Appl. Math. and Comp. Sci., 1996, vol.6, No.1, 123–136

THE RATE DISTORTION REGION FOR CODING IN STATIONARY SYSTEMS

ILAN SADEH*

We present the idea of process assignment as a basic tool to provide operational compression bounds. We give a simple proof for the operational rate distortion function using process assignment arguments for ergodic and stationary systems. We generalize the problem to multiterminal systems. A terse review of random process theory is followed by two examples in communications and computers industry, where this theory supplies key bounds to the network performance. Coupling the process definitions with the mathematical stochastic representation of process assignment provides a new and simple proof of the achievable rate region for a degraded diversity system, or a multiple-description system. The proof can be extended to more complicated communications networks. The equivalence of Shannon-type and operational bounds is also addressed.

1. Introduction

It is known from Shannon's work (Berger, 1971; Gray, 1975; Omura, 1973; Ornstein and Shields, 1990; Ziv, 1972) and many others that every block code for data compression of block-length l and an average per-letter distortion D for a finite-alphabet, stationary and ergodic source has a rate greater than or equal to $R(D)$. We show that the same result holds for any compression algorithm, not necessarily related to block coding, where stationarity is preserved. The proof is short and simple. The result is not new, since Ornstein and Shields (1990) mentioned it, but a simpler proof possesses its merits and clarity.

The main new idea is the mathematical representation of process assignment by stochastic terms. We use it in the proof for the rate distortion function for ergodic and stationary sources. Coupling the process definitions with the mathematical stochastic representation of process assignment provides new and simple proofs of various source coding theorems for ergodic sources. Gray *et al.* (1975) used ideas based on process definitions to prove some source coding theorems.

We assume that the system preserves stationarity. However, the block-code does not preserve stationarity but only L-stationarity. To include the block-code in the analysis, block stationarity or L-stationarity should be adapted by randomization as done in (Gray *et al.*, 1975).

* Centre for Technological Education and Research, Holon; affiliated with Tel Aviv University, Department of Communications Engineering, P.O. Box 305, Holon, 58102 Israel, e-mail: sade@math.tau.ac.il

We use the process assignment concept and extend the results for multi-rate multi-distortion systems (also known as multiple-description systems—MDS). The problem is discussed in (Ahlswede, 1985; Berger and Zhang, 1983; Cover and Gamal, 1982; Zhand and Berger, 1987) and presented in (Blahut, 1987). Such systems are usually defined as diversity ones. A diversity system for communication sends several copies of the same message to a user through several different channels, so that even if all but one channel are broken, the message will arrive at the receiver. A degraded diversity system is more subtle and requires only a part of the channel capacity. It sends a part of the message through each of two channels but it does so in such a way that either part suffices to reconstruct a degraded copy of the message. A possible application is in the voice telephony. A digitized voice signal could be halved and sent over two routes. If either route is blocked or disconnected, a reduced-fidelity reproduction is still available to the receiver. Another application is the case of image communication where a large amount of information is transmitted on parallel cables. Clearly, the same idea can be extended to n routes.

To illustrate the importance of diversity systems for communication, we describe the Galileo case as presented by Cheung and Tong (1993). The Galileo spacecraft is currently on its way to Jupiter and its moons. In April 1991, the high-gain antenna (HGA) failed to deploy as commanded. In the case when the current efforts to deploy the HGA fail, communications during the Jupiter encounter will be through one of the low-gain antennas (LGA) on an S-band (2.3 Ghz) carrier. A lot of efforts have been made to open the HGA. Also various options for improving Galileo's telemetry down link performance are evaluated in the event that the HGA does not open on the arrival in Jupiter. Among all available options the most promising one is to perform image and non-image data compression using software on board of the spacecraft. This involves in-flight re-programming of the existing flight software of the spacecraft processors which have very limited computational and memory resources.

The solution in that case was based on a lossy image compression scheme. The rest of the data comes from various spacecraft instruments. This can either be compressed by using instrument-specific compression schemes or by using a lossless universal compression algorithm.

Moreover, projects like the Galileo involve over 20 years of efforts. We believe there should be at least double links to secure that a hardware failure such as Galileo's S-band contingency will not destroy the whole mission. In such high-risk missions, a diversity system for communication is indispensable. It does not make sense to risk the whole mission because of a possible hardware failure.

Another important application of MDS (Multiple Description System) is in the computer industry. Suppose one wants to compress images on a shared memory machine where each memory portion has its own properties of access time and size. The requirement is that it is possible to obtain a "fast" reconstructed image with a distortion level D_1 while memory access is through the "cache" only. An almost perfect reconstructed image with a distortion level D_0 is required where all memory sections are accessible. Moreover, we can extend the problem to a multi-processing system where each processor has its own cache and can reconstruct the image with distortion level D_1. While they work simultaneously in parallel, all processors share the whole data in memory and can have almost perfect reconstruction.

We are interested in systems where more than one point-to-point information links are available. Generally speaking, a multi-terminal network allows for more than one data user and more than one channel between them. Information networks can have conflicting requirements imposed by several terminals.

We generalize the case of a point-to-point link to the case where the goal is to break a message into two (or more) distorted replicas which contain together enough information to reconstruct the original message. The question is to design and prove the bounds for the achievable rate regions where distortion levels are defined for the cases. It is known that codes exist for degraded diversity systems, though it is not known how to construct the code.

The main aim of this paper is to prove the extension of the rate distortion function $R(D)$ to the multi-terminal case. Data compression for diversity s ystems was treated by Cover and El Gamal (1982). They gave a region of achievable rates and distortions which describe the inner bound of the achievable rates. Berger and Zhang (1983) established tightness in a special case. We give an exact bound.

The paper is organized as follows. Section 2 contains definitions and the new proof of the theorem about the rate distortion function. Section 3 is devoted to the multi-rate multi-distortion functions. A new result is obtained for the minimal achievable rate region given a vector of tolerable distortions. Section 4 is a short summary of the presented results.

2. Rate Distortion Function for Stationary Systems

Let u be an ergodic finite-valued stationary sequence with entropy rate H. Let $\bar{u}$ denote a sample sequence or a block. Here, by the block we mean of course a block of consecutive symbols in the sequence. The notation u_i^j stands for a block between positions i and j in the sequence u. We shall require the standard definition of entropy rate

$$H(u) = \lim_{n \to \infty} \frac{1}{n} H(u_1^n) \tag{1}$$

Given alphabets U and V, a distortion measure is any function $d : |U \times V| \to \mathbb{R}^+$. The function d measures the distortion (cost, penalty, loss, etc.) suffered each time when the source produces letters $u \in U$ and the user receives letters $v \in V$. Usually, the range of d is finite and without loss of generality we may assume that, for all $u_i \in U$, $\min_j d(u_i, v_j) = 0$.

Let $\rho_n(\bar{u}; \bar{v})$ denote the average value of the 'per letter' distortions for the letters that comprise the block $\bar{u}$,

$$\rho_n(\bar{u}; \bar{v}) = \frac{1}{n} \sum_{k=1}^{n} d(\bar{u}_k; \bar{v}_k) \tag{2}$$

The pair $(\bar{u}_k; \bar{v}_k)$ denotes the letters at the k-th position at the source and the user, respectively. The distortion is assumed to be memoryless.

Let Q be the transition matrix which describes a channel where the input in the past u_1^n has been transformed into the output process v_1^n defined on the alphabet V. That Q approximates the data compression algorithm at the level of a single symbol, even though, as explained later, at the whole process level the transformation from the input to the output process is deterministic. Here and subsequently, i and j stand for the letters from the alphabets U and V, respectively. The transition matrix represents the uncertainty with respect to each pair $(u_k; v_k)$ at some position k, without any knowledge about the past, i.e.

$$Q(j \mid i) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \Pr(v_k = j \mid u_k = i) \tag{3}$$

For a joint stationary source reproduction pair process, we have $Q(j \mid i) = \Pr(v_k = j \mid u_k = i)$ for all k.

Due to the ergodicity of the source and the law of large numbers, it is almost sure that

$$\lim_{n \to \infty} \rho_n(\bar{u}; \bar{v}) = E_Q \rho_n(\bar{u}; \bar{v}) \tag{4}$$

The idea of a "process assignment" is the following: all the optimal encoders are "quasi-deterministic" in the sense that there is almost no randomness in the assignment of a time process v to a time process u. Even if there is some randomness in the assignment process, we impose that its contribution to the conditional entropy rate vanishes in the limit. The quasi-determinism of the encoder-decoder pair imposes the constraint that the conditional entropy of the process v, given the process u, is zero, i.e.

$$H(v \mid u) = \lim_{n \to \infty} \frac{1}{n} H(v_1^n \mid u_1^n) = 0 \tag{5}$$

It is justified by the fact that a non-deterministic machine yields a strictly positive conditional entropy rate $H(v \mid u) > 0$. Hence, if the non-deterministic encoder attains the minimal information $I(u, v) = R(D)$, the reproduction entropy rate is $H(v) = R(D) + H(v \mid u) > R(D)$. Hence, the rate and the channel capacity required for a reliable transmission with fidelity D is strictly larger than $R(D)$. This is a contradiction to Theorem 7.2.6 in (Berger, 1971), because the source output cannot be reproduced with fidelity D at the receiver end of a channel with capacity C such that $H(v) > C > R(D)$. The idea was discussed in (Berger, 1971; Gray *et al.*, 1975). The deterministic mapping agrees with the remarkable implication of the sliding-block coding theorem which asserts that there exists a joint measure with the property (5) that yields distortion and rate arbitrarily close to the infimum.

However, we show that if we restrict ourselves to quasi-deterministic mappings, the result for $R(D)$ in stationary systems is obtained in a very simple way. In the sequel, we shall minimize the compression ratio over the class of all encoder-decoder pairs. Obviously, this is equivalent to taking the infimum with respect to Q over all possible transition matrices while keeping the determinism condition (5).

Since the processes are jointly stationary, we can define the per-letter average mutual information rate as

$$I_Q(u;v) = \lim_{n\to\infty} \frac{1}{n} I(v_1^n; u_1^n)$$

where the subscript Q emphasizes the dependence on the transition matrix Q.

It is known that the minimum achievable rate of a code with average per-letter distortion D is denoted by $R(D)$ and is called the rate distortion function. The function $R(D)$ measures the abridged information content at distortion D. The function $R(D)$ divides the rate-distortion plane into the set of points for which good codes exist and the set of points for which no codes for data compression exist. The existence of a code implies the existence of Q such that (5) is valid and that ρ_Q defined by (4) is not greater than D. The rate distortion function $R(D)$ is given by the minimal mutual information per source symbol subject to the constraint on the average distortion. It is known (Berger, 1971) that, given an ergodic source,

$$\lim_{l\to\infty} \inf_{\widehat{Q}} \frac{1}{l} I_{\widehat{Q}}\left(u_0^{l-1}, v_0^{l-1}\right) = R(D) \tag{6}$$

where the infimum with respect to $\widehat{Q}$ is taken over all the conditional distributions on $U^l \times V^l$, having one l-dimensional marginal and satisfying

$$E_{\widehat{Q}} \rho_l(\bar{u}, \bar{v}) \leq D$$

Shannon's theorem shows (Berger, 1971) that, for ergodic sources, $R(D)$ is the lowest attainable rate by any block code with average distortion not exceeding D. It was also proved by Ornstein and Shields (1990) that $R(D)$ is the bound for variable-length coding. The equality of the operational and information-theoretic rate-distortion functions is well-known for all stationary (or invariant) codes, i.e. time-invariant deterministic mappings from input sequences to output sequences. However, we present here a simple proof for the stationary case. The result is not new, but we believe that the simple proof justifies its presentation. It is also a preparation for the multi-terminal case.

We modify the definition of the rate distortion function to fit the process assignment idea.

Definition 1. The *operational rate distortion function* $R(D)$ is defined to be the minimal entropy rate of the infinite output process v generated by the infinite input process u and a process assignment algorithm (encoder and decoder pair) such that the average per-letter distortion is at most D.

The assignment of another process to a process is identical to produce a block-coding of infinite length.

We assume a general ergodic stationary source u. We focus attention on sources with a finite alphabet U and we assume for each n the existence of a probability mass function (pmf in brief)

$$p^n(\bar{u}) = \Pr(u_1^n = \bar{u}) \quad \forall\, \bar{u} \in U^n$$

where $\bar{u}$ is a block of length n and U^n is the collection of all n-tuples with coordinates in U. We denote by i and j the letters in the alphabets U and V, respectively.

Theorem 1. *Given an ergodic stationary source u, denoted by $[U,p]$, suppose that a distortion measure $d(i,j)$ and a non-negative number D are selected. If the system preserves stationarity, then the operational rate distortion function is equal to the Shannon-type rate distortion function, i.e.*

$$\inf_{Q:\sum_i \sum_j p(i)Q(j\,|\,i)d(i,j)\le D} I(p,Q) = R(D) \tag{7}$$

Proof. Define the set of all output sequences v generated by the input sequence u with distribution p and an encoder-decoder pair simulated by the conditional transition matrix Q with the "quasi-determinism" property (5),

$$\mathcal{V}(p,Q) = \{v : H(v\,|\,u) = 0\}$$

This is because there is almost no randomness in the assignment of the time process v to the time process u. The process assignment of the encoder-decoder pair imposes the constraint that the conditional entropy of the process v, given the process u, vanishes in the limit (5).

$$H(v\,|\,u) = 0 \tag{8}$$

Due to the ergodicity of the source and the law of large numbers, it is almost sure that (4) holds and, by using the stationarity property, we obtain

$$\lim_{n\to\infty} \rho_n(u_1^n; v_1^n) = E_Q\rho_n(u_1^n; v_1^n) = \sum_i \sum_j p(i)Q(j\,|\,i)d(i,j) = \rho_Q \tag{9}$$

Since the pair (u,v) is required to be stationary, the joint probability distributions are invariant under translation of the time origin. Thus, the expected value of the average distortion needs only be computed at a single time, i.e.

$$E_Q\rho_n(u_1^n; v_1^n) = \frac{1}{n}\sum_{k=1}^{n} E_Q d(u_k; v_k) = E_Q d(u_0; v_0) = \rho_Q$$

The encoder-decoder pair is represented in the stationary case by the conditional empirical per-letter distribution Q. The rate-distortion function is obtained by minimization of the output sequence entropy over the class of all sequences that are obtained by a quasi-deterministic mapping under the fidelity constraint from the input sequence and over the class of all such mappings at the limit as the sequence length tends to infinity. The minimization over the class of all quasi-deterministic encoder-decoder pairs is equivalent to the minimizing with respect to Q over all possible rational transition matrices (or infimum over Q as a bound) while keeping the fidelity constraint

$$R(D) = \lim_{n\to\infty} \inf_{Q^n:\rho_n(u_1^n;v_1^n)\le D} \; \inf_{v_1^n:H(v_1^n\,|\,u_1^n)=0} \frac{1}{n}\left\{H(v_1^n)\right\}$$

By taking the limit, using (9) and the process definitions, we obtain

$$\inf_{Q:\sum_i \sum_j p(i)Q(j\mid i)d(i,j)\le D} \; \inf_{v\in\mathcal{V}(p,Q)} \left\{ H(v) \right\} = R(D) \tag{10}$$

Using the determinism condition (5) and (8), we have

$$\inf_{Q:\sum_i \sum_j p(i)Q(j\mid i)d(i,j)\le D} I(p,Q) = R(D)$$

∎

Corollary 1. *The operational rate distortion function in the sense of the minimal output sequence entropy over the class of all quasi-deterministic mappings from the input sequence, such that the average per-letter distortion is at most D, is equivalent in stationary systems to the Shannon-type definition of $R(D)$ as the minimum rate at which the source produces information subject to the requirement that its output must be reproduced with average fidelity D.*

3. Multi-Rate Multi-Distortion Functions

Similar results hold when not only one point-to-point information link is available. We generalize the previous case to the case where the goal is to break a message into two (or more) distorted replicas which contain together enough information to reconstruct the original message. The question is to design and prove the bounds for the achievable rates regions in networks where distortion levels are defined for the cases. We present the following figure to visualize the problem.

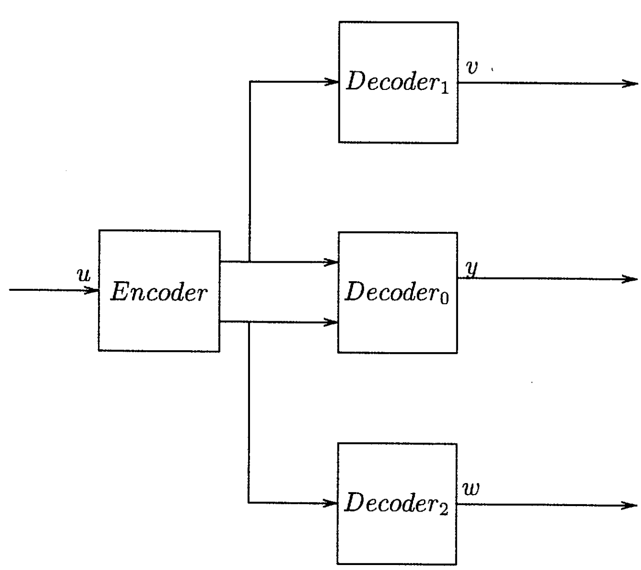

Fig. 1. Degraded diversity system.

Given alphabets U and Y, a distortion measure d_0 is defined as $d_0 : |U \times Y| \longrightarrow \mathrm{IR}^+$. Given alphabets U and V, a distortion measure d_1 is a function $d_1 : |U \times V| \longrightarrow \mathrm{IR}^+$. Given alphabets U and W, a distortion measure d_2 is defined as $d_2 : |U \times W| \longrightarrow \mathrm{IR}^+$. The functions d_0, d_1, and d_2 measure the distortion (cost, penalty, loss, etc.) suffered each time the source produces letters $u \in U$ and the user receives letters $v \in V$, $w \in W$, or $y \in Y$, respectively. Usually, the range of d_1 is finite and without loss of generality we may assume that for, all k, $\min_l d_1(u_k, v_l) = 0$. The same rules hold for d_2 and d_0.

Let $\rho_n^0(\bar{u}; \bar{y})$ denote the average value of the 'per letter' distortions for the letters that comprise the block

$$\rho_n^0(\bar{u}; \bar{y}) = \frac{1}{n} \sum_{i=1}^{n} d_0(\bar{u}_i; \bar{y}_i) \tag{11}$$

The pair $(\bar{u}_i; \bar{y}_i)$ denotes the letters at the i-th position at the source and the user, respectively. The distortion is assumed to be memoryless.

Let $\rho_n^1(\bar{u}; \bar{v})$ denote the average value of the 'per letter' distortions for the letters that comprise the block

$$\rho_n^1(\bar{u}; \bar{v}) = \frac{1}{n} \sum_{i=1}^{n} d_1(\bar{u}_i; \bar{v}_i) \tag{12}$$

The pair $(\bar{u}_i; \bar{v}_i)$ denotes the letters at the i-th position at the source and the user, respectively. The distortion is assumed to be memoryless.

Let $\rho_n^2(\bar{u}; \bar{w})$ denote the average value of the 'per letter' distortions for the letters that comprise the block

$$\rho_n^2(\bar{u}; \bar{w}) = \frac{1}{n} \sum_{i=1}^{n} d_2(\bar{u}_i; \bar{w}_i) \tag{13}$$

The pair $(\bar{u}_i; \bar{w}_i)$ denotes the letters at the i-th position at the source and the user, respectively. The distortion is assumed to be memoryless.

We formulate the general problem as in (Blahut, 1987). There are four alphabets—the source alphabet U and three reproducing alphabets V, W, and Y, possibly of different sizes. A stream of symbols from the source alphabet enters the encoder, and two bit streams leave the encoder at the rates R_1 bits per input symbol and R_2 bits per input symbol, respectively.

Both bit streams enter the central decoder. Only one bit stream enters each side decoder. Side decoder 1 is required to describe the source data using the alphabet V with average distortion D_1 under the distortion measure d_1. Side decoder 2 is required to describe the source data using the alphabet W with average distortion D_2 under the distortion measure d_2. The central decoder is required to describe the source data using the alphabet Y with average distortion D_0 under the distortion measure d_0.

For each value of distortion vector (D_0, D_1, D_2) there is an achievable rate region defined as the set of rate pairs R_1 and R_2 for which there exist codes of distortion at most (D_0, D_1, D_2). The achievable rate region is not known except for the simple case of a binary degraded diversity system for a binary symmetric source.

In this paper, we present an exact solution for the achievable rate region. So far only an outer and an inner bound have been known. A crude outer bound on the region is as follows.

Theorem 2. (Outer Bound) *The pair (R_1, R_2) is not in the achievable rate region (abbreviated as ARR) unless*

$$R_1 + R_2 \geq R(D_0), \qquad R_1 \geq R(D_1), \qquad R_2 \geq R(D_2)$$

However, we should not expect that all the rate pairs satisfying these inequalities are in the ARR. This is because a part of the information in the two codewords must be dependent so that the needs of the side decoders can be satisfied. The redundant portion of the information will be useless to the central decoder. In addition, the central decoder may need some detailed information that neither side decoder needs.

The inner bound on the ARR has been developed by using a single letter view of the data compression, even though they have assumed that the encoder and the decoder work on blocks. The proof of the inner bound has started by assuming codewords that satisfy the needs of the side decoders and appending to these codewords additional information that will satisfy the needs of the central decoder. However, this method is not guaranteed to be the most efficient one since it is not known as well and tight as the corresponding outer bound. Moreover, the method assumes only block-coding while we guess there are codes which are achievable only by a non-block-coding method.

The inner bound on the ARR developed by Cover and El Gamal (1982) by using random coding arguments is as follows.

Theorem 3. (Inner Bound) *The ARR contains the convex hull of the set of all (R_1, R_2) such that*

$$R_1 + R_2 \geq I(u; (v, w, y)) + I(v; w)$$

$$R_1 \geq I(u; v), \qquad R_2 \geq I(u; w)$$

for some conditional probability distribution $Q = Q_{vwy\,|\,u}$ such that

$$D_0 \geq Ed_0(u, y), \qquad D_1 \geq Ed_1(u, v), \qquad D_2 \geq Ed_2(u, w)$$

From the Inner Bound we conclude that the process assignment idea is also valid for the multi-terminal case. Suppose the pair *Encoder-Decoder*$_1$ is not quasi-deterministic and $H(v\,|\,u) > 0$. Such a stochastic source code might result in a satisfactory average distortion D_1 if the other channel is broken, but the entropy rate of the reproduction v is too large. Since the Inner Bound is defined by constraining the average mutual information rate $I(u, v)$ to be equal to the rate R_1 on its specific channel, the reproduction entropy $H(v) = I(u,v) + H(v\,|\,u)$ is greater than R_1 in general, since the conditional entropy rate is assumed to be strictly positive. Hence, the test channel on Channel 1 does not provide a good stochastic code since the rate and, therefore, the channel capacity required for reliable transmission are strictly larger than the constraint given by the Inner Bound. This is a contradiction to the Inner Bound that states that all pairs (R_1, R_2) that are equal to the bounds with average distortion levels in the constraints are included in the ARR. Since an ideal system achieves exactly the bound, the pair in question is not ideal. Thus, we conclude that the pair *Encoder Decoder*$_1$ is quasi-deterministic.

By a similar reasoning we conclude that also the pair *Encoder Decoder*$_2$ is quasi-deterministic.

We conclude this discussion by asserting that also *Decoder*$_0$ must be quasi-deterministic, since the Encoder is. It does not make sense that the decoding function is not quasi-deterministic while the encoding function is, and also the decoding functions, while one of the channels is broken, are quasi-deterministic. Thus, we obtain that

$$H(v\,|\,u) = H(w\,|\,u) = H(y\,|\,u) = 0 \tag{14}$$

in the ideal system. However, this result is based on random coding arguments. We prefer to define the *ARR*, as done in the previous section, as a deterministic mapping between processes. The so-called operational *ARR* is an extension of the operational rate-distortion function. The equivalence of the Shannon-type definition of the *ARR* and the operational *ARR* will be concluded.

Definition 2. The *ARR* (operational achievable rate region) is the region of all rate pairs (R_1, R_2) on the channels, such that the deterministic mapping from the input process u to the output processes y, v, and w yields average per-letter distortions not exceeding D_0, D_1, and D_2, respectively.

The following theorem will describe the exact bound. We denote this bound as the multi-rate multi-distortion function.

Theorem 4. (The Multi-Rate Multi-Distortion Function Theorem) *Let u be a stationary ergodic source with finite alphabet U and distribution p. Let v, w, and y be jointly stationary ergodic reproduction processes with finite alphabets V, W, and Y, respectively. Given a distortion vector (D_0, D_1, D_2), the achievable rate region (ARR) for a degraded diversity system is the region of all R_1, and R_2 defined by the convex hull bounded according to the following inequalities:*

$$R_1 + R_2 \geq I(u;y), \qquad R_1 \geq I(u;v), \qquad R_2 \geq I(u;w)$$

for all conditional probability distributions $Q = Q_{vwy\,|\,u}$ such that

$$D_0 \geq Ed_0(u,y), \qquad D_1 \geq Ed_1(u,v), \qquad D_2 \geq Ed_2(u,w)$$

Proof. Following the previous discussion, for all pairs $(R_1, R_2) \in ARR$, there exists an ideal process assignment machine (encoder-decoder system) that compresses the input u such that

$$R_1 + R_2 \geq H(y), \qquad R_1 \geq H(v), \qquad R_2 \geq H(w)$$

and the average distortions satisfy the inequalities

$$D_0 \geq \lim_{n\to\infty} \rho_n^0(\bar{u}, \bar{y}), \qquad D_1 \geq \lim_{n\to\infty} \rho_n^1(\bar{u}, \bar{v}), \qquad D_2 \geq \lim_{n\to\infty} \rho_n^2(\bar{u}, \bar{w})$$

On the other hand, for all pairs (R_1, R_2) that are not included in the ARR, there are no machines with the above property.

Let Q be a transition matrix which describes a channel where the input in the past u_1^n has been transformed into the output process y_1^n defined on the alphabet Y, into the output process v_1^n defined on the alphabet V and into the output process w_1^n defined on the alphabet W. That Q approximates the data compression algorithm at the level of a single symbol, even though, at the whole block level, the transformation from the input to the outputs is deterministic. Here and subsequently, the symbols i, j, m, and t represent letters from the alphabets U, V, W, and Y, respectively. The transition matrix represents the uncertainty with respect to each quadrature $(u_k; v_k; w_k; y_k)$ at some position k, without any knowledge about the past, i.e.

$$Q(j, m, t \mid i) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Pr(v_k = j, w_k = m, y_k = t \mid u_k = i) \tag{15}$$

Thus, we have

$$Q_1(j \mid i) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Pr(v_k = j \mid u_k = i) = \sum_m \sum_t Q(j, m, t \mid i)$$

$$Q_2(m \mid i) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Pr(w_k = m \mid u_k = i) = \sum_j \sum_t Q(j, m, t \mid i)$$

$$Q_0(t \mid i) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Pr(y_k = t \mid u_k = i) = \sum_j \sum_m Q(j, m, t \mid i)$$

In what follows, we assume that the system preserves stationarity and thus the conditional probability is constant in time.

We define the following set of transition matrices

$$\mathcal{Q}(p, D_0, D_1, D_2) = \left\{ Q \;\middle|\; D_0 \geq Ed_0(u, y),\; D_1 \geq Ed_1(u, v),\; D_2 \geq Ed_2(u, w) \right\}$$

Due to the ergodicity of the source and the stationarity of the system, following (4) and (9), it is almost sure that

$$\lim_{n\to\infty} \rho_n^0(\bar{u};\bar{y}) = \sum_i \sum_j \sum_m \sum_t p(i) Q_{j,m,t\,|\,i} d_0(i,t) = \rho^0(Q)$$

$$\lim_{n\to\infty} \rho_n^1(\bar{u};\bar{v}) = \sum_i \sum_j \sum_m \sum_t p(i) Q_{j,m,t\,|\,i} d_1(i,j) = \rho^1(Q)$$

$$\lim_{n\to\infty} \rho_n^2(\bar{u};\bar{w}) = \sum_i \sum_j \sum_m \sum_t p(i) Q_{j,m,t\,|\,i} d_2(i,m) = \rho^2(Q)$$

The three fidelity constraints yield, at the limit as the sequence length tends to infinity,

$$\lim_{n\to\infty} \rho_n^0(\bar{u};\bar{y}) = Ed_0(u,y) \le D_0$$

$$\lim_{n\to\infty} \rho_n^1(\bar{u};\bar{v}) = Ed_1(u,v) \le D_1$$

$$\lim_{n\to\infty} \rho_n^2(\bar{u};\bar{w}) = Ed_2(u,w) \le D_2$$

We construct, by the definition of the operational multi-rate multi-distortion function, a mapping from a process to processes where there is no randomness in the assignment of the time process v to the time process u. The determinism of the encoder-decoder pair (14) imposes an additional constraint that the conditional entropy of the process v, given the process u, is zero, i.e. $H(v\,|\,u) = 0$. The same reasoning yields $H(w\,|\,u) = 0$ and $H(y\,|\,u) = 0$.

The set of all machines that transform the input u to the outputs v, w, and y subject to the fidelity constraints is described by the intersection of the set $\mathcal{Q}$ and the set of all transition matrices that satisfy conditions (14). Let define that set as

$$\mathcal{P}(p, D_0, D_1, D_2) = \Big\{ Q \mid D_0 \ge Ed_0(u,y),\ \ D_1 \ge Ed_1(u,v),\ \ D_2 \ge Ed_2(u,w) \\ H(v\,|\,u) = 0,\quad H(w\,|\,u) = 0,\quad H(y\,|\,u) = 0 \Big\}$$

The rates are always greater than or equal to the entropy rates of the output processes v, w and y. The entropy rates of the outputs are determined by the transition matrix $Q \in \mathcal{P}$. Therefore the following inequalities hold:

$$\Big\{ H(v) \Big\} \le R_1 \tag{16}$$

$$\Big\{ H(w) \Big\} \le R_2 \tag{17}$$

$$\Big\{ H(y) \Big\} \le R_1 + R_2 \tag{18}$$

for some conditional probability distribution $Q \in \mathcal{P}$.

Since $H(v\,|\,u) = 0$, $H(y\,|\,u) = 0$ and $H(w\,|\,u) = 0$, it follows that $H(v) = I(u,v)$, $H(y) = I(u,y)$ and $H(w) = I(u,w)$, and the mutual information comes instead of the entropies. Thus, the region ARR is described by the following inequalities:

$$\Big\{ I(u,v) \Big\} \leq R_1 \tag{19}$$

$$\Big\{ I(u,w) \Big\} \leq R_2 \tag{20}$$

$$\Big\{ I(u,y) \Big\} \leq R_1 + R_2 \tag{21}$$

for some rational conditional probability distribution $Q \in \mathcal{Q}$. ■

Clearly, this bound is located between the old inner bound and the outer bound.

4. Summary

We have proved a coding bound for any stationary (time-invariant) encoding algorithms and for a stationary ergodic source. We have presented the idea of process assignment in a single- and multi-terminal system. Coupling the process definitions with the mathematical representation of process assignment provides a new and simple tool to analyse stationary multi-terminal systems and networks in general. We have solved an open problem which is the exact bound for multi-rate multi-distortion systems. The problem is also known as the degraded diversity system problem or the multiple-description system (MDS) problem. The equivalence between the Shannon-type bounds and the operational bounds, either in a single-link stationary system or the multi-terminal stationary system, was proved. The results are paving the way for solving difficult problems in multi-terminal information theory. In such cases block coding fails but other algorithms may attain the bounds as studied in this work.

Acknowledgement

The author is greatly indebted to Prof. Zeev Schuss and Prof. Benjy Weiss for their useful advice.

References

Ahlswede R. (1985): *The rate distortion region of a binary source for multiple descriptions without excess rate.* — IEEE Trans. Inform. Th., v.IT-31, pp.721–726.

Berger T. (1971): *Rate Distortion Theory: A Mathematical Basis for Data Compression.* — New York: Prentice-Hall.

Berger T. and Zhang Z. (1983): *Minimum breakdown degradation in binary source encoding.* — IEEE Trans. Inform. Th., v.IT-29, pp.807–814.

Blahut R.E. (1987): *Principles and Practice of Information Theory.* — Addison-Wesley Publishing Co.

Cheung K.M. and Tong K. (1993): *Proposed data compression schemes for the Galileo S-Band contingency mission.* — Proc. *Space and Earth Science Data Compression Workshop NASA, DCC' 93*, Snowbird, Utah.

Cover T.M. and El Gamal A. (1982): *Achievable rates for multiple descriptions.* — IEEE Trans. Inform. Th., v.IT-28, pp.851–857.

Gray R.M. (1975): *Sliding block source coding.* — IEEE Trans. Inform. Th., v.IT-21, pp.357–368.

Gray R.M., Neuhoff D.L. and Omura J.K. (1975): *Process definitions of distortion rate functions and source coding theorems.* — IEEE Trans. Inform. Th., v.IT-21, pp.524–532.

Omura J. (1973): *A coding theorem for discrete time sources.* — IEEE Trans. Inform. Th., v.IT-19, pp.490–498.

Ornstein D.S. and Shields P.C. (1990): *Universal almost sure data compression.* — The Annals of Probability, v.18, pp.441–452.

Zhang Z. and Berger T. (1987): *New results in binary multiple descriptions.* — IEEE Trans. Inform. Th., v.IT-33, pp.502–521.

Ziv J. (1972): *Coding of sources with unknown statistics- Part 1: Probability of encoding error; Part 2: Distortion relative to a fidelity criterion* . — IEEE Trans. Inform. Th., v.IT-18, pp.384–394.

Received: February 5, 1995
Revised: December 12, 1995