On Calculating Sakrison's Rate Distortion Function for Classes of Parameterized Sources

Laurence B. Wolfe

Abstract—Sakrison extended Shannon's notion of the rate distortion function to parameterized classes of sources by taking a minimax approach and defining a measure of the minimum rate required for information reconstruction subject to a prescribed fidelity level D . Unfortunately, calculation of Sakrison's rate distortion function may be very difficult because analytic solutions do not generally exist and there has been a lack of a constructive method for finding the rate. However, an approach presented in this correspondence may be used to calculate an approximation to Sakrison's rate distortion function for classes of sources with a finite, discrete input space and a continuous parameter space. The approach gives rise to an algorithm which is shown to be convergent and numerical examples are studied.

Index Terms—Rate distortion function, source matching, relative entropy.

I. INTRODUCTION

Sakrison in [6] extended the notion of a rate distortion function to classes of sources. Shannon [7] had previously defined the rate distortion function $R^\theta(D)$ for an individual source θ which measures the minimum amount of information that must be preserved by any code to allow reproduction of the compressed data with average distortion less than or equal to a given D . According to Sakrison [6], the rate distortion function $R^\Lambda(D)$ for a class of sources with compact parameter space Λ may be defined as the supremum over all rate distortion functions $R^\theta(D)$ in the class. However, calculation of $R^\Lambda(D)$ for a continuous Λ is not generally tractable because any solution would involve an uncountable number of integrals.

Apparently, Sakrison's $R^\Lambda(D)$ can be approximated for continuous Λ by performing an exhaustive search after selecting for some M , a finite set of parameters $\{\theta_j\}_{j=1}^M$ to represent the class of sources with finite, discrete input spaces, where $\theta_j \in \Lambda$. In fact, it will be shown later in this correspondence that under some general conditions and a specific value of M , there exists an optimal set $\{\theta_j^*\}_{j=1}^M$ yielding the most accurate approximation to Sakrison's $R^\Lambda(D)$. However, finding the $\{\theta_j^*\}_{j=1}^M$ is not trivial because it would involve computation of an infinite number of integrals and the lack of a general computational method. Therefore, we seek a procedure for selecting a set $\{\theta_j\}_{j=1}^N$ which contains a subset that accurately approximates $\{\theta_j^*\}_{j=1}^M$ for some $M \leq N$. We present such a procedure in this correspondence and employ an approach which selects $\{\theta_j\}_{j=1}^N$ from Λ by using relative entropy to group within N subclasses, all sources in Λ whose entropies are within a previously assigned level of similarity. In this correspondence, preliminary information is reviewed in Section II and an approach for calculating Sakrison's [6] rate distortion function is presented in Section III. An example is studied in Section IV, and conclusions are in Section V.

II. PRELIMINARIES

Sakrison [6] extended the notion of the rate distortion function for individual sources to parameterized classes of sources by taking

Manuscript received July 22, 1993; revised November 14, 1994.

The author is a Computer Systems Manager with the U.S. Government (GSA), Acquisition Reviews Branch, Washington DC 202405 USA.

IEEE Log Number 9411672.

a minimax approach and defining a measure of the minimum rate required for information reconstruction subject to a prescribed fidelity level D . Shannon [7] had previously defined the rate distortion function $R^\theta(D)$ for an individual, discrete ergodic memoryless source with probability measure p^θ with parameter θ and respective source and reproduction spaces $A = \{0, \dots, J-1\}$ and $B = \{0, \dots, K-1\}$ as

$$R^\theta(D) = \min_{Q \in Q(D)} I(p^\theta; Q) \quad (1)$$

where

$$I(p^\theta; Q) = \sum_{j \in A} \sum_{k \in B} p_j^\theta \cdot Q_{k|j} \cdot \log \left(\frac{Q_{k|j}}{\sum_{i \in A} p_i^\theta \cdot Q_{k|i}} \right) \quad (2)$$

and

$$Q(D) = \left\{ Q_{k|j} \left| \sum_{j \in A} \sum_{k \in B} p_j^\theta \cdot Q_{k|j} \cdot d_{jk} \leq D \right. \right\}. \quad (3)$$

A distortion matrix $[d_{jk}]$ assigns a nonnegative penalty value for the reproduction of input j by output k and $Q(D)$ is the set of all such matrices. Observe that (2) can also be written as

$$\begin{aligned} I(p^\theta; Q) &= \sum_{j \in A} p_j^\theta \cdot \sum_{k \in B} Q_{k|j} \cdot \log \frac{Q_{k|j}}{q_k}, \\ &= \sum_{j \in A} p_j^\theta \cdot H(Q; q) \end{aligned} \quad (4)$$

where

$$q_k \equiv \sum_{i \in A} p_i^\theta \cdot Q_{k|i}, \quad (5)$$

and $H(Q; q)$ is the relative entropy which measures the similarity (or difference) between probability measures Q and q over the reproduction space.

Apparently, analytic solutions to (1) do not exist in general, and a minimization technique must be used to find a solution. One such technique was given by Blahut [1] and is briefly reviewed for reference because it is used in Sections III and IV. Blahut's algorithm formulates an associated minimization problem in terms of a LaGrange multiplier s

$$\begin{aligned} R^\theta(D) &= \min_{Q \in Q(D)} \left[\sum_{j \in A} \sum_{k \in B} p_j^\theta \cdot Q_{k|j} \cdot \log \frac{Q_{k|j}}{\sum_{i \in A} p_i^\theta \cdot Q_{k|i}} \right. \\ &\quad \left. - s \left(\sum_{j \in A} \sum_{k \in B} p_j^\theta \cdot Q_{k|j} \cdot d_{jk} - D \right) \right] \end{aligned} \quad (6)$$

where

$$D = \sum_{j \in A} \sum_{k \in B} p_j^\theta \cdot Q_{k|j}^* \cdot d_{jk} \quad (7)$$

and Q^* is the channel transition matrix in a compact space $Q(D)$ which achieves the minimum. Using this parametric form, Blahut

formulated a double minimum problem by defining

$$F(p^\theta, Q, q) = \sum_{j \in A} \sum_{k \in B} p_j^\theta \cdot Q_{k|j} \cdot \log \frac{Q_{k|j}}{q_k} - s \sum_{j \in A} \sum_{k \in B} p_j^\theta \cdot Q_{k|j} \cdot d_{jk} \quad (8)$$

with the result that

1)

$$R^\theta(D) = s \cdot D + \min_{Q, q} F(p_j^\theta, Q_{k|j}, q).$$

2) For a fixed $Q_{k|j}$ in $Q(D)$, $F(p_j^\theta, Q_{k|j}, q)$ is minimized by

$$q_k = \sum_{j \in A} p_j^\theta \cdot Q_{k|j}. \quad (9)$$

3) For a fixed q , $F(p_j^\theta, Q_{k|j}, q)$ is minimized by

$$Q_{k|j} = \frac{q_k \cdot e^{s \cdot d_{jk}}}{\sum_{i \in A} q_i \cdot e^{s \cdot d_{ji}}}. \quad (10)$$

The results show that Blahut's calculation of Shannon's $R^\theta(D)$ must depend on a parameter s .

However, in order to extend Shannon's rate distortion function to a class of sources Λ , Sakrison [6] considered the mutual information $I(p^\theta, Q)$ for each $\theta \in \Lambda$ and defined the extended rate distortion function for a class of sources Λ as

$$R^\Lambda(D) = \inf_{Q \in Q(D)} \sup_{\theta \in \Lambda} I(p^\theta; Q). \quad (11)$$

Sakrison [6] also proved that if the parameter space Λ is compact, then

$$\begin{aligned} R^\Lambda(D) &\equiv \inf_{Q \in Q(D)} \sup_{\theta \in \Lambda} I(p^\theta; Q) \\ &= \sup_{\theta \in \Lambda} \inf_{Q \in Q(D)} I(p^\theta; Q) \\ &= \sup_{\theta \in \Lambda} R^\theta(D). \end{aligned} \quad (12)$$

Apparently, (11) and (12) show that calculating $R^\Lambda(D)$ is equivalent to solving a minimax problem. Although such problems are generally difficult to solve, the well-known minimax theorem as given in [4] shows that a solution may be obtained by finding a least favorable distribution and the corresponding Bayes risk. Using (12), consider a Bayes problem with a given prior distribution $\tau(\theta)$ over Λ and a weighted mutual information function $r(\tau, Q)$

$$r(\tau, Q) \equiv \int_{\theta \in \Lambda} I(p^\theta, Q) \cdot \tau(\theta) d\theta. \quad (13)$$

Following a source-matching approach presented by Davisson *et al.* in [3] and combining (12) and (13) it can be shown that a solution is given by

$$\begin{aligned} R^\Lambda(D) &= \inf_{Q \in Q(D)} \sup_{\theta \in \Lambda} I(p^\theta, Q) \\ &= \sup_{\tau \in \Xi} \inf_{Q \in Q(D)} r(\tau, Q) \end{aligned} \quad (14)$$

where Ξ is the class of all prior distributions defined on Λ .

When Λ and $Q(D)$ are compact and both the input and output spaces are discrete, the infimum and supremum in (14) can actually be replaced by the minimum and maximum. Furthermore, if the class of sources Λ is finite the minimax problem may be solved by calculating $R^\theta(D)$ for each $\theta \in \Lambda$. $R^\Lambda(D)$ may then be determined by an exhaustive search over all $\theta \in \Lambda$.

Unfortunately, this approach cannot be directly applied to finding Sakrison's $R^\Lambda(D)$ when the class of sources Λ is continuous since

an uncountable number of solutions of (11), (12), and (14) would be required; one for each θ . Furthermore, analytic solutions for $R^\Lambda(D)$ are not generally available and calculating $R^\Lambda(D)$ may be very difficult because of the lack of a general computational approach. However, an approach presented in the next section may be used to calculate an approximation to $R^\Lambda(D)$.

III. AN APPROACH FOR CALCULATING $R^\Lambda(D)$

The approach comes from the recognition that Sakrison's rate distortion function $R^\Lambda(D)$ for a continuous class of sources Λ can be approximated by performing an exhaustive search after selecting for some M , a finite set of parameters $\{\theta_j\}_{j=1}^M$ to represent the class, where $\theta_j \in \Lambda$. We consider only classes with a finite, discrete source space A of size J . Under certain conditions, specified later, it is shown this section that for a specific value of M , there exists a set $\{\theta_j^*\}_{j=1}^M$ which yields the most accurate approximation to Sakrison's $R^\Lambda(D)$. However, finding the $\{\theta_j^*\}_{j=1}^M$ is not trivial because of the lack of a general computational method. Therefore, we seek a procedure for selecting a set $\{\theta_j\}_{j=1}^N$ which contains a subset that accurately approximates $\{\theta_j^*\}_{j=1}^M$ for some $M < N$. We present such a procedure in this section. The procedure uses relative entropy (i.e., Kullback distance, cross entropy, discrimination information) as in [8] to group, within subclasses, all sources in a continuous parameter class whose entropies are within a previously assigned level of similarity. A finite set of N representative sources is then selected from the N subclasses, namely $\{\theta_j\}_{j=1}^N$, which is subsequently used to derive an approximation to Sakrison's rate distortion function.

We begin by describing an algorithm adapted from [8] to partition a continuous class of sources Λ into N subclasses and select a set of N representatives; one from each subclass. An error threshold ϵ_{SP} can be adjusted in the algorithm such that N can assume any positive integer value. For simplicity of discussion, the algorithm will be given assuming that each parameter $\theta \in \Lambda$ lies on the real line in the closed interval $[a, b]$. Furthermore, we assume that $H(\theta; \theta')$ varies smoothly with θ , for all $\theta \in \Lambda$. We also assume that Λ is compact; and if it is not, add the hull to the space for purposes of computation and subsequently restrict our choice to only those sources which lie in Λ for inclusion in the finite set of subclasses. Also, the algorithm can readily be adapted to higher dimensions under similar, general conditions.

The Source Partition algorithm given below groups all sources into N subclasses using relative entropy to measure the similarity between any two sources. Thus the algorithm produces a finite set of N parameters $\{\theta_\gamma\}_{\gamma=1}^N$, which partitions $[a, b]$ into N subclasses $\{S_\gamma\}_{\gamma=1}^N$ where $S_\gamma = [\theta_{\gamma-1}, \theta_\gamma]$. The $\{S_\gamma\}_{\gamma=1}^N$ are chosen such that the relative entropy between any two sources in a subclass is less than a given tolerance ϵ_{SP} . Clearly, the number of subclasses produced, N , can assume any positive integer value by appropriately varying ϵ_{SP} .

A. Algorithm SP (Source Partition)

- 1) Initialization: Set ϵ_{SP} = an assigned error tolerance; $\theta_1 = a$ and $N = 1$.
- 2) Set $N = N + 1$. Find $\theta_N > \theta_{N-1}$ such that $H(\theta_N; \theta_{N-1}) = \epsilon_{SP}$.
- 3) If $\theta_N > b$ output the set $b \cup \{\theta_\gamma\}$; $\gamma = 1, \dots, N - 1$.
- 4) If $\theta_N < b$, go to step 2.

An approximation to Sakrison's rate distortion function $R^\Lambda(D)$ could in concept be calculated with any finite set of representative sources $\{\theta_\gamma\}_{\gamma=1}^N = 1$, produced by Algorithm SP. In fact, $R_N^\Lambda(D)$, the rate distortion function calculated using these N representative sources, must converge to Sakrison's rate distortion function $R^\Lambda(D)$

as $N \rightarrow \infty$, as will be shown later in this section. However, the validity of approximating Sakrison's $R^\Lambda(D)$ in this way depends on the proper choice of N , as shown in the following theorem.

Theorem 1: Let Λ be a continuous, compact parameter space for a class of sources with input space A of size J . Further, assume that each parameter $\theta \in \Lambda$ lies on the real line in the closed interval $[a, b]$ and that $H(\theta, \theta')$ varies smoothly with θ , for all $\theta \in \Lambda$. For a given distortion measure d , there exists a finite set of sources $\{\theta_i^* \in \Lambda\}_{i=1}^M$ where $M \leq J$, which gives rise to Sakrison's rate distortion function $R^\Lambda(D)$.

Proof: Let Q^* be the channel given by solution of (11) thereby generating Sakrison's rate distortion function $R^\Lambda(D)$. However, Q^* along with all $\theta \in \Lambda$ gives rise to a new class of sources W with probability distributions given by

$$W_k^\theta = \sum_{j \in A} Q_{k|j}^* \cdot p_j^\theta. \quad (15)$$

Therefore, W is a class of discrete sources with parameter given by $\theta \in \Lambda$. Let $H(W^\theta)$ denote the average length of the best code for W^θ and define the redundancy between any two codes as

$$r(W^{\theta_i}, W^{\theta_j}) = |H(W^{\theta_i}) - H(W^{\theta_j})|. \quad (16)$$

The problem becomes one of minimizing the maximum redundancy over W by finding a source best matched to the class of sources. However, this is a source-matching problem which, as shown in [3], is solved by finding the channel capacity between the parameter space Λ and the output space A . The solution of this source-matching problem also produces an associated least favorable distribution τ^* over Λ . The validity of utilizing a finite set $\{\theta_i \in \Lambda\}_{i=1}^M$ chosen from an uncountable number of sources can now be seen for the source-matching problem by applying [5, Corollary 3, p. 96] which states that for a finite output space A there is a distribution τ^* over Λ that assigns a nonzero probability to only a minimal number of sources $\{\theta_i \in \Lambda\}_{i=1}^M$ and τ^* gives rise to the channel capacity between Λ and A . Furthermore, M can be no larger than the size of the output set, i.e., $M \leq J$.

Observe that there exists a one-to-one correspondence between each W^θ and some p^θ and that W maps onto the class of all $\{p^\theta\}$. By construction, the quantization and source-matching problems correspond directly and solution of one implies solution of the other.

QED

The suggested approach may now be summarized as follows:

Algorithm 1:

- 1) Apply *Algorithm SP* to produce N representative sources $\{\theta_j\}_{j=1}^N$, which contains a subset that approximates $\{\theta_i^*\}_{i=1}^M$, the optimal set of M source representatives, where $M \leq J \leq N$.
- 2) Calculate Shannon's rate distortion function for each $\theta_j \in \{\theta_j\}_{j=1}^N$ and find the maximum by exhaustive search, which is the approximation to Sakrison's rate distortion function $R^\Lambda(D)$.

One of the most well-known ways of calculating Shannon's rate distortion function as in step 2 is by using Blahut's algorithm [1]. Before showing convergence of the approach, we observe that there may be a difficulty in directly applying Blahut's rate distortion algorithm [1] to step 2. Blahut's algorithm uses a LaGrange multiplier s which is the slope of the rate distortion function, to determine the information rate required at distortion level D for a channel transition matrix Q . Therefore, a direct application of Blahut's algorithm to a set of representative sources will produce a comparison based upon slope s . Unfortunately, the respective rate distortion functions at slope s may actually have dissimilar distortions D ; thereby rendering the comparison invalid. That is, two different sources may require

two different slopes s in order to produce the same distortion D . Therefore, application of Blahut's algorithm to this problem must be modified to ensure that the rate distortion functions of any two sources are compared only at the same level D .

A modification in using Blahut's algorithm can apparently be made by selecting slope s_j as a function of D , for each source θ_j in the set of N representative sources $\{\theta_j\}_{j=1}^N$. There are simple techniques available to approximate s for a particular value of D . An example of one such method is obtainable from [10, p. 95] in the case of the difference distortion measure $[d_{jk}] \equiv \rho(k-j)$, for some measure ρ . Employing this or a similar method applicable to the given distortion measure d , enables calculation of $R^{\theta_j}(D)$ for each $j = 1, \dots, N$ which gives rise to the set $\{R^{\theta_j}(D)\}_{j=1}^N$. Taking the maximum over this set determines the sought after approximation to Sakrison's rate distortion function $R^\Lambda(D)$.

It is clear that any set of N representative sources $\{\theta_j\}_{j=1}^N$, $M \leq J \leq N$ produced by *Algorithm SP* does not necessarily contain the optimal M sources. However, since relative entropy is used as a measure of similarity to construct the subclasses, the j th optimal source must be in some subclass represented by θ_k , for some k . Therefore, the set $\{\theta_j\}_{j=1}^N$ contains a subset which approximates the optimal $\{\theta_i^*\}_{i=1}^M$ that determines Sakrison's rate distortion function $R^\Lambda(D)$. Clearly, the error depends upon the relative entropy threshold ϵ_{SP} used to determine the subclasses. Thus for each increasing value of N , $\{\theta_j\}_{j=1}^N$ must contain a subset which more closely approximates the optimal $\{\theta_i^*\}_{i=1}^M$ because of our assumption that $H(\theta; \theta')$ varies smoothly with θ , for all $\theta \in \Lambda$. Therefore, even though only M sources are required to represent the class of sources Λ , it is apparent that the accuracy of the approximation increases with smaller ϵ_{SP} and corresponding larger N , as shown in the following theorem.

Theorem 2: Let $R_N^\Lambda(D)$ be the rate distortion function for a compact class Λ , that is determined by exhaustive search over the set $\{R^{\theta_j}(D)\}_{j=1}^N$ produced when $N \geq M$ representative sources are selected by *Algorithm SP*. Further, assume that each parameter $\theta \in \Lambda$ lies on the real line in the closed interval $[a, b]$ and that $H(\theta; \theta')$ varies smoothly with θ , for all $\theta \in \Lambda$. The sequence $\{R_N^\Lambda(D)\}$ converges in the limit to $R^\Lambda(D)$ as $N \rightarrow \infty$.

Proof: It suffices to show that $R_{N+1}^\Lambda(D) \geq R_N^\Lambda(D)$ for all N because $R^\Lambda(D)$ achieves the supremum over all $R^\theta(D)$ in the compact space Λ . Consider the approximation produced by N representative sources $\{\theta_\gamma\}_{\gamma=1}^N$. Let $\theta_s \in \{\theta_\gamma\}_{\gamma=1}^N$ produce the smallest mutual information. Case 1: find $\theta_g \in \{\theta_\gamma\}_{\gamma=1}^{N+1}$ that produces the greatest mutual information, such that $\theta_g \notin \{\theta_\gamma\}_{\gamma=1}^N$ and where the mutual information produced by θ_g is greater than or equal to that produced by θ_s . Thus the following equation is true:

$$\begin{aligned} R_N^\Lambda(D) &= \sup_{\tau^N \in \Xi^N} \inf_{Q \in Q(D)} r(\tau^N, Q) \\ &= \sup_{\tau^N \in \Xi^N} \inf_{Q \in Q(D)} \int_{\theta \in \Lambda} I(p^\theta; Q) \cdot \tau^N(\theta) d\theta \\ &= \max_{\tau^N \in \Xi^N} \min_{Q \in Q(D)} \sum_{\theta \in \Lambda} i(p^\theta; Q) \cdot \tau^N(\theta) \\ &\leq \max_{\tau^N \in \Xi^N} \min_{Q \in Q(D)} \sum_{\theta \in \{\Lambda - \theta_s\}} \tau^N(\theta) \cdot I(p^\theta; Q) \\ &\quad + \tau^N(\theta_s) \cdot I(p^{\theta_g}; Q) \\ &\leq \max_{\tau^{N+1} \in \Xi^{N+1}} \min_{Q \in Q(D)} \sum_{\theta \in \Lambda} \tau^{N+1}(\theta) \cdot I(p^\theta; Q) \\ &= R_{N+1}^\Lambda(D). \end{aligned} \quad (17)$$

The third equality occurs because the N representative sources form a finite class. The first inequality is true because the mutual

Fig. 1. $R^\Lambda(D)$ for a class of memoryless sources.

information produced by θ_g is greater than that of θ_s . The second inequality is a result of (14). Case 2: There does not exist a $\theta_g \in \{\theta_\gamma\}_{\gamma=1}^{N+1}$ that produces the greatest mutual information, such that $\theta_g \notin \{\theta_\gamma\}_{\gamma=1}^N$ and where the mutual information produced by θ_g is greater than or equal to that produced by θ_s . Clearly in this case, $R_N^\Lambda(D) = R_{N+1}^\Lambda(D)$. In both cases the sequence $\{R_N^\Lambda(D)\}_{N=1}^\infty$ is nondecreasing; in fact $\{R_N^\Lambda(D)\}_{N=1}^\infty$ is monotonically increasing because of our assumption that $H(\theta; \theta')$ varies smoothly with θ , for all $\theta \in \Lambda$. $\{R_N^\Lambda(D)\}_{N=1}^\infty$ is also bounded since $R^\Lambda(D)$ achieves the supremum over all $R^\theta(D)$ in the compact space Λ . Therefore, $\{R_N^\Lambda(D)\}_{N=1}^\infty$ converges to $R^\Lambda(D)$ and the theorem follows. QED

IV. NUMERICAL EXAMPLES

Example 1: Algorithm 1 was used to calculate Sakrison's rate distortion function for the class of stationary sources with identical source and output spaces $A = B = \{1, \dots, N\}$, where $j \in A$ and $\Lambda = [0, 1]$. Let the distortion measure $[d_{jk}]$ be the probability of error matrix $[1 - \delta_{jk}]$ where $\delta_{jk} = 1$ if $j = k$ and equals 0, otherwise. The class of sources is given by

$$p_j^\theta = \binom{N}{j} \cdot \theta^j \cdot (1 - \theta)^{N-j}. \quad (18)$$

For this example, we let $N = 2$. Fig. 1 shows the results produced by the algorithm which are similar to an analytic result in [9].

Example 2: Algorithm 1 was used for a class of binary symmetric sources with memory where the parameter $\underline{\theta} = [\theta_0, \theta_1]$ is a two-dimensional variable in the unit square $[0, 1] \times [0, 1]$. The source alphabet is $A = \{0, 1\}$ and the stochastic transition matrix is given by

$$\begin{bmatrix} \theta_{00} & \theta_{01} \\ \theta_{10} & \theta_{11} \end{bmatrix} = \begin{bmatrix} 1 - \theta_0 & \theta_0 \\ \theta_1 & 1 - \theta_1 \end{bmatrix} \quad (19)$$

where θ_{ij} represents the transition probability from the previous state i to the present state j . For simplicity of notation, θ_{01} is denoted by θ_0 and θ_{10} is denoted by θ_1 . The stationary pmf is given by the following equation:

$$\begin{aligned} \pi &= \left(\frac{\theta_1}{\theta_0 + \theta_1}, \frac{\theta_0}{\theta_0 + \theta_1} \right) \\ &= (\pi_0, \pi_1). \end{aligned} \quad (20)$$

Let $\underline{x}^N = [x_1, \dots, x_N]$ be a message block of length N generated by the Markov source and

$$p^\theta(\underline{x}^N) = \pi_h \cdot (1 - \theta_0)^i \cdot \theta_0^j \cdot \theta_1^k \cdot (1 - \theta_1)^{N-i-j-k-1} \quad (21)$$

where i, j, k represent the number of state transitions as defined by the transition matrix and π_h is the steady-state probability of x_1 in

Fig. 2. $R^\Lambda(D)$ for a class of sources with memory.

state h . The probability of error distortion matrix $[d_{ij}]$ was also used in this example where $\Lambda = [0.3, 0.45]$ and $N = 3$. Fig. 2 shows the results.

V. CONCLUSIONS

An approach was presented in this correspondence for calculation of Sakrison's rate distortion function for classes of sources with a continuous parameter space. The approach gives rise to a computational algorithm which was shown to be convergent and numerical examples were studied to demonstrate its application.

REFERENCES

- [1] R. E. Blahut, "Computation of channel capacity and rate distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460-473, July 1972.
- [2] C.-I. Chang, S. C. Fan, and L. D. Davisson, "On numerical methods of calculating the capacity of a continuous-input discrete-output memoryless channel," *Inform. and Comput.*, vol. 86, no. 1, pp. 1-13, 1990.
- [3] L. D. Davisson and A. Leon-Garcia, "A source matching approach to finding minimax codes," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 166-174, Mar. 1980.
- [4] T. S. Ferguson, *Mathematical Statistics*. New York: Academic Press, 1967.
- [5] R. G. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.
- [6] D. J. Sakrison, "The rate distortion function for a class of sources," *Inform. and Contr.*, vol. 15, pp. 165-195, 1969.
- [7] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, July 1948.
- [8] C.-I. Chang, "Source matching problems revisited," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1391-1395, July 1992.
- [9] A. J. Viterbi and J. K. Omura, *Principles of Digital Communications and Coding*. New York: McGraw-Hill, 1979, p. 398.
- [10] R. M. Gray, *Source Coding Theory*. Norwell, MA: Kluwer, 1990.