INFORMATION AND CONTROL 24, 358–383 (1974)

Expurgated Bounds, Bhattacharyya Distance, and Rate Distortion Functions*

JIM K. OMURA

System Science Department, University of California, Los Angeles, California

We examine new low rate error upper bounds for M equally likely code words used over discrete input channels. When optimized over the code ensemble probability distribution, these bounds coincide with the optimized expurgated bounds and the error exponents satisfy rate distortion equations for natural Bhattacharyya distances. Proofs for these error bounds do not require expurgation of code words, and for certain "modular" channels including all binary input memoryless channels, the bounds extend to convolutional codes.

1. INTRODUCTION

Consider a code of blocklength n and rate $R = (1/n) \ln M$ for transmission over a channel with input and output alphabets $\{0, 1,..., K-1\}$ and $\{0, 1,..., J-1\}$, respectively. Let code words be represented by $\mathbf{x}_1, \mathbf{x}_2,..., \mathbf{x}_M$, the channel output sequence by $\mathbf{y}$, and the channel conditional probability for blocklength n by $p(\mathbf{y} \mid \mathbf{x})$. This is illustrated in Fig. 1.

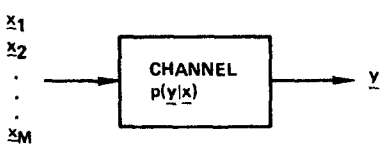

FIG. 1. General discrete channel.

Gallager (1965) has shown that there exist codes of blocklength n and rate R whose probability of error, P_e, is uniformly bounded by the expurgated bound given by

$$P_e < \exp\{-n[-\rho R + E_x(\rho, Q) - (\rho/n) \ln 4]\}, \qquad \rho \geqslant 1,$$

* This work was supported by the National Science Foundation under Grant GK-23982.

where

$$E_x(\rho, Q) = -\frac{\rho}{n} \ln \left\{ \sum_{\mathbf{x}} \sum_{\mathbf{x}'} Q(\mathbf{x}) Q(\mathbf{x}') \left[\sum_{\mathbf{y}} (p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{y} \mid \mathbf{x}'))^{1/2} \right]^{1/\rho} \right\}.$$

Here $Q(\mathbf{x})$ is the probability distribution over the ensemble of independently chosen code words. Achievement of the tightest bound minimization with respect to ρ and $Q(\mathbf{x})$ must be performed under the restrictions

$$\rho \geqslant 1,$$
$$Q(\mathbf{x}) \geqslant 0, \qquad \sum_{\mathbf{x}} Q(\mathbf{x}) = 1.$$

Jelinek (1968) investigated the problem of minimizing the expurgated bound for discrete memoryless channels and found quite general sufficient conditions under which, for a given ρ, the optimizing distribution $Q(\mathbf{x})$ satisfies

$$Q(\mathbf{x}) = \prod_{i=1}^{n} Q(x_i).$$

This condition is stated in the following theorem due to Jelinek.

THEOREM. *For a given* $\rho \in [1, \infty)$,

$$f(\rho, Q) = \exp\left[-\frac{n}{\rho} E_x(\rho, Q) \right]$$

is a convex function of $Q(\cdot)$ *if and only if the symmetric* $K \times K$ *matrix*

$$\left\{ \left[\sum_{j=0}^{J-1} (p(j \mid i)\, p(j \mid k))^{1/2} \right]^{1/\rho} \right\} \tag{1}$$

is nonnegative definite. In that case,

$$E_x(\rho) \triangleq \max_{Q} E_x(\rho, Q)$$

is given by

$$E_x(\rho) = -\rho \ln \left\{ \sum_{i,k} Q^*(i) Q^*(k) \left[\sum_{j=0}^{J-1} (p(j \mid i)\, p(j \mid k))^{1/2} \right]^{1/\rho} \right\},$$

where

$$Q^*(\mathbf{x}) = \prod_{i=1}^{n} Q^*(x_i)$$

and $Q^*(i)$ *is any distribution over inputs* $\{0, 1,..., K-1\}$ *that satisfies*

$$\sum_i Q^*(i) \left[\sum_{j=0}^{J-1} (p(j \mid i)\, p(j \mid k))^{1/2}\right]^{1/\rho}$$

$$\geqslant \sum_{i,k} Q^*(i) Q^*(k) \left[\sum_{j=1}^{J-1} (p(j \mid i)\, p(j \mid k))^{1/2}\right]^{1/\rho}$$

for all $k \in \{0, 1,..., K-1\}$, *with equality whenever* $Q^*(k) > 0$.

Jelinek also derived a necessary and sufficient condition for the matrix given in (1) to be nonnegative definite for all $\rho \geqslant 1$. He then examined a class of equidistant channels that includes all binary input channels and showed that for this class the optimal expurgated exponent is attained by the uniform distribution over the inputs. For this case he obtained closed form expressions for the expurgated error exponents.

In this paper we derive low rate error upper bounds that do not require expurgation of code words from a code and for a certain class of channels which includes all memoryless binary input channels they extend to convolutional codes. When minimized with respect to the code ensemble probability distribution, these bounds are the same as the optimized expurgated bounds. At first, we will treat general discrete input channels. For extensions to convolutional codes, we restrict ourselves to memoryless modular channels.

This paper is partly tutorial in that we explore how the natural Bhattacharyya distance associated with a channel is related to our upper bounds and in some cases to lower bounds of the error probability. Gallager (1965), Shannon, Gallager, and Berlekamp (1967), Berlekamp (1969), and Kailath (1967) have discussed the Bhattacharyya distance and its relation to error bounds. We present an interpretation from a rate distortion theory viewpoint and show that our error exponent in general satisfies a lower bound to the rate distortion function which is defined by the code ensemble probability distribution and the Bhattacharyya distance. The code distribution that minimizes our error bound (optimized expurgated bound) satisfies some rate distortion function equation exactly and the optimum probability distribution yields an upper bound to the natural rate distortion function for each value of distortion.

2. Error Bounds and Bhattacharyya Distance

We now present the error upper bound for a given code $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$. Following the notation associated with the general discrete channel of Fig. 1, we define the natural Bhattacharyya distance between code words $\mathbf{x}$ and $\mathbf{x}'$ as

$$d(\mathbf{x}, \mathbf{x}') = -\frac{1}{n} \ln \sum_{\mathbf{y}} (p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{y} \mid \mathbf{x}'))^{1/2} \tag{2}$$

$$= d(\mathbf{x}', \mathbf{x}).$$

Throughout this paper, we assume that $d(\mathbf{x}, \mathbf{x}') < \infty$ for all $\mathbf{x}, \mathbf{x}' \in A^n$, where $A = \{0, 1, \ldots, K-1\}$. This means that the zero-error capacity defined by Shannon (1956) is equal to zero. Letting $Y_m = \{\mathbf{y} : p(\mathbf{y} \mid \mathbf{x}_m) > p(\mathbf{y} \mid \mathbf{x}_{m'})$ all $m' \neq m\}$ and $Y_m{}^c$ be its complement, the probability of error when $\mathbf{x}_m$ is the transmitted code word is bounded by

$$\begin{aligned} P_{e,m} &= \sum_{\mathbf{y} \in Y_m{}^c} p(\mathbf{y} \mid \mathbf{x}_m) \\ &= \sum_{m' \neq m} \sum_{\mathbf{y} \in Y_{m'}} p(\mathbf{y} \mid \mathbf{x}_m) \\ &\leqslant \sum_{m' \neq m} \sum_{\mathbf{y} \in Y_{m'}} p(\mathbf{y} \mid \mathbf{x}_m) \left[\frac{p(\mathbf{y} \mid \mathbf{x}_{m'})}{p(\mathbf{y} \mid \mathbf{x}_m)}\right]^{1/2} \\ &\leqslant \sum_{m' \neq m} \sum_{\mathbf{y}} (p(\mathbf{y} \mid \mathbf{x}_m)\, p(\mathbf{y} \mid \mathbf{x}_{m'}))^{1/2} \\ &= \sum_{m' \neq m} e^{-nd(\mathbf{x}_m, \mathbf{x}_{m'})} \end{aligned}$$

For any $s \in [-1, 0]$, we have Holder's inequality

$$[P_{e,m}]^{-s} \leqslant \sum_{m' \neq m} e^{snd(\mathbf{x}_m, \mathbf{x}_{m'})} \tag{3}$$

for $m = 1, 2, \ldots, M$.

We now consider an ensemble of codes where each code word is chosen independently with probability distribution $\{Q(\mathbf{x}) : \mathbf{x} \in A^n\}$. To derive the expurgated bound, Gallager (1965) averaged (3) with respect to all M code words to obtain

$$\overline{[P_{e,m}]^{-s}} \leqslant M f_n(s, Q), \tag{4}$$

where

$$f(s, Q) = \sum_{\mathbf{x}} \sum_{\mathbf{x}} Q(\mathbf{x}') Q(\mathbf{x}) \, e^{snd(\mathbf{x}, \mathbf{x}')}.$$

We take a slightly different approach by averaging $[P_{e,m}]^{-s}$ over the ensemble of the mth code word only to get

$$\overline{P[_{e,m}]^{-s}}^{m} \leqslant \sum_{m' \neq m} \sum_{\mathbf{x}} Q(\mathbf{x}) \, e^{snd(\mathbf{x}, \mathbf{x}_{m'})}.$$

Now define

$$\gamma_n(s, Q) = \left\{ \max_{\substack{\mathbf{x}' \in A^n \\ Q(\mathbf{x}') > 0}} \sum_{\mathbf{x}} Q(\mathbf{x}) \, e^{snd(\mathbf{x}, \mathbf{x}')} \right\}^{1/n}.$$

Then

$$\begin{aligned} \overline{[P_{e,m}]^{-s}}^{m} &\leqslant M[\gamma_n(s, Q)]^n \\ &= e^{n[R + \ln \gamma_n(s, Q)]}. \end{aligned}$$

Hence, given any set of $M - 1$ code words $\{\mathbf{x}_{m'}, m' \neq m\}$, there exists a code word $\mathbf{x}_m \in A^n$ such that

$$P_{e,m} \leqslant e^{-nD_n(Q)}, \tag{5}$$

where

$$\begin{aligned} D_n(Q) &= (1/s)[R + \ln \gamma_n(s, Q)], \\ & s \in [-1, 0]. \end{aligned}$$

We next prove our upper bound on the error probability

$$P_e = \frac{1}{M} \sum_{m=1}^{M} P_{e,m}$$

without expurgation of code words from a code.

THEOREM. *There exists a code with probability of error satisfying the bound*

$$P_e \leqslant 2e^{-nD_n(Q)}.$$

Proof. We consider a sequence of M codes where each code is the same as the previous code except for one code word. In particular, let us begin

with any code in our ensemble $B_0 = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\}$ and consider a sequence of M codes $B_1, B_2, ..., B_M$ given by

$$\begin{aligned} B_1 &= \{\mathbf{x}_1^1, \mathbf{x}_2^1, ..., \mathbf{x}_M^1\} \\ B_2 &= \{\mathbf{x}_1^2, \mathbf{x}_2^2, ..., \mathbf{x}_M^2\} \\ &\vdots \\ B_k &= \{\mathbf{x}_1^k, \mathbf{x}_2^k, ..., \mathbf{x}_M^k\} \\ &\vdots \\ B_M &= \{\mathbf{x}_1^M, \mathbf{x}_2^M, ..., \mathbf{x}_M^M\}, \end{aligned}$$

where

$$\mathbf{x}_m^k = \begin{cases} \mathbf{x}_m & k < m, \\ \mathbf{x}_m^m & k \geqslant m. \end{cases}$$

Here, only one code word, $\mathbf{x}_k^k$, is changed in converting B_{k-1} to B_k. The error probability of the kth code word in B_k is bounded by

$$P_{e,k}(B_k) \leqslant \sum_{k' \neq k} e^{-nd(\mathbf{x}_k^k, \mathbf{x}_{k'}^k)}.$$

Given B_{k-1}, the arguments leading to (5) demonstrated that there exists an $\mathbf{x}_k^k \in A^n$ such that

$$P_{e,k}(B_k) \leqslant e^{-nD_n(Q)}. \tag{6}$$

Here $\mathbf{x}_k^k$ is chosen to satisfy this bound. We do this for each selection of $\mathbf{x}_1^1, \mathbf{x}_2^2, ..., \mathbf{x}_M^M$ and form the code sequence in this manner. For $m < k$, define

$$g_m(B_k) = \sum_{m' \neq m} e^{-nd(\mathbf{x}_m^k, \mathbf{x}_{m'}^k)}$$

which is the upper bound to the error probability of the mth code word in the code B_k. Except for the kth code word, $\mathbf{x}_k^k$, in B_k, all the code words in B_k are the same as those in B_{k-1} so that

$$\begin{aligned} g_m(B_k) &= \sum_{\substack{m' \neq m \\ m' \neq k}} e^{-nd(\mathbf{x}_m^{k-1}, \mathbf{x}_{m'}^{k-1})} + e^{-nd(\mathbf{x}_m^k, \mathbf{x}_k^k)} \\ &\leqslant g_m(B_{k-1}) + e^{-nd(\mathbf{x}_m^k, \mathbf{x}_k^k)} \\ &\leqslant g_m(B_m) + \sum_{l=m+1}^{k} e^{-nd(\mathbf{x}_m^l, \mathbf{x}_l^l)}. \end{aligned}$$

We are primarily interested in the last code in the sequence, B_M, so let $k = M$,

$$g_m(B_M) \leqslant g_m(B_m) + \sum_{l=m+1}^{M} e^{-nd(\mathbf{x}_m{}^l, \mathbf{x}_l{}^l)}.$$

(6) gives $g_m(B_m) \leqslant e^{-nD_n(Q)}$ so that $g_m(B_M) \leqslant e^{-nD_n(Q)} + \sum_{l=m+1}^{M} e^{-nd(\mathbf{x}_m{}^l, \mathbf{x}_l{}^l)}$. Consider

$$\frac{1}{M}\sum_{m=1}^{M} g_m(B_M) \leqslant e^{-nQ_n(D)} + \frac{1}{M}\sum_{m=1}^{M-1}\sum_{l=m+1}^{M} e^{-nd(\mathbf{x}_m{}^l, \mathbf{x}_l^l)}.$$

Interchanging orders of summation,

$$\begin{aligned}
\frac{1}{M}\sum_{m=1}^{M-1}\sum_{l=m+1}^{M} e^{-nd(\mathbf{x}_m{}^l, \mathbf{x}_l{}^l)} &\leqslant \frac{1}{M}\sum_{l=2}^{M}\sum_{m=1}^{l-1} e^{-nd(\mathbf{x}_m{}^l, \mathbf{x}_l{}^l)} \\
&\leqslant \frac{1}{M}\sum_{l=2}^{M}\sum_{m\neq l} e^{-nd(\mathbf{x}_m{}^l, \mathbf{x}_l{}^l)} \\
&= \frac{1}{M}\sum_{l=2}^{M} g_l(B_l) \\
&\leqslant e^{-nD_n(Q)}.
\end{aligned}$$

Hence for the code $B_M = \{\mathbf{x}_1{}^1, \mathbf{x}_2{}^2, \ldots, \mathbf{x}_M{}^M\}$, we have

$$\begin{aligned}
P_e &= \frac{1}{M}\sum_{m=1}^{M} P_{e,m} \\
&\leqslant \frac{1}{M}\sum_{m=1}^{M} g_m(B_M) \\
&\leqslant 2e^{-nD_n(Q)}. \quad \blacksquare
\end{aligned}$$

Although this theorem does not require expurgation of code words from a code for an arbitrary probability distribution, this bound is generally weaker than the well-known expurgated bound. To see this, recall from (4) that if we average $[P_{e,m}]^{-s}$ over all code words, we have

$$\overline{[P_{e,m}]^{-s}} \leqslant Mf_n(s, Q),$$

while averaging only over the mth code word, we have from (5)

$$\overline{[P_{e,m}]^{-s}}^{m} \leqslant M[\gamma_n(s, Q)]^n,$$

where in general for any $s \in [-1, 0]$,

$$\begin{aligned} f_n(s, Q) &= \sum_{\mathbf{x}'} \sum_{\mathbf{x}} Q(\mathbf{x}') Q(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')} \\ &\leqslant \max_{\substack{\mathbf{x}' \in A^n \\ Q(\mathbf{x}')>0}} \sum_{\mathbf{x}} Q(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')} \\ &= [\gamma_n(s, Q)]^n, \end{aligned}$$

We have equality, however, for any optimizing probability distribution.

THEOREM. *For fixed $s \in [-1, 0]$, a probability distribution $\{Q_s^*(\mathbf{x}) : \mathbf{x} \in A^n\}$ that minimizes $f_n(s, Q)$ also minimizes $\gamma_n(s, Q)$ and satisfies*

$$\min_Q f_n(s, Q) = f_n(s, Q_s^*) = [\gamma_n(s, Q_s^*)]^n.$$

Necessary conditions for $\{Q_s^*(\mathbf{x}) : \mathbf{x} \in A^n\}$ are

$$C \leqslant \sum_{\mathbf{x}} Q_s^*(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')} \tag{7}$$

for some constant C with equality for all $\mathbf{x}' \in A^n$ such that $Q_s^*(\mathbf{x}') > 0$.

Proof. Let $2C$ be a lagrange multiplier for the constraint $\sum_{\mathbf{x}} Q(\mathbf{x}) = 1$, and let

$$J(Q) = f_n(s, Q) - 2C \sum_{\mathbf{x}} Q(\mathbf{x}).$$

The probability distribution that minimizes $f_n(s, Q)$ must satisfy

$$\frac{\partial J(Q)}{\partial Q(\mathbf{x}')}\bigg|_{Q=Q_s^*} = 2 \sum_{\mathbf{x}} Q_s^*(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')} - 2C \geqslant 0$$

with equality for $\mathbf{x}'$ such that $Q_s^*(\mathbf{x}') > 0$. We now have for any distribution $\{Q(\mathbf{x}) : \mathbf{x} \in A^n\}$,

$$f_n(s, Q_s^*) \leqslant f_n(s, Q) \leqslant [\gamma_n(s, Q)]^n.$$

But

$$[\gamma_n(s, Q_s^*)]^n = \max_{\substack{\mathbf{x}' \in A^n \\ Q_s^*(\mathbf{x}') > 0}} \sum_{\mathbf{x}} Q_s^*(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')}$$
$$= C$$
$$= f_n(s, Q_s^*).$$

Hence, Q_s^* also minimizes $\gamma_n(s, Q)$. ∎

In general, there may be several local minima of $f(s, Q)$ with respect to the distribution $\{Q(\mathbf{x}) : \mathbf{x} \in A^n\}$. Each of these would satisfy the necessary conditions given above and yield $f(s, Q^*) = [\gamma_n(s, Q^*)]^n$. Jelinek (1968) pointed out that when the $K^n \times K^n$ matrix $| e^{nd(\mathbf{x},\mathbf{x}')} |$ is nonnegative definite, then the minimizing distribution is unique and the above conditions for $Q^*(\mathbf{x})$ are both necessary and sufficient. He examined this problem in more detail for memoryless channels.

Suppose we have two probability distributions that satisfy the necessary conditions of (7). In particular, let

$Q_s^1(\mathbf{x})$ satisfy

$$f(s, Q_s^1) = [\gamma(s, Q_s^1)]^n \leqslant \sum_{\mathbf{x}} Q_s^1(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')}$$

with equality when $Q_s^1(\mathbf{x}') > 0$, and $Q_s^2(\mathbf{x})$ satisfy

$$f(s, Q_s^2) = [\gamma(s, Q_s^2)]^n \leqslant \sum_{\mathbf{x}} Q_s^2(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')}$$

with equality when $Q_s^2(\mathbf{x}') > 0$.

Defining $B_1 = \{\mathbf{x} : Q_s^1(\mathbf{x}) > 0\}$ and $B_2 = \{\mathbf{x} : Q_s^2(\mathbf{x}) > 0\}$, we have

LEMMA. *If* $B_1 \subset B_2$, *then* $f(s, Q_s^1) \leqslant f(s, Q_s^2)$.

Proof.

$$\frac{f(s, Q_s^1)}{f(s, Q_s^2)} \leqslant \frac{\sum_{\mathbf{x}} Q_s^1(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')}}{f(s, Q_s^2)}.$$

Averaging this with respect to distribution $Q_s^2(\mathbf{x})$, we have

$$\frac{f(s, Q_s^1)}{f(s, Q_s^2)} = \sum_{\mathbf{x}'} Q_s^2(\mathbf{x}') \frac{f(s, Q_s^1)}{f(s, Q_s^2)}$$
$$\leqslant \frac{\sum_{\mathbf{x}'} Q_s^2(\mathbf{x}') \sum_{\mathbf{x}} Q_s^1(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')}}{f(s, Q_s^2)}$$
$$\leqslant \frac{\sum_{\mathbf{x}} Q_s^1(\mathbf{x}) \sum_{\mathbf{x}'} Q_s^2(\mathbf{x}')\, e^{snd(\mathbf{x},\mathbf{x}')}}{f(s, Q_s^2)}$$

But when $Q_s^1(\mathbf{x}) > 0$, $\mathbf{x} \in B_1 \subset B_2$ and

$$\sum_{\mathbf{x}'} Q_s^2(\mathbf{x}')\, e^{snd(\mathbf{x},\mathbf{x}')} = f(s, Q_s^2).$$

Hence

$$\frac{f(s, Q_s^1)}{f(s, Q_s^2)} \leqslant \sum_{\mathbf{x}} Q_s^1(\mathbf{x})$$

$$= 1. \quad \blacksquare$$

We conclude this section by noting that all the results in this section apply to discrete input continuous output channels when defining the Bhattacharyya distance for code words $\mathbf{x}, \mathbf{x}' \in A^n$ as

$$d(\mathbf{x}, \mathbf{x}') = -\frac{1}{n} \ln \int (p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{y} \mid \mathbf{x}'))^{1/2}\, d\mathbf{y},$$

where $p(\mathbf{y} \mid \mathbf{x})$ is the conditional probability density of the channel output given the input vector $\mathbf{x} \in A^n$. A common example is the additive gaussian noise channel with discrete amplitude modulated input pulses.

3. Rate Distortion Functions

We next present some relationships between our error exponent $D_n(Q)$ and rate distortion functions. For any code ensemble probability distribution $\{Q(\mathbf{x}) : \mathbf{x} \in A^n\}$, define $B(Q) = \{\mathbf{x} : Q(\mathbf{x}) > 0\}$. For this given probability distribution, choose $B(Q)$ as our source and representation alphabet and $d(\mathbf{x}, \mathbf{x}')$ given by (2) as a distortion measure between $\mathbf{x}, \mathbf{x}' \in B(Q)$. This source, $(B(Q), Q)$, and Bhattacharyya distance yield a natural rate distortion function originally defined by Shannon (1960),

$$R_n(D; Q) = \frac{1}{n} \min \sum_{\mathbf{x}'} \sum_{\mathbf{x}} Q(\mathbf{x}) Q(\mathbf{x}' \mid \mathbf{x}) \ln \left[\frac{Q(\mathbf{x}' \mid \mathbf{x})}{P(\mathbf{x}')}\right]$$

where

$$P(\mathbf{x}') = \sum_{\mathbf{x}} Q(\mathbf{x}' \mid \mathbf{x}) Q(\mathbf{x})$$

and the minimization is over all conditional probabilities that satisfy the constraint

$$\sum_{\mathbf{x}'}\sum_{\mathbf{x}} Q(\mathbf{x}' \mid \mathbf{x})\, Q(\mathbf{x})\, d(\mathbf{x}, \mathbf{x}') \leqslant D.$$

A well-known (Gallager (1968), Berger (1971)) convenient form for deriving bounds to $R_n(D; Q)$ is

$$R_n(D; Q) = \max_{\substack{s \leqslant 0 \\ \lambda \in \Lambda_s}} \left\{ sD + \frac{1}{n} \sum_{\mathbf{x}} Q(\mathbf{x}) \ln \lambda(\mathbf{x}) \right\},$$

where

$$\Lambda_s = \left\{ \lambda;\ \lambda(\mathbf{x}) \geqslant 0, \sum_{\mathbf{x}} Q(\mathbf{x})\, \lambda(\mathbf{x})\, e^{snd(\mathbf{x}, \mathbf{x}')} \leqslant 1\ \forall \mathbf{x}' \in B(Q) \right\}.$$

Let us now reexamine our upper bound to the error probability. Recall that there exists a code of blocklength n and rate R satisfying the bound

$$P_e \leqslant 2e^{-nD_n(Q)}, \tag{8}$$

where

$$D_n(Q) = (1/s)[R + \ln \gamma_n(s, Q)],$$

$$\gamma_n(s, Q) = \left\{ \max_{\substack{\mathbf{x}' \in A^n \\ Q(\mathbf{x}') > 0}} \sum_{\mathbf{x}} Q(\mathbf{x})\, e^{snd(\mathbf{x}, \mathbf{x}')} \right\}^{1/n},$$

$$s \in [-1, 0].$$

THEOREM. $R_n(D; Q) \geqslant \max_{s \leqslant 0} \{sD - \ln \gamma_n(s, Q)\} = R_L(D; Q)$.

Proof. Here we have

$$[\gamma_n(s, Q)]^n = \max_{\mathbf{x}' \in B(Q)} \sum_{\mathbf{x}} Q(\mathbf{x})\, e^{snd(\mathbf{x}, \mathbf{x}')}.$$

Letting

$$\lambda(\mathbf{x}) = 1/[\gamma_n(s, Q)]^n,$$

we get

$$\sum_{\mathbf{x}} Q(\mathbf{x})\, \lambda(\mathbf{x})\, e^{snd(\mathbf{x}, \mathbf{x}')} = \frac{\sum_{\mathbf{x}} Q(\mathbf{x})\, e^{snd(\mathbf{x}, \mathbf{x}')}}{[\gamma_n(s, Q)]^n} \leqslant 1,$$

and hence $\lambda \in \Lambda_s$. This choice of $\lambda \in \Lambda_s$ gives the lower bound. ∎

This bound applies to source and representation alphabets restricted to $B(Q) \subset A^n$. In most cases of interest, we will have $B(Q) = A^n$. The lower bound to the rate distortion function can be expressed in parametric form in terms of $s \in [-\infty, 0]$,[1]

$$R_L(D_s, Q) = sD_s - \ln \gamma_n(s, Q),$$
$$D_s = (1/\gamma_n(s, Q))(\partial/\partial s)\, \gamma_n(s, Q).$$

For parameter values $s \in [-1, 0]$, we see that our error exponent $D_n(Q)$ in (8) satisfies

$$R = R_L(D_n ; Q). \tag{9}$$

Hence the error exponent satisfies a lower bound to the natural rate distortion function for low rates corresponding to parameter values $s \in [-1, 0]$. This is illustrated in Fig. 2 where for $s = -1$ we define $R^* = R_L(D_{-1}, Q)$ and

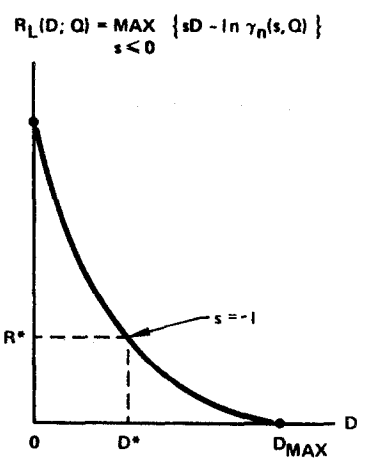

FIG. 2. Lower bound rate distortion function.

$D^* = D_{-1}$. For $R \leqslant R^*$, $D_n(Q)$ satisfies (9) where the exponent is already maximized with some $s \in [-1, 0]$.

We next examine how an optimum probability distribution is related to our discussion above. Recall that for each $s \in [-1, 0]$, a probability distri-

[1] $\gamma_n(s, Q)$ is defined for $s \in [-\infty, 0]$. It is an upper bound to error probabilities only for $s \in [-1, 0]$.

bution that minimizes the error bound satisfies the necessary condition (7),

$$[\gamma_n(s, Q_s^*)]^n \leqslant \sum_{\mathbf{x}} Q_s^*(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')} \tag{7}$$

with equality when $Q_s^*(\mathbf{x}') > 0$. Since an optimum distribution $\{Q_s^*(\mathbf{x}) : \mathbf{x} \in A^n\}$ depends on s, in the following discussion we consider source and representation alphabet given by A^n and let $\{Q(\mathbf{x}) : \mathbf{x} \in A^n\}$ be any source probability distribution. With the Bhattacharyya distance, we then have the rate distortion function $R_n(D; Q)$. Note that this rate distortion function differs from the previous one since in general $B(Q) \subset A^n$.

THEOREM.

$$R_n(D; Q) \leqslant \max_{s \leqslant 0} \{sD - \ln \gamma_n(s, Q_m^*)\},$$

where for each $s \in [-\infty, 0]$, $\{Q_s^*(\mathbf{x}) : \mathbf{x} \in A^n\}$ *satisfies* (7).

Proof. Consider

$$\begin{aligned}
\sum_{\mathbf{x}} Q(\mathbf{x}) \ln \lambda(\mathbf{x}) + \ln[\gamma_n(s, Q_s^*)]^n &= \sum_{\mathbf{x}} Q(\mathbf{x}) \ln\{\lambda(\mathbf{x})[\gamma_n(s, Q_s^*)]^n\} \\
&\leqslant \sum_{\mathbf{x}} Q(\mathbf{x})\{\lambda(\mathbf{x})[\gamma_n(s, Q_s^*)]^n - 1\} \\
&= \sum_{\mathbf{x}} Q(\mathbf{x})\, \lambda(\mathbf{x})[\gamma_n(s, Q_s^*)]^n - 1 \\
&\leqslant \sum_{\mathbf{x}} Q(\mathbf{x})\, \lambda(\mathbf{x}) \sum_{\mathbf{x}} Q_s^*(\mathbf{x}')\, e^{snd(\mathbf{x}',\mathbf{x})} - 1 \\
&= \sum_{\mathbf{x}'} Q_s^*(\mathbf{x}') \sum_{\mathbf{x}} Q(\mathbf{x})\, \lambda(\mathbf{x})\, e^{snd(\mathbf{x},\mathbf{x}')} - 1 \\
&\leqslant \sum_{\mathbf{x}'} Q_s^*(\mathbf{x}') - 1 \\
&= 0.
\end{aligned}$$

Here we used the inequalities

$$\ln x \leqslant x - 1,$$

$$[\gamma_n(s, Q_s^*)]^n \sum_{\mathbf{x}'} Q_s^*(\mathbf{x}')\, e^{snd(s,'\mathbf{x})} \qquad \forall \mathbf{x} \in A^n,$$

and

$$\sum_{\mathbf{x}} Q(\mathbf{x})\,\lambda(\mathbf{x})\,e^{snd(\mathbf{x},\mathbf{x}')} \leqslant 1 \qquad \forall \mathbf{x}' \in A^n.$$

Hence

$$\frac{1}{n}\sum_{\mathbf{x}} Q(\mathbf{x}) \ln \lambda(\mathbf{x}) \leqslant -\ln \gamma_n(s, Q_s{}^*)$$

for any $\lambda \in \Lambda_s$. ∎

This theorem applies to any distribution that satisfies the necessary condition of (7). Each such probability distribution yields a local minimum of $f_n(s, Q)$ and the condition $f_n(s, Q_s{}^*) = [\gamma_n(s, Q_s{}^*)]^n$. We now have a useful corollary which helps us in evaluating error exponents that may or may not be maximized.

COROLLARY. *There exists a code of blocklength n and rate R such that*

$$P_e \leqslant 2e^{-nD},$$

where D satisfies

$$R = R_n(D),$$

and $R_n(D)$ is any rate distortion function with source and representation alphabet A^n, Bhattacharyya distortion $d(\mathbf{x}, \mathbf{x}')$, and D corresponding to parameter values $s \in [-1, 0]$. In particular, D satisfies

$$R = \max_{Q} R_n(D; Q)$$

where the maximization is over all source probability distributions on A^n.

Proof. This follows from the previous theorem and the fact that if (D_1, Q_1) and (D_2, Q_2) satisfy

$$R = R_n(D_1; Q_1) = R_n(D_2; Q_2)$$

and at D_1

$$R_n(D_1; Q_1) \geqslant R_n(D_1; Q_2),$$

then

$$D_1 \geqslant D_2. \quad ∎$$

We now present a few examples of memoryless channels where

$$p(\mathbf{y} \mid \mathbf{x}) = \prod_{k=1}^{n} p(y_k \mid x_k)$$

and

$$d(\mathbf{x}, \mathbf{x}') = \frac{1}{n} \sum_{k=1}^{n} d(x_k , x_k'),$$

where

$$d(x_k , x_k') = -\ln \sum_{y_k} (p(y_k \mid x_k)\ p(y_k \mid x_k'))^{1/2}.$$

For this case, we restrict ourselves to choosing code word components independently, i.e., we assume

$$Q(\mathbf{x}) = \prod_{k=1}^{n} Q(x_k).$$

This is not necessary for any optimum distribution (see Jelinek (1968)) but generally simplifies our evaluation of $D_n(Q)$ and the corresponding $R_n(D; Q)$.

For the above assumptions, we now seek rate distortion functions for sources with alphabet $A = \{0, 1,..., K - 1\}$ and distortion

$$d(i, k) = -\ln \sum_{j=0}^{J-1} (p(j \mid i)\, p(j \mid k))^{1/2}$$

for each $i, k \in A$.

EXAMPLE. Binary Input Channels.

When $A = \{0, 1\}$, we have

$$d(i, k) = \begin{cases} 0, & i = k, \\ \alpha, & i \neq k, \end{cases}$$

where

$$\alpha = -\ln \sum_{j=1}^{J-1} (p(j \mid 0)\, p(j \mid 1))^{1/2}.$$

For the choice $Q(0) = Q(1) = 1/2$, we get the well-known rate distortion function (Berger (1971))

$$R_n(D; Q) = R(D) = \ln 2 - \mathscr{H}(D/\alpha),$$

where

$$\mathscr{H}(x) = -x \ln x - (1 - x) \ln(1 - x).$$

For all binary input memoryless channels, it turns out that $Q_s^*(\mathbf{x}) = 1/2^n$ is the optimum distribution. Hence, the tightest error exponent for such channels is D where D satisfies

$$R = \ln 2 - \mathscr{H}(D/\alpha)$$

for $R \leqslant \ln 2 - \mathscr{H}(1/(1 + e))$. This is also the optimum expurgated exponent.

EXAMPLE. Equidistant Channels.

Jelinek (1968) defined equidistant channels as channels where

$$d(i, k) = \begin{cases} 0, & i = k, \\ \alpha, & i \neq k. \end{cases}$$

For this channel, we also have the optimum probability distribution given by the uniform distribution $Q_s^*(\mathbf{x}) = 1/K^n$. The rate distortion function is

$$R_n(D; Q^*) = R(D) = \ln K - \mathscr{H}(D/\alpha) - (D/\alpha) \ln(K - 1).$$

Again the tightest error exponent satisfies

$$R = \ln K - \mathscr{H}(D/\alpha) - (D/\alpha) \ln(K - 1)$$

for low rates.

EXAMPLE. Balanced Channel.

We define a balanced channel as a channel with balanced Bhattacharyya distortion measure; i.e., $d(i, k)$ satisfies

$$\begin{aligned} \{d(i, k) : k \in A\} &= \{d(i, k) : i \in A\} \\ &= \{d_0, d_1, d_2, \ldots, d_{K-1}\} \end{aligned}$$

for all i, $k \in A$. For such channels, the uniform distribution $Q_s^*(\mathbf{x}) = 1/K^n$ gives a local minimum of $f(s, Q)$. For this choice, our error exponent $D_n(Q_s^*)$ coincides with the expurgated exponent for this distribution. Here

$$\gamma(s, Q^*) = \frac{1}{K} \sum_{k=0}^{K-1} e^{sd_k}$$

and the parametric equations for the rate distortion function are

$$R(D_s) = sD_s - \ln\left(\frac{1}{K}\sum_{k=0}^{K-1} e^{sd_k}\right)$$

$$D_s = \frac{\sum_{k=0}^{K-1} d_k e^{sd_k}}{\sum_{k=0}^{K-1} e^{sd_k}}.$$

Our error exponent satisfies $R = R(D)$ for rates corresponding to parameters $s \in [-1, 0]$. In this example, we only know that the resulting exponent is a lower bound to the largest error exponent.

4. Modular Channels and Convolutional Codes

Berger and Yu (1972) recently defined modular distortion measures for context-dependent fidelity criteria. For a class of channels that includes all binary input memoryless channels, the natural Bhattacharyya distances are modular and we present a simple proof of our error bound for linear codes transmitted over these channels. For this class of modular channels where the channel is also memoryless, the error bounds extend to convolutional codes.

We now assume that the channel input letters $A = \{0, 1, \ldots, q-1\}$[2] form a finite field with addition $\oplus$ and multiplication $\cdot$. A^n then forms a field of sequences of n letters from A. When we restrict ourselves to the binary input channel, $\oplus$ is the usual modulo -2 addition. Following Berger and Yu, we now define a mocular channel.

Definition. A channel is modular if the Bhattacharyya distance between any two elements $\mathbf{x}, \mathbf{x}' \in A^n$ is a function only of $\mathbf{x} \oplus \mathbf{x}' \in A^n$. That is

$$d(\mathbf{x}, \mathbf{x}') = -\frac{1}{n}\ln\sum_{\mathbf{y}}\left(p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{y} \mid \mathbf{x}')\right)^{1/2}$$

$$= d(\mathbf{x} \oplus \mathbf{x}') \qquad \forall \mathbf{x}, \mathbf{x}' \in A^n.$$

We see that the set $\{d(\mathbf{x}, \mathbf{x}') = d(\mathbf{x} \oplus \mathbf{x}') : \mathbf{x} \in A^n\}$ is the same for all $\mathbf{x}' \in A^n$, and hence modular channels are also balanced channels. This means that the distribution $Q^*(\mathbf{x}) = 1/q^n$ yields a local minimum of $f_n(s, Q)$ and that

[2] In this section, we use $K = q$ which is more common when discussing linear codes.

$f_n(s, Q^*) = [\gamma_n(s, Q^*)]^n$. Hence, $D_n(Q^*)$ is also the expurgated exponent for this probability distribution.

Throughout the rest of this section, we assume that the channel under consideration is modular. Suppose we pick p elements $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p \in A^n$ and denote $\mathscr{C}$ as the set of all linear combinations of these p basic elements. $\mathscr{C}$ is a linear code with $M = e^{nR} = q^p$ code words. We now have

LEMMA. $\sum_{\mathbf{x}' \in \mathscr{C}\, \mathbf{x}' \neq \mathbf{x}} e^{sn\hat{d}(\mathbf{x},\mathbf{x}')}$ *is independent of* $\mathbf{x} \in \mathscr{C}$.

Proof. By the linearity of $\mathscr{C}$, we have

$$\{d(\mathbf{x}, \mathbf{x}') = d(\mathbf{x} \oplus \mathbf{x}') : \mathbf{x} \in \mathscr{C}\} = \{d(\mathbf{x}, \mathbf{x}'') = d(\mathbf{x} \oplus \mathbf{x}'') : \mathbf{x} \in \mathscr{C}\}$$

for any $\mathbf{x}', \mathbf{x}'' \in \mathscr{C}$. Eliminating $d(\mathbf{x}', \mathbf{x}') = d(\mathbf{x}'', \mathbf{x}'') = 0$ from both sides does not change the equality of the two sets. ∎

Next consider an ensemble of linear codes of the above form where the basis elements are independently chosen according to the probability distribution $Q^*(\mathbf{b}) = 1/q^n$. This results in code words being pairwise independent and each nonzero code word having marginal probability distribution $Q^*(\mathbf{x}) = 1/q^n$.

THEOREM. *For modular channels there exist linear codes that satisfy*

$$P_e \leqslant e^{-nD_n(Q^*)},$$

where $Q^*(\mathbf{x}) = 1/q^n$.

Proof. Let $\mathscr{C}$ be any linear code. Then recall that for the mth code word being transmitted, we have from (3)

$$[P_{e,m}]^{-s} \leqslant \sum_{\substack{\mathbf{x}'_{m'} \in \mathscr{C} \\ m' \neq m}} e^{sn\hat{d}(\mathbf{x}_m, \mathbf{x}_{m'})}, \qquad s \in [-1, 0].$$

By our lemma, the upper bound is independent of $\mathbf{x}_m \in \mathscr{C}$ so

$$[P_{e,m}]^{-s} \leqslant \sum_{m'=2}^{M} e^{sn\hat{d}(\mathbf{x}_1, \mathbf{x}_{m'})}$$

for all $m = 1, 2, \ldots, M = e^{nR} = q^p$. Hence,

$$\max_m [P_{e,m}]^{-s} \leqslant \sum_{m'=2}^{M} e^{sn\hat{d}(\mathbf{x}_1, \mathbf{x}_{m'})}.$$

Here we take $\mathbf{x}_1 = \mathbf{O}$ as the zero code word. Averaging this over the ensemble of codes gives

$$\begin{aligned}\overline{\max_m [P_{e,m}]^{-s}} &\leqslant \sum_{m'=2}^{M} \sum_{\mathbf{x}} \frac{1}{q^n} e^{snd(x_1,\mathbf{x})} \\ &\leqslant M[\gamma_n(s, Q^*)]^n \\ &= e^{snD_n(Q^*)}.\end{aligned}$$

This means there exists a linear code where

$$\max_m [P_{e,m}]^{-s} \leqslant e^{snD_n(Q^*)}$$

or

$$P_{e,m} \leqslant e^{-nD_n(Q^*)}$$

for $m = 1, 2, \ldots, M$. Thus

$$P_e \leqslant e^{-nD_n(Q^*)}. \quad \blacksquare$$

This proof does not require expurgation of code words in a code or sequential selection of each code word as in our earlier proof in Section 2. It only requires ensemble pairwise independence of code words in a code. Thus we can use this bound for convolutional codes just as Viterbi and Odenwalder (1969) did for symmetric binary input memoryless channels. Following their approach, we now derive error bounds for convolutional codes used over a memoryless modular channel at low rates.

Let us assume a terminated L-branch tree generated by a ν-stage convolutional encoder whose input consists of L q-ary data symbols followed by $\nu - 1$ zeros. Assuming the maximum likelihood Viterbi decoding algorithm (Viterbi (1967), Forney (1967), Omura (1969), Viterbi (1971)), we consider the probability of error, $P_{e,m}$, when the mth sequence of L data symbols is transmitted. Letting $P_m(j)$ be the probability of eliminating the correct path at the jth step in the Viterbi algorithm, we get the union bound,

$$P_{e,m} \leqslant \sum_{j=1}^{L} P_m(j). \tag{10}$$

Eliminating the correct path at the jth step occurs if and only if some path that diverged from the correct path earlier and remerged again for the first time at the ith step has greater likelihood during the diverged interval.

Defining $A_m(j, l)$ as the error probability caused by some path that diverged from the correct path $\nu + l$ branches earlier and remerged for the first time after at step j, we have the union bound

$$P_m(j) \leqslant \sum_{l=0}^{j-1} A_m(j, l). \tag{11}$$

We now show that there exists a convolutional code for which $A_m(j, l)$ is upper bounded by our error bound for all choices of the transmitted sequence.

Assume a linear convolutional code where for each q-ary data input to the encoder there are b q-ary output symbols so that each branch of the tree of our code has b q-ary symbols. Hence, for the $\nu + l$ branches considered above, we have

$$n_l = (\nu + l)b$$

channel input symbols. Here we assume that there are q channel input alphabet symbols. Let $\mathbf{x}_m$ represent the n_l channel symbols corresponding to the correct path and $\mathbf{x}_{m'}$ the channel symbols that correspond to a path that diverged only during the interval of $\nu + l$ branches. Following our earlier arguments leading to (3), we have

$$[A_m(j, l)]^{-s} \leqslant \sum_{m' \neq m} e^{sn_l d(\mathbf{x}_m, \mathbf{x}_{m'})},$$
$$s \in [-1, 0].$$

Since our channel is memoryless and modular, we have

$$d(\mathbf{x}_m, \mathbf{x}_{m'}) = d(\mathbf{x}_m \oplus \mathbf{x}_{m'})$$
$$= \frac{1}{n_l} \sum_{k=1}^{n_l} d(\mathbf{x}_{m_k} \oplus x_{m_k'}).$$

The set $\{\mathbf{x}_m \oplus \mathbf{x}_{m'} : m' \neq m\}$ is the set of all possible difference sequences between the transmitted sequence and those sequences that diverge over the $\nu + l$ branches. Since the convolutional encoder is linear, this set is the set of all possible error sequences for paths that diverge over $\nu + l$ branches and it is independent of the particular sequence $\mathbf{x}_m$. Assuming that $\{\mathbf{x}_{m'}^0 : m' \neq 1\}$ correspond to those paths that diverge from the all-zero path for $\nu + l$ branches, we have

$$\max_{\mathbf{x}_m} [A(j, l)]^{-s} \leqslant \sum_{m'=1} e^{sn_l d(\mathbf{0}, \mathbf{x}_{m'}^0)} \tag{12}$$

where $\mathbf{x}_1^0 = \mathbf{O}$ is assumed. There are at most q^l distinct paths that diverge from the correct path for $\nu + l$ branches. We now consider an ensemble of time-varying convolutional encoders where the difference sequence $\mathbf{x}_{m'}^0$ has probability distribution

$$Q^*(\mathbf{x}_{m'}^0) = 1/q^{n_l}.$$

Averaging (12) over this ensemble, we get

$$\begin{aligned}
\overline{\max_{\mathbf{x}_m} [A(j,l)]^{-s}} &\leqslant \sum_{m' \neq 1} \sum_{\mathbf{x}} \frac{1}{q^{n_l}} e^{snd(\mathbf{0},\mathbf{x})} \\
&= \sum_{m' \neq 1} \sum_{\mathbf{x}} \frac{1}{q^{n_l}} \prod_{k=1}^{n_l} e^{sd(0,x_k)} \\
&\leqslant q^l \left\{ \sum_{k=0}^{q-1} \frac{1}{q} e^{sd(0,k)} \right\}^{n_l} \\
&= e^{sn_l D(l)}.
\end{aligned}$$

We know there exists a convolutional encoder such that for any transmitted data sequence we have

$$A_m(j,l) \leqslant e^{-n_l D(l)},$$

where

$$D(l) = \frac{1}{s} \left\{ \frac{l}{(\nu + l)b} \ln q + \ln \left(\sum_{k=0}^{q-1} \frac{1}{q} e^{sd(0,k)} \right) \right\}.$$

Note that $D(l)$ is the expurgated exponent for $\rho = -1/s$ and $Q = Q^*$ at rate $R_l = (l/(\nu + l)b) \ln q$. Recalling (10) and (11), we have

$$\begin{aligned}
P_e &\leqslant \max_m P_{e,m} \\
&\leqslant \sum_{j=1}^{L} \sum_{l=0}^{j-1} e^{-(\nu+l)bD(l)}.
\end{aligned}$$

Letting

$$E_x(s) = \frac{1}{s} \ln \left(\sum_{k=0}^{q-1} \frac{1}{q} e^{sd(0,k)} \right),$$

we have

$$\begin{aligned}(\nu + l)b\, D(l) &= l(\ln q/s) + (\nu + l)b\, E_x(s) \\ &= l\{(\ln q/s) + b\, E_x(s)\} + \nu b\, E_x(s).\end{aligned}$$

For the terminated tree code where $L \gg \nu$, the actual rate is given by $R = \ln q/b$ nats per channel symbol. Hence, as long as

$$\ln q/s + b\, E_x(s) > 0,$$

or equivalently,

$$E_x(s) + (1/s)\, R > 0,$$

the error probability satisfies

$$\begin{aligned}P_e &\leqslant \sum_{j=1}^{L} \sum_{l=0}^{j-1} e^{-l\{(\ln q/s)+bE_x(s)\}-\nu bE_x(s)} \\ &\leqslant (L/(1 - e^{-(E_x(s)+(1/s)R)}))\, e^{-\nu bE_x(s)}.\end{aligned}$$

$E_x(s)$ is a monotonically decreasing function of $s \in [-1, 0]$. For $\epsilon > 0$, choose s to satisfy

$$\epsilon = E_x(s) + (1/s)R,$$

and let

$$E_\epsilon(R) = E_x(s).$$

Then

$$P_e \leqslant (L/(1 - e^{-\epsilon}))\, e^{-\nu bE_\epsilon(R)},$$

where

$$R = -s[E_x(s) - \epsilon]$$

and

$$R < R_0 = -\ln\left(\sum_{k=0}^{q-1} \frac{1}{q}\, e^{-d(0,d)}\right).$$

For binary input memoryless channels, this gives the same error exponent obtained by Viterbi and Odenwalder (1969).

We conclude this section with several examples of memoryless modular channels.

Binary Input Channels

For binary input channels, we have

$$d(i, k) = \begin{cases} 0, & i = k, \\ \alpha, & i \neq k, \end{cases}$$

where

$$\alpha = -\ln \sum_{j=0}^{J-1} (p(j \mid 0)\, p(j \mid 1))^{1/2}.$$

For continuous output channels, α is simply

$$\alpha = -\ln \int (p(y \mid 0)\, p(y \mid 1))^{1/2}\, dy.$$

Clearly, over the binary field we have

$$d(i, k) = \alpha(i \oplus k),$$

and all binary input channels are modular and the exponent is

$$E_x(s) = (1/s) \ln((1 + e^{s\alpha})/2).$$

Equidistant Channels

For equidistant channels we have

$$d(i, k) = \begin{cases} 0 & i = k, \\ \alpha & i \neq k, \end{cases} \qquad i, k \in A = \{0, 1, \ldots, q - 1\}.$$

Letting

$$\delta(i \oplus k) = \begin{cases} 0 & i \oplus k = 0, \\ 1 & i \oplus k \neq 0, \end{cases}$$

we have

$$d(i, k) = \alpha\delta(i \oplus k)$$

which is modular. The error exponent is

$$E_x(s) = (1/s) \ln[(1 + (q - 1)\, e^{s\alpha})/q].$$

q orthogonal or simplex pulses over a white gaussian noise channel is a

typical example of an equidistant channel. For orthogonal pulses of energy E and a white gaussian noise channel of spectral density $N_0/2$, we have

$$\alpha = E/2N_0 .$$

The fading channels examined by Kennedy (1969) are also equidistant and hence modular.

Uniform Phase Modulation

Suppose q inputs to a white gaussian noise channel are represented by phase-modulated pulses where the phase angles are equally spaced on a unit circle. Let the pulses be $s_k(t)$, $k = 0, 1, \ldots, q-1$. Then

$$\begin{aligned} d(i,k) &= \| s_i - s_k \|^2/4N_0 \\ &= \frac{1}{4N_0} \int | s_i(t) - s_k(t)|^2 \, dt. \end{aligned}$$

Since $\| s_i - s_k \|^2$ depends only on $i \oplus k$, this channel is also modular. In particular, let

$$s_k(t) = \sqrt{E} \cos k(2\pi/q)\, \phi_1(t) + \sqrt{E} \sin k(2\pi/q)\, \phi_2(t),$$

where $\phi_1(t)$ and $\phi_2(t)$ are orthonormal carriers. Then

$$\| s_i - s_k \|^2 = 2E(1 - \cos[(k \oplus i)(2\pi/q)])$$

and

$$E_x(s) = \frac{1}{s} \ln \left(\sum_{k=0}^{q-1} \frac{1}{q} e^{s(E/2N_0)[1-\cos k(2\pi/q)]} \right).$$

5. Discussion

For low rates we have shown a close relationship between expurgated bounds and the natural Bhattacharyya distances defined by the channel. These upper bounds were then extended to convolutional codes for the class of modular channels.

The Bhattacharyya distance is a special case of a more general distance measure between code words. For the pair-wise reversible channels, this general distance measure coincides with the Bhattacharyya distance which

becomes the Hamming distance measure for the binary symmetric channel. We have recently shown (Omura (1973)) that for memoryless channels, this general distance measure is related to upper and lower bounds on the low rate error probability in the same way as with Hamming distance for binary block codes used over the binary symmetric channel. We then proved a general Gilbert bound for block codes using this distance measure. We hope that the relationships between these distances, error bounds at low rates, and rate distortion functions will give new insights into the performance of codes for memoryless channels. We conjecture that for the pair-wise reversible memoryless channel the expurgated bounds discussed in this paper are exponentially tight.

RECEIVED: August 18, 1972; REVISED: October 10, 1973

REFERENCES

BERGER, T. (1971), "Rate Distortion Theory: A Mathematical Basis for Data Compression," Chap. 2, Prentice-Hall, Englewood Cliffs, NJ.

BERGER, T., AND YU, W. C. (1972), Rate-distortion theory for context-dependent fidelity criteria, *IEEE Trans. Information Theory*, **18**, 378–384.

BERLEKAMP, E. R. (1964), Block coding with noiseless feedback, Ph.D. Thesis, Department of Electrical Engineering, MIT.

FORNEY, G. D. (1967), Final report on a coding system for advanced solar missions, Contract NASA-3637, submitted by Codex Corp.

GALLAGER, R. G. (1965), A simple derivation of the coding theorem and some applications, *IEEE Trans. Information Theory* **11**, 3–18.

GALLAGER, R. G. (1968), "Information Theory and Reliable Communication," Chap. 9, Wiley, New York.

JELINEK, F. (1968), Evaluation of expurgated bound exponents, *IEEE Trans. Information Theory* **14**, 501–505.

KAILATH, T. (1967), The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Communication Technology* **15**, 52–60.

KENNEDY, R. S. (1969), "Fading Dispersive Communication Channels," Wiley, New York.

OMURA, J. K. (1969), On the Viterbi decoding algorithm, *IEEE Trans. Information Theory* **15**, 177–179.

OMURA, J. K. (1973), On general Gilbert bounds, *IEEE Trans. Information Theory* **19**, 661–665.

SHANNON, C. E. (1956), The zero error capacity of a noisy channel, *IRE Trans. Information Theory* **2**, 8–19.

SHANNON, C. E. (1959), Coding theorems for a discrete source with a fidelity criterion, *IRE Nat. Conv. Rec.*, Part 4, pp. 142–163; also in R. E. MACHOL, Ed. (1960), "Information and Decision Processes" pp. 93–126, McGraw–Hill, New York.

SHANNON, C. E., GALLAGER, R. G., AND BERLEKAMP, E. R. (1967), Lower bounds to error probability for coding on discrete memoryless channels. Parts I and II, *Information and Control* **10**, 65–103 (Part I), pp. 522–552 (Part II).

VITERBI, A. J. (1967), Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Information Theory* **13**, 260–269.

VITERBI, A. J., AND ODENWALDER, J. P. (1969), Further results on optimal decoding of convolutional codes, *IEEE Trans. Information Theory* **15**, 732–734.

VITERBI, A. J. (1971), Convolutional codes and their performance in communication systems, *IEEE Trans. Communication Technology* **19**, 751–772.