Generalized Gaussian Multiterminal Source Coding in the High-Resolution Regime

GENERALIZED GAUSSIAN MULTITERMINAL SOURCE CODING IN THE HIGH-RESOLUTION REGIME

BY

XIAOLAN TU, B.Sc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2018)
(Mathematics & Statistics)

McMaster University
Hamilton, Ontario, Canada

TITLE:	Generalized Gaussian Multiterminal Source Coding in the High-Resolution Regime
AUTHOR:	Xiaolan Tu B.Sc. Mechatronic Engineering Beijing Institute of Technology
SUPERVISOR:	Prof. Narayanaswamy Balakrishnan Prof. Jun Chen
NUMBER OF PAGES:	viii, 49

To those who gave me help and support

Abstract

Source coding, a central concept in information theory, is the study of encoding and decoding data. Depending on the topological structure of the coding system, i.e. how the sources are connected with encoders, different rate distortion functions are used. In this thesis two different encoding schemes—distributed and decentralized—are discussed and compared with a benchmark (centralized) coding structure. Specifically, all structures for two and three sources are discussed and a special case for the multi-source (more than three sources) is calculated. This work gives a pathway to characterize the generalized multiterminal source coding systems by finding the difference in the rate distortion limits from the optimal centralized coding system. It is shown that in specific cases, some decentralized systems can achieve the Shannon lower bound in a high resolution regime.

Acknowledgements

I would first like to thank Dr. Jun Chen for supervising me and for all his help throughout my degree. I'm grateful for all the advice he gave me on this thesis and my coursework; he is a great engineer and mentor. I would also like to thank Dr. Balakrishnan for helping me transfer into the statistics department and his invaluable advice in my graduate degree. I could not have completed this degree without his support. I am also grateful for Dr. Feng Shui for being a member of my examination committee. Finally thanks to all my colleagues and friends who made my life at McMaster colourful.

Contents

List of Figures

Chapter 1

Introduction

1.1 Literature Review

Source coding, a central method in information theory, was rigorously developed to find mathematical models for data compression (Shannon, 1948). Today, source coding has three main frameworks: centralized, decentralized, and distributed.

Centralized source coding refers to a encoder-decoder scheme where all sources are received by a central encoder, which jointly encodes the sources. This is considered the optimum setup when sources are correlated because the encoder can model the dependent information (Shannon, 1948).

Distributed source coding (DSC) is a modelling framework that compresses multiple correlated sources separately. DSC's main advantage is transferring computational load from the encoder to the decoder side. Slepian and Wolf (1973) first showed that two correlated sources could efficiently be reconstructed provided a joint decoder was used. If joint encoding is unfeasible or there are computational constraints at the encoder level, DSC should be considered. For example, two satellites unable to communicate with each other can independently send images to earth where a joint decoder, unconstrained by computational stress or power, can receive the source (Aljohani *et al.*, 2016). Later, Cover (1975) extended the work done by Slepian and Wolf to the multiterminal system.

Decentralized source coding refers to the framework in which some combination of sources are jointly encoded and some are not. Particularly, decentralized source coding with three sources was studied by Chen *et al.* (2007) and Wang and Chen (2013). This type of scheme is discussed throughout the thesis and will be compared with the optimal coding scheme, centralized.

Within these coding frameworks lies two different methods for constructing the encoders: direct and indirect. Direct source coding assumes the sources are perfectly observed by the encoders without noise, while indirect assumes the sources are corrupted by some independent noise. Direct multiterminal source coding is an extension of Slepian and Wolf (1973) and has a tight inner bound for the distortion rate region referred to as the Berger-Tung bound (Berger, 1978). This thesis considers only the direct version, however much work has been done for the indirect case. The pivotal paper by Oohama (1997) who took advantage of Shannons entropy power inequality, led to important advances in the general indirect scalar Gaussian multiterminal source coding (GMSC) problem. In fact, Oohama's paper was the starting point for a complete characterization of the rate region of the scalar Gaussian CEO problem, a special case of GMSC. However, in a vector scenario, Oohama's methods can be unsuitable as shown in Wang and Chen (2013) and Wang and Chen (2014).

In the general case for multiterminal source coding, Chen *et al.* (2007) further explored the sum rate distortion function. Showing that a centralized systems rate distortion can be matched by ostensibly weaker decentralized ones. Chen *et al.* (2007) presented a novel graphic approach to investigate generalized multiterminal source coding systems and highlighted the role of shared inputs of different encoders. Through the use of Gaussian Markov networks (GMN) or a Gaussian Markov random field (GMRF) the effect of shared inputs on GMSC was highlighted and the sum-rate of the Berger-Tung scheme could match the Shannon lower bound under certain conditions.

1.2 Contribution and organization

In this thesis, a pathway is given to characterize the rate distortion limit of generalized multiterminal source coding by comparing it to a corresponding special source coding system with a centralized encoder. Eventually, the Shannon lower bound is obtained by optimizing the rate distortion function subject to some criterion. However, for some complex systems, the source coding problem can not be transformed into an optimization problem. In addition to this, some times even when the equivalent optimization form is found, a closed form of the solution is not obtainable or unfeasible. There are three common distortion criteria: distortion matrix criterion, vector distortion criterion, and sum distortion criterion. This thesis will only consider the sum distortion criterion, where the trace of the covariance matrix is bounded by some constant, D. Different coding systems with 2 sources, 3 sources and L sources are discussed separately. Surprisingly, it is found that some source sharing coding systems with non-centralized encoders can be as powerful as a centralized version.

The rest of the paper is organized as follows, the next two sections of Chapter 1 introduce the rate distortion problem and graphical models. In Chapter 2 a formal statement of the problem is presented. The two-sources source coding and three-sources case are then discussed in Chapter 3. Following that, another section in Chapter 3 shows two special cases of the generalized multiterminal coding systems with L sources. Finally, some discussion is made in Chapter 5.

1.3 Rate Distortion

1.3.1 Rate Distortion Theory

In this section, we will address a fundamental problem of information theory, rate distortion. More details and proofs can be found in Cover and Thomas (2006). Suppose we want to represent a continuous random variable by a sequence of bits, infinite bits will be required. Hence, if finite bits are used, distortion is inevitable. From this example, the rate distortion theory can be stated as an optimization problem. Based on the distribution of sources and a distortion measure, at a given rate, what is the minimum expected distortion achievable? The problem can also be stated in a different but equivalent way, for a fixed distortion, what is the minimum rate description.

1.3.2 Gaussian Source

When the source is normally distributed, $\mathcal{N}(0, \sigma^2)$, the rate distortion function with a squared distortion function is

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases}. \tag{1.1}$$

When there are L sources with a multivariate normal distribution, $\mathcal{N}(\mu, \boldsymbol{\Sigma})$, the rate distortion function is

$$\begin{aligned} R(\mathbf{D}) = \min_{\mathbf{D}} \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}|}{|\mathbf{D}|} \\ \text{subject to: } \mathbf{0} \preceq \mathbf{D}, \\ tr(\mathbf{D}) \leq Ld, \end{aligned} \tag{1.2}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix, $\mathbf{D}$ is a symmetric and semi-positive matrix called the distortion matrix which is different for coding systems with different structures, and d is a positive constant close to zero.

1.4 Graphical models

The purpose of this section is to introduce the Probabilistic Graphical Model (PGM), which is a graph-based representation. It allows us to compactly express the conditional dependence structure of a joint distribution over a high-dimensional space. More details can be found in the book of Koller and Friedman (2009).

1.4.1 Gaussian Markov Network

One important Probabilistic Graphical Model is called Markov Network, also known as Markov Random Field (MRF). It is an undirected graphical model where the interaction between variables is not directed. A node in a Markov Field represents a random variable and an edge in a Markov Field shows the conditional dependence between the two corresponding random variables. Specifically, when the joint distribution is a multivariate Gaussian distribution, a Markov Network is called a Gaussian

Markov Network (GMN).

Assume a set of n random variables $\mathcal{X} = \{X_1, X_2, \cdots, X_n\}$ are jointly multivariate Gaussian distributed with a density function:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} |\mathbf{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right], \tag{1.3}$$

where $\boldsymbol{\mu}$ is the mean vector with dimension $1 \times n$ and $\mathbf{\Sigma}$ is the covariance matrix. The inverse of $\mathbf{\Sigma}$ is called the information matrix which is $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$. The information form of the exponential part can then be expressed as:

$$\begin{aligned} -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Theta}(\mathbf{x}-\boldsymbol{\mu}) \\ &= -\frac{1}{2}\left[\underbrace{\mathbf{x}^T\mathbf{\Theta}\mathbf{x}}_{\text{scalar}} - \underbrace{2\mathbf{x}^T\mathbf{\Theta}\boldsymbol{\mu}}_{\text{scalar}} + \underbrace{\boldsymbol{\mu}^T\mathbf{\Theta}\boldsymbol{\mu}}_{\text{scalar}}\right]. \end{aligned} \tag{1.4}$$

The last term does not contain any variables i.e. a constant, therefore it can be seen that the density function is proportional to the first two terms and can be further expressed in so called information form:

$$p(\mathbf{x}) \propto \exp\left[-\frac{1}{2}\mathbf{x}^T\mathbf{\Theta}\mathbf{x} + (\mathbf{\Theta}\boldsymbol{\mu})^T\mathbf{x}\right]. \tag{1.5}$$

In this form $\mathbf{\Theta}$ needs to be a positive semi-definite matrix in order to define a valid Gaussian distribution. For the multivariate Gaussian distribution, the covariance matrix reveals independence and the information matrix shows conditional independence between variables.

Theorem 1: X_i and X_j are independent if and only if $\sigma_{ij} = 0$.

Theorem 2: $p(\mathbf{x}) \models (X_i \perp X_j | \mathcal{X} - \{X_i, X_j\})$ if and only if $\theta_{ij} = 0$.

Based on Theorem 2, one can see that the information form induces a pairwise Markov network, where if $\theta_{jk} = \theta_{kj} = 0$, then there is no edge between X_j and X_k.

1.5 A combinatorial Problem

Consider a set $\mathbf{B} = \{1, 2, \cdots, L\}$, where L is a positive integer. $\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_k$ are k nonempty subsets of set $\mathbf{B}$ (here k is not fixed). If $\mathbf{A}_1 \cup \mathbf{A}_2 \cup \cdots \mathbf{A}_k = \mathbf{B}$ and $\mathbf{A}_i \not\subset \mathbf{A}_j$ for all i and j, then we say that $\mathbb{A} \triangleq \Big\{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_k\Big\}$ is a set cover of $\mathbf{B}$ (i.e, a family of nonempty subsets of $\mathbf{B}$ whose union contains $\mathbf{B}$ itself).

Among all set covers of $\mathbf{B}$, some of them are isomorphic to each other via index relabeling. For example, when $\mathbf{B} = \{1, 2, 3\}$, set cover $\Big\{\{1, 2\}, \{3\}\Big\}$ and set cover $\Big\{\{1, 3\}, \{2\}\Big\}$ are isomorphic. Here the combinatorial problem can be stated as: given a set $\mathbf{B}$, how many non-equivalent set covers can be found?

When L is small, the answer is straightforward. For instance, when $L = 2$, there are only two non-equivalent set covers which are $\Big\{1, 2\Big\}$ and $\Big\{\{1\}, \{2\}\Big\}$, when $L = 3$, there are five non-equivalent set covers:

$$\Big\{\{1\}, \{2\}, \{3\}\Big\},$$
$$\Big\{\{1, 2\}, \{3\}\Big\},$$
$$\Big\{\{1, 2\}, \{2, 3\}\Big\},$$
$$\Big\{\{1, 2\}, \{2, 3\}, \{1, 3\}\Big\},$$
$$\Big\{1, 2, 3\Big\}.$$

When L increases, the number of non-equivalent set covers quickly increases, for $L = \{1, 2, 3, 4, 5, 6, 7, \cdots\}$, the corresponding number of set covers $= \{1, 2, 5, 20, 180, 16143,$

$489996795, \cdots\}$. Consider a generalized multiterminal source coding system with L sources and K encoders, where each source is connected with at least one encoder and each encoder observes a nonempty subset of sources. Clearly, the number of inequivalent possible source coding topological structures is the sequence discussed above.

Chapter 2

Problem statement and preliminaries

A formal statement of the central problem is given in this chapter and some previous results will be shown after. To begin with, we consider a generalized multiterminal source coding system of L correlated Gaussian sources $\mathbf{X}^L = \{X_1, X_2, \cdots, X_L\}$ which is an L dimensional random vector. The variance covariance matrix of $\mathbf{X}^L$ is denoted by $\boldsymbol{\Sigma}$ and the inverse of variance covariance matrix is called information matrix denoted by $\boldsymbol{\Theta}$. This coding system is shown in Figure 2.1, L Sources are observed by K encoders and then separately compressed and represented by bits. After compression, data are forwarded to a central decoder and are jointly reconstructed in this step.

It is worth mentioning that each encoder is allowed to observe a nonempty subset of sources, and each source is observed by at least one encoder and can be observed by multiple encoders at the same time. Thus, after connecting sources and encoders, we can achieve a bipartite graph. The number of non-equivalent bipartite graphs is the sequence found in Section 1.5. When the number of sources is two or three, we can find two or five non-equivalent source coding systems respectively which will be discussed in details in Chapter 3. Throughout this work, all the data compressed by different encoders will be forwarded to a centralized information processing center (decoder) which attempts to reconstruct the remote source $\mathbf{X}^L$.

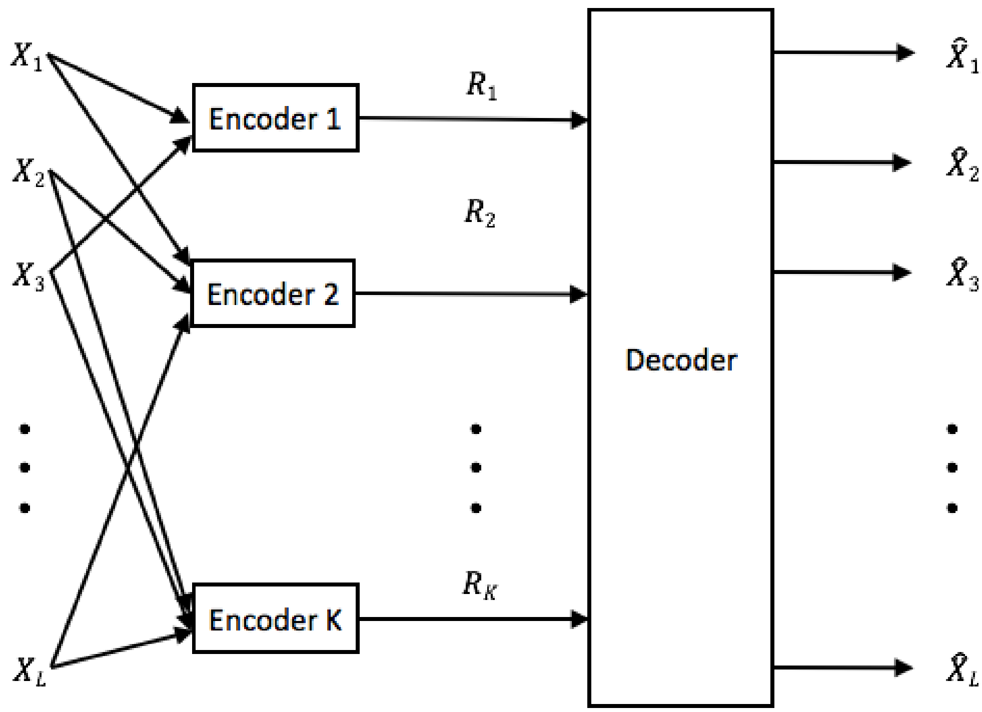

Figure 2.1: Generalized multiterminal source coding system with L sources and K encoders.

The rate distortion function of this coding system is given in Equation 1.2, in which the distortion criteria is a trace constraint called the sum distortion criteria, $tr(\mathbf{D}) \leq Ld$, i.e the trace of the distortion matrix must not exceed a prescribed positive number. The equivalent trace constraint can be expressed as

$$\mathbf{0} \preceq \mathbf{D} = \begin{pmatrix} \epsilon d_1 & .. & \cdots & .. \\ .. & \epsilon d_2 & \cdots & .. \\ \vdots & \vdots & \ddots & \vdots \\ .. & .. & \cdots & \epsilon d_L \end{pmatrix} \tag{2.1}$$

where the diagonal elements have to satisfy $\epsilon(d_1 + d_2 + \cdots + d_L) \leq Ld$ and ϵ and d are both small enough (i.e, close to zero) and non-negative.

A special case of this generalized multiterminal source coding system is shown in Figure 2.2 where there is only one centralized encoder which can simultaneously observe all sources $\mathbf{X}^L = \{X_1, X_2, \cdots, X_L\}$. Intuitively, the centralized encoder can gather all the information about the input sources completely, namely, it can not only observe each source but also the information between them, like correlation. Therefore, we can state that the source coding system with a centralized encoder is the most powerful among all the possible coding topological structures which means its' rate distortion function $(R(\mathbf{D})_{\{X_1, X_2, \cdots, X_L\}} = \min_{\mathbf{D}} \frac{1}{2} \log \frac{|\mathbf{\Sigma}|}{|\mathbf{D}|})$ is the smallest with a diagonal distortion matix

$$\mathbf{D} = \begin{pmatrix} \epsilon d_1 & & & \\ & \epsilon d_2 & & \\ & & \ddots & \\ & & & \epsilon d_L \end{pmatrix} \tag{2.2}$$

which can be found in this paper Chen *et al.* (2007).

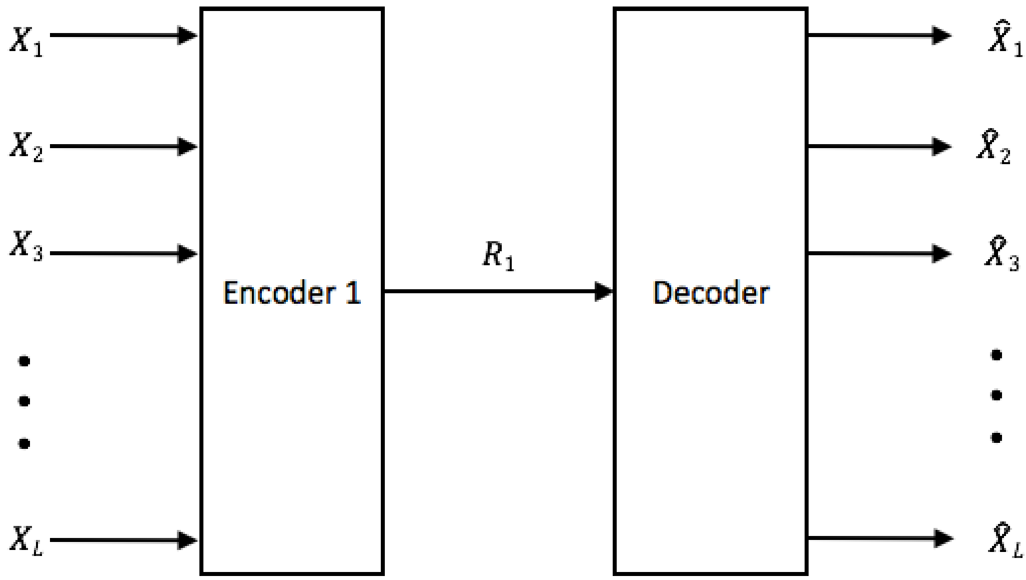

Figure 2.2: Centralized source coding system with L sources.

Apparently, in order to minimize $\frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{|\mathbf{D}|}$, $|\mathbf{D}|$ needs to be maximized subject to a lower bound of the distortion matrix and an upper bound of its trace ($tr(\mathbf{D}) \leq Ld$) which is expressed in (2.3)

$$\begin{aligned} &\max|\mathbf{D}| = \max \epsilon^L \prod_{i=1}^{L} d_i \\ &\text{subject to: } \mathbf{0} \preceq \mathbf{D} \Rightarrow d_i \text{ is nonnegative for } i = 1, 2, \cdots, L\ , \\ &\qquad tr(\mathbf{D}) = \epsilon \sum_{i=1}^{L} d_i \leq Ld, \end{aligned} \tag{2.3}$$

where ϵ is a small non-negative constant. It can be proved that the product is maximized if and only if each entry is equal to each other. A brief argument is given here: suppose any pair d_i and d_j are different and let d_m be their arithmetic average so that $d_i + d_j = 2d_m$ and $d_i = d_m + c$, $d_j = d_m - c$ for some $c \neq 0$. Consider the product we have:

$$d_i \times d_j = (d_m + c)(d_m - c) = d_m^2 - c^2 < d_m^2, \tag{2.4}$$

which means one can maintain the overall sum but increase the overall product by replacing d_i and d_j by d_ms. Therefore the determinant is maximized when all diagonal entries equal to each other which is $\epsilon d_1 = \epsilon d_2 = \cdots = \epsilon d_L = d$. Thus the rate distortion for this source coding system with a centralized encoder is

$$R(\mathbf{D})_{\{X_1,X_2,\cdots,X_L\}} \underset{d\to 0}{=} \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{d^L}. \tag{2.5}$$

Specifically, when there are two sources (i.e, $L = 2$), the rate distortion function is

$$R(\mathbf{D})_{\{X_1,X_2\}} \underset{d\to 0}{=} \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{d^2}, \tag{2.6}$$

and when there are three sources (i.e, $L = 3$), the rate distortion function is

$$R(\mathbf{D})_{\{X_1,X_2,X_3\}} \underset{d\to 0}{=} \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{d^3}. \tag{2.7}$$

Consider this most powerful coding system as a benchmark, we would like to find what is the differences between any other source coding systems with different encoder structures (non-centralized) in the value of rate distortion functions which can be mathematically expressed as

$$\lim_{\epsilon \to 0} \frac{R(\mathbf{D})_{\text{non-centralized encoders}} - R(\mathbf{D})_{\text{centralized encoder}}}{\epsilon^2}. \tag{2.8}$$

It is worth mentioning that both $R(\mathbf{D})_{\text{non-centralized encoders}}$ and $R(\mathbf{D})_{\text{centralized encoder}}$ approach zero when ϵ is small enough, thus

$$\lim_{\epsilon \to 0} R(\mathbf{D})_{\text{non-centralized encoders}} - R(\mathbf{D})_{\text{centralized encoder}}$$

does not imply anything valuable because it is always zero. One can find that

$$R(\mathbf{D})_{\text{non-centralized encoders}} - R(\mathbf{D})_{\text{centralized encoder}} = O(\epsilon^2),$$

where $O(\epsilon^2)$ means the order of this difference about ϵ is two. Therefore, finding the difference in (2.8) is meaningful.

Chapter 3

Generalized Gaussian Multiterminal Source Coding

In this chapter, we will discuss coding systems with two (X_1 and X_2), three (X_1, X_2 and X_3) and multiple (L) sources. The process of source coding are the same as mentioned before, encoded by encoders and decoded by a central decoder. In generalized multiterminal source coding, it is important to know that inputs can be shared by different encoders. There are two non-equivalent coding systems for two sources and five for three sources. Using the centralized encoder system as a reference, differences between this reference and other non-centralized encoder systems will be found.

3.1 The Two-Source Case

Assume the joint distribution of X_1 and X_2 is multivariate normal distribution with mean vector $\boldsymbol{\mu} = [\mu_1, \mu_2]^T$, variance covariance matrix $\mathbf{\Sigma}$, and information matrix $\mathbf{\Theta}$, where

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \qquad \mathbf{\Theta} = \mathbf{\Sigma}^{-1} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix} \tag{3.1}$$

and entries in $\mathbf{\Theta}$ are nonzero. The Markov network of X_1 and X_2 is shown here

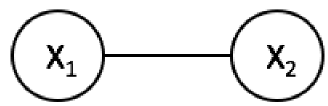

Figure 3.1: The Markov network of X_1 and X_2.

where there is an edge connecting X_1 and X_2 because off-diagonal entry θ_{12} in information matrix is nonzero. Figure 3.2 shows a source coding system with the most powerful encoder which can observe two sources X_1 and X_2 simultaneously. *Encoder 1* encodes the sources and forwards the compressed data R_1 to a central decoder where data is reconstructed and the output $\hat{X}_1$ and $\hat{X}_2$ are generated.

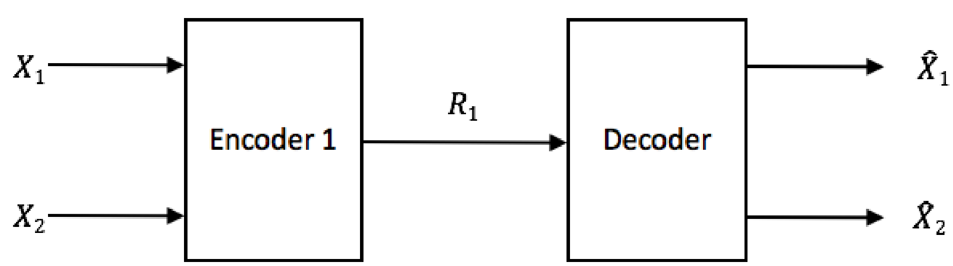

Figure 3.2: One encoder which observes X_1 and X_2 simultaneously.

Figure 3.3, the source coding system with the least powerful encoders which is called distributed source coding, observe two sources X_1 and X_2 individually. Sources are encoded by *Encoder 1* and *Encoder 2* to compressed data R_1 and R_2. They are further forwarded to a central decoder which decodes the data generates the output $\hat{X}_1$ and $\hat{X}_2$. The edge between X_1 and X_2 in the Markov network is not observed by this distributed encoding setup.

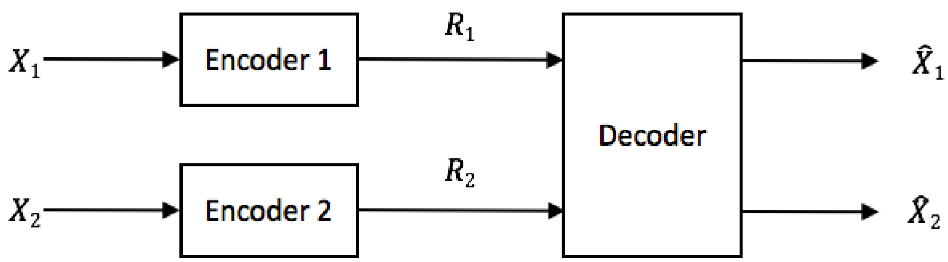

Figure 3.3: Two encoders which observe X_1 and X_2 individually.

The rate distortion function for this distributed coding system which can be found in the paper of A. B. Wagner and Viswanath (2008) is

$$\begin{aligned} R(\mathbf{D})_{\{X_1\},\{X_2\}} &\underset{d\to 0}{=} \min \frac{1}{2} \log \frac{|\mathbf{\Sigma}|}{|\mathbf{D}|} \\ \text{subject to: } & \mathbf{0} \preceq \mathbf{D} = (\mathbf{\Theta} + \mathbf{A_1^{-1}})^{-1}, \\ & tr(\mathbf{D}) \leq 2d, \end{aligned} \tag{3.2}$$

where $\mathbf{D}$ is the distortion matrix and $\mathbf{A_1}$ is the noise covariance matrix

$$\mathbf{0} \preceq \mathbf{A_1} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}. \tag{3.3}$$

By rearranging $\mathbf{D}$, we have

$$\begin{aligned}\mathbf{D} &= \left(\mathbf{\Theta} + \mathbf{A_1}^{-1}\right)^{-1} \\ &= \left(\mathbf{A_1}^{-1} + \mathbf{\Theta}\right)^{-1} \\ &= \left[\mathbf{A_1}^{-1}\left(\mathbf{I} + \mathbf{A_1}\mathbf{\Theta}\right)\right]^{-1} \\ &= \left(\mathbf{I} + \mathbf{A_1}\mathbf{\Theta}\right)^{-1}\mathbf{A_1}.\end{aligned} \tag{3.4}$$

According to Sherman/Morrison formula, one can further expand this inversion:

$$\begin{aligned}(3.4) &= \left[\mathbf{I} - \mathbf{A_1}\mathbf{\Theta} + \sum_{n=2}^{n=\infty}(-1)^n\left(\mathbf{A_1}\mathbf{\Theta}\right)^n\right]\mathbf{A_1} \\ &= \mathbf{A_1} - \mathbf{A_1}\mathbf{\Theta}\mathbf{A_1} + \sum_{n=2}^{n=\infty}(-1)^n\left(\mathbf{A_1}\mathbf{\Theta}\right)^n\mathbf{A_1}.\end{aligned} \tag{3.5}$$

Then substituting $\mathbf{\Theta}$ (3.1) and $\mathbf{A_1}$ (3.3) into (3.5), we have

$$\begin{aligned}\mathbf{D} &= \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} - \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}\begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix}\begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} + O(\mathbf{A_1}^3) \\ &= \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} - \begin{pmatrix} a_{11}^2\theta_{11} & a_{11}a_{22}\theta_{12} \\ a_{11}a_{22}\theta_{12} & a_{22}^2\theta_{22} \end{pmatrix} + O(\mathbf{A_1}^3) \\ &= \begin{pmatrix} a_{11} - a_{11}^2\theta_{11} + O(a^3) & -a_{11}a_{22}\theta_{12} + O(a^3) \\ -a_{11}a_{22}\theta_{12} + O(a^3) & a_{22} - a_{22}^2\theta_{22} + O(a^3) \end{pmatrix},\end{aligned} \tag{3.6}$$

where $O(\mathbf{A_1}^3)$ means the absolute-value of the error (the error when $\mathbf{D}$ is estimated by the first two terms in Equation 3.6) is at most some constant times $\mathbf{A_1}^3$, $O(a^3)$ has a similar meaning. As shown in (3.2), in order to minimize $\log\frac{|\mathbf{\Sigma}|}{|\mathbf{D}|}$, $|\mathbf{D}|$ needs to be maximized subject to the trace constraint: $tr(\mathbf{D}) \leq 2d$. The equivalent trace constraint can be expressed as

$$\mathbf{D} = \begin{pmatrix} \epsilon d_1 & .. \\ .. & \epsilon d_2 \end{pmatrix}, \tag{3.7}$$

where $d_1\epsilon + d_2\epsilon \leq 2d$. Based on (3.6) and (3.7), we can get an equation

$$\begin{pmatrix} a_{11} - a_{11}^2\theta_{11} + O(a^3) & -a_{11}a_{22}\theta_{12} + O(a^3) \\ -a_{11}a_{22}\theta_{12} + O(a^3) & a_{22} - a_{22}^2\theta_{22} + O(a^3) \end{pmatrix} = \begin{pmatrix} \epsilon d_1 & .. \\ .. & \epsilon d_2 \end{pmatrix}, \tag{3.8}$$

where the corresponding diagonal entries are equal

$$a_{ii} - a_{ii}^2\theta_{ii} + O(a^3) = \epsilon d_i, \qquad i = 1, 2. \tag{3.9}$$

The solution of this quadratic equation is

$$\begin{aligned} a_{ii} &= \frac{1 \pm \sqrt{1 - 4\theta_{ii}\left[\epsilon d_i - O(a^3)\right]}}{2\theta_{ii}} \\ &= \frac{1 \pm \sqrt{1 + 4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]}}{2\theta_{ii}}, \qquad i = 1, 2. \end{aligned} \tag{3.10}$$

As ϵ approaches 0, ϵd_i also approaches 0. Therefore the left hand of (3.9) reaches 0 at the same time, that is

$$\lim_{\epsilon \to 0} \epsilon d_i = \lim_{\epsilon \to 0} \left(a_{ii} - a_{ii}^2\theta_{ii} + O(a^3)\right) = 0. \tag{3.11}$$

One conclusion can be drawn from (3.11) which is

$$\lim_{\epsilon \to 0} a_{ii} = 0. \tag{3.12}$$

Thus, there is only one solution for a_{ii} which is

$$a_{ii} = \frac{1 - \sqrt{1 + 4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]}}{2\theta_{ii}}, \qquad i = 1, 2. \tag{3.13}$$

The Taylor series expansion of $\sqrt{1+x}$ is given by

$$\sqrt{1+x} = 1 + x\frac{1}{2} - x^2\frac{1}{8} + \sum_{n=3}^{\infty} x^n (-1)^{n-1} \frac{(2n-3)!}{n!(n-2)!2^{2n-2}}, \tag{3.14}$$

similarly, we can find the Taylor expansion of $\sqrt{1 + 4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]}$ which is

$$\begin{aligned}\sqrt{1 + 4\theta_{ii}\left[-\epsilon d_i + O(a_i^3)\right]} &= 1 + 4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]\frac{1}{2} - \left\{4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]\right\}^2\frac{1}{8} \\ &\quad + \sum_{n=3}^{\infty}\left\{4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]\right\}^n (-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}} \\ &= 1 + 2\theta_{ii}\left[-\epsilon d_i + O(a^3)\right] - 2\theta_{ii}^2\left[-\epsilon d_i + O(a^3)\right]^2 \\ &\quad + \sum_{n=3}^{\infty}\left\{4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]\right\}^n (-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}}.\end{aligned} \tag{3.15}$$

After substituting the expansion back into (3.13), we have

$$\begin{aligned}a_{ii} =& \frac{1}{2\theta_{ii}}\Bigg\{1 - 1 - 2\theta_{ii}\left[-\epsilon d_i + O(a^3)\right] + 2\theta_{ii}^2\left[-\epsilon d_i + O(a^3)\right]^2 \\ & - \sum_{n=3}^{\infty}\left\{4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]\right\}^n (-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}}\Bigg\} \\ =& \left[\epsilon d_i - O(a^3)\right] + \theta_{ii}\left[-\epsilon d_i + O(a_i^3)\right]^2 \\ & - \frac{1}{2\theta_{ii}}\sum_{n=3}^{\infty}\left\{4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]\right\}^n (-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}}.\end{aligned} \tag{3.16}$$

Consider the limit of $\frac{a_{ii}}{\epsilon}$, given by

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{a_{ii}}{\epsilon} &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\{\left[\epsilon d_i - O(a^3)\right] + b_{ii}\left[-\epsilon d_i + O(a^3)\right]^2 \\ &- \frac{1}{2\theta_{ii}}\sum_{n=3}^{\infty}\left\{4\theta_{ii}\left[-\epsilon d_i + O(a^3)\right]\right\}^n (-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}}\} \\ &= d_i.\end{aligned} \tag{3.17}$$

We are interested in the difference of these two rate distortion functions (Equation 2.6 and 3.2) when d (or ϵ) is small enough and this problem can be stated as follows

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{R(\mathbf{D})_{\{X_1\},\{X_2\}} - R(\mathbf{D})_{\{X_1,X_2\}}}{\epsilon^2} \\ &= \lim_{\epsilon\to 0}\frac{\frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{|(\mathbf{\Theta}+\mathbf{A_1}^{-1})^{-1}|} - \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{d^2}}{\epsilon^2} \\ &= \lim_{\epsilon\to 0}\frac{\frac{1}{2}\log\frac{d^2}{|(\mathbf{\Theta}+\mathbf{A_1}^{-1})^{-1}|}}{\epsilon^2},\end{aligned} \tag{3.18}$$

where the denominator in the logarithm $|(\mathbf{\Theta} + \mathbf{A_1}^{-1})^{-1}|$ can be calculated as

$$|(\mathbf{\Theta} + \mathbf{A_1}^{-1})^{-1}| = \epsilon^2 d_1 d_2 - \left[a_{11}a_{22}\theta_{12} + O(a^3)\right]^2, \tag{3.19}$$

this is because in (3.7) we rewrite the constraint and in (3.8) we found an equation about $\mathbf{D}$. By substituting the determinant in (3.19) and flipping the fraction in the logarithm, the limit in (3.18) becomes

$$\begin{aligned}
\lim_{\epsilon\to 0}\frac{\frac{1}{2}\log\frac{d^2}{|(\mathbf{\Theta}+\mathbf{A_1}^{-1})^{-1}|}}{\epsilon^2} &= \lim_{\epsilon\to 0}\frac{\frac{1}{2}\ln\frac{d^2}{\epsilon^2 d_1 d_2-\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}}{\epsilon^2}\\
&= \lim_{\epsilon\to 0}\frac{-\frac{1}{2}\ln\frac{\epsilon^2 d_1 d_2-\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}{\epsilon^2 d_1 d_2}}{\epsilon^2} \qquad (3.20)\\
&= -\frac{1}{2}\lim_{\epsilon\to 0}\frac{ln\left\{1-\frac{\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}{\epsilon^2 d_1 d_2}\right\}}{\epsilon^2}.
\end{aligned}$$

Note that ϵ and d approach zero at a similar speed and $\epsilon^2 d_1 d_2$ is replaced by d^2 in the second step, this is because the diagonal entries of matrix $\mathbf{D}$ are much greater than off-diagonal entries, thus, the product of diagonal elements contributes the most to the determinant of $\mathbf{D}$. Then, the determinant of $\mathbf{D}$ which can be approximated by the product of all diagonal entries needs to be maximized subject to a trace constraint, which is the same optimization problem in (2.4). Thus by a similar argument we can conclude that $\epsilon d_1 \approx \epsilon d_2 \approx d$ and $\epsilon d_1 d_2 \approx d^2$.

It is known that L'Hospital's Rule can help evaluate limits involving indeterminate forms, one example is

$$\lim_{x\to 0}\frac{\ln(1-x)}{x} = \lim_{x\to 0}\frac{\frac{-1}{1-x}}{1} = \lim_{x\to 0}-(1+\frac{x}{1-x}) = -1. \qquad (3.21)$$

Similarly L'Hospital's Rule can be applied to the limit in (3.20)

$$
\begin{aligned}
(3.20) &= -\frac{1}{2}\lim_{\epsilon\to 0}\frac{ln\left\{1-\frac{\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}{\epsilon^2 d_1 d_2}\right\}}{\frac{\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}{\epsilon^2 d_1 d_2}}\frac{\frac{\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}{\epsilon^2 d_1 d_2}}{\epsilon^2} \\
&= -\frac{1}{2}\lim_{\epsilon\to 0}(-1)\frac{\frac{\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}{\epsilon^2 d_1 d_2}}{\epsilon^2} \\
&= \frac{1}{2}\lim_{\epsilon\to 0}\frac{\frac{\left[a_{11}a_{22}\theta_{12}+O(a^3)\right]^2}{\epsilon^2 d_1 d_2}}{\epsilon^2} \\
&= \frac{1}{2}\lim_{\epsilon\to 0}\left\{\frac{a_{11}^2 a_{22}^2\theta_{12}^2+O(a^6)-2a_{11}a_{22}\theta_{jk}O(a^3)}{\epsilon^4 d_1 d_2}\right\} \\
&= \frac{1}{2}\lim_{\epsilon\to 0}\left\{\frac{a_{11}^2 a_{22}^2\theta_{12}^2+O(a^6)-O(a^5)}{\epsilon^4 d_1 d_2}\right\},
\end{aligned}
\tag{3.22}
$$

where the first step is rearranging the limit to form an indeterminate form of $\frac{0}{0}$ and the second step applies the result in (3.21), the remaining steps follow the square formula. As shown in (3.12) and (3.17), a approaches zero when ϵ is small enough and they approach zero at a similar speed. Thus the limit in (3.22) can be calculated as

$$
\begin{aligned}
\lim_{\epsilon\to 0}\frac{R(\mathbf{D})_{\{X_1\},\{X_2\}}-R(\mathbf{D})_{\{X_1,X_2\}}}{\epsilon^2} &= \frac{1}{2}\lim_{\epsilon\to 0}\left[\frac{a_{11}^2 a_{22}^2\theta_{12}^2}{\epsilon^4 d_1 d_2}+\frac{O(a^6)}{\epsilon^4 d_1 d_2}-\frac{O(a^5)}{\epsilon^4 d_1 d_2}\right] \\
&= \frac{1}{2}\left[\frac{d_1^2 d_2^2\theta_{12}^2}{d_1 d_2}+0-0\right] \\
&= \frac{1}{2}d_1 d_2\theta_{12}^2.
\end{aligned}
\tag{3.23}
$$

Where θ_{12} is the off-diagonal entry in the information matrix and d_1 and d_2 are the corresponding diagonal element in the distortion matrix.

3.2 The Three-Source Case

Assume the joint distribution of three sources X_1, X_3 and X_3 is multivariate normal with mean vector $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3]^T$, variance covariance matrix $\boldsymbol{\Sigma}$, and information matrix $\boldsymbol{\Theta}$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}, \qquad \boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{12} & \theta_{22} & \theta_{23} \\ \theta_{13} & \theta_{23} & \theta_{33} \end{pmatrix}$$

and entries in $\boldsymbol{\Theta}$ are nonzero. A Markov network of X_1, X_2 and X_3 can be drawn in Figure 3.4,

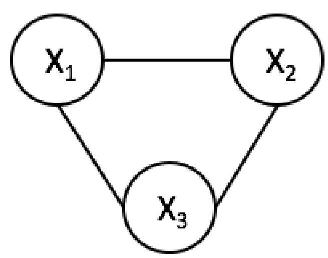

Figure 3.4: The Markov network of X_1, X_2 and X_3.

in which there are three nodes which are X_1, X_2 and X_3 and three edges between the nodes because none of the off-diagonal entries is zero.

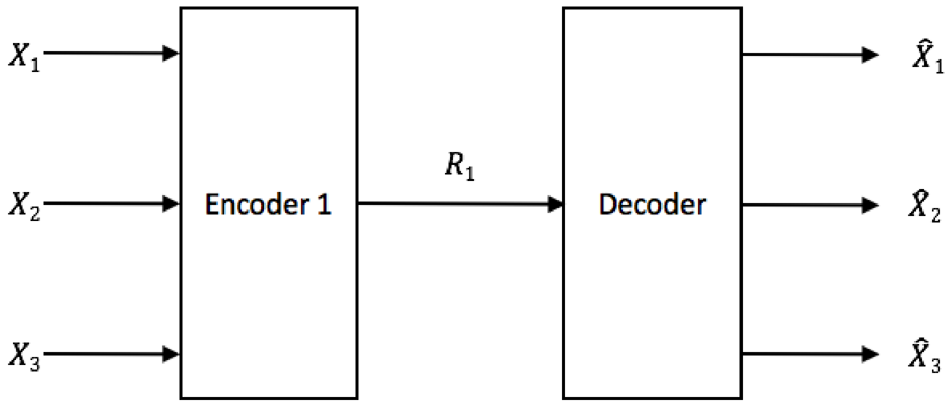

Figure 3.5: One encoder which observes X_1, X_2, and X_3 simultaneously.

Figure 3.5 shows the coding system with a centralized encoder which can also be considered as the most powerful encoder structure since it can observe all sources simultaneously. As a result, this centralized coding structure can observe all the edges in the Markov network between X_1, X_2 and X_3. The rate distortion function of this coding system is shown in (2.7). In this section, four coding systems with non-centralized encoders will be discussed.

3.2.1 Coding system 1

The second most powerful encoder is shown in Figure 3.6. There are three encoders and each one can observe two different sources. All the edges in the Markov network between X_1, X_2 and X_3 can be observed by this coding setup.

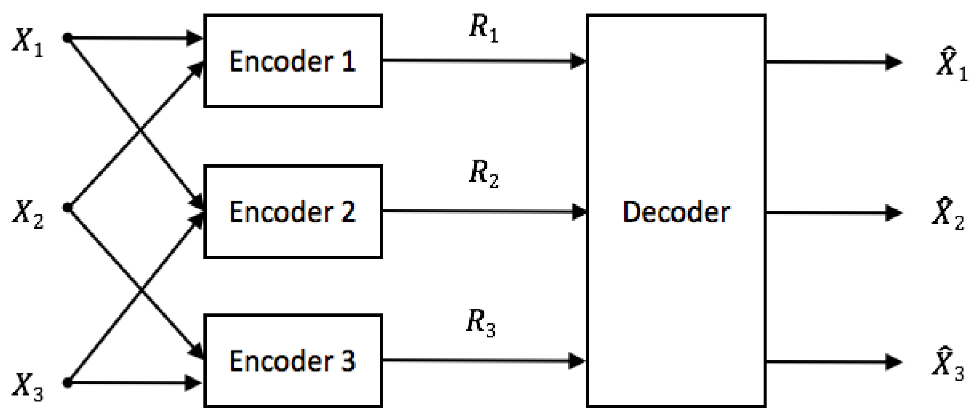

Figure 3.6: Three encoders which observe $\{X_1, X_2\}$, $\{X_1, X_3\}$, and $\{X_2, X_3\}$.

In Chen *et al.* (2007), it was shown that the rate distortion limit of this coding system is exactly the same with the centralized coding system which is

$$R(\mathbf{D})_{\{X_1,X_2\},\{X_1,X_3\},\{X_2,X_3\}} \underset{d \to 0}{=} \frac{1}{2} \log \frac{|\mathbf{\Sigma}|}{d^3}. \tag{3.24}$$

It is surprising to find that even though this decentralized coding system seems to be less powerful than the centralized coding system, the optimized results of rate distortion limit are the same; that is, we can say that they are equal powered for quadratic Gaussian source coding.

3.2.2 Coding system 2

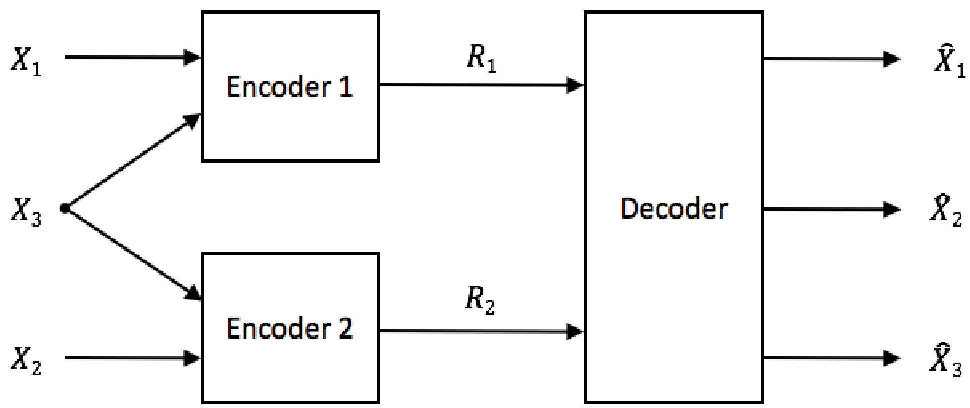

Figure 3.7: Two encoders which observe $\{X_1, X_3\}$ and $\{X_2, X_3\}$.

The coding system in Figure 3.7 consists of two encoders each of which observe two sources. The edges between X_1, X_3 and X_2, X_3 can be observed by this coding setup, but the edge between X_1 and X_2 is not observed. The rate distortion function of this system is

$$\begin{aligned} R(\mathbf{D})_{\{X_1,X_3\},\{X_2,X_3\}} &\underset{d\to 0}{=} \min \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{|\mathbf{D}|} \\ \text{subject to: } \mathbf{0} \preceq \mathbf{D} &= \text{diag}\Big(\big(\mathbf{\Theta}_{1,2} + \text{diag}(a_{11}, a_{22})^{-1}\big)^{-1}, a_{33}\Big), \\ tr(\mathbf{D}) &\leq 3d, \end{aligned} \tag{3.25}$$

where $\mathbf{\Theta}_{1,2}$ is the 2nd order leading principal submatrix of information matrix $\mathbf{\Theta}$, i.e.,

$$\mathbf{\Theta}_{1,2} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix}.$$

$\mathbf{D}$ can be listed in a matrix form which is

$$\mathbf{D} = \begin{pmatrix} \left(\begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}^{-1} \right)^{-1} & \mathbf{0} \\ \mathbf{0}^T & a_{33} \end{pmatrix} \tag{3.26}$$

where $\mathbf{0} = [0, 0]^T$. Similarly, the trace constraint here can be equivalently written in a matrix equation form like

$$\begin{pmatrix} \left(\begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}^{-1} \right)^{-1} & \mathbf{0} \\ \mathbf{0}^T & a_{33} \end{pmatrix} = \begin{pmatrix} \epsilon d_1 & .. & .. \\ .. & \epsilon d_2 & .. \\ .. & .. & \epsilon d_3 \end{pmatrix}, \tag{3.27}$$

which can be further split into two submatrix equations given by

$$\begin{aligned} &(1) \quad \left(\begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}^{-1} \right)^{-1} = \begin{pmatrix} \epsilon d_1 & .. \\ .. & \epsilon d_2 \end{pmatrix}, \\ &(2) \quad a_{33} = \epsilon d_3. \end{aligned} \tag{3.28}$$

Therefore, this system can be transferred into a 2×2 case which is exactly the same as the one in (3.2), thus the result should be the same with (3.23) which is

$$\lim_{\epsilon \to 0} \frac{R(\mathbf{D})_{\{X_1,X_3\},\{X_2,X_3\}} - R(\mathbf{D})_{\{X_1,X_2,X_3\}}}{\epsilon^2} = \frac{1}{2} d_1 d_2 \theta_{12}^2. \tag{3.29}$$

3.2.3 Coding system 3

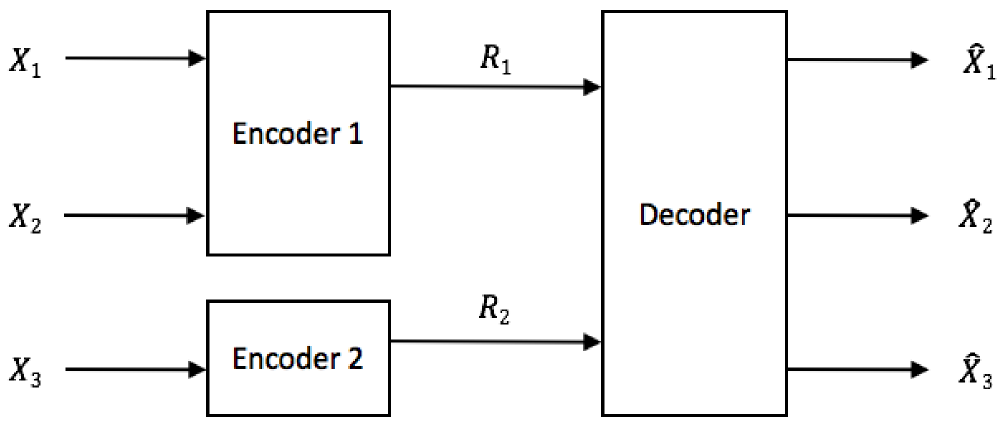

Figure 3.8: Two encoders which observe $\{X_1, X_2\}$ and $\{X_3\}$.

Figure 3.8 shows the coding system consisting of two encoders which observe $\{X_1, X_2\}$ and $\{X_3\}$ separately. This coding system can see the edge between X_1 and X_2, but not the edge between X_1, X_3 and X_2, X_3 in their Markov network. The rate distortion function of this system was found by Wang and Chen (2013) given by

$$\begin{aligned} R(\mathbf{D})_{\{X_1,X_2\},\{X_3\}} &\underset{d\to 0}{=} \min \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{|\mathbf{D}|} \\ \text{subject to: } \mathbf{0} \preceq \mathbf{D} &= \Big(\mathbf{\Theta} + \operatorname{diag}\big(\mathbf{A}_{1,2}, a_{33}\big)^{-1}\Big)^{-1}, \\ tr(\mathbf{D}) &\leq 3d, \end{aligned} \tag{3.30}$$

where $\mathbf{A}_{1,2}$ is a 2×2 positive semidefinite matrix. Denote $\operatorname{diag}\big(\mathbf{A}_{1,2}, a_3\big)$ as $\mathbf{A_2}$ with the form of

$$\mathbf{0} \preceq \mathbf{A_2} = \operatorname{diag}\big(\mathbf{A}_{1,2}, a_3\big) = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}. \tag{3.31}$$

By the Sherman/Morrison formula, $\mathbf{D}$ can be expanded as

$$\begin{aligned} \mathbf{D} &= (\mathbf{\Theta} + \mathbf{A_2}^{-1})^{-1} \\ &= \mathbf{A_2} - \mathbf{A_2}\mathbf{\Theta}\mathbf{A_2} + \sum_{n=2}^{n=\infty}(-1)^n(\mathbf{A_2}\mathbf{\Theta})^n\mathbf{A_2}, \end{aligned} \tag{3.32}$$

then we can substitute $\mathbf{A_2}$ and $\boldsymbol{\Theta}$ and calculate the matrix entries, (3.32) becomes

$$
\begin{aligned}
(3.22) &= \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{12} & \theta_{22} & \theta_{23} \\ \theta_{13} & \theta_{23} & \theta_{33} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} + O({\mathbf{A_2}}^3) \\
&= \begin{pmatrix} \begin{matrix} a_{11} - \Big(a_{11}^2\theta_{11} + \\ 2a_{11}a_{12}\theta_{12} + a_{12}^2\theta_{22}\Big) \end{matrix} & \begin{matrix} a_{12} - \Big(a_{11}a_{12}\theta_{11} + a_{12}^2\theta_{12} \\ a_{11}a_{22}\theta_{12} + a_{12}a_{22}\theta_{22}\Big) \end{matrix} & \begin{matrix} -\Big(a_{11}a_{33}\theta_{13} + \\ a_{12}a_{33}\theta_{23}\Big) \end{matrix} \\ \begin{matrix} a_{12} - \Big(a_{11}a_{12}\theta_{11} + a_{12}^2\theta_{12} \\ a_{11}a_{22}\theta_{12} + a_{12}a_{22}\theta_{22}\Big) \end{matrix} & \begin{matrix} a_{22} - \Big(a_{12}^2\theta_{11} + \\ 2a_{12}a_{22}\theta_{12} + a_{22}^2\theta_{22}\Big) \end{matrix} & \begin{matrix} -\Big(a_{12}a_{33}\theta_{13} + \\ a_{22}a_{33}\theta_{23}\Big) \end{matrix} \\ \begin{matrix} -\Big(a_{11}a_{33}\theta_{13} + \\ a_{12}a_{33}\theta_{23}\Big) \end{matrix} & \begin{matrix} -\Big(a_{12}a_{33}\theta_{13} + \\ a_{22}a_{33}\theta_{23}\Big) \end{matrix} & a_{33} - a_{33}^2\theta_{33} \end{pmatrix} \\
&\quad + O({\mathbf{A_2}}^3)
\end{aligned} \tag{3.33}
$$

by which one can get the order of each entry

$$
\mathbf{D} = \begin{pmatrix} O(a) & O(a_{12} - a^2 + a^3) & O(a^2) \\ O(a_{12} - a^2 + a^3) & O(a) & O(a^2) \\ O(a^2) & O(a^2) & O(a) \end{pmatrix}. \tag{3.34}
$$

where a_{11}, a_{22} and a_{33} are close and represented by a here. The trace constraint on $\mathbf{D}$ in (3.30) can be rewritten in matrix form, the equation for these two representations can be found as

$$\mathbf{D} = \begin{pmatrix} O(a) & O(a_{12} - a^2 + a^3) & O(a^2) \\ O(a_{12} - a^2 + a^3) & O(a) & O(a^2) \\ O(a^2) & O(a^2) & O(a) \end{pmatrix} = \begin{pmatrix} \epsilon d_1 & .. & .. \\ .. & \epsilon d_2 & .. \\ .. & .. & \epsilon d_3 \end{pmatrix} \tag{3.35}$$

where $\epsilon d_1 + \epsilon d_2 + \epsilon d_3 = 3d$. Based on this equation one can say that all the corresponding diagonal entries are equal to each other, specifically

$$O(a) = \epsilon d_i, \qquad i = 1, 2, 3 \tag{3.36}$$

which means that a_{ii} approaches zero at a similar speed with ϵ. In order to achieve the goal of reaching the rate distortion limit in (3.30), $\frac{1}{2} \log \frac{|\mathbf{\Sigma}|}{|\mathbf{D}|}$ needs to be minimized which is equivalent to maximizing $|\mathbf{D}|$. The determinant can then be calculated based on (3.34)

$$\begin{aligned} |\mathbf{D}| &= O(a)^3 - 2O(a^2)^2 O(a) - O(a)O(a_{12} - a^2 + a^3)^2 + 2O(a^2)^2 O(a_{12} - a^2 + a^3) \\ &= O(a^3) - 2O(a^5) - O(a)O(a_{12} - a^2 + a^3)^2 + 2O(a^4)O(a_{12} - a^2 + a^3) \end{aligned} \tag{3.37}$$

Since a is positive and close to zero, we know that the higher order a term is, the smaller it will be. Thus one can conclude that a_{12} is equivalent to a second order term of a in order to cancel the second order term in the entry $O(a_{12} - a^2 + a^3)$ and maximize the determinate of $\mathbf{D}$. The maximized determinant of $\mathbf{D}$ is

$$\begin{aligned} |\mathbf{D}|_{max} &= O(a^3) - 2O(a^5) + O(a^7) \\ &\approx O(a^3) \\ &= \epsilon d_1 d_2 d_3 \end{aligned} \tag{3.38}$$

Now we can calculate the limit of rate distortion difference given by

$$\begin{aligned} &\lim_{\epsilon \to 0} \frac{R(\mathbf{D})_{\{X_1,X_2\},\{X_3\}} - R(\mathbf{D})_{\{X_1,X_2,X_3\}}}{\epsilon^2} \\ &= \frac{1}{2} \lim_{\epsilon \to 0} \frac{\log \frac{d^3}{|\mathbf{D}|}}{\epsilon^2} \\ &= -\frac{1}{2} \lim_{\epsilon \to 0} \frac{\log \frac{|\mathbf{D}|}{\epsilon^3 d_1 d_2 d_3}}{\epsilon^2} \end{aligned} \tag{3.39}$$

where the last step is because $\epsilon d_1 d_2 d_3 \approx d^3$. The reason that we can replace d^3 by $\epsilon d_1 d_2 d_3$ is shown in (3.38), the main contribution to the determinate of $\mathbf{D}$ is $O(a^3)$, thus $|\mathbf{D}|$ can be approximated by the product of diagonal elements ($\epsilon d_1 d_2 d_3$), and this becomes a product optimization problem subject to a summation constraint. This type of optimization problem was discussed in Chapter 2 (Equation 2.4), where the result contains equal elements, i.e. $\epsilon d_1 \approx \epsilon d_2 \approx \epsilon d_3 \approx d$ and ${}_1 d_2 d_3 \approx d^3$. Then after calculating the determinant of $\mathbf{D}$ and substituting it into (3.39), we have

$$\begin{aligned} (3.39) &= -\frac{1}{2} \lim_{\epsilon \to 0} \frac{\log \frac{\epsilon^3 d_1 d_2 d_3 - d_{11} d_{23} d_{32} - d_{22} d_{13} d_{31} - d_{33} d_{12} d_{12} + d_{12} d_{23} d_{31} + d_{13} d_{21} d_{32}}{\epsilon^3 d_1 d_2 d_3}}{\epsilon^2} \\ &= -\frac{1}{2} \lim_{\epsilon \to 0} \frac{\log \left(1 - \frac{d_{11} d_{23}^2 + d_{22} d_{13}^2 + d_{33} d_{12}^2 - 2 d_{12} d_{23} d_{13}}{\epsilon^3 d_1 d_2 d_3}\right)}{\epsilon^2} \end{aligned} \tag{3.40}$$

where d_{ij} is the entry in the ith row and jth column. By rearranging the expression and applying L'Hospital's Rule, the limit becomes

$$\begin{aligned}(3.40) &= -\frac{1}{2}\lim_{\epsilon\to 0}\frac{\log\left(1-\frac{d_{11}d_{23}^2+d_{22}d_{13}^2+d_{33}d_{12}^2-2d_{12}d_{23}d_{13}}{\epsilon^3 d_1 d_2 d_3}\right)}{\frac{d_{11}d_{23}^2+d_{22}d_{13}^2+d_{33}d_{12}^2-2d_{12}d_{23}d_{13}}{\epsilon^3 d_1 d_2 d_3}} \times \frac{\frac{d_{11}d_{23}^2+d_{22}d_{13}^2+d_{33}d_{12}^2-2d_{12}d_{23}d_{13}}{\epsilon^3 d_1 d_2 d_3}}{\epsilon^2} \\ &= \frac{1}{2}\lim_{\epsilon\to 0}\frac{\frac{d_{11}d_{23}^2+d_{22}d_{13}^2+d_{33}d_{12}^2-2d_{12}d_{23}d_{13}}{\epsilon^3 d_1 d_2 d_3}}{\epsilon^2}.\end{aligned} \tag{3.41}$$

From (3.38), we can also see $\lim_{\epsilon\to 0}\frac{a_{ii}}{\epsilon} = d_i$, substituting this conclusion and the order of each entry from $\mathbf{D}$, we are able to get

$$\begin{aligned}(3.40) &= \frac{1}{2}\lim_{\epsilon\to 0}\frac{d_{11}d_{23}^2+d_{22}d_{13}^2+O(a^7)-2O(a^7)}{\epsilon^5 d_1 d_2 d_3} \\ &= \frac{1}{2}\lim_{\epsilon\to 0}\frac{d_{11}d_{23}^2+d_{22}d_{13}^2}{\epsilon^5 d_1 d_2 d_3} \\ &= \frac{1}{2}d_1 d_3 \theta_{13} + \frac{1}{2}d_2 d_3 \theta_{23}.\end{aligned} \tag{3.42}$$

3.2.4 Coding system 4

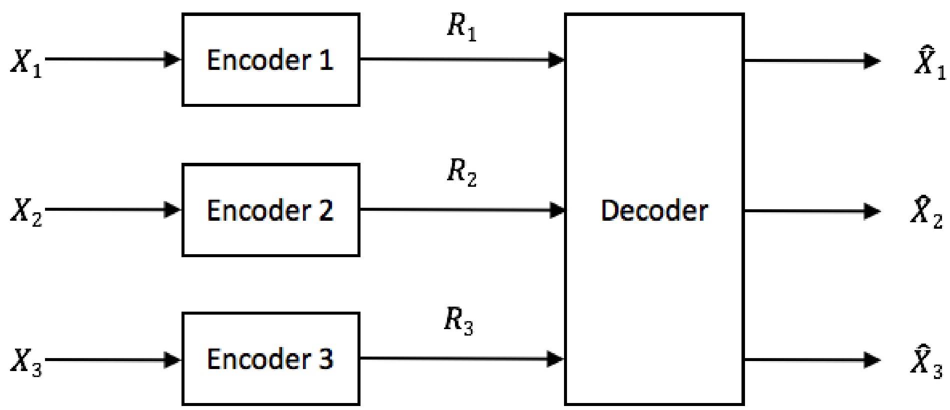

Figure 3.9: Three encoders which observe X_1, X_2, and X_3 individually.

The system in Figure 3.9 is considered the weakest one where X_1, X_2, and X_3 are observed by three different encoders individually. None of the edges in the Markov network of X_1, X_2 and X_3 is captured by this coding system. Wang *et al.* (2010) studied the distortion function of this system which is of similar form with (3.2)

$$\begin{aligned} R(\mathbf{D})_{\{X_1\},\{X_2\},\{X_3\}} &\underset{d\to 0}{=} \min \frac{1}{2} \log \frac{|\mathbf{\Sigma}|}{|\mathbf{D}|} \\ \text{subject to: } &\mathbf{0} \preceq \mathbf{D} = (\mathbf{\Theta} + \mathbf{A_3^{-1}})^{-1}, \\ &tr(\mathbf{D}) \leq 3d, \end{aligned} \tag{3.43}$$

where $\mathbf{D}$ is the distortion matrix and $\mathbf{A_3}$ is the noise covariance matrix with a diagonal form

$$\mathbf{0} \preceq \mathbf{A_3} = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}.$$

By following these steps one can find the limit: $\lim_{\epsilon \to 0} \frac{R(\mathbf{D})_{\{X_1\},\{X_2\},\{X_3\}} - R(\mathbf{D})_{\{X_1,X_2,X_3\}}}{\epsilon^2}$. First expand the matrix $\mathbf{D}$ by the Sherman/Morrison formula, then rewrite the constraint and solve the equation on this constraint. Finally, after solving the equation and finding it's Taylor expansion, the limit can be calculated as

$$\lim_{\epsilon \to 0} \frac{R(\mathbf{D})_{\{X_1\},\{X_2\},\{X_3\}} - R(\mathbf{D})_{\{X_1,X_2,X_3\}}}{\epsilon^2} = \frac{1}{2}(d_1 d_2 \theta_{12}^2 + d_1 d_3 \theta_{13}^2 + d_2 d_3 \theta_{23}^2). \quad (3.44)$$

Since this three sources coding system is a special case of the L sources coding system shown in Chapter 4 (when $L = 3$), more details about solving process can be found in the following section.

3.3 The Multi-Source Case

In this section, we will discuss multi-source generalized multiterminal source coding for quadratic Gaussian cases. The combinatorial problem discussed in Chapter 1 showed there are almost infinite non-equivalent coding structures. However, in this section only two cases will be discussed and compared, one is the most powerful coding system with a centralized encoder and the other is the least powerful system with L encoders which can only observe one distinct source. Assume the joint distribution of sources $X_1, X_2, \cdots, X_L$ is multivariate normal with mean vector $\boldsymbol{\mu}$, variance covariance matrix $\mathbf{\Sigma}$, and information matrix $\mathbf{\Theta}$ which is the inverse of $\mathbf{\Sigma}$ given by

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1L} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1L} & \sigma_{2L} & \dots & \sigma_{LL} \end{pmatrix}, \qquad \mathbf{\Theta} = \mathbf{\Sigma}^{-1} = \begin{pmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1L} \\ \theta_{12} & \theta_{22} & \dots & \theta_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{1L} & \theta_{2L} & \dots & \theta_{LL} \end{pmatrix} \tag{3.45}$$

and entries in $\mathbf{\Theta}$ are nonzero. A Markov network of X_1, X_2, $\cdots$, and X_3 can be found which is a pairwise Markov network, this is shown in Figure 3.10.

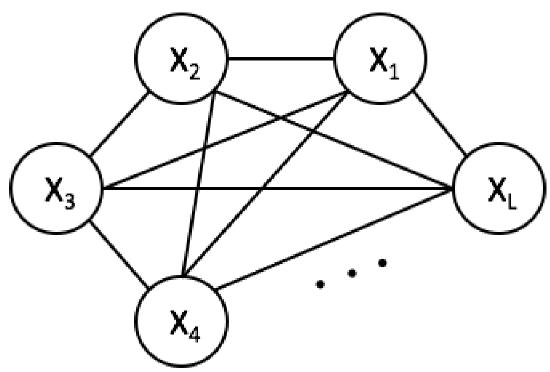

Figure 3.10: The Markov network of X_1, X_2, $\cdots$, X_L.

The most powerful coding system with L sources is shown in Figure 2.2, in which X_1, X_2, $\cdots$, X_L are observed jointly by one Encoder and forwarded to a centralized decoder where the compressed information is reconstructed. The least powerful coding system is shown in Figure 3.11 which can not observe any edge in the Markov network of X_1, X_2, $\cdots$, X_L. There are L encoders and each one observes one distinct source.

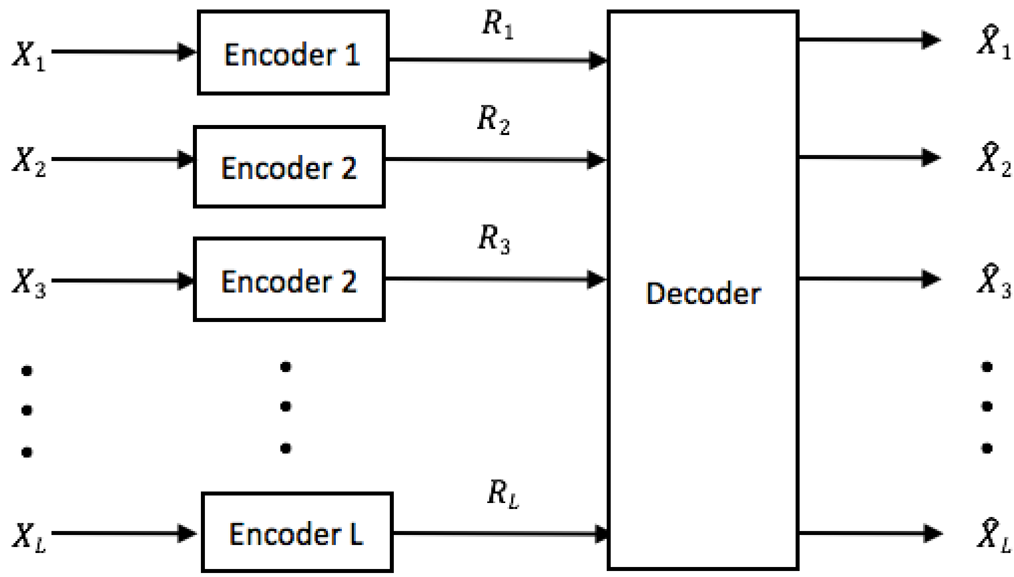

Figure 3.11: L encoders which observe X_1, X_2, $\cdots$, X_L individually.

The distortion function of this coding system was found by Wang *et al.* (2010) which is

$$\begin{aligned} R(\mathbf{D})_{\{X_1\},\{X_2\},\cdots,\{X_L\}} &\underset{d\to 0}{=} \min \frac{1}{2}\log\frac{|\mathbf{\Sigma}|}{|\mathbf{D}|} \\ \text{subject to: } &\mathbf{0} \preceq \mathbf{D} = (\mathbf{\Theta} + \mathbf{A_4^{-1}})^{-1}, \\ &tr(\mathbf{D}) \leq Ld, \end{aligned} \tag{3.46}$$

where $\mathbf{D}$ is the distortion matrix and $\mathbf{A_4}$ is the noise covariance matrix which captures the structure of encoders

$$\mathbf{0} \preceq \mathbf{A_4} = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{LL} \end{pmatrix}.$$

The equivalent trace constraint can be expressed as the following equation

$$\mathbf{D} = (\mathbf{\Theta} + \mathbf{A_4}^{-1})^{-1} = \begin{pmatrix} \epsilon d_1 & .. & \dots & .. \\ .. & \epsilon d_2 & \dots & .. \\ \vdots & \vdots & \ddots & .. \\ .. & .. & \dots & \epsilon d_L \end{pmatrix}, \tag{3.47}$$

where the diagonal elements $\epsilon d_1 + \epsilon d_2 + \cdots, +\epsilon d_L \leq Ld$. However, the non-diagonal elements are unknown. Similarly, the Sherman/Morrison formula can be used to expand the inversion, in order to apply Sherman/Morrison formula, $\mathbf{D}$ needs to be rearranged first. These steps are

$$\begin{aligned} \mathbf{D} &= (\mathbf{\Theta} + \mathbf{A_4}^{-1})^{-1} \\ &= (\mathbf{A_4}^{-1} + \mathbf{\Theta})^{-1} \\ &= \left[\mathbf{A_4}^{-1}(\mathbf{I} + \mathbf{A_4}\mathbf{\Theta})\right]^{-1} \\ &= (\mathbf{I} + \mathbf{A_4}\mathbf{\Theta})^{-1}\mathbf{A_4} \\ &= \mathbf{A_4} - \mathbf{A_4}\mathbf{\Theta}\mathbf{A_4} + \sum_{n=2}^{n=\infty}(-1)^n(\mathbf{A_4}\mathbf{\Theta})^n\mathbf{A_4}. \end{aligned} \tag{3.48}$$

After substituting the matrices **D** and $\mathbf{A_4}$ in to (3.48), we have

$$(3.48) = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{LL} \end{pmatrix} - \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{LL} \end{pmatrix} \times \begin{pmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1L} \\ \theta_{12} & \theta_{22} & \dots & \theta_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{1L} & \theta_{2L} & \dots & \theta_{LL} \end{pmatrix}$$
$$\times \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{LL} \end{pmatrix} + O({\mathbf{A_4}}^3), \tag{3.49}$$

where $O({\mathbf{A_4}}^3)$ refers to the big O notation. In (3.50) summation and multiplication is calculated as follows:

$$(3.49) = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{LL} \end{pmatrix} - \begin{pmatrix} a_{11}^2\theta_{11} & a_{11}a_{22}\theta_{12} & \dots & a_{11}a_{LL}b_{1L} \\ a_{11}a_{22}\theta_{12} & a_{22}^2\theta_{22} & \dots & a_{22}a_{LL}b_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11}a_{LL}\theta_{1L} & a_{22}a_{LL}\theta_{2L} & \dots & a_{LL}^2\theta_{LL} \end{pmatrix} + O({\mathbf{A_4}}^3)$$
$$= \begin{pmatrix} a_{11} - a_{11}^2\theta_{11} + O(a^3) & -a_{11}a_{22}\theta_{12} + O(a^3) & \cdots & -a_{11}a_{LL}\theta_{1L} + O(a^3) \\ -a_{11}a_{22}\theta_{12} + O(a^3) & a_2 - a_2^2\theta_{22} + O(a^3) & \dots & -a_2a_L\theta_{2L} + O(a^3) \\ \vdots & \vdots & \ddots & \vdots \\ -a_{11}a_{LL}\theta_{1L} + O(a^3) & -a_{22}a_{LL}\theta_{2L} + O(a^3) & \dots & a_{LL} - a_{LL}^2\theta_{LL} + O(a^3) \end{pmatrix} \tag{3.50}$$

Based on (3.47) and (3.50), an equation can be set

$$\begin{pmatrix} a_{11}-a_{11}^2\theta_{11}+O(a^3) & -a_{11}a_{22}\theta_{12}+O(a^3) & \cdots & -a_{11}a_{LL}\theta_{1L}+O(a^3) \\ -a_{11}a_{22}\theta_{12}+O(a^3) & A_4-a_2^2\theta_{22}+O(a^3) & \dots & -a_2a_L\theta_{2L}+O(a^3) \\ \vdots & \vdots & \ddots & \vdots \\ -a_{11}a_{LL}\theta_{1L}+O(a^3) & -a_{22}a_{LL}\theta_{2L}+O(a^3) & \dots & a_{LL}-a_{LL}^2\theta_{LL}+O(a^3) \end{pmatrix} = \begin{pmatrix} \epsilon d_1 & .. & \dots & .. \\ .. & \epsilon d_2 & \dots & .. \\ \vdots & \vdots & \ddots & .. \\ .. & .. & \dots & \epsilon d_L \end{pmatrix} \tag{3.51}$$

where corresponding diagonal entries are equal

$$a_{ii}-a_{ii}^2\theta_{ii}+O(a^3)=\epsilon d_i, \quad i=1,2,\cdots,L. \tag{3.52}$$

Solutions to this quadratic function are

$$a_{ii}=\frac{1\pm\sqrt{1+4\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]}}{2\theta_{ii}}, \quad i=1,2,\cdots,L. \tag{3.53}$$

As ϵ approaches 0, ϵd_i also approaches 0, thus we have

$$\lim_{\epsilon\to 0}\epsilon d_i=\lim_{\epsilon\to 0}\left(a_{ii}-a_{ii}^2\theta_{ii}+O(a^3)\right)=0. \tag{3.54}$$

From (3.54) we can conclude that

$$\lim_{\epsilon\to 0}a_{ii}=0, \tag{3.55}$$

this is because $(a_{ii}-a_{ii}^2\theta_{ii}+O(a^3))$ is a polynomial form of a_{ii}. Thus, there is only one solution for a_i which is

$$a_{ii}=\frac{1-\sqrt{1+4\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]}}{2\theta_{ii}}. \tag{3.56}$$

The Taylor expansion of $\sqrt{1+4\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]}$ can be found as

$$\begin{aligned}\sqrt{1+4\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]} &= 1+2\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]-2\theta_{ii}^2\left[-\epsilon d_i+O(a^3)\right]^2 \\ &+\sum_{n=3}^{\infty}\left\{4\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]\right\}^n(-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}}.\end{aligned} \tag{3.57}$$

After substituting this Taylor expansion into (3.56), a_{ii} becomes

$$\begin{aligned}a_{ii} &= \left[\epsilon d_i-O(a^3)\right]+\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]^2 \\ &-\frac{1}{2b_{ii}}\sum_{n=3}^{\infty}\left\{4\theta_{ii}\left[-\epsilon d_i+O(a^3)\right]\right\}^n(-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}}.\end{aligned} \tag{3.58}$$

We then consider the limit of $\frac{a_{ii}}{\epsilon}$,

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{a_{ii}}{\epsilon} &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Bigg\{\left[\epsilon d_i-O(a_i^3)\right]+b_{ii}\left[-\epsilon d_i+O(a_i^3)\right]^2 \\ &-\frac{1}{2b_{ii}}\sum_{n=3}^{\infty}\Big\{4b_{ii}\left[-\epsilon d_i+O(a_i^3)\right]\Big\}^n(-1)^{n-1}\frac{(2n-3)!}{n!(n-2)!2^{2n-2}}\Bigg\} \\ &= d_i,\end{aligned} \tag{3.59}$$

which shows that a_{ii} becomes small and approaches zero at a similar speed with ϵ and d. Our interest is in finding the limit of the difference of two distortion functions:

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{R(\mathbf{D})_{\{X_1\},\{X_2\},\cdots,\{X_L\}}-R(\mathbf{D})_{\{X_1,X_2,\cdots,X_L\}}}{\epsilon^2} \\ &=\lim_{\epsilon\to 0}\frac{\frac{1}{2}\left(\ln\frac{|\mathbf{\Sigma}|}{|(\mathbf{\Theta}+\mathbf{A_4}^{-1})^{-1}|}-\ln\frac{|\mathbf{\Sigma}|}{d^L|}\right)}{\epsilon^2} \\ &=\lim_{\epsilon\to 0}\frac{\frac{1}{2}\ln\frac{d^L}{|(\mathbf{\Theta}+\mathbf{A_4}^{-1})^{-1}|}}{\epsilon^2}\end{aligned} \tag{3.60}$$

where $R(\mathbf{D})_{\{X_1,X_2,\cdots,X_L\}}$ can be found in (2.5) and $|(\mathbf{\Theta}+\mathbf{A_4}^{-1})^{-1}|$ can becalculated as

$$|(\mathbf{\Theta}+\mathbf{A_4}^{-1})^{-1}| = \epsilon^L \prod_{i=1}^{L} d_i - \sum_{j\neq k} \frac{\epsilon^L \prod_{i=1}^{L} d_i}{\epsilon^2 d_j d_k} \Big[- a_{jj}a_{kk}\theta_{jk} + O(a^3) \Big]^2 + \sum_{n=3}^{L} O(a^{6+2n}\epsilon^{L-n}), \tag{3.61}$$

where the first part is all the diagonal terms, the second part is the summation of those terms with the two off-diagonal terms, and the last part is the summation of the rest of the terms with three or more off-diagonal terms. Substitute (3.61) into (3.60), and the limit becomes

$$\begin{aligned} (3.60) &= \lim_{\epsilon\to 0} \frac{\frac{1}{2}\ln \frac{d^L}{\epsilon^L \prod_{i=1}^{L} d_i - \sum_{j\neq k} \frac{\epsilon^L \prod_{i=1}^{L} d_i}{\epsilon^2 d_j d_k}\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2 + \sum_{n=3}^{L} O(a^{6+2n}\epsilon^{L-n})}}{\epsilon^2} \\ &= \lim_{\epsilon\to 0} \frac{-\frac{1}{2}\ln \frac{\epsilon^L \prod_{i=1}^{L} d_i - \sum_{j\neq k} \frac{\epsilon^L \prod_{i=1}^{L} d_i}{\epsilon^2 d_j d_k}\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2 + \sum_{n=3}^{L} O(a^{6+2n}\epsilon^{L-n})}{\epsilon^L \prod_{i=1}^{L} d_i}}{\epsilon^2} \\ &= -\frac{1}{2}\lim_{\epsilon\to 0} \frac{ln \left\{ 1 - \sum_{j\neq k} \frac{\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2}{\epsilon^2 d_j d_k} + \sum_{n=3}^{L} O(\frac{a^{6+2n}}{\epsilon^n}) \right\}}{\epsilon^2} \end{aligned} \tag{3.62}$$

where the second step is flipping the fraction in the logarithm and the third step is simplifying it further. Note that d^L is replaced by $\epsilon^L \prod_{i=1}^{L} d_i$ in the second step because of a similar reason with the 2×2 case on Page 22. Clearly, the limit is an indeterminate form of $\frac{0}{0}$, L'Hospital's Rule should be considered. Similarly, rearrange the limit to get the form $\lim\limits_{x\to 0} \frac{\ln(1-x)}{x}$ because this limit can be calculated by L'Hospital's

Rule which equals -1,

$$(3.62) = -\frac{1}{2}\lim_{\epsilon\to 0}\frac{ln\left\{1-\left[\sum_{j\neq k}\frac{\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2}{\epsilon^2 d_j d_k}-\sum_{n=3}^{L}O(\frac{a^{6+2n}}{\epsilon^n})\right]\right\}}{\sum_{j\neq k}\frac{\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2}{\epsilon^2 d_j d_k}-\sum_{n=3}^{L}O(\frac{a^{6+2n}}{\epsilon^n})}\times \frac{\sum_{j\neq k}\frac{\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2}{\epsilon^2 d_j d_k}-\sum_{n=3}^{L}O(\frac{a^{6+2n}}{\epsilon^n})}{\epsilon^2} \tag{3.63}$$

where the first term is the form of $\lim_{x\to 0}\frac{\ln(1-x)}{x}$, then one can substitute $\lim_{x\to 0}\frac{\ln(1-x)}{x}=-1$ to get

$$\begin{aligned}(3.63) &= -\frac{1}{2}\lim_{\epsilon\to 0}(-1)\frac{\sum_{j\neq k}\frac{\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2}{\epsilon^2 d_j d_k}-\sum_{n=3}^{L}O(\frac{a^{6+2n}}{\epsilon^n})}{\epsilon^2}\\ &= \frac{1}{2}\lim_{\epsilon\to 0}\frac{\sum_{j\neq k}\frac{\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2}{\epsilon^2 d_j d_k}-\sum_{n=3}^{L}O(\frac{a^{6+2n}}{\epsilon^n})}{\epsilon^2}.\end{aligned} \tag{3.64}$$

After applying L'Hospital's rule, we get a simplified expression without a logarithm term. We can further split the fraction into two terms and then expand the quadratic term which are shown here

$$\begin{aligned}(3.64) &= \frac{1}{2}\lim_{\epsilon\to 0}\left\{\sum_{j\neq k}\frac{\left[-a_{jj}a_{kk}\theta_{jk}+O(a^3)\right]^2}{\epsilon^4 d_j d_k}-\sum_{n=3}^{L}O\left(\frac{a^{6+2n}}{\epsilon^{n+2}}\right)\right\}\\ &= \frac{1}{2}\lim_{\epsilon\to 0}\left\{\sum_{j\neq k}\frac{a_{jj}^2a_{kk}^2\theta_{jk}^2-2a_{jj}a_{kk}\theta_{jk}O(a^3)+O(a^6)}{\epsilon^4 d_j d_k}-\sum_{n=3}^{L}O\left(\frac{a^{6+2n}}{\epsilon^{n+2}}\right)\right\}\\ &= \frac{1}{2}\lim_{\epsilon\to 0}\left\{\sum_{j\neq k}\left[\frac{a_{jj}^2a_{kk}^2\theta_{jk}^2}{\epsilon^4 d_j d_k}-\frac{2a_{jj}a_{kk}\theta_{jk}O(a^3)}{\epsilon^4 d_j d_k}+\frac{O(a^6)}{\epsilon^4 d_j d_k}\right]-\sum_{n=3}^{L}O\left(\frac{a^{6+2n}}{\epsilon^{n+2}}\right)\right\}\end{aligned} \tag{3.65}$$

Substitute (3.59) into (3.65), the limit becomes

$$\begin{aligned}(3.60) &= \frac{1}{2}\left\{\sum_{j\neq k}\left[\frac{d_j^2 d_k^2 \theta_{jk}^2}{d_j d_k} + 0 - 0\right] - \sum_{n=3}^{L} 0\right\} \\ &= \frac{1}{2}\sum_{j\neq k} d_j d_k \theta_{jk}^2\end{aligned} \tag{3.66}$$

where $j = 1, 2, \cdots, L, k = 1, 2, \cdots, L$ and $j \neq k$.

Chapter 4

Conclusion

This thesis started off by considering the effect of different encoding setups on the rate distortion function. Chapter 1 showed that as L increased, the number of unique inequivalent source coding systems quickly inflated. For this reason, rate distortion for all $L = 2$ and $L = 3$ sources coding systems were first explored, and only two special cases for the general L sources setup. By comparing all the rate distortion difference limits for two and three source coding systems, one conclusion can be made. If there is an edge between X_j and X_k in the Markov network which is not captured by any encoder, then the term $d_j d_k \theta_{jk}^2$ should be considered in the limit, where θ_{jk} is an off-diagonal entry in the information matrix $\mathbf{\Theta}$ and d_j and d_k are the corresponding diagonal element in the distortion matrix. In the 3 sources case, some coding setups were also shown to approach zero in the rate distortion limit difference which means the centralized system can be matched by ostensibly weaker decentralized ones. Specifically speaking, some decentralized coding systems can achieve the Shannon lower bound in the high-resolution regime.

For L sources, one special case had similar results in the difference of rate distortion function limit. For this case, the most powerful (centralized) was compared with a ostensibly less powerful (distributed) one. It was shown that the difference in rate distortion limits could approach the same term ($d_j d_k \theta_{jk}^2$) provided the necessary encoding setup.

One may want to extend this conclusion to generalized Gaussian multiterminal source coding systems with some other coding topological structures. Intuitively this theorem seems correct; however, the conjecture needs to be proved further for other systems which are not included here.

Bibliography

A. B. Wagner, S. T. and Viswanath, P. (2008). Rate region of the quadratic Gaussian two-encoder source-coding problem. *IEEE Transactions on Information Theory*, **54**, 1938–1961.

Aljohani, A. J., Ng, S. X., and Hanzo, L. (2016). Distributed source coding and its applications in relaying-based transmission. *IEEE Access*, **4**, 1940–1970.

Berger, T. (1978). Multiterminal source coding. *The Information Theory Approach to Communications*, **229**, 171–231.

Chen, J., Etezadi, F., and Khisti, A. (2007). Generalized Gaussian multiterminal source coding and probabilistic graphical models. *Journal of Applied Probability*, **44**, 938–949.

Cover, T. M. (1975). A proof of the data compression theorem of slepian and wolf for ergodic sources (corresp.). *IEEE Transactions on Information Theory*, **21**(2), 226–228.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory, Second Edition*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models Principles and Techniques*. MIT Press, Cambridge, MA.

Oohama, Y. (1997). Gaussian multiterminal source coding. *IEEE Transactions on Information Theory*, **43**, 1912–1923.

Shannon, C. E. (1948). A Mathematical Theory of Communication. **27**, 379423.

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

Wang, J. and Chen, J. (2013). Vector Gaussian two-terminal source coding. *IEEE Transactions on Information Theory*, **59**, 3693 – 3708.

Wang, J. and Chen, J. (2014). Vector Gaussian multiterminal source coding. *IEEE Transactions on Information Theory*, **60**, 55335552.

Wang, J., Chen, J., and Wu, X. (2010). On the sum rate of Gaussian multiterminal source coding: New proofs and results. *IEEE Transactions on Information Theory*, **56**, 3946–3960.