INFORMATION AND CONTROL 40, 181–191 (1979)

Abstract Alphabet Distortion-Rate Functions*

JAMES G. DUNHAM†

Department of Electrical Engineering, Washington University, St. Louis, Missouri 63130

Two definitions have been given for the distortion-rate function of a source-user pair—one involving test channel induced pair probability measures and the other involving general pair probability measures. It is established that both definitions are equivalent for all source-user pairs. Examples are given which exhibit some kinds of the possible pathological behavior of the distortion-rate function for general source-user pairs.

I. INTRODUCTION

The standard definition for the rate-distortion function of a source-user pair, as given by Shannon (1959) and Berger (1971), involves a limit of information theoretic minimizations over source-user pair probability measures on random vectors which are induced by test channels. The operational significance of this rate-distortion function was established by a theorem on block source coding with a fidelity criterion for discrete-time, abstract alphabet, stationary and ergodic sources and single-letter fidelity criteria with a reference letter (Berger (1971), Theorems 7.2.4 and 7.2.5). This theorem showed that the smallest attainable rate using block codes with a distortion constraint was given by this rate-distortion function evaluated at the distortion constraint.

In many communication situations, one has a rate constraint rather than a distortion constraint. For example, one desires to transmit a given source over a given finite capacity channel and the problem is to determine the minimum distortion which can be achieved by an ideal communication system. As is well known, the rate-distortion function is a convex nonincreasing function and, therefore, has a well defined inverse function called the distortion-rate function. The operational significance of this distortion-rate function is that the smallest attainable distortion using block codes with a rate constraint is given by this function evaluated at the rate constraint. Thus the distortion-rate function formulation of the communication problem yields the natural approach for solving problems with a rate constraint.

* This research was supported by NSF Grant ENG 76-02276.

† James G. Dunham was formerly with Stanford University, Stanford, Calif. 94305.

0019-9958/79/020181-11$02.00/0

Gallager (1968); Csiszár (1974); and Gray, Neuhoff and Omura (1975) have considered an alternate definition of the rate-distortion function or, equivalently, the distortion-rate function, which involves a limit of information theoretic minimizations over general source-user pair probability measures on random vectors. As shown by Csiszár (1974), this definition of the rate-distortion function allows one to generalize the familiar finite-alphabet case results for evaluating the rate-distortion function to the general case. As shown by Gray, Neuhoff and Omura (1975), this definition of the distortion-rate function is easily generalized to stationary process distortion-rate functions which are needed to evaluate the performance of stationary and block stationary data compression systems.

For the finite alphabet case, Gray, Neuhoff and Omura (1975) showed that both definitions of the distortion-rate function were the same since every source-user pair probability measure is equivalent to a pair probability measure induced by connecting the source to some test channel. However, in the abstract alphabet case this equivalence is not immediately clear. When the average mutual information between the source and user is finite, it is shown in section III that this equivalence holds and, therefore, both definitions of the rate-distortion function and the distortion-rate function are the same.

Properties of the rate-distortion function have been extensively studied for general source-user pairs, but less attention has been paid to the properties of the distortion-rate function. In section IV examples of source-user pairs are given which show some kinds of the possible pathological behavior of the distortion-rate function in the general case. One example shows that the distortion-rate function may have a discontinuity at rate zero even if it is finite everywhere, paralleling Csiszár's (1974) example where the rate-distortion function is finite everywhere, but discontinuous at distortion zero. Another example shows that the distortion-rate function may be infinite for an interval of rates.

II. Mathematical Preliminaries

Let $(A, \mathscr{A})$ be an abstract measurable space where the abstract space A is called the source alphabet. Define the product measurable space $(A^\infty, \mathscr{A}^\infty) = \times_{k=-\infty}^{\infty} (A_k, \mathscr{A}_k)$ where $(A_k, \mathscr{A}_k) = (A, \mathscr{A})$, all k; A^∞ is the space of all doubly infinite sequences $x = (..., x_{-1}, x_0, x_1, ...)$ from A; and $\mathscr{A}^\infty$ is the usual product σ-algebra. For $n = 1, 2,...$, define the product measurable space $(A^n, \mathscr{A}^n) = \times_{k=1}^{n} (A_k, \mathscr{A}_k)$ and a typical element from A^n is denoted by $x^n = (x_1, ..., x_n)$. Let X_k: $A^\infty \to A$ be the kth scalar coordinate function defined by $X_k(x) = x_k$, all $x \in A^\infty$. Let X^n: $A^n \to A^n$ be a vector coordinate function defined by the identity mapping $X^n(x^n) = x^n$.

We denote by $[A, \mu, X]$ or μ the discrete-time *source* or *process* with underlying alphabet A, probability measure μ on $(A^\infty, \mathscr{A}^\infty)$ and name X. A source is stationary (ergodic) if the probability measure μ is stationary (ergodic). If A is a finite alphabet, then $[A, \mu, X]$ is called a finite alphabet source.

Let $(B, \mathscr{B})$ be an abstract measurable space where the abstract space B is called the source reproduction alphabet. For $n = 1, 2, \ldots$, define, as before, the product measurable spaces $(B^n, \mathscr{B}^n)$ and $((A \times B)^n, (\mathscr{A} \times \mathscr{B})^n)$. A typical element from B^n and $(A \times B)^n$ is denoted by $y^n = (y_1, \ldots, y_n)$ and $(x, y)^n = (x^n, y^n) = (x_1, \ldots, x_n; y_1, \ldots, y_n)$, respectively. Let $Y^n: B^n \to B^n$ and $(X, Y)^n: (A \times B)^n \to (A \times B)^n$ be vector coordinate functions defined by the identity mappings $Y^n(y^n) = y^n$ and $(X, Y)^n(x, y)^n = (x, y)^n = (x^n, y^n)$, respectively.

A mapping $q^n: \mathscr{B}^n \times A^n \to [0, 1]$ is called a regular conditional probability measure if: (1), the set function $q(\cdot \mid x^n)$ is a probability measure on $\mathscr{B}^n$, all $x^n \in A^n$; and (2), the function $q(E \mid \cdot)$ is $\mathscr{A}^n$-measurable, all $E \in \mathscr{B}^n$. We denote by $[A^n, X^n, q^n, B^n, Y^n]$ or q^n the *test channel* connecting the source space A^n to the source reproduction space B^n through the regular conditional probability measure q^n. Let μ^n denote the restriction of the source probability measure μ to the source space A^n. If a source μ^n is "connected" to a test channel q^n, then a *source-user pair probability measure* p^n is indiced on $((A \times B)^n, (\mathscr{A} \times \mathscr{B})^n)$ where

$$p^n(E \times F) = \int_E q^n(F \mid x^n)\, d\mu^n(x^n), \qquad E \in \mathscr{A}^n, \quad F \in \mathscr{B}^n.$$

Denote the probability measure indiced on the source reproduction space B^n by $\nu^n(F) = p^n(A^n \times F)$, all $F \in \mathscr{B}^n$.

If the source-use pair probability measure p^n is absolutely continuous with respect to the product of the marginal probability measures μ^n and ν^n, $p^n \ll \mu^n \times \nu^n$, let $f_n \geqslant 0$ denote the Radon-Nikodym derivative of p^n with respect to $\mu^n \times \nu^n$, that is, $p^n(G) = \int_G f_n \, d\mu^n \times \nu^n$, all $G \in (\mathscr{A} \times \mathscr{B})^n$. Then $\log_2 f_n$ is called the information density and its expectation

$$E_{p^n}[\log_2 f_n] = \int_{(A \times B)^n} f_n \log_2 f_n \, d\mu^n \times \nu^n = I(X^n; Y^n)$$

is the average mutual information between the random vectors X^n and Y^n. If $p^n \ll \mu^n \times \nu^n$, then the information density is not defined and $I(X^n; Y^n) = +\infty$. Pinsker (1964) shows that this definition of the average mutual information $I(X^n; Y^n)$ is equivalent to the usual definition for abstract spaces which involves a supremum over finite measurable partitions.

III. Distortion-Rate Functions

For each integer n, let $\rho_n: (A \times B)^n \to [0, \infty]$ be a measurable function, called a word distortion measure, that specifies the cost in reproducing the source word x^n by the source reproduction word y^n. The family of word distortion measures generates a *fidelity criterion* $F_\rho = \{\rho_n \mid n = 1, 2, \ldots\}$ which is used to measure the accuracy of the user's reproduction of the source. A fidelity criterion F_ρ composed of word distortion measures of the form $\rho_n(x^n, y^n) = n^{-1} \sum_{k=1}^{n} \rho(x_k, y_k)$ is called a single-letter fidelity criterion.

In the literature, two different distortion-rate functions associated with a source-user pair $[A, \mu, X]$ and F_ρ have appeared. Let $\mathscr{P}_n(\mu, R)$ denote the set of all source-user pair probability measures p^n defined on $(A \times B)^n$ which are induced by connecting the source μ^n on A^n to a test channel q^n on $\mathscr{B}^n \times A^n$ and have $I(X^n; Y^n) \leqslant nR$. Then the *test channel distortion-rate function* is defined, following Shannon (1959) and Berger (1971), by

$$D(R) = \inf_n D_n(R),$$

where

$$D_n(R) = \inf_{p \in \mathscr{P}_n(\mu, R)} E_p[\rho_n(X^n, Y^n)]$$

where $D_n(R)$ is infinite of there is no pair measure meeting the condition. Let $\bar{\mathscr{P}}_n(\mu, R)$ denote the set of all source-user pair probability measures $\bar{p}^n$ defined on $(A \times B)^n$ which have $\bar{p}^n(E \times B^n) = \mu^n(E)$, all $E \in \mathscr{A}^n$; and have $I(X^n; Y^n) \leqslant nR$. Then the *pair measure distortion-rate function* is defined, following Gallager (1968) and Csiszár (1974), by

$$\bar{D}(R) = \inf_n \bar{D}_n(R),$$

where

$$\bar{D}_n(R) = \inf_{\bar{p} \in \bar{\mathscr{P}}_n(\mu, R)} E_{\bar{p}}[\rho_n(X^n, Y^n)],$$

where $\bar{D}_n(R)$ is infinite if there is no pair measure meeting the condition.

Since each $p^n \in \mathscr{P}_n(\mu, R)$ is a source-user pair probability measure that has $p^n(E \times B^n) = \mu^n(E)$, all $E \in \mathscr{A}^n$; and has $I(X^n; Y^n) \leqslant nR$, $p^n \in \bar{\mathscr{P}}_n(\mu, R)$. This implies that $\mathscr{P}_n(\mu, R) \subseteq \bar{\mathscr{P}}_n(\mu, R)$ and so $D_n(R) \geqslant \bar{D}_n(R)$ for all n. Thus $D(R) \geqslant \bar{D}(R)$ for all finite rates R.

If every source-user pair probability measure $\bar{p}^n \in \bar{\mathscr{P}}_n(\mu, R)$ were, in fact, induced by connecting the source μ^n to some test channel q^n, then $\mathscr{P}_n(\mu, R) = \bar{\mathscr{P}}_n(\mu, R)$ and we would have $D(R) = \bar{D}(R)$. For example, suppose A and B are finite alphabets. For any $\bar{p}^n \in \bar{\mathscr{P}}_n(\mu, R)$ define $q^n \colon \mathscr{B}^n \times A^n \to [0, 1]$ by

$$\begin{aligned} q^n(y^n \mid x^n) &= p^n(x^n, y^n)/\mu(x^n) \qquad &&\text{if} \quad \mu(x^n) \neq 0 \\ &= 0 &&\text{else} \end{aligned}$$

and it is clear that q^n is a test channel, that is, it is a regular conditional probability measure. Since connecting the source μ^n to q^n induces $\bar{p}^n$, $\mathscr{P}_n(\mu, R) = \bar{\mathscr{P}}_n(\mu, R)$ and so $D(R) = \bar{D}(R)$ for all finite rates. It is known (Parthasarathy (1967)) that if

$(A, \mathscr{A})$ and $(B, \mathscr{B})$ are separable standard Borel spaces, that is, the σ-algebras are σ-isomorphic to the σ-algebra of a complete separable metric space, then such a test channel can always be found since regular conditional probability measures always exist. However, not every abstract measurable space is a separable Borel space.

The following theorem shows that when the average mutual information $I(X^n; Y^n)$ is finite a regular conditional probability measure can always be found and, therefore, $D(R) = \bar{D}(R)$ for all finite rates. Thus both definitions of the distortion-rate function are the same for general source-user pairs.

THEOREM 1. *Let* $[A, \mu, X]$ *be a source and let* F_ρ *be a fidelity criterion. Then* $D(R) = \bar{D}(R)$ *for all rates* $R \in [0, \infty)$.

Proof. Fix a rate $R \in [0, \infty)$, a source-user pair probability measure $\bar{p}^n \in \bar{\mathscr{P}}_n(\mu, R)$ and it suffices to show that $\bar{p}^n \in \mathscr{P}_n(\mu, R)$. Since $I(X^n; Y^n) \leqslant nR < \infty$, $\bar{p}^n$ is absolutely continuous with respect to the product of the marginal measures μ^n and ν^n, $\bar{p}^n \ll \mu^n \times \nu^n$. Let $f_n \geqslant 0$ denote the Radon-Nikodym derivative of $\bar{p}^n$ with respect to $\mu^n \times \nu^n$, that is, $\bar{p}^n(G) = \int_G f_n \, d\mu^n \times \nu^n$, all $G \in (\mathscr{A} \times \mathscr{B})^n$. Define the function $\bar{q}^n\colon \mathscr{B}^n \times A^n \to [0, \infty]$ by $\bar{q}^n(F \mid x^n) = \int_F f(x^n, y^n) \, d\nu^n(y^n)$; all $F \in \mathscr{B}^n$, $x^n \in A^n$. Observe that for each $F \in \mathscr{B}^n$, $\bar{q}^n(F \mid \cdot)$ is an $\mathscr{A}^n$-measurable function (Halmos (1950), Th. 36.B) and for each $x^n \in A^n$, $\bar{q}^n(\cdot \mid x^n)$ is a positive measure on $\mathscr{B}^n$ (Halmos (1950), Sec. 23). However, $\bar{q}^n$ is not necessarily a regular conditional probability measure since $\bar{q}^n(B^n \mid x^n)$ is not necessarily equal to one for all $x^n \in A^n$. Our goal is to find a regular conditional probability measure q^n equivalent to $\bar{q}^n$.

Let $G_n = \{x^n \in A^n \mid \bar{q}^n(B^n \mid x^n) = 1\}$ and we claim that $\mu^n(G_n) = 1$. For each integer k, let $E_k = \{x^n \in A^n \mid \bar{q}^n(B^n \mid x^n) > 1 + 1/k\}$ and let $F_k = \{x^n \in A^n \mid \bar{q}^n(B^n \mid x^n) < 1 - 1/k\}$. Using the Fubini Theorem,

$$\begin{aligned}
\mu^n(E_k) &= \bar{p}^n(E_k \times B^n) \\
&= \int_{E_k \times B^n} f_n \, d\mu^n \times \nu^n \\
&= \int_{E_k} \left[\int_{B^n} f_n(x^n, y^n) \, d\nu^n(y^n) \right] d\mu^n(x^n) \\
&= \int_{E_k} \bar{q}^n(B^n \mid x^n) \, d\mu^n(x^n) \\
&\geqslant (1 + 1/k) \int_{E_k} d\mu^n(x^n).
\end{aligned}$$

Thus $\mu^n(E_k) \geqslant (1 + 1/k) \, \mu^n(E_k)$ which implies that $\mu^n(E_k) = 0$, all k. In a similar fashion, one shows that $\mu^n(F_k) \leqslant (1 - 1/k) \, \mu^n(F_k)$ which implies that $\mu^n(F_k) = 0$, all k. Since $G_n{}^c = \bigcup_{k=1}^{\infty} E_k \cup F_k$, $0 \leqslant \mu^n(G_n{}^c) \leqslant \sum_{k=1}^{\infty} \mu^n(E_k) +$

$\sum_{k=1}^{\infty} \mu^n(F_k) = 0$. Since $G_n{}^c$ is the complement of G_n, $\mu^n(G_n) = 1$, completing the proof of the claim.

Let σ^n be a point probability measure defined on $(B^n, \mathscr{B}^n)$ by

$$\begin{aligned} \sigma^n(F) &= 1 \qquad y^* \in F \\ &= 0 \qquad y^* \notin F \end{aligned}$$

where y^* is an arbitrary but fixed point in B^n. Define the mapping $q^n: \mathscr{B}^n \times A^n \to [0, 1]$ by

$$q^n(F \mid x^n) = \bar{q}^n(F \mid x^n) \chi_{G_n}(x^n) + \sigma^n(F) \chi_{G_n{}^c}(x^n)$$

where χ_G denotes the indicator function of the set G. We claim that q^n is a regular conditional probability measure: (1), since G_n is $\mathscr{A}^n$-measurable and since the sum and product of $\mathscr{A}^n$-measurable functions is $\mathscr{A}^n$-measurable, it follows that $q^n(F \mid \cdot)$ is $\mathscr{A}^n$-measurable for each $F \in \mathscr{B}^n$; and (2), for each $x^n \in A^n$, $q^n(\cdot \mid x^n)$ is a positive measure where $q^n(B^n \mid x^n) = 1$. Thus $q^n(\cdot \mid x^n)$ is a probability measure for each $x^n \in A^n$. Therefore, q^n is a regular conditional probability measure.

Let p^n be the pair probability measure induced by connecting the source μ^n to the test channel q^n, that is, for each $G \in (\mathscr{A} \times \mathscr{B})^n$, $p^n(G) = \int q^n(G(x^n) \mid x^n) \, d\mu^n(x^n)$ where $G(x^n)$ is the x^n section of G. We now show that $p^n = \bar{p}^n$.

For any set $E \times F$ where $E \in \mathscr{A}^n$ and $F \in \mathscr{B}^n$,

$$\begin{aligned} p^n(E \times F) &= \int_E q^n(F \mid x^n) \, d\mu^n(x^n) \\ &= \int_{E \cap G_n} q^n(F \mid x^n) \, d\mu^n(x^n) + \int_{E \cap G_n{}^c} q^n(F \mid x^n) \, d\mu^n(x_n) \qquad (1) \end{aligned}$$

Examining the first term of (1), the definitions of q^n and $\bar{q}^n$ show that

$$\begin{aligned} \int_{E \cap G_n} q^n(F \mid x^n) \, d\mu^n(x^n) &= \int_{E \cap G_n} \bar{q}^n(F \mid x^n) \, d\mu^n(x^n) \\ &= \int_{E \cap G_n} \left[\int_F f_n(x^n, y^n) \, d\nu^n(y^n) \right] d\mu^n(x^n) \\ &= \int_E \left[\int_F f_n(x^n, y^n) \, d\nu^n(y^n) \right] d\mu^n(x^n) \\ &= \bar{p}^n(E \times F), \end{aligned}$$

where the third equality follows from the fact that $0 = \mu^n(G_n{}^c) \geqslant \bar{p}((E \cap G_n{}^c) \times$

$F) \geqslant 0$. Examining the second term of (1), the definition of q^n shows that

$$\begin{aligned} 0 &\leqslant \int_{E\cap G_n^c} q^n(F \mid x^n)\, d\mu^n(x^n) \\ &= \int_{E\cap G_n^c} \sigma^n(F)\, d\mu^n(x^n) \\ &\leqslant \int_{E\cap G_n^c} d\mu^n(x^n) \\ &\leqslant \mu^n(G_n^c) = 0. \end{aligned}$$

Thus $p^n(E \times F) = \bar{p}^n(E \times F)$. Since the probability measures p^n and $\bar{p}^n$ agree on a class of sets which generates the σ-algebra $(\mathscr{A} \times \mathscr{B})^n$ (Halmos (1950), Sec. 33), the standard extension theory of measure theory (Halmos (1950), Th. 13.A) shows that $p^n = \bar{p}^n$. Therefore, $\bar{p}^n \in \mathscr{P}_n(\mu, R)$, completing the proof of the theorem.

IV. Examples of Distortion-Rate Functions

For many source-user pairs of interest, the rate-distortion functioning or, equivalently, the distortion-rate function, can be explicitly evaluated as discussed by Gallager (1968), Berger (1971) and Csiszar (1974). Properties of the rate-distortion have been extensively studied for general source-user pairs, but less attention has been paid to the properties of the distortion-rate function. Examples will be given which show some of the kinds of the possible pathological behavior of the distortion-rate function for general source-user pairs.

Let the source-user pair $[A, \mu, X]$ and F_ρ have a distortion-rate function $D(R)$. Define the following parameters of $D(R)$:

$$R_0 = \inf\{R \geqslant 0 \mid D(R) < \infty\}$$

$$d_0 = \lim_{\epsilon \downarrow 0} D(R_0 + \epsilon) \triangleq D(R_0 +).$$

Since $D(R)$ is a nonincreasing function in R, R_0 represents rate where if $R > R_0$, then $D(R) < \infty$ and if $0 \leqslant R < R_0$, then $D(R) = \infty$. $D(R_0)$ may be finite or infinite. If $D(R)$ is continuous from the right at R_0, then $d_0 = D(R_0)$. The general importance of R_0 is the $D(R)$ is a convex nonincreasing function for all rates $R \geqslant R_0$. The following example is typical of many distortion-rate function where $R_0 = 0$ and $D(R_0) = d_0$.

Example 1. The source is an independent, identically distributed equiprobable binary process with a Hamming distance distortion measure d_H defined by

$$d_H(x, y) = 0 \quad \text{if} \quad x = y$$
$$= 1 \quad \text{else.}$$

The rate-distortion function $R(D)$ is given by

$$R(D) = 1 + D \log_2 D + (1 - D) \log_2(1 - D) \quad \text{if} \quad D \in [0, \tfrac{1}{2}]$$
$$= 0 \quad \text{else.}$$

Letting $R^{-1}(D)$ denote the inverse function of $R(D)$, defined for $R \in [0, 1]$,

$$D(R) = R^{-1}(D) \quad \text{if} \quad R \in [0, 1]$$
$$= 0 \quad \text{else.}$$

Thus it follows that $R_0 = 0$ and $D(R_0) = d_0 = \frac{1}{2}$.

The next example of a source-user pair shows that the distortion-rate function $D(R)$ may have a discontinuity at the rate $R_0 = 0$, paralleling Csiszár's (1974) example where the rate-distortion function $R(D)$ is finite everywhere, but discontinuous at distortion $D = 0$. Since $D(R)$ is convex and nonincreasing for all rates $R \geqslant R_0$, it follows that the only finite rate where $D(R)$ can possibly have a discontinuity is at R_0. We note that this source-user pair is similar to Berger's (1975) example of an information-singular random process but different in that we randomize the initial phase while Berger randomizes the initial translation constant.

EXAMPLE 2. Let $A = \{x \mid x \text{ complex number}, \mid x \mid = 1\}$ where $\mid \cdot \mid$ denotes the modulus (length) of x. Then A is the unit circle in the complex plane and let $\mathscr{A}$ be the Borel subsets of A. Let $[A, \mu, X]$ be the process formed by irrational rotations on the unit circle, that is, let Z be a uniform random variable on $[0, 2\pi)$, let α be an irrational number, and define the complex random process $X_k = e^{i(\alpha k+Z)}$ and note that $X_k = e^{i\alpha k} \exp(iZ) = c^k X$. This process is easily shown to be stationary and ergodic (Billingsley (1965), pp. 9–11).

Let the source reproduction space $B = A$ and define a per-letter distortion measure ρ by $\rho(x, y) = \mid x - y \mid^2$. Let F_ρ be the single-letter fidelity criterion generated by the per-letter distortion measure ρ. For the source-user pair $[A, \mu, X]$ and F_ρ, let $D(R)$ be the associated distortion-rate function and the following theorem gives the explicit evaluation of $D(R)$.

THEOREM 2. *Let the source $[A, \mu, X]$ and the fidelity criterion F_ρ be given as above. Then*

$$D(R) = \begin{cases} 1/2\pi \int_0^{2\pi} \mid e^{iz} - 1 \mid^2 dz & R = 0 \\ 0 & R > 0. \end{cases}$$

Proof. Fix a rate $R \in (0, \infty)$ and, for each integer n, divide the interval $[0, 2\pi)$ into $[2^{nR}]$ equal-length subintervals, where $[z]$ denotes the greatest integer bounded by z. Let $\hat{Z}$ denote the center of the interval in which Z lies and then associate $Y = \exp(i\hat{Z})$ with the corresponding random variable X. Next, associate $X_k = c^k X$ with $Y_k = c^k Y$ and let p^n denote the induced source-user pair probability measure. Since Y^n can produce at most $[2^{nR}]$ distinct blocks of source reproduction symbols.

$$n^{-1}I(X^n; Y^n) \leqslant n^{-1}H(Y^n) \leqslant n^{-1}\log[2^{nR}] \leqslant R,$$

where $H(Y^n)$ is the usual definition of the entropy of a finite alphabet random variable Y^n. Computing the expected distortion,

$$E[\rho^n(X^n, Y^n)] = E[| e^{i(Z-Z)} - 1 |^2] \leqslant \pi^2 2^{-2nR}$$

since $| Z - \hat{Z} | \leqslant \pi 2^{-nR}$ and $| e^{iz} - 1 | \leqslant | z |$ for all real z. Thus $D_n(R) \leqslant \pi^2 2^{-2nR}$ and so $D(R) = 0$.

We now evaluate $D(0)$. Since $R = 0$, any $p^n \in \mathscr{P}_n(\mu, 0)$ has $I(X^n; Y^n) = 0$ and Pinsker (1964), Eq. (2.2.1), shows that $p^n = \mu^n \times \nu^n$. Evaluating the distortion, the Fubini theorem and the stationarity of μ show that

$$\begin{aligned} E[\rho_n(X^n, Y^n)] &= \int_{B^n} \left[n^{-1} \sum_{k=1}^{n} \int_{A_k} | X_k(x^n) - Y_k(y^n)|^2 \, d\mu^n(x^n) \right] d\nu^n(y^n) \\ &= \int_{B^n} \left[n^{-1} \sum_{k=1}^{n} 1/2\pi \int_0^{2\pi} | c^k e^{iz} - Y_k(y^n)|^2 \, dz \right] d\nu^n(y^n) \\ &= \int_{B^n} \left[n^{-1} \sum_{k=1}^{n} 1/2\pi \int_0^{2\pi} | e^{iz} - 1 |^2 \, dz \right] d\nu^n(y^n) \\ &= 1/2\pi \int_0^{2\pi} | e^{iz} - 1 |^2 \, dz. \end{aligned}$$

Therefore, $D(0) = 1/2\pi \int_0^{2\pi} | e^{iz} - 1 |^2 \, dz$, completing the proof of the theorem.

Next, we give an example of a finite-alphabet source-user pair which shows that R_0 may be finite and strictly greater than 0. Thus $D(R)$ may be infinite for the interval of rates $[0, R_0)$ and finite for the semi-infinite interval of rates $(R_0, \infty]$

EXAMPLE 3. Let $A = \{0, 1, 2, 3\}$, let $\lambda \in (0, 1)$ and let the source $[A, \mu, X]$ be an independent, identically distributed random process where

$$\begin{aligned} \mu(\{i\}) &= \lambda/2 & i &= 0, 1 \\ &= (1 - \lambda)/2 & i &= 2, 3. \end{aligned}$$

Let $B = A$ and define a per-letter distortion measure ρ by

$$\rho(x, y) = d_H(x, y) \qquad \text{if} \quad [y/2] \leqslant x/2 < [y/2] + 1$$
$$= \infty \qquad \text{else}$$

where d_H is the Hamming distance distortion measure of Example 1 and $[x]$ denotes the greatest integer bounded by x. Let F_ρ be the single-letter fidelity criterion generated by ρ. For the source-user pair $[A, \mu, X]$ and F_ρ, let $D(R)$ be the associated distortion-rate function and the following theorem gives the explicit evaluation of $D(R)$.

THEOREM 3. *Let the source $[A, \mu, X]$ and the fidelity criterion F_ρ be as above. Then*

$$D(R) = D'(R - h_b(\lambda)) \qquad R \geqslant h_b(\lambda)$$
$$= \infty \qquad \text{else}$$

where $D'(R)$ is the distortion-rate function of Example 1 and $h_b(\lambda) = -\lambda \log_2 \lambda - (1 - \lambda) \log_2(1 - \lambda)$.

Proof. Since the source is memoryless and the fidelity criterion is single-letter, $D(R) = D_1(R)$. For a rate R and a pair probability measure $p \in \mathscr{P}_1(\mu, R)$, suppose that $E_p[\rho(X^1, Y^1)] < \infty$. The distortion measure ρ implies that 0(1) must be reproduced as a 0 or 1 and that 2(3) must be reproduced as a 2 or 3. Thus the pair probability measure p can be expressed as $p = \lambda p_1 + (1 - \lambda) p_2$ where p_1 is a pair probability measure on $\{0, 1\} \times \{0, 1\}$ and p_2 is a pair probability measure on $\{2, 3\} \times \{2, 3\}$. Let $\mu_1(\mu_2)$ and $\nu_1(\nu_2)$ be the corresponding A and B marginal probability measures, respectively, of $p_1(p_2)$. Then $\mu = \lambda\mu_1 + (1 - \lambda) \mu_2$ and $\nu = \lambda\nu_1 + (1 - \lambda) \nu_2$. For the average mutual information,

$$I_p(X; Y) = \lambda I_{p_1}(X; Y) + (1 - \lambda) I_{p_2}(X; Y) + h_b(\lambda) \tag{2}$$

where $h_b(\lambda) = -\lambda \log_2\lambda - (1 - \lambda) \log_2(1 - \lambda)$ and for the expected distortion,

$$E_p[\rho(X, Y)] = \lambda E_{p_1}[\rho(X, Y)] + (1 - \lambda) E_{p_2}[\rho(X, Y)]. \tag{3}$$

The source-user pair μ_1 and F_ρ and the source-user pair μ_2 and F_ρ are equivalent to the source-user pair of Example 1. Let $D'(R)$ denote the distortion-rate function of Example 1. Then (2) and (3) together imply that

$$D(R) \leqslant \lambda D'(R_1) + (1 - \lambda) D'(R_2) \tag{4}$$

where $R \geqslant h_b(\lambda) + \lambda R_1 + (1 - \lambda) R_2$.

Let $D^*(R) = \inf\{\lambda D'(R_1) + (1 - \lambda) D'(R_2) \mid R \geqslant h_b(\lambda) + \lambda R_1 + (1 - \lambda)R_2\}$. We claim that $D(R) = D^*(R)$ for all rates $R \geqslant h_b(\lambda)$. First, (4) shows that

$D(R) \leqslant D^*(R)$, all $R \geqslant h_b(\lambda)$. Fix an $\epsilon > 0$, a rate $R \geqslant h_b(\lambda)$ and choose a pair probability measure $p \in \mathscr{P}_1(\mu, R)$ where $E_p[\rho(X, Y)] \leqslant D(R) + \epsilon < \infty$. Then

$$\begin{aligned} D(R) + \epsilon &\geqslant E_p[\rho(X, Y)] \\ &\geqslant \lambda E_{p_1}[\rho(X, Y)] + (1 - \lambda) E_{p_2}[\rho(X, Y)] \\ &\geqslant \lambda D'(I_{p_1}(X; Y)) + (1 - \lambda) D'(I_{p_2}(X; Y)) \\ &\geqslant D^*(R). \end{aligned}$$

Since this is true for all $\epsilon > 0$, $D(R) \geqslant D^*(R)$ and, therefore, $D(R) = D^*(R)$, $R \geqslant h_b(\lambda)$, completing the proof of the claim.

Since $D'(R)$ is a convex $\cup$ function for $R \geqslant R_0 = 0$,

$$\begin{aligned} \lambda D'(R_1) + (1 - \lambda) D'(R_2) &\geqslant D'(\lambda R_1 + (1 - \lambda) R_2) \\ &= D'(R - h_b(\lambda)). \end{aligned} \tag{5}$$

for all R_1 and R_2 where $R - h_b(\lambda) = \lambda R_1 + (1 - \lambda) R_2$. Equality holds in (5) when $R_1 = R_2 = R - h_b(\lambda)$. Therefore, $D'(R - h_b(\lambda)) = D^*(R) = D(R)$ for all rates $R \geqslant h_b(\lambda)$.

Suppose that $D(R) < \infty$ for some rate $R < h_b(\lambda)$. Since the average mutual information is always nonnegative, (2) is violated, producing a contradiction. Therefore, $D(R) = \infty$ for all rates $0 \leqslant R < h_b(\lambda)$, completing the proof of the theorem.

RECEIVED: May 27, 1978

REFERENCES

1. BERGER, T. (1971), "Rate-Distortion Theory," Prentice–Hall, Englewood Cliffs, N. J.
2. BERGER, T. (1975), Information-singular random processes, *IEEE Trans. Information Theory* **21**, 502–511.
3. BILLINGSLEY, P. (1965), "Ergodic Theory and Information," Wiley, New York.
4. CSISÁR, I. (1974), On an extremum problem of information theory, *Studia Sci. Math. Hungar.* **9**, 57–71.
5. GALLAGER, R. G. (1968), "Information Theory and Reliable Communication," Wiley, New York.
6. GRAY, R. M., NEUHOFF, D. C., AND OMURA, J. K. (1975), Process definitions of distortion-rate functions and source coding theorems, *IEEE Trans. Information Theory* **21**, 524–532.
7. HALMOS, P. (1950), "Measure Theory," Van Nostrand Reinhlld, New York.
8. PARTHASARATHY, K. R. (1967), "Probability Measures on Metric Spaces," Academic Press, New York.
9. Pinsker, M. S. (1964), "Information and Information Stability of Random Variables and Processes," Holden–Day, San Francisco.
10. SHANNON, C. E. (1959), Coding theorems for a discrete source with a fidelity criterion, *in IRE National Convention Record*, Part 4, 142–163.