. 1 SOURCE ENCODING PROBLEM

We consider two distortion measures. One, we refer to as the mean distortion measure, is defined by a distortion level in which the distance function $d(U,\tilde{U})$ is just the metric ρ

$$\bar{d} \triangleq E(D) \triangleq E(d(U,\tilde{U})) \triangleq E(\rho(U,\tilde{U})) \tag{1-1}$$

The second distortion measure, which we refer to as the indicator distortion measure, is defined by a distortion level given by

$$\bar{d}_{\rho_0} \triangleq E(D_{\rho_0}) \triangleq E(d_{\rho_0}(U,\tilde{U})) \tag{1-2}$$

in which the distance function d_{ρ_0} is given by

$$d_{\rho_0}(u,\tilde{u}) = \begin{cases} 0 \text{ if } \rho(u,\tilde{u}) < \rho_0 \\ 1 \text{ if } \rho(u,\tilde{u}) \geq \rho_0 \end{cases} \tag{1-3}$$

In most cases the statements we make apply to both distortion measures; only when it is necessary to distinguish between them will the subscript ρ_0 be included in our notation.

Let 0 denote a reference point in $\mathcal{U}$; if $\mathcal{U}$ is also a vector space we take 0 to be the zero vector. We assume for any class $\mathcal{A}$ of sources to be considered that

$$E_\alpha[\rho(U,0)] \leq K_0 < \infty \tag{1-4}$$

for all $\alpha \in \mathcal{A}$. We further assume that any class of sources considered is compact, where we say that $\mathcal{A}$ is compact if for any $\epsilon_1, \epsilon_2 > 0$ we can find a set $\mathcal{U}'$ which can be covered by a finite number of sets of radius at most ϵ_2. We then require that

$$E_\alpha(d(U,0)\, I_{\mathcal{U}'^c}(U)) \leq \epsilon_1 \tag{1-5}$$

for all $\alpha \in \mathcal{A}$, in which d is either the metric or indicator function (depending on which distortion measure is being considered), $\mathcal{U}'^c$ is the complement of $\mathcal{U}'$, and $I_{\mathcal{U}'^c}$ denotes the indicator function of $\mathcal{U}'^c$.

Having delineated the mathematical scope of our work, we now outline the results contained in this paper. In the balance of this section we introduce the necessary notation, define the rate distortion function for a class of sources, and state the coding theorem which is the principal result of this paper. In section 2 we prove the negative side of the coding theorem. In section 3 we prove the positive side of the coding theorem for a finite class of sources and in section 4 this is extended to include a compact class of sources. Section 5 concludes by considering several examples of compact classes of sources for which we can evaluate the rate-distortion function and also discusses limitations on the mathematical structure of our theory. A reader who wishes to first obtain the flavor of our results without coping with the details of the proofs can do so by reading sections 1 and 5 complete, and the statements of Lemma 1 and Theorem 1 and the proof of Theorem 2, all in section 3.

If L source outputs, $u_1, \ldots, u_L$ are blocked together, forming an L-extension of the source, the resulting L-tuple will be denoted by

$$\underset{\sim}{u} = (u_1, \ldots, u_L) \tag{1-6}$$

and the corresponding decoder output by

$$\tilde{\underset{\sim}{u}} = (\tilde{u}_1, \ldots, \tilde{u}_L) \tag{1-7}$$

Similarly, if channel inputs and outputs are denoted by x and y respectively, we denote blocks or N-tuples of such quantities, forming an N-extension of the channel, by

$$\underset{\sim}{x} = (x_1, \ldots, x_N) \qquad (1\text{-}8)$$

$$\underset{\sim}{y} = (y_1, \ldots, y_N) \qquad (1\text{-}9)$$

When L source outputs are blocked together, the random value of distortion averaged over the block is denoted by

$$\bar{D} \triangleq d(\underset{\sim}{U},\underset{\sim}{\tilde{U}}) \triangleq (1/L) \sum_{\ell=1}^{L} d(U_\ell,\tilde{U}_\ell) \triangleq (1/L) \sum_{\ell=1}^{L} D_\ell \qquad (1\text{-}10)$$

As stated earlier, we index source distributions $P_\alpha(u)$ by α and denote a class of possible source distributions by $\mathcal{A}$. We will index transition probabilities, corresponding to a test or hypothetical encoder-channel-decoder mapping, by γ

$$P_\gamma(\tilde{u}|u)$$

Distributions or expectations that depend on both the source and transition distributions will be so labeled

$$P_{\alpha\gamma}(u,\tilde{u}) = P_\gamma(\tilde{u}|u)P_\alpha(u) \qquad (1\text{-}11)$$

$$P_{\alpha\gamma}(\tilde{u}) = \int_{\mathcal{U}} P_\gamma(\tilde{u}|u)dP_\alpha(u) \qquad (1\text{-}12)$$

$$E_{\alpha\gamma}(D) = \int_{\mathcal{U}\times\mathcal{U}} \rho(u,\tilde{u})dP_{\alpha\gamma}(u,\tilde{u}) \qquad (1\text{-}13)$$

The Randon-Nikodym derivative of the measure $P_{\alpha\gamma}(u,\tilde{u})$ with respect to the measure $P_\alpha(u)P_{\alpha\gamma}(\tilde{u})$ we denote by

$$\Lambda_{\alpha\gamma}(u,\tilde{u}) \triangleq \frac{dP_{\alpha\gamma}(u,\tilde{u})}{d[P_\alpha(u)P_{\alpha\gamma}(\tilde{u})]} \qquad (1\text{-}14)$$

When the measure $P_{\alpha\gamma}$ is continuous with respect to Lebesque measure, this quantity is merely the ratio of the corresponding density functions. We denote by $I_{\alpha\gamma}(U;\tilde{U})$ the average mutual information between U and $\tilde{U}$ for the distribution defined by Eq. (1-11). In the only situations of interest this quantity will be finite, the derivative of Eq. (1-14) will exist, and we can take by Pinsker (1964, p. 10) the information density to be

$$I_{\alpha\gamma}(u;\tilde{u}) = \ln \Lambda_{\alpha\gamma}(u,\tilde{u}) \qquad (1\text{-}15)$$

Evaluated for specific values u and $\tilde{u}$, this quantity is a function of u and $\tilde{u}$; when the arguments U and $\tilde{U}$ are random, we denote the resulting random variable by $I_{\alpha\gamma}$. The mutual information $I_{\alpha\gamma}(U;\tilde{U})$ is given in terms of the information density by (Pinsker, 1964, p. 10)

$$I_{\alpha\gamma}(U;\tilde{U}) = E_{\alpha\gamma}(I_{\alpha\gamma}) = \int_{\mathcal{U}\times\mathcal{U}} I_{\alpha\gamma}(u,\tilde{u})dP_{\alpha\gamma}(u,\tilde{u}) \qquad (1\text{-}16)$$

The average mutual information for a block of L source and decoder outputs is denoted simply by

$$I_{\alpha\gamma}(\underset{\sim}{U};\underset{\sim}{\tilde{U}})$$

and the corresponding random variable by $I_{\alpha\gamma}^L$

$$I_{\alpha\gamma}(\underset{\sim}{U};\underset{\sim}{\tilde{U}}) = E_{\alpha\gamma}(I_{\alpha\gamma}^L) \qquad (1\text{-}17)$$

We now define a particular subset of transition distributions. For any non-negative number d we define

$$\Gamma_\alpha(d) = \{\gamma : E_{\alpha\gamma}(D) \le d\} = \{\gamma : E_{\alpha\gamma}(d(U;\tilde{U})) \le d\} \qquad (1\text{-}18)$$

and

$$\Gamma_{\mathcal{A}}(d) = \{\gamma : E_{\alpha\gamma}(D) \le d \text{ for all } \alpha \in \mathcal{A}\} = \bigcap_{\alpha\in\mathcal{A}} \Gamma_\alpha(d) \qquad (1\text{-}19)$$

We use the above notation for either the mean or indicator distortion measure when the threshold value of the indicator function is understood; when it is necessary to specify this threshold value ρ_0, we write

$$\Gamma_{\mathcal{A}}(d,\rho_0) = \{\gamma : E_{\alpha\gamma}(D_{\rho_0}) \le d \text{ all } \alpha \in \mathcal{A}\} \qquad (1\text{-}20)$$

and refer to this as a d, ρ_0 distortion level.

In the above notation, the rate distortion function for a single source α is simply

$$R_\alpha(d) = \inf_{\gamma\in\Gamma_\alpha(d)} I_{\alpha\gamma}(U;\tilde{U}) \qquad (1\text{-}21)$$

For a class $\mathcal{A}$ of sources, we define the rate distortion function to be

$$R_{\mathcal{A}}(d) = \inf_{\gamma\in\Gamma_{\mathcal{A}}(d)} \sup_{\alpha\in\mathcal{A}} I_{\alpha\gamma}(U;\tilde{U}) \qquad (1\text{-}22)$$

Note that in general this rate is greater than or equal to

$$\sup_{\alpha\in\mathcal{A}} \inf_{\gamma\in\Gamma_\alpha(d)} I_{\alpha\gamma}(U;\tilde{U}) = \sup_{\alpha\in\mathcal{A}} R_\alpha(d) \qquad (1\text{-}23)$$

This corresponds to the situation in finding the capacity of a class of channels, where the capacity of the class (even for a finite class) may be strictly less than the minimum capacity of any single channel in the class. In that setting, examples can be given (Blackman, Breiman, and Thomasian, 1959) for which such strict inequality holds. In the source ... show that for a <u>compact class of</u> <u>sources the two rates of Eqs. (1-22) and (1-23) are equal</u>. This is of considerable practical importance, because the rate of Eq. (1-22) is extremely difficult to evaluate, while the rate defined by Eq. (1-23) can be obtained easily in a number of instances.

As in the case of a single source, the function $R_{\mathcal{A}}(d)$ is monotone non-increasing and convex in d. The monotonicity follows simply from the fact that the set $\Gamma_{\mathcal{A}}(d)$ does not diminish with increasing d. The convexity is proven as for a single source (Gallager, 1968); select arbitrary d_1, d_2, and δ, all greater than zero. Let γ_1 and γ_2 be distributions in $\Gamma_{\mathcal{A}}(d_1)$ and $\Gamma_{\mathcal{A}}(d_2)$ respectively such that

$$\sup_{\alpha\in\mathcal{A}} I_{\alpha\gamma_i}(U;\tilde{U}) \le R_{\mathcal{A}}(d_i) + \delta \qquad i = 1, 2$$

Let the distribution γ be given by

$$P_\gamma(\tilde{u}|u) = \theta P_{\gamma_1}(\tilde{u}|u) + (1-\theta)P_{\gamma_2}(\tilde{u}|u)$$

Then from the linearity of the expectation

$$\gamma \in \Gamma_{\mathcal{A}}[\theta d_1 + (1-\theta)d_2]$$

Further, as we shall show in section 2, $I_{\alpha\gamma}(U;\tilde{U})$ is convex in γ so that for any α

$$I_{\alpha\gamma}(U;\tilde{U}) \le \theta I_{\alpha\gamma_1}(U;\tilde{U}) + (1-\theta) I_{\alpha\gamma_2}(U;\tilde{U})$$

Taking the sup with respect to α over both sides of this inequality yields

$$\sup_{\alpha\in\mathcal{A}} I_{\alpha\gamma}(U;\tilde{U}) \le \theta R_{\mathcal{A}}(d_1) + (1-\theta) R_{\mathcal{A}}(d_2) + \delta$$

Since this inequality holds for an arbitrary positive δ, we have

$$R_{\mathcal{A}}[\theta d_1 + (1-\theta)d_2] \le R_{\mathcal{A}}(d_1) + (1-\theta)R_{\mathcal{A}}(d_2) \qquad (1\text{-}24)$$

The significance of the rate $R_{\mathcal{A}}(d)$ lies in the following theorem, the negative side of which is proven in section 2 and the positive side in sections 3 and 4.

Source Encoding Theorem

Negative Statement: Consider designing an encoder which connects the source output U and the decoder output $\tilde{U}$ via a channel and yields a distortion level less than or equal to d for each source α in a class $\mathcal{A}$ (not necessarily compact). Then the capacity of the channel must be at least $R_{\mathcal{A}}(d)$ nats per source symbol.

Positive Statement: Consider blocking sequential source symbols into L-tuples of source symbols and encoding each generated L-tuple of source symbols into one of M possible L-tuples of source symbols. If the class of sources is compact, then given any $d > 0$, $\delta > 0$, it is possible to find a sufficiently large L and a set of M L-tuples such that mapping source outputs into this set of M L-tuples yields distortion

$$E_\alpha(\bar{D}) \le d + \delta \qquad (1\text{-}25)$$
$$\text{all } \alpha \in \mathcal{A}$$

with

$$M \le \exp\{L[R_{sup}(d) + \delta]\} \qquad (1\text{-}26)$$

and

$$R_{sup}(d) = \sup_{\alpha\in\mathcal{A}} R_\alpha(d) \qquad (1\text{-}27)$$

Note that since from the definition of $R_{\mathcal{A}}(d)$

$$R_{\mathcal{A}}(d) \ge \sup_{\alpha\in\mathcal{A}} R_\alpha(d) \qquad (1\text{-}28)$$

the positive and negative sides of the Theorem imply that

$$R_{\mathcal{A}}(d) = \sup_{\alpha\in\mathcal{A}} R_\alpha(d) \qquad (1\text{-}29)$$

for $\mathcal{A}$ a compact class of sources. Since equality holds, one might ask why we do not take the right hand side of Eq. (1-29) as the definition of the rate of a class, since it is a simpler entity than $R_{\mathcal{A}}(d)$. Our answer is that $R_{\mathcal{A}}(d)$ is more basic to the structure of the problem; the negative part of the Theorem holds with $R_{\mathcal{A}}(d)$ for an arbitrary class $\mathcal{A}$, while the positive part holds with $\sup R_\alpha(d)$ only for a compact class.

2. THE NEGATIVE SIDE OF THE CODING THEOREM

Consider encoding blocks of source outputs of length L onto a time-discrete channel of C nats per use. Let N denote the number of channel usages available in the time interval taken to generate L source vectors. If the encoding system yields an average distortion per source vector

$$E_\alpha(\bar{D}) = \frac{1}{L}\sum_{\ell=1}^{L} E_\alpha(D_\ell)$$

less than or equal to d for any $\alpha \in \mathcal{A}$, then the following relation must hold between C and $R_{\mathcal{A}}(d)$

$$NC \ge LR_{\mathcal{A}}(d) \qquad (2\text{-}1)$$

Proof.

$$\underset{\sim}{X} = f(\underset{\sim}{U}) \quad \text{and} \quad \tilde{\underset{\sim}{U}} = g(\underset{\sim}{Y})$$

for some measurable functions f and g. Thus $\underset{\sim}{U}$ is dense in $\underset{\sim}{X}$, $\underset{\sim}{U}$ and $\underset{\sim}{\tilde{U}}$ is subordinate to $\underset{\sim}{Y}$ and (Pinsker, 1964, pp. 37-38)

$$I(\underset{\sim}{U}, \underset{\sim}{X}; \underset{\sim}{Y}) = I(\underset{\sim}{U};\underset{\sim}{Y}) \geq I(\underset{\sim}{U};\underset{\sim}{\tilde{U}}) \tag{2-2}$$

Further, from Kolmogorov's formula (Pinsker, 1964, pp. 37-38) and the fact that $\underset{\sim}{Y}$ depends on $\underset{\sim}{U}$ only through $\underset{\sim}{X}$

$$I(\underset{\sim}{U}, \underset{\sim}{X}; \underset{\sim}{Y}) = I(\underset{\sim}{X};\underset{\sim}{Y}) + I(\underset{\sim}{Y};\underset{\sim}{U}|\underset{\sim}{X}) = I(\underset{\sim}{X};\underset{\sim}{Y}) \tag{2-3}$$

For N independent uses of the channel

$$NC \geq I(\underset{\sim}{X},\underset{\sim}{Y}) \tag{2-4}$$

Combining Eqs. (2-2)-(2-4) yields

$$NC \geq I(\underset{\sim}{U},\underset{\sim}{\tilde{U}}) \tag{2-5}$$

Making repeated use of Kolmogorov's formula

$$I(\underset{\sim}{U};\underset{\sim}{\tilde{U}}) = \sum_{\ell=1}^{L} I(U_\ell;\underset{\sim}{\tilde{U}}|U_1, \ldots, U_{\ell-1}) \tag{2-6}$$

and from Kolmogorov's formula and the independence of the U_ℓ

$$\begin{aligned} I(U_\ell;\underset{\sim}{\tilde{U}}|U_1, \ldots, U_{\ell-1}) &= I(U_\ell;\underset{\sim}{\tilde{U}},U_1, \ldots, U_{\ell-1}) - I(U_\ell;U_1, \ldots, U_{\ell-1}) \\ &= I(U_\ell;\underset{\sim}{\tilde{U}},U_1, \ldots, U_{\ell-1}) \geq I(U_\ell;\tilde{U}_\ell) \end{aligned} \tag{2-7}$$

Combining Eqs. (2-5)-(2-7), we obtain

$$NC \geq \sum_{\ell=1}^{L} I(U_\ell;\tilde{U}_\ell) \tag{2-8}$$

This inequality holds for any $\alpha\gamma$ distribution; the subscripts were omitted in the above for convenience. Now let γ_ℓ denote the distribution induced on U_ℓ, $\tilde{U}_\ell$ by the encoder-channel system and let $\alpha\bar{\gamma}$ denote the distribution

$$P_{\alpha\bar{\gamma}}(u,\tilde{u}) = \sum_{\ell=1}^{L} \frac{1}{L} P_{\gamma_\ell}(\tilde{u}|u)P_\alpha(u) \tag{2-9}$$

The corresponding measure on $\tilde{U}$ is

$$P_{\alpha\bar{\gamma}}(F) = \int_{\mathcal{U} \times F} dP_{\alpha\bar{\gamma}}(u,\tilde{u}) \tag{2-10}$$

for any measurable set F defined by $\tilde{u}$. The average distortion is given by

$$\begin{aligned} \frac{1}{L}\sum_{\ell=1}^{L} E_{\alpha\gamma_\ell}(D_\ell) &= \int d(u,\tilde{u}) \sum_{\ell=1}^{L} \frac{1}{L}\, dP_{\alpha\gamma_\ell}(u,\tilde{u}) \\ &= E_{\alpha\bar{\gamma}}(D) \end{aligned} \tag{2-11}$$

Thus by hypothesis the distribution $\bar{\gamma}$ must be in $\Gamma_{\mathcal{Q}}(d)$.

Now let us consider the right hand side of inequality (2-8). Let $E_j \times F_j$, j = 1, ..., J, denote a partition of $\mathcal{U} \times \mathcal{U}$. We have

$$\begin{aligned} &\sum_{\ell=1}^{L}\sum_{j=1}^{J} \ln\left[\frac{P_{\alpha\gamma_\ell}(E_j \times F_j)}{P_{\alpha\gamma_\ell}(F_j)P_\alpha(E_j)}\right] P_{\alpha\gamma_\ell}(E_j \times F_j) \\ &= L\sum_{j=1}^{J}\sum_{\ell=1}^{L} \ln\left[\frac{\frac{1}{L}P_{\alpha\gamma_\ell}(E_j \times F_j)}{\frac{1}{L}P_{\alpha\gamma_\ell}(F_j)P_\alpha(E_j)}\right]\frac{1}{L}P_{\alpha\gamma_\ell}(E_j \times F_j) \\ &\geq L\sum_{j=1}^{J} \ln\left[\frac{P_{\alpha\bar{\gamma}}(E_j \times F_j)}{P_{\alpha\bar{\gamma}}(F_j)P_\alpha(E_j)}\right] P_{\alpha\bar{\gamma}}(E_j \times F_j) \end{aligned} \tag{2-12}$$

use having been made of the well-known inequality (Pinsker, 1964, pp. 21-22)

$$\sum_{j=1}^{J} r_j \log \frac{r_j}{u_j} \geq (r_1 + \cdots + r_J) \log \frac{r_1 + \cdots + r_J}{u_1 + \cdots + u_J}$$

for any set of non-negative r_j and u_j. Pick a partition for which the right hand side of inequality (2-12) is within ϵ of L times $I_{\alpha\bar{\gamma}}(U;\tilde{U})$. Let the left hand side of this inequality be denoted by A and the right hand side by B. Then we have

$$\sum_{\ell=1}^{L} I_{\alpha\gamma_\ell}(U_\ell;\tilde{U}_\ell) \geq A \geq B \geq L I_{\alpha\bar{\gamma}}(U;\tilde{U}) - \epsilon$$

Since this holds for any ϵ, we have

$$\sum_{\ell=1}^{L} I_{\alpha\gamma_\ell}(U_\ell;\tilde{U}_\ell) \geq L I_{\alpha\bar{\gamma}}(U;\tilde{U}) \qquad (2\text{-}13)$$

Recall that this inequality holds for any L-tuple of $\alpha\gamma_\ell$ distributions, it only having been assumed that the U_ℓ were independent. A trivial modification of the above argument can be used to show that the mutual information is convex in γ. Combining inequalities (2-8) and (2-13) yields

$$NC \geq L I_{\alpha\bar{\gamma}}(U;\tilde{U}) \qquad \text{all } \alpha \in \mathcal{A} \qquad (2\text{-}14)$$

and hence

$$NC \geq L \sup_{\alpha \in \mathcal{A}} I_{\alpha\bar{\gamma}}(U;\tilde{U}) \qquad (2\text{-}15)$$

Finally, since $\bar{\gamma} \in \Gamma_{\mathcal{A}}(d)$

$$NC \geq L \sup_{\alpha \in \mathcal{A}} I_{\alpha\bar{\gamma}}(U;\tilde{U}) \geq L \inf_{\gamma \in \Gamma(d)} \sup_{\alpha \in \mathcal{A}} I_{\alpha\gamma}(U;\tilde{U}) = L R_{\mathcal{A}}(d) \qquad (2\text{-}16)$$

3. THE POSITIVE SIDE OF THE CODING THEOREM--FINITE CLASSES OF SOURCES

The positive side of the coding theorem is proven by a random coding argument. The starting point is a proof of the source coding theorem for a single source due to Gallager (1968). Let us define the following distributions on L-tuples of source outputs.

$$P_\alpha(\underset{\sim}{u}) = \prod_{\ell=1}^{L} P_\alpha(u_\ell) \qquad (3\text{-}1)$$

$$P_\gamma(\tilde{\underset{\sim}{u}}|\underset{\sim}{u}) = \prod_{\ell=1}^{L} P_\gamma(\tilde{u}_\ell|u_\ell) \qquad (3\text{-}2)$$

$$P_{\alpha\gamma}(\tilde{u}) = \int_{\mathcal{U}} P_\gamma(\tilde{u}|u)\,dP_\alpha(u) \qquad (3\text{-}3)$$

$$P_{\alpha\gamma}(\tilde{\underset{\sim}{u}}) = \int_{\mathcal{U}\times\cdots\times\mathcal{U}} P_\gamma(\tilde{\underset{\sim}{u}}|\underset{\sim}{u})\,dP_\alpha(\underset{\sim}{u}) = \prod_{\ell=1}^{L} P_{\alpha\gamma}(\tilde{u}_\ell) \qquad (3\text{-}4)$$

$$P_{\alpha\gamma}(\underset{\sim}{u},\tilde{\underset{\sim}{u}}) = P_\gamma(\tilde{\underset{\sim}{u}}|\underset{\sim}{u})P_\alpha(\underset{\sim}{u}) \qquad (3\text{-}5)$$

An (L,M) source code is a set of M L-tuples of source outputs, $\underset{\sim}{u}_1, \underset{\sim}{u}_2, \ldots, \underset{\sim}{u}_M$. An ensemble of source codes is generated by selecting sets of M L-tuples, each L-tuple in a set being drawn independently according to the distribution $P_{\alpha\gamma}(\tilde{\underset{\sim}{u}})$. The probability measure on this ensemble we denote by P_c. A source is encoded by selecting a set of M code words (L-tuples) from the ensemble and mapping a generated source L-tuple $\underset{\sim}{u}$ into the code word $\underset{\sim}{u}_m$ which minimizes

$$d(\underset{\sim}{u},\tilde{\underset{\sim}{u}}_m) \qquad m = 1, 2, \ldots, M$$

with arbitrary selection of a code word if the minimum is not unique.

At this point we relate the probability measure P_c governing source vectors and code words to the probability measure $P_{\alpha\gamma}$ generating the ensemble of codes.

Lemma 1 (Gallager 9.3.1): Let R^* and d^* be arbitrary positive numbers and $\alpha\gamma$ an arbitrary distribution. Define the set A by

$$A = \{\underset{\sim}{u},\tilde{\underset{\sim}{u}} : I_{\alpha\gamma}(\underset{\sim}{u},\underset{\sim}{u}) > LR^* \text{ or } d(\underset{\sim}{u},\tilde{\underset{\sim}{u}}) > d^*\} \tag{3-6}$$

Then

$$P_c(\bar{D} > d^*) \le P_{\alpha\gamma}(A) + \exp(-Me^{-LR^*}) \tag{3-7}$$

the probability $P_c(\bar{D} > d^*)$ being the probability over the ensemble of codes and ensemble of source outputs that a source output $\underset{\sim}{U}$ lies at distance greater than d^* from one of the M code words in the set.

For completeness we give a proof of this lemma here, our proof being a generalization of Gallager's proof from a discrete to an arbitrary measure space.

Proof:

For an arbitrary u define $A_{\underset{\sim}{u}}$ as the set of $\tilde{\underset{\sim}{u}}$ for which $u,\tilde{\underset{\sim}{u}} \in A$ (note that $A_{\underset{\sim}{u}}$ is measurable)

$$A_{\underset{\sim}{u}} = \{\tilde{\underset{\sim}{u}} : I_{\alpha\gamma}(\underset{\sim}{u},\tilde{\underset{\sim}{u}}) > LR^* \text{ or } d(\underset{\sim}{u},\tilde{\underset{\sim}{u}}) > d^*\} \tag{3-8}$$

In general the conditional probability $P_c(\bar{D} > d^*|\underset{\sim}{u})$ need not exist as an integrable function of $\underset{\sim}{u}$; in the case at hand we can show that it is bounded by an integrable function of $\underset{\sim}{u}$

$$P_c(\bar{D} > d^*|\underset{\sim}{u}) = P_c\{d(\underset{\sim}{u},\underset{\sim}{u}_m) > d^* \quad m = 1, 2, \ldots, M\}$$

$$= \prod_{m=1}^{M} P(d(\underset{\sim}{u},\underset{\sim}{u}_m) > d^*)$$

$$\le \left[1 - \int_{A_{\underset{\sim}{u}}^c} dP_{\alpha\gamma}(\tilde{\underset{\sim}{u}})\right]^M \tag{3-9}$$

Now for $\underset{\sim}{u}$ and $\tilde{\underset{\sim}{u}} \in A_{\underset{\sim}{u}}^c$

$$\Lambda_{\alpha\gamma}(\underset{\sim}{u},\tilde{\underset{\sim}{u}}) \le e^{LR^*} \tag{3-10}$$

so that inequality (3-9) can be weakened to

$$P_c(\bar{D} > d^*|\underset{\sim}{u}) \le \left[1 - e^{-LR^*}\int_{A_{\underset{\sim}{u}}^c} \Lambda_{\alpha\gamma}(\underset{\sim}{u},\tilde{\underset{\sim}{u}})dP_{\alpha\gamma}(\tilde{\underset{\sim}{u}})\right]^M \tag{3-11}$$

Now

$$(1 - \beta x)^M \le 1 - x + e^{-M\beta} \tag{3-12}$$

(see Gallager's Eqs. 9.3.22-23) identifying β with e^{-LR^*} and x with the integral, inequality (3-11) can thus be weakened to

$$P_c(\bar{D} > d^*|u) \le \int_{A_u} \Lambda_{\alpha\gamma}(\underset{\sim}{u},\tilde{\underset{\sim}{u}})dP_{\alpha\gamma}(\tilde{\underset{\sim}{u}}) + \exp(-Me^{-LR^*}) \tag{3-13}$$

Integrating both sides of this inequality with respect to $P_\alpha(\underset{\sim}{u})$ over $\mathcal{U}$ then yields

$$P_c(\bar{D} > d^*) \le \int_A dP_{\alpha\gamma}(\underset{\sim}{u},\tilde{\underset{\sim}{u}}) + \exp(-Me^{-LR^*})$$

$$= P_{\alpha\gamma}(A) + \exp(-Me^{-LR^*})$$

At this point we give a proof of the positive side of the coding theorem for a single source due to Gallager (1968). Although this is not our objective, we need to make use of these ideas at a later point.

Theorem 1 (Gallager 9.6.2). Let $R_\alpha(d)$ be the rate distortion function of a source α. Then for any $d > 0$ and any $\delta > 0$ there exists a (sufficiently large) L and an (M + [illegible]) code with

$$M \le \exp L[R_\alpha(d) + \delta] \tag{3-14}$$

and

$$E_\alpha(\bar{D}) \le d + \delta \tag{3-15}$$

Proof (after Gallager):

Let γ_0 be a distribution in $\Gamma_\alpha(d)$ such that

$$I_{\alpha\gamma_0}(U;\tilde{U}) \le R_\alpha(d) + \frac{\delta}{4} \tag{3-16}$$

We consider an ensemble of (L,M) codes generated by $P_{\alpha\gamma_0}(\tilde{u})$ and apply Lemma 1, selecting $d^* = d + \frac{\delta}{2}$

$$R^* = R_\alpha(d) + \frac{\delta}{2}$$

and

$$M = \exp\{L[R(d) + \tfrac{3}{4}\delta]\}$$

yielding

$$P_c\left\{\bar{D} > d + \frac{\delta}{2}\right\} \le P_{\alpha\gamma_0}(A) + \exp(-e^{L\delta/4}) \tag{3-17}$$

in which

$$P_{\alpha\gamma_0}(A) = P_{\alpha\gamma_0}\left\{I^L_{\alpha\gamma_0} > LR_\alpha(d) + \frac{\delta}{2} \quad \text{or} \quad \bar{D} > \left(d + \frac{\delta}{2}\right)\right\}$$

$$\le P_{\alpha\gamma_0}\left\{\frac{1}{L} I^L_{\alpha\gamma_0} > I_{\alpha\gamma_0}(U;\tilde{U}) + \frac{\delta}{4}\right\}$$

$$+ P_{\alpha\gamma_0}\left\{\bar{D} > \left(d + \frac{\delta}{2}\right)\right\} \tag{3-18}$$

Now by the independence of the $(U_\ell,\tilde{U}_\ell)$ $\ell = 1, \ldots, L$ under the distribution defined by Eqs. (3-1)-(3-5)

$$\frac{1}{L} I^L_{\alpha\gamma_0} = \frac{1}{L}\sum_{\ell=1}^{L} I_{\alpha\gamma_0}(u_\ell,\tilde{u}_\ell) \tag{3-19}$$

and

$$\bar{D} = \frac{1}{L}\sum_{\ell=1}^{L} D_\ell = \frac{1}{L}\sum_{\ell=1}^{L} d(U_\ell,\tilde{U}_\ell) \tag{3-20}$$

and the right hand sides of both equations are sums of independent identically distributed random variables. Further (Pinsker, [illegible], p. [illegible])

$$E_{\alpha\gamma_0}(|I_{\alpha\gamma_0}|) \le \frac{2}{e} + I_{\alpha\gamma_0}(U;\tilde{U}) < \infty$$

and

$$E_{\alpha\gamma_0}(|D_\ell|) = E_{\alpha\gamma_0}(D_\ell) \le d < \infty$$

Thus by the weak law of large numbers (Feller, 1966) $P_{\alpha\gamma_0}(A)$ can be made as small as desired by making L sufficiently large. Thus, from inequalities (3-17) and (3-18) we have

$$P_c\left\{\bar{D} > d + \frac{\delta}{2}\right\} \le \beta(L) \tag{3-21}$$

in which $\beta(L)$ can be made arbitrarily small by making L sufficiently large. For convenience let B denote the event $\bar{D} > d + \frac{\delta}{2}$. Then

$$E_{\alpha\gamma_0}(\bar{D}) \le \left(d + \frac{\delta}{2}\right) + P_{\alpha\gamma_0}(B)E_{\alpha\gamma_0}(\bar{D}|B) \qquad (3\text{-}22)$$

If when the event B occurs we map the source output into the zero vector

$$P_{\alpha\gamma_0}(B)E_{\alpha\gamma_0}(\bar{D}|B) = \frac{1}{L}\sum_{\ell=1}^{L} P_{\alpha\gamma_0}(B)E_{\alpha\gamma_0}(\rho(U_\ell,0)|B) \qquad (3\text{-}23)$$

Now consider any positive valued random variable V and any set B and let d_0 be picked such that

$$P(V > d_0) > P(B)$$

then

$$P(B)E(V|B) = \int_B v\,dP(v) \le \int_{v \ge d_0} v\,dP(v) \qquad (3\text{-}24)$$

Combining inequalities (3-23) and (3-24) with $V = \rho(U,0)$

$$P_{\alpha\gamma_0}(B)E_{\alpha\gamma_0}(\bar{D}|B) \le \int_{v \ge d_0} v\,dP(v) \qquad (3\text{-}25)$$

But $E(V) \le K_0 < \infty$, thus

$$\lim_{d_0 \to \infty} \int_{v \ge d_0} v\,dP(v) = 0 \qquad (3\text{-}26)$$

The right hand side of inequality (3-25) may thus be made less than $\frac{\delta}{2}$ by making d_0 sufficiently large; this is accomplished by making L sufficiently large and $\beta(L)$ sufficiently small. This remark coupled with inequality (3-22) completes the proof of the Theorem.

We now proceed to the source coding theorem for a finite class of sources, $\mathcal{A} = \{\alpha_k\}$, $k = 1, \ldots, K$.

Theorem 2: Consider a finite class of sources, $\mathcal{A} = \{\alpha_k\}$, $k = 1, \ldots, K$. Let

$$R_{max}(d) = \max_{k=1,\ldots,K} R_{\alpha_k}(d) \qquad (3\text{-}27)$$

Then for any $d > 0$ and $\delta' > 0$ there exists a sufficiently large L and an $(L, M + 1)$ code with

$$M \le \exp L[R_{max}(d) + \delta'] \qquad (3\text{-}28)$$

and

$$E_{\alpha_k}(\bar{D}) \le d + \delta' \qquad \text{all } \alpha_k \in \mathcal{A} \qquad (3\text{-}29)$$

Proof:

Consider α_k and apply Theorem 1 with $\delta = \delta'/2$. Then for α_k there is an $(L_k, M_k + 1)$ code with

$$M_k \le \exp L_k[R_{\alpha_k}(d) + \delta'/2] \qquad (3\text{-}30)$$

and

$$E_{\alpha_k}(\bar{D}) \le d + \delta'/2 \qquad (3\text{-}31)$$

Such a code may be constructed for each $\alpha_k \in \mathcal{A}$. Let

$$L = \max\{L_1, L_2, \ldots, L_K, (2 \log K)/\delta'\} \qquad (3\text{-}32)$$

and construct a code of block length L for each α_k in $\mathcal{A}$. Each such code consists of M_k L-tuples plus the L-tuple of zero "vectors." Let the <u>union code</u> for $\mathcal{A}$ be the set of all

$$M = \sum_{k=1}^{K} M_k \tag{3-33}$$

L-tuples plus the zero L-tuple. This union code yields distortion determined by Eq. (3-31) for any $\alpha_k \in \mathcal{A}$. Further this union code has a total number of L-tuples bounded by

$$M + 1 = 1 + \sum_{k=1}^{K} M_k$$

$$\le 1 + \sum_{k=1}^{K} \exp\{L[R_{\alpha_k}(d) + \delta'/2]\}$$

$$\le 1 + K \exp\{L[R_{max}(d) + \delta'/2]\}$$

$$= 1 + \exp\{L[R_{max}(d) + \delta'/2 + (\log K/L)]\}$$

$$\le 1 + \exp\{L[R_{max}(d) + \delta']\} \tag{3-34}$$

This completes the proof of the Theorem.

4. THE POSITIVE SIDE OF THE CODING THEOREM--COMPACT CLASSES OF SOURCES

We now extend our proof of the positive side of the coding theorem to compact classes of sources as defined in section 1. The import of our definition of compact is, that given a compact class $\mathcal{A}$, we can find a finite class $\mathcal{A}''$ of sources with a rate distortion function arbitrarily close to $R_{\mathcal{A}}(d)$ and such that a code for $\mathcal{A}''$ works for any $\alpha \in \mathcal{A}$.

Let us proceed to construct such a class $\mathcal{A}''$ for an arbitrary compact class $\mathcal{A}$. For arbitrary values of $\varepsilon_1, \varepsilon_2 > 0$, we pick a $\mathcal{U}'$ and a ...... $\mathcal{U}'$ with maximum radius ε_2. We denote the sets in this

covering by $\mathcal{U}_j$ and let u_j denote a fixed interior point in $\mathcal{U}_j$, $j = 1, 2, \ldots, J = J(\varepsilon_2) < \infty$. For any given $\alpha \in \mathcal{A}$, we then construct a probability measure $\alpha'\gamma'$ on $\{u_j\} \times \{u_j\}$ by

$$P_{\alpha'\gamma'}(j,k) \triangleq P_{\alpha'\gamma'}(u_j,\tilde{u}_k) = P_{\alpha\gamma}\{\mathcal{U}_j \times \mathcal{U}_k\} \tag{4-1}$$

and denote the class of such source measures by $\mathcal{A}'$. The random variables defined by the mappings $\mathcal{U}_j \to u_j$, $\mathcal{U}_k \to u_k$, we denote respectively by U' ar. $\tilde{U}'$.

Let γ' denote the transition probability corresponding to γ defined by Eq. (4-1). The triangle inequality and Eq. (1-5) then imply that if $\gamma \in \Gamma_{\mathcal{A}}(d)$, then

$$\gamma' \in \Gamma_{\mathcal{A}'}(d + \varepsilon_1 + 2\varepsilon_2) \tag{4-2}$$

for the mean distortion measure and if $\gamma \in \Gamma_{\mathcal{A}}(d,\rho_0)$, then

$$\gamma' \in \Gamma_{\mathcal{A}'}(d + \varepsilon_1, \rho_0 + 2\varepsilon_2) \tag{4-3}$$

for the indicator distortion measure. This follows simply by regarding the mapping from U' to $\tilde{U}'$ as the composition of a mapping from U' to U which induces P_α on U, the γ mapping from U to $\tilde{U}$, and the mapping from $\tilde{U}$ to $\tilde{U}'$. Further, given any $\alpha' \in \mathcal{A}'$, there is at least one $\alpha \in \mathcal{A}$ such that $\alpha'\gamma'$ is a partition of $\alpha\gamma$, and hence (Pinsker, 1964)

$$I_{\alpha'\gamma'}(U';\tilde{U}') \le I_{\alpha\gamma}(U;\tilde{U}) \tag{4-4}$$

We next generate a finite class $\mathcal{A}''$ of sources as follows. We pick an integer Q. Then for a given $\alpha' \in \mathcal{A}'$ let the largest of the

$$P_{\alpha'}(j) = P_\alpha(\mathcal{U}_j)$$

be denoted by $p_{\alpha'}(j_0)$. For any $j \neq j_0$ select $Qp_{\alpha''}(j)$ to be the integer such that

$$Qp_{\alpha'}(j) < Qp_{\alpha''}(j) \le (Q+1)p_{\alpha'}(j) \qquad j = 1, 2, \ldots, J; \quad j \neq j_0 \tag{4-5}$$

and

$$p_{\alpha''}(j_0) = 1 - \sum_{\substack{j=1 \\ j\neq j_0}}^{J} p_{\alpha''}(j)$$

Then

$$0 \le p_{\alpha''}(j) - p_{\alpha'}(j) \le \left(\frac{1}{Q}\right) \qquad j \neq j_0$$

$$0 \le p_{\alpha}(j_0) - p_{\alpha''}(j_0) \le \frac{J}{Q} \tag{4-6}$$

Further, if we require

$$Q > 2J^2 \tag{4-7}$$

Then from Lemma 4 of Blackwell, Breiman, and Thomasian (1959), we have

$$p_{\alpha'}(j) \le p_{\alpha''}(j)e^{2J^2/Q} \qquad \text{all } j \tag{4-8}$$

Let us now consider the joint measure

$$p_{\alpha''\gamma'}(j,k) = p_{\alpha''}(j)p_{\gamma'}(k|j)$$

Using inequalities (4-6), we have

$$|p_{\alpha''\gamma'}(j,k) - p_{\alpha'\gamma'}(j,k)| \le p_{\gamma'}(k|j)|p_{\alpha''}(j) - p_{\alpha'}(j)|$$

$$\le \begin{cases} J/Q & j = j_0 \\ \ldots & \end{cases} \tag{4-9}$$

Then

$$|p_{\alpha''\gamma'}(k) - p_{\alpha'\gamma'}(k)| \le \sum_{j=1}^{J} |p_{\alpha''\gamma'}(j,k) - p_{\alpha'\gamma'}(j,k)|$$

$$\le J/Q + (J-1)/Q \le 2J/Q \qquad \text{all } k \tag{4-10}$$

From Lemma 1 of Blackwell, Breiman, and Thomasian (1959), we have that for

$$|p - p'| \le \varepsilon < e^{-1} \tag{4-11}$$

then

$$|p \ln p - p' \ln p'| \le \varepsilon^{1/2} \tag{4-12}$$

Writing

$$I_{\alpha''\gamma'}(U';\tilde{U}') = \sum_{j,k} p_{\alpha''\gamma'}(j,k) \ln p_{\alpha''\gamma'}(j,k)$$

$$- \sum_{j} p_{\alpha''}(j) \ln p_{\alpha''}(j)$$

$$- \sum_{k} p_{\alpha''\gamma'}(k) \ln p_{\alpha''\gamma'}(k)$$

and using inequalities (4-6), and (4-9-12), we have

$$|I_{\alpha''\gamma'}(U';\tilde{U}') - I_{\alpha'\gamma'}(U';\tilde{U}')| \le 6J^2/Q^{1/2} \tag{4-13}$$

We also need a bound relating $E_{\alpha''\gamma'}(D)$ to $E_{\alpha'\gamma'}(D)$. Unfortunately this bound cannot be as tight. Let

$$d_{max} = \max_{j,k=1,2,\ldots,J} d(u_j,u_k) \quad (4\text{-}14)$$

(note that $d_{max} \leq J\varepsilon_2$ or 1). Then, from inequality (4-6)

$$E_{\alpha''\gamma'}(D) - E_{\alpha'\gamma'}(D) = \sum_{j=1}^{J} [P_{\alpha''}(j) - P_{\alpha'}(j)] E(d(u_j,\tilde{u}_k)|\tilde{u}_k)$$

$$\leq d_{max}(J-1)/Q \leq d_{max} J/Q \quad (4\text{-}15)$$

Let us now summarize the behavior of the class of sources $\mathcal{A}''$. Combining inequalities (4-2-5), (4-13), and (4-15), we see that for any $\alpha'' \in \mathcal{A}''$ and γ' there is a corresponding $\alpha \in \mathcal{A}$ and γ such that

$$I_{\alpha''\gamma'}(U',\tilde{U}') \leq I_{\alpha\gamma}(U;\tilde{U}) + 6J^2/Q^{1/2} \quad (4\text{-}16)$$

and that $\gamma \in \Gamma_{\mathcal{A}}(d)$ implies for the mean distortion measure

$$\gamma' \in \Gamma_{\mathcal{A}''}(d + \varepsilon_1 + 2\varepsilon_2 + d_{max} J/Q) \quad (4\text{-}17)$$

while for the indicator distortion measure $\gamma \in \Gamma_{\mathcal{A}}(d,\rho_0)$ implies

$$\gamma' \in \Gamma_{\mathcal{A}''}(d + \varepsilon_1 + d_{max} J/Q,\ \rho_0 + 2\varepsilon_2) \quad (4\text{-}18)$$

Further the number of sources in $\mathcal{A}''$ is bounded by

$$K \leq (Q+1)^J \quad (4\text{-}19)$$

We thus have a finite class of sources whose rate distortion performance can be made arbitrarily close to that of the original class $\mathcal{A}$. We must now examine how a code designed to handle $\mathcal{A}''$ would work for $\mathcal{A}$. We first consider the performance of such a code for $\mathcal{A}'$, the extension to $\mathcal{A}$ being easily handled later.

Consider any source $\alpha' \in \mathcal{A}'$. For any such source we can find an $\alpha'' \in \mathcal{A}''$ such that inequality (4-8) is satisfied. Thus on any point $\underset{\sim}{u}_j$ in the L-fold product space $\{u_{j_1}\} \times \ldots \times \{u_{j_L}\}$

$$P_{c\alpha'}(\underset{\sim}{u}_j) \leq P_{c\alpha''}(\underset{\sim}{u}_j)[\exp(2J^2/Q)]^L \quad (4\text{-}20)$$

Let $P_{c\alpha''}$ denote the probability measure over an ensemble of codes generated by the probability measure $\alpha''\gamma$. Then Eq. (4-20) implies that when this same code is used on the source α'

$$P_{c\alpha'}(\bar{D} > d^*) \leq \exp(2LJ^2/Q) P_{c\alpha''}(\bar{D} > d^*) \quad (4\text{-}21)$$

Thus to establish that the union of the K codes generated by the $\alpha'' \in \mathcal{A}''$ works for the class $\mathcal{A}'$ we require an exponential bound for $P_{c\alpha''}(\bar{D} > d^*)$. Such a bound was not developed earlier because the general structure of the probability space did not allow it; here we are working with a finite probability space.

Lemma 2: Let $P_{c\alpha''}$ denote the probability distribution over the ensemble of (L,M) codes generated by $\alpha''\gamma'$, with

$$M = \exp[L(\bar{I} + \delta_1)] \quad (4\text{-}22)$$

$$\bar{I} = I_{\alpha''\gamma'}(U';\tilde{U}') = E(I_{\alpha''\gamma'}) \quad (4\text{-}23)$$

and

$$\bar{d} = E_{\alpha''\gamma'}(D) \quad (4\text{-}24)$$

Then for $\delta_2/d_{max} < 1/4$, $\delta_3/J^2 < 1/2$

$$P_{\alpha''}(\bar{D} > \bar{d} + \delta_2) \le \exp[-L(\delta_1 - \delta_3)] + \exp[-L\delta_3^2/3J^2] + \exp[-L\delta_2^2/3d_{max}^2] \tag{4-25}$$

Proof:

Applying the union bound to Lemma 1 and setting

$$R^* = I + \delta_3, \quad d^* = \bar{d} + \delta_2, \tag{4-26}$$

yields

$$P_{\alpha''}(\bar{D} > \bar{d} + \delta_2) \le \exp[L(\delta_1 - \delta_3)] + P_{\alpha''\gamma'}(\bar{D} < \bar{d} + \delta_2) + P_{\alpha''\gamma'}(I^L_{\alpha''\gamma'} > L(\bar{I} + \delta_3)) \tag{4-27}$$

Let us first consider the last term in inequality (4-27). For convenience we drop the $\alpha_k''\gamma'$ subscript. Using the Chernof bound, for any $t > 0$ we have

$$P(I^L \ge L\zeta) \le E(\exp[t(I^L - L\zeta)]) = e^{-tL\zeta}E(e^{tI^L}) = e^{-tL\zeta}E^L(e^{tI}) \tag{4-28}$$

Making the necessary adjustments for change of sign, we have for $0 < t < 1$ from the proof of Theorem 3 of Blackwell, Breiman, and Thomasian (1959) that

$$E(e^{tI}) \le 1 + t\bar{I} + \frac{t}{2}\frac{J^2}{(1 - t^2)} \le \exp\left\{t\bar{I} + \frac{t^2J^2}{2(1 - t^2)}\right\} \tag{4-29}$$

Letting $\zeta = \bar{I} + \delta_3$, setting $t = \delta_3/J^2 < 1/2$, and combining inequalities (4-28) and (4-29) yields

$$P(I^L \ge L(\bar{I} + \delta_3)) \le \exp\left\{-L\left[t\delta_3 - \frac{t^2J^2}{2(1 - t^2)}\right]\right\} \le \exp(-L\delta_3^2/3J^2) \tag{4-30}$$

Next we consider the second term in inequality (4-27) and again omit the $\alpha''\gamma'$ subscript for convenience. Again using the Chernof bound, for any $t > 0$

$$P(\bar{D} \ge \zeta) \le E\exp[t(L\bar{D} - L\zeta)] = e^{-tL\zeta}E(e^{tL\bar{D}}) = e^{-tL\zeta}E^L(e^{tD}) \tag{4-31}$$

But, for some θ, $0 < \theta < 1$

$$E(e^{tD}) = E(1 + tD + \frac{t^2}{2}\theta^2e^{tD}) \le 1 + t\bar{d} + \frac{t^2}{2}d_{max}^2 e^{td_{max}} \le \exp(t\bar{d} + \frac{t^2}{2}d_{max}^2 e^{td_{max}}) \tag{4-32}$$

Letting $\zeta = \bar{d} + \delta_2$, $t = \delta_2/d_{max}^2 < 1/4d_{max}$, and combining inequalities (4-31) and (4-32) yields

$$P(\bar{D} \ge \bar{d} + \delta_2) \le \exp(-L[t\delta_2 - \frac{t^2}{2}d_{max}^2 e^{td_{max}}]) \le \exp(-L\delta_2/3d_{max}^2) \tag{4-33}$$

Combining inequalities (4-27), (4-30), and (4-33) then completes the proof of the Lemma.

At this point we are in a position to prove the main Theorem.

Theorem 3: Let $\mathcal{Q}$ denote a compact class of sources and $R_\alpha(d)$ (or $R_\alpha(d,\rho_0)$ for the indicator distortion measure) the rate distortion function for a source $\alpha \in \mathcal{A}$. Let $R_{sup}(d) = \sup_{\alpha \in \mathcal{A}} R_\alpha(d)$. Then for any $d > 0$, $\delta > 0$, there exists a sufficiently large L and an (L,M) code with

$$M \le \exp[L(R_{sup} + \delta)]$$

and

$$E_\alpha(\bar{D}) \le d + \delta \quad \text{all } \alpha \in \mathcal{A}$$

for the mean distortion measure and

$$E_\alpha(\bar{D}_{\rho_0 + 3\delta/10}) \le d + 3\delta/5 \quad \text{all } \alpha \in \mathcal{A}$$

for the indicator distortion measure.

Proof:

For a given δ, select

$$\epsilon_1 = \epsilon_2 = \delta/10 \tag{4-34}$$

This determines J, the number of ϵ_2 neighborhoods required to cover $\mathcal{U}'$. With J determined, set Q to be the smallest integer such that

$$\delta/10 > \max(6J^2/Q^{1/2},\ Jd_{max}/Q,\ 2J^2/Q,\ 24J^4/5\delta Q,\ 30J^2d_{max}^2/\delta Q) \tag{4-35}$$

Let us assume for reference that we have enumerated the K sources in $\mathcal{A}''$. Consider an arbitrary source in $\mathcal{A}''$, say the kth, α_k''. This source was generated in mapping $\mathcal{U}$ into the discrete space $\{u_j\}$ by some source or sources in $\mathcal{A}$. Pick any such source α that gets mapped into α_k'' and refer to it as α_k. Then for α_k there exists a γ in $\Gamma_{\alpha_k}(d)$ or $\Gamma_{-}(d,\rho_0)$ such that

$$I_{\alpha_k\gamma}(U;\tilde{U}) \le R_{\alpha_k} + \delta/10 \tag{4-36}$$

Thus, combining inequalities (4-16) and (4-36)

$$\bar{I} = I_{\alpha_k''\gamma'}(U';\tilde{U}') \le R_{\alpha_k} + \delta/5 \tag{4-37}$$

Similarly, from inequalities (4-17), (4-18), (4-34) and (4-35), we have that for the mean distortion measure

$$\gamma' \in \Gamma_{\alpha_k''}(d + 2\delta/5) \tag{4-38}$$

and for the indicator distortion measure

$$\gamma' \in \Gamma_{\alpha_k''}(d + \delta/5,\ \rho_0 + \delta/5) \tag{4-39}$$

If we now apply inequality (4-21) and Lemma 2 with

$$\delta_1 = 7\delta/10,\ \delta_2 = \delta/5,\ \delta_3 = \delta/2$$

we have for any source α' that gets mapped into α_k''

$$P_{c\alpha'}(\bar{D} > d + 3\delta/5) \le \exp(L2J^2/Q)$$

$$\{\exp(-L\delta/5) + \exp(-L\delta^2/12J^2)$$

$$+ \exp(-L\delta^2/75d_{max}^2)\}$$

$$\triangleq \beta(L) \tag{4-40}$$

for the mean distortion measure and

$$P_{c\alpha'}(\bar{D}_{\rho_0 + \delta/5} > d + 2\delta/5) \le \beta(L) \tag{4-41}$$

for the indicator distortion measure. Writing out the expression for $\beta(L)$ and using inequality (4-35), we have

$$\beta(L) = \exp[-L(\delta/5 - 2J^2/Q)] + \exp[-L(5\delta/12J^2)(\delta/5 - 24J^4/5\delta Q)]$$

$$+ \exp[-L(\delta/15d_{max}^2)(\delta/5 - 30J^2 d_{max}^2/\delta Q)]$$

$$\leq \exp(-L\delta/10) + \exp(-L\delta^2/24J^2)$$

$$+ \exp(-L\delta^2/150d_{max}^2) \qquad (4\text{-}42)$$

Let us last pick L sufficiently large that

$$\beta(L)\ d_{max} \leq \delta/10 \qquad (4\text{-}43)$$

and

$$(1/L) \log K \leq (J/L) \log (Q + 1) \leq \delta/10 \qquad (4\text{-}44)$$

Then for any source α' mapped into α_k'', the ensemble of codes generated by $\alpha_k''\gamma'$ yields

$$E_{c\alpha'}(\bar{D}) \leq d + 3\delta/d + \beta(L)d_{max} \leq d + 7\delta/10 \qquad (4\text{-}45)$$

for the mean distortion measure and

$$E_{c\alpha'}(\bar{I}_{\rho_0+\delta/5}) \leq d + 2\delta/5 + \beta(L)d_{max} \leq d + \delta/2 \qquad (4\text{-}46)$$

for the indicator distortion measure. Next consider all the sources $\alpha \in \mathcal{A}$ that get mapped into α_k''; when such a source has its output u mapped into $\{u_j, j = 1, \ldots, J\}$, yielding a source α', the additional error is no greater than

$$\epsilon_1 + \epsilon_2 = \delta/5$$

for the mean distortion measure, while for the indicator distortion measure no more than $\epsilon_1 = \delta/10$ is added to the average distortion level and $\epsilon_2 = \delta/10$ to the threshold value ρ_0. Thus for any source in $\mathcal{A}$ that is mapped into α_k'', the ensemble of codes generated by $\alpha''\gamma'$ yields

$$E_{c\alpha}(\bar{D}) \leq d + 9\delta/10 \qquad (4\text{-}47)$$

for the mean distortion measure and

$$E_{c\alpha}(\bar{I}_{\rho_0+3\delta/10}) \leq d + 3\delta/5 \qquad (4\text{-}48)$$

for the indicator distortion measure.

To complete the proof, we note that all the above statements hold for any $\alpha'' \in \mathcal{A}''$. Further, the code for α_k'' has

$$M = \exp\{L(\bar{I} + \delta_1)\} \leq \exp\{L(R_{\alpha_k} + 9\delta/10)\}$$

points. Then from inequality (4-44), the union of the codes for all K $\alpha'' \in \mathcal{A}''$ have less than

$$K \max_{k=1,\ldots,K} \exp\{L(R_{\alpha_k} + 9\delta/10)\} \leq \sup_{\alpha \in \mathcal{A}} \exp\{L(R_\alpha + \delta)\}$$

points. Further the union code guarantees performance given by inequality (4-47) or (4-48) for all $\alpha \in \mathcal{A}$.

5. EXAMPLES AND DISCUSSION

We now consider a number of examples and then briefly discuss the structure of our theory.

In all the examples discussed, we consider a mean-square distortion measure, or in the case of random processes or vector-valued random variables, weighted mean-square distortion measures.

First, let us consider examples of compact classes of sources. Let U be a real-valued random variable and let $\mathcal{A}$ be the class of all distributions such that for some fixed $\epsilon > 0$

$$E_\alpha(|U|^{2(1+\epsilon)}) \le K_1 < \infty \tag{5-1}$$

all $\alpha \in \mathcal{A}$

Then, using the Hölder inequality and a Chebyshef-style inequality

$$\int_{|u|>u_0} u^2 dF_\alpha(u) \le E_\alpha^{\frac{1}{1+\epsilon}}(|U|^{2(1+\epsilon)}) P^{\frac{\epsilon}{1+\epsilon}}(|U| > u_0)$$

$$\le (K_1)^{\frac{1}{1+\epsilon}} \left[\frac{K_1}{|u_0|^{2(1+\epsilon)}}\right]^{\frac{\epsilon}{1+\epsilon}}$$

$$\le K_1 \, |u_0|^{-2\epsilon} \tag{5-2}$$

all $\alpha \in \mathcal{A}$

The right-hand side of this inequality can be made smaller than any ϵ_1 by making u_0 sufficiently large. The resulting interval $[-u_0, u_0]$ is totally bounded (can be covered by a finite number of intervals of length $\epsilon_2 > 0$) and hence this class of sources is compact.

Similarly, consider a random process U(t) on the time interval [0,T]. Consider the N-dimensional subspace of $L_2[0,T]$ spanned by an orthonormal set $\varphi_1, \ldots, \varphi_N$ and let $U_p(t)$ denote the projection of U(t) on this subspace. Let us then express U(t) as the sum of $U_p(t)$ plus a remainder

$$U(t) = U_p(t) + U_r(t) \tag{5-3}$$

Consider the class of processes U(t) such that given an $\epsilon_1 > 0$ we can find an N sufficiently large that

$$E_\alpha\left\{\int_0^T |U_r(t)|^2 dt\right\} < \epsilon_1/2 \tag{5-4}$$

all $\alpha \in \mathcal{A}$

From the work of Root and Pitcher (1955) it follows that this condition will hold in the case in which the processes U(t) are stationary with square integrable correlation functions and the power spectral densities of all the processes in the class are suitably concentrated in some common frequency interval of finite extent.

In addition to the condition of inequality (5-4), we also require that the Fourier coefficients of U(t) corresponding to the $\varphi_1, \ldots, \varphi_N$ satisfy inequality (5-1). This will be satisfied, if for example

$$E_\alpha(|U_t|^4) \le K_1 < \infty \tag{5-5}$$

all $\alpha \in \mathcal{A}$ $\quad 0 \le t \le T$

This class of sources is then compact by extension of our argument for a real-valued random variable to Euclidean N-space.

Let us now evaluate the rate distortion function for several classes of sources. First consider a gaussian random variable source with mean m and variance σ^2. Let the class $\mathcal{A}$ be all such sources with $|m| \le m_0 < \infty$ and $\sigma^2 \le \sigma_0^2 < \infty$. This class is compact since Eq. (5-1) is satisfied and $R_{\mathcal{A}}(d)$ is the sup of the rate distortion functions of the distributions in this class, namely the rate distortion function for a gaussian variable of arbitrary mean and variance σ_0^2

$$R_{\mathcal{A}}(d) = (1/2) \ln [\sigma_0^2/d] \tag{5-6}$$

An example of more practical importance is to assume that the first and second moments of the random variable are known, $E(U) = m$, $\mathrm{var}\,(U) = \sigma^2$, since they can be estimated from measurements, but that nothing else is known about the distribution except that it satisfies inequality (5-1). Again this class of sources is compact and it is well known (Shannon, 1948; Sakrison, 1968, p. 481) that the largest rate distortion function in this class is that of a gaussian random variable of variance σ^2, so

$$R_{\mathcal{a}}(d) = (1/2) \ln [\sigma^2/d] \qquad (5\text{-}7)$$

Note that although the rate-distortion function for $\mathcal{a}$ is that of a gaussian variable, the code for $\mathcal{a}$ will be different than the code for a gaussian variable. The code for a gaussian distribution is a uniform distribution on the surface of an L-dimensional sphere of radius $(\sigma^2 - d)^{1/2}$ (Sakrison, 1968, p. 481); the code for $\mathcal{a}$ contains additional points at the angles determined by

$$u_1 = u_2 = \cdots = u_L$$

and

$$u_k^2 = \sigma^2 - d, \quad u_j = 0, \quad j \neq k$$

This is to account for singular probability distributions on the surface of the sphere caused respectively by the uniform distribution and distributions whose distribution function drops off slowly for large u.

The above example can be extended to the case in which u(t) is a sample function of a random process on the interval [0,T] and it is known that U(t) is zero-mean with known covariance function $R_u(t,s)$. Let $\mathcal{a}$ be the class of all such sources satisfying inequality (5-5). This class

. . . . denote the Karhunen-Loève expansion set for

the covariance function $R_u(t,s)$

$$U(t) = \underset{N\to\infty}{\mathrm{l.i.m.}} \sum_{k=1}^{N} U_k \phi_k(t) \quad 0 \le t \le T \qquad (5\text{-}6)$$

The U_k are then all mutually uncorrelated. The maximum $R_a(d)$ occurs when all of the U_k are statistically independent and each is gaussian; that is, when the U_k are all jointly gaussian and the process U(t) is gaussian. For a mean-integral-square error criterion the rate distortion function of a gaussian process with covariance function $R_u(t,s)$ can be determined parametrically in terms of the eigenvalues of $R_u(t,s)$ (Sakrison, 1968, p. 506) and this is the rate distortion function for the class $\mathcal{a}$. If the distortion measure is a weighted mean-square distortion measure, the same remarks apply, except the eigenvalues are the eigenvalues of the covariance operator composed with the weighting operator (Sakrison, 1968, p. 506).

Next, let us comment briefly on the complexity required in encoding for a class of sources. Consider the class of random variables of known mean and variance satisfying inequality (5-1). It is known (Sakrison, 1968, p. 481) that for a gaussian source a rate can be achieved which is only fractionally greater than the rate $R_a(d)$ by simply quantizing a single random variable and subsequently performing digital encoding on the quantized variable. However, for the class of sources under consideration, this is not so. The proofs of Theorems 2 and 3 indicate that efficient encoding in general requires block lengths L (simultaneous quantization of an L-tuple of random variables), where L is such that

$$(\log K)/L \ll R_a$$

and K is the number of sources dense in the class $\mathcal{A}$. For the example cited, there seems to be no way around this requirement. For the class of gaussian sources of unknown mean this is not so, for one can collect a long block, calculate the sample mean, and then encode the sample mean and the individual normalized random variables in a simple fashion.

It would be of possible interest to consider extending our theory to non-compact classes of sources; in so doing, a number of questions naturally arise. First are there physically interesting examples of such classes? Second, our proof of the positive side of the coding theorem depended heavily on compactness. Is it possible that there are non-compact classes of sources with finite $R_{\mathcal{A}}(d)$ for which no code at rate close to $R_{\mathcal{A}}(d)$ exists? Lastly, are there non-compact classes for which

$$R_{\mathcal{A}}(d) \gtrless \sup_{\alpha\in\mathcal{A}} R_\alpha(d) \quad ?$$

ACKNOWLEDGMENTS

This work depends heavily on the method used by Gallager (1968) to prove the coding theorem for a single source (in particular his Lemma 1) and on the structure of the work of Blackwell, Breiman, and Thomasian (1959) on the capacity of a class of channels. Helpful conversations with P. Varaiya are also gratefully acknowledged.

REFERENCES

Blackwell, D., Breiman, L., and Thomasian, A. (1959). "The Capacity of a Class of Channels," Ann. Math. Statist., 30, 1229-1241.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, John Wiley, New York, 231-232.

Gallager, R. G. (1968). Information Theory and Reliable Communication, John Wiley, New York, Chapter 9 (In press).

Pinsker, M. S. (1964). Information and Information Stability of Random Variables and Processes, Holden-Day, New York, Translated by Amiel Feinstein.

Root, W. L. and T. S. Pitcher, "On the Fourier-series Expansion of Random Functions," Annals of Math. Statistics, 26, June, 1955, pp. 313-318.

Sakrison, D. J. (1968). "A Geometric Treatment of the Source Encoding of a Gaussian Random Variable," IEEE Trans. on Information Theory, IT-14, pp. 481-486.

Sakrison, D. J. (1968). "The Rate Distortion Function of a Gaussian Process with a Weighted Square Error Criterion," IEEE Trans. on Inf. Theory, IT-14, pp. 506-508.

Shannon, C. E. (1948). "A Mathematical Theory of Communication," Bell System Technical Journal, 17, 623-656, see especially Part V.

Shannon, C. E. (1960). "Coding Theorems for a Discrete Source with a Fidelity Criterion," Information and Decision Processes, McGraw Hill, New York, 93-126.