RC22241 (W0111-026) November 8, 2001
Electrical Engineering

IBM Research Report

On Successive Refinement of the Binary Symmetric Markov Source

Luis A. Lastras
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Toby Berger
School of Electrical Engineering,
Cornell University, Ithaca, NY 14853

Research Division
Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich

On Successive Refinement of the Binary Symmetric Markov Source

Luis Lastras, Toby Berger *

October 20, 2005

Abstract

We show that for every $n > 2$ the standard nth order approximation $R_n(D)$, to the rate-distortion function of the binary symmetric Markov source is not successively refinable under the Hamming distortion measure in an open interval of the form $D_n < D < 1/2 = D_{\max}$.

1 Preliminaries

The phenomenon of successive refinement [1], [2], [3], [4], [5], [6], [7], [8] has attracted much attention in the information theory community because of the interesting non trivial classification that it enables for data sources. Roughly speaking, a source is said to be successively refinable (SR) with respect to a given distortion measure if for any given distortion levels D_1 and $D_2 < D_1$, it is possible to design a two stage lossy compression code the first stage of which is optimum at D_1 which can then be refined in the second stage so as to achieve the same performance of the best single-step code at distortion level D_2. An interesting recent finding is that non-SR sources cannot suffer from arbitrarily large performance degradation when using two-stage codes [9].

Another research endeavor to which our present work is strongly linked to is the theory of compression of sources with memory. The question of determining the exact achievable region for coding systems that process these sources was settled in a reasonably general setting in [10], [11], [12] (see also they survey of Kieffer [13] which includes references to his own work); for the progressive coding case see [7]. However, the nature of this solution is such that, with the exception of the important quadratic Gaussian case, it is difficult to obtain closed form expressions for said achievable region even for the simplest of sources. For example, consider the binary symmetric Markov source (BSMS) in Figure 1. In evaluating the quality of an approximation to a sample

*Luis Lastras is with the IBM TJ Watson Research Center, Yorktown Heights, NY 10598. Toby Berger is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA

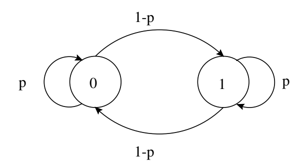

Figure 1: The binary symmetric Markov source

path of BSMS using a coding system, we will employ the standard Hamming distance. For every positive integer n, let BSMS_n denote the iid data source that generates symbols from the alphabet $\{0,1\}^n$ using the probability law obtained from the projection of the original measure along the first n symbols of BSMS. Also, let $R_n(D)$ denote the rate-distortion function of BSMS_n under the bitwise hamming distortion metric; it is well known that the minimum rate required to reproduce BSMS within distortion D is given by $R(D) = \lim_{n\to\infty} R_n(D)$. Previous efforts to obtain a closed form expression for the rate-distortion function of BSMS have not proved successful [14, 15] aside from the ground breaking work of Gray [16] in which such an expression was obtained for a nonzero range of small distortions $0 \le D \le D_c$. For $D > D_c$ the structure of the optimal forward channel that attains the optimization problem that defines $R_n(D)$ is quite complex if $n > 3$, as evidenced by the closed form expressions for said channel found for the $n = 4$ case [15].

This work is an attempt to establish a connection between these two important fields. A worthy goal would be to resolve whether or not BSMS is SR. It is not difficult to see that BSMS is SR in the range $(0, D_c]$ and that for $n = 1, 2$, BSMS_n is SR in $(0, D_{\max} = 1/2]$; these facts are proved in the Appendix. Note however that current tests for successive refinement [3] require us to verify a Markov condition on the optimal forward channels which, even in the single stage setting, are not currently analytically understood for $D > D_c$ and $n > 4$; therefore attacking this problem with our current tools promises to be difficult.

Our strategy was as follows: if we could show that for every $n > 2$, BSMS_n is SR at least in some subset of $(D_c, D_{\max}]$, this would suggest that it may be worthwile to try to prove that BSMS is indeed SR.

For every $n > 1$, there is a small region of distortions $(D_n, 1/2]$ in which there exists a simple characterization of the sequences of symbols that are used when building codewords of a good code. Exploring these regions seemed a natural next step. However, as we show below, for every $n > 2$, BSMS_n is *not* SR when refining from D_1 to D_2 for any $D_n < D_2 < D_1 < 1/2$. We point out that $R_n(D)$ is always a lower bound to the rate of any practical system that attempts to encode the "super-n-symbols" of BSMS_n using either "super-scalar" quantizers or "super-block" codes, so our result is of practical significance. Also, to our knowledge this is the first instance of an infinite number of examples of non-SR sources. We strongly believe that $D_n \to 1/2$ as $n \to \infty$ and even if it did not, it would still be possible for BSMS, the limit of the BSMS$_n$, to be SR despite the fact that BSMS$_n$ is non-SR for each $n > 2$. Thus the main problem remains open.

We have organized this correspondence as follows: in Section 2 we give a formal description of the source that we have targeted. Section 3 contains the main result. The Appendix contains other support material, including proofs of facts discussed in these Preliminaries.

2 Information source description

For any n, let S^n be the ordered set of all binary n-sequences. We assume the usual lexicographic order which is evidenced by the following recursive construction for S^n:

$$S^1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \qquad S^n = \left[\begin{array}{c|c} \mathbf{0} & S^{n-1} \\ \hline \mathbf{1} & S^{n-1} \end{array}\right]$$

where $\mathbf{0}$ (resp. $\mathbf{1}$) denotes a column vector of all zeros (resp. all ones) of the appropiate size. For each n the BSMS probability law specifies a probability mass function over the sequences in S^n, which can be compactly represented using the recursive arguments that are next given. Let $p > 1/2$ be the probability that $X_i = X_{i-1}$ given X_{i-1}, and let $q = 1 - p$; the assumption that $p > 1/2$ is made without any loss of generality. Let $\mathbf{p_n}$ denote the column vector with 2^n entries with jth element equal to the probability assigned to the jth n-sequence. Let $\mathbf{p_n^0}$ (resp. $\mathbf{p_n^1}$) denote the upper (resp. lower) half of $\mathbf{p_n}$; $\mathbf{p_n^0}$ contains the probability of those n-sequences with leftmost symbol equal to zero, and $\mathbf{p_n^1}$ contains the probability of those n-sequences with leftmost symbol equal to one. It is not difficult to see that the $\mathbf{p_n^0}$ and $\mathbf{p_n^1}$ vectors can be expressed recursively using $\mathbf{p_{n-1}^0}$, $\mathbf{p_{n-1}^1}$ as follows:

$$\mathbf{p_n^0} = \begin{bmatrix} \mathbf{p_{n-1}^0}p \\ \mathbf{p_{n-1}^1}q \end{bmatrix}; \qquad \mathbf{p_n^1} = \begin{bmatrix} \mathbf{p_{n-1}^0}q \\ \mathbf{p_{n-1}^1}p \end{bmatrix}$$

with initialization

$$\mathbf{p_1^0} = \mathbf{p_1^1} = \left[\frac{1}{2}\right]$$

The vector $\mathbf{p_n}$ is then obtained by stacking the two halves:

$$\mathbf{p_n} \quad = \quad \begin{bmatrix} \mathbf{p_n^0} \\ \mathbf{p_n^1} \end{bmatrix}.$$

We will define the function $p^n : \{0,1\}^n \to [0,1]$ to be the probability mass function that is naturally induced by the vector $\mathbf{p_n}$. Now assemble an iid source $\mathbf{X_1}, \mathbf{X_2}, \cdots$ with range $\{0,1\}^n$ and marginal probability mass function p^n. We refer to this source as BSMS$_n$. We show that BSMS$_n$ is not successively refinable under the Hamming distortion measure for all $n > 2$. Whether BSMS itself is successively refinable or not remains open.

3 Results

Let d_H denote the Hamming distance metric; that is, $d_H(\mathbf{x}, \mathbf{y})$ denotes the number of positions in which the binary sequences $\mathbf{x}$ and $\mathbf{y}$ differ. Let $R_n(D)$ denote the rate distortion function of $\mathbf{p_n}$ with respect to Hamming distance. Recall that $R_n(D)$ is defined as

$$R_n(D) = n^{-1} \min_{Q \in \mathcal{Q}} I(Q)$$

where

$$\begin{aligned} \mathcal{Q} &\triangleq \{Q : n^{-1} d(Q) \leq D\} \\ I(Q) &\triangleq \sum_{\mathbf{x}, \mathbf{y} \in S^n} p_{\mathbf{x}} Q_{\mathbf{y}|\mathbf{x}} \log_2 \frac{Q_{\mathbf{y}|\mathbf{x}}}{\sum_{\hat{\mathbf{x}} \in S^n} p_{\hat{\mathbf{x}}} Q_{\mathbf{y}|\hat{\mathbf{x}}}} \\ d(Q) &\triangleq \sum_{\mathbf{x}, \mathbf{y} \in S^n} p_{\mathbf{x}} Q_{\mathbf{y}|\mathbf{x}} d_H(\mathbf{x}, \mathbf{y}) \end{aligned}$$

This constrained minimization problem is difficult to solve even for modest values of n. However, for high distortion and for any n, the optimal Q^* has a simple structure, a fact which will be exploited to establish the desired result. The following theorem gives necessary and sufficient conditions for a test channel Q attaining the minimum in the definition of $R_n(D)$ [12] (p. 35):

Theorem 1 *A necessary and sufficient condition for a conditional probability assignment Q to yield a point on the $R(D)$ curve at which the slope equals $\log a$ is the existence of a marginal output q such that*

$$Q_{\mathbf{y}|\mathbf{x}} = \lambda_{\mathbf{x}} q_{\mathbf{y}} a^{d(\mathbf{x},\mathbf{y})} \tag{1}$$

$$c_{\mathbf{y}} := \sum_{\mathbf{x} \in S^n} \lambda_{\mathbf{x}} p_{\mathbf{x}} a^{d(\mathbf{x},\mathbf{y})} \begin{cases} = 1 \ \textit{if} \ q_{\mathbf{y}} > 0 \\ \leq 1 \ \textit{if} \ q_{\mathbf{y}} = 0 \end{cases} \quad (*) \tag{2}$$

where

$$\lambda_{\mathbf{x}}^{-1} = \sum_{\mathbf{y} \in S^n} q_{\mathbf{y}} a^{d(\mathbf{x},\mathbf{y})}. \tag{3}$$

Given some marginal q and some $a \in (0,1]$, call the conditional probability obtained through Equations (1) and (3) the *"channel" generated by q in a*.

We now pause to establish a basic property of the rate-distortion function of BSMS_n. The smallest value of D for which $R(D)$ is equal to zero is called $D_{\max}$. This number can be computed using the following formula: ([12], pag. 27)

$$D_{\max} = \min_{\mathbf{y}} \left(\sum_{\mathbf{x} \in S^n} p_{\mathbf{x}} d_{\mathbf{x},\mathbf{y}} \right) \tag{4}$$

We will now prove that for any n, $D_{\max} = n/2$ by showing that the summation inside the parenthesis in the preceding equation is independent of $\mathbf{y}$ and equal to $n/2$. Let $\mathbf{D}_n$ represent the Hamming distortion matrix for sequences of length n, this is, $\{\mathbf{D}_n\}_{j,k} = d_H(j,k)$; note that $\mathbf{D}_n$ is a symmetric matrix. Therefore, we claim that for all n,

$$\mathbf{D}_n \mathbf{p_n} = \frac{n}{2}\mathbf{1}$$

where $\mathbf{1}$ denotes a column vector of all ones of the appropiate size. This is easily shown by induction. Let $\mathcal{I}$ denote a matrix of all ones of the appropiate size, and note that $\mathcal{I}\mathbf{p_n^0} = \mathcal{I}\mathbf{p_n^1} = \frac{1}{2}\mathbf{1}$. The case $n = 1$ is clearly true; assume that $\mathbf{D}_{n-1}\mathbf{p_{n-1}} = \frac{n-1}{2}\mathbf{1}$. Then

$$\begin{aligned}
\mathbf{D}_n \mathbf{p_n} &= \begin{bmatrix} \mathbf{D}_{n-1} & \mathbf{D}_{n-1} + \mathcal{I} \\ \mathbf{D}_{n-1} + \mathcal{I} & \mathbf{D}_{n-1} \end{bmatrix} \begin{bmatrix} \mathbf{p_n^0} \\ \mathbf{p_n^1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{D}_{n-1}(\mathbf{p_n^0} + \mathbf{p_n^1}) + \frac{1}{2}\mathbf{1} \\ \mathbf{D}_{n-1}(\mathbf{p_n^0} + \mathbf{p_n^1}) + \frac{1}{2}\mathbf{1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{D}_{n-1}\mathbf{p_{n-1}} + \frac{1}{2}\mathbf{1} \\ \mathbf{D}_{n-1}\mathbf{p_{n-1}} + \frac{1}{2}\mathbf{1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{n}{2}\mathbf{1} \\ \frac{n}{2}\mathbf{1} \end{bmatrix} = \frac{n}{2}\mathbf{1}
\end{aligned}$$

In this correspondence we have directed our attention to the region where the slope of the rate-distortion function of BSMS_n is close to zero. It is not difficult to prove that, for any n, the slope of $R_n(\cdot)$ at $D_{\max}$ is equal to zero, thereby associating the high distortion region with the range of small negative slopes. The argument is as follows: let $\hat{q}$ be the probability mass function that assigns probability 1 to a binary n-sequence that attains the minimum in Equation (4) and zero mass to every other sequence and let $\hat{Q}_{\mathbf{y}|\mathbf{x}} = q_{\mathbf{y}}$. From Theorem 1 it is easily seen that $\hat{Q}$ satisfies the necessary conditions to yield a point on the rate-distortion function with rate and slope equal to zero and distortion $D_{\max} = 1/2$.

The following theorem characterizes the optimal output probabilities for every n at a certain high distortion region:

Theorem 2 *Let n be a fixed integer. If $p > 1/2$, there exists an interval $(a_n^*, 1)$ such that for any $a \in (a_n^*, 1)$, the assignment $q^*_{00\cdots 0} = q^*_{11\cdots 1} = 1/2$ has the property that $c^*_{00\cdots 0} = c^*_{11\cdots 1} = 1$ and $\forall\, \mathbf{y} \in S^n - \{00\cdots 0, 11\cdots 1\}$, $c^*_{\mathbf{y}} < 1$. Hence, the channel generated by q^* in a is optimal.*

For the proof we refer the reader to subsection 3.1 and the support material [17].

The marginal q^*, as defined in the statement of Theorem 2, also possesses the property of being the unique optimal solution in the high distortion region, a result that is proved in the Appendix.

We are now in position to prove that BSMS$_n$ is not successively refinable. Let a_1 and a_2 be such that $a_n^* < a_2 < a_1 < 1$. Let Q^1 and Q^2 denote the channels generated by q^* in a_1 and a_2, respectively. We show that the source is not successively refinable from $d(Q^1)$ to $d(Q^2)$. If successive refinement were achievable, there would exist a channel $Q_{\mathbf{Y_1},\mathbf{Y_2}|\mathbf{X}}$ such that the random variables $\mathbf{X}, \mathbf{Y_1}, \mathbf{Y_2}$ jointly defined through said channel and $\mathbf{p_n}$ satisfy

- $\mathbf{X} \longrightarrow \mathbf{Y_2} \longrightarrow \mathbf{Y_1}$
- $Ed(\mathbf{X}, \mathbf{Y_i}) = d(Q^i)$, $i \in \{1, 2\}$
- $I(\mathbf{X}; \mathbf{Y_i}) = I(Q^i)$, $i \in \{1, 2\}$

We will refer to the different marginal and conditional probabilities of the joint distribution $Q_{\mathbf{Y_1},\mathbf{Y_2}|\mathbf{X}}p^n$ using the appropriate subindex. Thus $Q_{\mathbf{Y_1}|\mathbf{Y_2}}$ denotes the conditional probability of $\mathbf{Y_1}$ given $\mathbf{Y_2}$, $Q_{\mathbf{Y_2}|\mathbf{X}}$ denotes the conditional probability of $\mathbf{Y_2}$ given $\mathbf{X}$, and $q_{\mathbf{Y_1}}$ and $q_{\mathbf{Y_2}}$ denote the marginals of $\mathbf{Y_1}$ and $\mathbf{Y_2}$, respectively.

By the uniqueness of the optimal output marginals at a_1 and a_2, $q_{\mathbf{Y_1}} = q_{\mathbf{Y_2}} = q^*$. From this fact it can be deduced that the forward channel $Q_{\mathbf{Y_1}|\mathbf{Y_2}}$ is a binary symmetric channel:

$Q_{\mathbf{Y_1}\|\mathbf{Y_2}}(\mathbf{y_1}\|\mathbf{y_2})$	$\mathbf{y_1} = 00\cdots 0$	$\mathbf{y_1} = 11\cdots 1$
$\mathbf{y_2} = 00\cdots 0$	$1 - t$	t
$\mathbf{y_2} = 11\cdots 1$	t	$1 - t$

(5)

with some crossover probability t (we would like to point out that Chow and Berger [8] attacked their problem in a similar manner). The Markov property implies that

$$Q_{\mathbf{Y_1},\mathbf{Y_2}|\mathbf{X}}(\mathbf{y_1}, \mathbf{y_2}|\mathbf{x})p^n(\mathbf{x}) = Q_{\mathbf{Y_1}|\mathbf{Y_2}}(\mathbf{y_1}|\mathbf{y_2})Q_{\mathbf{Y_2}|\mathbf{X}}(\mathbf{y_2}|\mathbf{x})p^n(\mathbf{x})$$

By summing over all $\mathbf{y_2}$ and dividing by $p^n(\mathbf{x})$ (which for BSMS$_n$ is always a positive number if $p \in (0,1)$), we see that

$$Q_{\mathbf{Y_1}|\mathbf{X}}(\mathbf{y_1}|\mathbf{x}) = \sum_{\mathbf{y_2}\in S^n} Q_{\mathbf{Y_1}|\mathbf{Y_2}}(\mathbf{y_1}|\mathbf{y_2}) Q_{\mathbf{Y_2}|\mathbf{X}}(\mathbf{y_2}|\mathbf{x}) \tag{6}$$

Invoking uniqueness again, we see that $Q_{\mathbf{Y_1}|\mathbf{X}}$ and $Q_{\mathbf{Y_2}|\mathbf{X}}$ are the channels generated by q^* at a_1 and a_2, respectively:

$$Q_{\mathbf{Y_i}|\mathbf{X}}(\mathbf{y_i}|\mathbf{x}) \quad = \quad \lambda_i(\mathbf{x}) q^*(\mathbf{y_i}) a_i^{d_H(\mathbf{x},\mathbf{y_i})} \qquad i \in \{1,2\} \tag{7}$$

Substituting (7) in (6) we obtain

$$\frac{\lambda_1(\mathbf{x})}{\lambda_2(\mathbf{x})} q^*(\mathbf{y_1}) a_1^{d_H(\mathbf{x},\mathbf{y_1})} = \sum_{\mathbf{y_2}\in S^n} Q_{\mathbf{Y_1}|\mathbf{Y_2}}(\mathbf{y_1}|\mathbf{y_2}) q^*(\mathbf{y_2}) a_2^{d_H(\mathbf{x},\mathbf{y_2})}$$

Define $d_{\mathbf{x}} \triangleq d_H(\mathbf{x}, 00\cdots 0)$. Substituting $\mathbf{y_1} = 00\cdots 0$ and using (5) we see that

$$\begin{aligned} \frac{\lambda_1(\mathbf{x})}{\lambda_2(\mathbf{x})} a_1^{d_{\mathbf{x}}} &= (1-t) a_2^{d_{\mathbf{x}}} + t a_2^{n-d_{\mathbf{x}}} \\ &= a_2^{d_{\mathbf{x}}} + t(a_2^{n-d_{\mathbf{x}}} - a_2^{d_{\mathbf{x}}}) \end{aligned}$$

After noting that

$$\frac{\lambda_1(\mathbf{x})}{\lambda_2(\mathbf{x})} = \frac{a_2^{d_{\mathbf{x}}} + a_2^{n-d_{\mathbf{x}}}}{a_1^{d_{\mathbf{x}}} + a_1^{n-d_{\mathbf{x}}}}$$

we obtain

$$\frac{1 + a_2^{n-2d_{\mathbf{x}}}}{1 + a_1^{n-2d_{\mathbf{x}}}} = 1 + t(a_2^{n-2d_{\mathbf{x}}} - 1) \tag{8}$$

If $n = 2d_{\mathbf{x}}$ then (8) simplifies to the trivial identity $1 = 1$. If $n \neq 2d_{\mathbf{x}}$, we obtain the following expression for t:

$$t \quad = \quad \phi\left(a_1, a_2, n - 2d_{\mathbf{x}}\right) \qquad \mathbf{x} \in \{\mathbf{x} : n \neq 2d_{\mathbf{x}}\} \tag{9}$$

where

$$\phi(a_1, a_2, b) = \frac{a_1^b - a_2^b}{(1 + a_1^b)(1 - a_2^b)}$$

Call a sequence $\mathbf{x}$ *valid* if $n \neq 2d_{\mathbf{x}}$. The left hand side of (9) is a constant while its right hand side has the potential to be equal to two different numbers when evaluating it at two different valid sequences $\mathbf{x_1}$ and $\mathbf{x_2}$. If $n = 2$, direct evaluation shows that

$$\phi\left(a_1, a_2, 2 - 2d_{00}\right) \quad = \quad \phi\left(a_1, a_2, 2 - 2d_{11}\right)$$

and since 01 and 10 are not valid sequences, we cannot obtain a contradiction; in fact in [17] it is shown that BSMS_2 is successively refinable everywhere. If $n > 2$ then a contradiction can always be found through appropiate selection of sequences $\mathbf{x_1}$ and $\mathbf{x_2}$. The main ingredient of this proof is the following Lemma which we establish in the Appendix:

Lemma 1 *If $a_2 < a_1$ then $\phi(a_1, a_2, b)$ is a strictly decreasing function of b in the domain $b \in (0, \infty)$.*

Assume that $n > 2$. Choose

$$\begin{aligned} \mathbf{x_1} &= 100 \cdots 0 \\ \mathbf{x_2} &= 000 \cdots 0 \end{aligned}$$

and note that

$$\begin{aligned} n - 2d_{\mathbf{x_1}} &= n - 2 \quad > \quad 0 \\ n - 2d_{\mathbf{x_2}} &= n \end{aligned}$$

and hence

$$0 < n - 2d_{\mathbf{x_1}} < n - 2d_{\mathbf{x_2}}$$

Using Lemma 1 we obtain

$$\phi\left(a_1, a_2, n - 2d_{\mathbf{x_1}}\right) > \phi\left(a_1, a_2, n - 2d_{\mathbf{x_2}}\right)$$

which yields the contradiction we sought to obtain.

Thus, we conclude that if $n > 2$, then BSMS_n is not successively refinable under the Hamming distortion measure. Note however that this fact does not imply that BSMS is not successively refinable, since it is possible for the positive rate losses to become arbitrarily small as n grows. It seems improbable that lower bounding the minimal rate losses possible in the high distortion region will yield the desired result, since it is strongly believed that $a_n^* \to 1$ as $n \to \infty$, making it impossible to choose interesting values for a_1 and a_2.

3.1 Proof of Theorem 2

We shall find it useful to introduce the convention that a sequence $b_1 b_2 \cdots b_n$ will be associated with the integer $1 + \sum_{i=1}^{n} b_i 2^{i-1}$. We need to verify that $c_{00\cdots 0} = 1, c_{11\cdots 1} = 1$ and $c_k < 1$ for any other output reproduction sequence $k \notin \{1, 2^n\}$. The expression for λ_j is as follows:

$$\lambda_j^{-1} = \frac{1}{2} a^{d_{j,1}} + \frac{1}{2} a^{n - d_{j,1}}$$

where $d_{j,1}$ denotes the Hamming weight of the sequence j, and hence

$$c_k = \sum_{j=1}^{2^n} \mathbf{p_{n,j}} \frac{a^{d_{j,k}}}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})}$$

Since the sequence indexed by $2^n - j + 1$ is the bit-wise complement of the sequence indexed by j, we have $\mathbf{p_{n,2^n-j+1}} = \mathbf{p_{n,j}}$ and $d_{2^n-j+1,k} = n - d_{j,k}$. Therefore, an alternative expression for c_k is:

$$c_k = \sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} \frac{a^{d_{j,k}} + a^{n-d_{j,k}}}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})} \tag{10}$$

In particular,

$$\begin{aligned} c_{00\cdots 0} &= \sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} \frac{a^{d_{j,1}} + a^{n-d_{j,1}}}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})} \\ &= 2\sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} = 2\frac{1}{2} = 1 \end{aligned}$$

It is clear that $c_{11\cdots 1} = 1$ can be shown in a similar manner. This holds for all values of a, in particular, it is true for a in the interval $(a_n^*, 1)$ identified in the discussion that follows.

The arguments that we provide to demonstrate that $c_k < 1$ for $k \notin \{00\cdots 0, 11\cdots 1\}$ when $p > \frac{1}{2}$ are considerably longer. Evaluating (10) at $a = 1$ we see that for all k,

$$c_k|_{a=1} = 1 \tag{11}$$

Note that

$$\frac{d}{da}c_k = \sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} \left(\frac{d}{da} \frac{a^{d_{j,k}} + a^{n-d_{j,k}}}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})} \right) \tag{12}$$

It is not hard to see that for all n and for all k,

$$\left.\frac{d}{da}c_k\right|_{a=1} = 0$$

Thus c_k has a critical point at $a = 1$, where we think of c_k as a function of a. It will be shown that

Lemma 2 *For all n, and for $k \notin \{1, 2^n\}$*

$$\left.\frac{d^2}{da^2}c_k\right|_{a=1} < 0$$

This lemma shows that the critical point is a maximum. Combining this fact with Equation (11) allows us to conclude the existence of an open interval $(a_{n,k}, 1)$ with $a_{n,k} < 1$ in which $c_k < 1$. Setting a_n to be the largest of the $a_{n,k}$ proves Theorem 2. It remains to prove Lemma 2. We give the essentials of the argument but relegate some of the lengthier (but more mechanical) calculations to a technical report [17].

Proof of Lemma 2. Let $\mathbf{D}_n$ represent the Hamming distortion matrix for sequences of length n (i.e.$\{\mathbf{D}_n\}_{j,k} = d_{j,k}$), $[A]^2$ denote the matrix with entries equal to the square of the corresponding element of the original matrix A, and $\mathbf{p_n}$ denote the probability vector of the 2^n source sequences. Let

$$\gamma_{\mathbf{n}} = [\mathbf{D}_n]^2 \, \mathbf{p_n}$$

We next show that

$$\left. \frac{d^2}{da^2} c_k \right|_{a=1} = \gamma_{\mathbf{n,k}} - \gamma_{\mathbf{n,1}}$$

where $\gamma_{\mathbf{n,j}}$ represents the jth element of the $\gamma_{\mathbf{n}}$ vector.

The derivative that appears inside of the parenthesis in (12) is equal to

$$\begin{aligned} \frac{d}{da} \frac{a^{d_{j,k}} + a^{n-d_{j,k}}}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})} &= \\ & \frac{(a^{d_{j,1}} + a^{n-d_{j,1}})(d_{j,k} a^{d_{j,k}-1} + (n-d_{j,k}) a^{n-d_{j,k}-1}) - (a^{d_{j,k}} + a^{n-d_{j,k}})(d_{j,1} a^{d_{j,1}-1} + (n-d_{j,1}) a^{n-d_{j,1}-1})}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})^2} \\ &\triangleq \frac{A}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})^2} \end{aligned}$$

Taking a derivative again:

$$\begin{aligned} \frac{d}{da} \frac{A}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})^2} &= \\ & \frac{(a^{d_{j,1}} + a^{n-d_{j,1}})^2 \frac{d}{da} A - A \frac{d}{da}(a^{d_{j,1}} + a^{n-d_{j,1}})^2}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})^4} \end{aligned} \tag{13}$$

It is easy to verify that

$$A|_{a=1} = 0$$

Hence, in order to evaluate (13) at $a = 1$, we only need to evaluate $\frac{d}{da} A$ at $a = 1$:

$$\left. \frac{d}{da} A \right|_{a=1} = 2\left((d_{j,k} - n)^2 - (d_{j,1} - n)^2 + d_{j,k}^2 - d_{j,1}^2\right)$$

therefore

$$\left. \frac{d^2}{da^2} \frac{a^{d_{j,k}} + a^{n-d_{j,k}}}{\frac{1}{2}(a^{d_{j,1}} + a^{n-d_{j,1}})} \right|_{a=1} = (d_{j,k} - n)^2 - (d_{j,1} - n)^2 + d_{j,k}^2 - d_{j,1}^2$$

thus concluding that

$$
\begin{aligned}
\left.\frac{d^2}{da^2}c_k\right|_{a=1} &= \sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} \left((d_{j,k}-n)^2 - (d_{j,1}-n)^2 + d_{j,k}^2 - d_{j,1}^2\right) \\
&= \sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} \left((d_{j,k}-n)^2 - (d_{j,1}-n)^2\right) + \sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} \left(d_{j,k}^2 - d_{j,1}^2\right) \\
&= \sum_{j=2^{n-1}+1}^{2^n} \mathbf{p_{n,j}} \left(d_{j,k}^2 - d_{j,1}^2\right) + \sum_{j=1}^{2^{n-1}} \mathbf{p_{n,j}} \left(d_{j,k}^2 - d_{j,1}^2\right) \\
&= \sum_{j=1}^{2^n} \mathbf{p_{n,j}} \left(d_{j,k}^2 - d_{j,1}^2\right) \\
&= \sum_{j=1}^{2^n} \mathbf{p_{n,j}} d_{k,j}^2 - \sum_{j=1}^{2^n} \mathbf{p_{n,j}} d_{1,j}^2 \\
&= \gamma_{\mathbf{n,k}} - \gamma_{\mathbf{n,1}}
\end{aligned}
$$

In order to prove the Lemma, we must show that

$$\gamma_{\mathbf{n,1}} > \gamma_{\mathbf{n,k}} \tag{14}$$

for all n and for all $k \notin \{1, 2^n\}$. This can be done using recursion arguments exploiting the symmetries in the definitions of the quantities involved but the complete proof is somewhat lengthy. Accordingly, we have relegated these details to a technical report [17].

4 Appendix

4.1 BSMS is refinable for low distortions

From Gray's [16] landmark paper (see also pags. 252-253 of [12]), it is known that the following decomposition holds for BSMS$_n$: there exists a constant $D_c^{(n)}$ such that for $0 \leq D \leq D_c^{(n)}$,

$$R_n(D) = n^{-1}H(\mathbf{X}) - H_b(D)$$

where $\mathbf{X}$ is a $\{0,1\}^n$-valued random vector distributed according to the law that governs BSMS$_n$. The critical distortions $D_c^{(n)}$ decrease with increasing n and have limit

$$\lim_{n\to\infty} D_c^{(n)} \triangleq D_c = \frac{1}{2}\left\{1 - \sqrt{1 - \left[\frac{m}{1-m}\right]^2}\right\}$$

where $m = \min(p, 1-p)$. Furthermore, for any n and letting $\mathbf{Y}$ be the $\{0,1\}^n$-valued random variable that attains the minimum in the variational problem that defines $R_n(D)$, it is known that in the range $0 \le D \le D_c^{(n)}$,

$$\mathbf{X} = \mathbf{Y} \oplus \mathbf{N}$$

where $\oplus$ denotes the element-wise XOR operation, $\mathbf{N}$ is a $\{0,1\}^n$-valued random vector with independent binary entries such that $\text{Prob}(N_i = 1) = D$ and also $\mathbf{N}$ is statistically independent from $\mathbf{Y}$.

Now let $0 < D_1 \le D_c$ be a first distortion level and $0 \le D_2 < D_1$ be a second distortion level to which we will refine BSMS_n. Define $\mathbf{N^{(2)}}$ to be a binary iid random n-vector with $\text{Prob}(N_i^{(2)} = 1) = D_2$. Also define $\mathbf{N^{(1)}}$ to be a binary iid random n-vector with $\text{Prob}(N_i^{(1)} = 1) = (D_1 - D_2)/(1-2D_2)$ and further assume that every entry in both the $\mathbf{N^{(1)}}$ and $\mathbf{N^{(2)}}$ random vectors is statistically independent of every other random variable that we have defined. Then

$$\mathbf{X} \to \mathbf{X} \oplus \mathbf{N^{(2)}} \to \mathbf{X} \oplus \mathbf{N^{(2)}} \oplus \mathbf{N^{(1)}} \tag{15}$$

and

$$n^{-1} I(\mathbf{X}; \mathbf{X} \oplus \mathbf{N^{(2)}}) = R_n(D_2); \qquad n^{-1} I(\mathbf{X}; \mathbf{X} \oplus \mathbf{N^{(2)}} \oplus \mathbf{N^{(1)}}) = R_n(D_1)$$
$$n^{-1} E d_H(\mathbf{X}, \mathbf{X} \oplus \mathbf{N^{(2)}}) = D_2; \qquad n^{-1} E d_H(\mathbf{X}, \mathbf{X} \oplus \mathbf{N^{(2)}} \oplus \mathbf{N^{(1)}}) = D_1$$

BSMS is a discrete alphabet, stationary and ergodic random process and therefore falls in the class of sources for which the distortion-rate region is completely characterized by the results of Effros [7]. A precise definition and characterization of the achievable distortion-rate region can be found in Effros' work; here we limit ourselves to demonstrate that the point

$$(R(D_1), R(D_2) - R(D_1), D_1, D_2)$$

is in her achievable distortion-rate region. Note that Effros identifies the second coordinate in the vector above with the additional rate needed to achieve the refinement distortion D_2. All further references to achievable vectors are made under the conventions of [7].

The principal result of interest from [7] is Theorem 2, which combined with the comments on the second column of page 1891 imply that the vector $(R_1, R_2 - R_1, D_1, D_2)$ is achievable by the class of fixed-rate codes if

$$J(\alpha_1, \alpha_2, \beta_1, \beta_2) \le \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 R_1 + \beta_2 (R_2 - R_1)$$

for all $\alpha_1, \alpha_2, \beta_1, \beta_2 \ge 0$. The function $J(\alpha_1, \alpha_2, \beta_1, \beta_2)$ admits the following characterization ([7] Lemma 6):

$$J(\alpha_1, \alpha_2, \beta_1, \beta_2) = \lim_{n \to \infty} J_n(\alpha_1, \alpha_2, \beta_1, \beta_2)$$

where $J_n(\cdot)$ is defined as ([7] Equation 16)

$$J_n(\alpha_1,\alpha_2,\beta_1,\beta_2) = \inf_{\mathbf{Y_1},\mathbf{Y_2}} n^{-1}\left(\alpha_1 E d_H(\mathbf{X},\mathbf{Y_1}) + \alpha_2 E d_H(\mathbf{X},\mathbf{Y_2}) + \beta_1 I(\mathbf{X};\mathbf{Y_1}) + \beta_2 I(\mathbf{X};\mathbf{Y_2}|\mathbf{Y_1})\right)$$

If $\alpha_1,\alpha_2,\beta_1,\beta_2 \geq 0$, an upper bound to $J_n(\alpha_1,\alpha_2,\beta_1,\beta_2)$ can be obtained by making the substitutions $\mathbf{Y_1} \to \mathbf{X} \oplus \mathbf{N_1}$, $\mathbf{Y_2} \to \mathbf{X} \oplus \mathbf{N_2} \oplus \mathbf{N_1}$ and using the Markov chain condition (15):

$$J_n(\alpha_1,\alpha_2,\beta_1,\beta_2) \leq \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 R_n(D_1) + \beta_2(R_n(D_2) - R_n(D_1))$$

Taking the limit as $n \to \infty$, we obtain

$$J(\alpha_1,\alpha_2,\beta_1,\beta_2) \leq \alpha_1 D_1 + \alpha_2 D_2 + \beta_1 R(D_1) + \beta_2(R(D_2) - R(D_1))$$

Since this expression is valid for all $\alpha_1,\alpha_2,\beta_1,\beta_2 \geq 0$, from our earlier comments one concludes that $(R(D_1), R(D_2) - R(D_1), D_1, D_2)$ is achievable.

4.2 Successive refinement of BSMS$_n$ for $n = 1, 2$

For $n = 1$, BSMS$_n$ is successively refinable because it is simply the binary iid source which emits the symbols 0 and 1 each with probability 1/2, known to be successively refinable under the Hamming distortion metric [3]. For $n = 2$, BSMS$_n$ is also successively refinable however the proof of said fact requires a longer argument. The probability mass function that governs the behavior of BSMS$_2$ along with its associated distortion matrix are

x	$P_X(X = x)$
00	$p/2$
01	$(1-p)/2$
10	$(1-p)/2$
11	$p/2$

$d_{j,k}$	00	01	10	11
00	0	1/2	1/2	1
01	1/2	0	1	1/2
10	1/2	1	0	1/2
11	1	1/2	1/2	0

In this section we shall temporarily use X to refer to a sample of BSMS$_2$ and P_X for the probability law that governs BSMS$_2$. Assume that $p > 1/2$. A complete analytical solution for the optimal output probabilities for every value of $0 \leq D \leq 1/2$ is known [12], pag. 49 (note however that there, it was assumed that $p < 1/2$). Let

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

Assuming $p > 1/2$, the optimal output probabilities for $D \in [0, D_c^{(2)}]$ and $D \in [D_c^{(2)}, 1/2]$ are

$$q = \frac{1}{2(1-\beta)^2}\begin{bmatrix} p(1+\beta)^2 - 2\beta \\ (1-p)(1+\beta)^2 - 2\beta \\ (1-p)(1+\beta)^2 - 2\beta \\ p(1+\beta)^2 - 2\beta \end{bmatrix} \qquad q = \begin{bmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{bmatrix}$$

respectively, where

$$\beta = \frac{D}{1-D}$$

It can be checked that for $D = D_c$, both solutions coincide. The slope s at distortion D satisfies

$$s = 2\ln\beta$$

For a given optimal output probability vector q, the associated optimal test channels are

$$Q_{k|j} = \lambda_j q_k \exp(sd_{j,k})$$

where

$$\lambda_j^{-1} = \sum_k q_k \exp(sd_{j,k})$$

Now let $\beta_i = D_i/(1-D_i)$ for $i = 1, 2$. Depending on the values of D_1 and D_2, three cases are possible: $0 < D_2 < D_1 \leq D_c^{(2)}$, $D_c^{(2)} \leq D_2 < D_1 < 1/2$, and $0 < D_2 < D_c^{(2)} < D_1 < 1/2$. In the first case, BSMS_2 is easily seen to be refinable using the arguments similar to the ones used in the previous section. For the second case, let Y_2 be the random variable defined by $P_{Y_2|X}$, the optimal test channel at distortion D_2:

$$\begin{bmatrix} P(Y_2 = 00|X = x) \\ P(Y_2 = 11|X = x) \end{bmatrix} = \begin{cases} \begin{bmatrix} \beta_2^{2d_{00,x}}/(1+\beta_2^2) \\ \beta_2^{2d_{11,x}}/(1+\beta_2^2) \end{bmatrix} & \text{if} \quad x \in \{00, 11\} \\ \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} & \text{if} \quad x \in \{01, 10\} \end{cases}$$

Also let Y_1 be the random variable obtained by passing Y_2 through a binary symmetric channel with transition probability t:

$$\begin{bmatrix} P(Y_1 = 00|X = x) \\ P(Y_1 = 11|X = x) \end{bmatrix} = \begin{cases} \begin{bmatrix} 1-t & t \\ t & 1-t \end{bmatrix} \begin{bmatrix} \beta_2^{2d_{00,x}}/(1+\beta_2^2) \\ \beta_2^{2d_{11,x}}/(1+\beta_2^2) \end{bmatrix} & \text{if} \quad x \in \{00, 11\} \\ \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} & \text{if} \quad x \in \{01, 10\} \end{cases}$$

Choosing

$$t = \frac{\beta_1^2 - \beta_2^2}{(1+\beta_1^2)(1-\beta_2^2)},$$

it can be checked that for $x \in \{00, 11\}$,

$$\begin{bmatrix} P(Y_1 = 00|X = x) \\ P(Y_1 = 11|X = x) \end{bmatrix} = \begin{bmatrix} \beta_1^{2d_{00,x}}/(1+\beta_1^2) \\ \beta_1^{2d_{11,x}}/(1+\beta_1^2) \end{bmatrix}$$

and therefore Y_1 attains the rate distortion function of BSMS$_2$ at distortion D_1. By definition $X \to Y_2 \to Y_1$ and consequently BSMS$_2$ is refinable from D_1 to D_2. Finally, for the third case it suffices to note that because BSMS$_2$ is refinable from $D_c^{(2)}$ to D_2, and is also refinable from D_1 to $D_c^{(2)}$, conditional probability distributions $Q_{Y_2,Y_c|X}$ and $R_{Y_c,Y_1|X}$ exist such that $I(Q_{Y_2|X}) = R(D_2)$, $I(Q_{Y_c|X}) = R(D_c^{(2)})$, $I(R_{Y_c|X}) = R(D_c^{(2)})$, $I(R_{Y_1|X}) = R(D_1)$, and such that

$$Q_{Y_2,Y_c|X} = Q_{Y_2|Y_c} Q_{Y_c|X} \tag{16}$$

$$R_{Y_c,Y_1|X} = R_{Y_c|Y_1} R_{Y_1|X} \tag{17}$$

Two conditional distributions that attain the rate distortion function of a source at $D_c^{(2)}$ and in general any other distortion level need not be equal to each other; nevertheless, in this case

$$R_{Y_c|X} = Q_{Y_c|X} \tag{18}$$

by construction. Let $Q^*_{Y_2,Y_c,Y_1|X}$ be the conditional probability distribution defined by

$$Q^*_{Y_2,Y_c,Y_1|X} = Q_{Y_2|Y_c} R_{Y_c|Y_1} R_{Y_1|X} \tag{19}$$

Let Y_1^*, Y_2^* be random variables jointly distributed with X according to $P_X Q^*_{Y_2,Y_1|X}$. Note that

$$\begin{aligned} Q^*_{Y_2|X} &\overset{(a)}{=} \sum_{y_c,y_1} Q_{Y_2|Y_c} R_{Y_c|Y_1} R_{Y_1|X} \\ &\overset{(b)}{=} \sum_{y_c} Q_{Y_2|Y_c} \sum_{y_1} R_{Y_c,Y_1|X} \\ &\overset{(c)}{=} \sum_{y_c} Q_{Y_2|Y_c} Q_{Y_c|X} \\ &\overset{(d)}{=} \sum_{y_c} Q_{Y_2,Y_c|X} \\ &= Q_{Y_2|X} \end{aligned} \tag{20}$$

where (a) follows from (19), (b) follows from (17), (c) follows from (18) and (d) follows from (16). Also

$$\begin{aligned} Q^*_{Y_1|X} &= \sum_{y_c,y_2} Q_{Y_2|Y_c} R_{Y_c|Y_1} R_{Y_1|X} \\ &= R_{Y_1|X} \sum_{y_c} R_{Y_c|X} \sum_{y_2} Q_{Y_2|Y_c} \\ &= R_{Y_1|X} \sum_{y_c} R_{Y_c|X} \\ &= R_{Y_1|X} \end{aligned} \tag{21}$$

Equations (20) and (21), combined with the fact $X \to Y_2^* \to Y_1^*$ as it is evident from (19) yield immediately that BSMS$_2$ is refinable from D_1 to D_2.

4.3 Proof of uniqueness of the solution for high distortions

We prove this fact in two steps: Lemma 3 shows that an optimal output marginal must assign zero probability to any sequence $\mathbf{y} \in S^n - \{00\cdots 0, 11\cdots 1\}$. Next, Lemma 4 shows that the assignment $q_{00\cdots 0} = r, q_{11\cdots 1} = 1 - r$ is suboptimal if $r \neq 1/2$. Let $a \in (a_n^*, 1)$ and let Q^* denote the channel generated by q^* at a. Then

Lemma 3 *Let Q be a conditional probability such that $d(Q) = d(Q^*)$ and for some $\mathbf{x} \in S^n, \mathbf{y} \in S^n - \{00\cdots 0, 11\cdots 1\}$, $Q_{\mathbf{y}|\mathbf{x}} > 0$. Then $I(Q) > I(Q^*)$.*

The following Lemma establishes that the output marginal $q_{00\cdots 0} = r, q_{11\cdots 1} = 1 - r$ with $r \neq 1/2$ cannot be optimal because the associated $c_{00\cdots 0}$ and $c_{11\cdots 1}$ cannot be equal (in particular, they cannot both be equal to 1).

Lemma 4 *Let $q_{00\cdots 0} = r, q_{11\cdots 1} = 1 - r$. Assume $a \in (0, 1)$, and $r \neq 1/2$. Then $c_{00\cdots 0} \neq c_{11\cdots 1}$.*

Proof of Lemma 3. Assume that $I(Q) = I(Q^*)$. For any $t \in [0, 1]$ and $\mathbf{x}, \mathbf{y} \in S^n$, define

$$\begin{aligned} Q^t_{\mathbf{y}|\mathbf{x}} &= (1-t)Q^*_{\mathbf{y}|\mathbf{x}} + tQ_{\mathbf{y}|\mathbf{x}} \\ &\triangleq Q^*_{\mathbf{y}|\mathbf{x}} + t\Delta Q_{\mathbf{y}|\mathbf{x}} \end{aligned} \quad (22)$$

Note that the assumption $d(Q) = d(Q^*)$ implies that $d(Q^t) = d(Q^*)$, in other words, for all $t \in [0, 1]$ the perturbation to Q^* defined by (22) leaves the average distortion unchanged. Further note that by the convexity $\cup$ of the mutual information for fixed source probabilities, $I(Q^t) \leq (1-t)I(Q^*) + tI(Q) = I(Q^*)$. By optimality of Q^*, we see that the reverse inequality $I(Q^t) \geq I(Q^*)$ is also true, thus concluding that

$$I(Q^t) - I(Q^*) = 0 \qquad \forall\, t \in [0, 1] \quad (23)$$

We shall obtain a contradiction by showing that for small enough t, $I(Q^t) - I(Q^*) > 0$. The change in mutual information due to the perturbation can be estimated using a Taylor series expansion in t around $t = 0$, thus obtaining the following lower bound: [12] (p. 36)

$$I(Q^t) - I(Q^*) \geq t \left(\sum_{\mathbf{x} \in S^n} \sum_{\mathbf{y} \notin \{00\cdots 0, 11\cdots 1\}} p_{\mathbf{x}} \Delta Q_{\mathbf{y}|\mathbf{x}} (1 - c^*_{\mathbf{y}}) \right) + o(t) \quad (24)$$

as $t \to 0$. Note that each of the terms in the summation is nonnegative. Moreover, by assumption, there exists $\mathbf{x} \in S^n, \mathbf{y} \in S^n - \{00\cdots 0, 11\cdots 1\}$ such that $\Delta Q^t_{\mathbf{y}|\mathbf{x}} > 0$. Also, by Theorem 2, we know that $\forall\, \mathbf{y} \in S^n - \{00\cdots 0, 11\cdots 1\}$, $c^*_{\mathbf{y}} < 1$ (a *strict* inequality). Therefore, the factor multiplying t

in the right hand side of (24) is strictly positive. Finally note that for any sufficiently small t, the first term dominates the second one, thus obtaining the desired contradiction. ♣

Proof of Lemma 4. Consider the case when n is odd. For any $\mathbf{x} \in S^n$, let $d_{\mathbf{x}} \triangleq d_H(\mathbf{x}, 00\cdots 0)$. Define

$$S_0^n = \left\{ \mathbf{x} \in S^n : d_{\mathbf{x}} \leq \frac{n-1}{2} \right\}$$

Note that if $\mathbf{x} \in S_0^n$, its bitwise inverse satisfies $\tilde{\mathbf{x}} \in (S_0^n)^c$. Further note that by symmetry, $p_{\mathbf{x}} = p_{\tilde{\mathbf{x}}}$. Using these facts, it is easy to see that

$$\begin{aligned} c_{00\cdots 0} &= \sum_{\mathbf{x} \in S^n} \frac{p_{\mathbf{x}} a^{d_{\mathbf{x}}}}{r a^{d_{\mathbf{x}}} + (1-r) a^{n-d_{\mathbf{x}}}} \\ &= \sum_{\mathbf{x} \in S_0^n} p_{\mathbf{x}} \left(\frac{a^{d_{\mathbf{x}}}}{r a^{d_{\mathbf{x}}} + (1-r) a^{n-d_{\mathbf{x}}}} + \frac{a^{n-d_{\mathbf{x}}}}{r a^{n-d_{\mathbf{x}}} + (1-r) a^{d_{\mathbf{x}}}} \right) \end{aligned}$$

A similar expression for $c_{11\cdots 1}$ can be obtained:

$$c_{11\cdots 1} = \sum_{\mathbf{x} \in S_0^n} p_{\mathbf{x}} \left(\frac{a^{n-d_{\mathbf{x}}}}{r a^{d_{\mathbf{x}}} + (1-r) a^{n-d_{\mathbf{x}}}} + \frac{a^{d_{\mathbf{x}}}}{r a^{n-d_{\mathbf{x}}} + (1-r) a^{d_{\mathbf{x}}}} \right)$$

The difference $c_{00\cdots 0} - c_{11\cdots 1}$ can thus be expressed as

$$\begin{aligned} &\sum_{\mathbf{x} \in S_0^n} p_{\mathbf{x}} (a^{d_{\mathbf{x}}} - a^{n-d_{\mathbf{x}}}) \left(\frac{1}{r a^{d_{\mathbf{x}}} + (1-r) a^{n-d_{\mathbf{x}}}} - \frac{1}{r a^{n-d_{\mathbf{x}}} + (1-r) a^{d_{\mathbf{x}}}} \right) = \\ &(1-2r) \sum_{\mathbf{x} \in S_0^n} \frac{(a^{d_{\mathbf{x}}} - a^{n-d_{\mathbf{x}}})^2}{(r a^{d_{\mathbf{x}}} + (1-r) a^{n-d_{\mathbf{x}}})(r a^{n-d_{\mathbf{x}}} + (1-r) a^{d_{\mathbf{x}}})} \end{aligned}$$

Note that the summation in the equation above is strictly positive if $a \in (0,1)$. Therefore, if $r \neq 1/2$, $c_{00\cdots 0} - c_{11\cdots 1} \neq 0$. The case when n is even can obviously be handled using an appropriately defined quantizer cell S_0^n. ♣

4.4 Proof of Lemma 1

Lemma 1 (repeated) *If $a_2 < a_1$ then $\phi(a_1, a_2, b)$ is a strictly decreasing function of b in the domain $b \in (0, \infty)$.*

Proof. During the course of this proof, we will use three times the following elementary result from calculus: if $f(z)$ is a differentiable function on the interval (x, y), and if $\frac{df}{dz} > 0$ (resp. $\frac{df}{dz} < 0$) for every $z \in (x, y)$, then f is strictly increasing (resp. decreasing) on the interval. Recall that

$$\phi(a_1, a_2, b) = \frac{a_1^b - a_2^b}{(1 + a_1^b)(1 - a_2^b)}$$

For fixed values of a_1 and a_2, $\phi(a_1, a_2, \cdot)$ is a differentiable function of b in the domain $b \in (0, \infty)$. Therefore, to establish the Lemma it suffices to prove that

$$\frac{\partial \phi}{\partial b} < 0 \qquad \text{if } \; b > 0$$

It can be shown that

$$\frac{\partial \phi}{\partial b} = \left(\frac{b(\log a_1)(\log a_2)}{\left(a_1^{b/2} + a_1^{-b/2}\right)^2 \left(a_2^{b/2} - a_2^{-b/2}\right)^2} \right) \left(\varphi\left(a_1^b\right) - \varphi\left(a_2^b\right) \right) \tag{25}$$

where

$$\varphi(\gamma) = \frac{\gamma - \gamma^{-1}}{\log \gamma}$$

Note that the the first factor in (25) is always positive for $b > 0$, so it suffices to show that if $0 < a_2 < a_1 < 1$ and $b > 0$ then

$$\varphi\left(a_1^b\right) - \varphi\left(a_2^b\right) < 0 \tag{26}$$

This will be accomplished through the aid of the following result:

Lemma 5 *$\varphi(\gamma)$ is strictly decreasing in the domain $\gamma \in (0, 1)$.*

Proof. Clearly, $\varphi(\gamma)$ is a differentiable function of γ in the domain $(0, 1)$, hence it suffices to show that

$$\frac{d\varphi}{d\gamma} < 0 \qquad \text{if } \; \gamma \in (0, 1) \tag{27}$$

It is easily seen that

$$\frac{d\varphi}{d\gamma} = \frac{\eta(\gamma^2)}{(\gamma \log \gamma)^2}$$

where

$$\eta(\xi) = 1 - \xi + \frac{1}{2}(1 + \xi) \log \xi$$

Therefore, to prove (27) it is enough to show that if $\xi \in (0, 1)$, then $\eta(\xi) < 0$. This, in turn, can be shown by noting that

$$\begin{aligned} \frac{d\eta}{d\xi} &= -\frac{1}{2} + \frac{1}{2\xi} + \frac{1}{2} \log \xi \\ &> -\frac{1}{2} + \frac{1}{2\xi} + \frac{1}{2}\left(1 - \frac{1}{\xi}\right) \\ &= 0 \qquad \xi \in (0, 1) \end{aligned} \tag{28}$$

It is worth pointing out that a strict inequality would not hold in the interval $(0,1]$. The differentiability of $\eta(\xi)$ in the domain $\xi \in (0,1)$ and Equation (28) allow us to conclude that $\eta(\xi)$ is a strictly increasing function in the domain $\xi \in (0,1)$. Next let $\xi \in (0,1)$ and let m^* be a positive integer such that $\xi < 1 - 1/m^*$. Then

$$\eta(\xi) < \eta\left(1 - \frac{1}{m^*}\right) \tag{29}$$

and for all $m > m^*$

$$\eta\left(1 - \frac{1}{m^*}\right) < \eta\left(1 - \frac{1}{m}\right)$$

Taking the limit as $m \to \infty$, we obtain

$$\begin{aligned} \eta\left(1 - \frac{1}{m^*}\right) &\leq \lim_{m\to\infty} \eta\left(1 - \frac{1}{m}\right) \\ &\overset{(a)}{=} \eta\left(\lim_{m\to\infty} 1 - \frac{1}{m}\right) \\ &= \eta(1) \\ &= 0 \end{aligned} \tag{30}$$

where (a) follows from the continuity of $\eta(\xi)$ in the domain $\xi \in (0,\infty)$. Combining (30) and (29) gives

$$\eta(\xi) < 0 \qquad \text{if } \xi \in (0,1)$$

This establishes (27) and in turn finishes the proof of Lemma 5. ♣

If $b > 0$ and if $0 < a_2 < a_1 < 1$, then $0 < a_2^b < a_1^b < 1$. Combining this with the fact that $\varphi(\gamma)$ is strictly decreasing in the domain $\gamma \in (0,1)$, we establish Equation (26) which together with Equation (25) implies Lemma 1. ♣

References

[1] V. Koshelev. Hierarchical coding of discrete sources. *Probl. Pered. Inform.*, 16(3):31–49, 1980.

[2] V. Koshelev. An evaluation of the average distortion for discrete scheme of sequential approximation. *Probl. Pered. Inform.*, 17(3):20–33, 1981.

[3] W. H. Equitz, T Cover. Successive refinement of information. *IEEE Trans. Inform. Theory*, 37(2):269–275, March 1991.

[4] R.M. Gray, A. Wyner. Source coding for a simple network. *The Bell System Technical Journal*, 53(9):1681–1721, 1974.

[5] R.M. Gray. Source coding a binary symmetric source for a simple network. In *Proc. of the Sixth Hawaii Int'l. Conf. on System Sciences*, pages 354–355, Jan 1973.

[6] B. Rimoldi. Successive refinement of information: Characterization of the achievable rates. *IEEE Trans. Inform. Theory*, 40(1):253–259, January 1994.

[7] M. Effros. Distortion-rate bounds for fixed- and variable-rate multiresolution source codes. *IEEE Trans. Inform. Theory*, 45(6):1887–1910, September 1999.

[8] J. Chow, T. Berger. Failure of successive refinement for symmetric gaussian mixtures. *IEEE Trans. Inform. Theory*, 43(1):350–352, January 1997.

[9] L. Lastras, T. Berger. All sources are nearly successively refinable. *IEEE Trans. Inform. Theory*, 43(3):918–926, 2001.

[10] C.E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *IRE Conv. Rec.*, pages 142–163, 1959.

[11] T.J. Goblick. A coding theorem for time-discrete analog data sources. *IEEE Trans. Inform. Theory*, 15:401–407, 1969.

[12] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Inc., 1971.

[13] J.C. Kieffer. A survey of the thoery of source coding with a fidelity criterion. *IEEE Trans. Inform. Theory*, 39:1473–1490, 1993.

[14] T. Berger. Explicit bounds to r(d) or a binary symmetric markov source. *IEEE Trans. Inform. Theory*, 23:52–59, 1977.

[15] Ellett, F.E. Jr. A rate distortion problem for the binary markov source. *M.S. thesis, Cornell University School of Electrical Engineering, Ithaca, NY*, 1970.

[16] R. Gray. Information rates of autoregressive processes. *IEEE Trans. Inform. Theory*, 16(4):412–421, 1970.

[17] L. Lastras, T. Berger. Support material for: On successive refinement of the binary symmetric markov source. Technical report, IBM TJ Watson Research Center, January 2004.