On the Asymptotic Rate-Distortion Function of Multiterminal Source Coding Under Logarithmic Loss

On the Asymptotic Rate-Distortion Function of Multiterminal Source Coding Under Logarithmic Loss

By Yanning Li,

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Master of Applied Science

McMaster University

Master of Applied Science (2021)

Hamilton, Ontario (Department of Electrical and Computer Engineering)

TITLE: On the Asymptotic Rate-Distortion Function of Multiterminal Source Coding Under Logarithmic Loss

AUTHOR: Yanning Li (McMaster University)

SUPERVISOR: Dr. Jun Chen

NUMBER OF PAGES: vii, 49

Abstract

We consider the asymptotic minimum rate under the logarithmic loss distortion constraint. More specifically, we find the asymptotic minimum rate expression when given distortions get close to 0. The problem under consideration is separate encoding and joint decoding of correlated two information sources, subject to a logarithmic loss distortion constraint. We introduce a test channel, whose transition probability (conditional probability mass function) captures the encoding and decoding process. Firstly, we find the expression for the special case of doubly symmetric binary sources with binary-output test channels. Then the result is extended to the case where the test channels are arbitrary. When given distortions get close to 0, the asymptotic rate coincides with that for the aforementioned special case. Finally, we consider the general case and show that the key findings for the special case continue to hold.

Key words: Multiterminal source coding, rate-distortion theory, logarithmic loss.

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my supervisor Dr. Jun Chen for his assistance at every stage of the research project. He guided me to a completely new theoretical research field, letting me know theoretical research could also be lively and interesting. And Dr. Chen is always very patient with me, especially during writing this thesis. He kindly gave me very precious advice and suggestions. It is my great luck to have him as my supervisor.

I would also like to thanks Dr. Sorina Dumitrescu for her valuable technical suggestions on this project. And I should also appreciate Dr.Dongmei Zhao for her insightful comments to help me understand my project more deeply.

In addition, I would like to thank all the staff in ECE department. It is their kind help and support that have made my study and life in the McMaster University a wonderful time.

Furthermore, I would like to express gratitude to Jingjing Qian for her treasured tutorial and support which was really influential in shaping my research methods. And I am so grateful to my classmates and roommates for a cherished time spent together.

Finally, my deep and sincere gratitude to my parents for their love, support and encouragement. I am forever indebted to my parents for giving me the opportunities and experiences that have made me who I am. And I wish to thank my friends for their company and for bringing joy and hope in my life over the years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Data Compression

Data compression is a method that reduces the amount of data without losing useful information. Its main purpose is to improve the efficiency of transmission, storage, and processing. In computer science and information theory, data compression or source coding is the process of representing information in fewer bits than the original representation. It could be broadly classified into two classes called lossless compression and lossy compression.

Lossless compression preserves all the information in the data being compressed, and the reconstruction is identical to the original data[1]. It is necessary for text, where every character is important.

By contrast, lossy data compression allows losing detail or introducing small errors upon the reversal in exchange for better compression rates. It may be acceptable for images or voice, where we can sacrifice the quality of images or voice to decrease the file size.

Furthermore, rate–distortion theory offers the theoretical basis for lossy data compression. We shall give a brief review of this theory.

1.2 Rate-Distortion Theory

Rate-distortion theory, also known as rate-distortion source coding theory, is a theory that studies data compression by using the basic viewpoints and methods of information theory.

The basic problem in rate-distortion theory can be stated as follows: Given a source distribution and a distortion measure, what is the minimum expected distortion achievable at a particular rate? Or, equivalently, what is the minimum rate description required to achieve a particular distortion[2]?

Apparently, there are two important elements in this theory. First is the source distribution, and the second is the distortion measure. Distortion measure is a measure of distance between a random variable and its representation. Mathematically, any norm or distance is a measure of distortion. But in choosing a specific distortion measure one should take into account the physical meaning and calculation convenience.

In rate-distortion theory, the encoding and decoding process is succinctly represented by a test channel with a suitably chosen transition probability (conditional probability mass function).

1.3 Multiterminal Source Coding

Multiterminal (MT) source coding refers to separate encoding and joint decoding of multiple correlated sources. The fundamental problem here is to characterize the optimal tradeoff between the compression rates and the reconstruction distortions. Slepian and Wolf first formulated the lossless case of the multiterminal source coding problem and solved it in [3]. Then this result was extended to the lossy case. Ahlswede-Körner[4] and Wyner[5] solved the problem of source coding with side information; Wyner-Ziv[6] first characterized rate-distortion function of source coding with side information at the decoder; Berger-Tung [7], [8] provided the best known region of achievable rates for the multiterminal source coding problem. And Berger-Yeung[9], [10], extended the Wyner-Ziv problem to a more general form.

In 1996, Berger et al. defined a particular formulation of multiterminal source coding, known as the Chief Executive Officer (CEO) problem[11]. In this problem, there are ℓ separate encoders, which observe independently corrupted versions of a source; these encoders compress their respective observations and forward the compressed data separately to a central decoder, which then produces a (lossy) reconstruction of the target source. The fundamental question is to obtain a computable characterization of the tradeoff between the encoder rates and the reconstruction distortions[12].

Later, more researches were conducted on this problem by choosing specific source distribution or specific distortion measure. In particular, there are a large number of papers devoted to the quadratic Gaussian version of the CEO problem.

Recently, logarithmic loss distortion measure has become more and more popular in multiterminal source coding. It has nice mathematical properties and is often referred to as self-information loss in the literature on prediction. Logarithmic loss plays a central role in settings in which reconstructions are allowed to be ‘soft’, rather than ‘hard’ or deterministic. That is, rather than just assigning a deterministic value to each sample of the source, the decoder also gives an assessment of the degree of confidence or reliability on each estimate, in the form of weights or probabilities[13].

Besides, logarithmic loss also has an important place in information theory, where many of the fundamental quantities (e.g., entropy, relative entropy, etc.) can be interpreted as the optimal prediction risk or regret under logarithmic loss[14]. There are also many research papers conducted on lossy source coding problems with logarithmic loss distortion[12, 15–18].

1.4 Distributed Source Coding

Distributed source coding (DSC) is an important problem in information theory and communication. DSC problems regard the compression of multiple correlated information sources that do not communicate with each other by exploiting that the receiver can perform joint decoding of the encoded signals[19]. There are two main properties in DSC, first is that the computational burden in encoders is shifted to the joint decoder, making the encoding calculation very simple and the decoding calculation relatively complex. Secondly, DSC theory proves that independent encoding can in fact be designed as efficiently as joint encoding, as long as joint decoding is allowed.

The main application fields of distributed source coding include sensor network and image, video and multimedia compression. Traditional image source coding algorithms, such as video encoding standards MPEG-X and H.26X or still image encoding standards JPEG2000, extract the statistical correlation of the source at the encoder for compression, and the operation complexity of the encoder is higher than that of the decoder. With the development of electronic technology, some emerging applications such as wireless video sensor networks and camera arrays have developed rapidly. Due to the limited resources and power consumption of the encoder, these new applications are not suitable for adopting traditional image source encoding algorithms, and pose new challenges to traditional image encoding algorithms and system architectures.

Different from traditional image coding algorithms, distributed source coding transfers the correlation extraction work from the encoder to the decoder, and the computational complexity of the encoder is greatly reduced. Because of its unique advantages, DSC has become a research hot-spot in recent years.

1.5 Thesis Structure

This thesis is organized as follows: Chapter 1 introduces the background and related works. Chapter 2 defines the problem. Chapter 3 gives the three main results of the problem. Theorem 1 is obtained in a special case and proved in Chapter 4. Chapter 5 gives the proof of Theorem 2, which is an extension of Theorem 1. Chapter 6 shows the proof of the general result in Theorem 3. Chapter 7 offers the numerical verification test for the conclusion. Finally is the conclusion of the work. A list of references is provided at the end of the thesis.

Chapter 2

Problem Definitions

Consider a communication system consisting of two distributed information sources. Let X_1^n, X_2^n denote the sequences of the sources. Suppose the distribution of X_1 is known, and the joint pmf of X_1 and X_2 is given as $p(x_1, x_2) = p(x_1) \cdot p(x_2|x_1)$. Note that X_1^n and X_2^n are encoded as U_1 and U_2, and $U_1 \leftrightarrow X_1^n \leftrightarrow X_2^n \leftrightarrow U_2$ form a Markov Chain in that order. That is, the joint pmf $p(x_1, x_2, u_1, u_2) = p(x_1) \cdot p(x_2|x_1) \cdot p(u_1|x_1) \cdot p(u_2|x_2)$. U_1 and U_2 are sent to the decoder, where $\hat{X}_1^n$ and $\hat{X}_2^n$ are reconstructed by using (U_1, U_2). This coding system is shown in Fig.2.1.

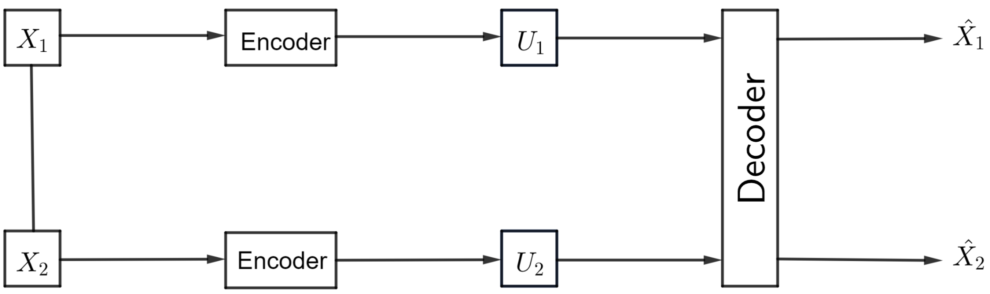

Figure 2.1: Coding system model

Each encoder consists of a function $f_i, (i = 1, 2)$,

$$f_i(X_i^n) = U_i, \tag{2.1}$$

where $X_i^n \in \mathcal{X}_i$ and $U_i \in \mathcal{U}_i$, for i =1, 2. And decoding functions are $g_i, (i = 1, 2)$, mapping U_1 and U_2 to the reconstructions $\hat{X}_1^n$ and $\hat{X}_2^n$,

$$g_i(U_1, U_2) = \hat{X}_i^n, \tag{2.2}$$

where $(U_1, U_2) \in \mathcal{U}_1 \times \mathcal{U}_2$, for i =1, 2.

Definition 1. *The Logarithmic loss distortion measure*

The logarithmic loss distortion between a source symbol x_j and a probability distribution $\hat{x_j}$ on $\mathcal{X}$ is defined as follows:

$$d(x_j, \hat{x_j}) = log(\frac{1}{\hat{x_j}(x_j)}), \qquad j = 1, 2, \ldots n, \tag{2.3}$$

where $\hat{x}(\cdot)$ designates a probability distribution on $\mathcal{X}$ and $\hat{x}(x)$ is the value of this distribution evaluated for the outcome $x \in \mathcal{X}$. And $\hat{x_j}(x_j)$ generally depends on (u_1, u_2). Throughout this thesis, the logarithm is the natural logarithm, with the base of the mathematical constant e. With this definition for symbol-wise distortion, we can easily define the total value of log-loss distortion between a sequence of symbols x_i^n and a sequence of distributions $\hat{x_i^n}$ as:

$$d(x_i^n, \hat{x_i^n}) = \frac{1}{n}\sum_{j=1}^{n} log(\frac{1}{\hat{x_j}(x_j)}), \qquad i = 1, 2. \tag{2.4}$$

Definition 2.

A rate distortion vector (R_1, R_2, D_1, D_2) is called strict-sense achievable for a distortion measure $d(\cdot,\cdot)$, if there exist encoding functions f_1, f_2 and decoding functions g_1, g_2 according to $Eq.2.1$ and $Eq.2.2$ such that for length n,

$$
\begin{aligned}
R_i &\geq \frac{1}{n} log|\mathcal{U}_i|, && for\ i = 1,2, \\
D_i &\geq \mathbb{E}d(X_i^n, \hat{X}_i^n), && for\ i = 1,2,
\end{aligned} \tag{2.5}
$$

where $\mathbb{E}(\cdot)$ denotes expectation function, $\hat{X}_i^n = g_i(f_1(X_1^n), f_2(X_2^n))$.

Definition 3.

The achievable rate-distortion region for a source is the closure of the set of all strict-sense achievable vectors (R_1, R_2, D_1, D_2), denoted by $\overline{\mathcal{RD}}^*$. Furthermore, we denote $\mathcal{RD}^i$ as the inner bound and $\mathcal{RD}^o$ as the outer bound of the rate-distortion region.

According to [15, Definition 3 and Theorem 1], $(R_1, R_2, D_1, D_2) \in \mathcal{RD}^i$ if and only if there exists a joint distribution of the form

$$
p(x_1)p(x_2|x_1)p(u_1|x_1)p(u_2|x_2), \tag{2.6}
$$

where $|\mathcal{U}_1| \geq |\mathcal{X}_1|$, $|\mathcal{U}_2| \geq |\mathcal{X}_2|$, which satisfies

$$\begin{aligned} R_1 &\geq I(X_1; U_1|U_2), \\ R_2 &\geq I(X_2; U_2|U_1), \\ R \triangleq R_1 + R_2 &\geq I(X_1, X_2; U_1, U_2), \\ D_1 &\geq H(X_1|U_1, U_2), \\ D_2 &\geq H(X_2|U_1, U_2). \end{aligned} \tag{2.7}$$

According to [15, Theorem 3], we have the following proposition.

Proposition 1.

$$\overline{\mathcal{RD}}^* = \mathcal{RD}^i = \mathcal{RD}^o. \tag{2.8}$$

Our problem is to find the minimum rate R with given distortion D_1 and D_2, according to proposition 1, now we can convert the problem into the following optimization problem:

$$\begin{aligned} min \quad & I(X_1, X_2; U_1, U_2), \\ s.t. \quad & D_1 \geq H(X_1|U_1, U_2), \\ & D_2 \geq H(X_2|U_1, U_2). \end{aligned} \tag{2.9}$$

Chapter 3

Main Results

3.1 Theorem 1

Theorem 1 (The minimum rate of binary case under logarithmic loss)**.**

Given D_1 and D_2, let $p(x_2|x_1)$ be a binary symmetric channel with crossover probability q, and let $p(u_1|x_1)$, $p(u_2|x_2)$ be binary-input binary-output channels, then the asymptotic minimum rate as $D_1,\ D_2 \to 0$ is:

$$I(X_1, X_2; U_1, U_2) = H(X_1, X_2) - D_1 - D_2 + I_{min}(X_1; X_2|U_1, U_2) + o(\frac{D_1 D_2}{logD_1 \cdot logD_2}), \tag{3.1}$$

where $I_{min}(X_1; X_2|U_1, U_2)$ is

$$I_{min}(X_1; X_2|U_1, U_2) = min \left\{ \begin{array}{c} \left[\frac{1}{2} \cdot \frac{2q-1}{1-q} + \frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right] \cdot \frac{4D_1D_2}{logD_1 \cdot logD_2}, \\ \left[\frac{1}{2} \cdot \frac{1-2q}{q} + \frac{1}{2} \cdot qlog\frac{q^2}{(1-q)^2}\right] \cdot \frac{4D_1D_2}{logD_1 \cdot logD_2} \end{array} \right\}. \tag{3.2}$$

3.2 Theorem 2

Theorem 2.

Given D_1 and $D_2 \to 0$, there is no change in the asymptotic rate when $|\mathcal{U}_i| > |\mathcal{X}_i|$ compared with the result when $|\mathcal{U}_i| = |\mathcal{X}_i|$. That means $|\mathcal{U}_i|$ could be arbitrarily large, but the asymptotic rate is always equal to the value calculated when $|\mathcal{U}_i| = |\mathcal{X}_i|$.

3.3 Theorem 3

Theorem 3.

Given D_1 and $D_2 \to 0$, suppose the distribution of X_1 is known, and $p(x_1, x_2) = p(x_1) \cdot p(x_2|x_1)$. The source alphabet of X_1 is $\mathcal{X}_1 = \{0, 1, \dots n-1\}$, $|\mathcal{X}_1| = n$, and the source alphabet of X_2 is $\mathcal{X}_2 = \{0, 1, \dots m-1\}$, $|\mathcal{X}_2| = m$, $(m \geq n)$. Given $p(u_1|x_1)$ with the probability $\epsilon_{i_1 i_2}$, $p(u_2|x_2)$ with the probability $\alpha_{j_1 j_2}$, suppose the alphabet of U_1 is $\mathcal{U}_1 = \{0, 1, \dots, u\}$, $|\mathcal{U}_1| = u + 1$, and the alphabet of U_2 is $\mathcal{U}_2 = \{0, 1, \dots, v\}$, $|\mathcal{U}_2| = v+1$. Moreover, $u+1 \geq n$, $v+1 \geq m$, that is $|\mathcal{U}_1| \geq |\mathcal{X}_1|$, $|\mathcal{U}_2| \geq |\mathcal{X}_2|$.

Then the asymptotic minimum rate as D_1, $D_2 \to 0$ is:

$$I(X_1, X_2; U_1, U_2) = H(X_1, X_2) - D_1 - D_2 + I_{min}(X_1; X_2|U_1, U_2) + o(\frac{D_1 D_2}{logD_1 \cdot logD_2}), \tag{3.3}$$

where $I_{min}(X_1; X_2|U_1, U_2)$ is

$$I_{min}(X_1; X_2|U_1, U_2) = min\left\{\frac{D_1D_2}{logD_1 \cdot logD_2} \cdot \frac{r_t}{p(x_1 = i_1 - 1)p(x_2 = j_1 - 1)}\right\}. \tag{3.4}$$

r_t is the coefficient of the cross-term, and the expression of r_t should be provided in the statement of Theorem 3. i_1 is the first subscript of $\epsilon_{i_1 i_2}$, while j_1 is the first subscript of $\alpha_{j_1 j_2}$.

Chapter 4

Proof of Theorem 1

4.1 Basic Part

Suppose that the source alphabets of X_1 and X_2 are just $\{0,1\}$, X_1 is uniformly distributed over $\{0,1\}$. Let $p(x_2|x_1)$ be a binary symmetric channel with crossover probability q, and let $p(u_1|x_1)$, $p(u_2|x_2)$ be binary-input binary-output channels with crossover probabilities ϵ_1 and ϵ_2, α_1 and α_2 respectively. The alphabet of U_i is equal to the sources alphabet X_i, that is $\mathcal{U}_i = \mathcal{X}_i = \{0,1\}$. The model is shown in Fig. 7.1.

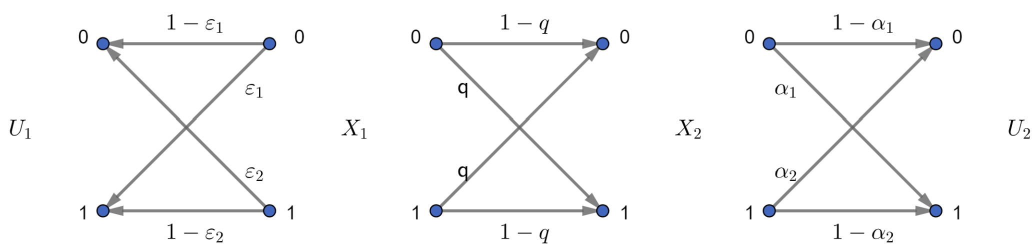

FIGURE 4.1: Binary case

Because X_1 is uniformly distributed, the probability of X_1 is

$$p(x_1) = \begin{cases} \frac{1}{2} & \text{(x=0)}, \\ \frac{1}{2} & \text{(x=1)}. \end{cases} \tag{4.1}$$

Then we can easily get

$$\begin{aligned} p(x_1, x_2) = p(x_1) \cdot p(x_2|x_1) &= \begin{pmatrix} x_1 = 0, x_2 = 0 & x_1 = 0, x_2 = 1 \\ x_1 = 1, x_2 = 0 & x_1 = 1, x_2 = 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}(1-q) & \frac{1}{2}q \\ \frac{1}{2}q & \frac{1}{2}(1-q) \end{pmatrix}, \end{aligned} \tag{4.2}$$

$$p(u_1|x_1) = \begin{pmatrix} u_1 = 0|x_1 = 0 & u_1 = 1|x_1 = 0 \\ u_1 = 0|x_1 = 1 & u_1 = 1|x_1 = 1 \end{pmatrix} = \begin{pmatrix} 1-\epsilon_1 & \epsilon_1 \\ \epsilon_2 & 1-\epsilon_2 \end{pmatrix}, \tag{4.3}$$

$$p(u_2|x_2) = \begin{pmatrix} u_2 = 0|x_2 = 0 & u_2 = 1|x_2 = 0 \\ u_2 = 0|x_2 = 1 & u_2 = 1|x_2 = 1 \end{pmatrix} = \begin{pmatrix} 1-\alpha_1 & \alpha_1 \\ \alpha_2 & 1-\alpha_2 \end{pmatrix}. \tag{4.4}$$

Note that $p(x_1, x_2, u_1, u_2) = p(x_1) \cdot p(x_2|x_1) \cdot p(u_1|x_1) \cdot p(u_2|x_2)$, we can derive $p(x_1, u_1, u_2) = \sum_{x_2} p(x_1, x_2, u_1, u_2)$, $p(x_2, u_1, u_2) = \sum_{x_1} p(x_1, x_2, u_1, u_2)$ and $p(u_1, u_2) = \sum_{x_1, x_2} p(x_1, x_2, u_1, u_2)$.

For the objective function in Eq.2.9, we have

$$\begin{aligned} I(X_1,X_2;U_1,U_2) &= H(X_1,X_2) - H(X_1,X_2|U_1,U_2) \\ &= H(X_1,X_2) - [H(X_1|U_1,U_2) + H(X_2|U_1,U_2) - I(X_1;X_2|U_1,U_2)] \\ &= H(X_1,X_2) - H(X_1|U_1,U_2) - H(X_2|U_1,U_2) + I(X_1;X_2|U_1,U_2). \end{aligned} \tag{4.5}$$

Given $p(x_1,x_2)$, the $H(X_1,X_2)$ is a constant, now let's calculate the rest part in Eq. 4.5. According to the definition of conditional entropy,

$$\begin{aligned} H(X_1|U_1,U_2) &= \sum_{u_1,u_2} p(u_1,u_2) \sum_{x_1} p(x_1|u_1,u_2) log \frac{1}{p(x_1|u_1,u_2)} \\ &= \sum_{x_1,u_1,u_2} p(x_1,u_1,u_2) log \frac{p(u_1,u_2)}{p(x_1,u_1,u_2)} \\ &= \sum_{x_1,u_1,u_2} p(x_1,u_1,u_2) \left[logp(u_1,u_2) - logp(x_1,u_1,u_2) \right]. \end{aligned} \tag{4.6}$$

The Taylor series expansion of $ln(1+x)$ is given by

$$ln(1+x) = x - \frac{1}{2}x^2 + o(x^2), \tag{4.7}$$

so we can take out a common factor in $logp(u_1,u_2)$ and in $logp(x_1,u_1,u_2)$ to construct $ln[factor \cdot (1+x)]$ and then apply the Taylor series expansion. In this way, the asymptotic expression of $H(X_1|U_1,U_2)$ is

$$\begin{aligned} H(X_1|U_1,U_2) = &-\frac{1}{2}\epsilon_1 log\epsilon_1 - \frac{1}{2}\epsilon_2 log\epsilon_2 \\ &+\frac{1}{2}\left[1+(1-2q)log\frac{q}{1-q}\right]\epsilon_1 + \frac{1}{2}\left[1+(1-2q)log\frac{q}{1-q}\right]\epsilon_2. \end{aligned} \tag{4.8}$$

Similarly, we can also get the asymptotic expression of $H(X_2|U_1,U_2)$

$$\begin{aligned} H(X_2|U_1,U_2) = &-\frac{1}{2}\alpha_1 log\alpha_1 - \frac{1}{2}\alpha_2 log\alpha_2 \\ &+\frac{1}{2}\left[1+(1-2q)log\frac{q}{1-q}\right]\alpha_1 + \frac{1}{2}\left[1+(1-2q)log\frac{q}{1-q}\right]\alpha_2, \end{aligned} \tag{4.9}$$

where we only keep up to the linear terms. According to the definition of conditional mutual information

$$\begin{aligned} I(X_1;X_2|U_1,U_2) &= \sum_{u_1,u_2} p(u_1,u_2) \sum_{x_1,x_2} p(x_1,x_2|u_1,u_2) log\frac{p(x_1,x_2|u_1,u_2)}{p(x_1|u_1,u_2)p(x_2|u_1,u_2)} \\ &= \sum_{x_1,x_2,u_1,u_2} p(x_1,x_2,u_1,u_2) log\frac{p(x_1,x_2,u_1,u_2)p(u_1,u_2)}{p(x_1,u_1,u_2)p(x_2,u_1,u_2)} \\ &= \sum_{x_1,x_2,u_1,u_2} p(x_1,x_2,u_1,u_2)[logp(x_1,x_2,u_1,u_2)p(u_1,u_2) - logp(x_1,u_1,u_2)p(x_2,u_1,u_2)]. \end{aligned} \tag{4.10}$$

With the same method in calculating conditional entropy, take out a common factor in $logp(x_1,x_2,u_1,u_2)p(u_1,u_2)$ and in $logp(x_1,u_1,u_2)p(x_2,u_1,u_2)$ to construct $ln[factor \cdot (1+x)]$ and then apply the Taylor series expansion. Through applying the Taylor expansion, we have

$$I(X_1;X_2|U_1,U_2) = \left[\frac{1}{2}\frac{2q-1}{1-q} + \frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right](\epsilon_1\alpha_1 + \epsilon_2\alpha_2) + \left[\frac{1}{2}\frac{1-2q}{q} + \frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right](\epsilon_1\alpha_2 + \epsilon_2\alpha_1), \quad (4.11)$$

where we retain up to the quadratic terms and drop the higher-order terms.

4.2 Optimization Part

Given D_1 and D_2, the optimization problem is

$$\begin{aligned} min \quad & I(X_1, X_2; U_1, U_2), \\ s.t. \quad & D_1 \geq H(X_1|U_1, U_2), \\ & D_2 \geq H(X_2|U_1, U_2). \end{aligned} \quad (4.12)$$

Substituting Eq.4.5 into Eq.4.12, we have

$$\begin{aligned} min \quad & H(X_1, X_2) - H(X_1|U_1, U_2) - H(X_2|U_1, U_2) + I(X_1; X_2|U_1, U_2), \\ s.t. \quad & D_1 \geq H(X_1|U_1, U_2), \\ & D_2 \geq H(X_2|U_1, U_2). \end{aligned} \quad (4.13)$$

As D_1, $D_2 \to 0$, $H(X_1|U_1, U_2)$ and $H(X_2|U_1, U_2)$ are also close to 0. Then we can simplify the expressions of $H(X_1|U_1, U_2)$ and $H(X_2|U_1, U_2)$ in Eq.4.8, Eq.4.9 to

$$H(X_1|U_1, U_2) = -\frac{1}{2}\epsilon_1 log\epsilon_1 - \frac{1}{2}\epsilon_2 log\epsilon_2, \quad (4.14)$$

$$H(X_2|U_1, U_2) = -\frac{1}{2}\alpha_1 log\alpha_1 - \frac{1}{2}\alpha_2 log\alpha_2. \tag{4.15}$$

Since $H(X_1, X_2)$ is a constant, we do not consider it in the optimization function. Substituting Eq.4.11, Eq.4.14, Eq.5.11 into the optimization function 4.13, we find that the objective is a function of ϵ_1, ϵ_2, α_1, and α_2.

Furthermore, comparing the order of $H(X_1|U_1, U_2)$, $H(X_2|U_1, U_2)$ and $I(X_1; X_2|U_1, U_2)$, it turns out that $H(X_1|U_1, U_2)$, $H(X_2|U_1, U_2)$ are much greater than $I(X_1; X_2|U_1, U_2)$. That is, the conditional entropy of X_1 and X_2 are the dominant terms of the objective function. Therefore, to minimize the objective function, our main target is to minimize $-H(X_1|U_1, U_2)$ and $-H(X_2|U_1, U_2)$. Note that $-H(X_1|U_1, U_2)$ and $-H(X_2|U_1, U_2)$ achieve their minimum values $-D_1$ and $-D_2$ respectively when constraints are active.

Now the original optimization problem Eq.4.13 is converted to the following optimization problem:

$$\begin{aligned} min \quad & I(X_1; X_2|U_1, U_2), \\ s.t. \quad & D_1 = H(X_1|U_1, U_2), \\ & D_2 = H(X_2|U_1, U_2). \end{aligned} \tag{4.16}$$

Substitute Eq.4.11 into Eq.4.16:

$$\begin{aligned} min \qquad & \left[\frac{1}{2}\frac{2q-1}{1-q} + \frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right](\epsilon_1\alpha_1 + \epsilon_2\alpha_2) \\ & + \left[\frac{1}{2}\frac{1-2q}{q} + \frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right](\epsilon_1\alpha_2 + \epsilon_2\alpha_1), \\ s.t. \qquad & D_1 = -\frac{1}{2}\epsilon_1 log\epsilon_1 - \frac{1}{2}\epsilon_2 log\epsilon_2, \\ & D_2 = -\frac{1}{2}\alpha_1 log\alpha_1 - \frac{1}{2}\alpha_2 log\alpha_2. \end{aligned} \tag{4.17}$$

Let's introduce a coefficient k, which represents the ratio of $-\frac{1}{2}\epsilon_1 log\epsilon_1$ to D_1, $k \in [0,1]$. Similarly, introduce a coefficient b, which represents the ratio of $-\frac{1}{2}\alpha_1 log\alpha_1$ to D_2, $b \in [0,1]$. Then we have the following equations:

$$\begin{cases} -\frac{1}{2}\epsilon_1 log\epsilon_1 = kD_1, \\ -\frac{1}{2}\epsilon_2 log\epsilon_2 = (1-k)D_1, \end{cases} \tag{4.18}$$

$$\begin{cases} -\frac{1}{2}\alpha_1 log\alpha_1 = bD_2, \\ -\frac{1}{2}\alpha_2 log\alpha_2 = (1-b)D_2. \end{cases} \tag{4.19}$$

By solving the system of equations 4.18,4.19, we can get the solutions of ϵ_1 and ϵ_2, α_1 and α_2

$$\begin{cases} \epsilon_1 = -\frac{2kD_1}{log(2kD_1)} = -\frac{2kD_1}{log(2k)+logD_1}, \\ \epsilon_2 = -\frac{2(1-k)D_1}{log[2(1-k)D_1]} = -\frac{2(1-k)D_1}{log[2(1-k)]+logD_1}, \end{cases} \tag{4.20}$$

$$\begin{cases} \alpha_1 = -\frac{2bD_2}{log(2bD_2)} = -\frac{2bD_2}{log(2b)+logD_2}, \\ \alpha_2 = -\frac{2(1-b)D_2}{log[2(1-b)D_2]} = -\frac{2(1-b)D_2}{log[2(1-b)]+logD_2}. \end{cases} \tag{4.21}$$

If $k \neq 0$ and $k \neq 1$, when $D_1 \to 0$, $logD_1 \to -\infty$, so we can ignore $log(2k)$ and $log[2(1-k)]$,

$$\begin{cases} \epsilon_1 \approx -\frac{2kD_1}{logD_1}, \\ \epsilon_2 \approx -\frac{2(1-k)D_1}{logD_1}. \end{cases} \tag{4.22}$$

Similarly, if $b \neq 0$ and $b \neq 1$, when $D_2 \to 0$, $logD_2 \to -\infty$, we can also get the approximate solutions of α_1 and α_2:

$$\begin{cases} \alpha_1 \approx -\frac{2bD_2}{logD_2}, \\ \alpha_2 \approx -\frac{2(1-b)D_2}{logD_2}. \end{cases} \tag{4.23}$$

The optimization problem can be written as:

$$\begin{aligned} min \quad & \left[\frac{1}{2}\frac{2q-1}{1-q} + \frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right] \cdot \frac{4D_1D_2}{logD_1 \cdot logD_2} \cdot kb \\ & + \left[\frac{1}{2}\frac{1-2q}{q} + \frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right] \cdot \frac{4D_1D_2}{logD_1 \cdot logD_2} \cdot k(1-b) \\ & + \left[\frac{1}{2}\frac{1-2q}{q} + \frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right] \cdot \frac{4D_1D_2}{logD_1 \cdot logD_2} \cdot (1-k)b \\ & + \left[\frac{1}{2}\frac{2q-1}{1-q} + \frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right] \cdot \frac{4D_1D_2}{logD_1 \cdot logD_2} \cdot (1-k)(1-b), \\ s.t. \quad & 0 < k < 1, \\ & 0 < b < 1. \end{aligned} \tag{4.24}$$

Because the products kb, $k(1-b)$, $(1-k)b$, $(1-k)(1-b)$ are all in the interval $(0,1)$ and the sum of them equals 1, these products can be regarded as the weight of each term in the total conditional mutual information. In this way, we just need to compare the values of the coefficients, find the minimum coefficient, adjust its corresponding weight w^* to the maximum value of 1, and set other terms' weights to be 0. If

$$\left[\frac{1}{2}\frac{2q-1}{1-q}+\frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right]\cdot\frac{4D_1D_2}{logD_1\cdot logD_2} < \left[\frac{1}{2}\frac{1-2q}{q}+\frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right]\cdot\frac{4D_1D_2}{logD_1\cdot logD_2}, \tag{4.25}$$

then let $kb=1$ or $(1-k)(1-b)=1$, that is $k=1$ and $b=1$ or $k=0$ and $b=0$, and the minimum of the objective function is

$$I(X_1;X_2|U_1,U_2)_{min}=\left[\frac{1}{2}\frac{2q-1}{1-q}+\frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right]\cdot\frac{4D_1D_2}{logD_1\cdot logD_2}. \tag{4.26}$$

If

$$\left[\frac{1}{2}\frac{1-2q}{q}+\frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right]\cdot\frac{4D_1D_2}{logD_1\cdot logD_2} < \left[\frac{1}{2}\frac{2q-1}{1-q}+\frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right]\cdot\frac{4D_1D_2}{logD_1\cdot logD_2}, \tag{4.27}$$

let $k(1-b)=1$ or $(1-k)b=1$, that is $k=1$ and $b=0$ or $k=0$ and $b=1$, then the minimum of the objective function is

$$I(X_1;X_2|U_1,U_2)_{min}=\left[\frac{1}{2}\frac{1-2q}{q}+\frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right]\cdot\frac{4D_1D_2}{logD_1\cdot logD_2}. \tag{4.28}$$

However, note that we assume $k \neq 0$ and $k \neq 1$, $b \neq 0$ and $b \neq 1$ when simplifying the solutions in Eq.4.22, Eq.4.23, so the minimum value of the $I(X_1; X_2|U_1, U_2)$ is obtained when the weight w^* corresponding to the term with the smallest coefficient is close to 1.

To sum up, the asymptotic minimum rate as D_1, $D_2 \to 0$ is:

$$I(X_1, X_2; U_1, U_2) = H(X_1, X_2) - D_1 - D_2 + I_{min}(X_1; X_2|U_1, U_2) + o(\frac{D_1 D_2}{log D_1 \cdot log D_2}), \tag{4.29}$$

where $I_{min}(X_1; X_2|U_1, U_2)$ is

$$I_{min}(X_1; X_2|U_1, U_2) = min \left\{ \begin{array}{c} \left[\frac{1}{2} \cdot \frac{2q-1}{1-q} + \frac{1}{2}(1-q) log \frac{(1-q)^2}{q^2}\right] \cdot \frac{4 D_1 D_2}{log D_1 \cdot log D_2}, \\ \left[\frac{1}{2} \cdot \frac{1-2q}{q} + \frac{1}{2} \cdot q log \frac{q^2}{(1-q)^2}\right] \cdot \frac{4 D_1 D_2}{log D_1 \cdot log D_2} \end{array} \right\}.$$

This completes the proof of Theorem 1.

Chapter 5

Proof of Theorem 2

Here we still consider binary sources X_1, and X_2. Suppose the source alphabets of X_1 and X_2 are $\mathcal{X}_i = \{0, 1\}$, X_1 is uniformly distributed over $\{0, 1\}$. Let $p(x_2|x_1)$ be a binary symmetric channel with crossover probability q. Given $p(u_1|x_1)$, $p(u_2|x_2)$, suppose the alphabet of U_1 is $\mathcal{U}_1 = \{0, 1, \ldots, u\}$, and the alphabet of U_2 is $\mathcal{U}_2 = \{0, 1, \ldots, v\}$. Moreover, $|\mathcal{U}_1| > |\mathcal{X}_1|$, $|\mathcal{U}_2| > |\mathcal{X}_2|$. The model is shown in Fig.5.1

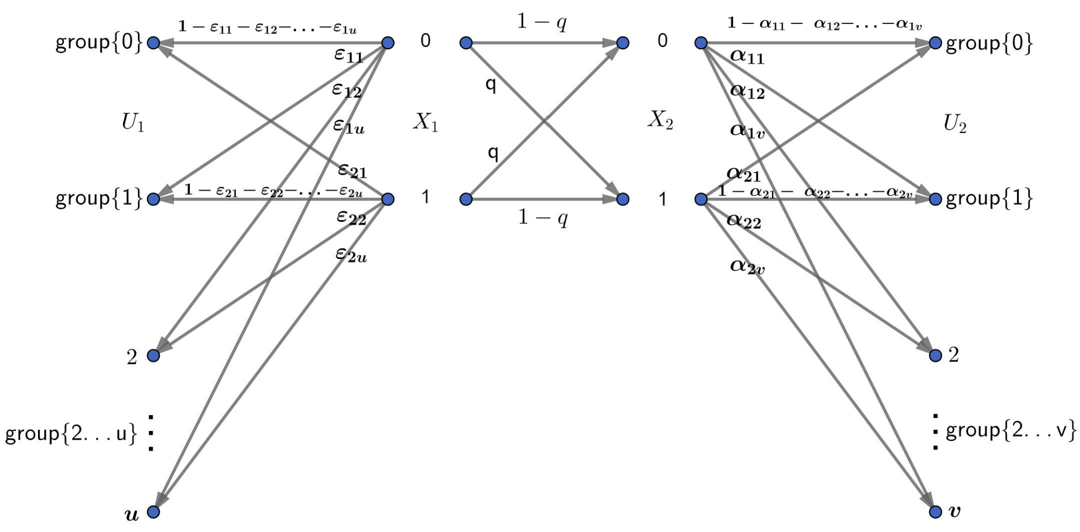

FIGURE 5.1: Binary sources with arbitrarily large U_1 and U_2

Classify the elements in the alphabets $\mathcal{U}_1$ and $\mathcal{U}_2$ into 3 categories respectively. With the pair of (u_1, u_2), those that have a great probability to be reconstructed to 0 are assigned into one category, renamed as group $\{0\}$. Those that have a great probability to be reconstructed to 1 are assigned into one category, renamed as group $\{1\}$. And the remaining elements that can not be reconstructed certainly are assigned into one category, renamed as group $\{2 \dots u\}$ and group$\{2 \dots v\}$, respectively.

We know that given D_1, $D_2 \to 0$, because $D_1 \geq H(X_1|U_1, U_2)$, $D_2 \geq H(X_2|U_1, U_2)$, $H(X_1|U_1, U_2)$, $H(X_2|U_1, U_2)$ also $\to 0$. Recall the definition of conditional entropy,

$$H(X_1|U_1, U_2) = \sum_{u_1, u_2} p(u_1, u_2) \sum_{x_1} p(x_1|u_1, u_2) log \frac{1}{p(x_1|u_1, u_2)}. \tag{5.1}$$

For group $\{0\}$, given the pair of (u_1, u_2), it has a great probability of being reconstructed to 0, So $p(x_1 = 0|u_1, u_2) \approx 1$, $p(x_1 = 1|u_1, u_2) \approx 0$. Hence,

$$\begin{aligned} &\sum_{x_1} p(x_1|u_1, u_2) log \frac{1}{p(x_1|u_1, u_2)} \\ &= p(x_1 = 0|u_1, u_2) log \frac{1}{p(x_1 = 0|u_1, u_2)} + p(x_1 = 1|u_1, u_2) log \frac{1}{p(x_1 = 1|u_1, u_2)} \\ &\approx 0. \end{aligned} \tag{5.2}$$

This means the value of the $p(u_1, u_2)$ could be arbitrary, and the value of $H(X_1|U_1, U_2)$ always $\to 0$. Similarly, $H(X_2|U_1, U_2)$ also always $\to 0$. For group $\{1\}$, the analysis process is the same, so we can get the conclusion that if u_1 and u_2 both belong to group $\{0\}$ or group $\{1\}$, it can be guaranteed that $H(X_1|U_1, U_2) \to 0$, $H(X_1|U_1, U_2)$ $\to 0$. For group $\{2 \dots u\}$ and group$\{2 \dots v\}$ we have the following Lemma.

Lemma 1. *As long as one of u_1 and u_2 belongs to group $\{2 \dots u\}$ or group$\{2 \dots v\}$, then the corresponding $p(u_1, u_2)$ must be close to 0.*

Let's prove it by reductio ad absurdum.

We should make a hypothesis: $p(u_1, u_2)$ is not close to 0.

Then suppose u_1 belongs to group $\{2 \dots u\}$, because we are not sure that the elements in group $\{2 \dots u\}$ could be reconstructed to 0 or 1. $p(x_1 = 0|u_1, u_2)$, $p(x_1 = 1|u_1, u_2)$ are also uncertain. We have

$$\sum_{x_1} p(x_1|u_1, u_2) log \frac{1}{p(x_1|u_1, u_2)} \neq 0. \tag{5.3}$$

According to the hypothesis: $p(u_1, u_2)$ is not close to 0. Then the product of $p(u_1, u_2)$ and $\sum_{x_1} p(x_1|u_1, u_2) log \frac{1}{p(x_1|u_1,u_2)}$ is also not close to 0. Thus, for group $\{2 \dots u\}$, what it contributes to $H(X_1|U_1, U_2)$ is a large value, which makes the total $H(X_1|U_1, U_2)$ bounded away from 0. Similarly, if u_2 belongs to group $\{2 \dots v\}$, we can get the same result that $H(X_2|U_1, U_2)$ is not close to 0.

Evidently, the results contradict with the fact that when D_1, $D_2 \to 0$, $H(X_1|U_1, U_2)$, $H(X_2|U_1, U_2)$ also $\to 0$. Therefore, the hypothesis is not true. We get the conclusion $p(u_1, u_2)$ must be close to 0.

Now let's compare the orders of the values that contribute to $H(X|U_1, U_2)$: when u_1 and u_2 are both in group $\{0\}$ or group $\{1\}$ and when one of u_1 and u_2 belongs to group $\{2 \dots u\}$ or group$\{2 \dots v\}$ or both u_1 and u_2 are in the group $\{2 \dots u\}$ and group$\{2 \dots v\}$.

Because X_1 is uniformly distributed, the probability of X_1 is

$$p(x_1) = \begin{cases} \frac{1}{2} & \text{(x=0)}, \\ \frac{1}{2} & \text{(x=1)}. \end{cases} \tag{5.4}$$

Then we can easily get

$$\begin{aligned} p(x_1, x_2) = p(x_1) \cdot p(x_2|x_1) &= \begin{pmatrix} x_1 = 0, x_2 = 0 & x_1 = 0, x_2 = 1 \\ x_1 = 1, x_2 = 0 & x_1 = 1, x_2 = 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}(1-q) & \frac{1}{2}q \\ \frac{1}{2}q & \frac{1}{2}(1-q) \end{pmatrix}. \end{aligned} \tag{5.5}$$

In this model, we have

$$\begin{aligned} p(u_1|x_1) &= \begin{pmatrix} u_1 = 0|x_1 = 0 & u_1 = 1|x_1 = 0 & u_1 = 2|x_1 = 0 \dots & u_1 = u|x_1 = 0 \\ u_1 = 0|x_1 = 1 & u_1 = 1|x_1 = 1 & u_1 = 2|x_1 = 1 \dots & u_1 = u|x_1 = 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \epsilon_{11} - \cdots - \epsilon_{1u} & \epsilon_{11} & \epsilon_{12} & \dots & \epsilon_{1u} \\ \epsilon_{21} & 1 - \epsilon_{21} - \cdots - \epsilon_{2u} & \epsilon_{22} & \dots & \epsilon_{2u} \end{pmatrix}, \end{aligned} \tag{5.6}$$

$$\begin{aligned} p(u_2|x_2) &= \begin{pmatrix} u_2 = 0|x_2 = 0 & u_2 = 1|x_2 = 0 & u_2 = 2|x_2 = 0 \dots & u_2 = v|x_2 = 0 \\ u_2 = 0|x_2 = 1 & u_2 = 1|x_2 = 1 & u_2 = 2|x_2 = 1 \dots & u_2 = v|x_2 = 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \alpha_{11} - \cdots - \alpha_{1v} & \alpha_{11} & \alpha_{12} & \dots & \alpha_{1v} \\ \alpha_{21} & 1 - \alpha_{21} - \cdots - \alpha_{2v} & \alpha_{22} & \dots & \alpha_{2v} \end{pmatrix}. \end{aligned} \tag{5.7}$$

Since $p(x_1, x_2, u_1, u_2) = p(x_1) \cdot p(x_2|x_1) \cdot p(u_1|x_1) \cdot p(u_2|x_2)$, we can derive $p(x_1, u_1, u_2) = \sum_{x_1} p(x_1, x_2, u_1, u_2)$, $p(x_2, u_1, u_2) = \sum_{x_2} p(x_1, x_2, u_1, u_2)$ and $p(u_1, u_2) = \sum_{x_1, x_2} p(x_1, x_2, u_1, u_2)$.

Given D_1, $D_2 \to 0$, let's calculate $H(X_1|U_1, U_2)$, $H(X_2|U_1, U_2)$, $I(X_1; X_2|U_1, U_2)$ in this case. The asymptotic expression of $H(X_1|U_1, U_2)$ is

$$\begin{aligned} H(X_1|U_1, U_2) = &- \frac{1}{2}\epsilon_{11} log \epsilon_{11} - \frac{1}{2}\epsilon_{21} log \epsilon_{21} \\ &+ \frac{1}{2}\left[1 + (1 - 2q) log \frac{q}{1-q}\right]\epsilon_{11} + \frac{1}{2}\left[1 + (1 - 2q) log \frac{q}{1-q}\right]\epsilon_{21} \\ &+ \frac{1}{2}(\epsilon_{12} + \cdots + \epsilon_{1u} + \epsilon_{21} + \cdots + \epsilon_{2u}), \end{aligned} \tag{5.8}$$

where terms $-\frac{1}{2}\epsilon_{11} log \epsilon_{11} - \frac{1}{2}\epsilon_{21} log \epsilon_{21} + \frac{1}{2}\left[1 + (1 - 2q) log \frac{q}{1-q}\right]\epsilon_{11} + \frac{1}{2}\left[1 + (1 - 2q) log \frac{q}{1-q}\right]\epsilon_{21}$ is induced by (u_1, u_2) which are in group $\{0\}$ and group $\{1\}$. This part is totally the same with the binary case. And the term $\frac{1}{2}(\epsilon_{12} + \cdots + \epsilon_{1u} + \epsilon_{21} + \cdots + \epsilon_{2u})$ is induced by (u_1, u_2) that one of u_1 and u_2 belongs to group $\{2 \ldots u\}$ or group $\{2 \ldots v\}$, and both u_1, u_2 are in group $\{2 \ldots u\}$ and group $\{2 \ldots v\}$.

As D_1, $D_2 \to 0$, we only keep the dominant terms of the $H(X_1|U_1, U_2)$

$$H(X_1|U_1, U_2) = -\frac{1}{2}\epsilon_{11} log \epsilon_{11} - \frac{1}{2}\epsilon_{21} log \epsilon_{21}. \tag{5.9}$$

Thus, when D_1, $D_2 \to 0$, the asymptotic expression of $H(X_1|U_1, U_2)$ is eventually the same as the expression in the binary case.

Similarly, we also calculate the asymptotic $H(X_2|U_1, U_2)$, and it is also equal to the expression in the binary case when D_1, $D_2 \to 0$.

$$\begin{aligned}H(X_2|U_1,U_2) =& -\frac{1}{2}\alpha_{11}log\alpha_{11} - \frac{1}{2}\alpha_{21}log\alpha_{21}\\ &+\frac{1}{2}\left[1+(1-2q)log\frac{q}{1-q}\right]\alpha_{11} + \frac{1}{2}\left[1+(1-2q)log\frac{q}{1-q}\right]\alpha_{21}\\ &+\frac{1}{2}(\alpha_{12}+\cdots+\alpha_{1v}+\alpha_{21}+\cdots+\alpha_{2v}).\end{aligned} \tag{5.10}$$

And when D_1, $D_2 \to 0$, we only keep the dominant terms of $H(X_2|U_1, U_2)$

$$H(X_2|U_1,U_2) = -\frac{1}{2}\alpha_{11}log\alpha_{11} - \frac{1}{2}\alpha_{21}log\alpha_{21}, \tag{5.11}$$

which is the same as the expression of the binary case.

Recall the definition of $I(X_1; X_2|U_1, U_2)$

$$I(X_1;X_2|U_1,U_2) = \sum_{u_1,u_2} p(u_1,u_2) \sum_{x_1,x_2} p(x_1,x_2|u_1,u_2)log\frac{p(x_1,x_2|u_1,u_2)}{p(x_1|u_1,u_2)p(x_2|u_1,u_2)}. \tag{5.12}$$

According to Lemma 1, as long as one of u_1 and u_2 belongs to group $\{2\ldots u\}$ or group$\{2\ldots v\}$, the corresponding $p(u_1, u_2) \approx 0$. Thus, for (u_1, u_2), one or two of them in group $\{2\ldots u\}$ or group$\{2\ldots v\}$, its corresponding $I(X_1; X_2|U_1, U_2) \approx 0$. That means $I(X_1; X_2|U_1, U_2)$ is mainly induced by (u_1, u_2) which are in group $\{0\}$ and group $\{1\}$. This is the same with the binary case. And we can get the following result directly,

$$\begin{aligned} I(X_1;X_2|U_1,U_2) = &\left[\frac{1}{2}\frac{2q-1}{1-q} + \frac{1}{2}(1-q)log\frac{(1-q)^2}{q^2}\right](\epsilon_{11}\alpha_{11}+\epsilon_{21}\alpha_{21}) \\ &+ \left[\frac{1}{2}\frac{1-2q}{q} + \frac{1}{2}qlog\frac{q^2}{(1-q)^2}\right](\epsilon_{11}\alpha_{21}+\epsilon_{21}\alpha_{11}). \end{aligned} \tag{5.13}$$

Now consider the optimization problem

$$\begin{aligned} min \quad & H(X_1,X_2) - H(X_1|U_1,U_2) - H(X_2|U_1,U_2) + I(X_1;X_2|U_1,U_2), \\ s.t. \quad & D_1 \geq H(X_1|U_1,U_2), \\ & D_2 \geq H(X_2|U_1,U_2). \end{aligned} \tag{5.14}$$

With the same $H(X_1|U_1,U_2)$, $H(X_2|U_1,U_2)$, $I(X_1;X_2|U_1,U_2)$ as in the binary case, we can get the same optimization solution. The asymptotic rate when $\mathcal{X}_i = \{0,1\}$, $\mathcal{U}_1 = \{0,1,\ldots,u\}$, $\mathcal{U}_2 = \{0,1,\ldots,v\}$ is same as the rate when $\mathcal{X}_i = \mathcal{U}_i = \{0,1\}$. Therefore, we can ignore the effects of group $\{2\ldots u\}$ and group $\{2\ldots v\}$.

In conclusion, given D_1 and $D_2 \to 0$, there is no change in the asymptotic rate when $|\mathcal{U}_i| > |\mathcal{X}_i|$ compared with the result when $|\mathcal{U}_i| = |\mathcal{X}_i|$. This completes the proof of Theorem 2.

Chapter 6

Proof of Theorem 3

6.1 Basic Part

Given D_1 and $D_2 \to 0$, suppose the distribution of X_1 is known, $p(x_1, x_2) = p(x_1) \cdot p(x_2|x_1)$. The source alphabet of X_1 is $\mathcal{X}_1 = \{0, 1, \dots n-1\}$, $|\mathcal{X}_1| = n$, and the source alphabet of X_2 is $\mathcal{X}_2 = \{0, 1, \dots m-1\}$, $|\mathcal{X}_2| = m$, $(m \geq n)$. Given $p(u_1|x_1)$ with the probability $\epsilon_{i_1 i_2}$, $p(u_2|x_2)$ with the probability $\alpha_{j_1 j_2}$, suppose the alphabet of U_1 is $\mathcal{U}_1 = \{0, 1, \dots, u\}$, $|\mathcal{U}_1| = u + 1$, and the alphabet of U_2 is $\mathcal{U}_2 = \{0, 1, \dots, v\}$, $|\mathcal{U}_2| = v+1$. Moreover, $u+1 \geq n$, $v+1 \geq m$, that is $|\mathcal{U}_1| \geq |\mathcal{X}_1|$, $|\mathcal{U}_2| \geq |\mathcal{X}_2|$.

Classify the elements in the alphabets $\mathcal{U}_1$ into n+1 categories. With the pair of (u_1, u_2), those that have a great probability to be reconstructed to the corresponding X_1 are assigned into n categories, renamed as group $\{0\}$ – group $\{n-1\}$ respectively. And the remaining elements that can not be reconstructed to X_1 certainly are assigned into one category, renamed as group $\{n \dots u\}$. Similarly, we also regroup the alphabets $\mathcal{U}_2$ into m+1 categories. Those that have a great probability to be reconstructed to the corresponding x_2 are assigned into m categories, renamed as group $\{0\}$ – group $\{m-1\}$ respectively. And the remaining elements that can not be reconstructed to x_2 certainly are assigned into one category, renamed as group $\{m \ldots v\}$.

This model is shown in Fig.6.1

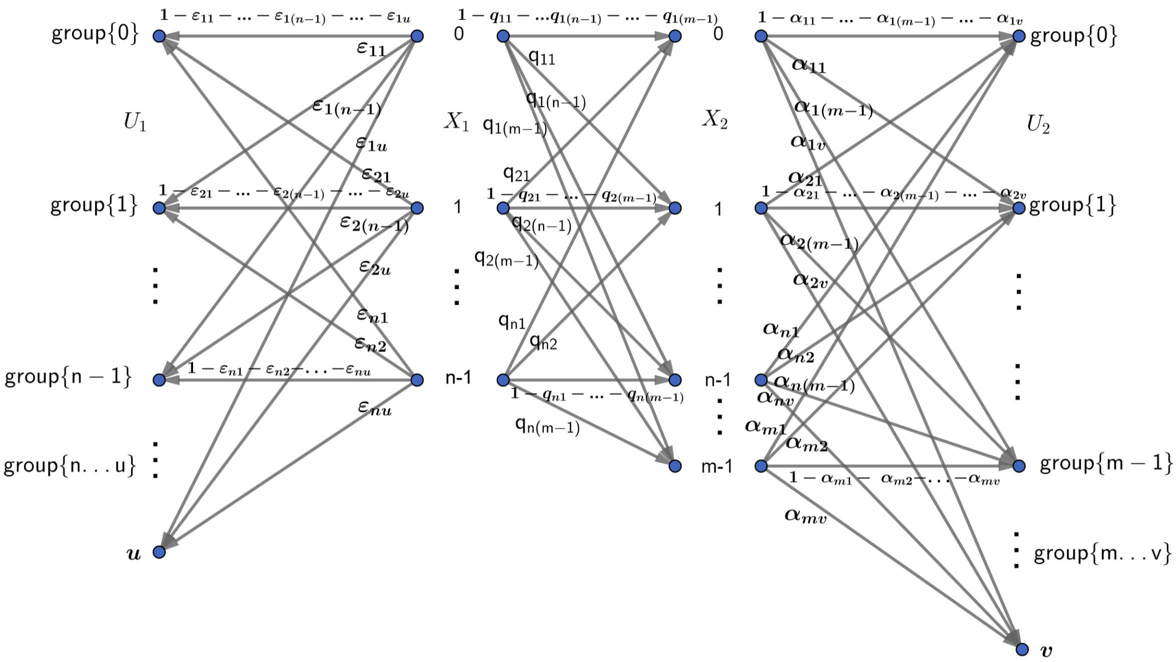

FIGURE 6.1: n dimensional X_1 and m dimensional X_2 with arbitrarily large U_1 and U_2

Because the distribution of X_1 is known, $p(x_1, x_2) = p(x_1) \cdot p(x_2|x_1)$, we can easily get

$$p(x_1, x_2) = \begin{pmatrix} (0,0) & (0,1) & (0,2) & \dots & (0,m-1) \\ (1,0) & (1,1) & (1,2) & \dots & (1,m-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (n-1,0) & (n-1,1) & (n-1,2) & \dots & (n-1,m-1) \end{pmatrix},$$

$$p(x_1, x_2) = \begin{pmatrix} p(x_1{=}0)(1-q_{11}-\ldots q_{1(m-1)}) & p(x_1{=}0)q_{11} & \ldots & p(x_1{=}0)q_{1(m-1)} \\ p(x_1{=}1)q_{21} & p(x_1{=}1)(1-q_{21}-\ldots q_{2(m-1)}) & \ldots & p(x_1{=}1)q_{2(m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ p(x_1{=}n-1)q_{n1} & p(x_1{=}n-1)q_{n2} & \ldots & p(x_1{=}n-1)(1-q_{n1}-\ldots q_{n(m-1)}) \end{pmatrix}_{n\times m}. \tag{6.1}$$

In this model, we have $p(u_1|x_1)$

$$= \begin{pmatrix} u_1=0|x_1=0 & u_1=1|x_1=0 & \dots & u_1=n-1|x_1=0 & \dots & u_1=u|x_1=0 \\ u_1=0|x_1=1 & u_1=1|x_1=1 & \dots & u_1=n-1|x_1=1 & \dots & u_1=u|x_1=1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ u_1=0|x_1=n-1 & u_1=1|x_1=n-1 & \dots & u_1=n-1|x_1=n-1 & \dots & u_1=u|x_1=n-1 \end{pmatrix}_{n\times(u+1)}$$
$$= \begin{pmatrix} 1-\epsilon_{11}-\cdots-\epsilon_{1u} & \epsilon_{11} & \dots & \epsilon_{1(n-1)} & \dots & \epsilon_{1u} \\ \epsilon_{21} & 1-\epsilon_{21}-\cdots-\epsilon_{2u} & \dots & \epsilon_{2(n-1)} & \dots & \epsilon_{2u} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \dots & 1-\epsilon_{n1}-\cdots-\epsilon_{nu} & \dots & \epsilon_{nu} \end{pmatrix}_{n\times(u+1)}, \tag{6.2}$$

and $p(u_2|x_2)$

$$
\begin{aligned}
&= \begin{pmatrix} u_2=0|x_2=0 & u_2=1|x_2=0 & \dots & u_2=m-1|x_2=0 & \dots & u_2=v|x_2=0 \\ u_2=0|x_2=1 & u_2=1|x_2=1 & \dots & u_2=m-1|x_2=1 & \dots & u_2=v|x_2=1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ u_2=0|x_2=m-1 & u_2=1|x_2=m-1 & \dots & u_2=m-1|x_2=m-1 & \dots & u_2=v|x_2=m-1 \end{pmatrix}_{m\times(v+1)} \\
&= \begin{pmatrix} 1-\alpha_{11}-\dots-\alpha_{1v} & \alpha_{11} & \dots & \alpha_{1(m-1)} & \dots & \alpha_{1v} \\ \alpha_{21} & 1-\alpha_{21}-\dots-\alpha_{2v} & \dots & \alpha_{2(m-1)} & \dots & \alpha_{2v} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & 1-\alpha_{m1}-\dots-\alpha_{mv} & \dots & \alpha_{mv} \end{pmatrix}_{m\times(v+1)} .
\end{aligned}
\tag{6.3}
$$

Since $p(x_1,x_2,u_1,u_2) = p(x_1)\cdot p(x_2|x_1)\cdot p(u_1|x_1)\cdot p(u_2|x_2)$, we can derive $p(x_1,u_1,u_2) = \sum_{x_1} p(x_1,x_2,u_1,u_2)$, $p(x_2,u_1,u_2) = \sum_{x_2} p(x_1,x_2,u_1,u_2)$ and $p(u_1,u_2) = \sum_{x_1,x_2} p(x_1,x_2,u_1,u_2)$.

According to Theorem 2, we know that given D_1, $D_2 \to 0$, the effect of the group $\{n\dots u\}$ and group $\{n\dots v\}$ could be ignored, then we have the asymptotic $H(X_1|U_1,U_2)$,

$$
\begin{aligned}
H(X_1|U_1,U_2) = {} & p(x_1=0)\left(-\epsilon_{11}log\epsilon_{11} - \epsilon_{12}log\epsilon_{12} - \dots - \epsilon_{1(n-1)}log\epsilon_{1(n-1)}\right) \\
& + p(x_1=1)\left(-\epsilon_{21}log\epsilon_{21} - \epsilon_{22}log\epsilon_{22} - \dots - \epsilon_{2(n-1)}log\epsilon_{2(n-1)}\right) \\
& \quad \vdots \qquad \vdots \qquad \vdots \qquad \qquad \vdots \\
& + p(x_1=n-1)\left(-\epsilon_{n1}log\epsilon_{n1} - \epsilon_{n2}log\epsilon_{n2} - \dots - \epsilon_{n(n-1)}log\epsilon_{n(n-1)}\right),
\end{aligned}
\tag{6.4}
$$

which consists of $n \times (n-1)$ terms. The asymptotic $H(X_2|U_1, U_2)$ is,

$$\begin{aligned} H(X_2|U_1,U_2) = p(x_2 = 0) &\left(-\alpha_{11} log\alpha_{11} - \alpha_{12} log\alpha_{12} - \cdots - \alpha_{1(m-1)} log\alpha_{1(m-1)}\right) \\ + p(x_2 = 1) &\left(-\alpha_{21} log\alpha_{21} - \alpha_{22} log\alpha_{22} - \cdots - \alpha_{2(m-1)} log\alpha_{2(m-1)}\right) \\ \vdots \quad & \quad \vdots \quad \quad \vdots \quad \quad \vdots \\ + p(x_2 = m-1) &\left(-\alpha_{m1} log\alpha_{m1} - \alpha_{m2} log\alpha_{m2} - \cdots - \alpha_{m(m-1)} log\alpha_{m(m-1)}\right), \end{aligned} \tag{6.5}$$

which consists of $m \times (m-1)$ terms. And the asymptotic $I(X_1; X_2|U_1, U_2)$ is

$$I(X_1; X_2|U_1, U_2) = \sum_{i1=1}^{n} \sum_{i2=1}^{n-1} \sum_{j1=1}^{m} \sum_{j2=1}^{m-1} r_t \cdot \epsilon_{i_1 i_2} \alpha_{j_1 j_2}, \tag{6.6}$$

where $i_1 = 1, 2, \ldots, n$, $i_2 = 1, 2, \ldots, n-1$, $j_1 = 1, 2, \ldots, m$, $j_2 = 1, 2, \ldots, m-1$. r_t is the coefficient of the cross-term, $t = 1, 2, \ldots, n \times (n-1) \times m \times (m-1)$.

For $\epsilon_{i_1 i_2}$, its coordinate in $p(u_1|x_1)$ matrix is

$$(a, b) = \begin{cases} (i_1, i_2) & when \quad i_1 < i_2, \\ (i_1, i_2 + 1) & when \quad i_1 \geq i_2. \end{cases} \tag{6.7}$$

For $\alpha_{j_1 j_2}$, its coordinate in $p(u_2|x_2)$ matrix is

$$(c, d) = \begin{cases} (j_1, j_2) & when \quad j_1 < j_2, \\ (j_1, j_2 + 1) & when \quad j_1 \geq j_2. \end{cases} \tag{6.8}$$

Selecting a and b rows, c and d columns from $p(x_1, x_2)$ matrix, we can get a 2×2 submatrix:

$$\begin{pmatrix} p_{ac} & p_{ad} \\ p_{bc} & p_{bd} \end{pmatrix}. \tag{6.9}$$

Now we have the coefficient r_t :

$$r_t = p_{ac} \cdot log(\frac{p_{ac} \cdot p_{bd}}{p_{ad} \cdot p_{bc}}) - \frac{p_{ac} \cdot p_{bd} - p_{ad} \cdot p_{bc}}{p_{bd}}, \tag{6.10}$$

where $t = 1, 2, \ldots, n \times (n-1) \times m \times (m-1)$.

6.2 Optimization Part

Consider the optimization problem in Eq.4.16:

$$\begin{aligned} min \quad & I(X_1; X_2|U_1, U_2), \\ s.t. \quad & D_1 = H(X_1|U_1, U_2), \\ & D_2 = H(X_2|U_1, U_2). \end{aligned} \tag{6.11}$$

Let's introduce a coefficient k_i, $(i = 1, \ldots, n^2 - n - 1)$, and $1 - k_1 - \cdots - k_{n^2-n-1}$ to represent the ratio of $-p(x_1 = i_1 - 1) \cdot \epsilon_{i_1 i_2} log \epsilon_{i_1 i_2}$ to D_1, $k_i \in [0, 1]$. Similarly, introduce a coefficient b_j,$(j = 1, \ldots, m^2 - m - 1)$ and $1 - b_1 - \cdots - b_{m^2-m-1}$ to represent the ratio of $-p(x_2 = j_1 - 1) \cdot \alpha_{j_1 j_2} log \alpha_{j_1 j_2}$ to D_2, $b_j \in [0, 1]$. For

$i_1 = 1, 2, \ldots, n$, $i_2 = 1, 2, \ldots, n-1$, we have following equations :

$$\text{A total of n(n-1)} \begin{cases} -p(x_1 = 0) \cdot \epsilon_{11} log \epsilon_{11} = k_1 D_1, \\ -p(x_1 = 0) \cdot \epsilon_{12} log \epsilon_{12} = k_2 D_1, \\ \quad \vdots \\ -p(x_1 = n-1) \cdot \epsilon_{n(n-2)} log \epsilon_{n(n-2)} = k_{n^2-n-1} D_1, \\ -p(x_1 = n-1) \cdot \epsilon_{n(n-1)} log \epsilon_{n(n-1)} = (1 - k_1 - \cdots - k_{n^2-n-1}) D_1. \end{cases} \tag{6.12}$$

For $j_1 = 1, 2, \ldots, m$, $j_2 = 1, 2, \ldots, m-1$

$$\text{A total of m(m-1)} \begin{cases} -p(x_2 = 0) \cdot \alpha_{11} log \alpha_{11} = b_1 D_2, \\ -p(x_2 = 0) \cdot \alpha_{12} log \alpha_{12} = b_2 D_2, \\ \quad \vdots \\ -p(x_2 = m-1) \cdot \alpha_{m(m-2)} log \alpha_{m(m-2)} = b_{m^2-m-1} D_2, \\ -p(x_2 = m-1) \cdot \alpha_{m(m-1)} log \alpha_{m(m-1)} = (1 - b_1 - \cdots - b_{m^2-m-1}) D_2. \end{cases} \tag{6.13}$$

By solving the system of equations 6.12, 6.13, we can get the solutions of $\epsilon_{i_1 i_2}$ and $\alpha_{j_1 j_2}$. Moreover, given $D_1, D_2 \to 0$, $logD_1$, $logD_2 \to -\infty$, if $k_i \neq 0$ and $k_i \neq 1$,

$b_j \neq 0$ and $b_j \neq 1$, we can get the approximate solutions,

$$\begin{cases} \epsilon_{11} = -\frac{k_1 D_1}{p(x_1=0)logD_1}, \\ \epsilon_{12} = -\frac{k_2 D_1}{p(x_1=0)logD_1}, \\ \quad \vdots \\ \epsilon_{n(n-2)} = -\frac{k_{n^2-n-1} D_1}{p(x_1=n-1)logD_1}, \\ \epsilon_{n(n-1)} = -\frac{(1-k_1-\cdots-k_{n^2-n-1})D_1}{p(x_1=n-1)logD_1}, \end{cases} \tag{6.14}$$

$$\begin{cases} \alpha_{11} = -\frac{b_1 D_2}{p(x_2=0)logD_2}, \\ \alpha_{12} = -\frac{b_2 D_2}{p(x_2=0)logD_2}, \\ \quad \vdots \\ \alpha_{m(m-2)} = -\frac{b_{m^2-m-1} D_2}{p(x_2=m-1)logD_2}, \\ \alpha_{m(m-1)} = -\frac{(1-b_1-\cdots-b_{m^2-m-1})D_2}{p(x_2=m-1)logD_2}. \end{cases} \tag{6.15}$$

Substitute $\epsilon_{i_1 i_2}$ and $\alpha_{j_1 j_2}$ in Eq.6.14, Eq.6.15 into Eq.6.6, now the optimization problem in Eq.4.16 can be converted to an optimization problem related to variables k_i and b_j. Each cross-term can be rewritten as

$$
\begin{aligned}
r_1\epsilon_{11}\alpha_{11} &= \frac{D_1D_2}{logD_1logD_2}\cdot\frac{r_1}{p(x_1=0)p(x_2=0)}\cdot b_1k_1,\\
r_2\epsilon_{11}\alpha_{12} &= \frac{D_1D_2}{logD_1logD_2}\cdot\frac{r_2}{p(x_1=0)p(x_2=0)}\cdot b_2k_1,\\
&\vdots\\
r_{m(m-1)}\epsilon_{11}\alpha_{m(m-1)} &= \frac{D_1D_2}{logD_1logD_2}\cdot\frac{r_{m(m-1)}}{p(x_1=0)p(x_2=m-1)}\cdot(1-b_1-\cdots-b_{m^2-m-1})k_1,\\
r_{m^2-m+1}\epsilon_{12}\alpha_{11} &= \frac{D_1D_2}{logD_1logD_2}\cdot\frac{r_{m^2-m+1}}{p(x_1=0)p(x_2=0)}\cdot b_1k_2,\\
&\vdots\\
r_{2m(m-1)}\epsilon_{12}\alpha_{m(m-1)} &= \frac{D_1D_2}{logD_1logD_2}\cdot\frac{r_{2m(m-1)}}{p(x_1=0)p(x_2=m-1)}\cdot(1-b_1-\cdots-b_{m^2-m-1})k_2,\\
&\vdots\\
&\vdots\\
r_{(n^2-n-1)m(m-1)+1}\epsilon_{n(n-1)}\alpha_{11} &= \frac{D_1D_2}{logD_1logD_2}\cdot\frac{r_{(n^2-n-1)m(m-1)+1}}{p(x_1=n-1)p(x_2=0)}\cdot b_1(1-k_1-\cdots-k_{n^2-n-1}),\\
&\vdots\\
r_{n(n-1)m(m-1)}\epsilon_{n(n-1)}\alpha_{m(m-1)} &= \frac{D_1D_2}{logD_1logD_2}\cdot\frac{r_{n(n-1)m(m-1)}}{p(x_1=n-1)p(x_2=m-1)}\\
&\quad\cdot(1-b_1-\cdots-b_{m^2-m-1})\cdot(1-k_1-\cdots-k_{n^2-n-1}).
\end{aligned}
$$

Totally there are $n\times(n-1)\times m\times(m-1)$ terms. And $I(X_1;X_2|U_1,U_2)$ is the sum of them.

Consider the optimization problem

$$\begin{aligned} min \quad & I(X_1;X_2|U_1,U_2), \\ s.t. \quad & 0 < k_i < 1, \\ & 0 < b_j < 1. \end{aligned} \tag{6.16}$$

The products $k_i b_j$, $(1-k_1-\cdots-k_{n^2-n-1})b_j$, $k_i(1-b_1-\cdots-b_{m^2-m-1})$ and $(1-k_1-\cdots-k_{n^2-n-1})(1-b_1-\cdots-b_{m^2-m-1})$ are all in the interval $(0,1)$. And note that the sum of them equals 1. Therefore, these products can be regarded as the weight of each term in the total $I(X_1;X_2|U_1,U_2)$. In this way, we just need to compare the values of the coefficients, find the minimum coefficient, adjust its corresponding weight w^* to the maximum value of 1, and set other terms' weights to be 0.

However, note that we assume $k_i \neq 0$ and $k_i \neq 1$, $b_j \neq 0$ and $b_j \neq 1$ when simplifying the solutions in Eq.6.12, Eq.6.13, so the minimum value of $I(X_1;X_2|U_1,U_2)$ is obtained when the weight w^* corresponding to the term with the smallest coefficient is close to 1.

Then the asymptotic minimum rate as D_1, $D_2 \to 0$ is:

$$I(X_1,X_2;U_1,U_2) = H(X_1,X_2) - D_1 - D_2 + I_{min}(X_1;X_2|U_1,U_2) + o(\frac{D_1 D_2}{logD_1 \cdot logD_2}), \tag{6.17}$$

where $I_{min}(X_1; X_2|U_1, U_2)$ is

$$I_{min}(X_1; X_2|U_1, U_2) = min\left\{\frac{D_1 D_2}{logD_1 \cdot logD_2} \cdot \frac{r_t}{p(x_1 = i_1 - 1)p(x_2 = j_1 - 1)}\right\}. \tag{6.18}$$

r_t is the coefficient of the cross-term in Eq.6.10. This completes the proof of Theorem 3.

Chapter 7

Numerical Test

Some numerical examples will be provided in this section to verify our main results.

1. Verification for keeping up to the dominant terms of $H(X|U_1, U_2)$.

For the Binary uniform case in Theorem 1, take $H(X_1|U_1, U_2)$ as an example. As D_1, $D_2 \to 0$, we simplify the expression of $H(X_1|U_1, U_2)$ in Eq.7.1, to Eq.7.2

$$\begin{aligned} H(X_1|U_1, U_2) = &-\frac{1}{2}\epsilon_1 log\epsilon_1 - \frac{1}{2}\epsilon_2 log\epsilon_2 \\ &+ \frac{1}{2}\left[1 + (1-2q)log\frac{q}{1-q}\right]\epsilon_1 + \frac{1}{2}\left[1 + (1-2q)log\frac{q}{1-q}\right]\epsilon_2, \end{aligned} \tag{7.1}$$

$$H(X_1|U_1, U_2) = -\frac{1}{2}\epsilon_1 log\epsilon_1 - \frac{1}{2}\epsilon_2 log\epsilon_2, \tag{7.2}$$

keeping up to the dominant terms instead of keeping up to linear terms. And when solving the optimization problem, we know the minimum value of $-H(X_1|U_1, U_2)$ is equal to $-D_1$. We can conduct an experiment to verify this approximation is acceptable.

First, suppose $q = \frac{1}{3}$, $\epsilon_1 = \epsilon_2 = \epsilon$, $D_1 = 2 \times 10^{-5}$, then we can solve the Eq.7.1 and Eq.7.2 by using Matlab. The solution of Eq.7.1 is

$$\epsilon = 0.0000014039870734622646750582710429794, \tag{7.3}$$

and the solution of Eq.7.2 is

$$\epsilon = 0.0000014907301328925392993290941286112, \tag{7.4}$$

where the difference of these two solutions is 8.6743×10^{-8}. The difference is so small that we think that if $D_1 \leq 2 \times 10^{-5}$, the $H(X_1|U_1, U_2)$ can only keep up to the dominant terms.

Then let's substitute the ϵ in 7.4 to the original expression of $H(X_1|U_1, U_2)$ that has not been applied Taylor series expansion and to the equation of Eq.7.2. The result of original expression is denoted by eq_1, and the result of dominant-terms equation is denoted by eq_2,

$$eq_1 = 2.1146e - 05, \qquad eq_2 = 2.0000e - 05. \tag{7.5}$$

Therefore, we hold the opinion that the approximation is acceptable.

2. Verification for the asymptotic minimum rate obtained when the weight w^* is close to 1.

To verify that the asymptotic minimum rate is obtained when the weight w^* is close to 1. Here we design a nested loop algorithm to iterate over all variables in the binary case of Theorem 1.

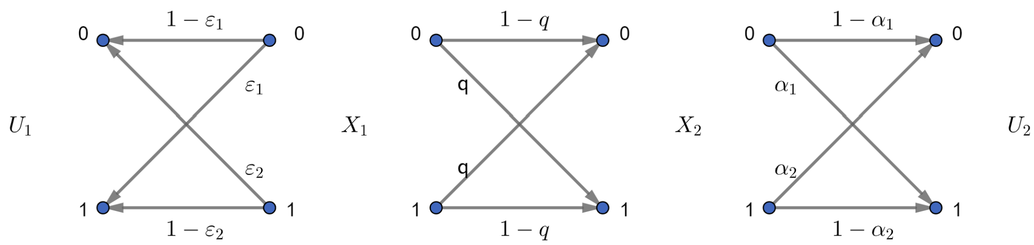

FIGURE 7.1: Binary case

The outer loop iterates ϵ_1, then ϵ_2 is obtained by solve the Eq.7.2. In the inner loop, we iterate α_1, then α_2 is obtained by solve the Eq.5.11. Next we compute the $I(X_1;X_2|U_1,U_2)$, and find the minimum value of it and print the corresponding ϵ_1, ϵ_2, α_1, α_2. Let I_1 denotes the $I(X_1;X_2|U_1,U_2)$ obtained by iteration, and I_2 denotes the $I(X_1;X_2|U_1,U_2)$ obtained by our theorem where we set the $w^* = 1$.

Suppose $q = \frac{1}{3}$, We can get the table 7.1. The table shows that the difference

$D_1 = D_2$	ϵ_1	ϵ_2	α_1	α_2	I_1	w^*	I_2
1×10^{-5}	7.000e-08	1.398e-06	7.000e-08	1.398e-06	4.683e-13	0.888	4.713e-13
5×10^{-6}	6.700e-07	2.729e-08	6.700e-07	2.729e-08	1.052e-13	0.907	1.057e-13
1×10^{-7}	1.000e-10	1.078e-08	1.0000e-10	1.078e-08	2.521e-17	0.977	2.524e-17
5×10^{-8}	5.000e-11	5.179e-09	5.000e-11	5.179e-09	5.829e-18	0.976	5.835e-18
1×10^{-8}	9.600e-10	2.488e-12	9.6000e-10	2.488e-12	1.968e-19	0.993	1.968e-19

TABLE 7.1: numerical test in the binary case

between $I(X_1;X_2|U_1,U_2)$ obtained by iteration and $I(X_1;X_2|U_1,U_2)$ obtained by our theorem where we set the $w^* = 1$ is very small. Moreover, as D_1 and $D_2 \to 0$, the w^* is getting closer to 1, which proves that our conclusion is correct.

To further prove our conclusion, we also conduct an experiment on another special case.

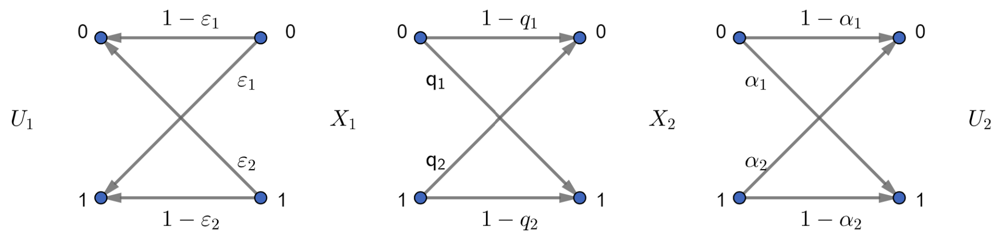

FIGURE 7.2: Asymmetric binary case

Suppose $q_1 = 0.1$, $q_2 = 0.2$, we can get the table 7.2, this test again proves

$D_1 = D_2$	ϵ_1	ϵ_2	α_1	α_2	I_1	w^*	I_2
1×10^{-5}	1.490e-06	4.199e-10	1.3400e-06	3.862e-10	2.357e-12	0.996	2.356e-12
5×10^{-6}	7.050e-07	6.439e-10	6.350e-07	1.802e-10	5.293e-13	0.995	5.287e-13
1×10^{-7}	1.090e-08	5.955e-11	9.8000e-09	5.687e-11	1.267e-16	0.993	1.264e-16
5×10^{-8}	5.200e-09	3.367e-11	4.700e-09	3.999e-11	2.954e-17	0.983	2.923e-17
1×10^{-8}	9.600e-10	2.488e-12	8.7000e-10	1.407e-12	9.904e-19	0.995	9.867e-19

TABLE 7.2: numerical test in the asymmetric binary case

asymptotic conditional mutual information is obtained when w^* is close to 1.

Chapter 8

Conclusion

We have studied the asymptotic minimum rate under given log-loss distortion D_1 and D_2, and D_1, $D_2 \to 0$. In order to attack the general case, we first studied the special case where two uniformly distributed sources are connected by a binary symmetric channel, and the alphabet of U_i is equal to the sources alphabet X_i Under this premise, we have a simple expression in terms of D_1 and D_2.

Then this result is extended in Theorem 2, where we consider enlarging the alphabet of U_i, making it greater than the sources alphabet X_i. It turns out that the impact of enlarging the alphabet of U_i on the final asymptotic result could be ignored when D_1, $D_2 \to 0$.

Finally, we derived the result of the most general case. The size of source alphabet X_1 is $|\mathcal{X}_1| = n$, the size of source alphabet X_2 is $|\mathcal{X}_2| = m$, $(m \geq n)$, the size of alphabet U_1 is $|\mathcal{U}_1| = u + 1$, $u + 1 \geq n$, and the size of alphabet U_2 is $|\mathcal{U}_2| = v + 1$, $v + 1 \geq m$. And the result also confirms the findings in the binary case. The asymptotic minimum rate is again expressed explicitly as a function of

D_1 and D_2.

In future work, we plan to extend our work to the more general case that includes noises and more sources.

Bibliography

[1] K. Sayood. Data Compression. In: *Encyclopedia of Information Systems*. Ed. by H. Bidgoli. New York: Elsevier, 2003, 423–444. ISBN: 978-0-12-227240-0.

[2] T. M. Cover and J. A. Thomas. Elements of information theory 2nd edition (wiley series in telecommunications and signal processing) (2006).

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

[4] R. Ahlswede and J. Korner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory* 21(6) (1975), 629–637.

[5] A. Wyner. On source coding with side information at the decoder. *IEEE Transactions on Information Theory* 21(3) (1975), 294–300.

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

[7] T. Berger. Multiterminal source coding. *The information theory approach to communications* (1978).

[8] S. Tung. Multiterminal source coding (ph. d. thesis abstr.) *IEEE Transactions on Information Theory* 24(6) (1978), 787–787.

[9] T. Berger and R. W. Yeung. Multiterminal source encoding with one distortion criterion. *IEEE Transactions on Information Theory* 35(2) (1989), 228–236.

[10] R. W. H. Yeung. Some results on multiterminal source coding. PhD thesis. Cornell University, 1988.

[11] T. Berger, Z. Zhang, and H. Viswanathan. The CEO problem [multiterminal source coding]. *IEEE Transactions on Information Theory* 42(3) (1996), 887–902.

[12] M. Nangir, R. Asvadi, J. Chen, M. Ahmadian-Attari, and T. Matsumoto. Successive Wyner-Ziv coding for the binary CEO problem under logarithmic loss. *IEEE Transactions on Communications* 67(11) (2019), 7512–7525.

[13] Y. Uğur, I. E. Aguerri, and A. Zaidi. Vector Gaussian CEO problem under logarithmic loss and applications. *IEEE Transactions on Information Theory* 66(7) (2020), 4183–4202.

[14] J. Jiao, T. A. Courtade, K. Venkat, and T. Weissman. Justification of logarithmic loss via the benefit of side information. *IEEE Transactions on Information Theory* 61(10) (2015), 5357–5365.

[15] T. A. Courtade and T. Weissman. Multiterminal Source Coding Under Logarithmic Loss. *IEEE Transactions on Information Theory* 60(1) (2014), 740–761.

[16] M. Nangir, R. Asvadi, M. Ahmadian-Attari, and J. Chen. Analysis and code design for the binary CEO problem under logarithmic loss. *IEEE Transactions on Communications* 66(12) (2018), 6003–6014.

[17] Y. Shkel, M. Raginsky, and S. Verdú. Universal lossy compression under logarithmic loss. In: *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2017, 1157–1161.

[18] D. Seo and L. R. Varshney. The CEO problem with rth power of difference and logarithmic distortions. *IEEE Transactions on Information Theory* 67(6) (2021), 3873–3891.

[19] Z. Xiong, A. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine* 21(5) (2004), 80–94.