INFORMATION AND CONTROL 27, 272–288 (1975)

Rate Distortion Functions of Countably Infinite Alphabet Memoryless Sources*

H. TAN†

Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08540

AND

K. YAO‡

System Science Department, School of Engineering and Applied Science, University of California, Los Angeles, Los Angeles, California 90024

The Shannon lower bound approach to the evaluation of rate distortion functions $R(D)$ for countably infinite alphabet memoryless sources is considered. Sufficient conditions based on the Contraction Mapping Theorem for the existence of the Shannon lower bound $R_L(D)$ to $R(D)$ in a region of distortion $[0, D_1]$, $D_1 > 0$ are obtained. Sufficient conditions based on the Schauder Fixed Point Theorem for the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$ are derived. Explicit evaluation of $R(D)$ is considered for a class of column balanced distortion measures. Other results for distortion measures with no symmetry conditions are also discussed.

1. INTRODUCTION

The rate-distortion function $R(D)$ of a source represents its equivalent rate subject to a fidelity criterion. Recently many results have appeared on the evaluation of $R(D)$ for various sources and distortion measures (Jelinek, 1967; Gallager, 1968; Pinkston, 1969; Berger, 1970, 1971; Wyner-Ziv,

* This research was supported by the Office of Naval Research under Task No. NR-042-285 and Contract No. N00014-69-a-0200-4043 and National Science Foundation Grant GK-42080.

† Formerly with System Science Department, School of Engineering and Applied Science, University of California, Los Angeles.

‡ On leave at Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, from September 1973 to April 1974.

1971; Gray, 1970, 1971a, 1971b, 1973a, 1973b; Rubin, 1973). Explicit evaluation of $R(D)$ for most sources and distortion measures remains a formidable task although a general method for the numerical computation of $R(D)$ for finite-alphabet memoryless sources is available (Blahut, 1972). A common approach to circumvent this difficulty is to derive the Shannon lower bound $R_L(D)$ to $R(D)$ and then to find conditions on source statistics and distortion measure for the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all values of distortion $D \in [0, D_c]$. Sufficient conditions for the existence of such a $D_c > 0$ have been derived for finite-alphabet discrete-time memoryless sources by Jelinek (1967) and Pinkston (1969), for certain finite-alphabet discrete-time sources with memory by Gray (1971a, 1971b, 1973a) and for certain Gaussian sources with memory under square-error distortion criterion by Berger (1970) and Gray (1970).

In this paper we consider the Shannon lower bound approach to the evaluation of rate-distortion functions $R(D)$ for countably infinite alphabet discrete-time memoryless sources. Specifically, in Section 2 we derive sufficient conditions for the existence of the Shannon lower bound $R_L(D)$ to $R(D)$ in a region of distortion $[0, D_1]$, where $D_1 > 0$. In Section 3, sufficient conditions for the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all distortion D in $[0, D_c]$ are derived. In Section 4, we consider in detail the evaluation of rate-distortion functions under a class of column balanced distortion measures for a range of distortion $[0, D_c]$, $D_c > 0$. In Section 5, we discuss a possible application of the results in Sections 2 and 3 to distortion measures with no symmetry conditions. Conclusions are given in Section 6.

2. The Shannon Lower Bound

We consider the source $\{X_t : t = 0, 1, 2,...\}$ to be a sequence of i.i.d. random variables taking on values in the source alphabet A. A will be taken to be either $I = \{0, 1, 2,...\}$ or $J = \{..., -1, 0, 1,...\}$. Let X_t have probability distribution $p(j) = \Pr[X_t = j]$, $j \in A$. Let the reproduced source alphabet be equal to the source alphabet A and the distortion measure $\rho: A \times \hat{A} \rightarrow [0, \infty)$ satisfy

$$\rho(j, k) > \rho(j, j) = 0 \qquad \forall j \neq k, j, k \in A. \tag{1}$$

As usual, the assumption $\rho(j, j) = 0$ involves no loss of generality (p. 26 of Berger, 1971). Moreover, since for $j \neq k$, $\rho(j, k)$ can be made arbitrarily small, Eq. (1) allows a broad class of distortion measures. Let Q_D be the set of all conditional probability distrituions $q(k \mid j) = \Pr[$output reproduced

$k \mid$ source produced $j]$, $j, k \in A$, such that $\sum_{k \in A} \sum_{j \in A} \rho(j, k)\, p(j)\, q(k \mid j) \leqslant D$ for each $D \geqslant 0$. Then the rate-distortion function $R(D)$ of the source $\{X_t\}$ with respect to the distortion measure ρ is defined by (Berger, 1971, p. 23)

$$R(D) = \inf_{q \in Q_D} I(q), \tag{2}$$

where

$$I(q) = \sum_{k \in A} \sum_{j \in A} q(k \mid j)\, p(j) \log \frac{q(k \mid j)}{\hat{q}(k)} \qquad \text{and} \qquad \hat{q}(k) = \sum_{j \in A} q(k \mid j)\, p(j).$$

We note that if there exists a $k^* \in A$ such that $\sum_{j \in A} p(j)\, \rho(j, k^*) < \infty$, then appropriate source coding theorems exist to give $R(D)$ as defined by Eq. (2) the desired operational significance (p. 281 of Berger, 1971).

The following theorem is useful for the evaluation of $R(D)$. This theorem is equivalent to Theorem 9.4.1 of Gallager (1968) and Theorem 2.5.3 of Berger (1971).

THEOREM 1. *For each $s \leqslant 0$, let Λ_s be the set of all sequences $\{\lambda_s(j): j \in A\}$ of nonnegative numbers satisfying*

$$\sum_{j \in A} \lambda_s(j)\, p(j) \exp(s\rho(j, k)) \leqslant 1 \qquad \forall k \in A. \tag{3}$$

Then for each $D \geqslant 0$, we have

$$R(D) = \sup_{s \leqslant 0,\, \{\lambda_s(j)\} \in \Lambda_s} \left[sD + \sum_{j \in A} p(j) \log \lambda_s(j) \right]. \tag{4}$$

Moreover, for each $s \leqslant 0$, a necessary and sufficient condition for $\{\lambda_s(j)\}$ to achieve the supremum in Eq. (4) *is the existence of a probability distribution $\{Q_s(k): k \in A\}$ such that*

$$\sum_{k \in A} Q_s(k) \exp(s\rho(j, k)) = [\lambda_s(j)]^{-1}, \qquad \forall j \in A, \tag{5}$$

and such that equality holds in Eq. (3) *for all $k \in A$ such that $Q_s(k) > 0$.*

Let $\mu_s(j) = \lambda_s(j)\, p(j)$ in Theorem 1 and suppose we require Eq. (3) to hold with equality for all $k \in A$. If, for each $s \leqslant 0$, there exists a sequence $\{\mu_s(j): j \in A\}$ of nonnegative numbers satisfying

$$\sum_{j \in A} \mu_s(j) \exp(s\rho(j, k)) = 1, \qquad \forall k \in A, \tag{6}$$

then for each $D \geqslant 0$, $R(D) \geqslant R_L(D)$, where $R_L(D)$ is called the Shannon lower bound to $R(D)$ and is given by

$$R_L(D) = \sup_{s \leqslant 0} \left[sD + \sum_{j \in A} p(j) \log \mu_s(j) + H(X) \right], \tag{7}$$

where the entropy rate of the source $H(X) = \sum_{j \in A} p(j) \log p(j)$ is assumed to be finite. For many probability distributions and distortion measures, the quantity to be maximized in the right-hand side of Eq. (7) is a strictly concave function of s (Berger, 1971, p. 92; and Jelinek, 1967). We will implicitly assume that this is true. Thus the supremum in Eq. (7) can be solved by differentiating the right-hand side of Eq. (7) with respect to s and setting it equal to zero, yielding the following parametrically defined equations for $R_L(D)$,

$$R_L(D_s) = sD_s + H(X) + \sum_{j \in A} p(j) \log \mu_s(j), \tag{8}$$

$$D_s = -\frac{\partial}{\partial s} \left\{ \sum_{j \in A} p(j) \log \mu_s(j) \right\}. \tag{9}$$

We note that $R_L(D)$ is well defined only for those values of distortion D_s in Eq. (9) with values of the parameter s for which there exist a sequence of nonnegative numbers $\{\mu_s(j): j \in A\}$ satisfying Eq. (6). Moreover, from Theorem 1 we conclude that $R(D)$ is equal to $R_L(D)$ for those values of distortion D_s in Eq. (9) with values of the parameter s for which there exist a sequence of nonnegative numbers $\{Q_s(k): k \in A\}$ satisfying Eq. (5). We note that the only condition required on this sequence is that it be a sequence of nonnegative numbers since any sequence $\{Q_s(k): k \in A\}$ satisfying Eq. (5) will sum to one provided that $\lambda_s(j) = \mu_s(j)/p(j)$ and $\{\mu_s(j): j \in A\}$ satisfies Eq. (6). Using Eqs. (8) and (9) and an argument similar to the proof of Theorem 2.5.1 of Berger (1971), it can be shown that the parameter s is the derivative of $R_L(D)$ at $D = D_s$. We note that by assumption, the quantity to be maximized on the right-hand side of Eq. (7) is strictly concave in s, which along with Eq. (9) implies that $dD_s/ds > 0$ for all values of s. We also note that for $s = -\infty$, since the distortion function ρ satisfies Eq. (1), the necessary and sufficient conditions given in Theorem 1 are satisfied by $\lambda_s(j) = 1/p(j)$ for all $j \in A$, which implies that $R(0) = R_L(0)$ and $D_s = 0$ at $s = -\infty$. Thus an interval of the parameter s, $(-\infty, s^*]$, where $s^* > -\infty$, corresponds to an interval of distortion D_s, $(0, D^*]$, where $D^* = D_{s^*} > 0$.

In order to evaluate $R(D)$ by the Shannon lower bound approach, conditions must be derived for the existence of $R_L(D)$ and for the existence of a $D_c > 0$

such that $R(D) = R_L(D)$ for all D in $[0, D_c]$. For finite-alphabet memoryless sources, Pinkston (1969) and Jelinek (1967) have shown that a sufficient condition for the existence of such a $D_c > 0$ is that the distortion measure satisfies Eq. (1) and that $\Pr[X_t = j] > 0$ for all letters j in the source alphabet. Jelinek's proof depends on the fact that under these assumptions, the finite matrix $\{\exp(s\rho(j, k))\}^{-1}$ exists and approaches the identity matrix I in the limit as s approaches $-\infty$. An elementary perturbation argument is then successfully applied due to the finite dimensionality of the matrix $\{\exp(s\rho(j, k))\}$. However, for the countably infinite alphabet case, such an argument fails due to the infinite dimensionality of the matrix $\{\exp(s\rho(j, k))\}$.

In Theorem 2 below, we will give sufficient conditions on the distortion measure ρ for the existence of a $D_1 > 0$ such that $R_L(D)$ is well-defined for all D in $[0, D_1]$. The proof of this theorem based on the Contraction Mapping Theorem (Luenberger, 1969, p. 272) is somewhat involved and is given in Appendix 1.

THEOREM 2. *Suppose the distortion measure ρ satisfies Eq.* (1) *and that there exists a $s^* \in (-\infty, 0)$ such that*

$$\sup_k \sum_{j \in A} \exp(s^* \rho(j, k)) < \infty. \tag{10}$$

For $s \leqslant s^*$, define

$$\alpha(s) = \sup_k \sum_{\substack{j \in A \\ j \neq k}} \exp(s\rho(j, k)). \tag{11}$$

Then there exists a $s_1 \in (-\infty, s^*]$ such that $\alpha(s_1) < 1$ and such that there exists a unique sequence $\{\mu_s(j) : j \in A\}$ of nonnegative numbers satisfying Eq. (6) *for each $s \leqslant s_1$. Moreover we have $1 - \alpha(s) \leqslant \mu_s(j) \leqslant 1$ for all $j \in A$ and $s \leqslant s_1$. Thus the Shannon lower bound $R_L(D)$ exists for all $D \in [0, D_1]$, where $D_1 > 0$ is given by Eq.* (9) *with $s = s_1$.*

It is interesting to note that the hypothesis of Eq. (10) prohibits all bounded distortion measures when A is not finite but allows all distortion measures which grow as fast as $\ln(|j - k|^a)$ for some $a > 0$.

In order to evaluate $R(D)$ by the Shannon lower bound approach, conditions must be obtained for $R(D)$ to be equal to $R_L(D)$. In the next section we derive sufficient conditions for the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$.

3. Region of Equality

From previous discussion in Section 2, $R(D) = R_L(D)$ for those values of distortion D_s in Eq. (9) with values of the parameter s for which there exists a probability distribution $\{Q_s(k): k \in A\}$ satisfying

$$\sum_{k \in A} \exp(s\rho(j, k))\, Q_s(k) = p(j)/\mu_s(j), \qquad \forall j \in A, \tag{12}$$

where $\{\mu_s(j): j \in A\}$ is the sequence of nonnegative numbers satisfying Eq. (6). In Theorem 3 we show the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$, by showing the existence of a $s_c \in (-\infty, s_1]$ such that Eq. (12) has a nonnegative solution $\{Q_s(k): k \in A\}$ for each $s \leqslant s_c$. The Contraction Mapping Theorem was used in the proof of Theorem 2 to show the existence of a nonnegative solution to Eq. (6). Here, we need to use the Schauder Fixed Point Theorem (p. 96 of Schwartz, 1969). The proof of Theorem 3 is somewhat involved and is given in Appendix 2.

Theorem 3. *Let the distortion measure ρ satisfy the hypothesis of Theorem 2, let s_1 be defined as in Theorem 2 and for each $s \leqslant s_1$, let $\{\mu_s(j): j \in A\}$ be the nonnegative solution of Eq. (6). Suppose there exists a $s_c \in (-\infty, s_1]$ such that for all $s \leqslant s_c$,*

$$\sum_{\substack{k \in A \\ k \neq j}} \exp(s\rho(j, k))\, p(k)/\mu_s(k) \leqslant p(j)/\mu_s(j), \qquad \forall j \in A. \tag{13}$$

Then for each $s \leqslant s_c$, there exists a probability distribution $\{Q_s(k): k \in A\}$ satisfying Eq. (12). Thus $R(D) = R_L(D)$ for all $D \in [0, D_c]$, where $D_c > 0$ is given by Eq. (9) with $s = s_c$.

Theorems 2 and 3 together have given sufficient conditions for the evaluation of $R(D)$ in a range of distortion $[0, D_c]$, $D_c > 0$, for countably infinite alphabet sources without memory. In particular, for finite-alphabet sources, hypotheses (10) and (13) of Theorems 2 and 3 are always satisfied, and thus results of Theorems 2 and 3 include the known results of Jelinek (1967). Of course, Theorems 2 and 3 are only existence theorems. In order to explicitly evaluate $R(D)$, the Shannon lower bound $R_L(D)$ must first be evaluated. Thus the nonnegative solution $\{\mu_s(j): j \in A\}$ to Eq. (6) must be obtained. Unfortunately, explicit evaluation of the solution does not appear to be possible in general due to the infinite dimensionality of the problem. In the next section we indicate a class of distortion measures for which the explicit nonnegative solution of Eq. (6) may be obtained.

4. Column Balanced Distortion Measures

We will define a class of distortion measures with symmetry conditions such that the explicit nonnegative solution to Eq. (6) can be easily obtained. Then sufficient conditions dependent only on source statistics and distortion measure are given for the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$.

Definition 1. The distortion measure ρ is said to be column balanced if the sets $A_k = \{\rho(j, k): j \in A\}$ are identical for every $k \in A$. For the case $A = J = \{..., -1, 0, 1,...\}$, the class of all column balanced distortion measures contains the class of all difference distortion measures, that is, the class of distortion measures of the form $\rho(j, k) = \rho(j - k)$. Thus, for this case, the class of column balanced distortion measures is sufficiently rich for applications. For the case $A = I = \{0, 1, 2,...\}$, the condition of column balanced appears to be fairly restrictive, yielding distortion measures which are not necessarily realistic for applications. For $A = I$, we give two cases of column balanced distortion measures below.

Example 1 (Column Balanced Distortion Measures).

(a)
$$\rho(j, k) = \begin{cases} \rho_j & \text{if } j < k \\ 0 & \text{if } j = k \\ \rho_{j-1} & \text{if } j > k \end{cases}$$

(b)
$$\rho(j, k) = \begin{cases} \rho_{k-j} & \text{if } j < k \\ 0 & \text{if } j = k \\ \rho_0 & \text{if } j = k + 1 \\ \rho_{j-1} & \text{if } j > k + 1. \end{cases}$$

The following lemma gives the solution to Eq. (6) when the distortion measure ρ is column balanced.

Lemma 1. *Let the distortion measure ρ be column balanced and suppose there exists a $s^* \in (-\infty, 0)$ such that $\sum_{j \in A} \exp(s^*\rho(j, 0)) < \infty$. Then for each $s \leqslant s^*$, the sequence $\{\mu_s(j): j \in A\}$ of nonnegative numbers given by*

$$\mu_s(j) = \Big[\sum_{m \in A} \exp(s\rho(m, 0))\Big]^{-1}, \qquad \forall j \in A \tag{14}$$

is a solution of Eq. (6).

Proof. It suffices to show that $\sum_{j\in A} \exp(s\rho(j, k)) = \sum_{j\in A} \exp(s\rho(j, 0))$ for each $s \leqslant s^*$ and $k \neq 0$. By the definition of column balanced, the series $\sum_{j\in A} \exp(s\rho(j, k))$ is a rearrangement (Rudin, 1964, p. 66) of the series $\sum_{j\in A} (s\rho(j, 0))$ for each $k \neq 0$. Since $\sum_{j\in A} \exp(s\rho(j, 0))$ converges absolutely for each $s \leqslant s^*$, by a well-known theorem (Rudin, 1964, p. 68) on rearrangements of series, every rearrangement of $\sum_{j\in A} \exp(s\rho(j, 0))$ converges to the same sum. This completes the proof of the lemma. Q.E.D.

For the remainder of this section we assume that the hypothesis of Lemma 1 is always satisfied. In order that the Shannon lower bound $R_L(D)$ be given by Eq. (8) and (9) we must determine if the quantity to be maximized in the right-hand side of Eq. (7) is strictly concave in s, that is to show that

$$m(s) = sD + H(X) + \sum_{j\in A} p(j) \log \Big(\sum_{j\in A} \exp(s\rho(j, 0))\Big)$$

is strictly concave in s for all $s \in (-\infty, s^*]$. This is established in the next lemma.

LEMMA 2. *Under hypothesis of Lemma* 1, $m(s)$ *is strictly concave for* $s \in (-\infty, s^*]$.

Proof. It suffices to prove that $m''(s) < 0$ for every $s \in (-\infty, s^*)$. Note that the series $\sum_{j\in A} \exp(s\rho(j, 0))$ converges uniformly on $(-\infty, s^*]$, and that $\sum_{j\in A} \rho(j, 0) \exp(s\rho(j, 0))$ converges whenever $\sum_{j\in A} \exp(s\rho(j, 0))$ converges (Hardy and Riesz, 1915, p. 5) and thus by a well-known theorem on differentiation (Rudin, 1964, p. 140), we have

$$\frac{d}{ds}\Big[\sum_{j\in A} \exp(s\rho(j, 0))\Big] = \sum_{j\in A} \rho(j, 0) \exp(s\rho(j, 0)) \qquad \text{for all } s \in (-\infty, s^*).$$

Similarly, we also have

$$\frac{d}{ds}\Big[\sum_{j\in A} \rho(j, 0) \exp(s\rho(j, 0))\Big] = \sum_{j\in A} \rho^2(j, 0) \exp(s\rho(j, 0))$$

for $s \in (-\infty, s^*)$. Thus we have

$$m''(s) = -\Big\{\sum_{j\in A} \rho^2(j, 0) \frac{\exp(s\rho(j, 0))}{\sum_{m\in A} \exp(s\rho(m, 0))} - \Big[\sum_{j\in A} \rho(j, 0) \frac{\exp(s\rho(j, 0))}{\sum_{m\in A} \exp(s\rho(m, 0))}\Big]^2\Big\}$$

or equivalently $m''(s) = -\mathrm{Var}(\rho(Z_s, 0))$, where Z_s is a random variable with distribution

$$\Pr[Z_s = j] = \frac{\exp(s\rho(j, 0))}{\sum_{m\in A} \exp(s\rho(m, 0))}, \qquad j \in A.$$

Since $\rho(j, 0) > 0$ for all $j \neq 0$ and $\Pr[Z_s = j] > 0$ for all $j \in A$, we conclude that $m''(s) < 0$ for all $s \in (-\infty, s^*)$. This completes the proof of the lemma.
Q.E.D.

Thus substitution of Eq. (14) into Eqs. (8) and (9) yield the following parametric equations for $R_L(D)$,

$$R_L(D_s) = H(X) + sD_s - \ln\Big(\sum_{j\in A} \exp(s\rho(j, 0))\Big), \tag{15}$$

$$D_s = \sum_{j\in A} \rho(j, 0) \frac{\exp(s\rho(j, 0))}{\sum_{m\in A} \exp(s\rho(m, 0))}, \tag{16}$$

where $s \leqslant s^*$. We note the general forms of Eqs. (15) and (16) are similar to those in the finite-alphabet case. We can now use Theorem 3 to obtain sufficient conditions for the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$. Note that the hypothesis of Theorem 3 contained also the hypothesis of Theorem 2. However, in the case of column balanced distortion measures, the hypothesis of Theorem 2 can be replaced by the hypothesis of Lemma 1 and Eq. (1) and s_1 replaced by s^*. The proof of Theorem 3 will still carry through replacing the role of f_s by $[\sum_{j\in A} \exp(s\rho(j, 0))]\, \underline{p}$, where $\underline{p} = \{p(j): j \in A\}$. We state this special case of Theorem 3 below as Theorem 4.

THEOREM 4. *Let the distortion measure ρ be column balanced, satisfy Eq.* (1) *and the hypothesis of Lemma* 1. *Suppose there exists a $s_c \in (-\infty, s^*]$ such that*

$$\sum_{\substack{k\in A\\ k\neq j}} \exp(s_c\rho(j, k))\, p(k) \leqslant p(j), \qquad \forall j \in A. \tag{17}$$

Then $R(D) = R_L(D)$ for all $D \in [0, D_c]$, where $R_L(D)$ is given by Eqs. (15) *and* (16) *and $D_c > 0$ by Eq.* (16) *with $s = s_c$.*

In the finite-alphabet case necessary and sufficient conditions for the existence of a $D_c > 0$ such that $R(D) = R_L(D)$, $\forall D \in [0, D_c]$ is that $p(j) > 0$, $\forall j$. In the countably infinite alphabet case, the condition $p(j) > 0$, $\forall j \in A$ is a necessary condition but not a sufficient condition (Rubin, 1973). We now consider some examples using Theorem 4 to determine the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$.

EXAMPLE 2. Consider the case $A = J$, the distortion measure ρ given by $\rho(j, k) = |j - k|^\nu$, where $\nu \geqslant 1$ and the source distribution $\{p(j): j \in J\}$ given by

$$p(j) = \frac{\theta(1-\theta)^{|j|}}{2-\theta}, \qquad 0 < \theta < 1, \quad \forall j \in J.$$

This is the distribution of the difference of two i.i.d. Geometrically distributed random variables. For this example, $R_L(D)$ is given by

$$R_L(D_s) = -\log\left(\frac{\theta}{2-\theta}\right) - \frac{2(1-\theta)}{(2-\theta)}\log(1-\theta)$$
$$+ sD_s - \log\left(1 + 2\sum_{j=1}^{\infty}\exp(sj^\nu)\right), \tag{18}$$

$$D_s = \frac{2\sum_{j=1}^{\infty} j^\nu \exp(sj^\nu)}{1 + 2\sum_{j=1}^{\infty}\exp(sj^\nu)}, \tag{19}$$

where $s < 0$. It can be easily shown that Eq. (17) is satisfied if s_c is given by

$$s_c = \log\left[\min\left\{\frac{1-\theta}{2(1+(1-\theta)^2)}, \frac{1}{3(1-\theta)}\right]\right\}. \tag{20}$$

Thus $R(D) = R_L(D)$ for all $D \in [0, D_{s_c}]$, where D_{s_c} given by Eq. (19) with s_c given by Eq. (20).

EXAMPLE 3. Consider the case $A = I$, the distortion measure ρ given by

$$\rho(j, k) = \begin{cases} \log[(j+2)!], & j < k, \\ 0, & j = k, \\ \log[(j+1)!], & j > k, \end{cases}$$

and the source distribution be Poisson with parameter λ. For this example, $R_L(D)$ is given by

$$R_L(D_s) = -\lambda - \lambda\log\lambda + e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k \log(k!)}{k!}$$
$$+ sD_s - \log\left(\sum_{k=1}^{\infty}(k!)^s\right), \tag{21}$$

$$D_s = \frac{\sum_{k=1}^{\infty} (k!)^s \log(k!)}{\sum_{k=1}^{\infty} (k!)^s}, \tag{22}$$

where $s < 0$. It can be shown that Eq. (17) is satisfied if s_c is given by

$$s_c = -1 - \lambda/\log 2. \tag{23}$$

Thus $R(D) = R_L(D)$ for all $D \in [0, D_{s_c}]$, where D_{s_c} is given by Eq. (22) with s_c given by Eq. (23).

These two examples show that the conditions in the hypothesis of Theorem 4 lead to nonvaccuous results.

5. Distortion Measures with No Symmetry Conditions

In the last section, we demonstrated the applicability of Theorem 3 to determine the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$. The primary constraint there was the assumption that the distortion measure be column balanced which was seen to be quite restrictive for the source alphabet $A = I$. This assumption was made in order to obtain the explicit solution of Eq. (6). It may be possible to obtain the explicit solution of Eq. (6) for some specific distortion measures which are not column balanced. Then Theorem 3 can still be used to determine the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$. An example of this situation is given by the case when $A = I$ and $\rho(j, k) = |j - k|$. It can be shown that the explicit solution of Eq. (6) for $s < 0$ is given by

$$\mu_s(j) = \begin{cases} \dfrac{1}{1 + e^s}, & j = 0, \\ \dfrac{1 - e^s}{1 + e^s}, & j > 0, \end{cases} \tag{24}$$

substitution of which into Eqs. (8) and (9) given the following parametric equations for $R_L(D)$,

$$R_L(D_s) = sD_s + H(X) - \log(1 + e^s) + (1 - p(0)) \log(1 - e^s), \tag{25}$$

$$D_s = \frac{e^s}{1 + e^s} + (1 - p(0)) \frac{e^s}{1 - e^s}, \tag{26}$$

where $s < 0$. Theorem 3 can now be used to determine the existence of a $D_c > 0$ such that $R(D) = R_L(D)$ for all $D \in [0, D_c]$.

Consider the case where the source distribution is Geometric, that is $p(j) = \theta(1-\theta)^j$, $0 < \theta < 1$, $j \in I$. Using Eq. (24), it can be shown for this example that Eq. (13) is satisfied for all $s \leqslant s_c$, where s_c is given by

$$e^{s_c} = \frac{(1-\theta)(1-(1-\theta)^2/2)}{2+(1-\theta)+2(1-\theta)^2-(1-\theta)^3/2}, \tag{27}$$

which implies that $R(D) = R_L(D)$ for all $D \in [0, D_{s_c}]$, where D_{s_c} is given by Eq. (26) with s_c by Eq. (27). This example demonstrates that the hypothesis of Theorem 3 leads to nonvaccuous results even when the distortion measure is not column balanced.

Thus far, we have considered only the problem of explicit evaluation of $R(D)$. If only numerical values of $R(D)$ are required, Theorems 2 and 3 may be used to obtain numerical values of $R(D)$ for a region of small distortion $(0, D_c]$, $D_c > 0$. We note that numerical values of $R_L(D)$ can be obtained by finding the numerical solution of Eq. (6). We also note Blahut's algorithm is valid for finite-alphabet case, but in the countably infinite case not much is known about that algorithm.

The proof of existence of a nonnegative solution to Eq. (6) for values of small distortion in Theorem 2 used the Contraction Mapping Theorem. It is well known that the fixed point guaranteed by the Contraction Mapping Theorem may be obtained by the method of successive approximations starting from any initial vector in the invariant set (Luenberger, 1969, p. 272). Thus for each $s \leqslant s_1$, where s_1 is given by Eq. (11) in Theorem 2, the solution of Eq. (6) may be obtained by the following method of successive approximations. Starting from the initial vector $\underline{x}_0 = \underline{1}$, we define the sequence $\{\underline{x}_n\}_{n=1}^{\infty} \subset l_\infty(A)$ by $\underline{x}_n = T_s^*(\underline{x}_{n-1})$, where T_s^* was defined in the proof of Theorem 2. Then from the Contraction Mapping Theorem, we have $\| \underline{x}_n - \underline{\mu}_s \|_\infty \to 0$ as $n \to \infty$. A numerical solution of Eq. (6) can thus be obtained.

Theorem 3 can still be used to determine if $R(D_s) = R_L(D_s)$ even if the explicit solution of Eq. (6) is not available. Since by Theorem 2 we know that $1 - \alpha(s) \leqslant \mu_s(j) \leqslant 1$ for all j in A and each $s \leqslant s_1$, Eq. (13) in Theorem 3 is satisfied if we can show that

$$\sum_{k \in A} \exp(s\rho(j,k))\, p(k) \leqslant (1-\alpha(s))\, p(j) \qquad \forall j \in A, \tag{28}$$

where $\alpha(s)$ is given by Eq. (11). Thus if Eq. (28) is satisfied for a value of the parameter s, $R(D_s) = R_L(D_s)$ and so the numerical value of $R(D_s)$ can be obtained by the method of successive approximations described above.

6. Conclusions

In this paper we have considered the problem of explicit evaluation of rate-distortion functions $R(D)$ for countably infinite alphabet memoryless sources by the Shannon lower bound approach. While the explicit evaluation of $R(D)$ by methods described in this paper is possible only for column balanced distortion measures in general, we also indicated the applicability of our methods to certain distortion measures with no symmetry conditions. The condition of column balanced distortion measures was not critical for the validity of the proved existence theorems, but was imposed to obtain explicit evaluation of $R_L(D)$, and thus explicit evaluation of $R(D)$. We have also shown the usefulness of these theorems to the problem of numerical calculation of $R(D)$ for values of small distortion. Finally it is interesting to investigate whether the functional analysis approach taken in this paper can be applied to treat the Shannon lower bound approach for sources with continuous alphabets.

Appendix 1

Proof of Theorem 2. Let $l_\infty(A)$ denote the Banach space of bounded sequences of real numbers $\underset{\sim}{x} = \{x(j): j \in A\}$ with sup norm $\| \underset{\sim}{x} \|_\infty = \sup_{j \in A} | x(j)|$. For $\underset{\sim}{x} = \{x(j): j \in A\}, \underset{\sim}{y} = \{y(j): j \in A\}$ in $l_\infty(A)$, we say $\underset{\sim}{x} \leqslant \underset{\sim}{y}$ iff $x(j) \leqslant y(j)$ for all $j \in A$. We say $\underset{\sim}{x}$ is nonnegative if $\underset{\sim}{x} \geqslant \underset{\sim}{0}$, where $\underset{\sim}{0}$ is the sequence of zeroes. Let $l_\infty^+(A) = \{\underset{\sim}{x} \in l_\infty(A): \underset{\sim}{x} \geqslant \underset{\sim}{0}\}$. For $s \leqslant s^*$, let B_s^* denote the linear operator on $l_\infty(A)$ defined by $B_s^* \underset{\sim}{x} = \{\sum_{j \in A} \exp(s\rho(j, k))\, x(j): k \in A\}$ for all $\underset{\sim}{x} \in l_\infty(A)$. If $\underset{\sim}{\mu}_s = \{\mu_s(j): j \in A\}$ and $\underset{\sim}{1} = \{1: j \in A\}$, then Eq. (6) may be written as $\underset{\sim}{\mu}_s = (I - B_s^*)\, \underset{\sim}{\mu}_s + \underset{\sim}{1}$, where I is the identity operator on $l_\infty(A)$. Define the mapping $T_s^*: l_\infty(A) \to l_\infty(A)$ by $T_s^*(\underset{\sim}{x}) = (I - B_s^*)\, \underset{\sim}{x} + \underset{\sim}{1}$ for all $\underset{\sim}{x} \in l_\infty(A)$.

In order to show that there exists a $s_1 \in (-\infty, s^*]$ such that Eq. (6) has a unique nonnegative solution $\{\mu_s(j): j \in A\}$ for each $s \leqslant s_1$, we will show the existence of a $s_1 \in (-\infty, s^*]$ such that T_s^* has a unique nonnegative fixed point in $l_\infty(A)$; that is, there exists a unique $\underset{\sim}{x} \in l_\infty^+(A)$ such that $T_s^*(\underset{\sim}{x}) = \underset{\sim}{x}$ for each $s \leqslant s_1$. This will prove the existence of a unique $l_\infty^+(A)$ solution to Eq. (6) for each $s \leqslant s_1$. But since every nonnegative solution of Eq. (6) must be bounded, we can conclude that this proves the existence of a unique nonnegative solution to Eq. (6) for each $s \leqslant s_1$.

Now, since the distortion measure ρ satisfies Eq. (1) and by assumption, there exists a $s^* > -\infty$ such that $\sup_k \sum_{j \in A} \exp(s^*\rho(j, k)) < \infty$, we can find a $s_1 \in (-\infty, s^*]$ such that $\alpha(s_1) < 1$. For each $s \leqslant s_1$ define $\mathscr{I}_s = \{\underline{x} \in l_\infty(A): (1 - \alpha(s))\, \underline{1} \leqslant \underline{x} \leqslant \underline{1}\}$. Since $\alpha(s) \leqslant \alpha(s_1)$ for all $s \leqslant s_1$, it is clear that $\mathscr{I}_s \subset l_\infty^+(A)$ for each $s \leqslant s_1$. Moreover, for each $s \leqslant s_1$, $\mathscr{I}_s$ is a closed subset of $l_\infty(A)$ and is therefore a complete metric space. First we claim that all possible $l_\infty^+(A)$ solutions of Eq. (6) for each $s \leqslant s_1$ must lie in $\mathscr{I}_s$. Clearly from Eqs. (1) and (6), $\mu_s(k) \leqslant 1$ for all $k \in A$ which together with Eq. (11) implies that $\mu_s(k) + \alpha(s) \geqslant \mu_s(k) + \sum_{j \neq k} \mu_s(j) \exp(s\rho(j, k)) = 1$ for each $s \leqslant s_1$. Then $\mu_s(k) \geqslant 1 - \alpha(s)$, $\forall k \in A$, and $s \leqslant s_1$. Thus for each $s \leqslant s_1$, we can restrict our consideration to $\mathscr{I}_s$ in the search of a unique $l_\infty^+(A)$ solution of Eq. (6).

We claim that for each $s \leqslant s_1$, T_s^* is a contraction (p. 272 of Luenberger, 1969) on $\mathscr{I}_s$. If this is true, then by the Contraction Mapping Theorem (p. 272 of Luenberger, 1969), T_s^* has a unique fixed point in $\mathscr{I}_s$ for each $s \leqslant s_1$ and thus Eq. (6) has a unique nonnegative solution $\{\mu_s(j): j \in A\}$ such that $1 - \alpha(s) \leqslant \mu_s(j) \leqslant 1$ for all $j \in A$ and $s \leqslant s_1$. To prove our claim, we first show that for each $s \leqslant s_1$, $T_s^*(\mathscr{I}_s) \subset \mathscr{I}_s$. We note that since $\rho(j, j) = 0$ for all $j \in A$, $T_s^*(\underline{x}) \geqslant T_s^*(\underline{y})$ whenever $\underline{x} \leqslant \underline{y}$, $\underline{x}, \underline{y} \in l_\infty(A)$. Thus for $s \leqslant s_1$ and $\underline{x} \in \mathscr{I}_s$, we have $T_s^*(\underline{1}) \leqslant T_s^*(\underline{x}) \leqslant T_s^*(0) = \underline{1}$. So in order to show that $T_s^*(\underline{x}) \in \mathscr{I}_s$, it suffices to show that $T_s^*(\underline{1}) \geqslant (1 - \alpha(s))\, \underline{1}$ or equivalently $(B_s^* - I)\, \underline{1} \leqslant \alpha(s)\, \underline{1}$. But this follows directly from Eq. (11). Thus we have shown that $T_s^*(\mathscr{I}_s) \subset \mathscr{I}_s$. For $s \leqslant s_1$ and $\underline{x}, \underline{y} \in l_\infty(A)$, we also have $\| T_s^*(\underline{x}) - T_s^*(\underline{y})\|_\infty \leqslant \alpha(s) \| \underline{x} - \underline{y} \|_\infty \leqslant \alpha(s_1) \| \underline{x} - \underline{y} \|_\infty$, where $\alpha(s_1) < 1$ by assumption. This completes the proof of our claim. The proof of the theorem is complete since by previous remarks, the interval $(-\infty, s_1]$ of the parameter s corresponds to the range of distortion $(0, D_1]$ where $D_1 = D_{s_1} > 0$. Q.E.D.

Appendix 2

In order to prove Theorem 3, we need to use Schauder Fixed Point Theorem (p. 96 of Schwartz, 1969) which is stated as Lemma A for reference.

Lemma A (Schauder Fixed Point Theorem). *Let X be a Banach space, K a compact convex subset of X. Let T: $K \to K$ be a continuous mapping, then T has a fixed point $x \in K$, that is, $T(x) = x$.*

To apply Schauder's Theorem we require conditions which identify the compact subsets of Banach Spaces. The following result (p. 37 of Liusternik and Sobolev, 1961) stated as Lemma B is sufficient for our purposes.

Lemma B. *A closed subset M of a Banach space X is compact if and only* *if it is totally bounded, that is, for every $\epsilon > 0$ there exists a finite set of points in X, say $\{x_1, x_2, \ldots, x_n\}$ such that for each point $x \in M$, there exists a point x_i, $1 \leqslant i \leqslant n$, so that $\| x - x_i \| < \epsilon$.*

Proof of Theorem 3. Let $l_1(A)$ denote the Banach space of absolutely summable sequences of real numbers $\underset{\sim}{x} = \{x(j): j \in A\}$ with norm $\| \underset{\sim}{x} \|_1 = \sum_{j\in A} | x(j)|$. Similar to the proof of Theorem 2, let the partial ordering on $l_1(A)$ be defined by $\underset{\sim}{x} \leqslant \underset{\sim}{y}$ iff $x(j) \leqslant y(j)$ for all $j \in A$. Let $l_1^+(A) = \{\underset{\sim}{x} \in l_1(A): \underset{\sim}{x} \geqslant \underset{\sim}{0}\}$. For $s \leqslant s_1$, let B_s denote the linear operator on $l_1(A)$ defined by $B_s\underset{\sim}{x} = \{\sum_{k\in A} \exp(s\rho(j, k))\, x(k): j \in A\}$ for all $\underset{\sim}{x} \in l_1(A)$. Let $\underset{\sim}{Q}_s = \{Q_s(k): k \in A\}$ and $\underset{\sim}{f}_s = \{p(j)/\mu_s(j): j \in A\}$. $\underset{\sim}{Q}_s \in l_1(A)$ since it is required to be a probability distribution. $\underset{\sim}{f}_s$ is also in $l_1(A)$ since $\sum_{j\in A} | p(j)/\mu_s(j)| \leqslant (1 - \alpha(s))^{-1} \sum_{j\in A} p(j) < \infty$ by virtue of Theorem 2. We can rewrite Eq. (11) as $\underset{\sim}{Q}_s = (I - B_s)\underset{\sim}{Q}_s + \underset{\sim}{f}_s$. Define the mapping $T_s : l_1(A) \to l_1(A)$ by $T_s(\underset{\sim}{x}) = (I - B_s)\, \underset{\sim}{x} + \underset{\sim}{f}_s$ for all $\underset{\sim}{x} \in l_1(A)$. For each $s \leqslant s_1$, since by Eq. (11), B_s is a continuous operator on $l_1(A)$, T_s is a continuous mapping on $l_1(A)$. To prove the theorem, it suffices to show that for each s $\leqslant s_c$, T_s has a fixed point in $l_1^+(A)$. By applying the Schauder Fixed Point Theorem, it suffices to show that for each $s \leqslant s_c$, there exists a convex compact subset $K_s \subset l_1(A)$ such that $K_s \subset l_1^+(A)$ and $T_s(K_s) \subset K_s$. For each $s \leqslant s_c$, let $K_s = [\underset{\sim}{0}, \underset{\sim}{f}_s] = \{\underset{\sim}{x} \in l_1(A): \underset{\sim}{0} \leqslant \underset{\sim}{x} \leqslant \underset{\sim}{f}_s\}$. We claim that this choice of K_s meets the above requirements. We first show that for each $s \leqslant s_c$, $T_s([\underset{\sim}{0}, \underset{\sim}{f}_s]) \subset [\underset{\sim}{0}, \underset{\sim}{f}_s]$. Note that since $\rho(j, j) = 0$ for all $j \in A$, $T_s(\underset{\sim}{x}) \geqslant T_s(\underset{\sim}{y})$ whenever $\underset{\sim}{x} \leqslant \underset{\sim}{y}$, $\underset{\sim}{x}$, $\underset{\sim}{y} \in l_1(A)$. Thus for $s \leqslant s_c$ and $\underset{\sim}{x} \in [\underset{\sim}{0}, \underset{\sim}{f}_s]$, $T_s(\underset{\sim}{f}_s) \leqslant T_s(\underset{\sim}{x}) \leqslant T_s(\underset{\sim}{0}) = \underset{\sim}{f}_s$. Thus to show that $T_s(\underset{\sim}{x}) \in [\underset{\sim}{0}, \underset{\sim}{f}_s]$ it suffices to show that $T_s(\underset{\sim}{f}_s) \geqslant \underset{\sim}{0}$ or equivalently $(B_s - I)\underset{\sim}{f}_s \leqslant \underset{\sim}{f}_s$ which follows from Eq. (13). This proves that $T_s([\underset{\sim}{0}, \underset{\sim}{f}_s]) \subset [\underset{\sim}{0}, \underset{\sim}{f}_s]$ for each $s \leqslant s_c$.

We will next show that for each $s \leqslant s_c$, $[\underset{\sim}{0}, \underset{\sim}{f}_s]$ is a compact subset of $l_1(A)$. It can be easily shown that $[\underset{\sim}{0}, \underset{\sim}{f}_s]$ is closed in $l_1(A)$. To show that $[\underset{\sim}{0}, \underset{\sim}{f}_s]$ is compact in $l_1(A)$, by Lemma B it suffices to show that it is totally bounded. For $i \in A$, define $\underset{\sim}{l}_i = \{l_k^{(i)}: k \in A\}$, where $l_k^{(i)} = 0$ if $k \neq i$, 1 if $k = i$. Then every $\underset{\sim}{x} \in l_1(A)$ is uniquely representable in the form $\sum_{i\in A} x(i)\, \underset{\sim}{l}_i$, where $x(i)$ is a real number. For $\underset{\sim}{x} = \sum_{i\in A} x(i)\, \underset{\sim}{l}_i \in [\underset{\sim}{0}, \underset{\sim}{f}_s]$ and integer $n \geqslant 1$, we have

$$\Big\| \sum_{|i|\geqslant n+1} x(i)\underset{\sim}{l}_i \Big\|_1 \leqslant \sum_{|i|\geqslant n+1} x(i)\, \| \underset{\sim}{l}_i \|_1 \leqslant \sum_{|i|\geqslant n+1} [p(i)/\mu_s(i)]$$

$$\leqslant (1 - \alpha(s))^{-1} \sum_{|i|\geqslant n+1} p(i).$$

Fix an $\epsilon > 0$. Thus, since $\sum_{i\in A} p(i) = 1$, there exists an integer $N_0 \geqslant 1$ such that $\| \sum_{|i|\geqslant N_0+1} x(i)\, \underset{\sim}{l}_i \|_1 < \epsilon/2$ uniformly for every $\underset{\sim}{x} \in [\underset{\sim}{0}, \underset{\sim}{f}_s]$. Now let

$S = \{\underline{y} = \{y(j): j \in A\}: 0 \leqslant y(j) \leqslant p(j)/\mu_s(j)$ for $|j| \leqslant N_0$ and $y(j) = 0$ for $|j| \geqslant N_0 + 1\}$. Now S is a closed and bounded finite-dimensional subset of $l_1(A)$ and thus by the Heine–Borel Theorem (p. 35 of Rudin, 1964) is compact. By Lemma B, S is totally bounded, so there exist $\{\underline{V}_1, \underline{V}_2, ..., \underline{V}_m\} \subset l_1(A)$ such that for each $\underline{y} \in S$, there exists a $\underline{V}_i$, $1 \leqslant i \leqslant m$, so that $\|\underline{y} - \underline{V}_i\|_1 < \epsilon/2$. Now take a $\underline{x} \in [\underline{0}, \underline{f}_s]$. We can write

$$\underline{x} = \sum_{|i| \leqslant N_0} x(i)\underline{l}_i + \sum_{|i| \geqslant N_0+1} x(i)\underline{l}_i .$$

Since $\sum_{|i| \leqslant N_0} x(i)\, \underline{l}_i \in S$, there exists a $\underline{V}_j$, $1 \leqslant j \leqslant m$, such that

$$\left\| \sum_{|i| \leqslant N_0} x(i)\underline{l}_i - \underline{V}_j \right\|_1 < \epsilon/2.$$

Since $\| \sum_{|i| \geqslant N_0+1} x(i)\, \underline{l}_i \|_1 < \epsilon/2$, it follows that $\| \underline{x} - \underline{V}_j \|_1 < \epsilon$ which proves that $[\underline{0}, \underline{f}_s]$ is totally bounded and therefore compact. Finally, the convexity of $[\underline{0}, \underline{f}_s]$ is clear. This completes the proof of the theorem. Q.E.D.

RECEIVED: January 16, 1974; REVISED: July 24, 1974

REFERENCES

BERGER, T. (1970), Information rates of Wiener Processes, *IEEE Trans. Inform. Theory, PGIT* **16**, 134–139.

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

BLAHUT, R. (1972), Computation of Channel Capacity and Rate-Distortion Functions, *IEEE Trans. Inform. Theory, PGIT* **18**, 460–473.

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

GRAY, R. M. (1970), Information Rates of Autoregressive Processes, *IEEE Trans. Inform. Theory, PGIT* **16**, 412–421.

GRAY, R. M. (1971a), Rate Distortion Functions for Finite-State Finite-Alphabet Markov Sources, *IEEE Trans. Inform. Theory, PGIT.* **17**, 127–134.

GRAY, R. M. (1971b), Information Rates of Stationary-Ergodic Finite-Alphabet Sources, *IEEE Trans. Inform. Theory, PGIT* **17**, 516–523.

GRAY, R. M. (1973a), A New Class of Lower Bounds to Information Rates of Stationary Sources Via Conditional Rate-Distortion Functions, *IEEE Trans. Inform. Theory, PGIT* **19**, 480–489.

GRAY, R. M. (1973b), Correction to "Information Rates of Stationary-Ergodic Finite-Alphabet Sources, *IEEE Trans. Inform. Theory, PGIT* **19**, 573.

HARDY, G. AND M. RIESZ (1915), "The General Theory of Dirichlet's Series," Cambridge Press, London.

JELINEK, F. (1967), Evaluation of Rate Distortion Functions for Low Distortions, *IEEE Proc.* **55**, 2067–2068.

LIUSTERNIK, L. A. AND V. J. SOBOLEV (1961), "Elements of Functional Analysis," Ungar, New York.

LUENBERGER, D. G. (1969), "Optimization by Vector Space Methods," Wiley, New York.

PINKSTON, J. T. (1969), An Application of Rate Distortion Theory to a Converse to the Coding Theorem, *IEEE Trans. Inform. Theory*, *PGIT* **15**, 66–71.

RUBIN, I. (1973), Information Rates for Poisson Sequences, *IEEE Trans. Inform. Theory*, *PGIT* **19**, 283–294.

RUDIN, W. (1964), "Principles of Mathematical Analysis," 2nd Ed., McGraw-Hill, New York.

SCHWARTZ, J. T. (1969), "Nonlinear Functional Analysis," Gordan and Breach, New York.

WYNER, A. D. AND ZIV, J. (1971), Bounds on the Rate-Distortion Function for Stationary Sources with Memory, *IEEE Trans. Inform. Theory*, *PGIT* **17**, 508–513.

Printed in Belgium