

Achievable Rates for Multiple Descriptions

ABBAS A. EL GAMAL, MEMBER, IEEE, AND THOMAS M. COVER, FELLOW, IEEE

Abstract—Consider a sequence of independent identically distributed (i.i.d.) random variables X_1, X_2, \dots, X_n and a distortion measure $d(X_i, \hat{X}_i)$ on the estimates \hat{X}_i of X_i . Two descriptions $i(X) \in \{1, 2, \dots, 2^{nR_1}\}$ and $j(X) \in \{1, 2, \dots, 2^{nR_2}\}$ are given of the sequence $X = (X_1, X_2, \dots, X_n)$. From these two descriptions, three estimates $\hat{X}(i(X))$, $\hat{X}(j(X))$, and $\hat{X}_0(i(X), j(X))$ are formed, with resulting expected distortions

$$E \frac{1}{n} \sum_{k=1}^n d(X_k, \hat{X}_{mk}) = D_m, \quad m = 0, 1, 2.$$

We find that the distortion constraints D_0, D_1, D_2 are achievable if there exists a probability mass distribution $p(x)p(\hat{x}_1, \hat{x}_2, \hat{x}_0 | x)$ with $E d(X, \hat{X}_m) \leq D_m$ such that

$$R_1 > I(X; \hat{X}_1),$$

$$R_2 > I(X; \hat{X}_2),$$

$$R_1 + R_2 > I(X; \hat{X}_1, \hat{X}_2, \hat{X}_0) + I(\hat{X}_1; \hat{X}_2),$$

where $I(\cdot)$ denotes Shannon mutual information. These rates are shown to be optimal for deterministic distortion measures.

I. INTRODUCTION

THE FOLLOWING problem of jointly good descriptions was posed by Gersho, Witsenhausen, Wolf, Wyner, Ziv, and Ozarow at the September 1979 IEEE Information Theory Workshop. Contributions to this problem can be found in Witsenhausen [5], Wolf, Wyner, and Ziv [6], Ozarow [7], and Witsenhausen and Wyner [8]. Suppose we wish to send a description of a stochastic process to a destination through a communication network. Assume that there is a chance that the description will be lost. Therefore we send two descriptions and hope that one of them will get through. Each description should be individually good. However, if both get through, then we wish the combined descriptive information to be as large as possible.

The difficulty is that good individual descriptions must be close to the process, by virtue of their goodness, and necessarily must be highly dependent. Thus the second description will contribute little extra information beyond one alone. On the other hand, two independent descriptions must be far apart and thus cannot in general be individually good.

The more general problem, stated precisely in the next section, is as follows. Consider a stochastic process X_1, X_2, \dots where the X_i 's are independent identically distributed (i.i.d.) according to some known distribution $p(x)$.

Manuscript received October 15, 1980; revised August 14, 1981. This work was supported in part under the Army Research Office, Grant DAAG29-79-C-0054, and it was supported in part by National Science Foundation Grant ENG78-23334 and Stanford Research Institute Contract DAHC-15-C-0187. This work was presented at the 1980 IEEE Symposium on Information Theory.

The authors are with Stanford University, Stanford, CA 94305.

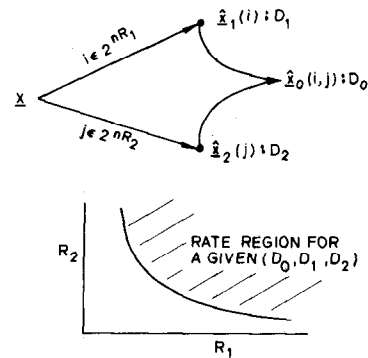


Fig. 1. Multiple descriptions.

Two individuals must describe X at respective rates R_1 and R_2 bits per transmission. Three single letter distortion measures d_1, d_2 , and d_0 are given. The question is, "What information should be sent at rates R_1 and R_2 so that a receiver given only description 1 can recover X with distortion D_1 , a receiver seeing only description 2 can recover X with distortion D_2 , and a receiver seeing both descriptions can recover X with distortion D_0 ?" For fixed distortions D_0, D_1 , and D_2 , what is the set of (R_1, R_2) necessary and sufficient to achieve these distortions?

In this paper, we shall exhibit an achievable rate region of (R_1, R_2) pairs as a function of the distortion vector $D = (D_0, D_1, D_2)$. See Fig. 1

The following examples consistent with Fig. 1 also motivate the search for a general theory of joint descriptions.

Communication Network: A communication network is used to send descriptions of X to New York with distortion D_1 at a cost c_1 dollars/bit, and to Boston with distortion D_2 at a cost c_2 dollars/bit. Given an acceptable distortion D_0 for the best estimate of x from the combined data base, minimize the cost $c = c_1 R_1 + c_2 R_2$.

Data Base: We wish to store the data $X = (X_1, \dots, X_n)$ with distortion D_1 in a New York computer with memory capacity nR_1 and store X with distortion D_2 in a Boston computer with memory capacity nR_2 . How should we do this so that when we combine the data we recover X with minimal distortion D_0 ?

Manager: A manager instructs two survey teams to gather information about X for their own use and for the subsequent use of the manager. What should he ask them to report?

II. DEFINITIONS

We shall first introduce the basic definitions of rate distortion theory and state Shannon's rate distortion theo-

rem. Then the definitions will be extended to multiple descriptions and their incurred distortions.

We assume that $X_i, i = 1, 2, \dots$ is a sequence of i.i.d. discrete random variables drawn according to a probability mass function $p(x)$. We are given a reconstruction space \hat{X} together with an associated distortion measure $d: \bar{X} \times \hat{X} \rightarrow R$. The distortion measure on n -sequences in $\bar{X}^n \times \hat{X}^n$ is defined by the average per symbol distortion

$$d(x, \hat{x}) = (1/n) \sum_{i=1}^n d(x_i, \hat{x}_i). \quad (2.1)$$

A *description* of $x \in \bar{X}^n$ is a map $i: \bar{X}^n \rightarrow \{1, 2, \dots, 2^{nR}\}$, where R is the *rate* of the description in bits per symbol of x . A *reconstruction* of X is a map $\hat{x}: \{1, 2, \dots, 2^{nR}\} \rightarrow \hat{X}^n$, and is said to incur *distortion* $D^{(n)}$ defined by

$$\begin{aligned} D^{(n)} &= Ed(X, \hat{x}(i(X))) \\ &= E(1/n) \sum_{k=1}^n d(X_k, \hat{x}_k(i(X))). \end{aligned} \quad (2.2)$$

The distortion D is said to be *achievable* with rate R for the stochastic process $\{X_i\}_{i=1}^\infty$ if for $n = 1, 2, \dots$, there exists sequence of rate R descriptions $i: \bar{X}^n \rightarrow \{1, 2, \dots, 2^{nR}\}$ and reconstructions $\hat{X}: \{1, 2, \dots, 2^{nR}\} \rightarrow \hat{X}^n$, such that $D^{(n)} \leq D$, for all n sufficiently large.

Let the rate distortion function $R(D)$ be the infimum of all rates R achieving distortion D on a given stochastic process $\{X_i\}$.

Theorem (Shannon [1]): If $X_i, i = 1, 2, \dots$, are i.i.d. discrete finite alphabet random variables with probability mass function $p(x)$, then

$$R(D) = \inf I(X; \hat{X}), \quad (2.3)$$

where

$$I(X; \hat{X}) = \sum_{x, \hat{x}} p(x, \hat{x}) \log(p(x, \hat{x})/p(x)p(\hat{x})), \quad (2.4)$$

and the infimum is taken over all joint probability mass functions $p(x)p(\hat{x}|x)$ such that

$$\sum_{x, \hat{x}} p(x, \hat{x})d(x, \hat{x}) \leq D. \quad (2.5)$$

Now we define the problem of multiple descriptions shown in Fig. 1. We are given three finite reconstruction spaces \hat{X}_0, \hat{X}_1 , and \hat{X}_2 , together with associated distortion measures

$$d_m: \bar{X} \times \hat{X}_m \rightarrow R, \quad m = 0, 1, 2.$$

The distortion measure on n -sequences in $\bar{X}^n \times \hat{X}_m^n$ is defined by the average per-letter distortion

$$d_m(x, \hat{x}_m) = (1/n) \sum_{i=1}^n d_m(x_i, \hat{x}_{mi}),$$

where $\hat{x}_m = (\hat{x}_{m1}, \hat{x}_{m2}, \dots, \hat{x}_{mn})$.

Definition: We shall say that (R_1, R_2) is an *achievable rate for distortion* $\mathbf{D} = (D_1, D_2, D_0)$ if there exists a sequence, indexed by n , of pairs of descriptions $i(x) \in \{1, 2, \dots, 2^{nR_1}\}, j(x) \in \{1, 2, \dots, 2^{nR_2}\}$, and reconstruction functions $\hat{x}_1(i), \hat{x}_2(j), \hat{x}_0(i, j)$ such that, for sufficiently

large n ,

$$E(1/n) \sum_{i=1}^n d_m(X_i, \hat{X}_{mi}) \leq D_m, \quad m = 0, 1, 2.$$

Definition: The *rate distortion region* $R(\mathbf{D})$ for distortion $\mathbf{D} = (D_1, D_2, D_0)$ is the closure of the set of achievable rate pairs (R_1, R_2) inducing distortion $\leq \mathbf{D}$. An *achievable rate region* is any subset of the rate distortion region.

In the next section, we shall exhibit an achievable rate region for the multiple description problem.

III. THEOREM AND CONSTRUCTION

The following achievable rate region for multiple descriptions has an information theoretic characterization.

Theorem 1: Let X_1, X_2, \dots be a sequence of i.i.d. finite alphabet random variables drawn according to a probability mass function $p(x)$. Let $d_i(\cdot, \cdot)$ be bounded. An achievable rate region for distortion $\mathbf{D} = (D_1, D_2, D_0)$ is given by the convex hull of all (R_1, R_2) such that

$$\begin{aligned} R_1 &> I(X; \hat{X}_1) \\ R_2 &> I(X; \hat{X}_2) \\ R_1 + R_2 &> I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_1; \hat{X}_2) \end{aligned} \quad (3.1)$$

for some probability mass function $p(\hat{x}, \hat{x}_0, \hat{x}_1, \hat{x}_2) = p(x)p(\hat{x}_0, \hat{x}_1, \hat{x}_2|x)$ such that

$$\begin{aligned} D_1 &\geq Ed_1(X; \hat{X}_1) \\ D_2 &\geq Ed_2(X; \hat{X}_2) \\ D_0 &\geq Ed_0(X; \hat{X}_0). \end{aligned} \quad (3.2)$$

Remark on convexification of region in Theorem 1: Since we have been unable to prove in general that the region in Theorem 1 is convex, it may be necessary to convexify this region by time-sharing. For each choice of $p = p(\hat{x}_0, \hat{x}_1, \hat{x}_2|x)$, let

$$\begin{aligned} V(p) &= (I(X; \hat{X}_1), I(X; \hat{X}_2), \\ &I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_1; \hat{X}_2), \\ &Ed_1(X, \hat{X}_1), Ed_2(X, \hat{X}_2), Ed_0(X, \hat{X}_0)). \end{aligned} \quad (3.3)$$

Let \mathcal{V} be the convex hull of $\cup V(p)$. (An arbitrary point in the convex hull can be obtained by mixing at most seven $V(p)$'s.) Let $(\mathbf{R}, \mathbf{D}) = (R_1, R_2, R_1 + R_2, D_1, D_2, D_0)$. Then it follows that (\mathbf{R}, \mathbf{D}) is achievable if there exists a $V \in \mathcal{V}$ such that

$$(\mathbf{R}, \mathbf{D}) > V. \quad (3.4)$$

Remark about method of proof: Let $p(x)$ be given, and fix a choice of $p(\hat{x}_1, \hat{x}_2, \hat{x}_0|x)$. Sequences x and \hat{x}_0 are said to be jointly typical if the empirical joint composition of (x, \hat{x}_0) is approximately equal to $p(x, \hat{x}_0)$. A subset B_0 of \hat{X}_0^n is said to *cover* \bar{X}^n if, with high probability, a randomly drawn $X \in \bar{X}^n$ is jointly typical with some $\hat{x}_0 \in B_0$. It can be shown that approximately $2^{nI(X; \hat{X}_0)}$ \hat{x}_0 's are both necessary and sufficient to cover \bar{X}^n . Moreover, if \hat{x}_0

is jointly typical with x , then the joint composition $p(x, \hat{x}_0)$ guarantees that $d_0(x, \hat{x}_0)$ will be close to D_0 .

Two extra ingredients in addition to the notion of covering are found in the proof of Theorem 1. The distribution $p(x, \hat{x}_1, \hat{x}_2, \hat{x}_0)$ is used to find a covering $\{(\hat{x}_1, \hat{x}_2)\}$ of \bar{X}^n so that the \hat{x}_1 's individually cover \bar{X}^n , the \hat{x}_2 's individually cover \bar{X}^n , and the (\hat{x}_1, \hat{x}_2) 's jointly cover \bar{X}^n . Finding the best rate region for the product covering involves a new result on lists. Second, for each (\hat{x}_1, \hat{x}_2) , a minimal conditional covering of \bar{X}^n is found. The conditional \hat{x}_0 information is to be distributed to the two describers and increases their rates beyond that which is necessary to individually recover \hat{x}_1 and \hat{x}_2 at distortions D_1 and D_2 . However, the extra information efficiently adds information to the two descriptions to yield a third description \hat{x}_0 with distortion D_0 .

Before proceeding to the proof in Section V, we give an example in the next section.

IV. EXAMPLE

Let X_1, X_2, \dots be independent identically distributed normal random variables with mean 0 and variance 1. From rate distortion theory [1], [2], we know that $R(D) = (1/2) \log 1/D$, is the rate distortion function for a squared error distortion

$$d(x, \hat{x}) = (x - \hat{x})^2. \tag{4.1}$$

Thus, $(n/2) \log 1/D$ bits are necessary and sufficient to describe a random sequence (X_1, X_2, \dots, X_n) with expected squared error

$$E \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2 \leq D. \tag{4.2}$$

$$\begin{aligned} R_1 &\geq (1/2) \log 1/D_1, \\ R_2 &\geq (1/2) \log 1/D_2, \\ R_1 + R_2 &\geq (1/2) \log 1/D_0 \\ &+ (1/2) \log \frac{(1 - D_0)^2}{(1 - D_0)^2 - ((1 - D_1)^{1/2}(1 - D_2)^{1/2} - (D_1 - D_0)^{1/2}(D_2 - D_0)^{1/2})^2}. \end{aligned} \tag{4.7}$$

Consequently, for the simultaneous description problem, an obvious inner bound on the rate region is

$$\begin{aligned} R_1 &\geq (1/2) \log 1/D_1 \\ R_2 &\geq (1/2) \log 1/D_2 \\ R_1 + R_2 &\geq (1/2) \log 1/D_0. \end{aligned} \tag{4.3}$$

The third inequality follows from the realization that the description rate is $R_1 + R_2$ for the \hat{X}_0 reconstruction. Surprisingly, for high distortions ($D_1 + D_2 - D_0 \geq 1$) these rates are actually achievable. The detailed calculations have been provided to the authors by M. Aref. We examine the region of Theorem 1¹ for the joint normal distribution on

$(\hat{X}, \hat{X}_1, \hat{X}_2, \hat{X}_0)$ obtained from $X \sim N(0, 1)$,

$$\begin{aligned} \hat{X}_1 &= \alpha_1 \hat{X}_0 + Z_1 \\ \hat{X}_2 &= \alpha_2 \hat{X}_0 + Z_2 \\ \hat{X}_0 &= \alpha_0 X + Z_0, \end{aligned} \tag{4.4}$$

where (Z_0, Z_1, Z_2) are jointly normal with mean $\mathbf{0}$ and covariance matrix

$$K = \begin{bmatrix} k_0 & 0 & 0 \\ 0 & k_1 & k_{12} \\ 0 & k_{12} & k_2 \end{bmatrix}, \tag{4.5}$$

and (Z_0, Z_1, Z_2) is independent of X . Optimizing (3.1) in Theorem 1 over $\alpha_0, \alpha_1, \alpha_2, K$ subject to the distortion constraints

$$\begin{aligned} E(X - \hat{X}_1)^2 &= D_1, \\ E(X - \hat{X}_2)^2 &= D_2, \\ E(X - \hat{X}_0)^2 &= D_0, \end{aligned}$$

yields the region described below.

Case 1: (High distortion) $D_1 + D_2 - D_0 \geq 1$.

The following region is achievable:

$$\begin{aligned} R_1 &> (1/2) \log 1/D_1, \\ R_2 &> (1/2) \log 1/D_2, \\ R_1 + R_2 &> (1/2) \log 1/D_0. \end{aligned} \tag{4.6}$$

This is trivially optimal by our previous remarks.

Case 2: (Low distortion) $D_1 + D_2 - D_0 < 1$.

The following region is achievable:

Apparently, for low distortions, the consequent dependence of the descriptions causes an increase in the total description rate $R_1 + R_2$ beyond the $(1/2) \log 1/D_0$ necessary for independent descriptions.

This region has recently been shown to be optimal by Ozarow [7].

V. PROOF OF THEOREM 1

Throughout this proof we shall assume that \bar{X} and the reconstruction spaces \bar{X}_i are finite sets.

Before proceeding with the proof, we shall define the idea of joint typicality.

Let $\{X^{(1)}, X^{(2)}, \dots, X^{(k)}\}$ denote a finite collection of discrete random variables with some fixed joint distribution $p(x^{(1)}, x^{(2)}, \dots, x^{(k)})$, $(x^{(1)}, x^{(2)}, \dots, x^{(k)}) \in \mathcal{X}^{(1)} \times$

¹It should be possible to extend the proof of Theorem 1 to the Gaussian source, using the techniques in Gallager [4].

$\mathcal{X}^{(2)} \times \cdots \times \mathcal{X}^{(k)}$. Let S denote an ordered subset of these random variables, and consider n independent copies of S . Thus,

$$\Pr \{S = s\} = \prod_{i=1}^n P \{S_i = s_i\}, \quad s \in \mathcal{S}^n. \quad (5.1)$$

For $s \in \mathcal{S}$, and $s \in \mathcal{S}^n$, let $N(s; s)$ be the number of indices $i \in \{1, 2, \dots, n\}$ such that $s_i = s$. By the law of large numbers, for any subset S of random variables and for all $s \in \mathcal{S}$,

$$(1/n)N(s; s) \rightarrow p(s). \quad (5.2)$$

Also,

$$-\frac{1}{n} \log p(S_1, S_2, \dots, S_n) = -\frac{1}{n} \sum_{i=1}^n \log p(S_i) \rightarrow H(S). \quad (5.3)$$

Convergence in (5.2) and (5.3) take place simultaneously with probability one for all 2^k subsets

$$S \subseteq \{X^{(1)}, X^{(2)}, \dots, X^{(k)}\}.$$

Definition: The set T_ϵ of strongly ϵ -typical n -sequences $(x^{(1)}, x^{(2)}, \dots, x^{(k)})$ is defined by

$$\begin{aligned} T_\epsilon(X^{(1)}, X^{(2)}, \dots, X^{(k)}) \\ = \left\{ (x^{(1)}, x^{(2)}, \dots, x^{(k)}) : \left| \frac{1}{n} N(s; s) - p(s) \right| < \epsilon, \forall s \in \mathcal{S}, \right. \\ \left. \forall S \subseteq \{X^{(1)}, X^{(2)}, \dots, X^{(k)}\} \right\}. \quad (5.4) \end{aligned}$$

Let $T_\epsilon(S)$ denote the restriction of T_ϵ to the coordinates corresponding to S . Let $\|A\|$ denote the cardinality of the set A . We need the following well-known properties of T_ϵ . There exists a $\delta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$, such that for sufficiently large n , and any $\epsilon \rightarrow 0$,

$$P \{T_\epsilon(S)\} \geq 1 - \epsilon, \quad \forall S \subseteq \{X^{(1)}, \dots, X^{(k)}\}, \quad (5.5)$$

$$2^{n(H(S)-\delta)} \leq \|T_\epsilon(S)\| \leq 2^{n(H(S)+\delta)} \quad (5.6)$$

$$2^{-n(H(S)+\delta)} \leq p(s) \leq 2^{-n(H(S)-\delta)}, \quad \text{for all } s \in T_\epsilon(S), \quad (5.7)$$

If $s_2 \in T_\epsilon(S_2)$, then the number of jointly ϵ -typical s_1 's, denoted by $\|T_\epsilon(S_1 | s_2)\|$, is bounded by

$$2^{n(H(S_1|S_2)-\delta)} \leq \|T_\epsilon(S_1 | s_2)\| \leq 2^{n(H(S_1|S_2)+\delta)}. \quad (5.8)$$

Proof of Theorem: We shall choose a net of \hat{x}_1 sequences covering \bar{X}^n with distortion D_1 , a net of \hat{x}_2 sequences covering \bar{X}^n with distortion D_2 , and a conditional net of \hat{x}_0 sequences (given (\hat{x}_1, \hat{x}_2)) covering \bar{X}^n with distortion D_0 . A random choice of the nets will suffice. Fix a joint probability mass function of the form $p(x)p(\hat{x}_1, \hat{x}_2, \hat{x}_0 | x)$ on $\bar{X} \times \bar{X}_1 \times \bar{X}_2 \times \bar{X}_0$ such that

$$\begin{aligned} Ed_1(X, \hat{X}_1) &< D_1 \\ Ed_2(X, \hat{X}_2) &< D_2 \\ Ed_0(X, \hat{X}_0) &< D_0. \end{aligned} \quad (5.9)$$

Choose real numbers $R'_1, R'_2, \Delta \geq 0$.

Random Coding: Let n -vectors $\hat{X}_1(1), \hat{X}_1(2), \dots, \hat{X}_1(2^{nR'_1})$ be drawn independently according to a uniform distribution over the set $T_\epsilon(\hat{X}_1)$ of ϵ -typical \hat{X}_1 n -vectors. That is, $P\{\hat{X}_1(i) = \hat{x}_1\} = 1/\|T_\epsilon(\hat{X}_1)\|$, if $\hat{x}_1 \in T_\epsilon$, and $= 0$ otherwise.

Similarly, let $2^{nR'_2}$ n -vectors $\hat{X}_2(1), \hat{X}_2(2), \dots, \hat{X}_2(2^{nR'_2})$ be drawn i.i.d. according to a uniform distribution over $T_\epsilon(\hat{X}_2)$.

Finally, for every jointly typical $(\hat{x}_1(i), \hat{x}_2(j))$ in the above list, let $\hat{X}_0(i, j, 1), \hat{X}_0(i, j, 2), \dots, \hat{X}_0(i, j, 2^{n\Delta})$, be drawn i.i.d. according to a uniform distribution over the set $T_\epsilon(\hat{X}_0 | \hat{x}_1(i), \hat{x}_2(j))$ of conditionally ϵ -typical \hat{x}_0 's, conditioned on $(\hat{x}_1(i), \hat{x}_2(j))$.

Encoding: Given an $x \in \bar{X}^n$, find, if possible, a triple (i, j, k) such that $(x, \hat{x}_1(i), \hat{x}_2(j), \hat{x}_0(i, j, k))$ is in the set T_ϵ of all jointly typical sequences. If no such (i, j, k) exists, simply set $(i, j, k) = (0, 0, 0)$.

We now divide the description of $k \in 2^{n\Delta}$ into two parts, $k = (k_1, k_2)$, $k_1 \in 2^{n\Delta_1}$, $k_2 \in 2^{n\Delta_2}$, $\Delta_1 \geq 0$, $\Delta_2 \geq 0$, $\Delta_1 + \Delta_2 = \Delta$. We use the obvious notation $2^{n\Delta}$ for the set $\{1, 2, \dots, [2^{n\Delta}]\}$, where $[t]$ denotes the greatest integer less than or equal to t . We send (i, k_1) to decoder 1, (j, k_2) to decoder 2, and consequently send (i, j, k) to decoder 0. The resulting rates of transmission become

$$\begin{aligned} R_1 &= R'_1 + \Delta_1, \\ R_2 &= R'_2 + \Delta_2. \end{aligned} \quad (5.10)$$

We shall prove that a successful encoding of X achieving distortions $\leq (D_1, D_2, D_0)$ is accomplished (with high probability by this random net) provided n is sufficiently large and

$$R'_1 > I(X; \hat{X}_1),$$

$$R'_2 > I(X; \hat{X}_2),$$

$$R'_1 + R'_2 - I(\hat{X}_1; \hat{X}_2) > I(X; \hat{X}_1, \hat{X}_2),$$

$$\Delta_1 \geq 0,$$

$$\Delta_2 \geq 0,$$

$$\Delta_1 + \Delta_2 > I(X; \hat{X}_0 | \hat{X}_1, \hat{X}_2). \quad (5.11)$$

Theorem 1 then follows from adding these inequalities with the result

$$R_1 = R'_1 + \Delta_1 \geq I(X; \hat{X}_1),$$

$$R_2 = R'_2 + \Delta_2 \geq I(X; \hat{X}_2),$$

$$R_1 + R_2 = R'_1 + R'_2 + \Delta_1 + \Delta_2$$

$$> I(X; \hat{X}_1, \hat{X}_2) + I(X; \hat{X}_0 | \hat{X}_1, \hat{X}_2) + I(\hat{X}_1, \hat{X}_2)$$

$$= I(X; \hat{X}_1, \hat{X}_2, \hat{X}_0) + I(\hat{X}_1; \hat{X}_2), \quad (5.12)$$

where the last equality is an information theoretic identity. These are the conditions of Theorem 1.

Reconstruction: Decoder 1, given (i, k_1) , announces $\hat{x}_1(i)$ as his reconstruction of x . Similarly decoder 2, given (j, k_2) , announces $\hat{x}_2(j)$. And decoder 0, given (i, j, k) , where $k = (k_1, k_2)$, announces $\hat{x}_0(i, j, k)$. These reconstructions will be shown to meet the distortion constraints.

Distortion: We first note that if (i, j, k) can be found satisfying the encoding step, then the distortions satisfy

$d_m(x, \hat{x}_m) \leq D_m$, $m = 0, 1, 2$, by definition of T_ϵ . Let us denote by E the error event that there does not exist (i, j, k) such that $(X, \hat{X}_1(i), \hat{X}_2(j), \hat{X}_0(i, j, k)) \in T_\epsilon$. We wish to show $P(E) \rightarrow 0$, as $n \rightarrow \infty$.

Probability of Error: An error E will occur if one or more of the following events occurs:

$$\begin{aligned} E_0: & X \notin T_\epsilon, \\ E_1: & (X, \hat{X}_1(i)) \notin T_\epsilon, \quad \text{for all } i \in 2^{nR_1}, \\ E_2: & (X, \hat{X}_2(j)) \notin T_\epsilon, \quad \text{for all } j \in 2^{nR_2}, \\ E_3: & (X, \hat{X}_1(i), \hat{X}_2(j)) \notin T_\epsilon, \\ & \text{for all } (i, j) \in 2^{nR_1} \times 2^{nR_2}, \\ E_4: & \exists (i, j) \in 2^{nR_1} \times 2^{nR_2}, \\ & \text{such that } (X, \hat{X}_1(i), \hat{X}_2(j)) \in T_\epsilon, \\ & \text{but } (X, \hat{X}_1(i), \hat{X}_2(j), \hat{X}_0(i, j, k)) \notin T_\epsilon, \\ & \text{for } k \in 2^{n\Delta}. \end{aligned}$$

Thus $E = \cup_{i=0}^4 E_i$ and the probability of (encoding) error is bounded above by

$$P(E) \leq P(E_0) + \sum_{i=1}^4 P(E_i \cap E_0^c). \quad (5.13)$$

Clearly $P(E_0) \rightarrow 0$, as $n \rightarrow \infty$. Also, it follows from known results in rate distortion theory that

$$\begin{aligned} R_1' > I(X; \hat{X}_1) & \text{ implies } P(E_1 \cap E_0^c) \rightarrow 0; \\ R_2' > I(X; \hat{X}_2) & \text{ implies } P(E_2 \cap E_0^c) \rightarrow 0; \\ \Delta > I(X; \hat{X}_0 | \hat{X}_1, \hat{X}_2) & \text{ implies } P(E_4 \cap E_0^c) \rightarrow 0. \end{aligned} \quad (5.14)$$

It remains to show that $P(E_3 \cap E_0^c) \rightarrow 0$, i.e., we wish to show that there are enough individually typical $\hat{X}_1(i), \hat{X}_2(j)$ so that we can find at least one jointly typical pair. Thus define, for each $x \in T_\epsilon$, the random set

$$C_x = \{(i, j) \in 2^{nR_1} \times 2^{nR_2}; (\hat{X}_1(i), \hat{X}_2(j)) \in T_\epsilon(\hat{X}_1, \hat{X}_2 | x)\}, \quad (5.15)$$

consisting of those (i, j) pairs for which the random code assignment yields $(x, \hat{X}_1(i), \hat{X}_2(j))$ jointly ϵ -typical. It is easily seen that

$$\begin{aligned} P(E_3 \cap E_0^c) &= P\{\|C_x\| = 0, X \in T_\epsilon\} \\ &\leq \max_{x \in T_\epsilon} P\{\|C_x\| = 0\}. \end{aligned} \quad (5.16)$$

But, for all $x \in T_\epsilon$, $0 < \alpha < 1$, we have

$$\begin{aligned} P\{\|C_x\| = 0\} &\leq P\{|\|C_x\| - E\|C_x\|| \geq \alpha E\|C_x\|\} \\ &\leq \text{var}\|C_x\| / \alpha^2 (E\|C_x\|)^2. \end{aligned} \quad (5.17)$$

Writing $\|C_x\|$ as the sum of indicator functions, it follows from arguments similar to the proof in [3] that

$$E\|C_x\| \geq 2^{n(R_1 + R_2 - H(\hat{X}_1) - H(\hat{X}_2) + H(\hat{X}_1, \hat{X}_2 | X) - o(1))} \quad (5.18)$$

and

$$\text{var}\|C_x\| \leq 2^{n(R_1 + R_2 - H(\hat{X}_1) - H(\hat{X}_2) + H(\hat{X}_1, \hat{X}_2 | X) + o(1))}. \quad (5.19)$$

Therefore, $P(E_3 \cap E_0^c) \rightarrow 0$ if

$$\begin{aligned} R_1' + R_2' &> H(\hat{X}_1) + H(\hat{X}_2) - H(\hat{X}_1, \hat{X}_2 | X) \\ &= I(X; \hat{X}_1, \hat{X}_2) + I(\hat{X}_1; \hat{X}_2). \end{aligned} \quad (5.20)$$

It remains to argue that the expected distortion is asymptotically unaffected when an encoding description (i, j, k) cannot be found, i.e., when the event E occurs. But E^c implies $\lim_{n \rightarrow \infty} E\{d | E^c\} \leq D + \epsilon \mathbf{1}$, by construction of T_ϵ . Finally,

$$Ed \leq P(E^c)(D + \epsilon \mathbf{1}) + p(E)\mathbf{1}d_{\max}, \quad (5.21)$$

where $\mathbf{1} = (1, 1, 1)$ and

$$d_{\max} = \max_{\substack{x \in \bar{X} \\ \hat{x}_m \in \hat{X}_m, m \in \{0, 1, 2\}}} d_m(x, \hat{x}_m).$$

Here we assume $d_{\max} < \infty$. Finally, $P(E) \rightarrow 0$ and $\epsilon \rightarrow 0$ imply that $\lim_{n \rightarrow \infty} Ed \leq D$, and the theorem is proved.

VI. FUNCTION DISTORTION MEASURES

Suppose that receivers 0, 1, and 2 wish only to know $f_0(x), f_1(x)$, and $f_2(x)$ respectively. Here we have a simpler rate region and a converse.

Theorem 2: Let $\hat{X}_0 = f_0(X)$, $\hat{X}_1 = f_1(X)$, and $\hat{X}_2 = f_2(X)$. Then (R_1, R_2) allows recovery of f_1 by 1, f_2 by 2, and f_0 by 0 if and only if

$$R_1 > I(X; \hat{X}_1) = H(f_1(X)), \quad (6.1)$$

$$R_2 > I(X; \hat{X}_2) = H(f_2(X)), \quad (6.2)$$

$$\begin{aligned} R_1 + R_2 &> I(f_1(X), f_2(X), f_0(X); X) \\ &\quad + I(f_1(X); f_2(X)) \\ &= H(f_1(X)) + H(f_2(X)) \\ &\quad + H(f_0(X) | f_1(X), f_2(X)). \end{aligned} \quad (6.3)$$

Proof: Achievability is immediate from Theorem 1.

For the converse, let there be maps

$$\begin{aligned} i: \bar{X}^n &\rightarrow 2^{nR_1}, \\ j: \bar{X}^n &\rightarrow 2^{nR_2}, \end{aligned}$$

and set

$$I = i(X), \quad J = j(X).$$

Also given are reconstruction functions $\hat{x}_0(\cdot), \hat{x}_1(\cdot), \hat{x}_2(\cdot)$, with the property

$$\begin{aligned} P(\hat{x}_0(I(X), J(X)) \neq f_0(X)) &\leq \epsilon, \\ P(\hat{x}_1(I(X)) \neq f_1(X)) &\leq \epsilon, \\ P(\hat{x}_2(J(X)) \neq f_2(X)) &\leq \epsilon, \end{aligned} \quad (6.4)$$

where

$$f_i(X) = (f_i(X_1), \dots, f_i(X_n)), \quad i = 0, 1, 2.$$

We shall show that the rates R_1, R_2 must satisfy (6.1), (6.2), and (6.3).

Inequalities (6.1) and (6.2) follow from Shannon. To prove (6.3) we first note that

$$\begin{aligned} nR_1 + nR_2 &\geq H(I) + H(J) \\ &= H(I, f_1(X)) + H(J, f_2(X)) \\ &\quad - H(f_1(X) | I) - H(f_2(X) | J). \end{aligned}$$

It follows from (6.4) and Fano's inequality that

$$\begin{aligned} H(f_1(X) | I) &\leq (H(I)\epsilon + h(\epsilon)) \leq n\delta_n, \\ H(f_2(X) | J) &\leq (H(J)\epsilon + h(\epsilon)) \leq n\delta_n, \end{aligned}$$

where $\delta_n \rightarrow 0$ as $\epsilon \rightarrow 0$. Therefore

$$\begin{aligned} nR_1 + nR_2 &\geq H(I, f_1(X)) + H(J, f_2(X)) - 2n\delta_n \\ &= nH(f_1(X)) + nH(f_2(X)) + H(I|f_1(X)) \\ &\quad + H(J|f_2(X)) - 2n\delta_n \\ &\geq nH(f_1(X)) + nH(f_2(X)) \\ &\quad + H(I, J, f_0(X) | f_1(X), f_2(X)) \\ &\quad - H(f_0(X) | I, J, f_1(X), f_2(X)) - 2n\delta_n. \end{aligned}$$

Again, by (6.4) and Fano's inequality, we have

$$H(f_0(X) | I, J) \leq n\delta_n.$$

Thus

$$\begin{aligned} nR_1 + nR_2 &\geq nH(f_1(X)) + nH(f_2(X)) \\ &\quad + nH(f_0(X) | f_1(X), f_2(X)) \\ &\quad + H(I, J | f_1(X), f_2(X), f_0(X)) - 3n\delta_n. \end{aligned} \tag{6.5}$$

Finally, by dropping the fourth term in (6.5), inequality (6.3) follows.

VII. ROBUST DESCRIPTIONS

We shall describe a problem concerning robust descriptions that seems at first to be unrelated to Theorem 1.

Suppose that we wish to describe a source $x \in \bar{X}^n$ by an index $i \in 2^{nR}$ in such a manner that the description $i(x)$ is good simultaneously for several given distortion measures. (See Fig. 2) For example, we may wish to describe a Gaussian source so that it can be recovered with low mean squared error distortion and low absolute error distortion. Or, for a Bernoulli source X , we might wish \hat{X}_1 to agree with X on the zeros of X , and for \hat{X}_2 to agree with X on the ones of X . Incidentally, this does not guarantee that X can be recovered perfectly from \hat{X}_1 and \hat{X}_2 .

A more precise specification of the problem is as follows. Given is a source $\{X_i\}_{i=1}^\infty$ where X_1, X_2, \dots are i.i.d. random variables drawn according to a probability mass function $p(x)$, $x \in \bar{X}$. Also given are m distortion mass functions $d_i(x, \hat{x}_i)$ defined on $\bar{X} \times \hat{X}_i$, $i = 1, 2, \dots, m$. Distortions $\mathbf{D} = (D_1, D_2, \dots, D_m)$ are to be achieved. What rate $R(\mathbf{D})$ is necessary? The solution to this problem is given in the following theorem. A version of this theorem can be found in Berger [2].

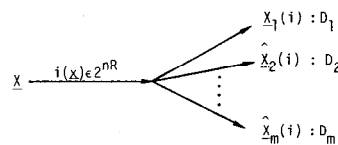


Fig. 2. Robust descriptions.

Theorem 2:

$$R(\mathbf{D}) = \min I(X; \hat{X}_1, \hat{X}_2, \dots, \hat{X}_m), \tag{7.1}$$

where the minimum is over all $p(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m | x)$ satisfying

$$Ed_i(X; \hat{X}_i) \leq D_i, \quad i = 1, 2, \dots, m. \tag{7.2}$$

Remark: As mentioned before, it appears that a description that can be used to make no mistakes on the 0's of X and can also be used to make no mistakes on the 1's of X must necessarily make no mistakes whatsoever. This would require rate $R = 1$, but since $R(D_1, D_2) < 1$, for all $D_1, D_2 < 1/2$, we see that the above idea is mistaken.

Remark: For $m = 2$, Theorem 2 is a special case of Theorem 1 obtained by constraining $R_2 = 0$.

We give some examples before proceeding with the proof for general m .

Example: Let X be Bernoulli with parameter $1/2$. Let d_1 and d_2 be given by

$$\begin{array}{c} \hat{x}_1 \\ x \begin{array}{|c|cc} \hline & 0 & 1 \\ \hline 0 & 0 & \infty \\ \hline 1 & 1 & 0 \\ \hline \end{array} \end{array} \quad \begin{array}{c} \hat{x}_2 \\ x \begin{array}{|c|cc} \hline & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline 1 & \infty & 0 \\ \hline \end{array} \end{array} \tag{7.3}$$

Thus d_1 allows no errors on the 0's of x , and d_2 allows no errors on the 1's of x . The optimizing distribution $p(x_1, x_2 | x)$ is given by inspection of Fig. 3. The resulting rate is

$$\begin{aligned} R(D_1, D_2) &= I(X; \hat{X}_1, \hat{X}_2) \\ &= h(D_1/2, D_2/2, (D_1 + D_2)/2) \\ &\quad - (1/2)h(D_1) - (1/2)h(D_2). \end{aligned} \tag{7.4}$$

Remark: It may happen that all distributions satisfying the distortion constraint (7.2) correspond to a Markov chain $X \rightarrow \hat{X}_1 \rightarrow \hat{X}_2$ in that order. In such a case, we have

$$I(X; \hat{X}_1, \hat{X}_2) = I(X; \hat{X}_1). \tag{7.5}$$

Thus, only the first distortion constraint is active, and a good description for 1 and 2 is given by the best description for 1 alone.

Proof of Theorem 2: Achievability follows immediately once the encoding is specified. Fix a choice of $p(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m | x)$ satisfying distortion constraints (7.2). This induces a distribution $p(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m)$. Choose 2^{nR} vectors $\hat{X} = (\hat{X}_1, \hat{X}_2, \dots, \hat{X}_m)$ independently drawn according to $\prod_{i=1}^m p(\hat{x}_{1i}, \hat{x}_{2i}, \dots, \hat{x}_{mi})$. Designate these vectors $\hat{X}(k)$, $k \in 2^{nR}$. By the usual arguments, these vectors will

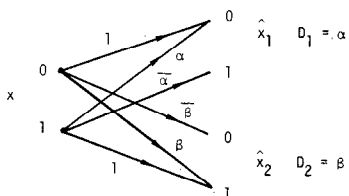


Fig. 3. Joint distribution for robust description.

“cover” \bar{X} ” if

$$R > I(X; \hat{X}_1, \hat{X}_2, \dots, \hat{X}_m). \quad (7.6)$$

Thus for large n , with probability arbitrarily close to 1, there will be an index $k \in 2^{nR}$ such that

$$d_i(X, \hat{X}_i(k)) \leq D_i, \quad i = 1, 2, \dots, m.$$

The converse is precisely along the lines of the converse for Shannon rate distortion theory.

VIII. CONCLUSION

It is a pleasant surprise that two descriptions are “twice as good” as one when the two descriptions need not be individually very good. This conclusion holds (for high distortions) for the Gaussian example in Section IV and also holds for the Hamming distortion on Bernoulli sequences, an example not given here. Of course, when we demand low distortions for the individual descriptions, the peculiar tension of this problem reveals itself. If the distor-

tion constraints are sufficiently severe, two descriptions are no better than one because they are in fact the same description.

ACKNOWLEDGMENT

Beyond our obvious debt to the formulators of this problem, we wish to thank M. Aref for working out the Gaussian example and C. Heegard for working out the Bernoulli case with Hamming distortion.

REFERENCES

- [1] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” *IRE Nat. Conv. Rec.*, Mar. 1959; reprinted in *Key Papers in the Development of Information Theory*. New York: IEEE.
- [2] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [3] A. El Gamal and E. van der Meulen, “A proof of Marton’s coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inform. Theory*, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.
- [4] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.
- [5] H. Witsenhausen, “On source networks with minimal breakdown degradation,” *Bell Syst. Tech. J.*, vol. 59, no. 6, pp. 1083–1087, July–Aug. 1980.
- [6] J. Wolf, A. Wyner, and J. Ziv, “Source coding for multiple descriptions,” *Bell Syst. Tech. J.*, vol. 59, no. 8, pp. 1417–1426, Oct. 1980.
- [7] L. Ozarow, “On a source coding problem with two channels and three receivers,” *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.
- [8] H. S. Witsenhausen and A. D. Wyner, “Source coding for multiple descriptions II: A binary source,” Bell Lab. Tech. Rep. TM-80-1217, Dec. 1980.