Rate Loss of Network Source Codes

Thesis by

Hanying Feng

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Defended May 6th, 2002)

To my parents

and

Qian

Acknowledgments

I wish to express my deepest appreciation to my advisor Professor Michelle Effros of the Data Compression Lab at Caltech. Her support, guidance and encouragement throughout the period of my graduate studies made the completion of this work possible.

I would like to thank Prof. Michelle Effros, Prof. Yaser S. Abu-Mostafa, Prof. Jehoshua (Shuki) Bruck, Prof. Steven Low, and Prof. Robert J. McEliece for serving on my Ph.D. examination committee, reading my thesis manuscript, and giving their valuable comments. I also want to thank Prof. P. P. Vaidyanathan and Dr. Dariush Divsalar for serving on my Ph.D. candidacy examination committee.

I would also like to thank my former and current fellow graduate students in Caltech's Communications Group for their friendship and help; special thanks go to Diego Dugatkin, Michael Fleming, Sidharth Jaggi, Chaitanya Rao, Qian Zhao, Hui Jin, Qingdi Liu, and Lifang Li.

Under Prof. Michelle Effros' supervision, I also worked on other projects, which are not included in this thesis. The first project treats the suboptimality of the Karhunen-Loève transform (KLT) for transform coding, and proposes both an improved version of the KLT and fast variations on the KLT [1], [2], [3], [4]. The second project treats multi-resolution

channel codes [5].

This material is based upon work partially supported by NSF Grant No. CCR-9909026 and Caltech's Lee Center for Advanced Networking.

Abstract

In this thesis, I present bounds on the performance of a variety of network source codes. These *rate loss* bounds compare the rates achievable by each network code to the rate-distortion bound $R(D)$ at the corresponding distortions. The result is a collection of optimal performance bounds that are easy to calculate.

I first present new bounds for the rate loss of multi-resolution source codes (MRSCs). Considering an M-resolution code with $M \geq 2$, 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. For 2-resolution 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 in scenarios (i) and (ii) and an asymptotic bound for scenario (iii). In this thesis, I show a constant bound for scenario (iii), tighten the bounds for scenario (i) and (ii), and generalize the bound for scenario (ii) to M-resolution greedy codes.

I also present upper bounds for the rate losses of additive MRSCs (AMRSCs), a special MRSC. I obtain two bounds on the rate loss of AMRSCs: one primarily good for low rate coding and another which depends on the source entropy.

I then generalize the rate loss definition and present upper bounds for the rate losses of multiple description source codes. I divide the distortion region into three sub-regions and bound the rate losses by small constants in two sub-regions and by the joint rate losses of a normal source with the same variance in the other sub-region.

Finally, I present bounds for the rate loss of multiple access source codes (MASCs). I show that lossy MASCs can be almost as good as codes based on joint source encoding.

Contents

List of Figures

Chapter 1

Introduction

1.1 Source Coding for One-to-One Communication

Shannon first stated the source coding theorem in his original paper [6], which considers the question of how short a description is possible if we want to represent information perfectly. This is usually referred as lossless data compression. In the same paper, he also introduced the idea of rate distortion, which deals with using the shortest possible description to represent information with some maximal allowed distortion. The rate distortion problem is also called lossy data compression, which is a superset of the lossless data compression. Shannon treated the lossy data compression question more exhaustively in [7], which proved the first rate distortion theorem. Since then, extensive theoretical and practical work has been done on source coding for one-to-one communication systems (e.g., [8], [9], [10], [11]). In one-to-one communication systems, only one sender and one receiver are allowed. The sender uses a single packet with rate R to describe the source without knowing other information, and the decoder decodes the single packet with distortion D without access to any other information, as shown in Figure 1.1.

Figure 1.1: A one-to-one communication system.

1.2 Source Coding for Network Communication

Recently, network systems and applications are becoming increasingly prevalent. The major services they provide are expanding from basic service to more versatile wide-band services. Unfortunately, further development of advanced network technologies is limited by the amount of information that can be sent through the corresponding networks. Therefore, it is imperative to use efficient data representations for optimizing network system performance.

A network is defined as any collection of nodes joined by communication links. Every node has a collection of sources to be encoded for transmission and a collection of sources for which it receives descriptions and builds reproductions.

At first look, it may be convenient to use one-to-one communication technique or independent coding for any two nodes with communication link. However, this is not an efficient way. In reality, a lot of information communicated in a network may be correlated or even identical. It does not make much sense to send the same message or very similar messages again and again. In other words, it might be much more efficient to remove the redundancy among sources sent or received by the same node, which is called *network source coding* or source coding for network communication. A more comprehensive treatment of network source coding can be found in [12].

Though network source coding generally improves the overall network performance, it comes at a price. First, joint encoding and joint decoding can increase both the design and the run-time complexity of practical coding algorithms. Second, the optimal performance limits of network codes are often more difficult to analyze than those of one-to-one systems.

In this thesis, I bound the optimal performance of a variety of lossy network source codes by bounding their *rate loss.* The following discussion introduces the notion of rate loss by considering the rate lossless for Wyner-Ziv Codes (WZSC), Multi-Resolution Source Codes (MRSCs), Multiple Description Source Codes (MDSCs), Multiple Access Source Codes (MASCs), and Additive MRSCs (AMRSCs). These results also appear in [13], [14], [15], [16], and [17].

1.3 Wyner-Ziv Source Codes

The concept of rate loss was first introduced by Zamir for WZSCs in [18]. A WZSC is a source code designed for one-to-one communication systems in which side information is available to the decoder but not available to the encoder, as shown in Figure 1.2. The theoretical performance of WZSCs, which is usually very difficult to analyze, has been determined by Wyner and Ziv in [19] and [20].

In [18], Zamir bounds the performance of WZSCs by finding an upper bound on the rate loss of a WZSC relative to a code with side information available to both the encoder and the decoder. More precisely, Zamir defines the rate loss of a WZSC as the maximal difference

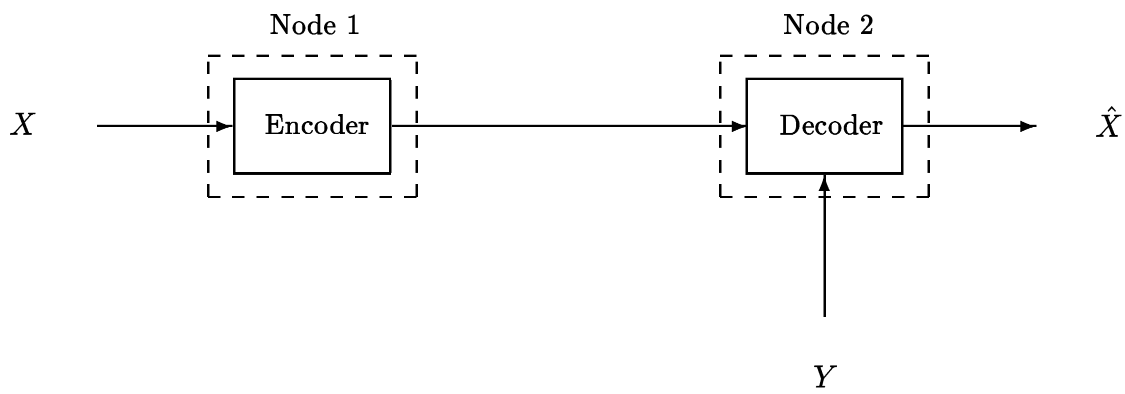

Figure 1.2: A WZSC. Node 1 describes source X to node 2, which uses side information Y to decode that description and build the reproduction $\hat{X}$.

between the rate of an optimal WZSC and the rate of the conditional rate-distortion function $R_{X|Y}(D)$ at the same distortion. By finding a source-independent bound of 0.5 bits per symbol (bps) for this difference in the squared-error distortion case, Zamir bounds the rate-distortion function of a WZSC as a function of the conditional rate-distortion function, which is generally much simpler to analyze.

1.4 Multi-Resolution Source Codes

Chapter 3 treats an extension of Zamir's rate loss to MRSCs. An MRSC is a source code in which simple, low-rate source descriptions are embedded in more complex, high-rate descriptions. Because of their ability to satisfy varying bandwidth constraints, computational capabilities, and performance requirements with a single code, MRSCs are playing an increasingly important role in research and in practice (e.g., [21], [22], [23], [24], [25], [26]). While MRSCs yield greater flexibility than traditional, single-resolution codes (1RSCs), that flexibility comes at a price in rate-distortion performance. In particular, for any $R_2 > R_1 \geq 0$ and any $D_1 > D_2 \geq 0$, we call the vector (R_1, R_2, D_1, D_2) *achievable* by multi-resolution coding on source X if there exists an MRSC that uses R_1 bps to describe X with distortion D_1 and then uses an additional $R_2 - R_1$ bps to refine the description to distortion D_2, as shown in Figure 1.3. The rate loss (L_1, L_2) 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. The rate loss describes the performance degradation associated with using a 2RSC rather than the best 1RSC with the same distortion.

A source is called *successively refinable* if the 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 [27]. Necessary and sufficient conditions for a source to be successively refinable appear in [28]. Examples of sources that are not successively refinable are shown for discrete-alphabet and continuous-alphabet sources in [28] and [29] respectively. The MRSC achievable rate-distortion region for non-successively refinable sources appears in [30] and [31].

In [32], Lastras and Berger consider the question of whether there exists a source in which the MRSC rate loss can be made arbitrarily large. Following the approach developed by Zamir for Wyner-Ziv codes in [18], they demonstrate that for *any* memoryless source and the mean squared error (mse) distortion measure, 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 $D_2 \rightarrow 0$, $L_1 \leq 0.5$ and $L_2 = 0$ is achievable.

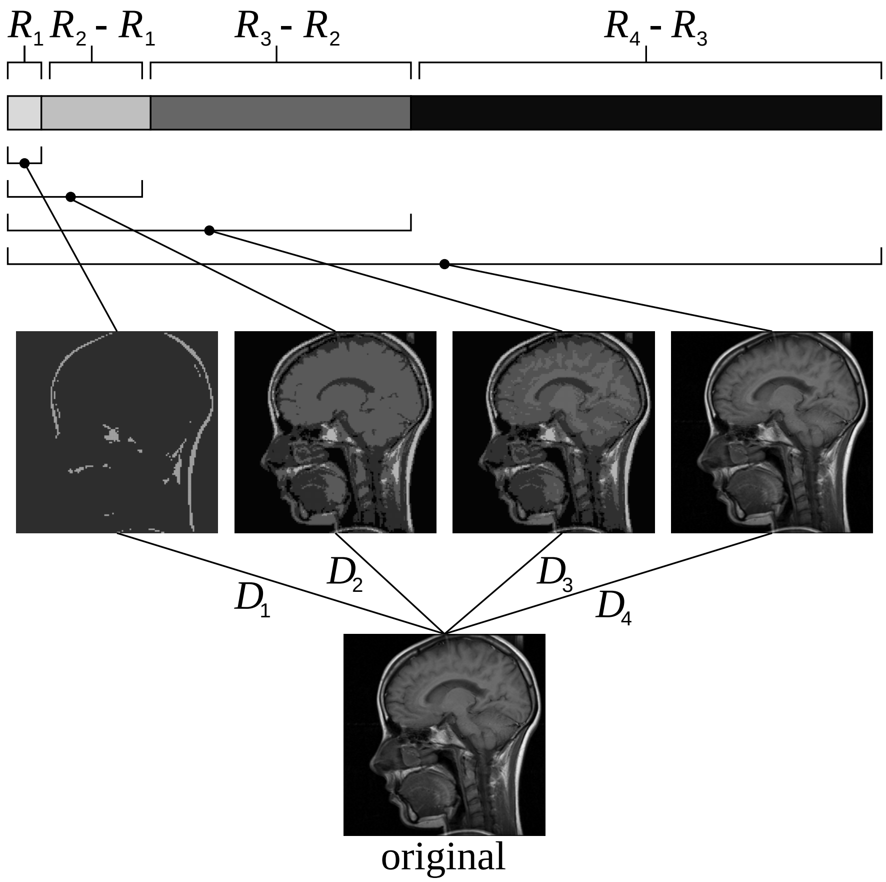

Figure 1.3: A 4RSC. Decoding the first R_1 bps of the binary description yields a reproduction with distortion D_1. Decoding an additional $R_2 - R_1$ bps, for a total rate of R_2 bps, yields a reproduction of distortion $D_2 < D_1$, and so on.

1.5 Multiple Description Source Codes

Chapter 4 treats the rate loss of MDSCs. In using packet-based communication systems for delay-sensitive data types, it may be necessary for the receiver to build a data reconstruction based on an incomplete data description. An MDSC is a code designed to give good rate-distortion performance under a variety of packet-loss scenarios. Figure 1.4 shows a 2-packet MDSC (2DSC). We measure 2DSC performance as $(R_1, R_2, D_0, D_1, D_2)$, where (R_i, D_i) $i \in$

$\{1, 2\}$ are the expected rate and distortion for packet i and D_0 is the expected distortion in jointly decoding the two packets.

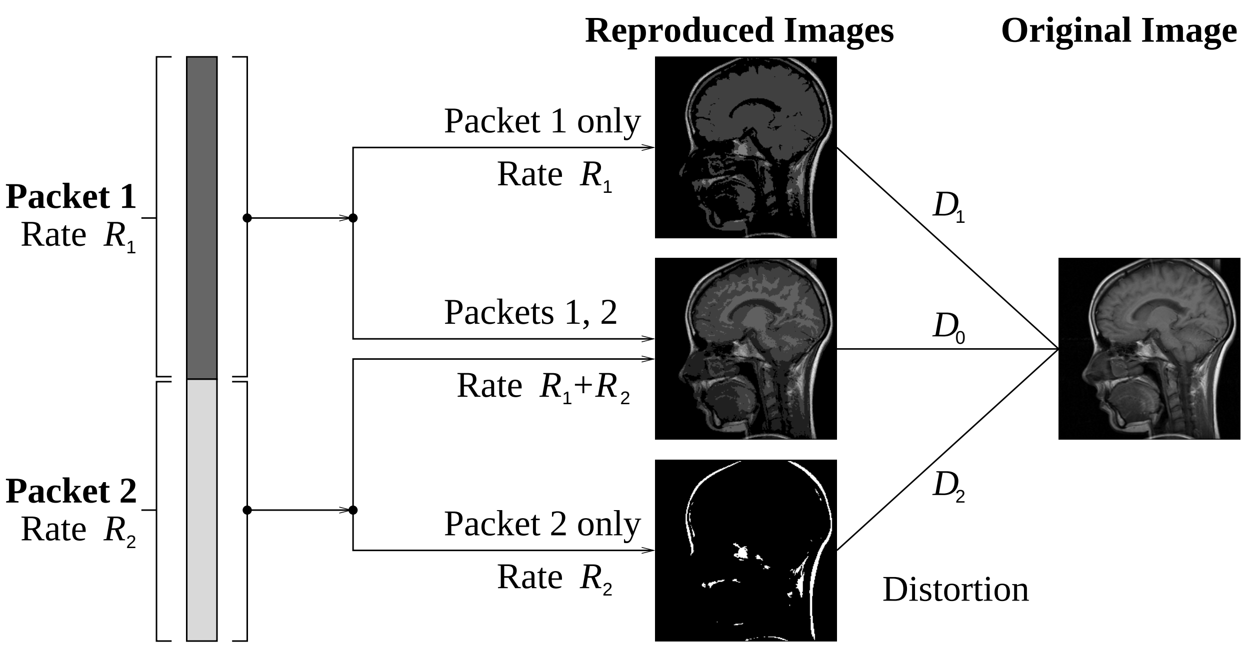

Figure 1.4: A 2DSC for image coding. Decoding the first binary description with R_1 bps yields a reproduction with distortion D_1, decoding another description with R_2 bps yields a reproduction of distortion D_2, and decoding both of them jointly yields a reproduction of distortion $D_0 < \min\{D_1, D_2\}$.

The rate loss of an MDSC describes the performance penalties of MDSCs relative to 1RSCs. I measure this penalty by the rate loss vector $L = (L_0, L_1, L_2)$ of an MDSC. Given rate-distortion function $R(D)$, the rate loss of a 2DSC is $L_i = R_i - R(D_i)$ $(i = 0, 1, 2)$ (here $R_0 = R_1 + R_2$).

1.6 Multiple Access Source Codes

In Chapter 5, I consider the problem of distributed data compression. In some situations, several senders may send information to a common receiver simultaneously. If communication among the senders is not allowed, then we call the system a multiple access system. Examples of multiple access systems include satellite systems where many independent satellites wish to communicate with a common ground receiver, and sensor array systems, where all the sensors in the system are sending information to a central processing unit which processes the data collected by all of the sensors. In such systems, the senders may transmit correlated information to the receiver, but cooperation among the senders is not possible or prohibitively expensive since the senders are remotely located. MASCs are source codes designed for multiple access systems. Figure 1.5 shows a 2-sender MASC (2ASC). We measure 2ASC performance as (R_1, R_2, D_1, D_2), where (R_1, D_1) and (R_2, D_2) are the expected rates and distortions for source X and Y respectively.

In 1973, Slepian and Wolf showed the surprising result that independent encoders in a lossless 2ASC can essentially achieve the same performance as a single encoder (or a pair of cooperating encoders) jointly encoding the pair of sources [33]. Tung and Berger investigate the lossy source coding problem for this system and give one inner bound and one outer bound on the achievable rate-distortion region in [34]. The inner bound is not tight in general, and the relationship between the inner bound and the achievable region for joint encoding is difficult to understand. Zamir and Berger show that the performance penalty associated with using a 2ASC rather than a joint encoder vanishes in the limit of high resolution [35]; however, this performance penalty is not zero in general (see, [20] and [18]).

Given joint rate-distortion function $R_{X,Y}(D_1, D_2)$ and conditional rate-distortion functions $R_{X|Y}(D_1)$ and $R_{Y|X}(D_2)$, the rate losses of a 2ASC are $L_1 = R_1 - R_{X|Y}(D_1)$, $L_2 = R_2 - R_{Y|X}(D_2)$, and $L_0 = R_1 + R_2 - R_{X,Y}(D_1, D_2)$.

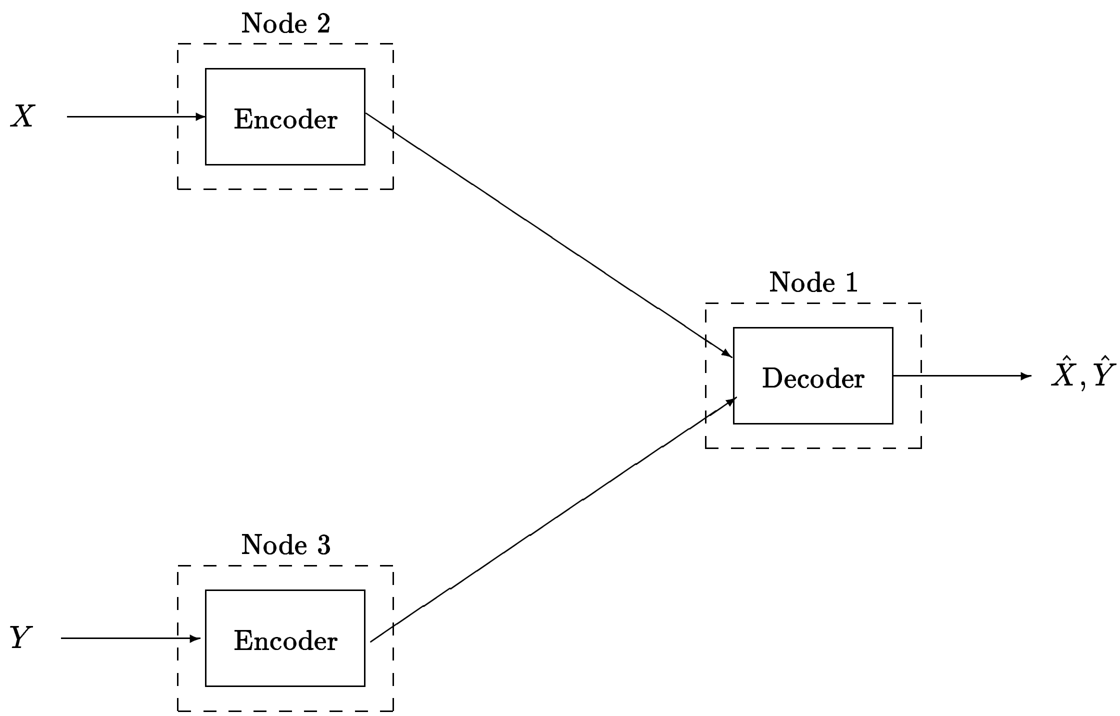

Figure 1.5: A 2ASC. Two senders at nodes 2 and 3 are independently sending messages X and Y to the receiver at node 1, which jointly decodes all the information he receives to construct the reproductions $\hat{X}$ and $\hat{Y}$.

1.7 Additive Multi-Resolution Source Codes

In Chapter 6, I derive two bounds on the rate losses of a special type of MRSCs called Additive Multi-Resolution Source Codes (AMRSCs). An AMRSC is an MDSC used as MRSC, which is also known as additive successive refinement codes. In particular, the kth-resolution reproduction of an AMRSC equals the sum of the independent reconstructions from the MRSC's first k packets [36]. A 2-stage AMRSC (A2RSC) encodes source X using two packets with expected 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 $D_2 < D_1$. The definition for the rate loss of an AMRSC is identical to the corresponding definition for an MRSC. AMRSCs are of potential interest since their codebook storage requirements are lower than those of other MRSCs.

1.8 Thesis Outline

This thesis focuses on finding source-independent upper bounds for the rate loss of MRSCs, MDSCs, MASCs, and AMRSCs. Rate loss bounds are useful for several reasons. (1) They describe the performance degradation associated with using the given code rather than the best single-resolution code with the same distortion(s). For example, small constant upper bounds on the rate loss of MRSCs (e.g., [32], [13], [15]) put to rest any fears that there might exist some source on which the cost of multi-resolution coding can be made arbitrarily large. (2) They give elegant and often tight inner bounds on the region of achievable rates and distortions. These bounds are much simpler to analyze than existing alternatives (e.g., [37] and [38]). (3) Since the exact rate-distortion regions for MDSCs, MASCs, and AMRSCs are not known in general, the rate loss also gives a good bound on the distance between the best existing inner and outer bounds. (The rate-distortion function provides a natural outer bound for general sources. For example, both

$$\begin{cases} R_1 \geq R(D_1), \\ R_2 \geq R(D_2), \end{cases}$$

(which has an additional implicit bound: $R_1 + R_2 \geq R(D_1) + R(D_2)$) and

$$\begin{cases} R_1 \geq R(D_1), \\ R_2 \geq R(D_2), \\ R_1 + R_2 \geq R(D_0), \end{cases}$$

are natural outer bounds for the achievable rate-distortion (RD) region for MDSCs. It can also be shown that

$$\begin{aligned} R_1 &\geq R_{X|Y}(D_1), \\ R_2 &\geq R_{Y|X}(D_2), \\ R_1 + R_2 &\geq R_{X,Y}(D_1, D_2), \end{aligned}$$

is an outer bound on the achievable RD region for MASCs [35])

The remainder of this thesis is organized in such a way that different chapters can be read independently, except Chapter 2, which contains material needed for all of the other chapters.

Chapter 2 introduces mathematical background material, notations, and definitions.

Chapter 3 lists my main results for MRSCs. I first present a non-asymptotic bound for L_1 when $L_2 = 0$. Second, I tighten an existing bound for L_2 with $L_1 = 0$ from $L_2 \leq 1$ to $L_2 < 0.7250$. Third, I tighten an existing bound for $L_1 = L_2$ from 0.5 to 0.3802. Finally, I generalize the result for L_2 when $L_1 = 0$ from 2-resolution to M-resolution source codes for any $M \geq 2$.

Chapter 4 gives my results for MDSCs. I divide the whole distortion region into three sub-regions and upper-bound the rate loss for each sub-region separately. For the lowest and highest values of D_0, I bound rate losses by small constants. For intermediate values of D_0, I bound the rate loss as a function of the rate loss of a Gaussian random variable, which can be arbitrarily large.

Chapter 5 describes my result for MASCs. I show that though there is a performance penalty to pay for using lossy MASCs, the performance penalty is bounded by a small constant. In other words, independent encoding can be almost as good as joint encoding in lossy coding.

I then turn to AMRSCs briefly. In Chapter 6, I introduce the two upper bounds I have obtained on the rate loss of AMRSCs: one primarily good for low rate coding and another which depends on the source entropy.

All theorems are proved after they are presented. All lemmas are proved in the Appendix.

Chapter 2

Background

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

Consider two random variables X and Y with a joint pdf $f_{X,Y}(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. The *mutual information* $I(X;Y)$ is defined as:

$$I(X;Y) = \int_{(x,y)} f_{X,Y}(x, y) \log \frac{f_{X,Y}(x, y)}{f_X(x) f_Y(y)} dx dy.$$

The *rate-distortion function* for source $\{X_i\}_{i=1}^{\infty}$ and 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, I 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 my results.

Following Zamir's approach from [18], [32] bounds the rate loss for MRSCs. This thesis relies on the same tool.

In this thesis, I assume an arbitrary iid real-alphabet source X with variance σ^2 and focus on the mse distortion measure, i.e., $d(x, y) = (x - y)^2$. I use $h(X)$ and $R(D)$ to denote the differential entropy and the rate-distortion function of source X, respectively. I assume $h(X)$ and σ^2 are finite.

Chapter 3

Multi-Resolution Source Codes

3.1 Preliminaries

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

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

with rate $(1/n)\log M_1$ and resulting distortion $Ed_n(X^n, g_n^{(1)}(f_n^{(1)}(X^n)))$ and a refinement pair $(f_n^{(2)}, g_n^{(2)})$

$$f_n^{(2)} : \mathbf{R}^n \to \{1, \ldots, M_2\} \quad \text{and} \quad g_n^{(2)} : \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \to \mathbf{R}^n$$

with total rate $(1/n)\log(M_1 M_2)$ and distortion $Ed_n(X^n, g_n^{(2)}(f_n^{(1)}(X^n), f_n^{(2)}(X^n)))$.

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

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

The achievable region for 2RSCs is defined as the set of all achievable rate-distortion vectors, which is described in the following theorem. The result for finite alphabets comes from [30]. A generalization to any Polish alphabet with an escape symbol appears in [31].

Theorem 1 *[30, Theorem 1],[31, Corollary 9] For any iid source $\{X_i\}_{i=1}^{\infty}$ with density $f_X(x)$ 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*

$$R_1 \geq I(X; Y_1), \qquad Ed(X, Y_1) \leq D_1$$

$$R_2 \geq I(X; Y_2, Y_1), \quad Ed(X, Y_2) \leq D_2.$$

The result generalizes to M-RSC with $M > 2$ [30], [31].

For any MRSC with $M \geq 2$ resolutions, if the rate-distortion vector $(R_1, R_2, \ldots, R_M, D_1, D_2, \ldots, D_M)$ is achievable for $0 < D_M < \ldots < D_1 \leq \sigma^2$, the rate loss at the ith resolution ($i \in \{1, \ldots, M\}$) is defined as: $L_i = R_i - R(D_i)$ bps.

3.2 Major results

3.2.1 Achieving the Rate-Distortion Bound in Resolution 2

Theorem 2 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 Gaussian. Use of several Gaussian approximations leads to both increasing and decreasing rate loss bounds, and the intersection of these bounds yields the desired constant bound on the rate loss. Figure 3.1 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 $\frac{1}{2}\log 5$.

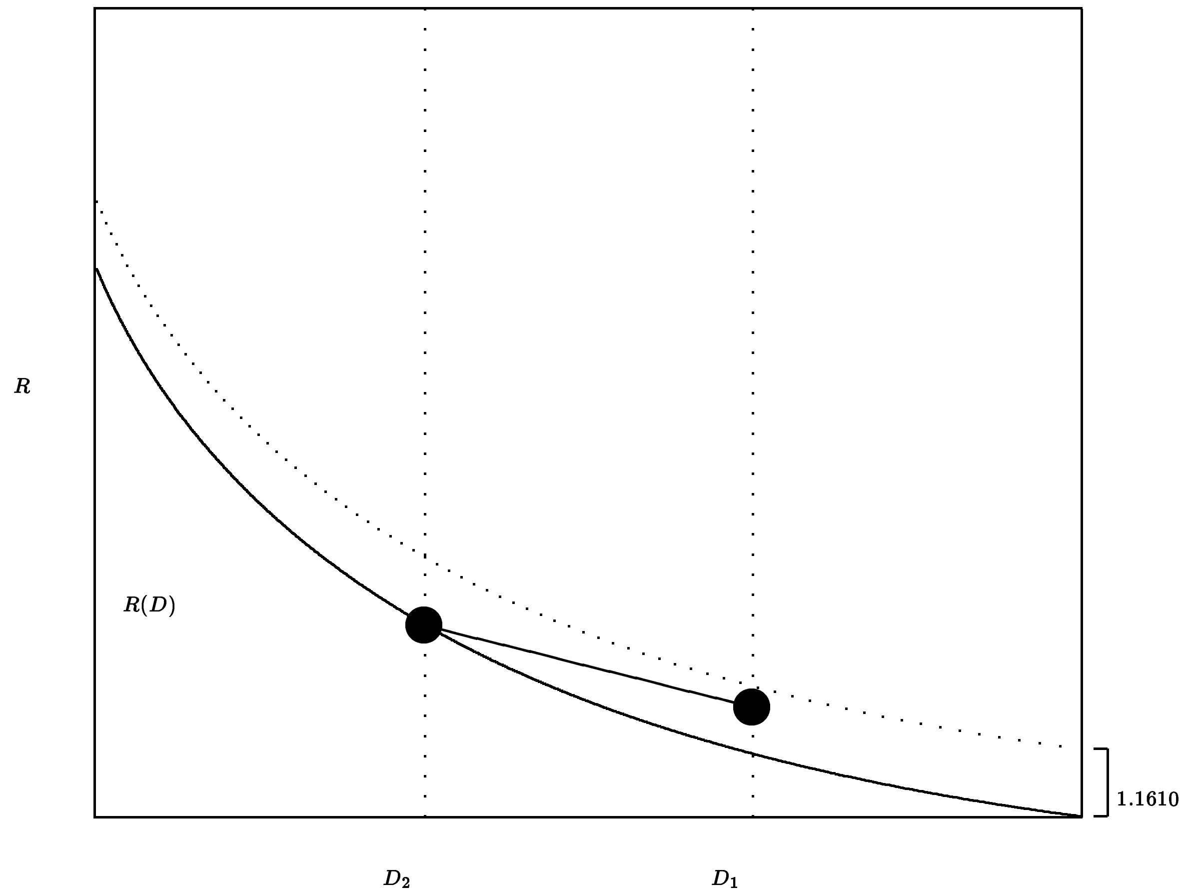

Figure 3.1: Graphical interpretation of Theorem 2.

Theorem 2 *For any iid source* $\{X_i\}_{i=1}^{\infty}$ *with an mse distortion measure, there exists a 2RSC-achievable rate-distortion vector* (R_1, R_2, D_1, D_2) *with* $L_2 = 0$ *and* $L_1 \leq \frac{1}{2}\log 5$ *bps.*

3.2.2 Achieving the Rate-Distortion Bound in Resolution 1

By applying the same basic strategy used in the proof of Theorem 2, I improve the bound given by [32, Theorem 5] from $L_1 = 0$ and $L_2 \leq 1$ to $L_1 = 0$ and $L_2 \leq \frac{1}{2}\log(\sqrt{3}+1)$.

Figure 3.2 shows the graphical interpretation of this result.

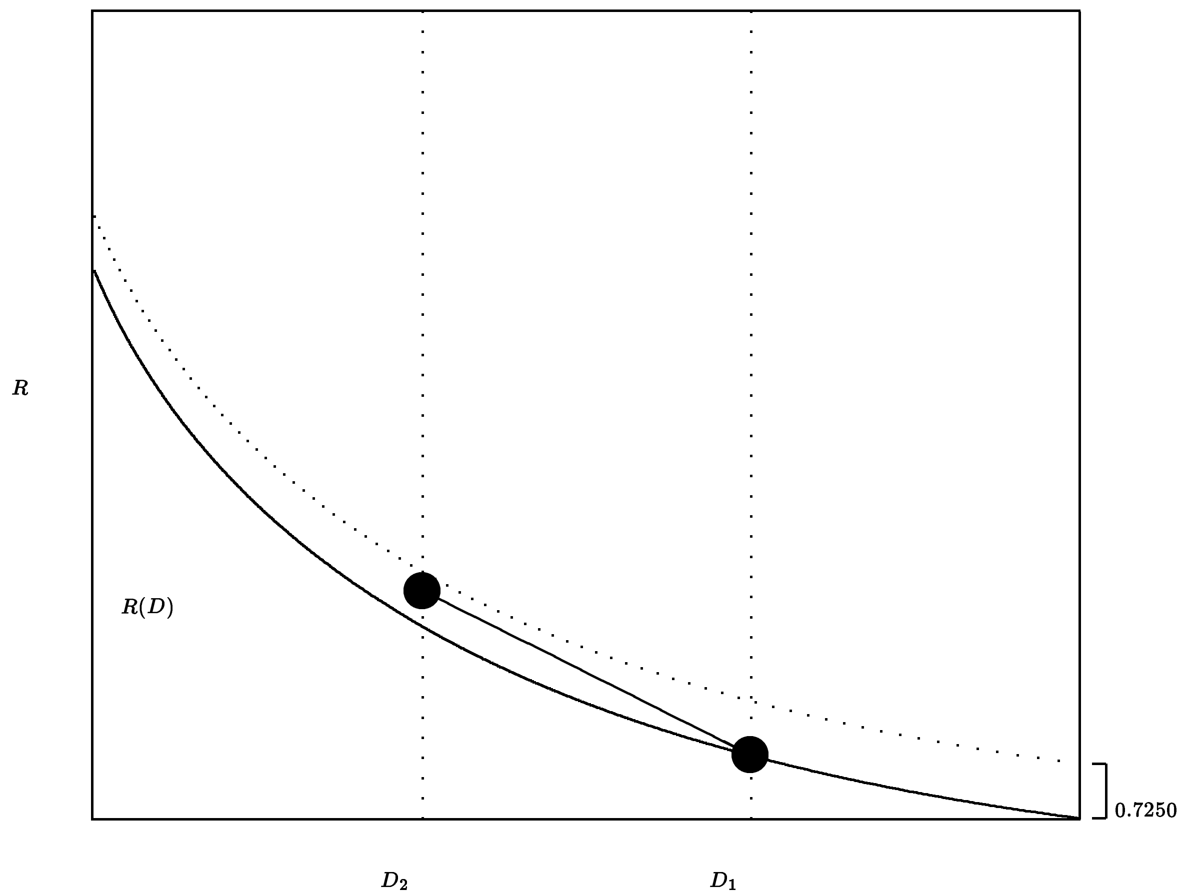

Figure 3.2: Graphical interpretation of Theorem 3.

Theorem 3 *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* $L_2 \le \frac{1}{2}\log(\sqrt{3}+1)$.

I next refine the shape of this bound using a technique employed in [32, Theorem 6].

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

The bound described in Theorem 4 is tight when $D_1 = \sigma^2$, where $L_2 = 0$. This bound can also be written as:

$$L_2 \leq 0.5 \log \left[\left(\sqrt{5 - \frac{4\sigma^2}{2\sigma^2 - D_1}} + 1 \right) \left(1 - \frac{D_1}{2\sigma^2} \right) \right],$$

which is a decreasing function of D_1 for a fixed value of σ^2. This bound achieves its maximum $L_2 \leq 0.5 \log(\sqrt{3}+1) < 0.7250$ when $D_1 = 0$, which is consistent with Theorem 3. The bound from Theorem 4 is less than the bound from Theorem 3 when $D_1 \neq 0$. Figure 3.3 shows the bound from Theorem 4. The shaded region shows the possible values of L_2 as a function of D_1/σ^2.

3.2.3 Identical Rate Losses in Both Resolutions

Lemma 1 shows the convexity of the rate loss. This result proves useful in Theorem 5, where I address the case where the rate losses at both resolutions are equal.

Lemma 1 *For any (D_1, D_2) with $D_2 < D_1$, if two 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 $0 \leq \alpha \leq 1$, 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 1 with the bounds on L_1 and L_2 from Theorem 2 and 3 gives new bounds for the case where $L_1 = L_2$. In particular, since for any $D_2 < D_1$ the rate losses $L_{10} \leq .5 \log 5$ and $L_{20} = 0$ and the rate losses $L_{11} = 0$ and $L_{21} \leq .5 \log(\sqrt{3}+1)$ are both achievable by multi-resolution coding with distortions (D_1, D_2),

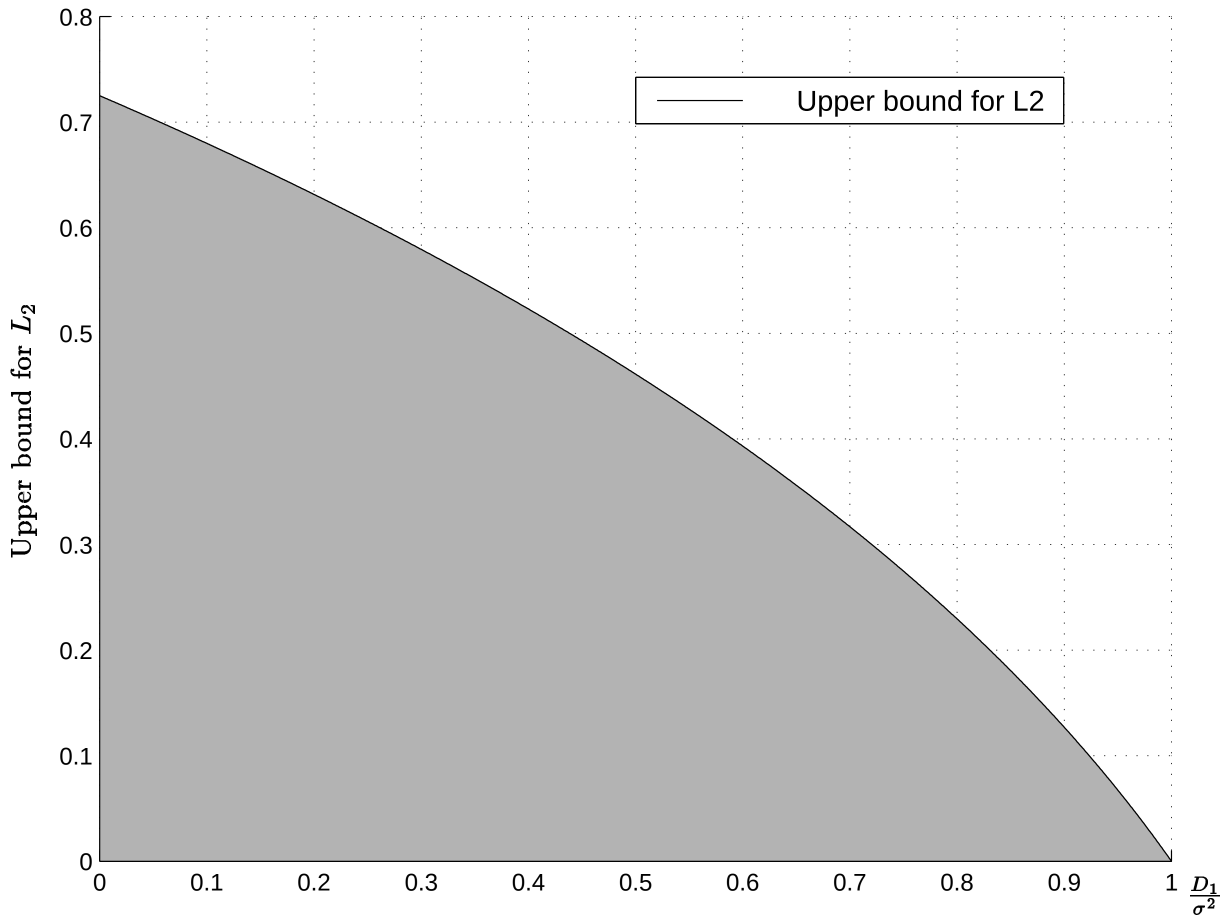

Figure 3.3: Possible values of L_2 in Theorem 4.

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

$$L_{1\alpha} = L_{2\alpha} = \frac{L_{10}L_{21}}{L_{10} + L_{21}} = \frac{1}{1/L_{10} + 1/L_{20}} \leq \frac{1}{1/(.5\log 5) + 1/(.5\log(\sqrt{3}+1))} < 0.4463$$

with distortions (D_1, D_2). This result slightly tightens the bound of [32, Theorem 3], which proves the achievability of $L_1 \leq 1/2$ and $L_2 \leq 1/2$ for distortions (D_1, D_2).

Theorem 5 improves the bound further. The graphical interpretation is shown in Figure 3.4.

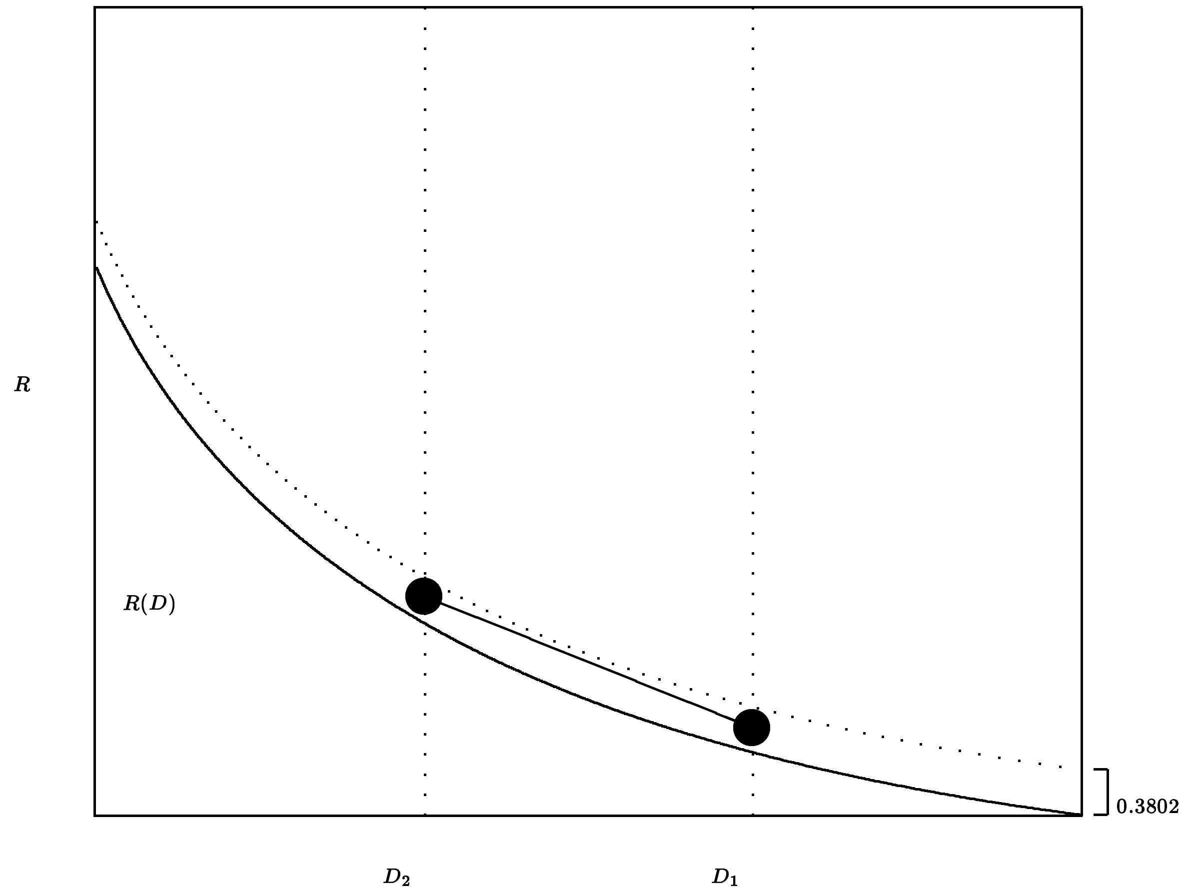

Figure 3.4: Graphical interpretation of Theorem 5.

Theorem 5 *For any $D_2 < D_1$, the rate losses $L_1 = L_2 < 0.3802$ are achievable.*

3.2.4 Tree-Structured Vector Quantization

For any $M \geq 2$ and $0 < D_M < \ldots < D_2 < D_1$, [32, 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. I 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) [39], 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 3.

Theorem 6 *For any* $D_M < \ldots < D_2 < D_1$*, there exists an achievable rate-distortion vector* $(R_1, R_2, \ldots, R_M, D_1, D_2, \ldots, D_M)$ *with* $L_1 = 0$, $L_i = I(X; U_1, U_2', U_3', \ldots, U_i') - R(D_i)$ *for all* $1 < i < M$*, and* $L_M \leq M/2$*, where* U_1 *is a random variable achieving* $R(D_1)$*,* $U_2', U_3', \ldots, U_M'$ *are successively defined as* $\arg\min_{U:Ed(X,U)\leq D_i} I(X; U, U_1, U_2', U_3', \ldots, U_{i-1}')$*,* $N_M \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_M')$*, and* $N_M \sim \mathcal{N}(0, D_M)$*.*

Theorem 6 gives a bound which increases as the number of resolutions increases. This is consistent with our intuition that the more resolutions compete for the resource, the higher the potential penalty in the rate loss for the last resolution. This bound suggests that the performance penalty associated with using greedily grown TSVQs [39] rather than the jointly optimized multi-resolution vector quantizers ([25], [26]) may be very large. Proving such a result would require a tight bound on the rate losses studied in Theorem 6. In addition, the proof of this theorem implies that $L_M \leq \frac{1}{2}\log\left(1 + \frac{D_M}{D_{M-1}}\right) + \frac{M-1}{2}$. I redefine Y_M as $(1-\alpha)X + \alpha U_{M-1}' + N_M'$, where $\alpha = D_M/D_{M-1}$, $N_M' \sim \mathcal{N}(0, D_M - \alpha^2 D_{M-1})$, and $N_M' \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_M')$. Since $h(Y_M|X, U_{M-1}') = h(Y_M|X, U_{M-1}', U_{M-2}', \ldots, U_2', U_1)$,

$$\begin{aligned}
L_M &= I(X; Y_M, U_{M-1}', \ldots, U_2', U_1) - R(D_M) \\
&= I(X; Y_M | U_{M-1}', \ldots, U_2', U_1) + L_{M-1} + [R(D_{M-1}) - R(D_M)] \\
&\leq I(X; Y_M | U_{M-1}') + L_{M-1}
\end{aligned}$$

$$\leq \frac{1}{2}\log\frac{D_{M-1}}{D_M} + L_{M-1} \tag{3.1}$$
$$\leq \frac{1}{2}\log\frac{D_{M-1}}{D_M} + \frac{M-1}{2},$$

then

$$L_M \leq \frac{M-1}{2} + \frac{1}{2}\log\left(\frac{\sqrt{5}+1}{2}\right) < \frac{M}{2} - .1528.$$

I also find that

$$L_M \leq \frac{1}{2}\log\frac{D_1}{D_M}$$

from (3.1).

3.3 Proofs of Theorems

I first introduce a new property of rate-distortion functions, which is used in the proof of Theorem 2.

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

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

Theorem 2 *For any iid source $\{X_i\}_{i=1}^{\infty}$ with an mse distortion measure, there exists a 2RSC-achievable rate-distortion vector (R_1, R_2, D_1, D_2) with $L_2 = 0$ and $L_1 \leq \frac{1}{2}\log 5$ bps.*

Proof: Let U_1 and U_2 be the random variables that achieve $R(D_1)$ and $R(D_2)$, respectively. 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 $N_1 \perp\!\!\!\perp N_2$). I define

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

Here Y_1 and Y_2 satisfy the distortion constraints $Ed(X, Y_2) \leq D_2$ and $Ed(X, Y_1) \leq D_1$.

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 $L_2 = I(X; Y_2, Y_1) - R(D_2) = I(X; U_2, U_2 + N_1) - I(X; U_2) = 0$ since $X \to U_2 \to U_2 + N_1$ forms a Markov chain. The rate loss at the first stage is

$$\begin{aligned} L_1 &= I(X; Y_1) - I(X; U_1) = I(X; U_2 + N_1) - I(X; U_1) \\ &= [I(X; U_2 + N_1) - I(X; X + N_2 + N_1)] + [I(X; X + N_2 + N_1) - I(X; U_1)]. \quad (3.2) \end{aligned}$$

Let L_{1A} denote the first difference on the right hand side of (3.2); then

$$\begin{aligned} L_{1A} &= I(X; U_2 + N_1) - I(X; X + N_2 + N_1) \\ &\leq I(X; U_2 + N_1) - I(X; U_2) + I(X; X + N_2) - I(X; X + N_2 + N_1) \quad (3.3) \\ &= I(X; U_2 + N_1) - I(X; U_2, U_2 + N_1) \\ &\quad + I(X; X + N_2, X + N_2 + N_1) - I(X; X + N_2 + N_1) \quad (3.4) \\ &= I(X; X + N_2 | X + N_2 + N_1) - I(X; U_2 | U_2 + N_1) \\ &= I(N_2 + N_1; N_1 | X + N_2 + N_1) - I(U_2 + N_1 - X; N_1 | U_2 + N_1) \quad (3.5) \\ &= I(N_1; N_2 + N_1, X + N_2 + N_1) - I(N_1; X + N_2 + N_1) \\ &\quad - I(U_2 + N_1 - X; N_1 | U_2 + N_1) \\ &= I(N_1; N_2 + N_1) - I(N_1; X + N_2 + N_1) - I(U_2 + N_1 - X; N_1 | U_2 + N_1) \quad (3.6) \\ &= I(N_1; N_2 + N_1) - I(N_1; X + N_2 + N_1) \\ &\quad + I(N_1; U_2 + N_1) - I(N_1; U_2 + N_1 - X) - I(N_1; U_2 + N_1 | U_2 + N_1 - X), \end{aligned}$$

where (3.3) follows since $Ed(X, X + N_2) = D_2$ implies that $I(X; U_2) = R(D_2) \leq I(X; X + N_2)$, (3.4) 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, (3.5) follows since $h(A|B) = h(A - B|B)$, and (3.6) 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 $J = I(N_1; U_2 + N_1 | U_2 + N_1 - X) = I(U_2 - X; X | U_2 + N_1 - X)$ and $K = I(N_1; U_2 + N_1) - I(N_1; X + N_2 + N_1) - J$, then $L_{1A} \leq I(N_1; N_2 + N_1) - I(N_1; U_2 + N_1 - X) + K$, and by the chain rule,

$$
\begin{aligned}
K &= I(N_1; U_2 + N_1 | X + N_2 + N_1) - I(N_1; X + N_2 + N_1 | U_2 + N_1) - J \\
&\leq I(N_1; U_2 + N_1 | X + N_2 + N_1) - J \\
&= I(X + N_2; X - U_2 + N_2 | X + N_2 + N_1) - J \\
&\leq I(X + N_2; X - U_2 + N_2) - J \qquad (3.7) \\
&\leq I(N_2, X + N_2; X - U_2 + N_2) - J \\
&= I(N_2; X - U_2 + N_2) + I(X - U_2 + N_2; X + N_2 | N_2) - J \\
&= I(N_2; X - U_2 + N_2) + I(U_2 - X; X | N_2) - I(U_2 - X; X | U_2 - X + N_1) \\
&= I(N_2; X - U_2 + N_2) + I(U_2 - X; X) - I(U_2 - X; X | U_2 - X + N_1) \qquad (3.8) \\
&= I(N_2; X - U_2 + N_2) + I(U_2 - X; U_2 - X + N_1) \\
&\quad - I(U_2 - X; U_2 - X + N_1 | X) \qquad (3.9) \\
&\leq I(N_2; X - U_2 + N_2) + I(U_2 - X; U_2 - X + N_1),
\end{aligned}
$$

where (3.7) 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, (3.8) follows since $N_2 \perp\!\!\!\perp (X, U_2)$, and (3.9) follows by the chain rule since

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

Thus,

$$\begin{aligned} L_{1A} &\leq [h(N_2 + N_1) - h(N_2)] - [h(U_2 + N_1 - X) - h(U_2 - X)] \\ &\quad + [h(X - U_2 + N_2) - h(X - U_2)] + [h(U_2 - X + N_1) - h(N_1)] \\ &= h(N_2 + N_1) - h(N_2) - h(N_1) + h(X - U_2 + N_2) \\ &= \frac{1}{2} \log \frac{D_1}{2\pi e D_2 (D_1 - D_2)} + h(X - U_2 + N_2) \\ &\leq \frac{1}{2} \log \frac{D_1}{2\pi e D_2 (D_1 - D_2)} + \frac{1}{2} \log(4\pi e D_2) \qquad (3.10) \\ &= \frac{1}{2} \log \frac{2D_1}{D_1 - D_2}, \end{aligned}$$

where (3.10) follows since $E(X - U_2 + N_2)^2 \leq 2D_2$, so

$$h(X - U_2 + N_2) \leq h(\mathcal{N}(0, 2D_2)) = \frac{1}{2} \log(4\pi e D_2).$$

I denote the second difference on the right hand side of (3.2) by L_{1B}, which is bounded by 0.5. The proof parallels that of [32, Theorem 3]. In particular,

$$\begin{aligned} L_{1B} &= I(X; X + N_2 + N_1) - I(X; U_1) \\ &= I(X; X + N_2 + N_1 | U_1) - I(X; U_1 | X + N_2 + N_1) \\ &\leq I(X; X + N_2 + N_1 | U_1) \\ &= I(X - U_1; X - U_1 + N_2 + N_1 | U_1) \\ &\leq I(X - U_1; X - U_1 + N_2 + N_1) \\ &\leq 0.5. \end{aligned}$$

Thus, I can bound L_1 as

$$L_1 = L_{1A} + L_{1B} \leq \frac{1}{2}\log\frac{2D_1}{D_1 - D_2} + \frac{1}{2} = 1 + \frac{1}{2}\log\frac{D_1}{D_1 - D_2}.$$

On the other hand, the rate loss at the first stage can also be bounded as:

$$L_1 = I(X;Y_1) - R(D_1) \leq I(X;Y_2,Y_1) - R(D_1) \leq \frac{1}{2}\log\frac{D_1}{D_2},$$

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

Putting these bounds for L_1 together gives

$$L_1 \leq \min\left\{\frac{1}{2}\log\frac{D_1}{D_2}, 1 + \frac{1}{2}\log\frac{D_1}{D_1 - D_2}\right\}.$$

The shaded region of Figure 3.5 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 $L_1 \leq \frac{1}{2}\log 5 < 1.1610$. □

Theorem 3 *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* $L_2 \leq \frac{1}{2}\log(\sqrt{3}+1)$.

Proof: [32, Theorem 5] actually shows that

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

I here derive a second bound that decreases in D_2/D_1. Together, the two bounds give the desired result. 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 $N_2 \perp\!\!\!\perp (X, U_1, U_2)$ as in Lemma 2 and defining $Y_2 = (1-\alpha)X + \alpha U_1 + N_2$ and $Y_1 = U_1$ gives $L_1 = 0$ and

$$\begin{aligned} L_2 &= I(X;Y_2,Y_1) - R(D_2) = I(X;Y_2,U_1) - R(D_2) \\ &= I(X;Y_2|U_1) + I(X;U_1) - R(D_2) \leq I(X;Y_2|U_1) \qquad (3.11) \\ &\leq \frac{1}{2}\log\frac{D_1}{D_2}. \qquad (3.12) \end{aligned}$$

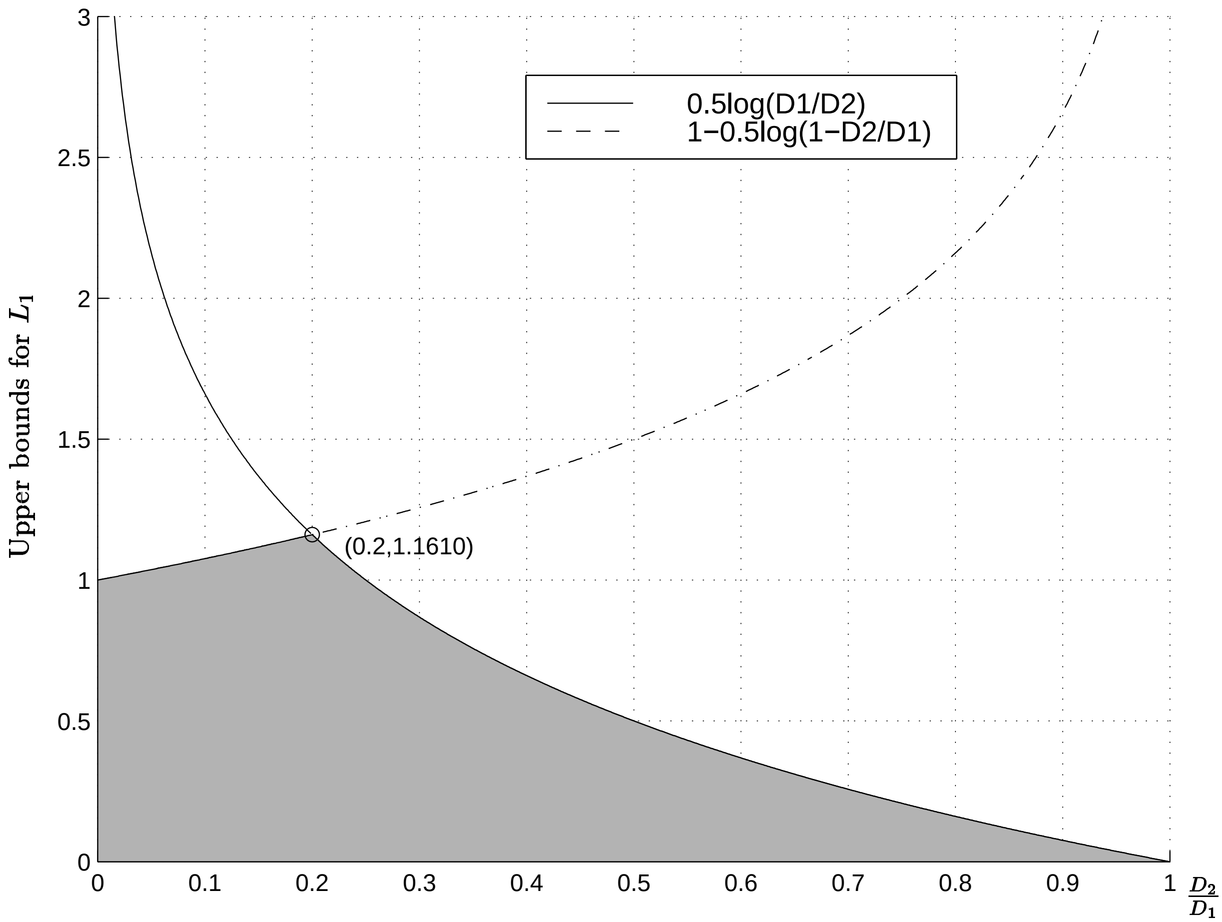

Figure 3.5: Possible values of L_1 in Theorem 2.

Here (3.11) follows since $I(X;U_1) = R(D_1)$, $D_2 < D_1$ and the rate-distortion function $R(D)$ is a non-increasing function of D, and (3.12) follows from steps (8.2)-(8.5) of the proof of Lemma 2.

Thus for any $D_2/D_1 < 1$, I have two upper bounds for L_2:

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

The first bound dominates when $D_1/D_2 \geq \sqrt{3}+1$ while the second dominates when $1 < D_1/D_2 < \sqrt{3}+1$ (See Figure 3.6). Together these bounds give $L_2 \leq (1/2)\log(\sqrt{3}+1) <$

0.7250. □

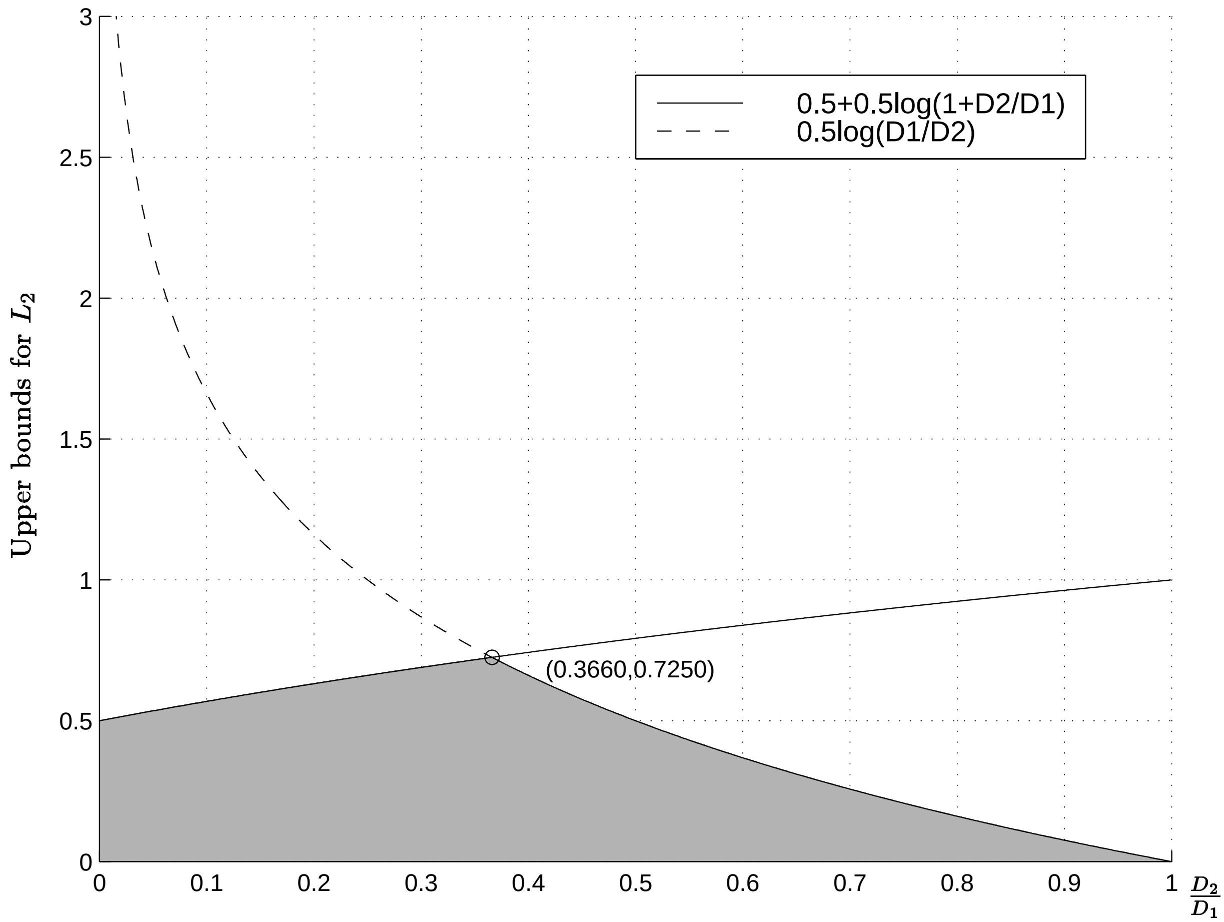

Figure 3.6: Possible values of L_2 in Theorem 3.

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

Proof: From Theorem 3,

$$L_2 \leq \frac{1}{2}\log\frac{D_1}{D_2}.$$

For any fixed $0 < D_1 \leq \sigma^2$, this bound is a decreasing function of $D_2 \in (0, D_1)$. I next derive another 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 [32, Theorem 5] gives

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

The second bound dominates when $D_2 \in \left(0, \frac{D_1\sigma^2\sqrt{6\sigma^2-5D_1}-D_1\sigma^2\sqrt{2\sigma^2-D_1}}{2(\sigma^2-D_1)\sqrt{2\sigma^2-D_1}}\right]$, while the first bound dominates for the remainder of the $(0, D_1)$ region. The maximal value of the combined bound is achieved at $D_2 = \frac{D_1\sigma^2\sqrt{6\sigma^2-5D_1}-D_1\sigma^2\sqrt{2\sigma^2-D_1}}{2(\sigma^2-D_1)\sqrt{2\sigma^2-D_1}}$, giving the desired result. □

Theorem 5 *For any $D_2 < D_1$, the rate losses $L_1 = L_2 < 0.3802$ are achievable.*

Proof: As shown in Theorem 2, the rate losses L_{10} and L_{20} are achievable, where

$$L_{20} = 0 \quad \text{and} \quad L_{10} \leq \begin{cases} A_1 = 1 + \frac{1}{2}\log\frac{D_1}{D_1-D_2}, & \text{if } D_1 \geq 5D_2 \\ A_2 = \frac{1}{2}\log\frac{D_1}{D_2}, & \text{if } D_1 < 5D_2. \end{cases}$$

From Theorem 3, the rate losses L_{11} and L_{12} are achievable, where

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

Thus, from Lemma 1, I 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

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

Figure 3.7 shows this bound for $L_{1\alpha} = L_{2\alpha}$. The maximum occurs when $D_2/D_1 = 0.2$, giving

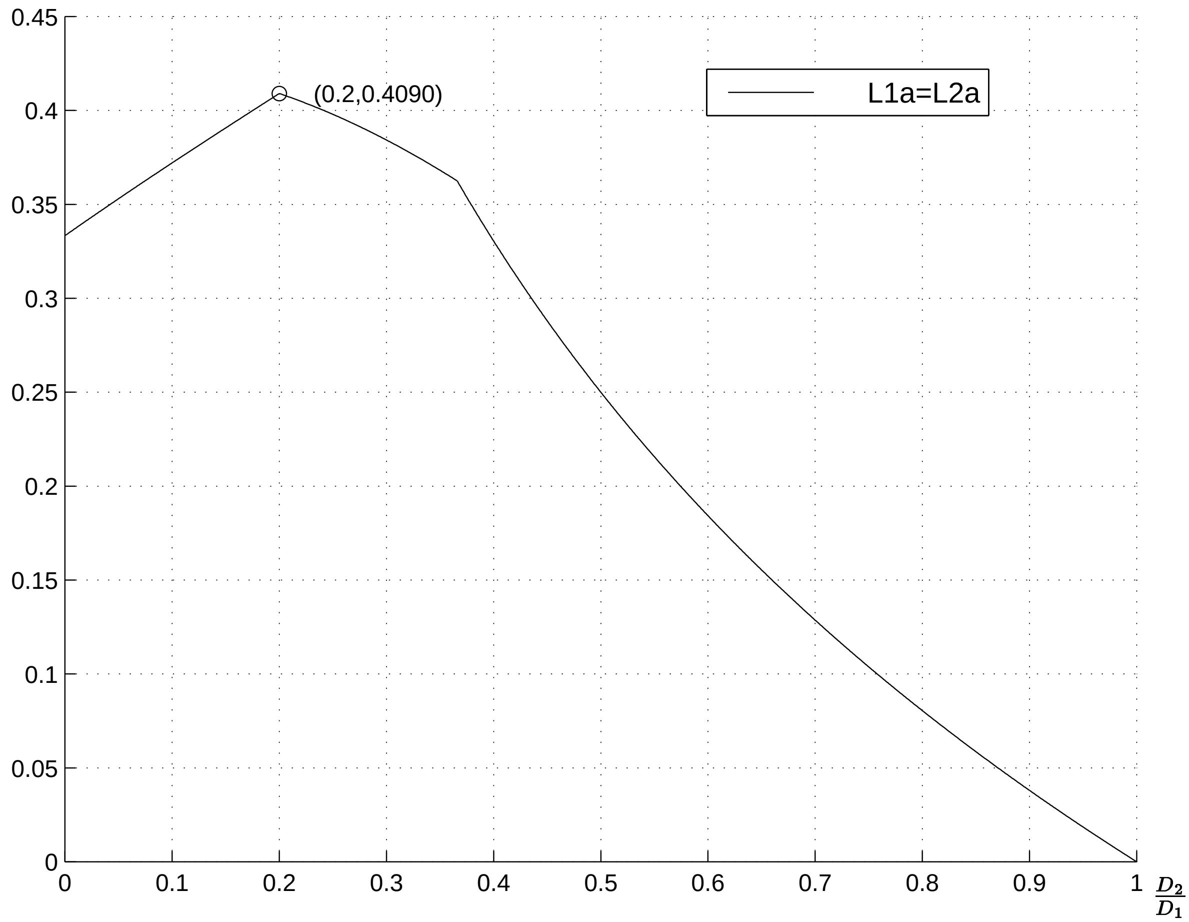

Figure 3.7: Upper bound I for $L_{1\alpha} = L_{2\alpha}$ in Theorem 5.

$$L_{1\alpha} = L_{2\alpha} \leq \frac{1}{1/(\frac{1}{2}\log 5) + 1/(\frac{1}{2}\log(1+0.2)+\frac{1}{2})} < .4091.$$

Finally, I use the results of Theorems 2 and 3 in a different way to get the desired result. By Theorem 2, $L_{10} \leq R(D_2) - R(D_1)$ and $L_{20} = 0$ are achievable. By Theorem 3, $L_{11} = 0$ and $L_{21} \leq \frac{1}{2}\log\frac{D_1}{D_2} + R(D_1) - R(D_2)$ are achievable. Thus by convexity, I can achieve

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

Figure 3.8 combines this new bound with the bound of Figure 3.7. 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. □

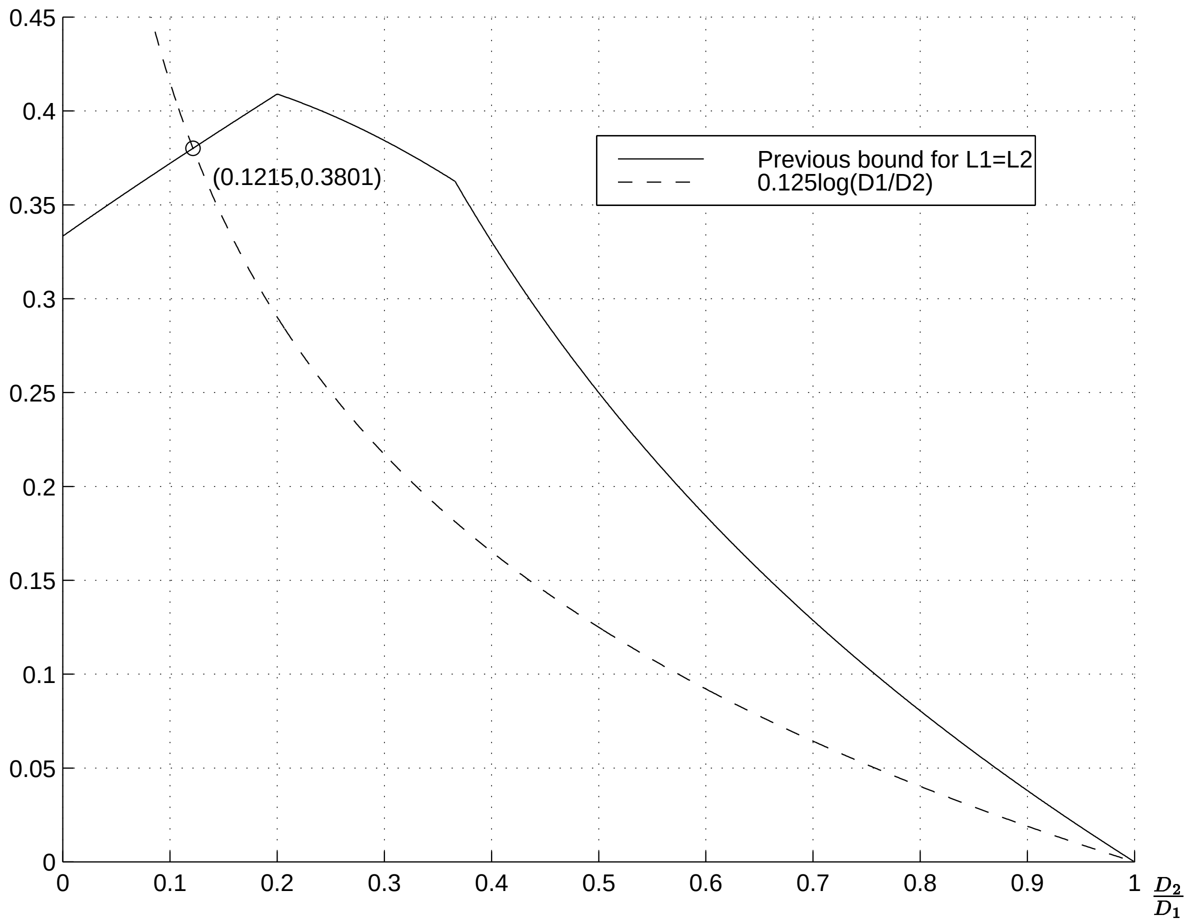

Figure 3.8: Upper bound II for $L_{1\alpha} = L_{2\alpha}$ in Theorem 5.

I first introduce Lemma 3, which is useful for proving this theorem.

Lemma 3 *For any $D_M < \ldots < D_3 < D_2 < D_1$,*

$$I(X; U_1, U_2', U_3', \ldots, U_M' | X + N_M) \leq M/2,$$

where $U_1, U_2', U_3', \ldots, U_{M-1}'$ are defined in Theorem 6.

Theorem 6 *For any $D_M < \ldots < D_2 < D_1$, there exists an achievable rate-distortion vector $(R_1, R_2, \ldots, R_M, D_1, D_2, \ldots, D_M)$ with $L_1 = 0$, $L_i = I(X; U_1, U_2', U_3', \ldots, U_i') - R(D_i)$ for all $1 < i < M$, and $L_M \leq M/2$, where U_1 is a random variable achieving $R(D_1)$, $U_2', U_3', \ldots, U_M'$ are successively defined as* $\arg\min_{U:Ed(X,U)\leq D_i} I(X; U, U_1, U_2', U_3', \ldots, U_{i-1}')$, *$N_M \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_M')$, and $N_M \sim \mathcal{N}(0, D_M)$.*

Proof: Let U_i be the random variable that achieves $R(D_i)$ for any $i \in \{1, 2, \ldots, M\}$, $N_{M-1} \sim \mathcal{N}(0, D_{M-1})$, and $N_{M-1} \perp\!\!\!\perp (X, U_1, U_2', U_3', \ldots, U_{M-1}', U_M)$. Further I define the joint distribution of $(X, U_1, U_2', U_3', \ldots, U_{M-1}', U_M)$ as $f_X Q_{U_1, U_2', U_3', \ldots, U_{M-1}'|X} Q_{U_M|X}$, therefore, $(U_1, U_2', U_3', \ldots, U_{M-1}') \to X \to U_M$ forms a Markov chain.

I define $Y_1 = U_1$, $Y_M = U_M$, and $Y_i = U_i'$ for any $i \in \{2, 3, \ldots, M-1\}$. Then the rate loss of this code at resolution M is

$$\begin{aligned} L_M &= I(X; Y_M, \ldots, Y_2, Y_1) - I(X; U_M) \\ &= I(X; U_M, U_{M-1}', \ldots, U_2', U_1) - I(X; U_M) \\ &= I(X; U_{M-1}', \ldots, U_2', U_1 | U_M) \\ &\leq I(X; U_{M-1}', \ldots, U_2', U_1, X + N_{M-1} | U_M) \\ &= I(X; X + N_{M-1} | U_M) + I(X; U_{M-1}', \ldots, U_2', U_1 | U_M, X + N_{M-1}). \end{aligned} \tag{3.13}$$

Here,

$$I(X; X + N_{M-1} | U_M) \leq \frac{1}{2} \log\left(1 + \frac{D_M}{D_{M-1}}\right) \leq \frac{1}{2}, \tag{3.14}$$

where the first inequality follows the approach of steps (8.2)-(8.5) from the Appendix, and the second inequality follows since $D_M < D_{M-1}$. Further,

$$I(X; U_{M-1}', \ldots, U_2', U_1 | U_M, X + N_{M-1})$$

$$
\begin{aligned}
&= h(U'_{M-1},\ldots,U'_2,U_1|U_M,X+N_{M-1}) \\
&\quad -h(U'_{M-1},\ldots,U'_2,U_1|U_M,X+N_{M-1},X) \\
&\le h(U'_{M-1},\ldots,U'_2,U_1|X+N_{M-1}) - h(U'_{M-1},\ldots,U'_2,U_1|U_M,X+N_{M-1},X) \quad (3.15) \\
&= h(U'_{M-1},\ldots,U'_2,U_1|X+N_{M-1}) - h(U'_{M-1},\ldots,U'_2,U_1|X+N_{M-1},X) \quad (3.16) \\
&= I(X;U'_{M-1},\ldots,U'_2,U_1|X+N_{M-1}) \\
&\le (M-1)/2, \quad (3.17)
\end{aligned}
$$

where (3.15) follows since conditioning reduces entropy, (3.16) follows since given $(X, X + N_{M-1})$, U_M and $(U_1, U'_2, U'_3, \ldots, U'_{M_1})$ are independent, and (3.17) follows from Lemma 3. Combining (3.13), (3.14) and (3.17) proves the theorem. □

Chapter 4

Multiple Description Source Codes

4.1 Preliminaries

An (n, M_1, M_2) 2DSC consists of two individual encoder/decoder pairs and a joint decoder. For packet 1, the encoder/decoder pair $(f_n^{(1)}, g_n^{(1)})$ are mappings

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

with rate $(1/n) \log M_1$ and distortion $Ed_n(X^n, g_n^{(1)}(f_n^{(1)}(X^n)))$. For packet 2, the encoder and decoder $(f_n^{(2)}, g_n^{(2)})$ are mappings

$$f_n^{(2)} : \mathbf{R}^n \to \{1, \ldots, M_2\} \text{ and } g_n^{(2)} : \{1, \ldots, M_2\} \to \mathbf{R}^n$$

with rate $(1/n) \log M_2$ and distortion $Ed_n(X^n, g_n^{(2)}(f_n^{(2)}(X^n)))$. The joint decoder $g_n^{(0)}$ is a mapping

$$g_n^{(0)} : \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \to \mathbf{R}^n$$

with total rate $(1/n) \log(M_1 M_2)$ and distortion $Ed_n(X^n, g_n^{(0)}(f_n^{(1)}(X^n), f_n^{(2)}(X^n)))$.

We say that the rate-distortion vector $(R_1, R_2, D_0, D_1, D_2)$ is 2DSC-achievable if for any $\epsilon > 0$ and for sufficiently large n, there exists an (n, M_1, M_2) 2DSC such that

$$\begin{aligned} \tfrac{1}{n}\log M_1 \le R_1 + \epsilon, &\quad Ed_n(X^n, g_n^{(1)}(f_n^{(1)}(X^n))) \le D_1 + \epsilon, \\ \tfrac{1}{n}\log M_2 \le R_2 + \epsilon, &\quad Ed_n(X^n, g_n^{(2)}(f_n^{(2)}(X^n))) \le D_2 + \epsilon, \\ &\quad Ed_n(X^n, g_n^{(0)}(f_n^{(1)}(X^n), f_n^{(2)}(X^n))) \le D_0 + \epsilon. \end{aligned}$$

A set of 2DSC-achievable rate-distortion vectors is described in the following theorem. The result for finite alphabets comes from [37]. A generalization to well-behaved continuous sources appears in [40]. The result is not tight in general [41]. A more general (but not single-letter) form for stationary sources appears in [38].

Theorem 7 *[37, Theorem 1] [40, Theorem 1] For any iid source $\{X_i\}_{i=1}^{\infty}$ with density $f_X(x)$ and distortion measure d, $(R_1, R_2, D_0, D_1, D_2)$ is 2DSC-achievable if there exists a conditional probability $Q_{Y_0,Y_1,Y_2|X}$ such that*

$$\begin{aligned} R_1 &\ge I(X;Y_1), & Ed(X,Y_1) &\le D_1, \\ R_2 &\ge I(X;Y_2), & Ed(X,Y_2) &\le D_2, \\ R_1 + R_2 &\ge I(X;Y_0,Y_1,Y_2) + I(Y_1;Y_2), & Ed(X,Y_0) &\le D_0. \end{aligned}$$

For any 2DSC, if the rate-distortion vector $(R_1, R_2, D_0, D_1, D_2)$ is achievable, the individual rate loss of description i $(i \in \{1,2\})$ is defined as $L_i = R_i - R(D_i)$ bps. The joint rate loss is measured in two different ways by $L_0 = R_1 + R_2 - R(D_0)$ bps and $L_{12} = R_1 + R_2 - [R(D_1) + R(D_2)] = L_1 + L_2$ bps. I denote the normalized distortions as $d_1 = D_1/\sigma^2$, $d_2 = D_2/\sigma^2$, and $d_0 = D_0/\sigma^2$. I assume $0 < D_0 < D_1, D_2 \le \sigma^2$, i.e., $0 < d_0 < d_1, d_2 \le 1$.

4.2 Major results

4.2.1 Partition of the distortion region

For each (D_0, D_1, D_2), define

$$\mathcal{R}(D_0, D_1, D_2) = \{(R_1, R_2) : (R_1, R_2, D_0, D_1, D_2) \text{ is 2DSC-achievable}\}.$$

The following lemma is useful in partitioning the achievable region.

Lemma 4 *If* $0 < D_1, D_2 \leq \sigma^2$, *then*

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

I partition the space of possible distortion vectors (D_0, D_1, D_2) into 3 regions:

$$\begin{aligned}
\mathcal{D}_1 &= \{(D_0, D_1, D_2) : 0 \leq D_0 \leq D_1 + D_2 - \sigma^2\} \\
\mathcal{D}_2 &= \{(D_0, D_1, D_2) : D_1 + D_2 - \sigma^2 < D_0 < (1/D_1 + 1/D_2 - 1/\sigma^2)^{-1}\} \\
\mathcal{D}_3 &= \{(D_0, D_1, D_2) : D_0 \geq (1/D_1 + 1/D_2 - 1/\sigma^2)^{-1}\}.
\end{aligned}$$

Thus $\mathcal{D}_1$, $\mathcal{D}_2$, and $\mathcal{D}_3$ correspond to the low-D_0, medium-D_0, and high-D_0 regions, respectively.

In [42], Ozarow shows that Theorem 7 is tight for any Gaussian source. From [17], the achievable region for a Gaussian source is

$$\begin{cases} R_1 \geq R(D_1) \\ R_2 \geq R(D_2) \\ R_1 + R_2 \geq R(D_0) \end{cases} \tag{4.1}$$

for $\mathcal{D}_1$;

$$\begin{cases} R_1 \geq R(D_1) \\ R_2 \geq R(D_2) \\ R_1 + R_2 \geq R(D_0) + L_{G0} \end{cases}$$

for $\mathcal{D}_2$; and

$$\begin{cases} R_1 \geq R(D_1) \\ R_2 \geq R(D_2) \end{cases} \tag{4.2}$$

for $\mathcal{D}_3$, where

$$R(D) = \frac{1}{2} \log \frac{\sigma^2}{D}$$

and L_{G0} denotes the joint rate loss (L_0) of a Gaussian source with $(D_0, D_1, D_2) \in \mathcal{D}_2$. By [37],

$$L_{G0} = \frac{1}{2} \log \frac{(1-d_0)^2}{(1-d_0)^2 - (\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)})^2}.$$

The following results give bounds on the values of $\mathcal{R}(D_0, D_1, D_2)$ for the three regions. By symmetry, each statement about rate loss L_1 can also be made to apply to rate loss L_2.

4.2.2 The Low-D_0 Region

I define a function

$$L_{\min}(d_1, d_2) = \min\left\{\frac{1}{2}\log\frac{1}{d_1}, \frac{1}{2}\log\frac{1}{d_2}\right\} = \frac{1}{2}\log\frac{\sigma^2}{\max\{D_1, D_2\}}. \tag{4.3}$$

Theorem 8 *For any $(D_0, D_1, D_2) \in \mathcal{D}_1 \bigcup \mathcal{D}_2$, there exists an $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ with $L_1 = R_1 - R(D_1) \leq 0.5 \log(2 - d_1) \leq 0.5$ and*

$$L_0 \quad = \quad R_1 + R_2 - R(D_0)$$

$$\leq 0.5\log(2-d_0)+\min\{L_{\min}(d_1,d_2),R(\max\{D_1,D_2\})+0.5\log(2-\max\{d_1,d_2\})$$

$$\leq \min\{L_{\min}(d_1,d_2)+0.5,R(\max\{D_1,D_2\})+1\}.$$

I apply this theorem to obtain more specific results for regions $\mathcal{D}_1$ and $\mathcal{D}_2$ separately, which are listed in the following corollaries.

Corollary 1 *For any* $(D_0,D_1,D_2)\in\mathcal{D}_1$*, there exists an* $(R_1,R_2)\in\mathcal{R}(D_0,D_1,D_2)$ *with* $L_1\leq 0.5\log(2-d_1)\leq 0.5$ *and* $L_0\leq 0.5\log[2(2-d_0)/(1+d_0)]\leq 1$.

The bound for L_0, which depends only on d_0, is tight when $D_0=\sigma^2$. Figure 4.1 plots this bound for L_0.

If X is Gaussian, then (4.1) describes $\mathcal{R}(D_0,D_1,D_2)$ for any $(D_0,D_1,D_2)\in\mathcal{D}_1$ [37]. Corollary 1 gives universal meaning to the definition of $\mathcal{D}_1$ for all iid sources: if $D_1+D_2-D_0\geq\sigma^2$, then

$$\begin{cases} R_1\geq R(D_1)+0.5 \\ R_2\geq R(D_2)+0.5 \\ R_1+R_2\geq R(D_0)+1 \end{cases} \tag{4.4}$$

is an inner bound on $\mathcal{R}(D_0,D_1,D_2)$, i.e., the rates can be "close" to their rate-distortion bounds in $\mathcal{D}_1$.[1] Figure 4.2 illustrates this property. The dashed lines trace the outer bound from (4.1); any rate pair below this bound is not MDSC-achievable. The shaded region surrounded by solid lines is the inner bound described in (4.4); any rate pair inside this region is MDSC-achievable. For any $(\breve{R}_1,\breve{R}_2)$ on the outer bound characterized by (4.1), the rate pair $(\breve{R}_1+0.5,\breve{R}_2+0.5)$ is achievable.

[1] In [43], Ahlswede shows that El Gamal and Cover's inner bound (see Theorem 7) is tight in the case of no excess rate at D_0, i.e., $R_1+R_2=R(D_0)$. However, this region may be difficult to evaluate in general.

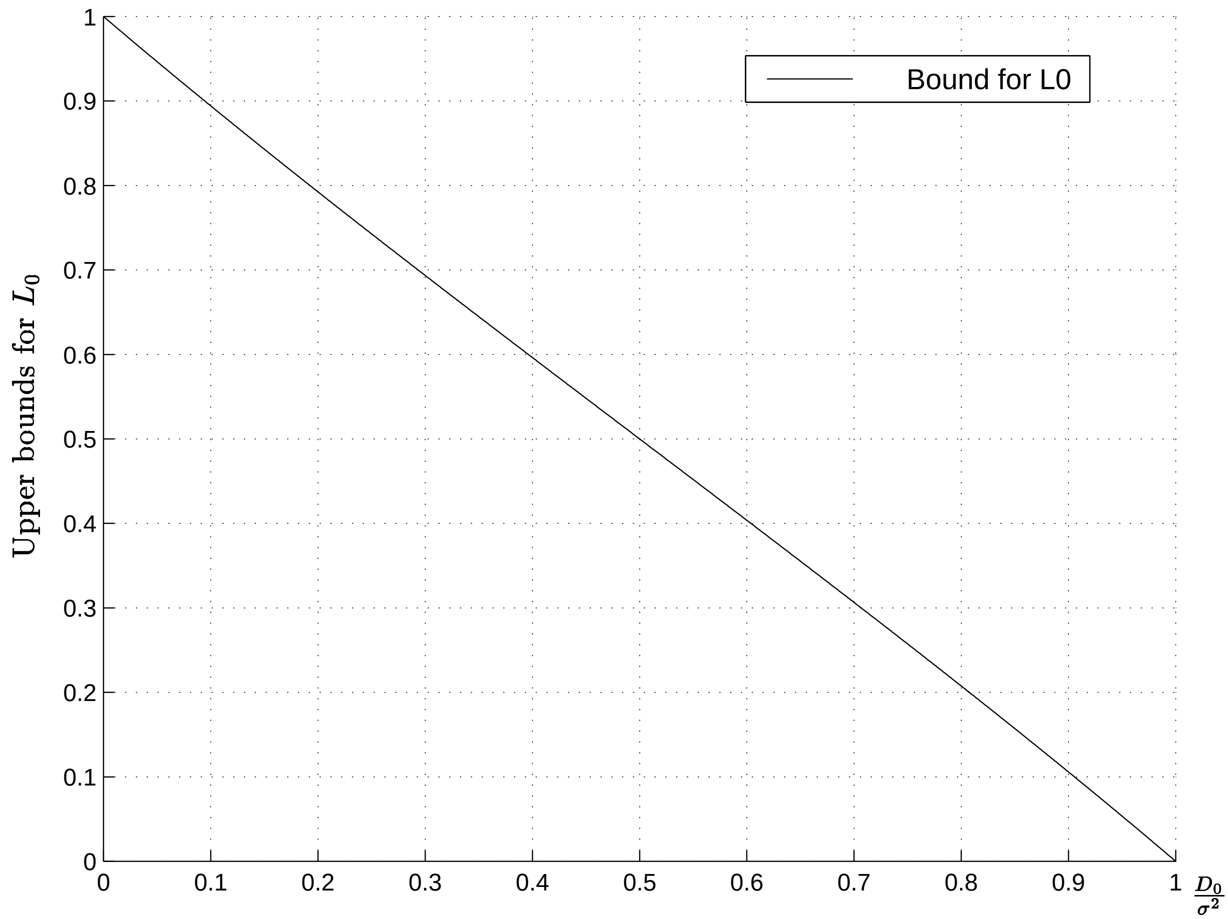

Figure 4.1: Bound for L_0 in Corollary 1.

In the special case where $D_2 = \sigma^2$, I can tighten this result to $L_0 = L_1 < 0.3802$, or $L_0 = 0$ and $L_1 < 1.1610$, or $L_0 < 0.7250$ and $L_1 = 0$, as shown in [15].

4.2.3 The Medium-D_0 Region

Theorem 8 also leads to a new inner bound on the achievable region in $\mathcal{D}_2$.

Corollary 2 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$ *there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5$ *and* $L_0 \leq L_{G0} + 1.5$.

I can get a tighter bound in part of $\mathcal{D}_2$ immediately from Theorem 8.

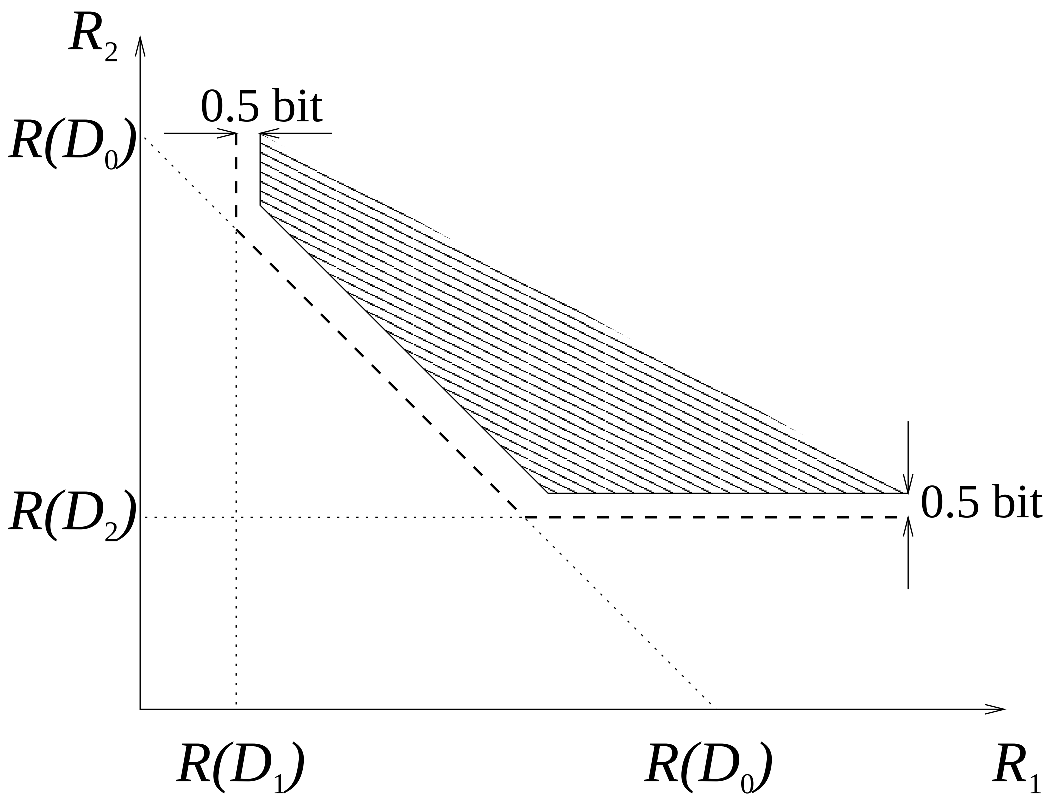

Figure 4.2: Outer and inner bounds on $\mathcal{R}(D_0, D_1, D_2)$ for $(D_0, D_1, D_2) \in \mathcal{D}_1$.

Corollary 3 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$ *with* $d_1 \geq 0.5$ *or* $d_2 \geq 0.5$*, there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5$ *and* $L_0 \leq 1$.

In $\mathcal{D}_2$, the joint rate loss of the Gaussian source is nonzero, but in some sub-regions, e.g., $\max(d_1, d_2) \geq 0.5$, it is still small (e.g., see Corollary 5 in Section 4.3). The above corollary generalizes this property to all iid sources.

In fact, I can also compare $R_1 + R_2$ to $R(D_1) + R(D_2)$ instead of $R(D_0)$, though both of them are natural outer bounds. I define a new joint rate loss as $L_{12} = R_1 + R_2 - [R(D_1) + R(D_2)]$. The relation between L_0 and L_{12} is

$$L_0 + R(D_0) = L_{12} + R(D_1) + R(D_2), \tag{4.5}$$

Corollary 4 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$*, there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5$ *and* $L_{12} \leq 0.5\log[\min\{D_1, D_2\}/D_0] + 1$.

To prove the preceding results, I use reconstructions Y_0, Y_1, and Y_2 for which $X \to Y_0 \to (Y_1, Y_2)$. I then turn to a different approach, which is similar to that used in [44]; using reconstructions that satisfy $X \to (Y_1, Y_2) \to Y_0$, leads to a tighter bound in terms of L_{G12} for L_{12} in $\mathcal{D}_2$, where L_{G12} is defined in [44] as

$$L_{G12} = \frac{1}{2}\log\frac{(1-d_0)^2 d_1 d_2}{(1-d_0)^2 d_1 d_2 - (d_0\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)})^2}.$$

Lemma 5 describes the relation between L_{G12} and L_{G0}, which also implies that L_{G12} equals L_{12} for the Gaussian source.

Lemma 5 *In* $\mathcal{D}_2$, $L_{G12} = 0.5\log(d_1 d_2) - 0.5\log(d_0) + L_{G0}$.

Theorem 9 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$*, there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5$ *and* $L_{12} \leq L_{G12} + 1$.

From the definition of L_{G12} and the proof of Corollary 2, I can see that Theorem 9 is tighter than both Corollary 4 and another bound

$$L_{12} \leq 1 + \frac{1}{2}\log\frac{d_1 d_2}{d_0} + \frac{1}{2}\log\frac{1}{d_1 + d_2 - d_1 d_2}. \tag{4.6}$$

Corollary 4 and the bound in (4.6) are easier to analyze. The difference between Theorem 9 and Corollary 4 or (4.6) is less than 1 bit.

4.2.4 The High-D_0 Region

Finally, I turn to $\mathcal{D}_3$.

Theorem 10 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_3$*, there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5 \log(2 - d_1) \leq 0.5$ *and* $L_2 \leq 0.5 \log(2 - d_2) \leq 0.5$*.*

We know that if we force the side rates to be low, i.e., $R_1 = R(D_1)$ and $R_2 = R(D_2)$,[2] we can certainly get a reproduction at the joint decoder with distortion $D_0 = \min\{D_1, D_2\}$. Intuitively, we should be able to do better than that and achieve distortion $D_0 \leq \min\{D_1, D_2\}$. This theorem justifies this intuition in some sense, telling us that for any source and any fixed D_1 and D_2, there always exists a multiple description code with $R_1 \approx R(D_1)$, $R_2 \approx R(D_2)$, and D_0 no more than $(1/D_1 + 1/D_2 - 1/\sigma^2)^{-1}$. Figure 4.3 illustrates $\mathcal{R}(D_0, D_1, D_2)$ for $(D_0, D_1, D_2) \in \mathcal{D}_3$. The dashed lines depict the outer bound characterized by (4.2); the shaded region surrounded by solid lines is the inner bound given by $R_1 \geq R(D_1) + 0.5$ and $R_2 \geq R(D_2) + 0.5$.

4.2.5 Tightness of the Bounds for the Medium-D_0 Region

In $\mathcal{D}_1$ and $\mathcal{D}_3$, I give constant upper bounds on the rate loss. In $\mathcal{D}_2$, such constant upper bounds do not exist in general (e.g., the rate loss of a Gaussian source can be arbitrarily large). I therefore rely on the lower bounds which can be derived from [44]. In the following theorems, I bound the distance between the upper bounds and lower bounds on the rate loss. I focus on region $\mathcal{D}_2$ with $\max\{D_1, D_2\} < \sigma^2/2$, since this is the only region where I have not found a constant bound.

Theorem 11 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$ *with* $\max\{D_1, D_2\} < \sigma^2/2$*, there exists a rate pair* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5 \log(2 - d_1) \leq 0.5$*, and the distance between the*

[2] This is also known as the no excess marginal rate case. Inner and outer bounds, which may be hard to compute in general, can be found in [45].

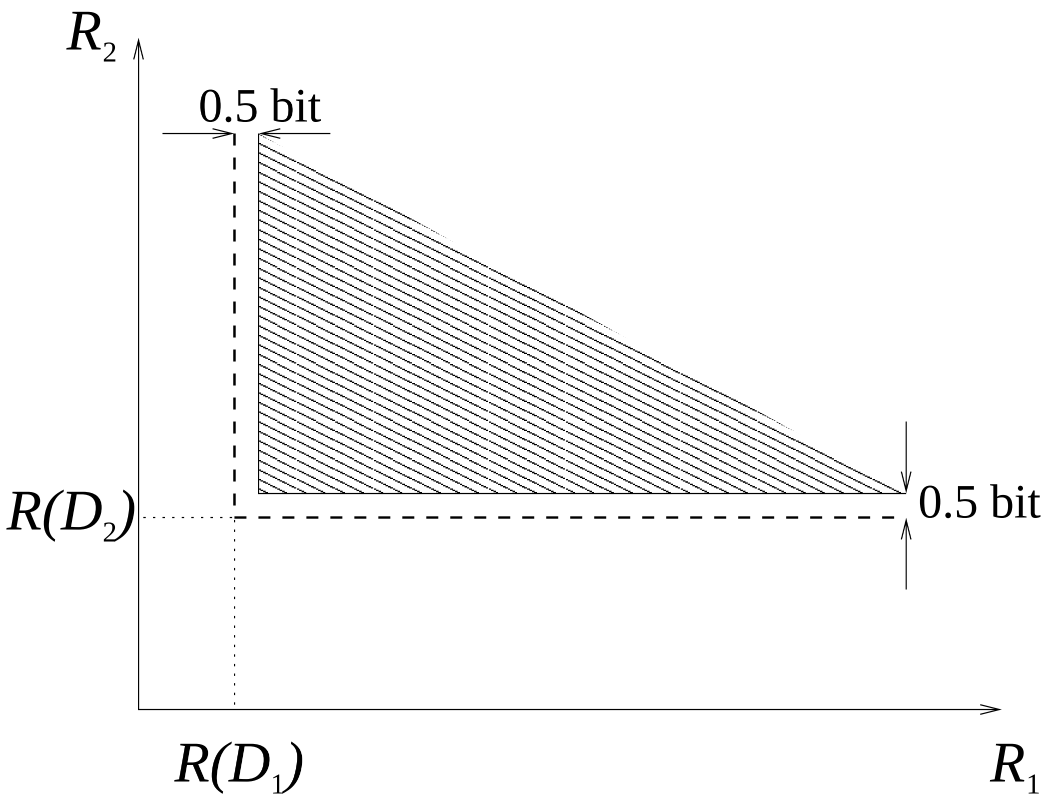

Figure 4.3: Outer and inner bounds on $\mathcal{R}(D_0, D_1, D_2)$ for $(D_0, D_1, D_2) \in \mathcal{D}_3$.

upper bound and the lower bound for L_0 *is less than or equal to* $\min\{[\log(2\pi e\sigma^2) - 2h(X)] + 1, 0.5\log(2\pi e\sigma^2) - h(X) + 1.5\}$.

This bound is relatively tight for sources whose differential entropy is close to that of Gaussian source with the same variance.

Constant bounds are possible in some special case. For example, if the Shannon lower bound (SLB) is tight at distortion D_1 and D_2, i.e.,

$$\begin{aligned} R(D_1) &= h(X) - \frac{1}{2}\log(2\pi e D_1), \\ R(D_2) &= h(X) - \frac{1}{2}\log(2\pi e D_2), \end{aligned}$$

then I can show the following theorem.

Theorem 12 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$ *with* $\max\{D_1, D_2\} < \sigma^2/2$, *if the SLB is tight at distortion* D_1 *and* D_2, *then there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5\log(2 - d_1) \leq 0.5$, *and the distance between the upper bound and the lower bound for* L_0 *is less than or equal to 2.*

It can be shown that for smooth sources, if the SLB is not tight for $R(D)$ at any distortion $D > 0$, it is asymptotically tight as $D \to 0$. So Theorem 12 generally works for the high-resolution scenario, i.e., $D_1, D_2 \to 0$ and D_1/D_0 and D_2/D_0 are held fixed. However, for this specific scenario, a better result has been shown in [44]: if the source is smooth and $D_1, D_2 \to 0$ and D_1/D_0 and D_2/D_0 are held fixed, then the upper bound and the lower bound for L_0 coincide asymptotically, i.e., the distance between the upper and lower bounds approaches 0 in the high-resolution limit.

4.2.6 Properties of the Rate-Distortion Function

The above results yield as a by-product some interesting inequalities about the rate-distortion function of an arbitrary iid source:

$$R(D_1) + R(D_2) \leq R(D_0) + 1, \text{ if } (D_0, D_1, D_2) \in \mathcal{D}_1; \quad (4.7)$$

$$R(D_0) - L_{G12} - 1 \leq R(D_1) + R(D_2) \leq R(D_0) + L_{G0} + 1.5, \text{ if } (D_0, D_1, D_2) \in \mathcal{D}_2; \quad (4.8)$$

$$R(D_1) + R(D_2) \leq R(D_0) + 1, \text{ if } (D_0, D_1, D_2) \in \mathcal{D}_2 \text{ and } \max\{D_1, D_2\} \geq \sigma^2/2; \quad (4.9)$$

$$R(D_1) + R(D_2) \geq R(D_0) - 1, \text{ if } (D_0, D_1, D_2) \in \mathcal{D}_3. \quad (4.10)$$

Here (4.7) comes from Theorem 8; (4.8) comes from Corollary 2 and Theorem 9; (4.9) comes from Corollary 3; and (4.10) comes from Theorem 10. In fact, tighter versions of (4.7) and

(4.9) can be drawn immediately from the properties of the rate-distortion function (e.g., convexity), as shown in Lemmas 6 and 7. The other inequalities are less obvious.

Lemma 6 *Suppose $R(D)$ is the rate-distortion function of an arbitrary iid source and $0 < D_0 < D_1, D_2 \le \sigma^2$, then*

$$R(D_1) + R(D_2) \le R(D_0), \quad \textit{if } D_1 + D_2 - D_0 \ge \sigma^2.$$

Lemma 7 *Suppose $R(D)$ is the rate-distortion function of an arbitrary iid source and $0 < D_0 < D_1, D_2 \le \sigma^2$, then*

$$R(D_1) + R(D_2) \le R(D_0) + 0.5,$$

if $D_1 + D_2 - \sigma^2 < D_0 < (1/D_1 + 1/D_2 - 1/\sigma^2)^{-1}$ and $\max\{D_1, D_2\} \ge \sigma^2/2$.

4.3 Proofs of Theorems

Generally speaking, the proofs again involve finding a Gaussian approximation of the optimizing reproduction distribution and bounding the optimal rate loss by the rate loss of the Gaussian.

I define $\beta_i = 1 - d_i$ for $i = 0, 1, 2$.

I define $\sigma_1^2 = \beta_1 D_1 - \beta_1^2 D_0/\beta_0$, $\sigma_2^2 = \beta_2 D_2 - \beta_2^2 D_0/\beta_0$, and $\Gamma = -\beta_1\beta_2 D_0/\beta_0$. The following Lemma is useful in the proof of Theorem 8.

Lemma 8 *For any $(D_0, D_1, D_2) \in \mathcal{D}_1 \bigcup \mathcal{D}_2$, $\Gamma^2 \le \sigma_1^2\sigma_2^2$.*

Theorem 8 *For any $(D_0, D_1, D_2) \in \mathcal{D}_1 \bigcup \mathcal{D}_2$, there exists an $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ with $L_1 = R_1 - R(D_1) \le 0.5 \log(2 - d_1) \le 0.5$ and*

$$L_0 = R_1 + R_2 - R(D_0)$$

$$\leq 0.5\log(2-d_0)+\min\{L_{\min}(d_1,d_2), R(\max\{D_1,D_2\})+0.5\log(2-\max\{d_1,d_2\})$$

$$\leq \min\{L_{\min}(d_1,d_2)+0.5, R(\max\{D_1,D_2\})+1\}.$$

Proof: I make the following assumptions:

1. U_0, U_1 and U_2 are the random variables that achieve $R(D_0)$, $R(D_1)$, and $R(D_2)$, respectively, i.e., $Ed(X,U_i) \leq D_i$, and $I(X;U_i) = R(D_i)$ for $i = 0,1,2$.

2. N_0, N_1, and N_2 are Gaussian random variables with zero means and variances $\sigma_0^2 = \beta_0 D_0$, σ_1^2, and σ_2^2, respectively.

3. $(N_0,N_1,N_2) \perp\!\!\!\perp (X,U_0,U_1,U_2)$. Further, $N_0 \perp\!\!\!\perp (N_1,N_2)$ and $E(N_1N_2) = \Gamma$.

First, I verify that these assumptions are reasonable by showing that $\sigma_1^2,\sigma_2^2 \geq 0$ and $[E(N_1N_2)]^2 \leq \sigma_1^2\sigma_2^2$, i.e., Cauchy's inequality holds for N_1 and N_2. The latter is true from Lemma 8. To prove the former, I have

$$\begin{aligned}
\beta_1 D_1 - \beta_1^2 D_0/\beta_0 &= \frac{\beta_1}{\beta_0}(\beta_0 D_1 - \beta_1 D_0) \\
&= \frac{\beta_1}{\beta_0}\left(\left(1-\frac{D_0}{\sigma^2}\right)D_1 - \left(1-\frac{D_1}{\sigma^2}\right)D_0\right) \\
&= \frac{\beta_1}{\beta_0}(D_1 - D_0) \geq 0,
\end{aligned}$$

or equivalently, $\sigma_1^2 \geq 0$. Similarly, $\sigma_2^2 \geq 0$.

Second, I define

$$\begin{aligned}
Y_0 &= \beta_0 X + N_0, \\
Y_1 &= \frac{\beta_1}{\beta_0}Y_0 + N_1 = \beta_1 X + \frac{\beta_1}{\beta_0}N_0 + N_1, \\
Y_2 &= \frac{\beta_2}{\beta_0}Y_0 + N_2 = \beta_2 X + \frac{\beta_2}{\beta_0}N_0 + N_2.
\end{aligned}$$

It can be shown that $Ed(X, Y_i) \leq D_i$, $i = 0, 1, 2$. Therefore, according to Theorem 7, the rate pair (R_1, R_2) is achievable if $R_1 \geq I(X;Y_1)$, $R_2 \geq I(X;Y_2)$, and $R_1 + R_2 \geq I(X;Y_0, Y_1, Y_2) + I(Y_1;Y_2)$. Following the proof of [32, Theorem 6], the rate loss at decoder 1 is

$$\begin{aligned}
L_1 &= I(X;Y_1) - I(X;U_1) \\
&= I(X;Y_1|U_1) - I(X;U_1|Y_1) && (4.11) \\
&\leq I(X;Y_1|U_1) \\
&= I\left(X;\beta_1 X + \frac{\beta_1}{\beta_0}N_0 + N_1 \middle| U_1\right) \\
&= I\left(X - U_1;\beta_1(X - U_1) + \frac{\beta_1}{\beta_0}N_0 + N_1 \middle| U_1\right) && (4.12) \\
&\leq I\left(X - U_1;\beta_1(X - U_1) + \frac{\beta_1}{\beta_0}N_0 + N_1\right) && (4.13) \\
&= h\left(\beta_1(X - U_1) + \frac{\beta_1}{\beta_0}N_0 + N_1\right) - h\left(\frac{\beta_1}{\beta_0}N_0 + N_1\right) \\
&\leq 0.5\log(2 - d_1) && (4.14) \\
&\leq 0.5,
\end{aligned}$$

where (4.11) follows by applying the chain rule twice to $I(X;Y_1, U_1)$ to obtain

$$\begin{aligned}
I(X;Y_1, U_1) &= I(X;Y_1) + I(X;U_1|Y_1) \\
&= I(X;U_1) + I(X;Y_1|U_1);
\end{aligned}$$

(4.12) follows since $h(A|B) = h(A-B|B)$ and $h(A|B, C) = h(A-\alpha B|B, C)$ for any constant α, (4.13) follows since $(N_0, N_1) \perp\!\!\!\perp (X, U_1)$ implies that $U_1 \to X - U_1 \to \beta_1(X - U_1) + \beta_1 N_0/\beta_0 + N_1$ forms a Markov chain, and (4.14) follows since the Gaussian distribution maximizes the differential entropy under the constraint that $Ed(X, U_1) \leq D_1$.

Similarly, the rate loss at decoder 2 is bounded as

$$L_2 = I(X;Y_2) - R(D_2) \leq 0.5\log(2 - d_2) \leq 0.5. \quad (4.15)$$

Finally, the total rate at the joint decoder of the optimal code is

$$\begin{aligned} R_1 + R_2 &\leq I(X; Y_0, Y_1, Y_2) + I(Y_1; Y_2) \\ &= I(X; Y_0) + I(Y_1; Y_2) \end{aligned} \tag{4.16}$$

because $X \to Y_0 \to (Y_1, Y_2)$ form a Markov chain. Thus, the rate loss at the joint decoder is

$$\begin{aligned} L_0 &= R_1 + R_2 - I(X; U_0) \\ &\leq [I(X; Y_0) - I(X; U_0)] + I(Y_1; Y_2) \\ &\leq 0.5 \log(2 - d_0) + I(Y_1; Y_2), \end{aligned} \tag{4.17}$$

where the derivation of (4.17) parallels that of (4.14).

Since $\beta_1 N_0/\beta_0 + N_1$ and $\beta_2 N_0/\beta_0 + N_2$ are Gaussian random variables and

$$E\left\{\left(\frac{\beta_1}{\beta_0}N_0 + N_1\right)\left(\frac{\beta_2}{\beta_0}N_0 + N_2\right)\right\} = \frac{\beta_1\beta_2}{\beta_0^2}\beta_0 D_0 + E(N_1 N_2) = 0,$$

then these two random variables are independent. From the definitions of Y_1 and Y_2, we can see that $Y_1 \to X \to Y_2$ forms a Markov chain. Therefore, I bound $I(Y_1; Y_2)$ as

$$\begin{aligned} I(Y_1; Y_2) &\leq I(X; Y_2) \qquad (4.18) \\ &= I\left(X; \beta_2 X + \frac{\beta_2}{\beta_0}N_0 + N_2\right) \\ &= h\left(\beta_2 X + \frac{\beta_2}{\beta_0}N_0 + N_2\right) - h\left(\frac{\beta_2}{\beta_0}N_0 + N_2\right) \\ &\leq \frac{1}{2}\log\frac{\beta_2^2\sigma^2 + \beta_2 D_2}{\beta_2 D_2} \\ &= \frac{1}{2}\log\frac{\sigma^2}{D_2}. \qquad (4.19) \end{aligned}$$

On the other hand, by the symmetry of (4.19), I have

$$I(Y_1; Y_2) \leq \frac{1}{2}\log\frac{\sigma^2}{D_1}.$$

Thus, in summary, I have

$$I(Y_1; Y_2) \leq L_{\min}(d_1, d_2). \tag{4.20}$$

As a consequence,

$$L_0 \leq \frac{1}{2}\log(2 - d_0) + L_{\min}(d_1, d_2).$$

I can also compare $I(Y_1; Y_2)$ to $R(D_1)$ or $R(D_2)$, i.e., by combining (4.17), (4.18), and (4.15), I can show that

$$L_0 \leq \frac{1}{2}\log(2 - d_0) + R(D_2) + \frac{1}{2}\log\left(2 - \frac{D_2}{\sigma^2}\right),$$

which is a decreasing function of D_2 for fixed D_0 and σ^2.

By symmetry,

$$L_0 \leq \frac{1}{2}\log(2 - d_0) + R(\max\{D_1, D_2\}) + \frac{1}{2}\log\left(2 - \frac{\max\{D_1, D_2\}}{\sigma^2}\right).$$

This concludes the proof. □

Corollary 1 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_1$*, there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5\log(2 - d_1) \leq 0.5$ *and* $L_0 \leq 0.5\log[2(2 - d_0)/(1 + d_0)] \leq 1$.

Proof: From Theorem 8, I know that for any $(D_0, D_1, D_2) \in \mathcal{D}_1$,

$$\begin{cases} L_1 \leq 0.5\log(2 - d_1), \\ L_2 \leq 0.5\log(2 - d_2), \\ L_0 \leq 0.5\log(2 - d_0) + L_{\min}(d_1, d_2). \end{cases} \tag{4.21}$$

is achievable.

From the definition of $\mathcal{D}_1$, $D_1 + D_2 \geq D_0 + \sigma^2$, thus,

$$\max\{D_1, D_2\} \geq \frac{D_1 + D_2}{2} \geq \frac{D_0 + \sigma^2}{2}.$$

Then from (4.3),

$$L_{\min}(d_1, d_2) = \frac{1}{2}\log\frac{\sigma^2}{\max\{D_1, D_2\}} \le \frac{1}{2}\log\frac{2}{1+d_0},$$

which proves this corollary. □

The two following inequalities are useful in proving Corollary 2.

Lemma 9 *In $\mathcal{D}_2$, $0 < L_{G0} < L_{\min}(d_1, d_2) < L_{G0} + 1$.*

Corollary 5 follows immediately from this lemma.

Corollary 5 *For any $(D_0, D_1, D_2) \in \mathcal{D}_2$, if $d_1 \le 0.5$ or $d_2 \le 0.5$, then $0 < L_{G0} < 0.5$.*

Corollary 2 *For any $(D_0, D_1, D_2) \in \mathcal{D}_2$ there exists an $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ with $L_1 \le 0.5$ and $L_0 \le L_{G0} + 1.5$.*

Proof: This corollary is an immediate result of Lemma 9 and Theorem 8. □

Corollary 4 *For any $(D_0, D_1, D_2) \in \mathcal{D}_2$, there exists an $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ with $L_1 \le 0.5$ and $L_{12} \le 0.5\log[\min\{D_1, D_2\}/D_0] + 1$.*

Proof: From Theorem 8, there exists an achievable rate pair (R_1, R_2) with $L_1 \le 0.5$ and $L_0 \le \min\{R(D_1), R(D_2)\} + 1$. From (4.5),

$$L_{12} \le 1 + \min\{R(D_0) - R(D_1), R(D_0) - R(D_2)\} \le 1 + \frac{1}{2}\log\frac{\min\{D_1, D_2\}}{D_0},$$

where the last inequality comes from Lemma 2 in Section 3.3. □

I define these quantities

$$\begin{aligned}
\Gamma' &= \frac{\beta_1\beta_2 D_0 - \sqrt{\beta_1\beta_2(D_1 - D_0)(D_2 - D_0)}}{\beta_0}, \\
\alpha_1 &= \frac{\beta_0(\beta_1 D_2 - \Gamma')}{\beta_1(\beta_2 D_1 + \beta_1 D_2 - 2\Gamma')}, \\
\alpha_2 &= \frac{\beta_0(\beta_2 D_1 - \Gamma')}{\beta_2(\beta_2 D_1 + \beta_1 D_2 - 2\Gamma')}.
\end{aligned}$$

Lemmas 10 and 11, which describe the properties of α_1, α_2, and Γ', are useful in the proof of Theorem 9.

Lemma 10 *In* $\mathcal{D}_2$, $-\sqrt{\beta_1\beta_2 D_1 D_2} < \Gamma' < 0$.

Lemma 11 *In* $\mathcal{D}_2$, $\alpha_1^2\beta_1 D_1 + \alpha_2^2\beta_2 D_2 + 2\alpha_1\alpha_2\Gamma' = \beta_0 D_0$.

Theorem 9 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$, *there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5$ *and* $L_{12} \leq L_{G12} + 1$.

Proof: In this proof, I define

$$\begin{aligned} Y_1 &= \beta_1 X + N_1', \\ Y_2 &= \beta_2 X + N_2', \end{aligned}$$

where $N_i' \sim \mathcal{N}(0, \beta_i D_i)$ and $N_i' \perp\!\!\!\perp (X, U_0, U_1, U_2)$ for $i = 1, 2$, and $E(N_1' N_2') = \Gamma'$. From Lemma 10, these definitions are reasonable and $\alpha_1, \alpha_2 > 0$.

Next, I define $Y_0 = \alpha_1 Y_1 + \alpha_2 Y_2$. Since $\alpha_1\beta_1 + \alpha_2\beta_2 = \beta_0$, $Y_0 = \beta_0 X + N_0'$, where $N_0' = \alpha_1 N_1' + \alpha_2 N_2'$. Notice that $N_0' \perp\!\!\!\perp (X, U_0, U_1, U_2)$, and from Lemma 11, $N_0' \sim \mathcal{N}(0, \beta_0 D_0)$. Therefore, it can be shown that $Ed(X, Y_i) \leq D_i$, $i = 0, 1, 2$.

Again, the rate pair (R_1, R_2) is achievable if $R_1 \geq I(X; Y_1)$, $R_2 \geq I(X; Y_2)$, and $R_1 + R_2 \geq I(X; Y_0, Y_1, Y_2) + I(Y_1; Y_2)$, and

$$\begin{aligned} L_1 = I(X; Y_1) - I(X; U_1) &\leq 0.5 \log(2 - d_1) \leq 0.5, \\ L_2 = I(X; Y_2) - I(X; U_2) &\leq 0.5 \log(2 - d_2) \leq 0.5. \end{aligned}$$

In addition, since Y_0 is a function of Y_1 and Y_2, $I(X; Y_0, Y_1, Y_2) = I(X; Y_1, Y_2)$, thus

$$L_{12} \leq I(X; Y_1, Y_2) + I(Y_1; Y_2) - R(D_1) - R(D_2)$$

$$
\begin{aligned}
&= I(X;Y_1) + I(X;Y_2|Y_1) + I(Y_1;Y_2) - R(D_1) - R(D_2) \\
&= I(X;Y_1) + I(X,Y_1;Y_2) - R(D_1) - R(D_2) \\
&= [I(X;Y_1) - R(D_1)] + [I(X;Y_2) - R(D_2)] + I(Y_1;Y_2|X) \\
&= [I(X;Y_1) - R(D_1)] + [I(X;Y_2) - R(D_2)] + I(N_1';N_2') \\
&\leq 0.5\log(2-d_1) + 0.5\log(2-d_2) + I(N_1';N_2') \\
&< 1 + I(N_1';N_2') \\
&= 1 - \frac{1}{2}\log\left[1 - \left(\frac{\Gamma'}{\sqrt{\beta_1 D_1}\sqrt{\beta_2 D_2}}\right)^2\right] \\
&= 1 - \frac{1}{2}\log\left[1 - \left(\frac{\beta_1\beta_2 D_0 - \sqrt{\beta_1\beta_2(D_1-D_0)(D_2-D_0)}}{\beta_0\sqrt{\beta_1\beta_2 D_1 D_2}}\right)^2\right] \\
&= 1 - \frac{1}{2}\log\left[1 - \left(\frac{d_0\sqrt{\beta_1\beta_2} - \sqrt{(d_1-d_0)(d_2-d_0)}}{\beta_0\sqrt{d_1 d_2}}\right)^2\right] \\
&= 1 + \frac{1}{2}\log\frac{(1-d_0)^2 d_1 d_2}{(1-d_0)^2 d_1 d_2 - (d_0\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)})^2} \\
&= 1 + L_{G12}.
\end{aligned}
$$

This proves the theorem. □

I introduce three new quantities:

$$
\begin{aligned}
\Delta &= \beta_0\beta_1\beta_2 D_0 D_1 D_2\left(\frac{1}{D_1} + \frac{1}{D_2} - \frac{1}{D_0} - \frac{1}{\sigma^2}\right) \\
&= \beta_0\beta_1\beta_2(D_0D_2 + D_0D_1 - D_1D_2 - d_0D_1D_2), \\
\alpha_1' &= \frac{\beta_0\beta_1 D_2 + \sqrt{\Delta}}{\beta_1(\beta_2 D_1 + \beta_1 D_2)}, \\
\alpha_2' &= \frac{\beta_0\beta_2 D_1 - \sqrt{\Delta}}{\beta_2(\beta_2 D_1 + \beta_1 D_2)},
\end{aligned}
$$

where $\Delta \geq 0$ in $\mathcal{D}_3$. The following lemma is useful in proving Theorem 10.

Lemma 12 *In $\mathcal{D}_3$, $\alpha_1'^2\beta_1 D_1 + \alpha_2'^2\beta_2 D_2 = \beta_0 D_0$.*

Theorem 10 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_3$*, there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5 \log(2 - d_1) \leq 0.5$ *and* $L_2 \leq 0.5 \log(2 - d_2) \leq 0.5$*.*

Proof: Let

$$
\begin{aligned}
Y_1 &= \beta_1 X + N_1' \\
Y_2 &= \beta_2 X + N_2' \\
Y_0 &= \alpha_1' Y_1 + \alpha_2' Y_2,
\end{aligned}
$$

where $N_i' \sim \mathcal{N}(0, \beta_i D_i)$ and $N_i' \perp\!\!\!\perp (X, U_0, U_1, U_2)$ for $i = 1, 2$, and $N_1' \perp\!\!\!\perp N_2'$ (note that N_1' and N_2' were not independent in the previous proof). It can be verified that $Y_0 = \alpha_1' Y_1 + \alpha_2' Y_2 = \beta_0 X + N_0$, where $N_0 = \alpha_1' N_1' + \alpha_2' N_2' \sim \mathcal{N}(0, \beta_0 D_0)$ (from Lemma 12) and $N_0 \perp\!\!\!\perp (X, U_0, U_1, U_2)$. Thus, it can be shown that Y_1, Y_2, and Y_0 satisfy the distortion requirements. From Theorem 7, the rate pair (R_1, R_2) is achievable if $R_1 \geq I(X; Y_1)$, $R_2 \geq I(X; Y_2)$, and $R_1 + R_2 \geq I(X; Y_0, Y_1, Y_2) + I(Y_1; Y_2) = I(X; Y_1, Y_2) + I(Y_1; Y_2) = I(X; Y_1) + I(X; Y_2) + I(Y_1; Y_2|X)$.

Again, $L_1 \leq 0.5 \log(2 - d_1) \leq 0.5$ and $L_2 \leq 0.5 \log(2 - d_2) \leq 0.5$. Bounding L_{12} gives us

$$
\begin{aligned}
L_{12} &= R_1 + R_2 - R(D_1) - R(D_2) \\
&\leq I(X; Y_1) + I(X; Y_2) + I(Y_1; Y_2|X) - R(D_1) - R(D_2) \\
&\leq 0.5 \log(2 - d_1) + 0.5 \log(2 - d_2) + I(N_1'; N_2') \\
&= 0.5 \log(2 - d_1) + 0.5 \log(2 - d_2) \qquad (4.22) \\
&\leq 1,
\end{aligned}
$$

where (4.22) comes from the fact that N_1' and N_2' are independent, therefore, $I(N_1'; N_2') = 0$.

Here, the bound for L_{12} is redundant from the bounds on L_1 and L_2. □

Before proving Theorem 11, I first define the entropy power of X as:

$$P_X = 2^{2h(X)}/(2\pi e).$$

and define $LG(\delta_0, \delta_1, \delta_2)$ as

$$LG(\delta_0, \delta_1, \delta_2) = \frac{1}{2}\log \frac{(1-\delta_0)^2}{(1-\delta_0)^2 - (\sqrt{(1-\delta_1)(1-\delta_2)} - \sqrt{(\delta_1-\delta_0)(\delta_2-\delta_0)})^2}.$$

Notice that $P_X \leq \sigma^2$ and $L_{G0} = LG(D_0/\sigma^2, D_1/\sigma^2, D_2/\sigma^2)$. I also need the following lemma in the proof of Theorem 11.

Lemma 13 *If* $0 < \delta_0 < \delta_1 \leq \delta_2 < 0.5$, *then*

$$LG(\delta_0, \delta_1, \delta_2) > \frac{1}{2}\log \frac{1}{4\delta_2(1-\delta_2)}.$$

Theorem 11 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$ *with* $\max\{D_1, D_2\} < \sigma^2/2$, *there exists a rate pair* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5\log(2-d_1) \leq 0.5$, *and the distance between the upper bound and the lower bound for* L_0 *is less than or equal to* $\min\{[\log(2\pi e\sigma^2) - 2h(X)] + 1, 0.5\log(2\pi e\sigma^2) - h(X) + 1.5\}$.

Proof: Here, I want to compare the lower bound and upper bound for L_0 and bound from above the distance between them. I use K_0 to denote this distance. Without loss of generality, I assume that $D_1 \leq D_2$. We are only interested in region $\mathcal{D}_2$ with $D_2 < \sigma^2/2$, i.e., $D_2 < \sigma^2/2$ and $D_1 + D_2 - \sigma^2 < 0 < D_0 < (1/D_1 + 1/D_2 - 1/\sigma^2)^{-1}$.

From Lemma 9, $LG(d_0, d_1, d_2) > 0$ in region $\mathcal{D}_2$. From the proof of Theorem 8 (i.e., (4.14), (4.15), (4.16), and (4.19)), there exists an $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ with

$$\begin{aligned} L_1 &\leq 0.5\log(2-d_1) \leq 0.5, \\ L_2 &\leq 0.5\log(2-d_2) \leq 0.5, \\ L_0 &\leq I(X;Y_0) + \frac{1}{2}\log\frac{\sigma^2}{D_2} - R(D_0), \end{aligned} \tag{4.23}$$

where Y_0 is defined as $Y_0 = \beta_0 X + N_0$, $N_0 \sim \mathcal{N}(0, \beta_0 D_0)$ and $N_0 \perp\!\!\!\perp X$. Thus

$$I(X;Y_0) \le \frac{1}{2}\log\frac{\sigma^2}{D_0}. \tag{4.24}$$

I first consider the case where $D_2 < P_X/2$. In this configuration, $D_1 + D_2 - P_X < 0 < D_0$. I also have

$$\frac{1}{D_1} + \frac{1}{D_2} - \frac{1}{D_0} < \frac{1}{\sigma^2} \le \frac{1}{P_X}.$$

In [44], it is shown that if $0 < D_0 < D_1, D_2 \le P_X$ and $D_1 + D_2 - P_X < D_0 < (1/D_1 + 1/D_2 - 1/P_X)^{-1}$, then the region described by

$$\begin{aligned} R_1 &\ge R(D_1) \\ R_2 &\ge R(D_2) \\ R_1 + R_2 &\ge \frac{1}{2}\log\frac{P_X}{D_0} + LG(D_0/P_X, D_1/P_X, D_2/P_X) \end{aligned}$$

is an outer bound for the achievable region, which also implies that

$$L_0 \ge 0.5\log(P_X/D_0) + LG(D_0/P_X, D_1/P_X, D_2/P_X) - R(D_0). \tag{4.25}$$

By combining (4.23), (4.24), and (4.25), I get the following bound on the distance K_0 between the upper and lower bounds for L_0

$$\begin{aligned} K_0 &\le \frac{1}{2}\log\frac{\sigma^2}{D_0} + \frac{1}{2}\log\frac{\sigma^2}{D_2} - \frac{1}{2}\log\frac{P_X}{D_0} - LG(D_0/P_X, D_1/P_X, D_2/P_X), \\ &= \frac{1}{2}\log\frac{\sigma^2}{P_X} + \frac{1}{2}\log\frac{\sigma^2}{D_2} - LG(D_0/P_X, D_1/P_X, D_2/P_X). \end{aligned}$$

From Lemma 13,

$$LG(D_0/P_X, D_1/P_X, D_2/P_X) > \frac{1}{2}\log\frac{P_X^2}{4D_2(P_X - D_2)}.$$

As a result,

$$\begin{aligned} K_0 &\leq \frac{1}{2}\log\frac{\sigma^2}{P_X} + \frac{1}{2}\log\frac{4\sigma^2(P_X - D_2)}{P_X^2} \\ &\leq \frac{1}{2}\log\frac{\sigma^2}{P_X} + \frac{1}{2}\log\frac{4\sigma^2}{P_X} \\ &= \log\frac{\sigma^2}{P_X} + 1. \end{aligned}$$

If $D_2 \geq P_X/2$, then I use 0 as the lower bound for L_0 and use the SLB

$$R(D_0) \geq \frac{1}{2}\log\frac{P_X}{D_0}.$$

From (4.23) and (4.24),

$$\begin{aligned} K_0 &\leq L_0 \\ &\leq \frac{1}{2}\log\frac{\sigma^2}{D_0} + \frac{1}{2}\log\frac{\sigma^2}{D_2} - R(D_0) \\ &\leq \frac{1}{2}\log\frac{\sigma^2}{D_0} + \frac{1}{2}\log\frac{\sigma^2}{D_2} - \frac{1}{2}\log\frac{P_X}{D_0} \\ &= \frac{1}{2}\log\frac{\sigma^2}{P_X} + \frac{1}{2}\log\frac{\sigma^2}{D_2} \\ &\leq \frac{1}{2}\log\frac{\sigma^2}{P_X} + \frac{1}{2}\log\frac{2\sigma^2}{P_X} \\ &= \log\frac{\sigma^2}{P_X} + 0.5. \end{aligned}$$

Thus, in both cases, $K_0 \leq \log(\sigma^2/P_X) + 1 = 2[0.5\log(2\pi e\sigma^2) - h(X)] + 1$.

In fact, I can show a better outer bound for the achievable region. It has been shown that (e.g., [42] and [44]), the inequality $R_1 + R_2 \geq SR(e_0, e_1, e_2, R_1 + R_2)$ is equivalent to the inequality $R_1 + R_2 \geq -0.5\log e_0 + LG(e_0, e_1, e_2)$ if $0 < e_0 < e_1, e_2 \leq 1$ and $e_0 + e_1 - 1 < e_0 < (1/e_1 + 1/e_2 - 1)^{-1}$, where $SR(e_0, e_1, e_2, R_1 + R_2)$ is a function defined as:

$$SR(e_0, e_1, e_2, R_1 + R_2) = \frac{1}{2}\log\frac{1}{e_0} - \frac{1}{2}\log\left(1 - \left(\sqrt{(1-e_1)(1-e_2)} - \sqrt{e_1e_2 - 2^{-2(R_1+R_2)}}\right)^2\right).$$

From [44],

$$\begin{aligned} R_1 &\geq R(D_1) \\ R_2 &\geq R(D_2) \\ R_1 + R_2 &\geq SR(d_0^*, D_1/P_X, D_2/P_X, R_1 + R_2) \end{aligned}$$

is also an outer bound for the achievable region if $0 < D_0 < D_1, D_2 \leq P_X$ and $D_1 + D_2 - P_X < P_X d_0^* < (1/D_1 + 1/D_2 - 1/P_X)^{-1}$, where d_0^* is the "effective distortion" at the joint decoder defined as

$$d_0^* = 2^{-2R(D_0)}.$$

Note that since $R(D_0) = 0.5 \log(1/d_0^*)$, this outer bound is equivalent to

$$\begin{aligned} R_1 &\geq R(D_1) \\ R_2 &\geq R(D_2) \\ R_1 + R_2 &\geq \frac{1}{2} \log \frac{1}{d_0^*} + LG(d_0^*, D_1/P_X, D_2/P_X) \\ &= R(D_0) + LG(d_0^*, D_1/P_X, D_2/P_X). \end{aligned} \tag{4.26}$$

I will bound the distance between the inner bound and this new outer bound. Again, I assume $D_1 \leq D_2$ and I first verify that (4.26) is an outer bound in region $\mathcal{D}_2$. In region $\mathcal{D}_2$,

$$D_0 < \frac{1}{1/D_1 + 1/D_2 - 1/\sigma^2} \leq \frac{1}{1/D_1 + 1/D_2 - 1/P_X}$$

because $P_X \leq \sigma^2$. Since

$$R(D_0) \geq h(X) - \frac{1}{2} \log(2\pi e D_0),$$

I have

$$P_X 2^{-2R(D_0)} \leq D_0 < \frac{1}{1/D_1 + 1/D_2 - 1/P_X}.$$

Thus, if $D_2 < P_X/2$, I have $D_1 + D_2 - P_X < 0 < D_0$, and from Lemma 13,

$$\begin{aligned} L_0 &\geq LG(d_0^*, D_1/P_X, D_2/P_X) \\ &\geq \frac{1}{2}\log\frac{P_X^2}{4D_2(P_X - D_2)}. \end{aligned}$$

From Theorem 8, there exists (R_1, R_2) with $L_1 \leq 0.5\log(2-d_1) \leq 0.5$, $L_2 \leq 0.5\log(2-d_2) \leq 0.5$ and

$$\begin{aligned} K_0 &\leq \frac{1}{2} + \frac{1}{2}\log\frac{\sigma^2}{D_2} - LG(d_0^*, D_1/P_X, D_2/P_X) \\ &\leq \frac{1}{2} + \frac{1}{2}\log\frac{4\sigma^2(P_X - D_2)}{P_X^2} \\ &\leq \frac{1}{2} + \frac{1}{2}\log\frac{4\sigma^2}{P_X}. \end{aligned}$$

Thus, $K_0 \leq 0.5\log(2\pi e\sigma^2) - h(X) + 1.5$. Similarly, if $D_2 \geq P_X/2$,

$$K_0 \leq L_0 \leq 0.5 + \frac{1}{2}\log\frac{\sigma^2}{D_2} \leq 0.5\log(2\pi e\sigma^2) - h(X) + 1.$$

In summary, in this region, K_0 is always bounded by $K_0 \leq 0.5\log(2\pi e\sigma^2) - h(X) + 1.5$, which is better than the previous bound if $h(X)$ is much less than $0.5\log(2\pi e\sigma^2)$. □

Theorem 12 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_2$ *with* $\max\{D_1, D_2\} < \sigma^2/2$, *if the SLB is tight at distortion* D_1 *and* D_2, *then there exists an* $(R_1, R_2) \in \mathcal{R}(D_0, D_1, D_2)$ *with* $L_1 \leq 0.5\log(2 - d_1) \leq 0.5$, *and the distance between the upper bound and the lower bound for* L_0 *is less than or equal to 2.*

Proof: Since the SLB is tight at distortion D_1,

$$D_1 = \frac{1}{2\pi e}2^{2[h(X)-R(D_1)]} = P_X d_1^*,$$

where $d_1^* = 2^{-2R(D_1)}$ is the "effective distortion" at decoder 1. Similarly, $D_2 = P_X d_2^*$, where $d_2^* = 2^{-2R(D_2)}$. Then, from (4.26),

$$R_1 + R_2 \geq R(D_0) + LG(d_0^*, D_1/P_X, D_2/P_X) = R(D_0) + LG(d_0^*, d_1^*, d_2^*).$$

Again, I assume that $D_1 \leq D_2$ without loss of generality. Because the rate-distortion function $R(D)$ is a non-increasing function of D, so $0 < d_0^* < d_1^* \leq d_2^*$. Thus, if $d_2^* < 0.5$, then from Lemma 13,

$$\begin{aligned} LG(d_0^*, d_1^*, d_2^*) &> \frac{1}{2}\log\frac{1}{d_2^*} + \frac{1}{2}\log\frac{1}{4(1-d_2^*)} \\ &= R(D_2) + \frac{1}{2}\log\frac{1}{4(1-d_2^*)} \\ &\geq R(D_2) - 1. \end{aligned}$$

From Theorem 8, $R_1 + R_2 = R(D_0) + R(D_2) + 1$ is achievable; thus, the difference between the upper bound and this new lower bound for L_0 is

$$K_0 \leq [R(D_0) + R(D_2) + 1] - [R(D_0) + R(D_2) - 1] = 2.$$

If $d_2^* \geq 0.5$, then $R(D_2) \leq 0.5$ according to the definition of d_2^*. Thus, $R_1 + R_2 = R(D_0) + R(D_2) + 1 \leq R(D_0) + 1.5$ is achievable, which implies that

$$K_0 \leq L_0 \leq R_1 + R_2 - R(D_0) < 1.5.$$

In either case, the distance between the upper and lower bound is no larger than 2. □

Chapter 5

Multiple Access Source Codes

5.1 Preliminaries

Let $\{X_i, Y_i\}_{i=1}^{\infty}$ be a real-valued iid vector source with joint pdf $f_{X,Y}(x,y)$.

Define the *joint rate-distortion function* for this vector source as:

$$R_{X,Y}(D_1, D_2) = \inf I(X,Y;\hat{X},\hat{Y}),$$

where the infimum is over all random vectors $(\hat{X},\hat{Y})$ satisfying

$$Ed(X,\hat{X}) \leq D_1, Ed(Y,\hat{Y}) \leq D_2.$$

This joint rate-distortion function is the minimum total rate for jointly describing source $\{X_i, Y_i\}_{i=1}^{\infty}$ with distortions no greater than D_1 for X and D_2 for Y ([46], [47], [48]).

Define

$$R_{X|Y}(D_1) = \inf I(X;\hat{X}|Y)$$

as the *conditional rate-distortion function*, where the infimum is over all random variables $\hat{X}$ such that $Ed(X,\hat{X}) \leq D_1$.

An (n, M_1, M_2) 2ASC consists of two encoders:

$$f_n^{(1)} : \mathbf{R}^n \to \{1, \ldots, M_1\}$$

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

with rates $(1/n) \log M_1$ and $(1/n) \log M_2$ respectively; and one decoder:

$$g_n : \{1, \ldots, M_1\} \times \{1, \ldots, M_2\} \to \mathbf{R}^n \times \mathbf{R}^n$$

with distortions $Ed_n(X^n, g_n^{(1)}(f_n^{(1)}(X^n), f_n^{(2)}(Y^n)))$ and $Ed_n(Y^n, g_n^{(2)}(f_n^{(1)}(X^n), f_n^{(2)}(Y^n)))$ for X^n and Y^n respectively, where $g_n^{(1)}$ stands for the first output of function g_n and $g_n^{(2)}$ stands for the second output.

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

$$\frac{1}{n} \log M_1 \le R_1 + \epsilon, \quad Ed_n(X^n, g_n^{(1)}(f_n^{(1)}(X^n), f_n^{(2)}(Y^n))) \le D_1 + \epsilon,$$

$$\frac{1}{n} \log M_2 \le R_2 + \epsilon, \quad Ed_n(Y^n, g_n^{(2)}(f_n^{(1)}(X^n), f_n^{(2)}(Y^n))) \le D_2 + \epsilon.$$

The achievable region for 2ASCs is defined as the set of all achievable rate-distortion vectors.

The achievable rate region for lossless coding with an MASC ($D_1 = D_2 = 0$) is characterized by

$$\begin{aligned} R_1 &\ge H(X|Y), \\ R_2 &\ge H(Y|X), \\ R_1 + R_2 &\ge H(X, Y), \end{aligned}$$

which is shown by Slepian and Wolf in [33].

In contrast, if the two senders can jointly encode X and Y, then the achievable region is

$$R_1 + R_2 \ge H(X, Y).$$

Thus, independent coding can achieve a total rate of $H(X,Y)$, which is as good as the total rate for joint coding. Figure 5.1 compares the rate region of Slepian-Wolf encoding to that of joint encoding.

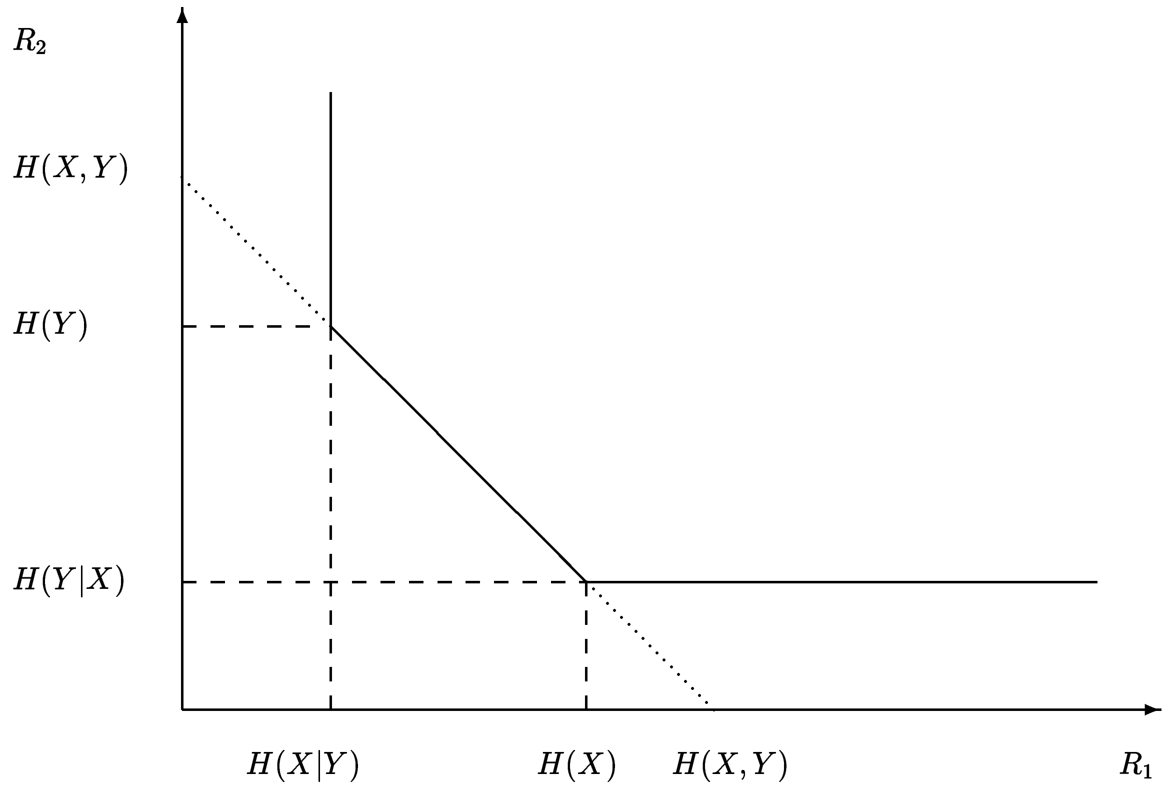

Figure 5.1: The achievable region for a lossless MASC compared to that for joint encoding. The solid lines shows the MASC-achievable region (also known as the Slepian-Wolf region), while the dotted line shows that for joint encoding.

For lossy coding scenarios, i.e., if D_1 and D_2 may be nonzero, then the achievable region of joint encoders is

$$R_1 + R_2 \geq R_{X,Y}(D_1, D_2).$$

For lossy MASCs, we do not know the exact achievable rate-distortion region. The following theorem gives an inner bound on the achievable region, which is not tight in general. The result for finite alphabets comes from [49] (see also [34]). It can be generalized to real-valued sources with an escape symbol following the method of [50] and [19] (see also [51], [52], [35]).

Theorem 13 *[49, Theorem 6.1],[34, Theorem 4.1]*

For any iid vector source $\{X_i, Y_i\}_{i=1}^{\infty}$ with density $f_{X,Y}(x, y)$ and distortion measure d, the vector (R_1, R_2, D_1, D_2) is 2ASC achievable if there exists a conditional probability $Q_{\hat{X},\hat{Y}|X,Y}$ such that

$$\hat{X} \to X \to Y \to \hat{Y},$$
$$R_1 \geq I(X,Y;\hat{X}|\hat{Y}), \qquad Ed(X,\hat{X}) \leq D_1,$$
$$R_2 \geq I(X,Y;\hat{Y}|\hat{X}), \qquad Ed(Y,\hat{Y}) \leq D_2,$$
$$R_1 + R_2 \geq I(X,Y;\hat{X},\hat{Y}).$$

For any 2ASC, if the rate-distortion vector (R_1, R_2, D_1, D_2) is achievable, the individual rate losses of encoder of X and Y are defined as $L_1 = R_1 - R_{X|Y}(D_1)$ and $L_2 = R_2 - R_{Y|X}(D_2)$ bps. The joint rate loss is measured by $L_0 = R_1 + R_2 - R_{X,Y}(D_1, D_2)$ bps. I further assume $0 < D_1 \leq \sigma_X^2$ and $0 < D_2 \leq \sigma_Y^2$, where σ_X^2 and σ_Y^2 are the variances of X and Y, respectively.

5.2 Major results

Theorem 14 compares lossy MASCs to the corresponding joint encoding. The statement about rate loss L_1 can also be made to apply to L_2.

Theorem 14 *For any distortions (D_1, D_2), there exists a 2ASC-achievable rate-distortion vector (R_1, R_2, D_1, D_2) with $L_1 \leq R_{X,Y}(D_1, D_2) - R_Y(D_2) - R_{X|Y}(D_1) + 1$, and $L_0 \leq 1$.*

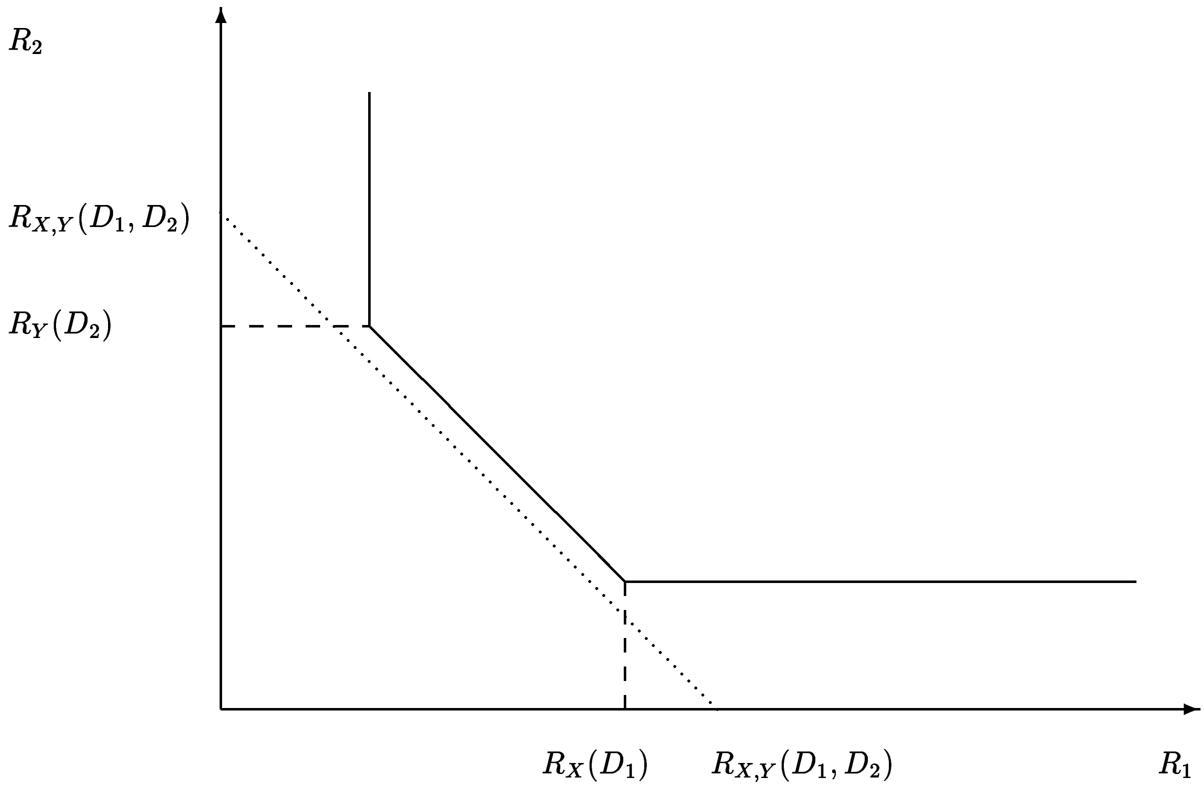

Figure 5.2: The inner bound on achievable region of lossy 2ASCs compared to the achievable region of joint encoding. The solid lines shows the inner bound, while the dotted line shows the achievable region. The difference between the sum of rates is bounded by 1.

Figure 5.2 shows this bound graphically.

The achievable region for lossy joint encoding is

$$R_1 + R_2 \geq R_{X,Y}(D_1, D_2).$$

Theorem 14 gives a new inner bound on the achievable region for lossy 2ASCs: any rate-distortion vector (R_1, R_2, D_1, D_2) satisfying

$$\begin{aligned} R_1 &\geq R_{X,Y}(D_1, D_2) - R_Y(D_2) + 1, \\ R_2 &\geq R_{X,Y}(D_1, D_2) - R_X(D_1) + 1, \end{aligned}$$

$$R_1 + R_2 \geq R_{X,Y}(D_1, D_2) + 1,$$

is achievable. The two corners of this inner bound are $(R_X(D_1), R_{X,Y}(D_1, D_2) - R_X(D_1) + 1)$ and $(R_{X,Y}(D_1, D_2) - R_Y(D_2) + 1, R_Y(D_2))$. In the region between these two corners, the difference between the total rate of MASCs and that of joint encoding is bounded by 1, which implies that separate encoding is almost as good as joint encoding in the lossy sense.

A trivial case occurs when

$$R_{X,Y}(D_1, D_2) - R_X(D_1) + 1 + R_{X,Y}(D_1, D_2) - R_Y(D_2) + 1 > R_{X,Y}(D_1, D_2) + 1.$$

In this case, the inequality $R_1 + R_2 \geq R_{X,Y}(D_1, D_2) + 1$ is redundant. However, the above condition is equivalent to

$$R_X(D_1) + R_Y(D_2) < R_{X,Y}(D_1, D_2) + 1,$$

which implies that independent codes for X and Y (at rates $R_X(D_1)$ and $R_Y(D_2)$, respectively) give total rate $R_X(D_1) + R_Y(D_2) < R_{X,Y}(D_1, D_2) + 1$, again giving a rate loss $L_0 < 1$ bps. In this case, X and Y are nearly independent, therefore, joint decoding does not help much.

Theorem 15 bounds the rate loss as a function of D_1, D_2, σ_X^2, and σ_Y^2.

Theorem 15 *For any distortions (D_1, D_2), there exists a 2ASC-achievable rate-distortion vector (R_1, R_2, D_1, D_2) with $L_1 \leq R_{X,Y}(D_1, D_2) - R_Y(D_2) - R_{X|Y}(D_1) + 0.5\log(2 - D_1/\sigma_X^2) + 0.5\log(2 - D_2/\sigma_Y^2)$, and $L_0 \leq 0.5\log(2 - D_1/\sigma_X^2) + 0.5\log(2 - D_2/\sigma_Y^2)$.*

Theorem 14 is consistent with Theorem 15 when $D_1 = D_2 = 0$, however, the latter is always tighter than the former when $D_1, D_2 > 0$; in particular, when $D_1 = D_2 = \sigma^2$, Theorem 15 is tight.

5.3 Proofs of Theorems

In this section, I will use the methods used throughout this thesis to prove the results for MASCs.

Theorem 14 *For any distortions* (D_1, D_2)*, there exists a 2ASC-achievable rate-distortion vector* (R_1, R_2, D_1, D_2) *with* $L_1 \leq R_{X,Y}(D_1, D_2) - R_Y(D_2) - R_{X|Y}(D_1) + 1$*, and* $L_0 \leq 1$.

Proof: Suppose U_1 and U_2 are two random variables such that $Ed(X, U_1) \leq D_1$, $Ed(Y, U_2) \leq D_2$ and $R_{X,Y}(D_1, D_2) = I(X, Y; U_1, U_2)$.

Define

$$\begin{aligned} \hat{X} &= X + N_1, \\ \hat{Y} &= Y + N_2, \end{aligned}$$

where $N_1 \sim (0, D_1)$, $N_2 \sim (0, D_2)$, $(N_1, N_2) \perp\!\!\!\perp (X, Y, U_1, U_2)$ and $N_1 \perp\!\!\!\perp N_2$.

It can be shown that $Ed(X, \hat{X}) \leq D_1$, $Ed(Y, \hat{Y}) \leq D_2$, and $\hat{X} \to X \to Y \to \hat{Y}$. Thus, by Theorem 13, any (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\geq I(X, Y; \hat{X}|\hat{Y}) \\ R_2 &\geq I(X, Y; \hat{Y}|\hat{X}), \\ R_1 + R_2 &\geq I(X, Y; \hat{X}, \hat{Y}), \end{aligned}$$

is achievable. For this choice of $(\hat{X}, \hat{Y})$,

$$\begin{aligned} L_0 &= R_1 + R_2 - R_{X,Y}(D_1, D_2) \\ &= I(X, Y; X + N_1, Y + N_2) - I(X, Y; U_1, U_2) \\ &= I(X, Y; X + N_1, Y + N_2|U_1, U_2) - I(X, Y; U_1, U_2|X + N_1, Y + N_2) \end{aligned}$$

$$
\begin{aligned}
&\leq I(X,Y;X+N_1,Y+N_2|U_1,U_2) \\
&= I(X-U_1,Y-U_2;X-U_1+N_1,Y-U_2+N_2|U_1,U_2) \\
&\leq I(X-U_1,Y-U_2;X-U_1+N_1,Y-U_2+N_2) \\
&\leq 1.
\end{aligned}
$$

Therefore,

$$R_{X,Y}(D_1,D_2) \leq I(X,Y;\hat{X},\hat{Y}) \leq R_{X,Y}(D_1,D_2)+1.$$

We also know that

$$
\begin{aligned}
R_X(D_1) \leq I(X;\hat{X}) \leq R_X(D_1)+0.5, \\
R_Y(D_2) \leq I(Y;\hat{Y}) \leq R_Y(D_2)+0.5.
\end{aligned}
$$

Since

$$
\begin{aligned}
I(X,Y;\hat{X},\hat{Y}) &= I(X,Y;\hat{Y})+I(X,Y;\hat{X}|\hat{Y}) \\
&= I(Y;\hat{Y})+I(X,Y;\hat{X}|\hat{Y}),
\end{aligned}
$$

where the last equality comes from the fact that $X \to Y \to \hat{Y}$. Therefore,

$$R_{X,Y}(D_1,D_2)-R_Y(D_2)-0.5 \leq I(X,Y;\hat{X}|\hat{Y}) \leq R_{X,Y}(D_1,D_2)-R_Y(D_2)+1.$$

Similarly,

$$R_{X,Y}(D_1,D_2)-R_X(D_1)-0.5 \leq I(X,Y;\hat{Y}|\hat{X}) \leq R_{X,Y}(D_1,D_2)-R_X(D_1)+1,$$

which provides the desired bounds on L_1 and L_2. □

Theorem 15 *For any distortions (D_1,D_2), there exists a 2ASC-achievable rate-distortion vector (R_1,R_2,D_1,D_2) with $L_1 \leq R_{X,Y}(D_1,D_2)-R_Y(D_2)-R_{X|Y}(D_1)+0.5\log(2-D_1/\sigma_X^2)+0.5\log(2-D_2/\sigma_Y^2)$, and $L_0 \leq 0.5\log(2-D_1/\sigma_X^2)+0.5\log(2-D_2/\sigma_Y^2)$.*

Proof: I define $\hat{X}$ and $\hat{Y}$ as:

$$\begin{aligned} \hat{X} &= \beta_1 X + N_1', \\ \hat{Y} &= \beta_2 Y + N_2', \end{aligned}$$

where $\beta_1 = 1 - D_1/\sigma_X^2$, $\beta_2 = 1 - D_2/\sigma_Y^2$, $N_i' \sim \mathcal{N}(0, \beta_i D_i)$ for i=1, 2. Further, I assume that $N_1' \perp\!\!\!\perp N_2'$ and $(N_1', N_2') \perp\!\!\!\perp (X, Y, U_1, U_2)$.

Repeating the steps in the previous proof gives us the desired results. □

Chapter 6

Additive Multi-Resolution Source Codes

Finally, we turn our attention briefly to A2RSCs.

6.1 Preliminaries

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

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

The following theorem from [36] describes an achievable region for AMRSCs. This region is not tight. The AMRSC theorem from [36] is for discrete memoryless sources with finite alphabets. The result extends to continuous memoryless sources with an escape symbol [53].

Theorem 16 *[36, Theorem 1] For any iid source $\{X_i\}_{i=1}^{\infty}$ with density $f_X(x)$ 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*

$$R_1 \geq I(X;Y_1), \qquad Ed(X,Y_1) \leq D_1$$

$$\Delta R \geq I(X;\Delta Y), \qquad Ed(X,Y_2) \leq D_2$$

$$R_2 \geq I(X;Y_1,\Delta Y) + I(Y_1;\Delta Y),$$

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

For any AMRSC with $M \geq 2$ resolutions, if the rate-distortion vector $(R_1, R_2, \ldots, R_M, D_1, D_2, \ldots, D_M)$ is achievable for $0 < D_M < \ldots < D_1 \leq \sigma^2$, the rate loss at the ith resolution ($i \in \{1, \ldots, M\}$) is defined as: $L_i = R_i - R(D_i)$ bps.

6.2 Major results

Theorem 17 *For any* $D_2 < D_1$*, there exists an A2RSC-achievable vector* (R_1, R_2, D_1, D_2) *with* $L_1 \leq 0.5$ *and*

$$L_2 \leq \min\left\{\frac{1}{2} + \frac{1}{2}\log\frac{D_1}{D_2}, \frac{1}{2} + \frac{1}{2}\log\left(1 + \frac{\sigma^2 - D_1}{D_2}\right), 1 + \frac{1}{2}\log\frac{D_1}{D_2}\left(1 - \frac{D_1 - D_2}{\sigma^2}\right)\right\}.$$

I can combine the first two bounds ($0.5 + 0.5\log(D_1/D_2)$ and $0.5 + 0.5\log(1 + (\sigma^2 - D_1)/D_2)$) 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 $L_2 \leq 0.5\log(1 + \sigma^2/D_2)$.

These bounds are good for the low rate region, especially for large D_2. For example, if either $D_2 \geq D_1/2$ or $D_2 \geq \sigma^2/3$, I have $L_1 \leq 0.5$ and $L_2 \leq 1$.

Based on the proof of Theorem 17, I next obtain a new bound that depends on $h(X)$, the differential entropy of the source. This bound is tight if X is Gaussian.

Theorem 18 *For any $D_2 < D_1$, there exists an A2RSC-achievable vector (R_1, R_2, D_1, D_2) with $L_1 \leq \min\{0.5, 0.5\log 2\pi e\sigma^2 - h(X)\}$, and $L_2 \leq 0.5\log 2\pi e\sigma^2 - h(X)$.*

6.3 Proofs of Theorems

Lemma 6 from Chapter 4 is useful in bounding the rate losses.

Theorem 17 *For any $D_2 < D_1$, there exists an A2RSC-achievable vector (R_1, R_2, D_1, D_2) with $L_1 \leq 0.5$ and*

$$L_2 \leq \min\left\{\frac{1}{2}+\frac{1}{2}\log\frac{D_1}{D_2}, \frac{1}{2}+\frac{1}{2}\log\left(1+\frac{\sigma^2-D_1}{D_2}\right), 1+\frac{1}{2}\log\frac{D_1}{D_2}\left(1-\frac{D_1-D_2}{\sigma^2}\right)\right\}.$$

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

I 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 16, $(I(X;Y_1), \max\{I(X;Y_1) + I(X;\Delta Y), I(X;Y_1,\Delta Y) + I(Y_1;\Delta Y)\}, D_1, D_2)$ 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 (8.1)-(8.5),

$$L_1 = I(X;Y_1) - I(X;U_1) \leq 0.5\log(2 - D_1/\sigma^2) \leq 0.5, \tag{6.1}$$

$$I(X;Y_2) - I(X;U_2) \leq 0.5\log(2 - D_2/\sigma^2) \leq 0.5,$$

$$I(X;\Delta Y) - I(X;U_0') \leq 0.5. \tag{6.2}$$

The rate loss at the second stage is

$$\begin{aligned} L_2 &= I(X;Y_1,\Delta Y)+I(Y_1;\Delta Y)-R(D_2) \\ &= I(X;Y_1)+I(X;\Delta Y)+I(Y_1;\Delta Y|X)-I(X;U_2), \end{aligned} \tag{6.3}$$

where

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

Thus L_2 can be bounded as:

$$\begin{aligned} L_2 &= [I(X;Y_1)-I(X;U_2)]+I(X;\Delta Y)+I(Y_1;\Delta Y|X) \\ &\le [I(X;Y_1)-I(X;U_1)]+I(X;(\beta_2-\beta_1)X+(\beta_2-\beta_1)N_2'/\beta_2-N_1) \\ &\quad +I(Y_1;\Delta Y|X) \\ &\le \frac{1}{2}+\frac{1}{2}\log\frac{1}{1-(D_1-D_2)/\sigma^2}+\frac{1}{2}\log\frac{D_1}{D_2}\left(1-\frac{D_1-D_2}{\sigma^2}\right) \\ &= \frac{1}{2}+\frac{1}{2}\log\frac{D_1}{D_2}. \end{aligned} \tag{6.4}$$

On the other hand,

$$\begin{aligned} L_2 &= I(X;Y_1)+[I(X;\Delta Y)-I(X;U_2)]+I(Y_1;\Delta Y|X) \\ &\le I(X;\beta_1 X+\beta_1 N_2'/\beta_2+N_1)+1/2+I(Y_1;\Delta Y|X) \qquad (6.5) \\ &\le \frac{1}{2}\log\frac{\sigma^2}{D_1}+\frac{1}{2}+\frac{1}{2}\log\frac{D_1}{D_2}\left(1-\frac{D_1-D_2}{\sigma^2}\right) = \frac{1}{2}+\frac{1}{2}\log\left(1+\frac{\sigma^2-D_1}{D_2}\right), \qquad (6.6) \end{aligned}$$

where (6.5) comes from the fact that $I(X;U_0') = R(\sigma^2-D_1+D_2) \le R(D_2) = I(X;U_2)$ and (6.2).

Combining Lemma 6, (6.1), (6.2) and (6.3) gives the last bound immediately. □

Theorem 18 *For any $D_2 < D_1$, there exists an A2RSC-achievable vector (R_1, R_2, D_1, D_2) with $L_1 \le \min\{0.5, 0.5\log 2\pi e\sigma^2 - h(X)\}$, and $L_2 \le 0.5\log 2\pi e\sigma^2 - h(X)$.*

Proof: From the SLB, $I(X;U_2) = R(D_2) \geq h(X) - 0.5\log 2\pi e D_2$. From (6.3), there exists an A2RSC-achievable vector (R_1, R_2, D_1, D_2) with $L_1 \leq 0.5$ and

$$\begin{aligned} L_2 &= I(X;Y_1) + I(X;\Delta Y) + I(Y_1;\Delta Y|X) - I(X;U_2) \\ &\leq \frac{1}{2}\log\frac{\sigma^2}{D_1} + \frac{1}{2}\log\frac{\sigma^2}{\sigma^2 - D_1 + D_2} + \frac{1}{2}\log\frac{D_1}{D_2}\left(1 - \frac{D_1 - D_2}{\sigma^2}\right) - h(X) + \frac{1}{2}\log 2\pi e D_2 \\ &= 0.5\log 2\pi e\sigma^2 - h(X). \end{aligned}$$

Similarly, L_1 is bounded by $I(X;Y_1) - I(X;U_1) \leq 0.5\log 2\pi e\sigma^2 - h(X)$. □

An outer bound on the achievable region for A2RSC also appears in [36]. 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, $L_1 = I(X;Y_1) - R(D_1) \leq 0.5$. Since $I(X;\Delta Y) = 0$, $L_2 = I(X;Y_1, \Delta Y) - R(D_2) = I(X;Y_2) - R(D_2) \leq 0.5$.

Chapter 7

Summary and Topics for Future Studies

In this thesis, I improve the existing rate loss bounds for 2RSCs and derive rate loss bounds for TSVQs, 2DSCs, 2ASCs, and A2RSCs for general iid sources and the mse distortion measure. The results include both constant bounds and bounds that depend on either the differential entropy of the source or the rate loss of a Gaussian source of the same variance.

The rate loss bounds are useful both because they characterize the potential penalties associated with using these network source codes and because they provide new inner bounds on the MRSC-, MDSC-, MASC-, and AMRSC-achievable rate-distortion regions. In the MDSC, MASC, and AMRSC cases, the achievable regions are not yet fully known. In all cases, these new inner bounds can be easily analyzed for any source for which we can find the rate-distortion bound. The results are quite tight in some cases.

While most cases lead to small, constant bounds, I believe that these bounds are not tight in general. For example, in the proof of Theorem 2, I show that for 2RSC, the rate loss

L_1 is bounded by two functions of D_1/D_2, which results in a constant bound. However, the minimum of these two bounds is 1 bps at $D_2/D_1 = 0$. But the asymptotic bound obtained by Lastras and Berger is only 0.5 bps as $D_2 \to 0$. Similarly, as stated at the end of Chapter 3, Theorem 6 *suggests* that the performance degradation associated with using greedily grown TSVQs rather than the jointly optimized multi-resolution vector quantizers may be very large. Proving or disproving this intuition would require a tight bound. So, one immediate future direction is to find tighter bounds in all of the scenarios.

For MDSC, it is also important to find a better outer bound for the MDSC-achievable rate-distortion region in region $\mathcal{D}_2$ so that we can have a more accurate estimate on the the rate loss. My conjecture is that the inner bounds found in this thesis, for example,

$$\begin{cases} R_1 \leq R(D_1), \\ R_2 \leq R(D_2), \\ R_1 + R_2 \leq R(D_0) + \min\{L_{G0} + 1.5, R(D_1) + 1, R(D_2) + 1\} \end{cases}$$

or

$$\begin{cases} R_1 \leq R(D_1), \\ R_2 \leq R(D_2), \\ R_1 + R_2 \leq R(D_1) + R(D_2) + L_{G12} + 1, \end{cases}$$

are very tight in the sense that they may be within a few bps of the exact achievable region. Therefore, it is very possible that there exists a better outer bound which is very close to the inner bounds in this thesis.

Similarly, for MASC, though I prove that lossy MASCs can do almost as well as joint encoding, it would be nice to give a better estimate on the achievable region. It can be shown that

$$
\begin{aligned}
R_1 &\geq R_{X|Y}(D_1), \\
R_2 &\geq R_{Y|X}(D_2), \\
R_1 + R_2 &\geq R_{X,Y}(D_1, D_2),
\end{aligned}
$$

is an outer bound on the achievable region (e.g., [35]). The total rate $(R_1 + R_2)$ of the inner bound obtained in this thesis is very close to the outer bound (within 1 bps). In contrast, the distance between the individual rate (R_1 or R_2) and the above outer bound ($R_{X|Y}(D_1)$ or $R_{Y|X}(D_2)$) may be large since Gray shows that

$$R_{X,Y}(D_1, D_2) \geq R_{X|Y}(D_1) + R_Y(D_2)$$

in [48]. My conjecture is that the inner bound is quite tight in the sense that the distances between that inner bound and the individual rates at an optimal (R_1, R_2) may be bounded by small constants. It is likely that we can show a better outer bound such that the distance between the inner and outer bounds is always bounded by a small constant, e.g., 1 bps.

I believe that the most interesting direction for future work is to generalize the concept of rate loss to more network source codes such as broadcast system source codes (BSSCs). Figure 7.1 shows a two-receiver BSSC (2BSSC). In these systems, the sender (node 1) transmits messages to two receivers (nodes 2 and 3). The transmitter at node 1 sends "private" information X and Y to receivers at nodes 2 and 3 respectively and "common" information Z to both receivers. Let $\hat{X}$ and $\hat{Y}$ be the reproductions of X and Y at nodes 2 and 3 respectively, $\hat{Z}_2$ and $\hat{Z}_3$ be the reproductions of Z at nodes 2 and 3 respectively. Gray and Wyner derive the achievable region for this code in [54]. It will be interesting to compare the rates for X, Y, and Z to the corresponding rate-distortion functions at the same distortions. The resulting rate loss for BSSCs would measure the performance degradation of this code. I believe the concept of rate loss is of great potential interest.

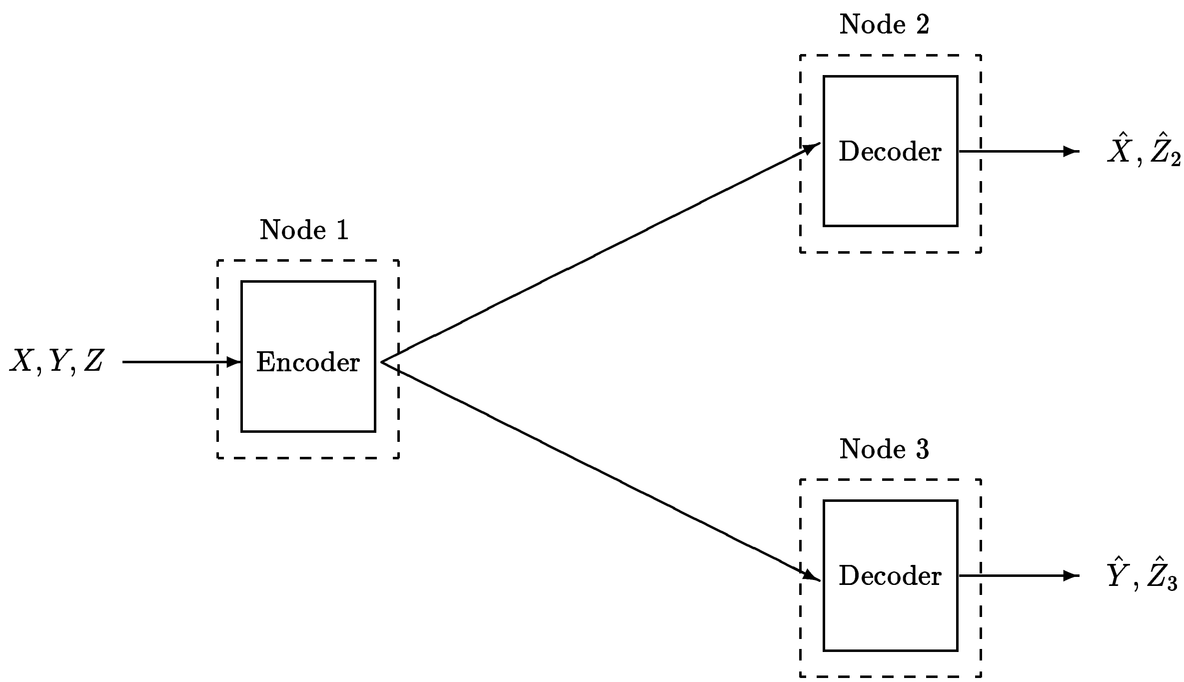

Figure 7.1: A 2BSSC.

Chapter 8

Appendix

Lemma 1 *For any* (D_1, D_2) *with* $D_2 < D_1$*, if two 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* $0 \leq \alpha \leq 1$*, 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* $\beta \in \{0, \alpha, 1\}$*.*

Proof: Following [31, Lemma 2], if $(R_{10}, R_{20}, D_1, D_2)$ and $(R_{11}, R_{21}, D_1, D_2)$ 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 multi-resolution 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 [31, 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 [31, Lemma 2]. The corresponding total rate is $R_{1\alpha} + \Delta R_\alpha = \alpha R_{11} + (1-\alpha)R_{10} + \alpha\Delta R_1 + (1-\alpha)\Delta R_0 = \alpha R_{21} + (1-\alpha)R_{20} = R_{2\alpha}$, 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. □

Lemma 4 *If* $0 < D_1, D_2 \leq \sigma^2$, *then*

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

Proof: I show the lemma by the following inequalities:

$$\begin{aligned} (d_1 + d_2 - 1)\left(\frac{1}{d_1} + \frac{1}{d_2} - 1\right) &= (d_1 + d_2 - 1)\frac{d_1 + d_2 - d_1 d_2}{d_1 d_2} \\ &= \frac{(d_1 + d_2)^2 - (d_1 + d_2)(1 + d_1 d_2) + d_1 d_2}{d_1 d_2} \\ &= \frac{-(d_1 + d_2)(1 - d_1)(1 - d_2) + d_1 d_2}{d_1 d_2} \\ &\leq 1, \end{aligned}$$

where the last inequality comes from the fact that $0 < d_1, d_2 \leq 1$. □

Lemma 5 *In* $\mathcal{D}_2$, $L_{G12} = 0.5\log(d_1 d_2) - 0.5\log(d_0) + L_{G0}$.

Proof: first, I can show that

$$\begin{aligned} & \left[(1-d_0)^2 d_1 d_2 - \left(d_0\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)}\right)^2\right] \\ & - \left[d_0(1-d_0)^2 - d_0\left(\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)}\right)^2\right] \\ = \; & (1-d_0)^2 d_1 d_2 - d_0^2(1-d_1)(1-d_2) - (d_1-d_0)(d_2-d_0) - d_0(1-d_0)^2 \\ & + d_0(1-d_1)(1-d_2) + d_0(d_1-d_0)(d_2-d_0) \\ = \; & (1-d_0)^2 d_1 d_2 + d_0(1-d_0)(1-d_1)(1-d_2) - d_0(1-d_0)^2 - (1-d_0)(d_1-d_0)(d_2-d_0) \end{aligned}$$

$$= \quad (1-d_0)\left[(1-d_0)d_1d_2 + d_0(1-d_1)(1-d_2) - d_0(1-d_0) - (d_1-d_0)(d_2-d_0)\right]$$

$$= \quad 0,$$

where the last equation comes from the fact that

$$(1-d_0)d_1d_2 - d_0(1-d_0) + d_0(1-d_1)(1-d_2) = d_1d_2 + d_0^2 - d_0d_1 - d_0d_2 = (d_1-d_0)(d_2-d_0).$$

Therefore,

$$\begin{aligned} L_{G12} &= \frac{1}{2}\log\frac{(1-d_0)^2d_1d_2}{(1-d_0)^2d_1d_2 - (d_0\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)})^2} \\ &= \frac{1}{2}\log\frac{(1-d_0)^2d_1d_2}{d_0(1-d_0)^2 - d_0(\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)})^2} \\ &= \frac{1}{2}\log\frac{d_1d_2}{d_0} + \frac{1}{2}\log\frac{(1-d_0)^2}{(1-d_0)^2 - (\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)})^2} \\ &= \frac{1}{2}\log\frac{d_1d_2}{d_0} + L_{G0}, \end{aligned}$$

and the lemma is proved. □

Lemma 6 *Suppose* $R(D)$ *is the rate-distortion function of an arbitrary iid source and* $0 < D_0 < D_1, D_2 \le \sigma^2$, *then*

$$R(D_1) + R(D_2) \le R(D_0), \quad \text{if } D_1 + D_2 - D_0 \ge \sigma^2.$$

Proof: Since

$$D_1 = \frac{\sigma^2 - D_1}{\sigma^2 - D_0}D_0 + \left(1 - \frac{\sigma^2 - D_1}{\sigma^2 - D_0}\right)\sigma^2,$$

from the convexity of the rate-distortion function,

$$R(D_1) \le \frac{\sigma^2 - D_1}{\sigma^2 - D_0}R(D_0) + \left(1 - \frac{\sigma^2 - D_1}{\sigma^2 - D_0}\right)R(\sigma^2) = \frac{\sigma^2 - D_1}{\sigma^2 - D_0}R(D_0).$$

By symmetry,

$$R(D_2) \le \frac{\sigma^2 - D_2}{\sigma^2 - D_0}R(D_0).$$

Thus,

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

which is non-positive if $D_1 + D_2 - D_0 \geq \sigma^2$. □

Lemma 7 *Suppose $R(D)$ is the rate-distortion function of an arbitrary iid source and $0 < D_0 < D_1, D_2 \leq \sigma^2$, then*

$$R(D_1) + R(D_2) \leq R(D_0) + 0.5,$$

if $D_1 + D_2 - \sigma^2 < D_0 < (1/D_1 + 1/D_2 - 1/\sigma^2)^{-1}$ and $\max\{D_1, D_2\} \geq \sigma^2/2$.

Proof: Without loss of generality, I assume $D_1 \geq \sigma^2/2$. From the Shannon's upper bound for the rate-distortion function,

$$R(D_1) \leq \frac{1}{2} \log \frac{\sigma^2}{D_1} \leq 0.5.$$

Since $R(D)$ is a non-increasing function of D and $D_2 \geq D_0$, $R(D_2) \leq R(D_0)$. Thus, $R(D_1) + R(D_2) \leq 0.5 + R(D_0)$. □

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

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

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

$$Ed(X,Y_2) = E(\alpha(X-U_1)-N_2)^2 = \alpha^2 E(X-U_1)^2 + D_2 - \alpha^2 D_1 \leq \alpha^2 D_1 + D_2 - \alpha^2 D_1 = D_2.$$

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

$$\begin{aligned}
R(D_2) - R(D_1) &\leq I(X;Y_2) - I(X;U_1) \\
&= I(X;Y_2|U_1) - I(X;U_1|Y_2) \qquad (8.1)\\
&\leq I(X;Y_2|U_1) \\
&= I(X;(1-\alpha)X + \alpha U_1 + N_2|U_1) \\
&= I(X-U_1;(1-\alpha)(X-U_1) + N_2|U_1) \qquad (8.2)\\
&\leq I(X-U_1;(1-\alpha)(X-U_1) + N_2) \qquad (8.3)\\
&\leq \sup I(W;(1-\alpha)W + N_2) \qquad (8.4)\\
&\leq \frac{1}{2}\log\frac{D_1}{D_2}, \qquad (8.5)
\end{aligned}$$

where (8.1) follows by applying the chain rule twice to $I(X;Y_2,U_1)$, (8.2) follows since $h(A|B) = h(A-B|B)$, and (8.3) 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 (8.4), I 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.

Lemma 3 *For any $D_M < \ldots < D_3 < D_2 < D_1$,*

$$I(X;U_1,U_2',U_3',\ldots,U_M'|X+N_M) \leq M/2,$$

where $U_1, U_2', U_3', \ldots, U_{M-1}'$ *are defined in Theorem 6.*

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

$$I(X; X+N_1|U_1) - I(X; U_1|X+N_1) = I(X; X+N_1) - I(X; U_1) = I(X; X+N_1) - R(D_1) \geq 0.$$

Thus, $I(X; U_1|X + N_1) \leq I(X; X + N_1|U_1) \leq 1/2$ by the argument from steps (8.2)-(8.5).

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

$$I(X; U_1, U_2', \ldots, U_k'|X + N_k) \leq k/2, \tag{8.6}$$

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 $N_{k+1} \sim \mathcal{N}(0, D_{k+1})$, $N_k' \sim \mathcal{N}(0, D_k - D_{k+1})$, $N_k' \perp\!\!\!\perp N_{k+1}$ and $(N_k', N_{k+1}) \perp\!\!\!\perp (X, U_1, U_2', \ldots, U_{k+1}')$. By the chain rule,

$$\begin{aligned} & I(X; U_1, U_2', \ldots, U_{k+1}'|X + N_{k+1}) \\ = \;& I(X; U_1, U_2', \ldots, U_k'|X + N_{k+1}) + I(X; U_{k+1}'|U_1, U_2', \ldots, U_k', X + N_{k+1}). \end{aligned} \tag{8.7}$$

I bound the first term on the right hand side of (8.7) as:

$$\begin{aligned} & I(X; U_1, U_2', \ldots, U_k'|X + N_{k+1}) \\ = \;& h(U_1, U_2', \ldots, U_k'|X + N_{k+1}) - h(U_1, U_2', \ldots, U_k'|X, X + N_{k+1}) \\ = \;& h(U_1, U_2', \ldots, U_k'|X + N_{k+1}, X + N_{k+1} + N_k') - h(U_1, U_2', \ldots, U_k'|X) && (8.8) \\ \leq \;& h(U_1, U_2', \ldots, U_k'|X + N_{k+1} + N_k') - h(U_1, U_2', \ldots, U_k'|X) && (8.9) \\ = \;& h(U_1, U_2', \ldots, U_k'|X + N_{k+1} + N_k') - h(U_1, U_2', \ldots, U_k'|X, X + N_{k+1} + N_k') && (8.10) \\ = \;& I(X; U_1, U_2', \ldots, U_k'|X + N_{k+1} + N_k') \\ = \;& I(X; U_1, U_2', \ldots, U_k'|X + N_k) \end{aligned}$$

$$\leq \; k/2, \tag{8.11}$$

where (8.8) and (8.10) follow since $(U_1, U_2', \ldots, U_k') \to X \to X + N_{k+1} \to X + N_{k+1} + N_k'$ forms a Markov chain, (8.9) follows since conditioning reduces differential entropy, and (8.11) follows from the previous assumption (8.6).

The second term on the right hand side of (8.7) can be bounded as follows:

$$
\begin{aligned}
& I(X; U_{k+1}' | U_1, U_2', \ldots, U_k', X + N_{k+1}) \\
& \quad \leq \; I(X; X + N_{k+1} | U_1, U_2', \ldots, U_k', U_{k+1}') \qquad (8.12) \\
& \quad = \; I(X - U_{k+1}'; X - U_{k+1}' + N_{k+1} | U_1, U_2', \ldots, U_k', U_{k+1}') \\
& \quad \leq \; I(X - U_{k+1}'; X - U_{k+1}' + N_{k+1}) \qquad (8.13) \\
& \quad \leq \; 1/2, \qquad (8.14)
\end{aligned}
$$

where (8.12) 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

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

(8.13) 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 (8.14) parallels the chain of inequalities in steps (8.4)-(8.5).

Combining (8.7), (8.11) and (8.14) proves this lemma for $M = k+1$. Finally, by induction, this lemma holds for any $M \geq 1$. □

Lemma 8 *For any* $(D_0, D_1, D_2) \in \mathcal{D}_1 \bigcup \mathcal{D}_2$, $\Gamma^2 \leq \sigma_1^2 \sigma_2^2$.

Proof: In $\mathcal{D}_1$ and $\mathcal{D}_2$,

$$\begin{aligned}
&\beta_0 D_2 D_1 - \beta_1 D_2 D_0 - \beta_2 D_1 D_0 \\
&= D_2 D_1 - \frac{D_2 D_1 D_0}{\sigma^2} - D_2 D_0 + \frac{D_2 D_1 D_0}{\sigma^2} - D_1 D_0 + \frac{D_2 D_1 D_0}{\sigma^2} \\
&= D_2 D_1 - D_2 D_0 - D_1 D_0 + \frac{D_2 D_1 D_0}{\sigma^2} \\
&= D_2 D_1 D_0 \left(\frac{1}{D_0} + \frac{1}{\sigma^2} - \frac{1}{D_1} - \frac{1}{D_2} \right) \\
&> 0.
\end{aligned}$$

As a consequence,

$$\begin{aligned}
\Gamma^2 &= \frac{\beta_1^2 \beta_2^2 D_0^2}{\beta_0^2} \\
&\leq \frac{\beta_1^2 \beta_2^2 D_0^2}{\beta_0^2} + \frac{\beta_1 \beta_2}{\beta_0} (\beta_0 D_2 D_1 - \beta_1 D_2 D_0 - \beta_2 D_1 D_0) \\
&= \left(\beta_1 D_1 - \frac{\beta_1^2 D_0}{\beta_0} \right) \left(\beta_2 D_2 - \frac{\beta_2^2 D_0}{\beta_0} \right) \\
&= \sigma_1^2 \sigma_2^2.
\end{aligned}$$

This proves the lemma. □

Lemma 9 *In $\mathcal{D}_2$, $0 < L_{G0} < L_{\min}(d_1, d_2) < L_{G0} + 1$.*

Proof: From the definition of $\mathcal{D}_2$,

$$\frac{1}{d_0} > \frac{1}{d_1} + \frac{1}{d_2} - 1$$

and $0 < d_0 < d_1, d_2 < 1$ (otherwise, for example, if $d_1 = 1$, then $d_1 + d_2 - 1 = d_2 > d_0$, which contradicts the definition of $\mathcal{D}_2$), which implies

$$d_0^2 (1 - d_1)(1 - d_2) < d_1 d_2 - d_2 d_0 - d_1 d_0 + d_0^2, \tag{8.15}$$

i.e.,

$$d_0 \sqrt{(1 - d_1)(1 - d_2)} \quad < \quad \sqrt{(d_1 - d_0)(d_2 - d_0)},$$

or equivalently, $\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)} \;<\; (1-d_0)\sqrt{(1-d_1)(1-d_2)}$.

On the other hand, $\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)} > 0$ since $1+d_0 > d_1+d_2$ for any $(D_0, D_1, D_2) \in \mathcal{D}_2$. Therefore,

$$\begin{aligned}
1 \;&<\; \frac{(1-d_0)^2}{(1-d_0)^2 - (\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)})^2} \\
&<\; \frac{(1-d_0)^2}{(1-d_0)^2 - (1-d_0)^2(1-d_1)(1-d_2)} \\
&=\; \frac{1}{1-(1-d_1)(1-d_2)} \\
&=\; \frac{1}{d_1+d_2-d_1d_2} \\
&<\; \frac{1}{\max\{d_1, d_2\}},
\end{aligned}$$

where the last inequality is from the fact that $0 < d_1, d_2 < 1$. Thus, $0 < L_{G0} < L_{\min}(d_1, d_2)$.

To prove the other inequality, I first assume $d_2 \le d_1$ without loss of generality. Note that for fixed d_0, d_1, $\sqrt{(1-d_1)(1-d_2)} - \sqrt{(d_1-d_0)(d_2-d_0)}$ is a positive and monotonically decreasing function of d_2 in $\mathcal{D}_2$, therefore

$$\begin{aligned}
L_{G0} \;&\ge\; \frac{1}{2}\log\frac{(1-d_0)^2}{(1-d_0)^2 - (\sqrt{(1-d_1)(1-d_1)} - \sqrt{(d_1-d_0)(d_1-d_0)})^2} \\
&=\; \frac{1}{2}\log\frac{(1-d_0)^2}{(1-d_0)^2-(1+d_0-2d_1)^2} \\
&=\; \frac{1}{2}\log\frac{(1-d_0)^2}{4(1-d_1)(d_1-d_0)},
\end{aligned}$$

thus,

$$\begin{aligned}
L_{\min}(d_1, d_2) - L_{G0} \;&\le\; \frac{1}{2}\log\frac{1}{d_1} - \frac{1}{2}\log\frac{(1-d_0)^2}{4(1-d_1)(d_1-d_0)} \\
&=\; 1 - \log(1-d_0) + \frac{1}{2}\log\frac{(1-d_1)(d_1-d_0)}{d_1} \\
&=\; 1 - \log(1-d_0) + \frac{1}{2}\log\left[1 + d_0 - \left(d_1 + \frac{d_0}{d_1}\right)\right]
\end{aligned}$$

$$\begin{aligned} &\leq 1-\log(1-d_0)+\frac{1}{2}\log(1+d_0-2\sqrt{d_0}) && (8.16)\\ &= 1-\log(1-d_0)+\log(1-\sqrt{d_0})\\ &< 1, \end{aligned}$$

where (8.16) follows from the arithmetic/geometric mean inequality $a+b \geq 2\sqrt{ab}$ for any non-negative a and b, and the last inequality comes from the fact that $0 < d_0 < 1$ thus $d_0 < \sqrt{d_0}$. □

Lemma 10 *In* $\mathcal{D}_2$, $-\sqrt{\beta_1\beta_2 D_1 D_2} < \Gamma' < 0$.

Proof: From (8.15), we can see that

$$d_0^2\beta_1\beta_2 < (d_1-d_0)(d_2-d_0).$$

Therefore, $\Gamma' < 0$. On the other hand, since $0 < d_0 < d_1, d_2 < 1$ in $\mathcal{D}_2$,

$$\begin{aligned} & 2\sqrt{\beta_1\beta_2 d_1 d_2} > 0 > d_1(d_2-1)+d_2(d_1-1) = 2d_1d_2-d_1-d_2,\\ \Rightarrow\ & 2d_0(1-d_0)\sqrt{\beta_1\beta_2 d_1 d_2} > 2d_1d_2d_0-2d_1d_2d_0^2-d_1d_0+d_1d_0^2-d_2d_0+d_2d_0^2,\\ \Rightarrow\ & (1-d_0)^2 d_1d_2+(1-d_1)(1-d_2)d_0^2+2d_0(1-d_0)\sqrt{\beta_1\beta_2 d_1 d_2} > (d_1-d_0)(d_2-d_0),\\ \Rightarrow\ & (1-d_0)^2 d_1d_2+2d_0(1-d_0)\sqrt{\beta_1\beta_2 d_1 d_2}+\beta_1\beta_2 d_0^2 > (d_1-d_0)(d_2-d_0) > 0,\\ \Rightarrow\ & (1-d_0)\sqrt{d_1d_2}+d_0\sqrt{\beta_1\beta_2} > \sqrt{(d_1-d_0)(d_2-d_0)}. \end{aligned}$$

As a consequence,

$$\Gamma' = \frac{\beta_1\beta_2 D_0-\sqrt{\beta_1\beta_2(D_1-D_0)(D_2-D_0)}}{\beta_0} > \frac{\sqrt{\beta_1\beta_2}}{\beta_0}[-(1-d_0)\sqrt{D_1D_2}] = -\sqrt{\beta_1\beta_2 D_1 D_2},$$

which concludes the proof. □

Lemma 11 *In* $\mathcal{D}_2$, $\alpha_1^2\beta_1 D_1+\alpha_2^2\beta_2 D_2+2\alpha_1\alpha_2\Gamma' = \beta_0 D_0$.

Proof: I define $D_{10} = D_1 - D_0$, $D_{20} = D_2 - D_0$, and $\Lambda = \sqrt{\beta_1 D_{20}} + \sqrt{\beta_2 D_{10}}$, then the numerator of α_1 is:

$$\beta_0\beta_1 D_2 - \beta_0\Gamma' = \beta_0\beta_1 D_2 - \beta_1\beta_2 D_0 + \sqrt{\beta_1\beta_2 D_{10} D_{20}} = \beta_1 D_{20} + \sqrt{\beta_1\beta_2 D_{10} D_{20}} = \Lambda\sqrt{\beta_1 D_{20}}.$$

By symmetry, $\beta_0\beta_2 D_1 - \beta_0\Gamma' = \Lambda\sqrt{\beta_2 D_{10}}$. Further $\beta_2 D_1 + \beta_1 D_2 - 2\Gamma' = \Lambda^2/\beta_0$. Thus,

$$\begin{aligned} \alpha_1 &= \frac{\beta_0(\beta_1 D_2 - \Gamma')}{\beta_1(\beta_2 D_1 + \beta_1 D_2 - 2\Gamma')} = \frac{\beta_0\sqrt{\beta_1 D_{20}}}{\beta_1 \Lambda}, \\ \alpha_2 &= \frac{\beta_0(\beta_2 D_1 - \Gamma')}{\beta_2(\beta_2 D_1 + \beta_1 D_2 - 2\Gamma')} = \frac{\beta_0\sqrt{\beta_2 D_{10}}}{\beta_2 \Lambda}. \end{aligned}$$

Since

$$\Gamma' = \frac{\beta_1\beta_2 D_0 - \sqrt{\beta_1\beta_2 D_{10} D_{20}}}{\beta_0} = \frac{\sqrt{\beta_1\beta_2}}{\beta_0}\left[D_0\sqrt{\beta_1\beta_2} - \sqrt{D_{10} D_{20}}\right],$$

I have

$$\begin{aligned} &\alpha_1^2\beta_1 D_1 + \alpha_2^2\beta_2 D_2 + 2\alpha_1\alpha_2\Gamma' \\ &= \frac{\beta_0^2 D_1 D_{20} + \beta_0^2 D_2 D_{10} + 2\beta_0^2\sqrt{\beta_1\beta_2 D_{10} D_{20}}\Gamma'/(\beta_1\beta_2)}{\Lambda^2} \\ &= \frac{\beta_0}{\Lambda^2}\left[\beta_0 D_1 D_{20} + \beta_0 D_2 D_{10} + 2D_0\sqrt{\beta_1\beta_2 D_{10} D_{20}} - 2D_{10} D_{20}\right] \\ &= \frac{\beta_0}{\Lambda^2}\left[\beta_1 D_0 D_{20} + \beta_2 D_0 D_{10} + 2D_0\sqrt{\beta_1\beta_2 D_{10} D_{20}}\right] \qquad (8.17) \\ &= \beta_0 D_0, \end{aligned}$$

where (8.17) comes from the fact that

$$\beta_0 D_1 - D_{10} = (1 - D_0/\sigma^2)D_1 - (D_1 - D_0) = D_0(1 - D_1/\sigma^2) = \beta_1 D_0,$$

and the last equation follows from the fact that $\beta_1 D_0 D_{20} + \beta_2 D_0 D_{10} + 2D_0\sqrt{\beta_1\beta_2 D_{10} D_{20}} = D_0(\sqrt{\beta_1 D_{20}} + \sqrt{\beta_2 D_{10}})^2 = D_0\Lambda^2$. □

Lemma 12 *In $\mathcal{D}_3$, $\alpha_1'^2\beta_1 D_1 + \alpha_2'^2\beta_2 D_2 = \beta_0 D_0$.*

Proof:

$$
\begin{aligned}
&\alpha_1'^2\beta_1 D_1 + \alpha_2'^2\beta_2 D_2 \\
= &\left[\frac{\beta_0\beta_1 D_2 + \sqrt{\Delta}}{\beta_1(\beta_2 D_1 + \beta_1 D_2)}\right]^2 \beta_1 D_1 + \left[\frac{\beta_0\beta_2 D_1 - \sqrt{\Delta}}{\beta_2(\beta_2 D_1 + \beta_1 D_2)}\right]^2 \beta_2 D_2 \\
= &\left[\frac{\beta_0^2\beta_1^2 D_2^2 + 2\beta_0\beta_1 D_2\sqrt{\Delta} + \Delta}{\beta_1(\beta_2 D_1 + \beta_1 D_2)^2}\right] D_1 + \left[\frac{\beta_0^2\beta_2^2 D_1^2 - 2\beta_0\beta_2 D_1\sqrt{\Delta} + \Delta}{\beta_2(\beta_2 D_1 + \beta_1 D_2)^2}\right] D_2 \\
= &\frac{\beta_0^2 D_1 D_2(\beta_1 D_2 + \beta_2 D_1) + \Delta(\beta_1 D_2 + \beta_2 D_1)/\beta_1\beta_2}{(\beta_2 D_1 + \beta_1 D_2)^2} \\
= &\frac{\beta_0^2 D_1 D_2 + \beta_0(D_0 D_2 + D_0 D_1 - D_1 D_2 - d_0 D_1 D_2)}{\beta_2 D_1 + \beta_1 D_2} \\
= &\beta_0 \frac{D_0 D_2 + D_0 D_1 - 2 d_0 D_1 D_2}{\beta_2 D_1 + \beta_1 D_2} \\
= &\beta_0 D_0 \frac{D_2 + D_1 - 2 D_1 D_2/\sigma^2}{\beta_2 D_1 + \beta_1 D_2} \\
= &\beta_0 D_0,
\end{aligned}
$$

where the last equation comes from the fact that $D_2 - D_1 D_2/\sigma^2 = D_2 - d_1 D_2 = \beta_1 D_2$, and $D_1 - D_1 D_2/\sigma^2 = \beta_2 D_1$. □

Lemma 13 *If* $0 < \delta_0 < \delta_1 \le \delta_2 < 0.5$, *then*

$$LG(\delta_0, \delta_1, \delta_2) > \frac{1}{2}\log\frac{1}{4\delta_2(1-\delta_2)}.$$

Proof: First, since $0.5 < 1 - \delta_2 \le 1 - \delta_1 < 1$ and $0 < \delta_2 - \delta_0 \le \delta_1 - \delta_0 < 0.5$, I have $\sqrt{(1-\delta_1)(1-\delta_2)} - \sqrt{(\delta_1-\delta_0)(\delta_2-\delta_0)} > 0$. Thus $LG(\delta_0, \delta_1, \delta_2)$ is a decreasing function of δ_1 for fixed δ_0 and δ_2. Since $\delta_1 \le \delta_2$,

$$
\begin{aligned}
LG(\delta_0, \delta_1, \delta_2) &= \frac{1}{2}\log\frac{(1-\delta_0)^2}{(1-\delta_0)^2 - (\sqrt{(1-\delta_1)(1-\delta_2)} - \sqrt{(\delta_1-\delta_0)(\delta_2-\delta_0)})^2} \\
&\ge \frac{1}{2}\log\frac{(1-\delta_0)^2}{(1-\delta_0)^2 - (\sqrt{(1-\delta_2)(1-\delta_2)} - \sqrt{(\delta_2-\delta_0)(\delta_2-\delta_0)})^2} \\
&= \frac{1}{2}\log\frac{(1-\delta_0)^2}{4(1-\delta_2)(\delta_2-\delta_0)}.
\end{aligned}
$$

Taking a partial derivative of the argument inside the logarithm gives

$$\frac{\partial}{\partial \delta_0}\left(\frac{(1-\delta_0)^2}{4(1-\delta_2)(\delta_2-\delta_0)}\right) = \frac{(1-\delta_0)(1+\delta_0-2\delta_2)}{4(1-\delta_2)(\delta_2-\delta_0)^2}.$$

Since $\delta_0 < \delta_2 < 1/2$ and $1+\delta_0 > 2\delta_2$, this derivative is always positive. Therefore,

$$\begin{aligned} LG(\delta_0,\delta_1,\delta_2) &> \frac{1}{2}\log \left.\frac{(1-\delta_0)^2}{4(1-\delta_2)(\delta_2-\delta_0)}\right|_{\delta_0=0} \\ &= \frac{1}{2}\log \frac{1}{4(1-\delta_2)\delta_2}, \end{aligned}$$

which gives us the desired result. $\square$

Bibliography

[1] H. Feng and M. Effros. Separable karhunen loeve transforms for the weighted universal transform coding algorithm. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, AZ, March 1999. IEEE.

[2] H. Feng and M. Effros. On the rate-distortion performance and computational efficiency of the Karhunen-Loève transform for lossy data compression. *IEEE Transactions on Image Processing*, 11:113–122, February 2002.

[3] H. Feng and M. Effros. A klt can be arbitrarily worse than the optimal transform for transform coding. In *Proceedings of Conference on Information Sciences and Systems*, Princeton, NJ, March 2002.

[4] M. Effros, H. Feng, and K. Zeger. Suboptimality of the Karhunen-Loève transform for transform coding. Submitted to the *IEEE Transactions on Information Theory*, April 2002.

[5] H. Feng and M. Effros. Multi-resolution channel codes. In *Proceedings of the IEEE International Symposium on Information Theory*, Sorrento, Italy, June 2000. IEEE.

[6] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423,623–656, 1948.

[7] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *IRE National Convention Record*, volume 4, pages 142–163, 1959.

[8] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[9] R. J. McEliece. *The Theory of Information and Coding*, volume 3 of *Encylcopedia of Mathematics and its Applications.* Addison-Wesley, Reading, Mass., 1977.

[10] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, 1991.

[11] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, Boston, 1992.

[12] M. Fleming, Q. Zhao, and M. Effros. Network vector quantization. Submitted to the *IEEE Transactions on Information Theory*, January 2002.

[13] H. Feng and M. Effros. Improved bounds for the rate loss of multi-resolution source codes. In *Proceedings of the IEEE International Symposium on Information Theory*, Washington, D.C., June 2001. IEEE.

[14] H. Feng and M. Effros. On the rate loss of multiple description source codes and additive successive refinement codes. In *Proceedings of the IEEE International Symposium on Information Theory*, Palais de Beaulieu, Switzerland, June 2002. IEEE. To appear.

[15] H. Feng and M. Effros. Improved bounds for the rate loss of multi-resolution source codes. Submitted to the *IEEE Transactions on Information Theory*, March 2002.

[16] H. Feng and M. Effros. On the rate loss of multiple description and multiple access source codes. Submitted to the *IEEE Transactions on Information Theory*, May 2002.

[17] H. Feng and M. Effros. On the achievable region of multiple description source codes for Gaussian sources. In preparation.

[18] R. Zamir. The rate loss in the Wyner-Ziv problem. *IEEE Transactions on Information Theory*, 42(6):2073–2084, November 1996.

[19] A. D. Wyner. The rate-distortion function for source coding with side information at the decoder–ii: General sources. *Information & Control*, 38(1):60–80, July 1978.

[20] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22:1–10, January 1976.

[21] E. A. Riskin and R. M. Gray. A greedy tree growing algorithm for the design of variable rate vector quantizers. *IEEE Transactions on Signal Processing*, 39:2500–2507, November 1991.

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

[23] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41:3445–3462, December 1993.

[24] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. on Circuits and Syst. Video Technol.*, 6(3):243–250, June 1996.

[25] M. Effros. Practical multi-resolution source coding: TSVQ revisited. In *Proceedings of the Data Compression Conference*, pages 53–62, Snowbird, UT, March 1998. IEEE.

[26] M. Effros and D. Dugatkin. Multi-resolution vector quantization. Submitted to the *IEEE Transactions on Information Theory*, October 2000. In review.

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

[28] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Transactions on Information Theory*, 37:269–275, March 1991.

[29] J. Chow and T. Berger. Failure of successive refinement for symmetric Gaussian mixtures. *IEEE Transactions on Information Theory*, 43:350–352, January 1997.

[30] B. Rimoldi. Successive refinement of information: Characterization of achievable rates. *IEEE Transactions on Information Theory*, 40:253–259, January 1994.

[31] M. Effros. Distortion-rate bounds for fixed- and variable-rate multiresolution source codes. *IEEE Transactions on Information Theory*, 45:1887–1910, September 1999.

[32] L. Lastras and T. Berger. All sources are nearly successively refinable. *IEEE Transactions on Information Theory*, 47(3):918–926, March 2000.

[33] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973.

[34] S. Y. Tung. *Multiterminal Source Coding*. Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1978.

[35] R. Zamir and T. Berger. Multiterminal source coding with high resolution. *IEEE Transactions on Information Theory*, 45(1):106–117, January 1999.

[36] E. Tuncel and K. Rose. Additive successive refinement. In *Proceedings of the IEEE International Symposium on Information Theory*, Washington, D.C., June 2001. IEEE.

[37] A. A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, 28(6):851–857, November 1982.

[38] M. Fleming and M. Effros. The rate distortion region for the multiple description problem. In *Proceedings of the IEEE International Symposium on Information Theory*, Sorrento, Italy, June 2000. IEEE.

[39] A. Buzo, A. H. Gray Jr., R .M. Gray, and J .D. Markel. Speech coding based upon vector quantization. *IEEE Transactions on Acoustics Speech and Signal Processing*, 28:562–574, October 1980.

[40] R. Venkataramani, G. Kramer, and V. K. Goyal. Bounds on the achievable region for certain multiple description coding problems. In *Proceedings of the IEEE International Symposium on Information Theory*, Washington, D.C., June 2001. IEEE.

[41] Z. Zhang and T. Berger. New results for multiple descriptions. *IEEE Transactions on Information Theory*, 33:502–521, July 1987.

[42] L. Ozarow. On a source-coding problem with two channels and three receivers. *Bell System Technical Journal*, 59(10):1909–1921, December 1980.

[43] R. Ahlswede. The rate-distortion region for multiple description without excess rate. *IEEE Transactions on Information Theory*, 31(6):721–726, November 1985.

[44] R. Zamir. Gaussian codes and shannon bounds for multiple descriptions. *IEEE Transactions on Information Theory*, 45(7):2629–2636, November 1999.

[45] Z. Zhang and T. Berger. Multiple description source coding with no excess marginal rate. *IEEE Transactions on Information Theory*, 41(2):349–357, March 1995.

[46] J. Goblick Jr. *Coding for a discrete information source with a distortion measure*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1963.

[47] R. M. Gray. Conditional rate-distortion theory. Technical Report 6502-2, Information Systems Lab, Stanford Univ., October 1972.

[48] R. M. Gray. A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions. *IEEE Transactions on Information Theory*, 19(4):480–489, July 1973.

[49] T. Berger. *Multiterminal source coding*. The Information Theory Approach to Communications. Springer-Verlag, New York, 1977.

[50] T. Berger and S. Y. Tung. Encoding of correlatedanalog sources. In *Proceedings of IEEE-USSR Joint Workshop on Information Theory*, pages 7–10, 1975.

[51] K. B. Housewright. *Source coding studies for multiterminal systems.* Ph.D. dissertation, University of California, Los Angeles, Los Angeles, CA, 1977.

[52] J. K. Omura and K. B. Housewright. Source coding studies for information networks. In *Proceedings of the IEEE International Conference on Communication*, Chicago, IL, 1977. IEEE.

[53] E. Tuncel and K. Rose. Personal communication, March 2002.

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