Improved Bounds for the Rate Loss of Multiresolution Source Codes

Hanying Feng, *Member, IEEE*, and Michelle Effros, *Senior Member, IEEE*

***Abstract*—In this paper, we present new bounds for the rate loss of multiresolution source codes (MRSCs). Considering an M-resolution code, the rate loss at the ith resolution with distortion D_i is defined as $L_i = R_i - R(D_i)$, where R_i is the rate achievable by the MRSC at stage i. This rate loss describes the performance degradation of the MRSC compared to the best single-resolution code with the same distortion. For two-resolution source codes, there are three scenarios of particular interest: i) when both resolutions are equally important; ii) when the rate loss at the first resolution is 0 ($L_1 = 0$); iii) when the rate loss at the second resolution is 0 ($L_2 = 0$). The work of Lastras and Berger gives constant upper bounds for the rate loss of an arbitrary memoryless source in scenarios i) and ii) and an asymptotic bound for scenario iii) as D_2 approaches 0. In this paper, we focus on the squared error distortion measure and a) prove that for scenario iii) $L_1 < 1.1610$ for all $D_2 < D_1$; b) tighten the Lastras–Berger bound for scenario ii) from $L_2 \leq 1$ to $L_2 < 0.7250$; c) tighten the Lastras–Berger bound for scenario i) from $L_i \leq 1/2$ to $L_i < 0.3802$, $i \in \{1, 2\}$; and d) generalize the bounds for scenarios ii) and iii) to M-resolution codes with $M \geq 2$. We also present upper bounds for the rate losses of additive MRSCs (AMRSCs). An AMRSC is a special MRSC where each resolution describes an incremental reproduction and the kth-resolution reconstruction equals the sum of the first k incremental reproductions. We obtain two bounds on the rate loss of AMRSCs: one primarily good for low-rate coding and another which depends on the source entropy.**

***Index Terms*—Additive successive refinement code, progressive transmission, tree-structured vector quantizer.**

I. Introduction

BECAUSE of their ability to satisfy varying bandwidth, computation, and performance constraints with a single code, multiresolution source codes (MRSCs) are playing an increasingly important role in research and in practice (e.g., [1]–[6]). A key question about MRSCs concerns the performance penalty associated with using MRSCs rather than single-resolution codes (1RSCs). In particular, for any $R_2 > R_1 > 0$ and any $D_1 > D_2 > 0$, we call the vector (R_1, R_2, D_1, D_2) *achievable* by multiresolution coding on source X if there exists an MRSC that uses R_1 bits per symbol (b/s) to describe X with distortion D_1 and then uses an additional $R_2 - R_1$ b/s to refine the description to distortion D_2, as shown in Fig. 1. The *rate loss* of the given two-resolution code (2RSC) is defined as $L_i = R_i - R(D_i)$ $(i = 1, 2)$, where $R(D)$ is the rate-distortion function for source X. Rate loss quantifies the performance degradation associated with using a 2RSC rather than the best 1RSC with the same distortion.

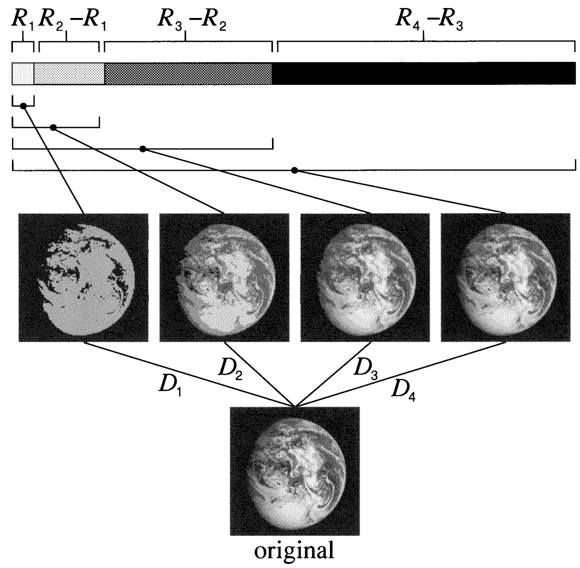


Fig. 1. A 4RSC. Decoding the first R_1 b/s of the binary description yields a reproduction with distortion D_1. Decoding an additional $R_2 - R_1$ b/s, for a total rate of R_2 b/s, yields a reproduction of distortion $D_2 < D_1$, and so on.

A source is called *successively refinable* if an optimal MRSC for any distortions (D_1, D_2) achieves the rate-distortion bound at *both* stages, i.e., $L_i = 0$ for $i = 1, 2$ [7]. Necessary and sufficient conditions for a source to be successively refinable appear in [8]. Examples of sources that are not successively refinable are shown for discrete-alphabet and continuous-alphabet sources in [8], [9] and [10], respectively. The MRSC achievable rate-distortion region for nonsuccessively refinable sources appears in [11] and [12].

In [13], Lastras and Berger consider the question of whether there exists a source in which the MRSC rate loss can be made arbitrarily large. Following an approach developed by Zamir in [14], they demonstrate that for *any* memoryless source, the squared error distortion measure, and any distortions (D_1, D_2), there exists an achievable vector such that $L_i \leq 1/2$, for $i \in \{1, 2\}$. Moreover, they show that an achievable vector can be found with $L_1 = 0$ and $L_2 \leq 1$. They also show that as $D_2 \to 0$, $L_1 \leq 1/2$ with $L_2 = 0$ is achievable.

Manuscript received April 4, 2002; revised November 8, 2002. This work was supported in part by the National Science Foundation under Grant CCR-9909026 and in part by the Caltech's Lee Center for Advanced Networking. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Washington, DC, June, 2001.

H. Feng is with Teradyne, Inc., Agoura Hills, CA 91301 USA (e-mail: fhy@z.caltech.edu).

M. Effros is with the Department of Electrical Engineering (MC 136-93), California Institute of Technology, Pasadena, CA 91125 USA (e-mail: effros@z.caltech.edu).

Communicated by R. Zamir, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2003.809604

In this paper, we first present a nonasymptotic bound for L_1 when $L_2 = 0$; second, we tighten the bound for L_2 with $L_1 = 0$ from $L_2 \leq 1$ to $L_2 < 0.7250$; third, we tighten the bound for $L_1 = L_2$ from $1/2$ to 0.3802; then, we generalize the result for L_2 when $L_1 = 0$ and the result for L_1 when $L_2 = 0$ from two-resolution to M-resolution source codes for any $M \geq 2$.

We also consider a special type of MRSC called an additive MRSC (AMRSC). AMRSCs, also known as additive successive refinement codes, are multiple-description codes used as MRSCs. The kth-resolution reproduction of an AMRSC equals the sum of the independent reconstructions from the multiple description code's first k packets [15]. A two-stage AMRSC (A2RSC) encodes source X using two packets with rates R_1 and $\Delta R = R_2 - R_1$, respectively. The reproduction from packet 1 has expected distortion D_1, and the sum of the reproductions from both packets yields expected distortion $D_2 < D_1$. AMRSCs are of potential interest since their codebook storage requirements are lower than those of other MRSCs and they provide a simple framework for low-complexity (greedy) encoding. Multistage vector quantizers are a practical example of AMRSCs. We obtain two bounds on the rate loss of AMRSCs.

II. Preliminaries

Let $\{X_i\}_{i=1}^{\infty}$ be a real-valued independent and identically distributed (i.i.d.) source with probability density function (pdf) $f_X(x)$. Let d be a real-valued nonnegative difference distortion measure, i.e., $d(x, y) = \rho(x - y)$ for any $x, y \in \boldsymbol{R}$ and some function $\rho: \boldsymbol{R} \to [0, \infty)$. Assume that ρ is continuous and that there exists a reference letter $y^* \in \boldsymbol{R}$ such that $E_X d(X, y^*) < \infty$. For any $x^n, y^n \in \boldsymbol{R}^n$, define

$$d_n(x^n, y^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i).$$

The rate-distortion function for source $\{X_i\}_{i=1}^{\infty}$ and the distortion measure d is

$$R(D) = \inf_{f(y|x):\ \int_{(x,y)} f(y|x) f_X(x)\, d(x,y)\, dx\, dy \leq D} I(X; Y)$$

which characterizes the minimum rate required to describe source X with distortion not exceeding D. In the arguments that follow, we frequently assume that there exists a conditional pdf $f(y|x)$ that achieves $R(D)$. This assumption simplifies the exposition considerably but is not a necessary condition for any of our results.

An (n, M_1, M_2) 2RSC consists of two encoder/decoder pairs: a) a coarse pair $(f_n^{(1)}, g_n^{(1)})$

$$f_n^{(1)}: \boldsymbol{R}^n \to \{1, \ldots, M_1\} \quad \text{and} \quad g_n^{(1)}: \{1, \ldots, M_1\} \to \boldsymbol{R}^n$$

with rate $(1/n) \log M_1$ and distortion

$$Ed_n\left(X^n, g_n^{(1)}\left(f_n^{(1)}(X^n)\right)\right)$$

and b) a refinement pair $(f_n^{(2)}, g_n^{(2)})$

$$f_n^{(2)}: \boldsymbol{R}^n \to \{1, \ldots, M_2\}$$

and

$$g_n^{(2)}: \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \to \boldsymbol{R}^n$$

with total rate $(1/n) \log(M_1 M_2)$ and distortion

$$Ed_n\left(X^n, g_n^{(2)}\left(f_n^{(1)}(X^n), f_n^{(2)}(X^n)\right)\right).$$

We say that the rate-distortion vector (R_1, R_2, D_1, D_2) is 2RSC-achievable if for any $\epsilon > 0$ and for sufficiently large n, there exists an (n, M_1, M_2) 2RSC such that

$$\begin{aligned} \frac{1}{n} \log M_1 &\leq R_1 + \epsilon \\ Ed_n\left(X^n, g_n^{(1)}\left(f_n^{(1)}(X^n)\right)\right) &\leq D_1 + \epsilon \\ \frac{1}{n} \log(M_1 M_2) &\leq R_2 + \epsilon \\ Ed_n\left(X^n, g_n^{(2)}\left(f_n^{(1)}(X^n), f_n^{(2)}(X^n)\right)\right) &\leq D_2 + \epsilon. \end{aligned}$$

The achievable region for 2RSCs, which is defined as the set of all achievable rate-distortion vectors, is described in the following theorem. The result for finite alphabets comes from [11]. A generalization to any Polish alphabet with an escape symbol appears in [12].

Theorem 1 [11, Theorem 1], [12, Corollary 9]: For any i.i.d. source $\{X_i\}_{i=1}^{\infty}$ with pdf $f_X(x)$ and distortion measure d, the vector (R_1, R_2, D_1, D_2) is 2RSC achievable if there exists a conditional probability $Q_{Y_1, Y_2|X}$ such that

$$\begin{aligned} R_1 &\geq I(X; Y_1), & Ed(X, Y_1) &\leq D_1 \\ R_2 &\geq I(X; Y_2, Y_1), & Ed(X, Y_2) &\leq D_2. \end{aligned}$$

The result generalizes to M RSCs with $M > 2$ [11], [12].

The definition of an (n, M_1, M_2) A2RSC is similar except that the refinement decoder is defined as $g_n^{(2)}: \{1, \ldots, M_2\} \to \boldsymbol{R}^n$ and the corresponding distortion is

$$Ed_n\left(X^n, g_n^{(1)}\left(f_n^{(1)}(X^n)\right) + g_n^{(2)}\left(f_n^{(2)}(X^n)\right)\right).$$

The following theorem from [15] describes an achievable region for AMRSCs.[1] This region is not tight. The AMRSC theorem from [15] is for discrete memoryless sources with finite alphabets. The result extends to continuous memoryless sources with an escape symbol [18].

Theorem 2 [15, Theorem 1]: For any i.i.d. source $\{X_i\}_{i=1}^{\infty}$ with pdf $f_X(x)$ and distortion measure d, (R_1, R_2, D_1, D_2) is A2RSC-achievable if there exists a conditional probability $Q_{Y_1, Y_2|X}$ such that

$$\begin{aligned} R_1 &\geq I(X; Y_1), & Ed(X, Y_1) &\leq D_1 \\ \Delta R &\geq I(X; \Delta Y), & Ed(X, Y_2) &\leq D_2 \\ R_2 &\geq I(X; Y_1, \Delta Y) + I(Y_1; \Delta Y) \end{aligned}$$

where $\Delta R = R_2 - R_1$ and $Y_1 + \Delta Y = Y_2$.

In this paper, we assume an arbitrary i.i.d. real-alphabet source with variance $\sigma^2 < \infty$ and focus on the squared error distortion measure, i.e., $d(x, y) = (x - y)^2$. For any MRSC or AMRSC with M resolutions, if rate-distortion vector $(R_1, \ldots, R_M, D_1, \ldots, D_M)$ is achievable for

$$0 < D_M < \cdots < D_1 \leq \sigma^2$$

[1]Note that the notation used here is a little different from that in [15]: $\hat{X}_1$ is replaced by Y_1, $\hat{X}_1 + \hat{X}_2$ is replaced by $Y_2 = Y_1 + \Delta Y$, and incremental rates are replaced by total rates. The proof of this theorem, which appears in [16], is very similar to the one used by El Gamal and Cover for the multiple descriptions problem [17].

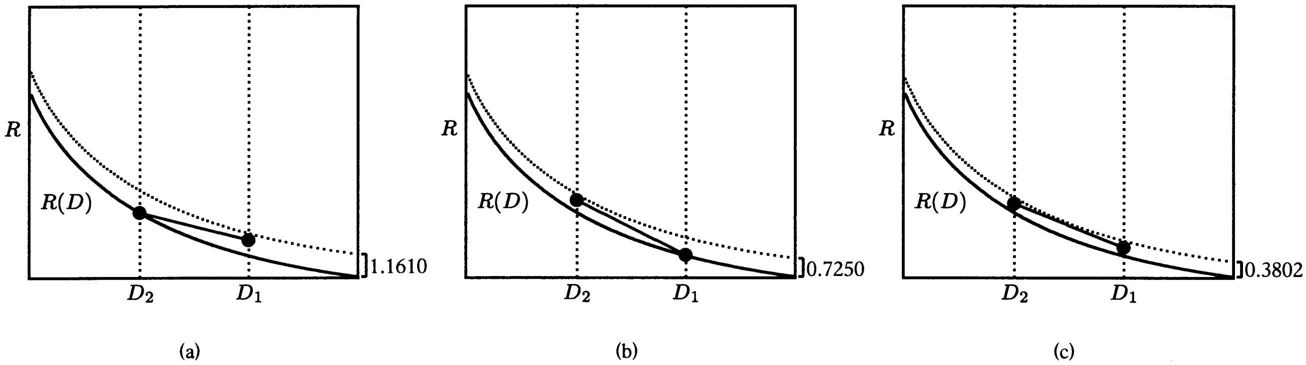


Fig. 2. Graphical interpretation of (a) Theorem 3, (b) Theorem 4, and (c) Theorem 6. In each graph, the lower curve represents $R(D)$, the upper bound represents the sum of $R(D)$ and the rate-loss bound given in the corresponding theorem, and the straight line shows the first- and second-resolution performances of a single code satisfying the corresponding bounds with equality. This approach for illustrating rate-loss bounds is the same as the one used in [13, Fig. 1].

the rate loss at the ith resolution $(i \in \{1, \dots, M\})$ is defined as $L_i = R_i - R(D_i)$ b/s.

III. Techniques

Since Theorem 1 proves the achievability of

$$(R_1, R_2, D_1, D_2) = (I(X; Y_1), I(X; Y_1, Y_2), Ed(X, Y_1), Ed(X, Y_2))$$

for *any* (Y_1, Y_2), proving our main results involves choosing a particular (Y_1, Y_2) that satisfies a desired distortion constraint (D_1, D_2) and then bounding the corresponding rate losses $L_1 = I(X; Y_1) - R(D_1)$ and $L_2 = I(X; Y_1, Y_2) - R(D_2)$. We rely on a few simple, linear constructions (or "test channels") for Y_1 and Y_2. In particular, given some fixed distortions $D_1 > D_2$, we define U_i to be a random variable that achieves the rate-distortion function at distortion D_i. We use four main constructions in building Y_i for source X

$$Y_i = U_i \qquad Y_i = X + N$$
$$Y_i = (1-\alpha)X + \alpha U_{i-1} + N, \quad Y_i = U_{i+1} + N$$

where N is zero-mean Gaussian noise independent of U_{i-1}, U_i, U_{i+1}, and X, α is a constant, and both α and the variance of N are chosen to satisfy the desired rate-distortion constraint. Note that the third and the fourth build a reconstruction for resolution i from the rate distortion achieving reconstruction for resolution $i-1$ or $i+1$. The first two constructions were also used in [13]. Our application of both new constructions and new combinations of old constructions yields the new results. In all cases, bounding the rate loss requires bounding the difference $I(X; Y_1, \dots, Y_i) - I(X; U_i)$. In several cases, the results given involve finding bounds associated with several choices of the vector (Y_1, Y_2) and then combining them. Combinations either apply different bounds for different values of (D_1, D_2) or involve convex combinations of several bounds (since the rate loss turns out to be convex, as shown in Lemma 2).

IV. Results

Throughout this section we assume an i.i.d. source $X_1, X_2, \dots$ and the squared error distortion measure.

We first establish a previously unknown property of rate-distortion functions, which is used in the proof of Theorem 3. All lemmas are proved in the Appendix.

Lemma 1: Suppose $R(D)$ is the rate-distortion function of an arbitrary i.i.d. source $\{X_i\}_{i=1}^{\infty}$ with distortion measure $d(x, y) = (x-y)^2$ and $0 < D_2 < D_1$, then

$$R(D_2) - R(D_1) \le \frac{1}{2} \log \frac{D_1}{D_2}.$$

Theorem 3 bounds the first-resolution rate loss of a 2RSC achieving $R_2 = R(D_2)$. Roughly, the proof involves finding a Gaussian approximation of the optimizing reproduction distribution and bounding the optimal rate loss by the rate loss of the approximation. Use of several Gaussian approximations leads to both increasing and decreasing rate loss bounds for one code, and the intersection of these bounds yields the desired constant bound on the rate loss. Fig. 2(a) shows the graphical interpretation of this theorem. Achieving performance on the rate-distortion curve in resolution 2 requires a rate penalty in resolution 1 that never exceeds $(1/2)\log 5$.

Theorem 3: For any (D_1, D_2) with $D_2 < D_1$, there exists a 2RSC-achievable rate-distortion vector (R_1, R_2, D_1, D_2) with $L_2 = 0$ and $L_1 \le (1/2)\log 5$ b/s.

Proof: We here design a code and derive two bounds on the rate loss. Both bounds are functions of D_2/D_1; one decreases as a function of D_2/D_1 and applies for all $D_2/D_1 < 1$ ($D_2 < D_1$ by assumption), while the other increases as a function of D_2/D_1 and applies when $D_2/D_1 < 1/2$. We combine these functions by choosing the best applicable bound for each value of D_2/D_1. We begin with the increasing bound.

Let U_1 and U_2 be the random variables that achieve $R(D_1)$ and $R(D_2)$, respectively, i.e., $I(X; U_i) = R(D_i)$ and $Ed(X, U_i) \le D_i$ for $i = 1, 2$. Next, let N_1 be a Gaussian random variable with mean 0 and variance $D_1 - D_2$ and N_2 be another Gaussian random variable with mean 0 and variance D_2 (written $N_1 \sim \mathcal{N}(0, D_1 - D_2)$ and $N_2 \sim \mathcal{N}(0, D_2)$). Set (N_1, N_2) to be independent of (X, U_1, U_2) and N_1 to be independent of N_2 (written $(N_1, N_2) \perp\!\!\!\perp (X, U_1, U_2)$ and $N_1 \perp\!\!\!\perp N_2$). Thus, $Ed(X, X + N_1 + N_2) = D_1$ and $Ed(X, U_2 + N_1) \le D_1$. We define

$$Y_1 = U_2 + N_1 \quad \text{and} \quad Y_2 = U_2.$$

Here Y_1 and Y_2 satisfy the distortion constraints, i.e., $Ed(X, Y_2) \le D_2$ and $Ed(X, Y_1) \le D_1$, and, by Theorem 1, the vector $(I(X;Y_1), I(X;Y_2,Y_1), D_1, D_2)$ is achievable. The rate loss at the second stage is

$$\begin{aligned} L_2 &= I(X;Y_2,Y_1) - R(D_2) \\ &= I(X;U_2,U_2+N_1) - I(X;U_2) = 0 \end{aligned}$$

since $X \to U_2 \to U_2+N_1$ forms a Markov chain. The rate loss at the first stage is

$$\begin{aligned} L_1 &= I(X;Y_1) - I(X;U_1) = I(X;U_2+N_1) - I(X;U_1) \\ &= [I(X;U_2+N_1) - I(X;X+N_2+N_1)] \\ &\quad + [I(X;X+N_2+N_1) - I(X;U_1)]. \qquad (1) \end{aligned}$$

We bound the difference between $I(X;U_2+N_1)$ and $I(X;X+N_2+N_1)$ in addition to the difference between $I(X;X+N_2+N_1)$ and $R(D_1) = I(X;U_1)$ to bound the rate loss. Let L_{1A} denote the first difference on the right-hand side of (1); then

$$\begin{aligned} L_{1A} &= I(X;U_2+N_1) - I(X;X+N_2+N_1) \\ &\le I(X;U_2+N_1) - I(X;U_2) + I(X;X+N_2) \\ &\quad - I(X;X+N_2+N_1) \qquad (2) \\ &= I(X;U_2+N_1) - I(X;U_2,U_2+N_1) \\ &\quad + I(X;X+N_2,X+N_2+N_1) \\ &\quad - I(X;X+N_2+N_1) \qquad (3) \\ &= I(X;X+N_2|X+N_2+N_1) - I(X;U_2|U_2+N_1) \\ &= I(N_2+N_1;N_1|X+N_2+N_1) \\ &\quad - I(U_2+N_1-X;N_1|U_2+N_1) \qquad (4) \\ &= I(N_1;N_2+N_1,X+N_2+N_1) \\ &\quad - I(N_1;X+N_2+N_1) \\ &\quad - I(U_2+N_1-X;N_1|U_2+N_1) \\ &= I(N_1;N_2+N_1) - I(N_1;X+N_2+N_1) \\ &\quad - I(U_2+N_1-X;N_1|U_2+N_1) \qquad (5) \\ &= I(N_1;N_2+N_1) - I(N_1;X+N_2+N_1) \\ &\quad + I(N_1;U_2+N_1) - I(N_1;U_2+N_1-X) \\ &\quad - I(N_1;U_2+N_1|U_2+N_1-X) \end{aligned}$$

where (2) follows since $Ed(X, X+N_2) = D_2$ implies that

$$I(X;U_2) = R(D_2) \le I(X;X+N_2)$$

(3) follows since $N_1 \perp\!\!\!\perp (X, U_2, N_2)$ implies

$$X \to U_2 \to U_2+N_1$$

and

$$X \to X+N_2 \to X+N_2+N_1$$

form Markov chains, (4) follows since $h(A|B) = h(A-B|B)$, and (5) follows since $(N_1, N_2) \perp\!\!\!\perp X$ implies

$$N_1 \to N_2+N_1 \to X+N_2+N_1$$

forms a Markov chain.

Let

$$\begin{aligned} J &= I(N_1;U_2+N_1|U_2+N_1-X) \\ &= I(U_2-X;X|U_2+N_1-X) \end{aligned}$$

and

$$K = I(N_1;U_2+N_1) - I(N_1;X+N_2+N_1) - J.$$

Then $L_{1A} \le I(N_1;N_2+N_1) - I(N_1;U_2+N_1-X) + K$, and by the chain rule

$$\begin{aligned} K &= I(N_1;U_2+N_1|X+N_2+N_1) \\ &\quad - I(N_1;X+N_2+N_1|U_2+N_1) - J \\ &\le I(N_1;U_2+N_1|X+N_2+N_1) - J \\ &= I(X+N_2;X-U_2+N_2|X+N_2+N_1) - J \\ &\le I(X+N_2;X-U_2+N_2) - J \qquad (6) \\ &\le I(N_2,X+N_2;X-U_2+N_2) - J \\ &= I(N_2;X-U_2+N_2) \\ &\quad + I(X-U_2+N_2;X+N_2|N_2) - J \\ &= I(N_2;X-U_2+N_2) + I(U_2-X;X|N_2) \\ &\quad - I(U_2-X;X|U_2-X+N_1) \\ &= I(N_2;X-U_2+N_2) + I(U_2-X;X) \\ &\quad - I(U_2-X;X|U_2-X+N_1) \qquad (7) \\ &= I(N_2;X-U_2+N_2) + I(U_2-X;U_2-X+N_1) \\ &\quad - I(U_2-X;U_2-X+N_1|X) \qquad (8) \\ &\le I(N_2;X-U_2+N_2) + I(U_2-X;U_2-X+N_1) \end{aligned}$$

where (6) follows since $N_1 \perp\!\!\!\perp (X, U_2, N_2)$ implies that $X-U_2+N_2 \to X+N_2 \to X+N_2+N_1$ forms a Markov chain, (7) follows since $N_2 \perp\!\!\!\perp (X, U_2)$, and (8) follows by the chain rule since

$$\begin{aligned} &I(U_2-X;X,U_2-X+N_1) \\ &\quad = I(U_2-X;X) + I(U_2-X;U_2-X+N_1|X) \\ &\quad = I(U_2-X;U_2-X+N_1) \\ &\qquad + I(U_2-X;X|U_2-X+N_1). \end{aligned}$$

Thus, if $2D_2 < D_1$ and we define $N_2' \sim \mathcal{N}(0, D_1 - 2D_2)$ and $N_2' \perp\!\!\!\perp (X, U_2, N_1, N_2)$, then

$$\begin{aligned} &I(N_1;U_2+N_1-X) \\ &\quad = I(N_2+N_2';X-U_2+N_2+N_2') \\ &\quad = I(N_2',N_2+N_2';X-U_2+N_2+N_2') \qquad (9) \\ &\quad = I(N_2';X-U_2+N_2+N_2') \\ &\qquad + I(N_2+N_2';X-U_2+N_2+N_2'|N_2') \\ &\quad = I(N_2';X-U_2+N_2+N_2') + I(N_2;X-U_2+N_2) \end{aligned}$$

where (9) follows since $N_2' \to N_2+N_2' \to X-U_2+N_2+N_2'$ forms a Markov chain.

As a result, if $2D_2 < D_1$

$$\begin{aligned} L_{1A} &\le I(N_1;N_1+N_2) + I(U_2-X;U_2-X+N_1) \\ &\quad - I(N_1;U_2+N_1-X) + I(N_2;X-U_2+N_2) \\ &= I(N_1;N_1+N_2) + I(U_2-X;U_2-X+N_1) \\ &\quad - I(N_2';X-U_2+N_2+N_2') \\ &\le \frac{1}{2}\log\frac{D_1}{D_2} + \frac{1}{2}\log\frac{D_1}{D_1-D_2} - \frac{1}{2}\log\frac{D_1}{2D_2} \qquad (10) \\ &= \frac{1}{2}\log\frac{2D_1}{D_1-D_2} \end{aligned}$$

where (10) follows from [19, p. 263, Problem 1] and $E(X-U_2)^2 \le D_2$.

We denote the second difference on the right-hand side of (1) by L_{1B}, which is bounded by $1/2$. The proof parallels that of [13, Theorem 3]. In particular

$$\begin{aligned} L_{1B} &= I(X;\, X+N_2+N_1) - I(X;\, U_1) \\ &= I(X;\, X+N_2+N_1|U_1) - I(X;\, U_1|X+N_2+N_1) \\ &\le I(X;\, X+N_2+N_1|U_1) \\ &= I(X-U_1;\, X-U_1+N_2+N_1|U_1) \\ &\le I(X-U_1;\, X-U_1+N_2+N_1) \\ &\le \frac{1}{2}. \end{aligned}$$

Thus, if $2D_2 < D_1$, we can bound L_1 as

$$\begin{aligned} L_1 &= L_{1A} + L_{1B} \le \frac{1}{2}\log\frac{2D_1}{D_1-D_2} + \frac{1}{2} \\ &= 1 + \frac{1}{2}\log\frac{D_1}{D_1-D_2}. \end{aligned}$$

We derive the second bound on L_1 by noting that for any $D_2 < D_1$, the rate loss at the first stage can also be bounded as

$$\begin{aligned} L_1 &= I(X;\, Y_1) - R(D_1) \le I(X;\, Y_2,\, Y_1) - R(D_1) \\ &= R(D_2) - R(D_1) \le \frac{1}{2}\log\frac{D_1}{D_2} \end{aligned} \tag{11}$$

by Lemma 1 and Theorem 1 since zero rate loss at the second stage implies $I(X;\, Y_2,\, Y_1) = R(D_2)$.

The shaded region of Fig. 3(a) shows the possible values of L_1 as a function of D_2/D_1. The maximal value occurs when $D_1 = 5D_2$, giving $L_1 \le (1/2)\log 5 < 1.1610$. □

By applying the same basic strategy used in the proof of Theorem 3, we improve the bound given by [13, Theorem 5] from $L_1 = 0$ and $L_2 \le 1$ to $L_1 = 0$ and $L_2 \le (1/2)\log(\sqrt{3}+1)$. Fig. 2(b) shows the graphical interpretation of this result.

Theorem 4: For any $(D_1,\, D_2)$ with $D_2 < D_1$, there exists an achievable rate-distortion vector $(R_1,\, R_2,\, D_1,\, D_2)$ with $L_1 = 0$ and $L_2 \le (1/2)\log(\sqrt{3}+1)$.

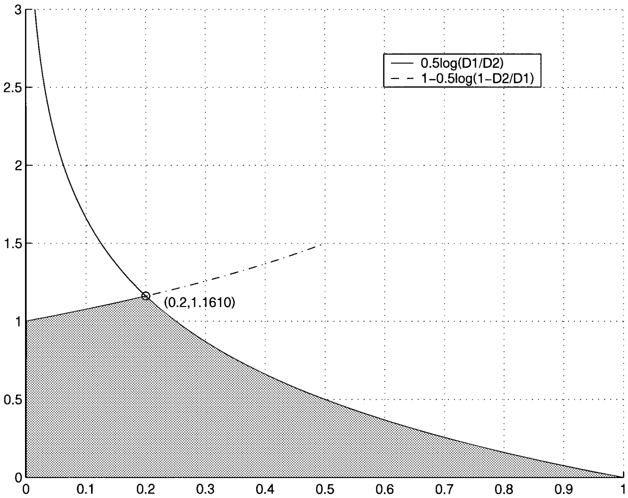
Proof: The proof of [13, Theorem 5] actually shows that

$$L_2 \le \frac{1}{2}\log\left(1+\frac{D_2}{D_1}\right) + \frac{1}{2}$$

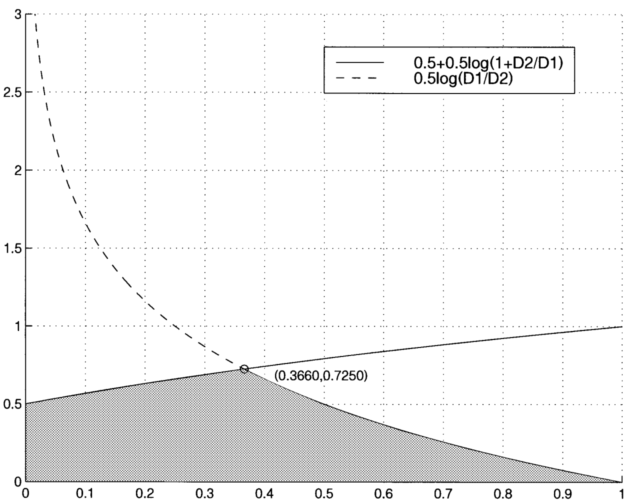
for a code with $Y_1 = U_1$ and $Y_2 = U_2$, where U_1 and U_2 are the random variables that achieve $R(D_1)$ and $R(D_2)$, respectively, and $U_1 \to X \to U_2$.

We here design a different code and derive a bound for the new code that decreases in D_2/D_1. We then put the increasing and decreasing bounds together by choosing, for each value of D_2/D_1, the smaller of the two bounds. Setting U_1 and U_2 to be the random variables that achieve $R(D_1)$ and $R(D_2)$, respectively, $\alpha = D_2/D_1$, and $N_2 \sim \mathcal{N}(0,\, D_2 - \alpha^2 D_1)$, with $N_2 \perp\!\!\!\perp (X,\, U_1)$ as in the proof of Lemma 1 and defining $Y_2 = (1-\alpha)X + \alpha U_1 + N_2$ and $Y_1 = U_1$ gives $L_1 = 0$ and

$$\begin{aligned} L_2 &= I(X;\, Y_2,\, Y_1) - R(D_2) = I(X;\, Y_2,\, U_1) - R(D_2) \\ &= I(X;\, Y_2|U_1) + I(X;\, U_1) - R(D_2) \\ &\le I(X;\, Y_2|U_1) \end{aligned} \tag{12}$$



(a)



(b)

Fig. 3. Possible values of (a) L_1 in Theorem 3 and (b) L_2 in Theorem 4.

$$\le \frac{1}{2}\log\frac{D_1}{D_2}. \tag{13}$$

Here (12) follows since $I(X;\, U_1) = R(D_1)$, $D_2 < D_1$, and the rate-distortion function $R(D)$ is a nonincreasing function of D, and (13) follows from steps (32)–(35) of the proof of Lemma 1.

Thus, for any $D_2/D_1 < 1$, we have

$$L_2 \le \begin{cases} \frac{1}{2}\log\left(1+\frac{D_2}{D_1}\right) + \frac{1}{2} \\ \frac{1}{2}\log\frac{D_1}{D_2}. \end{cases}$$

The first bound dominates when $D_1/D_2 \ge \sqrt{3}+1$; the second dominates when $1 < D_1/D_2 < \sqrt{3}+1$ (see Fig. 3(b)). Together these bounds give $L_2 \le (1/2)\log(\sqrt{3}+1) < 0.7250$. □

We can obtain a looser bound which can be generalized to any difference distortion measure by using $Y_2 = X + N$, where $Ed(N,\, 0) \le D_2$ and $N \perp\!\!\!\perp (X,\, U_1)$; in this case, we can bound L_2 by $(1/2)\log(1+D_1/D_2)$, which leads to the constant bound $(1/2)\log 3 < 0.7925$.

We next refine the shape of this bound using a technique employed in [13, Theorem 6].

Theorem 5: For any (D_1, D_2) with $D_2 < D_1$, there exists an achievable rate-distortion vector (R_1, R_2, D_1, D_2) with $L_1 = 0$ and

$$L_2 \le \frac{1}{2}\log\left[\left(\sqrt{\frac{6\sigma^2-5D_1}{2\sigma^2-D_1}}+1\right)\left(1-\frac{D_1}{2\sigma^2}\right)\right].$$

Proof: From Theorem 4, $L_2 \le (1/2)\log(D_1/D_2)$. For any fixed $0 < D_1 \le \sigma^2$, this bound is a decreasing function of $D_2 \in (0, D_1)$. We next design another code and derive its corresponding bound, which is an increasing function of D_2. Let $U_1 \to X \to U_2$, $\beta_1 = 1 - D_1/\sigma^2$, $N_1' \sim \mathcal{N}(0, \beta_1 D_1)$, and $N_1' \perp\!\!\!\perp (X, U_1, U_2)$. Then, setting $Y_2 = U_2$ and $Y_1 = U_1$ and using an approach from the proof of [13, Theorem 5] gives

$$\begin{aligned} L_2 &= I(X; U_1|U_2) \le I(X; U_1, \beta_1 X + N_1'|U_2)\\ &= I(X; \beta_1 X + N_1'|U_2) + I(X; U_1|U_2, \beta_1 X + N_1')\\ &\le I(X-U_2; \beta_1(X-U_2)+N_1') + I(X; \beta_1 X + N_1'|U_1)\\ &\le \frac{1}{2}\log\left(1+\frac{D_2}{D_1}-\frac{D_2}{\sigma^2}\right)+\frac{1}{2}\log\left(2-\frac{D_1}{\sigma^2}\right). \end{aligned}$$

The second bound dominates when $D_2 \in (0, \Gamma]$, where

$$\Gamma = \frac{D_1\sigma^2\sqrt{6\sigma^2-5D_1} - D_1\sigma^2\sqrt{2\sigma^2-D_1}}{2(\sigma^2-D_1)\sqrt{2\sigma^2-D_1}}$$

while the first dominates for the remainder of the $(0, D_1)$ region. The maximal value of the combined bound is achieved at $D_2 = \Gamma$, giving the desired result. □

The bound described in Theorem 5 is tight when $D_1 = \sigma^2$, where $L_2 = 0$. This bound can also be written as

$$L_2 \le \frac{1}{2}\log\left[\left(\sqrt{5-\frac{4\sigma^2}{2\sigma^2-D_1}}+1\right)\left(1-\frac{D_1}{2\sigma^2}\right)\right],$$

which is a decreasing function of D_1 for a fixed value of σ^2. This bound achieves its maximum when $D_1 = 0$, giving $L_2 \le (1/2)\log(\sqrt{3}+1) < 0.7250$, which is consistent with Theorem 4. The bound from Theorem 5 is less than the bound from Theorem 4 when $D_1 \ne 0$. Fig. 4 shows the bound from Theorem 5. The shaded region shows the possible values of L_2 as a function of D_1/σ^2.

Lemma 2 shows the convexity of the rate loss. This result proves useful in Theorem 6, where we address the case where the rate losses at both resolutions are equal.

Lemma 2: For any $D_2 < D_1$, if rate-distortion vectors $(R_{10}, R_{20}, D_1, D_2)$ and $(R_{11}, R_{21}, D_1, D_2)$ are 2RSC-achievable on source X, then for any $0 \le \alpha \le 1$, there exists an achievable rate-distortion vector $(R_{1\alpha}, R_{2\alpha}, D_1, D_2)$ for the same source with $L_{1\alpha} = \alpha L_{11} + (1-\alpha)L_{10}$ and $L_{2\alpha} = \alpha L_{21} + (1-\alpha)L_{20}$. Here $L_{1\beta} = R_{1\beta} - R(D_1)$ and $L_{2\beta} = R_{2\beta} - R(D_2)$, for all $\beta \in \{0, \alpha, 1\}$.

Using the convexity of the rate loss proved in Lemma 2 with the bounds on L_1 and L_2 from Theorems 3 and 4 gives new bounds for the case where $L_1 = L_2$. In particular, since for

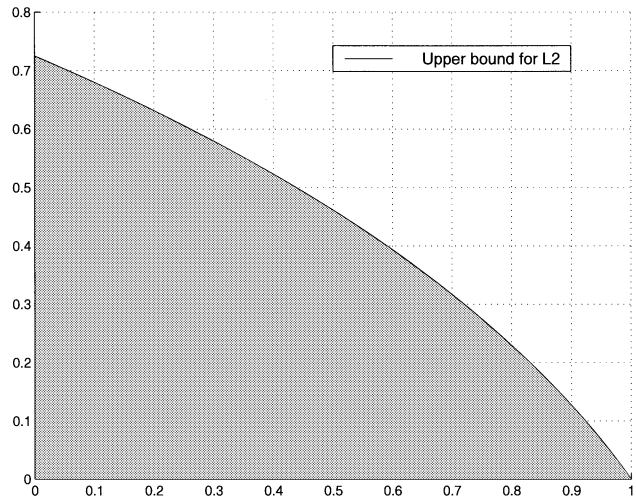


Fig. 4. Possible values of L_2 in Theorem 5.

any $D_2 < D_1$, the rate losses $L_{10} \le (1/2)\log 5$ and $L_{20} = 0$ and the rate losses $L_{11} = 0$ and $L_{21} \le (1/2)\log(\sqrt{3}+1)$ are both achievable by multiresolution coding with distortions (D_1, D_2), the rate losses $L_{1\alpha} = (1-\alpha)L_{10}$ and $L_{2\alpha} = \alpha L_{21}$ are also achievable at these distortions. Setting

$$\alpha = \frac{L_{10}}{L_{10}+L_{21}}$$

proves the achievability of

$$\begin{aligned} L_{1\alpha} &= L_{2\alpha} = \frac{L_{10}L_{21}}{L_{10}+L_{21}} = \frac{1}{1/L_{10}+1/L_{21}}\\ &\le \frac{1}{1/\left(\frac{1}{2}\log 5\right)+1/\left(\frac{1}{2}\log\left(\sqrt{3}+1\right)\right)} < 0.4463 \end{aligned}$$

with distortions (D_1, D_2). This result slightly tightens the bound of [13, Theorem 3], which proves the achievability of $L_1 \le 1/2$ and $L_2 \le 1/2$ for distortions (D_1, D_2).

Theorem 6 improves the bound further. The graphical interpretation is shown in Fig. 2(c).

Theorem 6: For any $D_2 < D_1$, the rate losses $L_1 = L_2 < 0.3802$ are achievable.

Proof: As shown in Theorem 3, the rate losses L_{10} and L_{20} are achievable, where

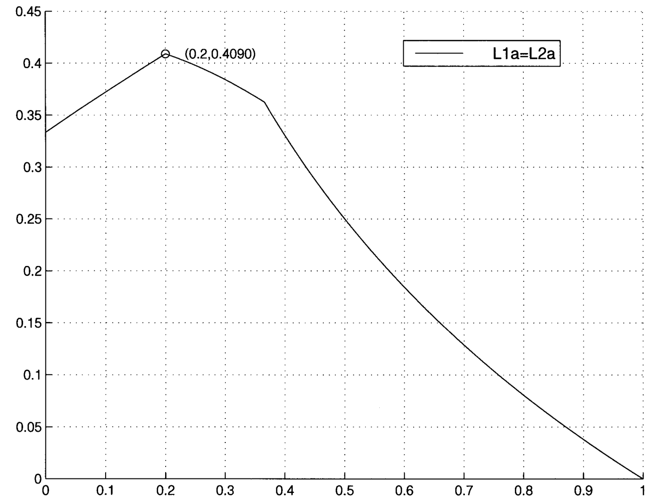
$$L_{20} = 0$$

and

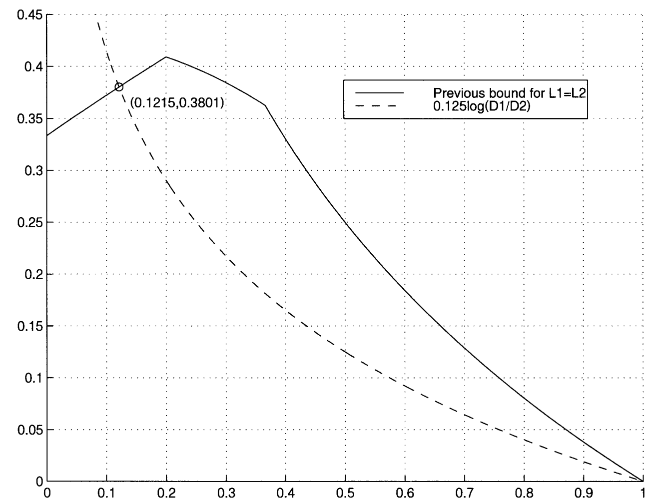
$$L_{10} \le \begin{cases} A_1 = \frac{1}{2}\log\frac{4D_1}{D_1-D_2}, & \text{if } D_1 \ge 5D_2\\ A_2 = \frac{1}{2}\log\frac{D_1}{D_2}, & \text{if } D_1 < 5D_2. \end{cases}$$

From Theorem 4, the rate losses L_{11} and L_{21} are achievable, where

$$\begin{aligned} L_{11} &= 0\\ L_{21} &\le \begin{cases} B_1 = \frac{1}{2}\log\left(2+\frac{2D_2}{D_1}\right), & \text{if } D_1 \ge \left(\sqrt{3}+1\right)D_2\\ B_2 = \frac{1}{2}\log\frac{D_1}{D_2}, & \text{if } D_1 < \left(\sqrt{3}+1\right)D_2. \end{cases} \end{aligned}$$



(a)



(b)

Fig. 5. Upper bounds for $L_{1\alpha} = L_{2\alpha}$ in Theorem 6.

Thus, from Lemma 2, we can draw the conclusion that the rate losses $L_{1\alpha} = (1-\alpha)L_{10}$ and $L_{2\alpha} = \alpha L_{21}$ are achievable. Setting $\alpha = L_{10}/(L_{10}+L_{21})$ gives

$$L_{1\alpha} = L_{2\alpha} \leq \begin{cases} \frac{1}{1/A_1+1/B_1}, & \text{if } \sigma^2 \geq D_1 \geq 5D_2 \\ \frac{1}{1/A_2+1/B_1}, & \text{if } \left(\sqrt{3}+1\right) D_2 \leq D_1 < 5D_2 \\ \frac{1}{1/A_2+1/B_2}, & \text{if } D_2 < D_1 < \left(\sqrt{3}+1\right) D_2. \end{cases}$$

Fig. 5(a) shows the bound for $L_{1\alpha} = L_{2\alpha}$. The maximum occurs when $D_2/D_1 = 0.2$, giving

$$L_{1\alpha} = L_{2\alpha} \leq \frac{1}{1/\left(\frac{1}{2}\log 5\right) + 1/\left(\frac{1}{2}\log 2.4\right)} < 0.4091.$$

Finally, we use the results of Theorems 3 and 4 in a different way to get the desired result. From (11), $L_{10} \leq R(D_2) - R(D_1)$ and $L_{20} = 0$ are achievable. By (12) and (13), $L_{11} = 0$ and

$$L_{21} \leq (1/2)\log(D_1/D_2) + R(D_1) - R(D_2)$$

are achievable. Thus, by convexity, we can achieve

$$\begin{aligned} L_{1\alpha} = L_{2\alpha} &= \frac{L_{10}L_{21}}{L_{10}+L_{21}} \\ &\leq \frac{[R(D_2)-R(D_1)]\left[\frac{1}{2}\log\frac{D_1}{D_2} + R(D_1) - R(D_2)\right]}{\frac{1}{2}\log\frac{D_1}{D_2}} \\ &\leq \frac{\left(R(D_2)-R(D_1)+\frac{1}{2}\log\frac{D_1}{D_2}+R(D_1)-R(D_2)\right)^2}{4\left(\frac{1}{2}\log\frac{D_1}{D_2}\right)} \\ &= \frac{1}{8}\log\frac{D_1}{D_2} \end{aligned} \tag{14}$$

where (14) follows since $4ab \leq (a+b)^2$ for $a, b \geq 0$. Fig. 5(b) combines this bound with the bound of Fig. 5(a). The new maximum is achieved when $D_2/D_1 \approx 0.1215$ and gives $L_{1\alpha} = L_{2\alpha} < 0.3802$, thus, $L_{1\alpha} = L_{2\alpha} < 0.3802$ is achievable. □

For any $M \geq 2$ and $0 < D_M < \cdots < D_2 < D_1$, [13, Corollary 1] shows that there exists an achievable rate-distortion vector $(R_1, \ldots R_M, D_1, \ldots, D_M)$ with $L_i \leq 1/2$, $i \in \{1, \ldots, M\}$. This solution suggests approximately identical priorities at all resolutions. We next consider the case where we minimize the rate loss at the first resolution, then minimize the rate loss at the second resolution subject to the first rate loss and so on. This greedy approach, used in the design of tree-structured vector quantizers (TSVQs) [20], apparently maximizes the rate loss at the last resolution. The next theorem provides an upper bound for this scenario. This result can also be regarded as a generalization of Theorem 4. We first introduce Lemma 3 and Lemma 4, which are useful for proving this theorem.

Lemma 3: Let $A, B, C,$ and D be random variables such that $A \to B \to C$, $A \to C \to D$, and $A \to (B, C) \to D$ form Markov chains. Then $I(A; B|D) \geq I(A; B|C)$.

Lemma 4: For any $D_M < \cdots < D_3 < D_2 < D_1$, let U_1 be a random variable achieving $R(D_1)$ and U_i' for $i \in \{2, 3, \ldots, M\}$ be successively defined as

$$U_i' = \arg \min_{U:\, Ed(X,U) \leq D_i} I(X; U_1, U_2', \ldots, U_{i-1}', U). \tag{15}$$

Then

$$I(X; U_1, U_2', \ldots, U_M' | X + N_M) \leq M/2,$$

where $N_M \sim \mathcal{N}(0, D_M)$ and $N_M \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_M')$.

Theorem 7: For any $D_M < \cdots < D_2 < D_1$, let U_1 be a random variable achieving $R(D_1)$, and define $U_2', \ldots, U_{M-1}'$ as in (15), and $L_i = I(X; U_1, U_2', \ldots, U_i') - R(D_i)$ for all $1 < i \leq M$. Then the rate losses for a greedily designed MRSC with distortions $(D_1, \ldots, D_M)$ are bounded as $L_1 = 0$ and $L_i \leq i/2$ for all $1 < i \leq M$.

Proof: For each $i \in \{2, \ldots, M\}$, let U_i be the random variable that achieves $R(D_i)$, and define $N_{i-1} \sim \mathcal{N}(0, D_{i-1})$, where

$$N_{i-1} \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_{i-1}', U_i).$$

Further let

$$(U_1, U_2', \ldots, U_{i-1}') \to X \to U_i$$

i.e., define the joint distribution of $(X, U_1, U_2', \ldots, U_{i-1}', U_i)$ as $f_X Q_{U_1, U_2', \ldots, U_{i-1}'|X} Q_{U_i|X}$.

The greedy approach uses $Y_1 = U_1$ and $Y_i = U_i'$ for all $i \in \{2, 3, \ldots, i-1\}$. The rate loss of this code at resolution i is

$$\begin{aligned} L_i &= I(X; Y_i, \ldots, Y_2, Y_1) - I(X; U_i) \\ &= I(X; U_i', U_{i-1}', \ldots, U_2', U_1) - I(X; U_i) \\ &\le I(X; U_i, U_{i-1}', \ldots, U_2', U_1) - I(X; U_i) \qquad (16) \\ &= I(X; U_{i-1}', \ldots, U_2', U_1|U_i) \\ &\le I(X; U_{i-1}', \ldots, U_2', U_1, X + N_{i-1}|U_i) \\ &= I(X; X + N_{i-1}|U_i) \\ &\quad + I(X; U_{i-1}', \ldots, U_2', U_1|U_i, X + N_{i-1}) \qquad (17) \end{aligned}$$

where (16) follows from the definition of U_i'. We bound the first term in (17) as

$$I(X; X + N_{i-1}|U_i) \le \frac{1}{2}\log\left(1 + \frac{D_i}{D_{i-1}}\right) \le \frac{1}{2} \qquad (18)$$

where the first inequality follows from the approach of steps (32)–(35) from the Appendix, and the second inequality follows since $D_i < D_{i-1}$. We bound the second term in (17) as

$$\begin{aligned} & I(X; U_{i-1}', \ldots, U_2', U_1|U_i, X + N_{i-1}) \\ &\quad \le I(X; U_{i-1}', \ldots, U_2', U_1|U_i, X + N_{i-1}) \\ &\qquad + I(U_i; U_{i-1}', \ldots, U_2', U_1|X + N_{i-1}) \\ &\quad = I(X, U_i; U_{i-1}', \ldots, U_2', U_1|X + N_{i-1}) \\ &\quad = I(X; U_{i-1}', \ldots, U_2', U_1|X + N_{i-1}) \\ &\qquad + I(U_i; U_{i-1}', \ldots, U_2', U_1|X, X + N_{i-1}) \\ &\quad = I(X; U_{i-1}', \ldots, U_2', U_1|X + N_{i-1}) \qquad (19) \\ &\quad \le \frac{i-1}{2} \qquad (20) \end{aligned}$$

where (19) follows since U_i and $(U_1, U_2', \ldots, U_{i-1}')$ are independent given $(X, X + N_{i-1})$, and (20) follows from Lemma 4. Combining (17), (18), and (20) proves the theorem. □

Theorem 7 gives a collection of rate loss bounds that increase with the increasing resolution. This is consistent with our intuition that performance degrades at higher resolutions for greedily designed codes and suggests that the performance penalty associated with using greedily grown TSVQs [20] rather than jointly optimized multiresolution vector quantizers [5], [6] may be large. Proving such a result would require a tight bound on the rate losses studied in Theorem 7. We next show how to obtain a tighter bound on the rate losses in all stages of this greedily designed code by refining some of the previous arguments. We define $U_i'' = (1-\alpha)X + \alpha U_{i-1}' + N_i'$, where $\alpha = D_i/D_{i-1}$, $N_i' \sim \mathcal{N}(0, D_i - \alpha^2 D_{i-1})$, and $N_i' \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_i')$. Since U_i'' and $(U_{i-2}', \ldots, U_2', U_1)$ are independent given (X, U_{i-1}')

$$\begin{aligned} L_i &= I(X; U_i', U_{i-1}', \ldots, U_2', U_1) - R(D_i) \\ &\le I(X; U_i'', U_{i-1}', \ldots, U_2', U_1) - R(D_i) \\ &= I(X; U_i''|U_{i-1}', \ldots, U_2', U_1) + L_{i-1} \\ &\quad + [R(D_{i-1}) - R(D_i)] \\ &\le I(X; U_i''|U_{i-1}', \ldots, U_2', U_1) + L_{i-1} \\ &\le I(X - U_{i-1}'; (1-\alpha)(X - U_{i-1}') \\ &\quad + N_i'|U_{i-1}', \ldots, U_2', U_1) + L_{i-1} \\ &\le \frac{1}{2}\log\frac{D_{i-1}}{D_i} + L_{i-1} \qquad (21) \\ &\le \frac{1}{2}\log\frac{D_{i-1}}{D_i} + \frac{i-1}{2} \end{aligned}$$

where (21) follows from the approach of steps (32)–(35) from the Appendix. Note that (18) also implies that

$$L_i \le (1/2)\log(1 + D_i/D_{i-1}) + (i-1)/2.$$

Together, these bounds give

$$L_i \le (i-1)/2 + (1/2)\log((\sqrt{5}+1)/2) < i/2 - 0.1528.$$

We also find from (21) that $L_i \le (1/2)\log(D_1/D_i)$. Unfortunately, it may be difficult to extend these new bounds to distortion measures other than squared error.

We can also generalize Theorem 7 in a different way. The following theorem suggests that the penalty of the greedy approach is in some sense constrained to the resolutions in which it is applied.

Theorem 8: For any $D_M < \cdots < D_2 < D_1$ and any $K \in \{2, \ldots, M\}$, let U_1 be a random variable achieving $R(D_1)$ and define $U_2', \ldots, U_{K-1}'$ as in (15). Set $Y_1 = U_1$ and $Y_i = U_i'$ for all $1 < i < K$, but constrain Y_i for $K \le i \le M$ only to guarantee distortion D_i. Then we can achieve rate losses bounded as $L_1 = 0$, $L_i \le i/2$ for all $1 < i < K$, and $L_i \le K/2$ for all $K \le i \le M$.

Proof: Let U_i be a random variable that achieves $R(D_i)$ for any $1 \le i \le M$, $N_i \sim \mathcal{N}(0, D_i - D_{i+1})$ for $K - 1 \le i < M$, $N_M \sim \mathcal{N}(0, D_M)$, $N_i \perp\!\!\!\perp N_j$ for any $j \ne i$, and $N_i \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_{K-1}')$. Further, define $Y_1 = U_1$, $Y_i = U_i'$ for any $1 < i < K$ and $Y_i = X + \sum_{j=i}^M N_j$ for $K \le i \le M$.

Then $L_1 = 0$, $L_i \le i/2$ for all $1 < i < K$ by Theorem 7, and for any $K \le i \le M$

$$\begin{aligned} L_i &= I(X; Y_1, \ldots, Y_i) - R(D_i) \\ &= I\left(X; U_1, U_2', \ldots, U_{K-1}', X + \sum_{j=K}^M N_j, \ldots, X + \sum_{j=i}^M N_j\right) - I(X; U_i) \\ &= I\left(X; U_1, U_2', \ldots, U_{K-1}', X + \sum_{j=i}^M N_j\right) - I(X; U_i) \end{aligned} \qquad (22)$$

$$= I\left(X; U_1, U_2', \ldots, U_{K-1}' | X + \sum_{j=i}^{M} N_j\right) + I\left(X; X + \sum_{j=i}^{M} N_j\right) - I(X; U_i)$$

$$\leq I\left(X; U_1, U_2', \ldots, U_{K-1}' | X + \sum_{j=i}^{M} N_j\right) + \frac{1}{2} \tag{23}$$

$$\leq I\left(X; U_1, U_2', \ldots, U_{K-1}' | X + \sum_{j=K-1}^{M} N_j\right) + \frac{1}{2} \tag{24}$$

$$\leq \frac{K}{2} \tag{25}$$

where (22) follows since

$$X \to X + \sum_{j=i}^{M} N_j \to \left(X + \sum_{j=K}^{M} N_j, \ldots, X + \sum_{j=i+1}^{M} N_j\right)$$

forms a Markov chain, (23) follows since

$$\sum_{j=i}^{M} N_j \sim \mathcal{N}(0, D_i)$$

(24) follows from Lemma 3, and (25) follows from Lemma 4. □

We next obtain a bound for the scenario where we first set L_M to 0, then minimize L_{M-1} subject to L_M, and so on. This result, which can be viewed as a generalization of Theorem 3, mirrors the approach of code designs like [21].

Theorem 9: For any $D_M < \cdots < D_2 < D_1$ and any $K \in \{1, \ldots, M-1\}$, let U_M be a random variable achieving $R(D_M)$ and for all $i \in \{M-1, M-2, \ldots, K\}$ let V_i' be sequentially defined as

$$V_i' = \arg \min_{V:\, Ed(X,V) \leq D_i,\, X \to U_M \to V_{M-1}' \cdots V_{i+1}' \to V} I(X; V).$$

Set $Y_i = V_i'$ for all $K < i < M$ and constrain Y_i for $1 \leq i \leq K$ only to guarantee distortion D_i. Then the rate losses

$$L_i \leq ((M-i)/2) \log 5, \qquad \text{for all } K < i < M$$
$$L_i \leq ((M-K)/2) \log 5, \qquad \text{for all } 1 \leq i \leq K$$

are achievable.

Proof: For any $1 \leq i \leq M$, let U_i be a random variable that achieves $R(D_i)$ and $N_i \sim \mathcal{N}(0, D_i - D_{i+1})$, where $N_i \perp\!\!\!\perp (X, U_M, V_{M-1}', \ldots, V_{K+1}')$ and $N_i \perp\!\!\!\perp N_j$ for any $j \neq i$. Further, define $Y_M = U_M$, $Y_i = V_i'$ for any $K < i < M$, and $Y_i = V_{K+1}' + \sum_{j=i}^{K} N_j$ for $1 \leq i \leq K$. Since $X \to Y_M \to \cdots \to Y_1$, $R_i = I(X; Y_1, \ldots, Y_i) = I(X; Y_i)$ for all $1 \leq i \leq M$. Thus, $L_M = 0$, $L_i = I(X; V_i') - R(D_i)$ for all $M > i > K$, and $L_i = I(X; V_{K+1}' + \sum_{j=i}^{K} N_j) - R(D_i)$ for any $1 \leq i \leq K$.

For all $K + 1 < i < M$

$$L_{i-1} = I(X; V_{i-1}') - R(D_{i-1}) \leq I(X; V_i' + N_{i-1}) - R(D_{i-1}).$$

Let $N_i' \sim \mathcal{N}(0, D_i)$ be independent of all other random variables; then from the proof of Theorem 3

$$\begin{aligned}
L_{i-1} &\leq I(X; V_i' + N_{i-1}) - I(X; U_{i-1}) \\
&= [I(X; V_i' + N_{i-1}) - I(X; V_i')] \\
&\quad + [I(X; X + N_i') - I(X; X + N_i' + N_{i-1})] \\
&\quad + [I(X; X + N_i' + N_{i-1}) - I(X; U_{i-1})] \\
&\quad + [I(X; V_i') - I(X; X + N_i')] \\
&\leq 1 + \frac{1}{2} \log \frac{D_{i-1}}{D_{i-1} - D_i} + I(X; V_i') - I(X; X + N_i') \\
&\leq 1 + \frac{1}{2} \log \frac{D_{i-1}}{D_{i-1} - D_i} + I(X; V_i') - R(D_i) \\
&= 1 + \frac{1}{2} \log \frac{D_{i-1}}{D_{i-1} - D_i} + L_i.
\end{aligned}$$

On the other hand, since $X \to V_i' \to V_{i-1}'$

$$R_{i-1} = I(X; V_{i-1}') \leq I(X; V_i').$$

Therefore,

$$\begin{aligned}
L_{i-1} &\leq I(X; V_i') - R(D_{i-1}) \\
&= I(X; V_i') - R(D_i) + R(D_i) - R(D_{i-1}) \\
&\leq L_i + \frac{1}{2} \log \frac{D_{i-1}}{D_i}.
\end{aligned}$$

Together, these bounds imply that $L_{i-1} \leq L_i + (1/2) \log 5$, giving $L_i \leq ((M-i)/2) \log 5$ for all $K < i < M$ by induction. Similarly, $L_i \leq ((M-K)/2) \log 5$ for all $1 \leq i \leq K$. □

Finally, we turn our attention briefly to A2RSCs. Before bounding the rate losses, we first demonstrate another property of the rate-distortion function.

Lemma 5: Suppose $R(D)$ is a rate-distortion function and $D_0 \leq \min\{D_1, D_2\}$, then

$$R(D_1) + R(D_2) \leq R(D_0), \qquad \text{if } D_1 + D_2 - D_0 \geq \sigma^2.$$

Theorem 10: For any $D_2 < D_1$, there exists an A2RSC-achievable vector (R_1, R_2, D_1, D_2) with $L_1 \leq 1/2$ and $L_2 \leq (1/2) \log \Lambda$, where

$$\Lambda = \min\left\{1 + \frac{D_1}{D_2} - \frac{D_1}{\sigma^2}, \frac{\sigma^2(\sigma^2 - D_1 + D_2) + (D_1 - D_2)D_2}{\sigma^2 D_2}, \frac{D_1}{D_2}\left(2 - \frac{D_1}{\sigma^2}\right)\left(1 + \frac{D_1 - D_2}{\sigma^2}\right)\left(1 - \frac{D_1 - D_2}{\sigma^2}\right)\right\}.$$

Proof: Assume without loss of generality that $E(X) = 0$. Let U_1, U_2, and U_0' be the random variables that achieve $R(D_1)$, $R(D_2)$, and $R(\sigma^2 - D_1 + D_2)$, respectively. Set $\beta_i = 1 - D_i/\sigma^2$, $i \in \{1, 2\}$

$$N_1 \sim \mathcal{N}(0, \beta_1 D_1 - \beta_1^2 D_2 / \beta_2)$$
$$N_2' \sim \mathcal{N}(0, \beta_2 D_2)$$

and $N_1 \perp\!\!\!\perp N_2'$. Further, let

$$(N_1, N_2') \perp\!\!\!\perp (X, U_0, U_1, U_2).$$

We define $Y_2 = \beta_2 X + N_2'$ and

$$Y_1 = \beta_1 Y_2 / \beta_2 + N_1 = \beta_1 X + \beta_1 N_2' / \beta_2 + N_1.$$

Then, by Theorem 2

$$(I(X;Y_1), \max\{I(X;Y_1)+I(X;\Delta Y), I(X;Y_1,\Delta Y)+I(Y_1;\Delta Y)\}, D_1, D_2)$$

is achievable, where

$$\Delta Y = Y_2 - Y_1 = (\beta_2-\beta_1)X + (\beta_2-\beta_1)N_2'/\beta_2 - N_1.$$

Since

$$\begin{aligned} I(X;Y_1,\Delta Y) &+ I(Y_1;\Delta Y) - I(X;Y_1) \\ &= I(X;\Delta Y|Y_1) + I(Y_1;\Delta Y) \\ &= I(\Delta Y; X, Y_1) \geq I(X;\Delta Y) \end{aligned}$$

then

$$(I(X;Y_1), I(X;Y_1,\Delta Y)+I(Y_1;\Delta Y), D_1, D_2)$$

is achievable. Using the argument from steps (31)–(35) in the Appendix and the fact that $Ed(X,\Delta Y) = \sigma^2 - D_1 + D_2$

$$\begin{aligned} L_1 &= I(X;Y_1) - I(X;U_1) \\ &\leq \frac{1}{2}\log\left(2-\frac{D_1}{\sigma^2}\right) \leq \frac{1}{2}, \end{aligned} \tag{26}$$

$$\begin{aligned} I(X;\Delta Y) - I(X;U_0') &\leq \frac{1}{2}\log\left(2-\frac{\sigma^2-D_1+D_2}{\sigma^2}\right) \\ &= \frac{1}{2}\log\left(1+\frac{D_1-D_2}{\sigma^2}\right). \end{aligned} \tag{27}$$

The rate loss at the second stage is

$$\begin{aligned} L_2 &= I(X;Y_1,\Delta Y) + I(Y_1;\Delta Y) - R(D_2) \\ &= I(X;Y_1) + I(X;\Delta Y) + I(Y_1;\Delta Y|X) - I(X;U_2) \end{aligned} \tag{28}$$

where

$$\begin{aligned} I(Y_1;\Delta Y|X) &= I\left(\frac{\beta_1}{\beta_2}N_2' + N_1; \frac{\beta_2-\beta_1}{\beta_2}N_2' - N_1\right) \\ &= \frac{1}{2}\log\frac{D_1}{D_2}\left(1-\frac{D_1-D_2}{\sigma^2}\right). \end{aligned}$$

Then L_2 can be bounded as

$$\begin{aligned} L_2 &= [I(X;Y_1) - I(X;U_2)] + I(Y_1;\Delta Y|X) + I(X;\Delta Y) \\ &\leq I(X;Y_1|U_2) + I(Y_1;\Delta Y|X) \\ &\quad + I(X;(\beta_2-\beta_1)X + (\beta_2-\beta_1)N_2'/\beta_2 - N_1) \\ &\leq \frac{1}{2}\log\left(1+\frac{D_2}{D_1}-\frac{D_2}{\sigma^2}\right) + \frac{1}{2}\log\frac{D_1}{D_2}\left(1-\frac{D_1-D_2}{\sigma^2}\right) \\ &\quad + \frac{1}{2}\log\frac{1}{1-(D_1-D_2)/\sigma^2} \\ &= \frac{1}{2}\log\left(1+\frac{D_1}{D_2}-\frac{D_1}{\sigma^2}\right). \end{aligned} \tag{29}$$

On the other hand

$$\begin{aligned} L_2 &= I(X;Y_1) + [I(X;\Delta Y) - I(X;U_2)] + I(Y_1;\Delta Y|X) \\ &\leq I(X;\beta_1 X + \beta_1 N_2'/\beta_2 + N_1) + I(X;\Delta Y|U_2) \\ &\quad + I(Y_1;\Delta Y|X) \\ &\leq \frac{1}{2}\log\frac{\sigma^2}{D_1} + \frac{1}{2}\log\left(1+\frac{(D_1-D_2)D_2}{\sigma^2(\sigma^2-D_1+D_2)}\right) \\ &\quad + \frac{1}{2}\log\frac{D_1}{D_2}\left(1-\frac{D_1-D_2}{\sigma^2}\right) \\ &= \frac{1}{2}\log\frac{\sigma^2(\sigma^2-D_1+D_2)+(D_1-D_2)D_2}{\sigma^2 D_2}. \end{aligned} \tag{30}$$

Combining Lemma 5, (26), and (27) with (28) gives the last bound. □

We can combine the bounds given in (29) and (30) for fixed D_2. The first one dominates when $D_1 \leq (\sigma^2+D_2)/2$, while the second one dominates when $D_1 > (\sigma^2+D_2)/2$. The maximal value of the combined bound is achieved at $D_1 = (\sigma^2+D_2)/2$, giving $L_2 \leq (1/2)\log(1+(\sigma^4-D_2^2)/(2D_2\sigma^2))$.

These bounds are good for the low-rate region, especially for large D_2. For example, if either $D_2/D_1 \geq 1/3$ or $D_2/\sigma^2 \geq \sqrt{10}-3 \approx 0.1623$, we have $L_1 \leq 1/2$ and $L_2 \leq 1$. Although they depend on D_1, D_2, and σ^2, they may still be interesting since no tight characterization is known for the achievable region, and it is generally difficult to compute from the original result in [15].

Based on the proof of Theorem 10, we next obtain a new bound that depends on $h(X)$, the differential entropy of the source. The bound can be easily computed using only the variance and the differential entropy of the source. This bound is tight if X is Gaussian.

Theorem 11: For any $D_2 < D_1$, there exists an A2RSC-achievable vector (R_1, R_2, D_1, D_2) with

$$\begin{aligned} L_1 &\leq \min\left\{\frac{1}{2}, \frac{1}{2}\log(2\pi e\sigma^2) - h(X)\right\} \\ L_2 &\leq \frac{1}{2}\log(2\pi e\sigma^2) - h(X). \end{aligned}$$

Proof: From the Shannon lower bound

$$I(X;U_2) = R(D_2) \geq h(X) - \frac{1}{2}\log(2\pi e D_2).$$

From (28), there exists an A2RSC-achievable vector (R_1, R_2, D_1, D_2) with $L_1 \leq 1/2$ and

$$\begin{aligned} L_2 &= I(X;\Delta Y) + I(Y_1;\Delta Y|X) + I(X;Y_1) - I(X;U_2) \\ &\leq \frac{1}{2}\log\frac{\sigma^2}{\sigma^2-D_1+D_2} + \frac{1}{2}\log\frac{D_1}{D_2}\left(1-\frac{D_1-D_2}{\sigma^2}\right) \\ &\quad + \frac{1}{2}\log\frac{\sigma^2}{D_1} - h(X) + \frac{1}{2}\log(2\pi e D_2) \\ &= \frac{1}{2}\log(2\pi e\sigma^2) - h(X). \end{aligned}$$

Similarly,

$$L_1 \leq I(X;Y_1) - I(X;U_1) \leq \frac{1}{2}\log(2\pi e\sigma^2) - h(X). \quad \square$$

An outer bound on the achievable region for A2RSCs also appears in [15]. It is easy to show the existence of a constant bound on the rate loss if the outer bound is tight. Set $Y_2 = X + N_2$, $Y_1 = X + N_2 + N_1$, where $(N_1, N_2) \perp\!\!\!\perp X$, $N_1 \perp\!\!\!\perp N_2$, $N_2 \sim \mathcal{N}(0, D_2)$, and $N_1 \sim \mathcal{N}(0, D_1 - D_2)$. Then, $\Delta Y = -N_1$. Therefore, $L_1 = I(X;Y_1) - R(D_1) \leq 1/2$. Since $I(X;\Delta Y) = 0$,

$$L_2 = I(X;Y_1,\Delta Y) - R(D_2) = I(X;Y_2) - R(D_2) \leq 1/2.$$

V. Summary

In this paper, we derive new rate loss bounds for MRSCs and A2RSCs, in some cases tightening existing results and in other cases treating cases where bounds did not exist previously. For 2RSCs, from the convexity of the rate loss and Theorems 3 and 4, given any rate budget at one resolution, we can immediately upper-bound the rate at the other resolution of the optimal code. While most cases lead to small, constant bounds, we believe that these bounds are not tight in general. For example, Lastras and Berger show that $L_1 \leq 1/2$ and $L_2 = 0$ are achievable as $D_2 \to 0$, our bound is $L_1 \to 1$ in this configuration.

Unlike most of Lastras and Berger's results, it may be difficult to extend most of our results to other distortion measure. For example, we construct $Y_1 = U_2 + N_1$ in the proof of Theorem 3, but this construction cannot be extended to other distortion measure directly, e.g.,

$$E[(X-(U_2+N_1))^4] \neq E[(X-U_2)^4] + E[N_1^4].$$

However, the results of Theorems 7 and 8 and a looser version of Theorem 4 can be extended to any difference distortion measure.

Another interesting issue is the implication of the rate loss bounds for code design. The entropy-coded dithered-lattice quantizer (ECDQ) [22], [23] is a uniform quantizer followed by a noiseless variable-rate encoder. The input of the quantizer is the sum of the source and an independent, uniformly distributed random variable and the output is the quantizer reproduction minus the same random noise. If we apply the approach developed in [24] to MRSCs, we can in practice design a code with rate loss [25]

$$\begin{aligned} L_i &\leq \frac{1}{2}\sum_{j=1}^{i} \log(2\pi e G_{Kj}) + \frac{1}{2}\log\left(2 - \frac{D_i}{\sigma^2}\right) \\ &\leq \frac{1}{2}\sum_{j=1}^{i} \log(2\pi e G_{Kj}) + 0.5 \end{aligned}$$

where K is the codeword length and G_{Kj} is the normalized second moment of the K-dimensional lattice at the jth resolution (e.g., see [22]). For example, if we use one-dimensional lattices, then $G_1 = 1/12$, and $L_1 < 0.7547$, $L_2 < 1.0094, \ldots, L_i < 0.2547i + 0.5$. If we use optimal lattices and let $K \to \infty$, then $G_K \to 1/(2\pi e)$, and $L_i \leq 0.5$ at any resolution. Unfortunately, we have not found corresponding practical coding schemes for other scenarios.

APPENDIX

Proof of Lemma 1: Let U_1 be the random variable that achieves $R(D_1)$, and let $N_2 \sim \mathcal{N}(0, D_2 - \alpha^2 D_1)$, with $N_2 \perp\!\!\!\perp (X, U_1)$ and $\alpha = D_2/D_1$. Note that for any $D_2 > 0$

$$D_2 - \alpha^2 D_1 = D_2 - \left(\frac{D_2}{D_1}\right)^2 D_1 = \frac{D_2}{D_1}(D_1 - D_2) > 0$$

implies that we have a legitimate distribution. Next, let $Y_2 = (1-\alpha)X + \alpha U_1 + N_2$. Notice that

$$\begin{aligned} Ed(X, Y_2) &= E(\alpha(X-U_1) - N_2)^2 \\ &= \alpha^2 E(X-U_1)^2 + D_2 - \alpha^2 D_1 \\ &\leq \alpha^2 D_1 + D_2 - \alpha^2 D_1 = D_2. \end{aligned}$$

Thus, $R(D_2) \leq I(X; Y_2)$, which implies

$$\begin{aligned} &R(D_2) - R(D_1) \\ &\quad \leq I(X; Y_2) - I(X; U_1) \\ &\quad = I(X; Y_2|U_1) - I(X; U_1|Y_2) && (31) \\ &\quad \leq I(X; Y_2|U_1) \\ &\quad = I(X; (1-\alpha)X + \alpha U_1 + N_2|U_1) \\ &\quad = I(X-U_1; (1-\alpha)(X-U_1) + N_2|U_1) && (32) \\ &\quad \leq I(X-U_1; (1-\alpha)(X-U_1) + N_2) && (33) \\ &\quad \leq \sup I(W; (1-\alpha)W + N_2) && (34) \\ &\quad = \frac{1}{2}\log\frac{(1-\alpha)^2 D_1 + D_2 - \alpha^2 D_1}{D_2 - \alpha^2 D_1} \\ &\quad = \frac{1}{2}\log\frac{D_1}{D_2} && (35) \end{aligned}$$

where (31) follows by applying the chain rule twice to $I(X; Y_2, U_1)$, (32) follows since $h(A|B) = h(A-B|B)$, and (33) follows since $U_1 \to X - U_1 \to (1-\alpha)(X-U_1) + N_2$ forms a Markov chain by the independence assumptions. In (34), we take the supremum over all random variables $W \perp\!\!\!\perp N_2$ such that $E(W^2) \leq D_1$; the supremum is achieved by $W \sim \mathcal{N}(0, D_1)$, giving the desired result. □

This bound is tight. The Gaussian source achieves this bound. If we set $Y_2 = X + N$, we can show a looser bound which can be extended to any difference distortion measure, where $Ed(N, 0) = D_2$ and $N \perp\!\!\!\perp (X, U_1)$. For example, the looser bound for the squared error distortion measure would be

$$R(D_2) - R(D_1) \leq \frac{1}{2}\log\left(1 + \frac{D_1}{D_2}\right).$$

In this case, we can also notice that $f_1(D_2) = R(D_2) - R(D_1)$ is a convex function of D_2 for fixed D_1 and this function achieves 0 if $D_2 = D_1$. Therefore, the combination of function $f_2(D_2) = (1/2)\log(1 + D_1/D_2)$ and its tangent which passes the point $(D_1, 0)$ provides a better bound.

Proof of Lemma 2: Following [12, Lemma 2], if

$$(R_{10}, R_{20}, D_1, D_2) \quad \text{and} \quad (R_{11}, R_{21}, D_1, D_2)$$

are achievable, then $(R_{1\alpha}, R_{2\alpha}, D_1, D_2)$ is achievable because of the convexity of the achievable rate-distortion region of the multiresolution codes, where

$$R_{1\alpha} = \alpha R_{11} + (1-\alpha)R_{10}$$

and

$$R_{2\alpha} = \alpha R_{21} + (1-\alpha)R_{20}.$$

(While the proof in [12, Lemma 2] uses incremental rates, the result generalizes immediately to total rates. In particular, if the vectors $(R_{10}, \Delta R_0 = R_{20} - R_{10}, D_1, D_2)$ and $(R_{11}, \Delta R_1 = R_{21} - R_{11}, D_1, D_2)$ of incremental rates and total distortions are achievable, then $(R_{1\alpha}, \Delta R_\alpha = \alpha \Delta R_1 + (1-\alpha)\Delta R_0, D_1, D_2)$ is achievable by [12, Lemma 2]. The corresponding total rate is

$$\begin{aligned} R_{1\alpha} + \Delta R_\alpha &= \alpha R_{11} + (1-\alpha)R_{10} + \alpha \Delta R_1 + (1-\alpha)\Delta R_0 \\ &= \alpha R_{21} + (1-\alpha)R_{20} = R_{2\alpha} \end{aligned}$$

which gives the convexity result used above.)

The corresponding rate losses are

$$\begin{aligned} L_{1\alpha} &= R_{1\alpha} - R(D_1) = \alpha R_{11} + (1-\alpha)R_{10} - R(D_1) \\ &= \alpha L_{11} + (1-\alpha)L_{10} \\ L_{2\alpha} &= R_{2\alpha} - R(D_2) = \alpha R_{21} + (1-\alpha)R_{20} - R(D_2) \\ &= \alpha L_{21} + (1-\alpha)L_{20} \end{aligned}$$

giving the desired result. □

Proof of Lemma 3: By using the chain rule for mutual information twice, we get

$$I(A; B, C|D) = I(A; B|D) + I(A; C|B, D) \quad (36)$$
$$= I(A; C|D) + I(A; B|C, D). \quad (37)$$

Now we can show $I(A; D|B) = 0$ since

$$\begin{aligned} I(A; D|B) &\le I(A; D|B) + I(A; C|B, D) \\ &= I(A; C, D|B) \\ &= I(A; C|B) + I(A; D|B, C) = 0. \end{aligned}$$

This also shows that $I(A; C|B, D) = 0$. In addition, $I(A; D|C) = 0$, therefore,

$$\begin{aligned} I(A; B|C, D) &= I(A; B|C, D) + I(A; D|C) \\ &= I(A; B, D|C) \\ &= I(A; B|C) + I(A; D|B, C) \\ &= I(A; B|C). \end{aligned} \quad (38)$$

Equations (36), (37) and $I(A; C|B, D) = 0$ imply that

$$I(A; B|D) = I(A; C|D) + I(A; B|C, D)$$

Thus, from (38)

$$I(A; B|D) = I(A; C|D) + I(A; B|C) \ge I(A; B|C). \quad \square$$

Proof of Lemma 4: We first prove that this lemma is true for $M = 1$. By applying the chain rule twice to $I(X; U_1, X+N_1)$, we obtain

$$\begin{aligned} &I(X; X+N_1|U_1) - I(X; U_1|X+N_1) \\ &\quad = I(X; X+N_1) - I(X; U_1) \\ &\quad = I(X; X+N_1) - R(D_1) \ge 0. \end{aligned}$$

Thus, $I(X; U_1|X+N_1) \le I(X; X+N_1|U_1) \le 1/2$ by the argument from steps (32)–(35).

Now suppose that this lemma holds for $M = k \ge 1$, i.e.,

$$I(X; U_1, U_2', \ldots, U_k'|X+N_k) \le k/2 \quad (39)$$

where $N_k \sim \mathcal{N}(0, D_k)$ and $N_k \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_k')$. Then for $M = k+1$, let $N_{k+1} \sim \mathcal{N}(0, D_{k+1})$, $N_k' \perp\!\!\!\perp N_{k+1}$, $N_k' \sim \mathcal{N}(0, D_k - D_{k+1})$, and

$$(N_k', N_{k+1}) \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_{k+1}').$$

By the chain rule

$$\begin{aligned} &I(X; U_1, U_2', \ldots, U_{k+1}'|X+N_{k+1}) \\ &\quad = I(X; U_1, U_2', \ldots, U_k'|X+N_{k+1}) \\ &\qquad + I(X; U_{k+1}'|U_1, U_2', \ldots, U_k', X+N_{k+1}). \end{aligned} \quad (40)$$

We bound the first term on the right-hand side of (40) as

$$\begin{aligned} &I(X; U_1, U_2', \ldots, U_k'|X+N_{k+1}) \\ &\quad \le I(X; U_1, U_2', \ldots, U_k'|X+N_{k+1}+N_k') && (41) \\ &\quad = I(X; U_1, U_2', \ldots, U_k'|X+N_k) \\ &\quad \le k/2 && (42) \end{aligned}$$

where (41) follow from Lemma 3, and (42) follows from the previous assumption (39).

The second term on the right-hand side of (40) can be bounded as follows:

$$\begin{aligned} &I(X; U_{k+1}'|U_1, U_2', \ldots, U_k', X+N_{k+1}) \\ &\quad \le I(X; X+N_{k+1}|U_1, U_2', \ldots, U_k', U_{k+1}') && (43) \\ &\quad = I(X-U_{k+1}'; X-U_{k+1}'+N_{k+1}|U_1, U_2', \ldots, U_{k+1}') \\ &\quad \le I(X-U_{k+1}'; X-U_{k+1}'+N_{k+1}) && (44) \\ &\quad \le 1/2 && (45) \end{aligned}$$

where (43) follows by applying the chain rule twice to $I(X; U_1, U_2', \ldots, U_k', U_{k+1}', X+N_{k+1})$ to get

$$\begin{aligned} &I(X; X+N_{k+1}|U_1, U_2', \ldots, U_k', U_{k+1}') \\ &\quad - I(X; U_{k+1}'|U_1, U_2', \ldots, U_k', X+N_{k+1}) \\ &= I(X; U_1, U_2', \ldots, U_k', X+N_{k+1}) \\ &\quad - I(X; U_1, U_2', \ldots, U_k', U_{k+1}') \ge 0 \end{aligned}$$

(44) follows since

$$(U_1, U_2', \ldots, U_k', U_{k+1}') \to X - U_{k+1}' \to X - U_{k+1}' + N_{k+1}$$

forms a Markov chain, and (45) parallels the chain of inequalities in steps (34)–(35).

Combining (40), (42), and (45) proves this lemma for $M = k+1$. Finally, by induction, this lemma holds for any M. $\square$

Proof of Lemma 5: Let

$$a = \frac{\sigma^2 - D_1}{\sigma^2 - D_0} \quad \text{and} \quad b = \frac{\sigma^2 - D_2}{\sigma^2 - D_0}$$

then $D_1 = aD_0 + (1-a)\sigma^2$. From the convexity of the rate-distortion function

$$R(D_1) \le aR(D_0) + (1-a)R(\sigma^2) = aR(D_0).$$

By symmetry, $R(D_2) \le bR(D_0)$, thus,

$$\begin{aligned} R(D_1) + R(D_2) - R(D_0) &\le aR(D_0) + bR(D_0) - R(D_0) \\ &= \frac{\sigma^2 + D_0 - D_1 - D_2}{\sigma^2 - D_0} R(D_0) \end{aligned}$$

which is nonpositive if $D_1 + D_2 - D_0 \ge \sigma^2$. $\square$

ACKNOWLEDGMENT

The authors are grateful to the Associate Editor and anonymous reviewers for their detailed suggestions and comments; in particular, Lemma 3 and the current improved result in Theorem 10 result from their insights.

REFERENCES

[1] E. A. Riskin and R. M. Gray, "A greedy tree growing algorithm for the design of variable rate vector quantizers," *IEEE Trans. Signal Processing*, vol. 39, pp. 2500–2507, Nov. 1991.

[2] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.

[3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[4] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996.

[5] M. Effros, "Practical multiresolution source coding: TSVQ revisited.," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 1998, pp. 53–62.

[6] M. Effros and D. Dugatkin, "Multiresolution vector quantization.," *IEEE Trans. Inform. Theory*, submitted for publication.

[7] V. Koshelev, “Hierarchical coding of discrete sources,” *Probl. Pered. Inform.*, vol. 16, no. 3, pp. 31–49, 1980.
[8] W. H. R. Equitz and T. M. Cover, “Successive refinement of information,” *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.
[9] L. Lastras and T. Berger, “On the refinement of the binary symmetric Markov source,” in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, p. 195.
[10] J. Chow and T. Berger, “Failure of successive refinement for symmetric Gaussian mixtures,” *IEEE Trans. Inform. Theory*, vol. 43, pp. 350–352, Jan. 1997.
[11] B. Rimoldi, “Successive refinement of information: Characterization of achievable rates,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.
[12] M. Effros, “Distortion-rate bounds for fixed- and variable-rate multiresolution source codes,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 1887–1910, Sept. 1999.
[13] L. Lastras and T. Berger, “All sources are nearly successively refinable,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 918–926, Mar. 2001.
[14] R. Zamir, “The rate loss in the Wyner–Ziv problem,” *IEEE Trans. Inform. Theory*, vol. 42, pp. 2073–2084, Nov. 1996.
[15] E. Tuncel and K. Rose, “Additive successive refinement,” in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, p. 31.
[16] ——, “Additive successive refinement,” *IEEE Trans. Inform. Theory*, submitted for publication.
[17] A. A. El Gamal and T. M. Cover, “Achievable rates for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.
[18] E. Tuncel and K. Rose, private communication, Mar. 2002.
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[20] A. Buzo, A. H. Gray, Jr, R. M. Gray, and J. D. Markel, “Speech coding based upon vector quantization,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–28, pp. 562–574, Oct. 1980.
[21] E. A. Riskin, R. Ladner, R. Y. Wang, and L. E. Atlas, “Index assignment for progressive transmission of full-search vector quantization,” *IEEE Trans. Image Processing*, vol. 3, pp. 307–312, May 1994.
[22] J. Ziv, “On universal quantization,” *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 344–347, May 1985.
[23] R. Zamir and M. Feder, “On lattice quantization noise,” *IEEE Trans. Inform. Theory*, vol. 42, pp. 1152–1159, July 1996.
[24] Y. Frank-Dayan and R. Zamir, “Dithered lattice-based quantizers for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 192–204, Jan. 2002.
[25] H. Feng, Q. Zhao, and M. Effros, “Network source coding using entropy constrained dithered quantization,” in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2003.