

New Bounds on the Rate-Distortion Function of a Binary Markov Source

Shirin Jalali

Department of Electrical Engineering
Stanford University
Stanford, CA, 94305, USA
shjalali@stanford.edu;

Tsachy Weissman

Department of Electrical Engineering
Stanford University
Stanford, CA, 94305, USA
tsachy@stanford.edu;

Abstract—This paper addresses the problem of bounding the rate-distortion function of a binary symmetric Markov source. We derive a sequence of upper and lower bounds on the rate-distortion function of such sources. The bounds are indexed by k , which corresponds to the dimension of the optimization problem involved. We obtain an explicit bound on the difference between the derived upper and lower bounds as a function of k . This allows to identify the value of k that suffices to compute the rate distortion function to a given desired accuracy. In addition to these bounds, a tighter lower bound which is also a function of k is derived. Our numerical results show that the new bounds improve on the Berger's upper and lower bounds even with small values of k .

I. INTRODUCTION

Shannon's well-known rate-distortion theory for i.i.d. sources answers the following question [1]: what is the minimum number of bits required per source symbol for describing the source within average distortion D to the decoder. The answer can be numerically found within desired precision by solving an optimization problem [2]. A natural extension of this theorem would be to the case where the source is no longer memoryless, e.g. a general, not necessarily i.i.d., stationary ergodic source. Although it may seem surprising, this problem has not been solved yet. Even for the simple case when the source is binary symmetric Markov with state transition probability of q , BSMS(q), the rate-distortion function can only be explicitly computed on a small-distortion region [4]. Beyond this region, even for this simple case, currently there only exist lower and upper bounds on $R(D)$.

In this paper, we first review the known results on the computation of rate-distortion function of sources with memory, and then present the derivation of the new upper and lower bounds on $R(D)$ of a BSMS(q) which are found applying the methods used in [7]. Finally, we consider some examples of the computation of the new bounds.

II. THE RATE-DISTORTION PROBLEM

Consider the well-known rate-distortion problem where one is interested in finding the minimum number of bits per source symbol required for describing the source to the decoder within a given fidelity constraint. Assume that given is a discrete memoryless source whose output, $\{X_n\}_{n=1}^{\infty}$, is an

i.i.d. process. X_n 's take values in a finite set \mathcal{X} , with $|\mathcal{X}| = N$ and for every n , $X_n \sim \mathbf{p}$, where $\mathbf{p} \in \mathbb{R}^N$, $p_i \geq 0$ for $1 \leq i \leq N$, and $\langle \mathbf{1}, \mathbf{p} \rangle = 1$. Furthermore, assume that the symbols of the decoded sequence $\{\hat{X}_n\}_{n=1}^{\infty}$, take values in $\hat{\mathcal{X}}$, with $|\hat{\mathcal{X}}| = M$. The distortion between an n -block source and reconstructed sequence is defined to be

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{x}_i),$$

where $d(x, \hat{x})$ is a bounded function that measures the distortion between the symbols of the source and reconstruction alphabets. Shannon's rate distortion theory states that the minimum rate R such that (R, D) is achievable is given by,

$$R(D) = \min_{\substack{p(\hat{x}|x): \sum_{(x, \hat{x})} p(x)p(\hat{x}|x)d(x, \hat{x}) \leq D}} I(X; \hat{X}). \quad (1)$$

Consequently, computing the rate-distortion function of a given discrete memoryless i.i.d. source at any point D is equivalent to finding the optimal test channel that optimizes the convex optimization problem described in (1). In [3], by forming the Lagrange dual of (1), and using strong duality theorem, it is shown that for finding $R(D)$, instead of (1), one can solve the following geometric program (in convex form):

$$\begin{aligned} & \text{maximize} && p\alpha - \gamma D \\ & \text{subject to} && \log \sum_{j=1}^N \exp(\log p_i + \alpha_i - \gamma d_{ij}) \leq 0, \\ & && j = 1, 2, \dots, M, \\ & && \gamma \geq 0, \end{aligned} \quad (2)$$

where the optimization variables are $\alpha \in \mathbb{R}^N$ and γ . In (2), d_{ij} is defined to represent the distortion between the i -th symbol in \mathcal{X} and the j -th symbol in $\hat{\mathcal{X}}$. The geometric program in

(2), can be reformulated in standard form as follows

$$\begin{aligned} & \text{maximize} && w^{-D} \prod_{i=1}^N z_i^{p_i} \\ & \text{subject to} && \sum_{i=1}^N p_i z_i w^{-d_{ij}} \leq 1, j = 1, 2, \dots, M, \\ & && w \geq 1, z_i \geq 0, i = 1, 2, \dots, N. \end{aligned} \quad (3)$$

For going from (2) to (3), one needs to do the following change of variables: $z_i = e^{\alpha_i}$, and $w = e^\gamma$. These characterizations would be helpful in computing the k -th order rate-distortion functions which will be defined in the next section, and are required for finding the lower and upper bounds on $R(D)$ of BSMS(q).

III. RATE-DISTORTION FUNCTION OF BSMS(q): KNOWN RESULTS

Consider a discrete stationary ergodic source with source alphabet \mathcal{X} and reconstruction alphabet $\hat{\mathcal{X}}$. The rate-distortion function of such source is given by [9]

$$R(D) = \lim_{n \rightarrow \infty} R_n(D), \quad (5)$$

where $R_n(D)$ is the n -th order rate-distortion function of the source defined as

$$R_n(D) = \min_{p(\hat{x}^n | x^n): E[d(x^n, \hat{x}^n)] \leq D} \frac{1}{n} I(x^n; \hat{x}^n). \quad (6)$$

Although it might not be clear at first glance, the above equation does not yield the rate-distortion function of a given stationary ergodic source explicitly. The problem is that the computational complexity required for solving the convex optimization problem given in (6) grows exponentially with n . Moreover, the convergence rate of $R_n(D)$ to $R(D)$ is typically slow, and at each step it is not clear how far we are from the optimal solution.

One of the simplest models for a source with memory is the pre-described BSMS(q). Computation of the rate-distortion function of this source has been investigated by Gray in [4], where it was shown that for $0 \leq D \leq D_c$,

$$R(D) = H_b(q) - H_b(D), \quad (7)$$

where $p = 1 - q$, and for $q \leq 1/2$,

$$D_c = \frac{1}{2} \left(1 - \sqrt{1 - (q/p)^2} \right),$$

and $H_b(x) = -x \log 2 - (1 - x) \log(1 - x)$. Beyond D_c , even for this simple case, currently only lower and upper bounds on $R(D)$ are known.

In 1977, Berger found *explicit* lower and upper bounds on $R(D)$, $R_\ell(D)$ and $R_u(D)$ respectively, which do not depend on n [6]. Eq. (4) gives the lower bound, where,

$$\begin{aligned} D_2 &= 0.5(1 - \sqrt{1 - 2q}), \\ r &= q/p, \\ p_\theta &= 1 - q_\theta = (1 + r^\theta)^{-1}, \end{aligned}$$

and θ and α are related as follows

$$\frac{r^\theta}{p(1 + r^\theta)^2(1 + r^\theta)} = \frac{\alpha}{(1 + \alpha)^2}.$$

It can be observed that for small distortion region, (4) coincides with Gray's result, but as distortion increases it deviates from his result and can be proven to be a strictly better lower bound. The upper bound, $R_u(D)$, is given by

$$R_u(D, \alpha) = D_\alpha \log \alpha - \log(1 + \alpha) - p \log p_\alpha - q \log q_\alpha,$$

where,

$$\begin{aligned} q_\alpha &= 2\sqrt{\alpha}(1 + \sqrt{\alpha})^{-2}, \\ r_\alpha &= q_\alpha/p_\alpha = 2\sqrt{\alpha}(1 + \alpha)^{-1}, \\ D_\alpha &= \left(\frac{\alpha}{1 - \alpha^2} \right) [(pr_\alpha + qr_\alpha^{-1})^2 - \alpha]. \end{aligned}$$

The advantage of these bounds is that they can be easily computed with little computational effort.

IV. RATE-DISTORTION FUNCTION OF BSMS(q): NEW RESULTS

In this section, based on the ideas presented in [7], new upper and lower bounds on the rate-distortion of a BSMS(q) will be derived.

Let $\{X_n\}_{n=-\infty}^{\infty}$ be the output sequence of a BSMS(q) source. Consider every $k + 1$ source output symbols $\{X_{i(k+1)}\}_{i=-\infty}^{\infty}$. Define $S_i \triangleq X_{i(k+1)}$, and super-symbol $Y_i \triangleq X_{i(k+1)+1}^{i(k+1)+k} = (X_{i(k+1)+1}, \dots, X_{i(k+1)+k})$:

$$\dots, X_{-2}, X_{-1}, \underbrace{X_0}_{S_0}, \underbrace{X_1, \dots, X_k}_{Y_0}, \underbrace{X_{k+1}}_{S_1}, \dots$$

Given the $\{S_i\}_{i=-\infty}^{\infty}$ sequence, from the Markovity of the source, Y_i 's would be independent random variables which are not identically distributed, and

$$P(Y_i | \{S_i\}_{i=-\infty}^{\infty}) = P(Y_i | S_i, S_{i+1}).$$

Since the source is binary, (S_i, S_{i+1}) can take on only $2 \times 2 = 4$ different values. As a result, given $\{S_i\}_{i=-\infty}^{\infty}$, Y_i 's would consist of four types of i.i.d. processes. Now given all these, a simple scheme for describing the source within average distortion D , would be as follows:

- 1) describe the side information sequence $\{S_i\}_{i=-\infty}^{\infty}$ losslessly,
- 2) describe the $\{Y_i\}_{i=-\infty}^{\infty}$ sequence within distortion $\tilde{D} = \frac{1}{r_k} D$, where $r_k = \frac{k}{k+1}$.

The total expected per symbol distortion of this scheme would be

$$0 \cdot \frac{1}{k+1} + \tilde{D} \frac{k}{k+1} = D, \quad (8)$$

which satisfies the desired average distortion constraint. Therefore, the rate required for this scheme gives an upper bound on $R(D)$ of a BSMS(q).

The average number of bits per source symbol required for the first step is equal to $\frac{H(X_k | X_0)}{k+1}$. The reason is that

$$R_\ell(D) = \begin{cases} H_b(q) - H_b(D) & 0 \leq D \leq D_2 \\ \max_{q/2 \leq \alpha \leq 1} [D \log \alpha - \log(1 + \alpha) - p \log p_\theta - q \log q_\theta] & D_2 \leq D \leq \frac{1}{2} \end{cases} \quad (4)$$

S_i is a first-order Markov process with a different transition probability than the original sequence, and consequently its lossless description requires $H(X_k|X_0)$ bits per S_i symbol, which in turn is divided by $k + 1$ to give the cost per source symbol. For the second step, where we should describe a mixture of four types of i.i.d. processes within average distortion \tilde{D} , the average required number of bits per source symbol would be

$$r_k \tilde{R}_k(\tilde{D}) = r_k \sum_{(i,j) \in \{0,1\}^2} P(S_0 = i, S_1 = j) R_k^{(i,j)}(\tilde{D}), \quad (9)$$

where, $R_k^{(i,j)}(\tilde{D})$ is the k -th order rate-distortion when X^k is distributed according to $P(X^k|X_0 = i, X_{k+1} = j)$. Each of the k -th order rate-distortion functions in (9) is multiplied by the corresponding $P(S_0 = i, S_1 = j)$ which denotes the proportion of the Y_i sequence that is allocated to the (i, j) -th mode. For each Y_i , the two source symbols that embrace it, namely $X_{i(k+1)}$ and $X_{(i+1)(k+1)}$, determine its mode. Note that for the Markov source considered here, $\tilde{R}_1(D)$ coincides with the erasure rate distortion function of Verdu and Weissman [8].

Combining the rates required for the first and second steps, one gets the following upper bound on the rate-distortion function of a BSMS(q),

$$R(D) \leq r_k \tilde{R}_k(\tilde{D}) + \frac{H(X_k|X_0)}{k+1}. \quad (10)$$

For proving the lower bound, consider the following problem where the decoder desires to describe the $\{Y_i\}$ sequence within a given fidelity constraint, while the $\{S_i\}$ sequence is known by both the encoder and the decoder losslessly beforehand. Any given coding scheme that achieves average distortion D for the BSMS(q), can also be considered as a scheme for encoding the underlying $\{Y_i\}$ sequence as well. In the sequel, it is shown how the average expected loss incurred by the $\{Y_i\}$ sequence using this technique can be upper bounded. Note that by definition,

$$\begin{aligned} \frac{1}{n(k+1)} \sum_{i=0}^{n(k+1)-1} E[d(X_i, \hat{X}_i)] &= \frac{k}{n(k+1)} \sum_{i=0}^{n-1} E[d(Y_i, \hat{Y}_i)] \\ &+ \frac{1}{n(k+1)} \sum_{i=0}^{n-1} E[d(S_i, \hat{S}_i)] \leq D. \end{aligned}$$

Since both terms in the above sum are greater than zero, each of them individually should be less than D as well. This implies that

$$\frac{k}{n(k+1)} \sum_{i=0}^{n-1} E[d(Y_i, \hat{Y}_i)] \leq D,$$

or,

$$\frac{1}{n} \sum_{i=0}^{n-1} E[d(Y_i, \hat{Y}_i)] \leq \frac{k+1}{k} D = \tilde{D}.$$

Thus, the given scheme describes the $\{Y_i\}$ sequence within an average distortion less than \tilde{D} . As a result, from the converse in the rate-distortion coding theorem for mixed i.i.d. sources with known state at both encoder and decoder, since $\tilde{R}_k(\tilde{D})$ is the infimum of the rates required for describing the Y_i sequence with average distortion \tilde{D} at the decoder, one gets the following lower bound on $R(D)$,

$$r_k \tilde{R}_k(\tilde{D}) \leq R(D). \quad (11)$$

Combining (10) and (11), we can bound the rate-distortion function of a BSMS(q) as follows

$$r_k \tilde{R}_k(\tilde{D}) \leq R(D) \leq r_k \tilde{R}_k(\tilde{D}) + \frac{H(X_k|X_0)}{k+1}. \quad (12)$$

A point to note about these bounds is that the difference between the upper bound and lower bound given in (12) goes to zero as k goes to infinity, and can separately be computed without computing the bounds themselves. For instance, for a given δ , choosing $k > \lceil 1/\delta - 1 \rceil$, guarantees $|r_k \tilde{R}_k(\tilde{D}) - R(D)| < \delta$.

Although the difference between the two sides of (12) goes to zero as k goes to infinity, still the lower bound seems to be a loose bound, because in its derivation, it was assumed that the decoder has access to the so-called side information sequence losslessly without spending any rate. In the sequel, an alternative lower bound is derived which will be seen in the next section to be tighter than the previous lower bound, and even outperforming Berger lower bound for moderate values of k .

Define the conditional rate distortion function $R^-(D)$ as follows

$$R^-(D) \triangleq \min_{p(\hat{X}_1|X_1, X_0): E[d(X_1, \hat{X}_1)] \leq D} I(X_1; \hat{X}_1|X_0) \quad (13)$$

and similarly,

$$R_k^-(D) \triangleq \min_{p(\hat{X}^k|X^k, X_0): E[d(X^k, \hat{X}^k)] \leq D} \frac{1}{k} I(X^k; \hat{X}^k|X_0). \quad (14)$$

In the following, it will first be shown that for any first order Markov source $R(D)$ is lower bounded by $R^-(D)$, and then by a simple argument it will be proven that for a BSMS(q), the rate distortion function would also be lower bounded by $R_k^-(D)$.

For a given first-order finite-alphabet Markov source, consider \mathcal{C} to be a source code of length n , rate R , and average expected distortion per symbol less than D . Similar to the

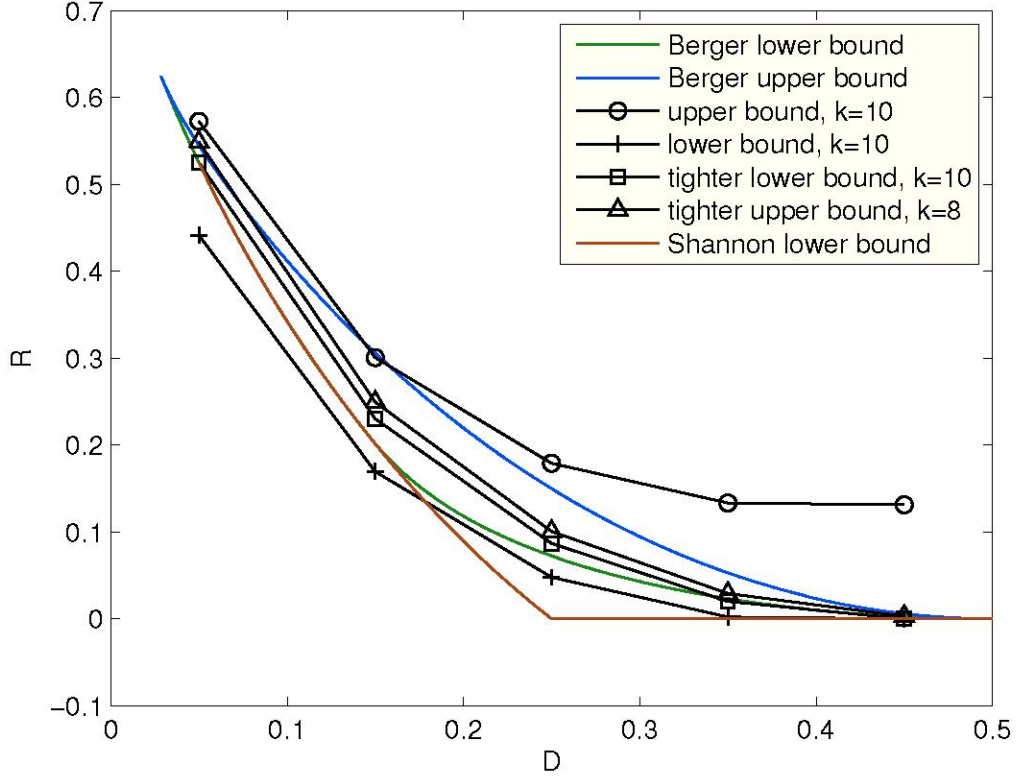


Fig. 1. Comparing the upper/lower bounds on $R(D)$ for a BSMS(q) with $q = 0.25$

inverse proof of the rate-distortion theory, the following series of inequalities holds,

$$\begin{aligned}
 nR &\geq H(\hat{X}^n), \\
 &= I(X^n; \hat{X}^n), \\
 &= \sum_{i=1}^n \left[H(X_i|X^{i-1}) - H(X_i|X^{i-1}, \hat{X}^n) \right], \\
 &\geq \sum_{i=1}^n \left[H(X_i|X_{i-1}) - H(X_i|X_{i-1}, \hat{X}_i) \right], \\
 &\geq \sum_{i=1}^n I(X_i; \hat{X}_i|X_{i-1}), \\
 &\geq I(X_1; \hat{X}_1) + \sum_{i=2}^n R^-(Ed(X_i, \hat{X}_i)), \\
 &\geq \sum_{i=1}^n R^-(Ed(X_i, \hat{X}_i)), \\
 &\geq nR^- \left(\frac{1}{n} \sum_{i=2}^n Ed(X_i, \hat{X}_i) \right) \geq nR^-(D).
 \end{aligned} \tag{15}$$

Therefore, $R(D)$ is lower bounded by $R^-(D)$. Now define

the sequence $\{Z_i\}$, where $Z_i \triangleq X_{(i-1)k+1}^{ik}$. Since the two $\{X_i\}$ and $\{Z_i\}$ sequences are essentially the same, they should have the same rate-distortion function. But since the defined sequence is also a first-order Markov source, the lower bound given by (13) applies to the new source as well. Consequently, since the two source have the same rate-distortion function, we conclude that

$$R(D) \geq \min_{p(\hat{Z}_1|Z_1, Z_0): Ed(Z_1, \hat{Z}_1) \leq D} \frac{1}{k} I(Z_1; \hat{Z}_1|Z_0),$$

or,

$$\begin{aligned}
 R(D) &\geq \min_{p(\hat{X}^k|X^k, X_{-k+1}^0): Ed(X^k, \hat{X}^k) \leq D} \frac{1}{k} I(X^k; \hat{X}^k|X_{-k+1}^0) \\
 &= \min_{p(\hat{X}^k|X^k, X_0): Ed(X^k, \hat{X}^k) \leq D} \frac{1}{k} I(X^k; \hat{X}^k|X_0) \\
 &= R_k^-(D).
 \end{aligned}$$

V. COMPUTING THE BOUNDS

As mentioned in the previous section, the new bounds have the property that for a given δ , one can easily compute the k that guarantees finding $R(D)$ within precision δ . The undesirable fact about these bounds is that for computing $\tilde{R}_k(\tilde{D})$, four k -th order rate distortion functions have to be

computed. Note that the symmetry involved in the structure of the problem implies

$$\begin{aligned} R_k^{(0,0)}(\tilde{D}) &= R_k^{(1,1)}(\tilde{D}), \\ R_k^{(0,1)}(\tilde{D}) &= R_k^{(1,0)}(\tilde{D}). \end{aligned}$$

For every (i, j) , in order to compute $R_k^{(i,j)}(\tilde{D})$, first the corresponding induced probability distribution on the binary vectors of length k , namely $P(X^k | X_0 = i, X_{k+1} = j)$ should be computed, and then solving the geometric programming given in (3), or equivalently in (2) would give the desired result.

Fig. 1 compares the new bounds with Berger upper and lower bounds for a BSMS(q), with $q = 0.25$. The tighter lower bound mentioned in the figure refers to the $R_k^-(D)$, which is computed for $k = 10$. For tighter upper bound, we have used the normal k -th order rate distortion function defined in (6) for $k = 8$. Note that $R_k(D)$ converges to $R(D)$ from above and consequently, for each value of k , can be considered as an upper bound to it.

VI. CONCLUSION AND FUTURE WORK

The rate-distortion function of an i.i.d. source can be computed numerically within desired precision. But in practice the more interesting case would be computing the rate-distortion function of a source that has memory. The reason is that the sources encountered in source coding context are naturally sources with memory. In this paper, binary symmetric Markov sources as one of the simplest sources with memory were investigated, and new upper and lower bounds on their rate-distortion function were derived. The point to note about the new bounds is that they are a sequence of bounds able to bound the rate-distortion function within any precision. Still the problem is that computing the bounds requires computing k dimensional rate-distortion functions which has a complexity growing with k exponentially. As a future work, one can look for new algorithms for computing these k -th order rate-distortion functions more efficiently, using the symmetries and structures involved in the problem.

ACKNOWLEDGMENT

The authors would like to thank Prof. Abbas El Gamal for the helpful discussions. They also acknowledge Stanford Graduate Fellowship supporting the first author, and NSF CAREER grant supporting the second author.

REFERENCES

- [1] C. Shannon, "Coding theorems for a discrete source with a fidelity criterion," IRE Nat. Conv. Rec, 1959.
- [2] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, 1972.
- [3] M. Chiang, S. Boyd "Geometric programming duals of channel capacity and rate distortion," *Information Theory, IEEE Transactions on*, 2004
- [4] R. M. Gray, "Information Rates of Autoregressive Processes," *IEEE Trans. Inform. Theory*, vol. 16, July 1970, pp. 412–421.
- [5] R. M. Gray, "Rate distortion functions for finite-state finite-alphabet Markov sources," *Information Theory, IEEE Transactions on*, 1971
- [6] T. Berger, "Explicit bounds to $R(D)$ for a binary symmetric Markov source," *IEEE Trans. on Inform. Theory*, 1977.

- [7] T. Weissman, A. El Gamal, "Source Coding With Limited-Look-Ahead Side Information at the Decoder," *IEEE Trans. on Inform. Theory*, Dec. 2006, pp. 5218–5239.
- [8] S. Verdú and T. Weissman, "Erasure entropy," in *Proc. Int. Symp. Inf. Th.*, Seattle, USA, July 2006, pp. 98–102.
- [9] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.