AD-753 260

CONDITIONAL RATE-DISTORTION THEORY

Robert M. Gray

Stanford University

STANFORD UNIVERSITY

CENTER FOR SYSTEMS RESEARCH

Conditional Rate-Distortion Theory

by

Robert M. Gray

October 1972

Information Systems Laboratory

DOCUMENT CONTROL. DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Stanford Electronics Laboratories Stanford University, Stanford, California	Unclassified **2b. GROUP**

3. REPORT TITLE

CONDITIONAL RATE-DISTORTION THEORY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report No. 6502-2

5. AUTHOR(S) (First name, middle initial, last name)

Robert M. Gray

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1972	~~17~~ 21	8

8a. CONTRACT OR GRANT NO N00014-67-A-0112-0044 & NSF Grant b. PROJECT NO GK -31630 6502 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) SU-SEL-72-047 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) Technical Report No. 6502-2

10. DISTRIBUTION STATEMENT This document has been approved for public release & sale; its distribution is unlimited.
Reproduction in whole or in part is permitted for any purpose of the United States Government.
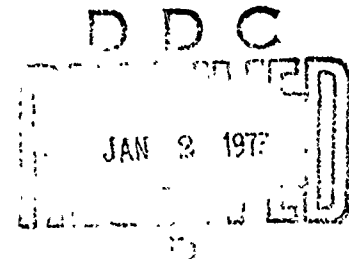
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Joint Services Electronics Program (U.S. Army, U.S. Navy & &.S. Air Force) and the National Science Foundation

13. ABSTRACT

The basic definitions, coding theorems, and properties of joint, marginal, and conditional rate-distortion functions are presented.

-i-

DD FORM 1473
, NOV 65

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
RATE-DISTORTION THEORY						
INFORMATION THEORY						
SOURCE CODING THEOREMS						
COMPOSITE SOURCES						

- ii -

CONDITIONAL RATE-DISTORTION THEORY

by

Robert M. Gray

October 1972

DDC

JAN 2 1973

-iii-

Information Systems Laboratory
Stanford Electronics Laboratories
Stanford University   Stanford, California

Introduction

The conditional rate-distortion function has proved useful in source coding problems involving the possession of side information such as in simple networks and sources with memory. Many of the basic properties of conditional rates, however, are lengthy but relatively straightforward extensions of the usual theory. Hence, these results have not appeared in the papers concerned with the applications of conditional rates. The purpose of this report is to present these basic definitions, coding theorems and bounds so as to provide a complete background reference for the journal papers on composite bounds for data compression performance attainable with sources with memory [8] and on source coding for simple networks [5].

Definitions and Preliminaries

Joint Rates

The <u>joint rate-distortion function</u> of a vector source $\underset{\sim}{U} = (U_1, U_2, \ldots, U_n)$ and a vector-valued distortion measure $\underset{\sim}{D}(\underset{\sim}{u}, \underset{\sim}{\hat{u}}) = \{D_k(\underset{\sim}{\hat{u}}, \underset{\sim}{u}), k=1, \ldots, m\}$ defined on $A(\underset{\sim}{U}) \times \Omega(\underset{\sim}{\hat{U}})$ is defined as follows:

$$R_{\underset{\sim}{U}}(\underset{\sim}{\delta}) = \inf_{p_t(\underset{\sim}{\hat{u}}|\underset{\sim}{u}) \in \mathscr{P}_{\underset{\sim}{U}}(\delta)} I(\underset{\sim}{U}; \underset{\sim}{\hat{U}})$$

where

$$p_t(\underset{\sim}{\hat{u}}|\underset{\sim}{u}) \in \mathscr{P}_{\underset{\sim}{U}}(\underset{\sim}{\delta})$$

1

if

a) $p_t(\hat{u}|u) \geq 0$

b) $\displaystyle\sum_{\hat{u}\in\Omega(\hat{U})} p_t(\hat{u}|u) = 1$

c) $\displaystyle\sum_{u\in A(U)} \sum_{\hat{u}\in\Omega(\hat{U})} p_t(\hat{u}|u)Q(u)D_k(u,\hat{u}) \leq \delta_k$

$$k = 1,2,\ldots,m$$

and

$$I(U;\hat{U}) = \sum_{u\in A(U)} \sum_{\hat{u}\in\Omega(\hat{U})} p_t(\hat{u}|u)Q(u) \, \log\frac{p_t(\hat{u}|u)}{\omega(\hat{u})}$$

$$\omega(\hat{u}) = \sum_{u} p_t(\hat{u}|u)Q(u)$$

Where $\Omega(\hat{U})$ is the set of letters for which $\omega(\hat{u})$ may be strictly positive.

It can be shown straightforwardly that $R_U(\delta)$ is a convex $\cup$ function of δ as Gallager has shown for the scalar source and distortion measure case [1, pp. 445-446].

Since the argument of the multiple-constraint rate is a vector, the inverse rate-distortion function or distortion-rate function will be a surface in m-space for each value of its scalar argument. The distortion-rate function is given by the surface $\mathcal{D}_U(R) = \{\Delta : R_U(\Delta) = R\}$. Since $R_U(\Delta)$ is a convex function of Δ, the surface $\mathcal{D}_U(R)$ will be the lower boundary of a convex region in m-space, i.e., if $\Delta_1, \Delta_2 \in \mathcal{D}_U(R)$, then $\lambda\Delta_1 + (1-\lambda)\Delta_2 \geq \mathcal{D}_U(R)$ for any $0 \leq \lambda \leq 1$.

2

Although $R_{\underset{\sim}{U}}(\underset{\sim}{\delta})$ is defined for the same source as is $H(\underset{\sim}{U})$, coding theorems can be proved showing it to be the equivalent entropy rate (rather than entropy) of a reproduction (compressed) sequence $\{\hat{\underset{\sim}{u}}_k\}$ of the i.i.d. vector source sequence $\{\underset{\sim}{u}_k\}$ subject to a vector-valued fidelity criterion. Such coding theorems are immediate extensions of those for the usual rate distortion function with a single argument 3 , [1, Section 9.3], [2, Chap. 3] and hence are stated here without proof, after the necessary notation is introduced.

Let $\{\underset{\sim}{u}_k\}_{k=1}$ be an i.i.d. sequence of vectors (n-tuples) $\underset{\sim}{u}_k =$ $(u_{k,1}, u_{k,2}, \ldots, u_{k,n})$. Denote a block of N vectors by $\underset{\sim}{u}^{(N)} = (\underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_N)$; the superscript N will be suppressed whenever possible. Assume the distortion measure between blocks of vectors is a single-vector distortion measure, i.e.,

$$\underset{\sim}{D}_N(\underset{\sim}{u}, \hat{\underset{\sim}{u}}) = N^{-1} \sum_{k=1}^{N} \underset{\sim}{D}(\underset{\sim}{u}_k, \hat{\underset{\sim}{u}}_k)$$

where $\underset{\sim}{D}(\underset{\sim}{u}_k, \hat{\underset{\sim}{u}}_k)$ is the per-vector distortion measure.

An <u>encoder</u> with parameters (N, M) is defined as a mapping

$$f_E : A(\underset{\sim}{U})^N \rightarrow I_M \overset{\triangle}{=} \{1, 2, \ldots, M\} .$$

A <u>decoder</u> corresponding to f_E is defined by the mapping

$$f_D : I_M \rightarrow \Omega(\hat{\underset{\sim}{U}})^N .$$

An encoder-decoder is applied as follows: If $f_E\left(\underset{\sim}{u}^{(N)}\right) = i \in I_M$, then $\hat{\underset{\sim}{u}}^{(N)} = f_D(i) \in \Omega(\hat{\underset{\sim}{U}})^N$. The encoder-decoder has average distortion $\underset{\sim}{\delta}$ if $E \underset{\sim}{D}_N(\underset{\sim}{u}, \hat{\underset{\sim}{u}}) \leqq \underset{\sim}{\delta}$, in which case we have a <u>code</u> with parameters $(\underset{\sim}{\delta}, N, M)$. A nonnegative number R is called $\underset{\sim}{\delta}$-attainable if for arbitrary $\epsilon > 0$

3

and N sufficiently large, there exists a code $(\underset{\sim}{\rho} + \underset{\sim}{\delta} \ \underset{\sim}{\epsilon}, N, M)$, $\underset{\sim}{\epsilon} = (\epsilon,$ $\epsilon, \ldots, \epsilon)$, where

$$M \leq \exp\{N[R + \epsilon]\} \ .$$

Theorem 1: (Source Coding Theorem)

Given an i.i.d. vector source $\{\underset{\sim}{u}_k\}$ and a single-vector vector-valued distortion measure, then the rate $R_{\underset{\sim}{U}}(\underset{\sim}{\delta}) + \epsilon$ is $\underset{\sim}{\delta}$-attainable.

Theorem 2: (Converse Source Coding Theorem)

Given the source and distortion measure of Theorem 1, no rate smaller than $R_{\underset{\sim}{U}}(\underset{\sim}{\delta})$ is $\underset{\sim}{\delta}$-attainable. The converse theorem can also be stated as follows: If $\underset{\sim}{\delta} < \mathcal{D}(R)$, then the rate R is not $\underset{\sim}{\delta}$-attainable.

Unlike as in Berger [2, Chap. 3], the above theorems are stated in terms of $\underset{\sim}{\delta}$-attainable rates rather than $\underset{\sim}{\delta}$-admissible codes. The above terminology adapts more readily to the coding theorems and examples considered here and in [5] and [8].

The evaluation of $R_{\underset{\sim}{U}}(\underset{\sim}{\delta})$ is in general complicated. The immediate extension of the Kuhn-Tucker minimization of single constraint rate distortion functions [1, p. 459], [2, p. 37] yields

$$R_{\underset{\sim}{U}}(\underset{\sim}{\delta}) = H(\underset{\sim}{U}) + \max_{\underset{\sim}{\rho} \geq 0, \ \underset{\sim}{f} \in \mathcal{F}(\underset{\sim}{\rho})} \left\{ \sum_{\underset{\sim}{u} \in A(\underset{\sim}{U})} Q(\underset{\sim}{u}) \log f_{\underset{\sim}{u}}(\underset{\sim}{\rho}) - \underset{\sim}{\delta} \ \underset{\sim}{\rho}^T \right\} \qquad (1)$$

where T stands for transpose and where $\underset{\sim}{f} = \underset{\sim}{f}(\underset{\sim}{\rho}) = \{f_{\underset{\sim}{u}}(\underset{\sim}{\rho}), \ \underset{\sim}{u} \in A(\underset{\sim}{U})\} \in \mathcal{F}(\underset{\sim}{\rho})$ if

   (i) $\underset{\sim}{f}(\underset{\sim}{\rho}) \geq \underset{\sim}{0}$

   (ii) $\displaystyle\sum_{\underset{\sim}{u} \in A(\underset{\sim}{U})} f_{\underset{\sim}{u}}(\underset{\sim}{\rho}) \ e^{-\underset{\sim}{D}(\underset{\sim}{u},\hat{\underset{\sim}{u}})\underset{\sim}{\rho}^T} \leq 1, \ \hat{\underset{\sim}{u}} \in \Omega(\hat{\underset{\sim}{U}}) \ .$

4

Necessary and sufficient conditions on $\underset{\sim}{f}$ to yield the maximum in (1) are

(iii) that there exist a nonnegative solution $\omega(\underset{\sim}{\hat{u}})$ to

$$\sum_{\underset{\sim}{\hat{u}} \in \Omega(\underset{\sim}{\hat{U}})} \omega(\underset{\sim}{\hat{u}}) \; f_{\underset{\sim}{u}}(\underset{\sim}{\rho}) \; e^{-D(u,\hat{u})\rho^T} = Q(\underset{\sim}{u}); \quad \underset{\sim}{u} \in A(\underset{\sim}{U})$$

and

(iv) that (ii) hold with equality for all $\underset{\sim}{\hat{u}}$ such that $\omega(\underset{\sim}{\hat{u}}) > 0$.

When $A(\underset{\sim}{U}) = \Omega(\underset{\sim}{\hat{U}})$, we can lower bound $R_{\underset{\sim}{U}}(\underset{\sim}{\delta})$ by solving (ii) with equality for all $\underset{\sim}{\hat{u}} \in A(\underset{\sim}{U})$ and inserting the result into (1). This $\underset{\sim}{f}$ will yield $R_{\underset{\sim}{U}}(\underset{\sim}{\delta})$ if the auxillary conditions (iii) and (iv) are satisfied. The remaining maximization over $\underset{\sim}{\rho}$ simply involves taking derivatives and will yield parametric expressions for $R_{\underset{\sim}{U}}(\underset{\sim}{\delta})$ and $\underset{\sim}{\delta}$. The $A(\underset{\sim}{U}) = \Omega(\underset{\sim}{\hat{U}})$ condition is crucial since only in this case does (ii) with equality become $\|A(\underset{\sim}{U})\|$ equations in $\|A(\underset{\sim}{U})\|$ unknowns, where $\|A(\underset{\sim}{U})\|$ denotes the size of the alphabet. The properties of this bound--called the extended Shannon lower bound (ESLB)--are an immediate extension of the Kuhn-Tucker minimization and are summarized below sans proof.

Theorem 3: Define the vector $\underset{\sim}{\varphi}(\underset{\sim}{\rho}) = \left\{\varphi_{\underset{\sim}{u}}(\underset{\sim}{\rho}), \; \underset{\sim}{u} \in A(\underset{\sim}{U})\right\}$ as the nonnegative solution (if it exists) to

$$\underset{\sim}{\varphi}(\underset{\sim}{\rho}) \; \underset{\sim}{E} = \underset{\sim}{1} \tag{2a}$$

where

$$\underset{\sim}{E} = \left\{\exp\left\{-D(\underset{\sim}{u},\underset{\sim}{\hat{u}})\underset{\sim}{\rho}^T\right\}; \quad \underset{\sim}{u},\underset{\sim}{\hat{u}} \in A(\underset{\sim}{U})\right\} \tag{2b}$$

5

and $\underset{\sim}{\rho} = \{\rho_1, \ldots, \rho_M\}$ is a vector having nonnegative entries. Define

$$R_{\underset{\sim}{U}}^{(L)}(\underset{\sim}{\delta}) = H(\underset{\sim}{U}) + \max_{\underset{\sim}{\rho} \geqq \underset{\sim}{0}} \left\{ \sum_{\underset{\sim}{u} \in A(\underset{\sim}{U})} Q(\underset{\sim}{u}) \log \varphi_{\underset{\sim}{u}}(\underset{\sim}{\rho}) - \underset{\sim}{\delta} \, \underset{\sim}{\rho}^T \right\} \qquad (2c)$$

Since $\underset{\sim}{\varphi}(\underset{\sim}{\rho}) \in \mathfrak{F}(\underset{\sim}{\rho})$ we have

$$R_{\underset{\sim}{U}}(\underset{\sim}{\delta}) \geqq R_{\underset{\sim}{U}}^{(L)}(\underset{\sim}{\delta}) \qquad (3)$$

Furthermore, (3) will hold with equality iff there exists a nonnegative solution $\omega(\underset{\sim}{\hat{u}})$ to (iii) with $\underset{\sim}{f} = \underset{\sim}{\varphi}$.

To be properly careful we should worry about the existence of a nonnegative solution to the $\|A(\underset{\sim}{U})\|$ equations in $\|A(\underset{\sim}{U})\|$ unknowns described by (2). A straightforward extension of arguments in [6] based on a fundamental result of Jelinek [4] yield the following sufficient condition for tightness of $R_{\underset{\sim}{U}}(\underset{\sim}{\delta})$ for a region of small $\underset{\sim}{\delta}$.

Theorem 4: Given the vector source $\underset{\sim}{U}$, a distortion measure satisfying $D(\underset{\sim}{u}, \underset{\sim}{\hat{u}}) > D(\underset{\sim}{u}, \underset{\sim}{u}) = \underset{\sim}{\varphi}$, and an available reproduction alphabet $Q(\underset{\sim}{\hat{U}}) = A(\underset{\sim}{U})$, there exists a surface $\underset{\sim}{\mathfrak{D}}_c(\underset{\sim}{U})$ containing strictly positive elements such that

$$R_{\underset{\sim}{U}}(\underset{\sim}{\delta}) = R_{\underset{\sim}{U}}^{(L)}(\underset{\sim}{\delta}), \qquad \underset{\sim}{\delta} \leqq \underset{\sim}{\mathfrak{D}}_c(\underset{\sim}{U})$$

The surface $\underset{\sim}{\mathfrak{D}}_c(\underset{\sim}{U})$ is called the cutoff, or critical distortion surface.

As discussed in [8], the most useful types of per-vector distortion measures are compound and weighted-average distortion measures. The joint rate with a weighted-average distortion measure with weights $(\alpha_1, \ldots, \alpha_n)$ can be obtained from the joint rate with a compound distortion measure as follows: Consider a weighted-average m-valued distortion measure with per-letter distortion measures $d_i(u_i, \hat{u}_i)$, $i=1,\ldots,n$. We have from the definitions that

6

$$R_U(\underset{\sim}{\Delta}) = \inf_{\{\underset{\sim}{\Delta}_1,\ldots,\underset{\sim}{\Delta}_n\} \in \mathfrak{D}(\underset{\sim}{\Delta})} R_U(\underset{\sim}{\Delta}_1,\ldots,\underset{\sim}{\Delta}_n) \qquad (4a)$$

$$\mathfrak{D}(\underset{\sim}{\Delta}) = \left\{ \{\underset{\sim}{\Delta}_1,\ldots,\underset{\sim}{\Delta}_n\} : \sum_{k=1}^{n} \alpha_k \underset{\sim}{\Delta}_k = \underset{\sim}{\Delta} \right\} \qquad (4b)$$

Since $R_U(\underset{\sim}{\Delta}_1,\ldots,\underset{\sim}{\Delta}_n)$ is a convex $\cup$ function of $(\underset{\sim}{\Delta}_1,\ldots,\underset{\sim}{\Delta}_n)$, Gallager's Theorem 4.4.1 [1, p. 81] can be extended to show that the above infimum occurs at the value of $\{\underset{\sim}{\Delta}_1,\ldots,\underset{\sim}{\Delta}_n\}$, say $\{\underset{\sim}{\Delta}_1^*,\ldots,\underset{\sim}{\Delta}_n^*\}$, which is in $\mathfrak{D}(\underset{\sim}{\Delta})$ and satisfies the following condition: Define $\underset{\sim}{\Delta}_k = \left(\Delta_{k,1}, \Delta_{k,2}, \cdots, \Delta_{k,m}\right)$, then

$$\frac{d}{d\Delta_{k,j}} R_U(\underset{\sim}{\Delta}_1,\ldots,\underset{\sim}{\Delta}_n)\bigg|_{\{\underset{\sim}{\Delta}_1^*,\ldots,\underset{\sim}{\Delta}_n^*\}} \geqq \frac{d}{d\Delta_j} R_U(\underset{\sim}{\Delta})\bigg|_{\underset{\sim}{\Delta}}$$

$$k=1,\ldots,n$$

with equality for all k such that $\Delta_{k,j}^* > 0$. The above condition can be abbreviated to the statement that the slopes of $R_U(\underset{\sim}{\Delta}_1^*,\ldots,\underset{\sim}{\Delta}_n^*)$ in the jth coordinate in each of its n-vector arguments equals the slope of $R_U(\underset{\sim}{\Delta})$ in its jth coordinate.

Despite the apparently circuitous way of finding a weighted-average distortion measure, (4) later proves quite useful in [8]. It should be pointed out that, even though the equal-slope condition appears horrendous, rate-distortion functions are usually evaluated as parametric expressions for the rate and distortion in terms of the slope. Hence, in actuality, this condition usually simplifies such evaluations.

Conditional Rates

Given the two-dimensional source XY described by $Q(x,y)$ and $A(XY)$, the induced marginal source Y described by $Q(y)$ and $A(y)$, and for each $y \in A(Y)$ the marginal source X_y described by $Q(x|y)$ and $A(X|y)$, define the <u>conditional rate-distortion function</u> of X given Y as

$$R_{X|Y}(\underset{\sim}{\Delta}) = \underset{p_t(\hat{x}|x,y) \in \mathcal{P}_{X|Y}(\underset{\sim}{\Delta})}{\text{Inf}} I(X;\hat{X}|Y)$$

where

$$\mathcal{P}_{X|Y}(\underset{\sim}{\Delta}) = \left\{ \text{pmf's } p(\hat{x}|x,y) : \underset{\sim}{E} \, \underset{\sim}{D}(x,\hat{x}) = \sum_{\hat{x},x,y} p_t(\hat{x}|x,y) \, Q(x,y) \, \underset{\sim}{D}(x,\hat{x}) \leqq \underset{\sim}{\Delta} \right\}$$

The following theorem relates the conditional rate-distortion function to a weighted sum of marginal rate distortion functions.

Theorem 5

$$R_{X|Y}(\underset{\sim}{\Delta}) = \underset{\{\underset{\sim}{\Delta}_y\} \in \mathcal{D}(\underset{\sim}{\Delta})}{\text{Inf}} \sum_{y \in A(Y)} R_{X|y}(\underset{\sim}{\Delta}_y) Q(y)$$

where $R_{X|y}(\underset{\sim}{\Delta})$ is the marginal rate-distortion function of the single source X_y, and

$$\mathcal{D}(\underset{\sim}{\Delta}) = \left\{ \{\underset{\sim}{\Delta}_y, \, y \in A(Y)\} : \sum_{y \in A(Y)} \underset{\sim}{\Delta}_y \, Q(y) = \underset{\sim}{\Delta} \right\}$$

Proof:

First choose a set $\{\underset{\sim}{\Delta}_y\}$ such that

$$\sum_y \underset{\sim}{\Delta}_y \, Q(y) = \underset{\sim}{\Delta} \qquad (5)$$

8

and then choose $\|A(Y)\|$ test channels $p_t(\hat{x}|x,y) \in P_{X|y}(\underset{\sim}{\Delta}_y)$, $y \in A(Y)$, where $P_{X|y}(\underset{\sim}{\Delta}_y) = \left\{ p_t(\hat{x}|x,y) : \underset{\sim}{E} \underset{\sim}{D}(x,\hat{x}) \leqq \underset{\sim}{\Delta}_y \right\}$. Then

$$\sum_y Q(y) \sum_{x,\hat{x}} \underset{\sim}{D}(x,\hat{x}) \, p_t(\hat{x}|x,y) \, Q(x|y) \leqq \sum_y Q(y) \underset{\sim}{\Delta}_y = \underset{\sim}{\Delta}$$

and hence $p_t(\hat{x}|x,y) \in P_{X|Y}(\underset{\sim}{\Delta})$. Thus for any such set of test channels

$$\sum_y I(X;\hat{X}|y) \, Q(y) = I(X;\hat{X}|Y) \geqq R_{X|Y}(\underset{\sim}{\Delta})$$

so that choosing each test channel $p(\hat{x}|x,y)$, $y \in A(Y)$, to yield $R_{X|y}(\underset{\sim}{\Delta}_y)$, we have

$$\sum_{y \in A(Y)} Q(y) \, R_{X|y}(\underset{\sim}{\Delta}_y) \geqq R_{X|Y}(\underset{\sim}{\Delta}) \qquad (6)$$

for any set $\left\{ \underset{\sim}{\Delta}_y , \ y \in A(Y) \right\}$ satisfying (5).

Next choose a test channel $p_t(\hat{x}|x,y) \in P_{X|Y}(\underset{\sim}{\Delta})$. This test channel will result in some set of conditional distortions defined by

$$\underset{\sim}{\Delta}_y^* \triangleq \sum_{\hat{x},x} p_t(\hat{x}|x,y) \, Q(x|y) \, \underset{\sim}{D}(x,\hat{x})$$

such that

$$\sum_{y \in A(Y)} \underset{\sim}{\Delta}_y^* \, Q(y) \leqq \underset{\sim}{\Delta} .$$

For any such test channel

9

$$I(X;\hat{X}|Y) = \sum_{y \in A(Y)} I(X;\hat{X}|y)\, Q(y)$$

$$\geq \sum_{y \in A(Y)} R_{X|y}(\underset{\sim}{\Delta}_y^*)\, Q(y)$$

$$\geq \underset{\{\underset{\sim}{\Delta}_y\} \in \mathcal{D}(\underset{\sim}{\Delta})}{\text{Inf}} \sum_{y \in A(Y)} R_{X|y}(\underset{\sim}{\Delta}_y)\, Q(y)$$

Choosing $p(\hat{X}|x,y)$ to yield $R_{X|Y}(\underset{\sim}{\Delta})$ gives

$$R_{X|Y}(\underset{\sim}{\Delta}) \geq \underset{\{\underset{\sim}{\Delta}_y\} \in \mathcal{D}(\underset{\sim}{\Delta})}{\text{Inf}} \sum_{y \in A(Y)} R_{X|y}(\underset{\sim}{\Delta}_y)\, Q(y)$$

which, with (6) proves the theorem.

Similar to (4), Gallager's Theorem 4.4.1, [1, p. 87] implies that the infimum in Theorem 5 is achieved by adding up the rate-distortion functions at points of equal slope in all coordinates, i.e.,

$$R_{X|Y}(\underset{\sim}{\Delta}) = \sum_{y \in A(Y)} R_{X|y}(\underset{\sim}{\Delta}_y)\dot{Q}(y) \tag{7a}$$

where

$$\{\underset{\sim}{\Delta}_y\} \in \mathcal{D}(\underset{\sim}{\Delta}) \tag{7b}$$

and for each $k=1,\ldots,\dot{m}$

$$R'_{X|y}(\underset{\sim}{\Delta}_y) = R'_{X|Y}(\underset{\sim}{\Delta}) , \quad y \in A(Y). \tag{7c}$$

In general X and Y may themselves be vector sources. For simplicity, however, we shall here remain with scalar notation.

10

The appropriate coding theorems for conditional rate-distortion functions are slightly more complicated to state than Theorem 1 due to the presence of the side information. Some of the definitions must be modified accordingly. An <u>encoder</u> with parameters (n,M) is now defined as a mapping

$$f_E : A(XY)^n \to I_M$$

A <u>decoder</u> corresponding to f_E is defined by the mapping

$$f_D : I_M \times A(Y)^n \to \Omega(X)^n .$$

An <u>encoder-decoder</u> with parameters (n,M) is applied as follows: Let $f_E(\underset{\sim}{x},\underset{\sim}{y}) = i \in I_M$, then $\underset{\sim}{\hat{x}} = f_D(i,\underset{\sim}{y})$. The encoder-decoder has average distortion $\underset{\sim}{\Delta}$ if

$$\underset{\sim}{E} \left\{ \underset{\sim}{D}(\underset{\sim}{x},\underset{\sim}{\hat{x}}) \right\} \leq \underset{\sim}{\Delta}$$

where the expectation is now over the joint ensemble XY. In such a case we have a code $(\underset{\sim}{\Delta}, n, M)$. As before, a rate R is said to be $\underset{\sim}{\Delta}$-attainable (conditioned on $\{y_k\}$) if for arbitrary $\epsilon > 0$ and n-sufficiently large, there exists a code $(\underset{\sim}{\Delta} + \underset{\sim}{\epsilon}, n, M)$ where

$$M \leq \exp\left\{n[R + \epsilon]\right\}$$

<u>Theorem 6</u>: (Conditional Source Coding Theorem)

Given an i.i.d. sequence of dependent pairs $\{(x_k, y_k)\}$ and a single-vector distortion measure, assume that both encoder and decoder are allowed to observe perfectly the sequence $\{y_k\}$. Then, the rate $R_{X|Y}(\underset{\sim}{\Delta})$ is $\underset{\sim}{\Delta}$-attainable, and no rate smaller than $R_{X|Y}(\Delta)$ is $\underset{\sim}{\Delta}$-attainable.

11

The above theorem can be proved using a fairly straightforward extension of the usual techniques [2, Section 3.2]. We get a coding theorem for free, however, by noting that (7) is almost identical to Berger's (6.1.21) [2, p. 184] and hence is the (multiple-constraint) rate-distortion function of an i.i.d. composite source with "switch" pmf $Q(y)$ and $\|A(Y)\|$ subsource pmf's $Q(x|y)$, $y \in A(Y)$.

Note the obvious similarity between (7) and the corresponding entropy relation

$$H(X|Y) = \sum_{y \in A(Y)} H(X|y)Q(y)$$

One implicit difficulty with conditional rates is the choice of the appropriate reproducing alphabet for each $y \in A(Y)$. There are two natural choices--either the corresponding conditional source alphabet $A(X|y)$ or the full alphabet

$$A(X) = \bigcup_{y \in A(Y)} A(X|y).$$

For greatest ease in evaluating $R_{X|y}(\underset{\sim}{\Delta})$ it is desirable to have identical source and available reproducing alphabets, as previously noted. Specifically, the ESLB is well defined only for this case. Thus, the happiest possible state of affairs would be if $\Omega(\hat{X}|y) = A(X|y)$ for each y. Unfortunately, however, the assumption usually required is that the reproducing alphabet $\Omega(\hat{X}|y)$ be $A(X)$ for each y. Thus, in general, we may have

$$A(X|y) \subsetneq \Omega(\hat{X}|y) = A(X)$$

12

and hence actual evaluations of $R_{X|y}(\Delta)$ may be quite complicated because the matrix $\underset{\sim}{E}$ of (2b) is not square and not invertible.

If $A(X) = A(X|y)$, all $y \in A(Y)$, then $R_{X|y}(\underset{\sim}{\Delta})$ usually can be evaluated straightforwardly. An equivalent assumption is that the alphabet $A(XY)$ is the cartesian product of the marginal alphabets, i.e., $A(XY) = A(X) \times A(Y)$. In order to obtain some reasonably general bounds and evaluations for the rate-distortion functions in the theorems, we occasionally assume that $A(XY) = A(X) \times A(Y)$. This assumption is not made in the more general theorems. IF $A(XY) = A(X) \times A(Y)$ the ESLB of $R_{X|Y}(\underset{\sim}{\Delta})$ is well defined as the weighted sum of the individual ESLB's:

$$R_{X|Y}(\underset{\sim}{\Delta}) = \underset{\{\underset{\sim}{\Delta}_y\} \in \mathcal{D}(\underset{\sim}{\Delta})}{\text{Inf}} \sum_y R_{X|y}(\underset{\sim}{\Delta}_y) \, Q(y)$$

$$\geq \underset{\{\underset{\sim}{\Delta}_y\} \in \mathcal{D}(\underset{\sim}{\Delta})}{\text{Inf}} \sum R_{X|y}^{(L)}(\underset{\sim}{\Delta}_y) \, Q(y) \triangleq R_{X|Y}^{(L)}(\underset{\sim}{\Delta})$$

where $R_{X|y}^{(L)}(\underset{\sim}{\Delta})$ is given by (2) with $\underset{\sim}{U} = X_y$. The infimum is obtained by adding up the functions at points of equal slope in each component. As in Theorem 4, we have the following result:

Corollary 1: If $A(XY) = A(X) \times A(Y)$ and (4) is satisfied, then there exists a cutoff distortion surface $\underset{\sim}{\mathcal{D}}_c(X|Y) > \underset{\sim}{0}$ such that

$$R_{X|Y}(\underset{\sim}{\Delta}) = R_{X|Y}^{(L)}(\underset{\sim}{\Delta}), \qquad \underset{\sim}{\Delta} \leqq \underset{\sim}{\mathcal{D}}_c(X|Y).$$

Difference Distortion Measures and Examples

The calculations involved in evaluating conditional rates are often simpler when dealing with difference distortion measures. Roughly

13

speaking, given a conditional rate-distortion function and a difference distortion measure, there exists a simple upper bound which actually yields the conditional rate iff the source satisfies a certain property.

Theorem 7: Let a product ensemble XY and a difference distortion measure be given. Then for any function $f(y)$ defined on $A(Y)$

$$R_{X|Y}(\underset{\sim}{\triangle}) = R_{X-f(Y)|Y}(\underset{\sim}{\triangle}) \leqq R_{X-f(Y)}(\underset{\sim}{\triangle})$$

with equality iff y and $x - f(y)$ are independent.

Proof: Define $z = x - f(y)$. Given a source pmf $Q_{X|Y}(x|y)$ we will have also the source pmf

$$Q_{Z|Y}(z|y) = Q_{X|Y}(z + f(y)|y)$$

since $x = z + f(y)$. Furthermore, any test channel $p_t^{(1)}(\hat{x}|x,y) \in \mathscr{P}_{X|Y}(\underset{\sim}{\triangle})$ induces a test channel $p_t^{(2)}(\hat{z}|z,y) = p_t^{(1)}(\hat{z} + f(y)|z + f(y),y) \in \mathscr{P}_{Z|y}(\underset{\sim}{\triangle})$ since

$$E\{d(\hat{z},y)|y\} = \sum_{\substack{z \in A(Z|y) \\ \hat{z} \in \Omega(\hat{Z}|y)}} d(\hat{z} - z)\, p_t^{(2)}(\hat{z}|z,y)\, Q_{Z|Y}(z|y)$$

$$= \sum_{\substack{x \in A(X|y) \\ \hat{z} \in \Omega(\hat{X}|y)}} d(\hat{x} - x)\, p_t^{(1)}(\hat{x}|x,y)\, Q_{X|Y}(x|y) \leqq \underset{\sim}{\triangle}$$

For the test channel $p_t^{(1)}(\hat{x}|x,y)$ we have

$$I(X;\hat{X}|y) = H(X|y) - H(X|\hat{X}y) = H(X + f(y)|y) - H(X + f(y)|\hat{X} + f(y),y)$$

$$= I(Z;\hat{Z}|y)$$

14

Choosing $p_t^{(1)}(\hat{x}|x,y)$ to yield $R_{X|y}(\underset{\sim}{\Delta})$ we have

$$R_{X|y}(\underset{\sim}{\Delta}) = I(Z; \hat{Z}|y) \geqq R_{Z|y}(\underset{\sim}{\Delta}) \tag{8}$$

By defining $x = z + f(y)$ and repeating the previous procedure with and z interchanged we obtain

$$R_{X|y}(\underset{\sim}{\Delta}) \leqq R_{Z|y}(\underset{\sim}{\Delta}) \tag{9}$$

Eq. (9) implies the left hand equality in Theorem 7. The right hand inequality follows from Theorem 2.1 of [8].

This theorem can be used to provide an alternate evaluation of the conditional rates in the examples of [8]. In the binary example $f(y) = y$ and the source x can be viewed as the mod-2 sum of two independent random variables y and z where

$$Q(y) = \frac{1}{2}, \qquad y = 0, 1$$

$$Q(z) = (1 - p)^{\delta_{z,0}} p^{1-\delta_{z,0}}$$

The Gaussian case can be viewed similarly. Perhaps a more interesting view of the Gaussian case is to choose the function $f(y)$ such that $x - f(y)$ and y are independent. It is well known that this is accomplished by choosing the conditional expectation

$$f(y) = E\left[x|y\right] = m_X + r \frac{\sigma_X}{\sigma_Y} (y - m_Y) .$$

Thus $R_{X|Y}(\Delta) = R_{X-f(Y)}(\Delta)$ is simply the marginal rate-distortion function of a zero mean Gaussian random variable with variance $\sigma_Y^2(1 - r^2)$.

It is worth observing that in both of the preceding examples the conditional rate of x given y is the marginal rate of the "innovation" $\nu = x - \hat{x}(y)$, where $\hat{x}(y)$ is the best estimate of x given y, i.e., the estimate that minimizes $\underset{\sim}{E}\, d(x, \hat{x}(y))$.

REFERENCES

1. R. G. Gallager, _Information Theory and Reliable Communication_, New York: Wiley, 1968.

2. Toby Berger, _Rate Distortion Theory_, Englewood, New Jersey: Prentice-Hall, 1971.

3. C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," _IRE Nat'l. Conv. Rec._, Pt. 4, 1959, pp. 142-163.

4. F. Jelinek, "Evaluation of Distortion-Rate Functions for Low Distortions," _Proc. IEEE_, Vol. 55, November 1967, pp. 2067-2068.

5. R. M. Gray and A. D. Wyner, "Source Coding for Simple Networks," submitted to _Bell Systems Technical Journal_.

6. R. M. Gray, "Information Rates of Stationary Ergodic Sources," _IEEE Trans. Inform. Theory_, Vol. IT-17, September 1971, pp. 516-523.

7. B. M. Leiner, Ph.D. dissertation (in preparation), Stanford University.

8. R. M. Gray, "A New Class of Lower Bounds to Information Rates of Sources with Memory via Conditional Rate-Distortion Functions," to appear, _Trans. IEEE-IT_, 1973.