Vector Multiple Description Source Coding

By
Soheil Mohajer

SUBMITTED TO THE
DEPARTMENT OF COMPUTER AND COMMUNICATION SCIENCES
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
AT
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
LAUSANNE, SWITZERLAND
AUGUST 2005

Thesis Supervisor: Suhas N. Diggavi
Professor of Information and Communication Systems
EPFL – LICOS

## Contents

Abstract iii

Acknowledgements v

1 Introduction 1

2 Preliminary 3
   2.1 Rate-Distortion Theory 3
       2.1.1 Rate Distortion Function for Jointly Gaussian Sources 6
       2.1.2 Distortion Rate Function for Jointly Gaussian Sources 8
   2.2 Strong Typicality 14
   2.3 Entropy Power Inequality 16
       2.3.1 Conditional Entropy Power Inequality 16

3 History and Previous Results 19

4 Ozarow’s Theorem for Scalar Gaussian Multiple Descriptions 25
   4.1 Achievability 30
   4.2 Analysis of the Proof 32

5 Achievable Rates for Vector Multiple Descriptions 35
   5.1 Problem Definition and the Main Theorem 35
   5.2 Proof of the vector EGC Theorem 37
       5.2.1 Random Coding 38
5.2.2 Encoding .................................................. 39
5.2.3 Reconstruction ........................................ 40
5.2.4 Analysis of Error ...................................... 41
5.3 Extension of Zhang-Berger Theorem for Vector Multiple Descriptions 46

6 Vector Multiple Descriptions for Jointly Gaussian Sources 49
   6.1 The Outer Bound ...................................... 49
   6.2 Gaussian Encoding .................................... 53
   6.3 Entropy Power Inequality Constraint ................. 57
   6.4 Conditional Markov Chain ............................ 60

7 Conclusion and Future Work 63

Bibliography 67
Abstract

In his famous papers in the late 1940’s, Shannon showed that there is a fundamental trade-off in data compression, between the quality of reconstruction and the rate at which compression can be done. Multiple Description Source Coding is an important method for sending compressed data over unreliable networks, where a lossy reconstruction of information is acceptable. *Multiple route diversity* is a subclass of *diversity schemes*, which are based on sending information over several routes, in order to reduce the information loss; with two or more channels, multiple descriptions of the source are produced and sent over the channels, in such a way that we get guaranteed approximation of the source when any subset of the channels have not failed.

Although designing good codes and finding information-theoretic bounds on the rate of the descriptions and the achievable distortions are considered in many papers, the general problem is still unsolved except for a few special cases. Gaussian source multiple descriptions for two routes is the only case for which the rate-distortion region is known.

In this work, we study the formal notion of multiple description source coding and the rate-distortion theory, and survey several major results in this area. We also investigate the case of jointly Gaussian source multiple descriptions, and try to characterize the achievable rates and distortions for this case. We try to find out whether the correlation between the sources can be used to decrease the rates, in comparison to the case in which each of the sources are encoded individually.

**Keywords:** Multiple Description Source Coding, Rate-Distortion Theory, Jointly Gaussian Sources, Lossy Source Coding, Achievable Rates and Distortions.
Acknowledgements

I am so grateful to my supervisor, Suhas N. Diggavi, for introducing the problem of Multiple Description Source Coding to me and for his advice and encouragement. Thanks to Chao Tian form Cornell University for co-refereeing the thesis. I also thank Etienne Perron, who helped me by providing some of his results, and other members of LICOS and Muriel Bardet for providing me with a friendly and creative work environment. Additional thanks go to Payam Pakzad and Amir F. Dana for many helpful and interesting discussions. Also special thanks to my friend and office-mate Mahdi Cheraghchi, for providing an enjoyable environment in the office and his deep knowledge about almost all computer programs! Finally, the last but not the least, I would like to thank my family for their love and support.
Chapter 1

Introduction

Diversity, is one of the main tools used to forming reliable network communication. Informally speaking, Diversity is a method of conveying information through multiple independent instantiations. Multiple routes diversity, is one of the main classes of diversity. In transmission over networks, random route failures and packet losses degrade performance. Using multiple routes and transmitting over these routes is an important solution for this problem. A fundamental question that arises is how we can best utilize the presence of such route diversity. In order to utilize these conduits, multiple description source coding generates multiple codeword streams to describe a source [5].

If the same information about the source is transmitted over both routes, then this is a form of repetition coding. However, when both routes are successful, there is no advantage in performance of the communication system. Perhaps a more sophisticated technique would be to send correlated descriptions of the source in the two routes such that each description is individually good, but they are different from one another so that if both routes are successful one gets a better approximation of the source. This is the main idea behind the concept of Multiple Description Source
We formalize the MD problem, using the set-up shown in Figure 1.1. Although we will just consider the problem for two descriptions, the similar problem can be posed for any number of descriptions. Given a source sequence $x^n = x(1), x(2), \ldots, x(n)$, we want to design an encoder that sends two descriptions of the source at rates $R_1$ and $R_2$ over the two routes such that we get guaranteed approximations of the source when either route fails, or when both succeed.

In order to describe fundamental bounds on the performance of such techniques precisely, we need to examine the problem from an information theoretic point of view. The main tool to do this is given in rate-distortion theory [2] and [1]. We also need some other useful information theoretic tools during the thesis. So, we will have the preliminaries in the next section and then will survey the history of the problem and obtained results.
Chapter 2
Preliminary

We begin by reviewing some basic definitions and the tools and the theorems in rate-distortion theory and Typicality that will became handy in later chapters. For a comprehensive study of the topic, refer to many standard texts available such as [2],[3] and [1].

2.1 Rate-Distortion Theory

This theory describes fundamental limits of the trade-off between the rate of the representation of a source and the quality of the approximation. Not surprisingly, the origins of this theory are in Shannon’s famous papers [18] and [19].

Given a source sequence

\[ x^n = \{x(1), x(2), \ldots, x(n)\} \]

from a given alphabet \( \mathcal{X} \), the source encoder needs to describe it using \( R \) bits per source sample (i.e., with a total of \( nR \) bits for the sequence). Equivalently, we map the source to the index set

\[ \mathcal{I} = \{1, 2, \ldots, 2^{nR}\} . \]
The goal is that given this description, a decoder is able to approximately reconstruct the source sequence by the sequence
\[ \hat{x}^n = \{ \hat{x}(1), \hat{x}(2), \ldots, \hat{x}(n) \}. \]

This is accomplished by constructing a function
\[ f : \mathcal{I} \rightarrow \hat{X}^n \]
where \( \hat{X} \) is the alphabet over which the reconstruction is done. The distortion measure \( \tilde{d}(x^n, \hat{x}^n) \), quantifies the quality of the approximation between the reconstructed and original source sequence. Typically, the distortion measure is a single-letter function constructed as
\[ \tilde{d}(x^n, \hat{x}^n) = \frac{1}{2} \sum_{i=1}^{n} d(x(i), \hat{x}(i)) \]
where \( d(x, \hat{x}) \) denotes the quality of the approximation for each sample. Square error distance,
\[ d(x, \hat{x}) = |x - \hat{x}|^2, \]
and Hamming distance function are two famous examples for such distortion functions.

Typically, the interest is in the average distortion over the set of input sequences, for the given probability distribution associated with the source sequence. Therefore, the average distortion is \( \mathbb{E}[\tilde{d}(x^t, \hat{x}^t)] \), and the problem becomes one of quantifying the smallest rate \( R \) that be used to describe the source with average fidelity \( D \), asymptotically in the block length \( n \). This is called the rate-distortion function \( R(D) \) and can be given an operational meaning by proving that there exist source codes
that can achieve this fundamental bound. The central result in single source rate-
distortion theory is that $R(D)$ is characterized as \[ R(D) = \min_{p(\hat{x}|x) \colon \mathbb{E}[d(x, \hat{x})] \leq D} I(X; \hat{X}) \] \hspace{1cm} (2.1)

where $I(X; \hat{X})$ represents the mutual information between $X$ and $\hat{X}$. Lossless source

coding, $D = 0$, can be considered as a simple case of this problem. It can be seen
that in this case, $R(0) = H(X)$, where $H(X)$ is the entropy of the discrete source.
Another important special case is when the source sequence comes from a Gaussian
distribution,

\[ X \sim \mathcal{N}(0, \sigma^2), \]

and we are interested in the squared error distortion metric, i.e., $d(x, \hat{x}) = |x - \hat{x}|^2$.
In this case, the solution of 2.1 is

\[ R(D) = \begin{cases} \frac{1}{2} \log \frac{\sigma^2}{D} & \text{if } D \leq \sigma^2 \\ 0 & \text{otherwise} \end{cases} \]

Another way of writing this is in terms of the distortion-rate function $D(R)$, which
characterizes the smallest distortion achievable for a given rate. In the Gaussian case,
we see that

\[ D(R) = \sigma^2 \exp(-2R). \]

We will interchangeably consider these two quantities. The rate defined in (2.1)
guarantees only that the average distortion does not exceed $D$. However, under some
regularity conditions, the rate-distortion function remains the same even when we
require that the probability of the distortion $\tilde{d}(x^n, \hat{x}^n)$ exceeding $D$ to go to zero ([2]
and [1]).
As we are going to consider the MD problem for the jointly Gaussian sources, we will need the rate-distortion and distortion-rate functions. We will describe these functions in the next subsection.

2.1.1 Rate Distortion Function for Jointly Gaussian Sources

Let \( x^n \) and \( y^n \) are two source sequences drawn due Gaussian probability mass functions, \( \mathcal{N}(0, \sigma_x^2) \) and \( \mathcal{N}(0, \sigma_y^2) \), respectively. Given \((x^n, y^n)\), the encoder has to produce a description of them at rate \( R \), such that we can get an estimation for each of the sequences, \( \hat{x}^n \) and \( \hat{y}^n \) which guarantee some individual distortion constraints,

\[
\mathbb{E}[\tilde{d}(x^n, \hat{x}^n)] \leq D_x
\]

and

\[
\mathbb{E}[\tilde{d}(y^n, \hat{y}^n)] \leq D_y
\]

The minimum required rate for this problem is

\[
R(D_x, D_y) = \min_{p(\hat{S}|S)} I(S; \hat{S}) \tag{2.2}
\]

\[
\mathbb{E}[\tilde{d}(x^n, \hat{x}^n)] \leq D_x
\]

\[
\mathbb{E}[\tilde{d}(y^n, \hat{y}^n)] \leq D_y
\]

where \( S = [X, Y]^\dagger \) and \( \hat{S} = [\hat{X}, \hat{Y}]^\dagger \). It is clear that if \( X \) and \( Y \) are independent, the solution of (2.2) is

\[
R(D_x, D_y) = R_x(D_x) + R_y(D_y) = \frac{1}{2} \log \frac{\sigma_x^2 \sigma_y^2}{D_x D_y}. \tag{2.3}
\]

If there is some correlation between the two sources, it is possible to encode \((x^n, y^n)\) using joint encoding by some rate less than (2.3). This problem has been solved by Perron et al.\[9\] and we state the result in the following theorem.
Theorem 1. Let \((X, Y)\) be jointly Gaussian random variables with covariance matrix

\[
K = \begin{pmatrix}
\sigma_x^2 & \rho \sigma_x \sigma_y \\
\rho \sigma_x \sigma_y & \sigma_y^2
\end{pmatrix}.
\]

If either \(D_x > \sigma_x^2\) or \(D_y > \sigma_y^2\), then the problem is trivial and convert to one source rate-distortion problem. If \(D_x < \sigma_x^2\) and \(D_y < \sigma_y^2\), the rate distortion function is given by

\[
R(D_x, D_y) = \begin{cases} 
\frac{1}{2} \log \frac{\sigma_x^2 \sigma_y^2 (1-\rho^2)}{\sigma_x^2 \sigma_y^2 (1-\gamma^2)} & \text{if } |\rho| < \rho_t \\
\max \left\{ \frac{1}{2} \log \frac{\sigma_x^2}{D_x}, \frac{1}{2} \log \frac{\sigma_y^2}{D_y} \right\} & \text{if } |\rho| \geq \rho_t 
\end{cases}
\]

where

\[
\rho_t = \min \left\{ \sqrt{\frac{D_y}{\sigma_y^2} - 1}, \sqrt{\frac{D_x}{\sigma_x^2} - 1} \right\}
\]

and

\[
\gamma = \frac{\sigma_x \sigma_y \left[ |\rho| - \rho_c \right]^+}{\sqrt{D_x D_y}},
\]

where

\[
\rho_c = \sqrt{\frac{(\sigma_x^2 - D_x)(\sigma_y^2 - D_y)}{\sigma_x^2 \sigma_y^2}}.
\]

Here we just give some intuition about the proof and refer to [9] for complete proof. Actually, the problem can be considered in three cases in the sense of how large the correlation between the sources is. If there is a large enough correlation between the sources, such that \(\text{Var}(Y|\hat{X}) \leq D_y\), it suffices to just send a description about \(X\), by rate

\[
R_x(D_X) = \frac{1}{2} \log \frac{\sigma_x^2}{D_x}.
\]

A similar argument can be done in the case of \(\text{Var}(X|\hat{Y}) \leq D_x\). So, for this regime of \(\rho\), which is \(|\rho| > \rho_t\), we just need a rate of

\[
\max \left\{ \frac{1}{2} \log \frac{\sigma_x^2}{D_x}, \frac{1}{2} \log \frac{\sigma_y^2}{D_y} \right\}.
\]
For the other two regimes of $\rho$, the description should contain information about both of the sources. In these cases, the encoder can add some Gaussian noises to the source and encode the noisy version of them which have a larger variance. If the correlation is small enough, i.e., $|\rho| \leq \rho_c$, the covariance of this noise can be diagonal (independent noises). Otherwise, there should be some correlation between the noises.

### 2.1.2 Distortion Rate Function for Jointly Gaussian Sources

In this section we will find the achievable distortion region for compressing a jointly Gaussian source. As we stated in the past part, the inverse problem, finding the minimum rate required for compressing a jointly Gaussian source with a given distortion pair, has been solved recently by Perron et al. [9].

Let $S = [X, Y]^\dagger$ is a zero-mean jointly Gaussian source with covariance matrix

$$K = \begin{pmatrix}
\sigma_x^2 & \mu \sigma_x \sigma_y \\
\mu \sigma_x \sigma_y & \sigma_y^2
\end{pmatrix}.$$

The problem is to find all pairs of $(D_x, D_y)$ such that there exists a coding scheme with deterministic encoding and decoding functions $(\mathcal{E}, \mathcal{D})$

$$U = \mathcal{E}(X, Y)$$

$$\hat{S} = [\hat{X}, \hat{Y}]^\dagger = \mathcal{D}(U)$$

such that rate of the code be less than or equal to some fix rate $R$, and

$$\mathbb{E}[(\hat{X} - X)^2] \leq D_x$$

$$\mathbb{E}[(\hat{Y} - Y)^2] \leq D_y$$

It is easy to use time-sharing argument to show that the region is convex. It is also clear that we just have to find the boundary points of the region. For any $D_x \geq \sigma_x^2$, we
don’t need to send any information about $X$ and we can use all the rate to send the description of $Y$. Thus, using the argument for the single source distortion region, the minimum achievable value for $D_y$ is $\sigma^2_y \exp(-2R)$. Using the similar argument, we can show that for $D_y \geq \sigma^2_y$, $D_x$ is achievable if and only if $D_x \geq \sigma^2_x \exp(-2R)$. Therefore we can focus only on the rectangular region $0 \leq D_x \leq \sigma^2_x$ and $0 \leq D_y \leq \sigma^2_y$. It has been shown that for the Gaussian sources, the optimal encoding scheme is producing Gaussian codewords and using MMSE criterion for the decoding. So, we can assume $S = \hat{S} + N$ where $N$ is a zero-mean Gaussian noise with covariance matrix

$$K_N = \begin{pmatrix} D_x & \alpha \\ \alpha & D_y \end{pmatrix}.$$ 

Obviously we have

$$R = I(S, \hat{S}) = h(S) - h(S|\hat{S}) = h(S) - h(S - \hat{S}|\hat{S}) = h(S) - h(N|\hat{S}) = h(S) - h(N) = \frac{1}{2} \log \det K - \frac{1}{2} \log \det K_N$$

Thus,

$$\sigma^2_x \sigma^2_y (1 - \mu^2) \exp(-2R) = D_x D_y - \alpha^2 \quad (2.5)$$
The other necessary condition is $K_N \preceq K$. This matrix condition, can be expressed as below scalar conditions.

\[
D_x \leq \sigma_x^2 \tag{2.6}
\]
\[
D_y \leq \sigma_y^2 \tag{2.7}
\]
\[
(\sigma_x^2 - D_x)(\sigma_y^2 - D_y) - (\mu \sigma_x \sigma_y - \alpha)^2 \geq 0 \tag{2.8}
\]

For simplicity, by using $p = D_x/\sigma_x^2$, $q = D_y/\sigma_y^2$ and $\beta = \alpha/\sigma_x \sigma_y$ we normalize all the conditions (2.5) and (2.6)-(2.8) as

\[
(1 - \mu^2) \exp(-2R) = pq - \beta^2 \tag{2.9}
\]
\[
(1 - p)(1 - q) - (\mu - \beta)^2 \geq 0 \tag{2.10}
\]

in the square region $0 \leq p \leq 1$ and $0 \leq q \leq 1$. For the rest of the argument we assume $\mu \geq 0$. For negative correlation, the similar argument is valid. It is also easy to show that if some negative value of $\beta$ satisfies the above conditions, $|\beta|$ also satisfies for the same pair of $(p, q)$. So we can assume $\beta \geq 0$. Note that the value of $\beta$ is not important in our solution and is just some auxiliary variable which has to be chosen such that the above conditions are satisfied. From (2.10) we have

\[
\beta \leq \mu + \sqrt{(1 - p)(1 - q)} \tag{2.11}
\]

and

\[
\beta \geq \mu - \sqrt{(1 - p)(1 - q)} \tag{2.12}
\]

where (2.11) holds whenever (2.9) holds. The LHS of (2.12) is negative for some values of $p$ and $q$ and so it will be a trivial condition for those pairs of $(p, q)$. For
other pairs of \((p, q)\), by mixing (2.9) and 2.12, we have

\[
p^2 + q^2 + 2(1 - 2\mu^2)pq - 2(1 - \mu^2)(1 + \exp(-2R))(p + q) +
\]
\[
(1 - \mu^2)^2(1 + \exp(-2R))^2 + 4\mu^2\exp(-2R) \leq 0
\]  

(2.13)

which contains the points inside a rotated ellipse. An important property of this ellipse is that it is tangent to vertical lines \(q = \exp(-2R)\) and \(q = 1\) and horizontal lines \(p = \exp(-2R)\) and \(p = 1\) in the \(p - q\) plane. If we denote by \(A\), \(B\), and \(C\), the points which satisfy the conditions \(pq \geq (1 - \mu^2)\exp(-2R)\), \((1 - p)(1 - q) \geq \mu^2\), and (2.13), respectively, then the boundary points of the desired region is the boundary of \(A \cap (B \cup C)\). We divide the rest of the argument into two separate cases.
Case 1: $R < \frac{1}{2} \log \frac{1+\mu}{1-\mu}$: In this case $A \cap B = \emptyset$ and so we have just to find $A \cap C$. On the other hand, in this case, the ellipse always lies above the homographic function $pq = (1 - \mu^2) \exp(-2R)$. So the boundary pints of the achievable region are given by the ellipse and vertical and horizontal lines at $\exp(-2R)$. This region is plotted for some values of $R$ and $\mu$ in Figure 2.1 and Figure 2.2. This region can be characterized as

$$q \geq p(2\mu^2 - 1) + (1 - \mu^2)(1 + \exp(-2R))$$

$$-2\rho \sqrt{(1 - \mu^2)(1 - p)(p - \exp(-2R))}$$

Figure 2.2: Distortion region for jointly Gaussian sources, $\mu = 0.8$ and various rates.
Figure 2.3: Intersection of three boundary curves; Two lower bounds are tangent to each other

if $p \in [\exp(-2R), 1 - \mu^2(1 - \exp(-2R))]$ and

$q \geq \exp(-2R)$

if $p \geq 1 - \mu^2(1 - \exp(-2R))$.

• Case 2: $R \geq \frac{1}{2} \log \frac{1+\mu}{1-\mu}$: In the second case, although region $A$ intersects both of regions $B$ and $C$, but their boundaries are tangent to each other as it is shown in Figure(2.3). So the central part of the distortion regions restricted by $pq < (1 - \mu^2) \exp(-2R)$ and the marginal parts of the bound coincides the ellipse boundary.
2.2 Strong Typicality

One of the most important techniques which is used frequently in this problem and other problems in the literature is the strong typicality. In the following, we will introduce the main concept and some of its properties which will be used in next sections.

Definition 1. The sequence \( x^n \in \mathcal{X}^n \) is said to be \( \epsilon \)-strong typical with respect to a distribution \( p(x) \) if

1. For all \( a \in \mathcal{X} \) with \( p(a) > 0 \), we have

\[
\left| \frac{1}{n} N(a ; x^n) - p(a) \right| < \frac{\epsilon}{|\mathcal{X}|} \quad (2.14)
\]

2. For all \( a \in \mathcal{X} \) with \( p(a) = 0 \), \( N(a ; x^n) = 0 \).

where \( N(a ; x^n) \) is the number of the symbol \( a \) in the sequence \( x^n \).

The set of all \( \epsilon \)-strong typical vectors of length \( n \) is denoted by \( T_\epsilon(X) \). Similarly we can define strong typicality for joint random variables.

Definition 2. A pair of sequence \( (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \) is said to be \( \epsilon \)-strong typical with respect to a distribution \( p(x, y) \) if

1. For all \( (a, b) \in \mathcal{X}^n \times \mathcal{Y}^n \) with \( p(a, b) > 0 \), we have

\[
\left| \frac{1}{n} N(a, b ; x^n, y^n) - p(a) \right| < \frac{\epsilon}{|\mathcal{X}| |\mathcal{Y}|} \quad (2.15)
\]

2. For all \( (a, b) \in \mathcal{X}^n \times \mathcal{Y}^n \) with \( p(a, b) = 0 \), \( N(a, b ; x^n, y^n) = 0 \).

where \( N(a, b ; x^n, y^n) \) is the number of pairs \( (x, y) \) in the sequence \( (x^n, y^n) \) which equal to \( (a, b) \).

There are many properties about number of typical sequences and their probabilities [1]. We will discuss some of these properties, but ignore for the proofs.
1. For all $\epsilon > 0$, $\Pr(T_{\epsilon}(X)) \to 1$ as $n \to \infty$ for an i.i.d. source, i.e., almost every i.i.d. sequences are typical for large enough $n$.

2. For any $x^n \in T_{\epsilon}(X)$, $\Pr(x^n) = 2^{-n[H(X)\pm o(\epsilon)]}$ where $o(\epsilon) \to 0$ as $n \to \infty$.

3. $|T_{\epsilon}(X)| \simeq 2^{n[H(X)\pm o(\epsilon)]}$.

4. If $(x_i, y_i)$ are chosen independently from joint distribution $p(x, y)$ for $i = 1, \ldots, n$, then $\Pr[(x^n, y^n) \in T_{\epsilon}(X, Y)] \to 1$ as $n \to \infty$.

5. If $(x, y) \in T_{\epsilon}(X, Y)$, then $x \in T_{\epsilon}(X)$ and $y \in T_{\epsilon}(Y)$. (but the converse is not true necessary.)

6. For any $(x, y) \in T_{\epsilon}(X, Y)$,

$$p(x^n, y^n) = 2^{-n[H(X,Y)\pm o(\epsilon)]}$$

$$p(x^n|y^n) = 2^{-n[H(X|Y)\pm o(\epsilon)]}$$

$$p(y^n|x^n) = 2^{-n[H(Y|X)\pm o(\epsilon)]}$$

7. $|T_{\epsilon}(X, Y)| \simeq 2^{n[H(X,Y)\pm o(\epsilon)]}$.

8. For any $x^n \in T_{\epsilon}(X)$, if $y_i$’s are chosen according to distribution $p(y|x_i)$ for $i = 1, \ldots, n$, then $\Pr[(x^n, y^n) \in T_{\epsilon}(X, Y)] \to 1$ as $n \to \infty$.

9. For any $x^n \in T_{\epsilon}(X)$, $|\{y^n : (x^n, y^n) \in T_{\epsilon}(X, Y)\}| = 2^{n[H(Y|X)\pm o(\epsilon)]}$.

10. if $x^n$ and $y^n$ are chosen according to their marginal distribution $(p(x^n)$ and $p(y^n))$, then $\Pr[(x^n, y^n) \in T_{\epsilon}(X, Y)] = 2^{-n[I(X,Y)\pm o(\epsilon)]}$.

11. If $(x^n, y^n, z^n) \in T_{\epsilon}(X, Y, Z)$, $(x^n, y^n) \in T_{\epsilon}(X, Y)$, $(x^n, z^n) \in T_{\epsilon}(X, Z)$ and $(y^n, z^n) \in T_{\epsilon}(Y, Z)$ but the converse is not true in general.
12. (Markov Lemma) If $X \leftrightarrow Y \leftrightarrow Z$, then

$$(x^n, y^n) \in T_\epsilon(X, Y)$$

and

$$(y^n, z^n) \in T_\epsilon(Y, Z)$$

imply that $(x^n, y^n, z^n) \in T_\epsilon(X, Y, Z)$.

2.3 Entropy Power Inequality

Entropy power inequality (EPI) which in known as Blachman inequality [11] is one of the powerful tools to deal with this kind of problems. This inequality gives a lower bound on the entropy of a sum of independent random variables [2].

**Theorem 2.** If $X$ and $Y$ are independent random $n$-vectors, then

$$e^{\frac{2}{n} h(X + Y)} \geq e^{\frac{2}{n} h(X)} + e^{\frac{2}{n} h(Y)}$$

(2.16)

2.3.1 Conditional Entropy Power Inequality

Here we introduce a conditional version of Blachman inequality which will become handy in next chapters.

**Theorem 3.** Let $W \leftrightarrow X \leftrightarrow X + Y$ be a Markov chain, where $Y$ is independent of $W$. Then

$$\exp \left( \frac{2}{n} h(X + Y \mid W) \right) \geq \exp \left( \frac{2}{n} h(X \mid W) \right) + \exp \left( \frac{2}{n} h(Y) \right)$$

(2.17)

For the proof, note that the pointwise conditional EPI holds for any instance of $W = w$:

$$\exp \left( \frac{2}{n} h(X + Y \mid W = w) \right) \geq \log \left[ \exp \left( \frac{2}{n} h(X \mid W = w) \right) + \exp \left( \frac{2}{n} h(Y) \right) \right]$$
where the last condition to $W = w$ is dropped because $Y$ is independent of $W$. Using the convexity of function $\log(e^x + k)$, we can average both sides over $W$ and preserve the direction of the inequality using Jensen’s inequality.

\[
\left( \frac{2}{n} h(X + Y | W) \right) \geq \log \left[ \exp \left( \frac{2}{n} h(X | W) \right) + \exp \left( \frac{2}{n} h(Y) \right) \right]
\]

Finally, by exponentiating both sides, we will have (2.17)
Chapter 3

History and Previous Results

The problem of “Multiple descriptions” which is also known as optimum diversity problem, was propounded by A. D. Wyner at the IEEE Shannon Theory Workshop, in September 1980. The problem of source coding for two encoders and three decoders, was formulated by L. H. Ozarow. At that workshop El Gamal and Cover deduced a set of conditions sufficient for \((R_1, R_2, D_0, D_1, D_2)\) to be achievable and subsequently published their famous result \([7]\) which we shall refer to as the EGC Theorem.

**Theorem 4 (EGC[7]).** Let \(X_{1, 2, \ldots}\) is a sequence of i.i.d. random variables drawn randomly according to a probability mass function \(p(x)\) from a finite size alphabet \(\mathcal{X}\). Let \(d_i(\cdot, \cdot), i = 0, 1, 2\), be bounded distance functions. An achievable rate region for distortion tuple \(D = (D_1, D_2, D_0)\) is given by the convex hull of all \((R_1, R_2)\) such that

\[
\begin{align*}
R_1 & \geq I(X; \hat{X}_1) \\
R_2 & \geq I(X; \hat{X}_2) \\
R_1 + R_2 & \geq I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_1; \hat{X}_2)
\end{align*}
\]

(3.1)

for some probability mass function

\[
p(x, \hat{x}_0, \hat{x}_1, \hat{y}_2) = p(x)p(\hat{x}_0, \hat{x}_1, \hat{y}_2 \mid x)
\]

(3.2)

such that

\[
\begin{align*}
\mathbb{E}[d_1(X, \hat{X}_1)] & \leq D_1 \\
\mathbb{E}[d_2(X, \hat{X}_2)] & \leq D_2 \\
\mathbb{E}[d_0(X, \hat{X}_0)] & \leq D_0
\end{align*}
\]
We denote the region defined in EGC Theorem by $R_{\text{EGC}}$. The following theorem by Witsenhausen [12] is also introduce an inner bound for the achievable rates of multi description problem.

**Theorem 5 (Witsenhausen [12]).** The quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable if there exist random variables $U$ and $V$ jointly distributed with a generic source random variable $X$, such that

\[
\begin{align*}
R_1 & \geq I(X;U) \\
R_2 & >\geq I(X;V) \\
R_1 + R_2 & \geq I(X;U,V) + I(U;V)
\end{align*}
\]

and there exist random variables of the forms

\[
\begin{align*}
\hat{X}_1 &= g_1(U) \\
\hat{X}_2 &= g_2(V) \\
\hat{X}_0 &= g_0(U,V)
\end{align*}
\]

such that the distortion constraints are satisfied.

By the way, this theorem had been abandoned in favor of the EGC Theorem. The reason of that is not only the characterization of EGC region is simpler, but it can be shown that $R_{\text{EGC}}$ includes the achievable region of this theorem.

In the case of a Gaussian source with $d(x, \hat{x}) = (x - \hat{x})^2$, Ozarow [6] showed that the EGC Theorem is tight; that is, all achievable quintuples satisfy the EGC conditions. He characterized the achievable rates of MD problem for Gaussian source as the following theorem.

**Theorem 6.** Let $x(1), x(2), \ldots$ be a sequence of i.i.d. random variables drawn according to Normal distribution, $X \sim \mathcal{N}(0,1)$. The achievable set of quintuples $(R_1, R_2, D_0, D_1, D_2)$ of MD problem is given by the set of points satisfying

\[
\begin{align*}
D_1 & \geq \exp(-2R_1) \\
D_2 & \geq \exp(-2R_2) \\
D_0 & \geq \exp[-2(R_1 + R_2)]\frac{1}{1-(\sqrt{\Pi}-\sqrt{\Delta})^2}
\end{align*}
\]

where $\Pi = (1 - D_1)(1 - D_2)$ and $\Delta = D_1D_2 - \exp[-2(R_1 + R_2)]$. 
We will see the proof of Ozarow’s theorem in Chapter 6. The natural question arises is whether the EGC Theorem is tight more generally.

Z. Zhang and T. Berger [13] proved the following theorem which is another characterization for the achievable rates.

**Theorem 7 (Zhang and Berger [13]).** Any quintuple \((R_1, R_2, D_0, D_1, D_2)\) is achievable if there exist random variables \(\hat{X}_0, \hat{X}_1, \hat{X}_2\) jointly distributed with a generic source random variable \(X\) such that

\[
\begin{align*}
R_1 &\geq I(X; \hat{X}_1, \hat{X}_0) \\
R_2 &\geq I(X; \hat{X}_2, \hat{X}_0) \\
R_1 + R_2 &\geq 2I(X; \hat{X}_0) + I(X; \hat{X}_1, \hat{X}_2 | \hat{X}_0) + I(\hat{X}_1; \hat{X}_2 | \hat{X}_0)
\end{align*}
\]

(3.3)

and there exist deterministic functions \(\phi_1, \phi_2, \phi_2\) which satisfy

\[
\begin{align*}
\mathbb{E}\{d_1(X, \phi_1(\hat{X}_0, \hat{X}_1))\} &\leq D_1 \\
\mathbb{E}\{d_2(X, \phi_2(\hat{X}_0, \hat{X}_2))\} &\leq D_2 \\
\mathbb{E}\{d_0(X, \phi_0(\hat{X}_0, \hat{X}_1, \hat{X}_2))\} &\leq D_0
\end{align*}
\]

Denoting the region of Zheng and Berger theorem by \(R_{ZB}\), it be can be shown that

\[
R_{EGC} \subset R_{ZB}.
\]

There is also a binary counterexample founded by Zhang and Berger[13] which shows that the EGC Theorem is not tight in general and \(R_{ZB}\) strictly includes \(R_{EGC}\).

There is also an outer bound for the case of binary symmetric sources with Hamming distance function by J. Wolf et al. [8]. The comparison between this bound and the inner bound of EGC Theorem shows that the outer bound not surprisingly contain the region of EGC theorem, but the bound exceeds the achievable points.

Another result for an outer bound of achievable points in the case of binary symmetric symmetric source and Hamming distortion function is derived by Zheng and Berger [13].
Theorem 8. For the equiprobable binary sources and error frequency distortion measure, if \( R(D_0) = R_1 + R_2 \), then all achievable quintuples 
\[(R_1, R_2, D_0, D_1, D_2)\]
satisfy the following conditions:
\[
\left(\frac{1}{2} + D_1 - 2D_0\right) \left(\frac{1}{2} + D_2 - 2D_0\right) \geq (1 - 2D_0)^2
\] (3.4)
\[D_1 \geq D(R_1)\] (3.5)
\[D_2 \geq D(R_2)\] (3.6)
where \( D(R) \) is the distortion-rate function.

Furthermore, Ahlswede[14] showed that EGC region is tight for the case of no excess rate for the joint description in which \( R_1 + R_2 = R(D_0) \). There is also a new paper by Fu and Yeung [15] which shows that EGC region is tight if one of the marginal decoders is required to recover a function of the source perfectly in the usual Shannon sense.

Multiple descriptions with side information is an important class MD problem. Indeed, this problem marries the multiple description source coding problem with the Wyner-Ziv question. There are two theorems which give inner bounds for the MD problem with side information in the recent paper by S. N. Diggavi and V. A. Vaishampayan [16].

Theorem 9 (MD with side information at encoder and decoder). Let \((x(1), s(1)), (x(2), s(2)), \ldots\) be a sequence of i.i.d. finite alphabet random variables drawn according to probability mass function \( Q(x, s) \). Let the encoder and decoders have access to side information \( s(1), s(2), \ldots \) (Figure 3.1, switch on). Then any tuple \((R_1, R_2, D_0, D_1, D_2)\) is achievable if there exist random variables \( \hat{Y}, \hat{X}_0, \hat{X}_1, \) and \( \hat{X}_2 \) jointly distributed with the source random variable \( X \) and the side-information random variable \( S \) such that,
\[
R_1 \geq I(X; \hat{X}_1, \hat{Y} | S)
\]
\[
R_2 \geq I(X; \hat{X}_2, \hat{Y} | S)
\]
\[
R_1 + R_2 \geq 2I(X; \hat{Y} | S) + I(X; \hat{X}_1, \hat{X}_2, \hat{X}_0 | \hat{Y}, S) + I(\hat{X}_1; \hat{X}_2 | \hat{Y}, S)
\] (3.7)
Figure 3.1: Multiple Descriptions problem with side information.

for some probability mass function

\[ p(x, s, \hat{y}, \hat{x}_0, \hat{x}_1, \hat{x}_2) = Q(x, s)p(\hat{y}, \hat{x}_0, \hat{x}_1, \hat{x}_2 \mid x, s) \]

and

\[
\mathbb{E}[d_1(X, \hat{X}_1)] \leq D_1 \\
\mathbb{E}[d_2(X, \hat{X}_2)] \leq D_2 \\
\mathbb{E}[d_0(X, \hat{X}_0)] \leq D_0
\]

The following theorem [16] considers the case in which the side-information is available only at the decoders.

**Theorem 10 (MD with decoder-only side information).** Let

\[(x(1), s(1)), (x(2), s(2)), \ldots\]

be a sequence of i.i.d. finite alphabet random variables drawn according to probability mass function \(Q(x, s)\). Let only the decoders have access to side information \(s(1), s(2), \ldots\) (Figure 3.1, switch off). Then any tuple

\[(R_1, R_2, D_0, D_1, D_2)\]

is achievable if there exist random variables \((Y, W_0, W_1, W_2)\) with probability mass function

\[ p(x, s, y, w_0, w_1, w_2) = Q(x, s)p(y, w_0, w_1, w_2 \mid x, s) \]
such that

\[
R_1 \geq I(X; W_1, Y | S)
\]

\[
R_2 \geq I(X; W_2, Y | S)
\]

\[
R_1 + R_2 \geq 2I(X; Y | S) + I(X; W_1, W_2, W_0 | Y, S) + I(W_1; W_2 | Y, S)
\]

(3.8)

and there exist reconstruction functions \(f_0, f_1, \) and \(f_2\) which satisfy

\[
\mathbb{E}[d_1(X, f_1(S, Y, W_1))] \leq D_1
\]

\[
\mathbb{E}[d_2(X, f_2(S, Y, W_2))] \leq D_2
\]

\[
\mathbb{E}[d_0(X, f_0(S, Y, W_0, W_1, W_2))] \leq D_0.
\]

It has been shown in [21] that in the single description Gaussian case, the decoder-only side-information rate distortion function coincides with that when both encoder and decoder were informed of the side-information. The same result has been derived [16] for the case of Gaussian two-description problem with common decoder side-information if the the distortion measures are mean-squared error function.

Successive refinement source coding [17] is also another related topic to this problem. Informally speaking, we can consider that as the multiple description source coding where there is only one marginal decoders and the second description is just used for refining the estimation in the central decoder.
Chapter 4

Ozarow’s Theorem for Scalar Gaussian Multiple Descriptions

As we stated before, the multiple descriptions problem has been solved only for the one Gaussian source by Ozarow [6]. We will try to apply the same techniques for the case of vector multiple descriptions. Before starting the argument about the vector case, we need to review the main theorem and analyze the proof.

Theorem 11. Let \(s(1), s(2), \ldots\) be a sequence of i.i.d. random variables drawn according to Gaussian distribution, \(X \sim \mathcal{N}(0, P)\). The achievable set of quintuples \((R_1, R_2, D_0, D_1, D_2)\) of MD problem is given by the set of points satisfying

\[
D_1 \geq P \exp(-2R_1)
\]
\[
D_2 \geq P \exp(-2R_2)
\]
\[
D_0 \geq P \exp[-2(R_1 + R_2)] \frac{1}{1-(\sqrt{\Pi - \Delta})^2}
\]

(4.1) (4.2)

where \(\Pi = (1 - \frac{D_1}{P})(1 - \frac{D_2}{P})\) and \(\Delta = \frac{D_1}{P} \frac{D_2}{P} - \exp[-2(R_1 + R_2)]\).

In order to bound the rates required in terms of given distortions, we can interpret the Theorem, as the following. For a sequence of i.i.d. random variables drawn according to Gaussian distribution, \(X \sim \mathcal{N}(0, P)\), the minimum rates required for
satisfying the distortion tuple \((D_0, D_1, D_2)\) are given as

\[
\begin{align*}
R_1 & \geq \frac{1}{2} \log \frac{P}{D_1} \\
R_2 & \geq \frac{1}{2} \log \frac{P}{D_2}
\end{align*}
\] (4.3)

\[
R_1 + R_2 \geq \frac{1}{2} \log \left[ \frac{P(P - D_0)^2}{D_0 [(P + D_0)(D_1 + D_2) - 2(PD_0 + D_1D_2)] + PD_0 \sqrt{\Pi(D_1 - D_0)(D_2 - D_0)}} \right]
\]

As the proof of this theorem is too important to our purpose, and we need to use its techniques later, we state the whole proof.

**Proof.** Using rate-distortion theory, it is easy to see that for Gaussian source and single description decoding, we have

\[
\begin{align*}
D_1 & \geq P \exp(-2R_1) \\
D_2 & \geq P \exp(-2R_2)
\end{align*}
\]

In the following we are going to bound \(D_0\) in terms of \(R_1, R_2, D_1,\) and \(D_2\). We denote the encoding and decoding functions by \((g_1, g_2)\) and \((f_0, f_1, f_2)\), respectively. The mutual information between the source block \(S\) and the reconstructed version of source by the central decoder \(\hat{S}_0\), satisfies the following inequalities:

\[
\begin{align*}
I(S; \hat{S}_0) & \leq I(S; g_1(S), g_2(S)) \\
& \leq H(g_1(S), g_2(S)) \\
& = H(g_1(S)) + H(g_2(S)) - I(g_1(S); g_2(S)) \\
& \overset{(b)}{=} I(S; g_1(S)) + I(S; g_2(S)) - I(g_1(S); g_2(S)) \\
& \overset{(c)}{=} n(R_1 + R_2) - I(g_1(S); g_2(S)) \\
& \overset{(a)}{=} n(R_1 + R_2) - I(\hat{S}_1; \hat{S}_2).
\end{align*}
\] (4.4)

In the above inequalities, those who labeled by \((a)\), follow by the date-processing theorem. Also \((b)\) is true because \(g_1\) and \(g_2\) are deterministic functions of \(S\), and \((c)\) comes from the channel constraints.
by converse of rate-distortion theorem, we have
\[
D_0 \geq D \left( \frac{1}{n} I(S; \hat{S}_0) \right) \\
\overset{(d)}{=} P \exp \left( -\frac{2}{n} I(S; \hat{S}_0) \right) \\
\geq P \exp \left[ -2(R_1 + R_2) \right] \exp \left( \frac{2}{n} I(\hat{S}_1; \hat{S}_2) \right)
\]
where \( (d) \) follows from the fact that we are dealing with Gaussian sources. For the rest of the argument we denote the last exponent in (4.5) by \( t \). In order to bound the mutual information between \( \hat{S}_1 \) and \( \hat{S}_2 \), we define an auxiliary random variable \( \tilde{S} \), formed by adding to \( S \) a zero-mean Gaussian vector \( Z \), whose components are independent and have common variance \( \lambda \). Informally speaking, \( \tilde{S} \) includes the common information of \( \hat{S}_1 \) and \( \hat{S}_2 \). Using the fact
\[
I(\hat{S}_1; \hat{S}_2) = I(\hat{S}_1; \hat{S}_2 \mid \tilde{S}) + I(\hat{S}_1; \tilde{S} \mid \hat{S}_2)
\]
we have
\[
I(\hat{S}_1; \hat{S}_2) = I(\hat{S}_1; \hat{S}_2 \mid \tilde{S}) + I(\hat{S}_1; \tilde{S} \mid \hat{S}_2) \\
\geq I(\hat{S}_1; \tilde{S}) - I(\hat{S}_1; \tilde{S} \mid \hat{S}_2) \\
= I(\hat{S}_1; \tilde{S}) + I(\hat{S}_2; \tilde{S}) - I(\hat{S}_1, \hat{S}_2; \tilde{S})
\]
where the last equality follows from the fact that \( I(A; B, C) = I(A; B) + I(A; C \mid B) \).
In order to bound the first and the second terms of the above expression, for \( i = 1, 2 \), we can write
\[
\frac{1}{n} \sum_{k=0}^{n} \mathbb{E} \left[ (\hat{S}_{ik} - \tilde{S}_k)^2 \right] = \frac{1}{n} \sum_{k=0}^{n} \mathbb{E} \left[ (\hat{S}_{ik} - S_k + S_k - \tilde{S}_k)^2 \right] \\
= \frac{1}{n} \sum_{k=0}^{n} \left[ \mathbb{E}(\hat{S}_{ik} - S_k)^2 + \mathbb{E}(S_k - \tilde{S}_k)^2 \right] \\
= D_i + \lambda
\]
where the cross term vanishes since \( Z_k \) are independent of all else. As \( \tilde{S} \) is a Gaussian vector with independent components, each of variance \( 1 + \lambda \), the rate-distortion function for \( \tilde{Y} \) is given by
\[
\tilde{R}(D) = \frac{1}{2} \log \frac{P + \lambda}{D}
\]
So using the converse of rate-distortion theory, we have

\[
\frac{1}{n} I(\hat{S}_i; \tilde{S}) \geq \frac{1}{2} \log \frac{P + \lambda}{D_i + \lambda} \tag{4.7}
\]

for \(i = 1, 2\). For bounding the last term in (4.6), we will use the conditional entropy power inequality form (2.17).

\[
I(\hat{S}_1, \hat{S}_2; \tilde{S}) = h(\tilde{S}) - h(\tilde{S} | \hat{S}_1, \hat{S}_2)
\]

\[
= \frac{n}{2} \log(2\pi e(k + \lambda)) - h(\tilde{S} | \hat{S}_1, \hat{S}_2)
\]

\[
\leq \frac{n}{2} \log(2\pi e(k + \lambda))
\]

\[
- \frac{n}{2} \log \left[ \exp \left( \frac{2}{n} h(S | \hat{S}_1, \hat{S}_2) \right) + \exp \left( \frac{2}{n} h(Z | \hat{S}_1, \hat{S}_2) \right) \right]
\]

\[
= \frac{n}{2} \log(2\pi e(k + \lambda)) - \frac{n}{2} \log \left[ \exp \left( \frac{2}{n} h(S | \hat{S}_1, \hat{S}_2) \right) + 2\pi e\lambda \right] \tag{4.8}
\]

We also have

\[
h(S | \hat{S}_1, \hat{S}_2) = h(S) - I(S; \hat{S}_1, \hat{S}_2)
\]

\[
= h(S) - H(\hat{S}_1, \hat{S}_2) + H(\hat{S}_1, \hat{S}_2 | S)
\]

\[
\leq h(S) - H(\hat{S}_1, \hat{S}_2)
\]

\[
= h(S) - H(\hat{S}_1) - H(\hat{S}_2) + I(\hat{S}_1, \hat{S}_2)
\]

\[
\geq n \log(2\pi e k) - nR_1 - nR_2 + \frac{n}{2} \log t \tag{4.9}
\]

where \((e)\) follows from the fact that \(\hat{S}_1\) and \(\hat{S}_2\) are deterministic function of \(S\). By substituting (4.9) into (4.8), we have

\[
I(\hat{S}_1, \hat{S}_2; \tilde{S}) \leq \frac{n}{2} \log(2\pi e(P + \lambda)) - \frac{n}{2} \log [(2\pi e Pt) \exp (- 2(R_1 + R_2)) + 2\pi e\lambda]
\]

\[
= \frac{n}{2} \log(P + \lambda) - \frac{n}{2} \log [Pt \exp (- 2(R_1 + R_2)) + \lambda]
\]

Substituting (4.7) and (4.10) in (4.6), and exponentiating both sides, we have

\[
t \geq \frac{Pt \exp (- 2(R_1 + R_2)) + \lambda(k + \lambda)}{(D_1 + \lambda)(D_2 + \lambda)}
\]

Isolating \(t\),

\[
t \geq \frac{\lambda(P + \lambda)}{\lambda^2 + \lambda[D_1 + D_2 - P \exp(-2(R_1 + R_2))] + [D_1 D_1 - P^2 \exp(-2(R_1 + R_2))]^{(4.10)}}
\]
Thus, from (4.5) we have

\[ D_0 \geq P \exp[-2(R_1 + R_2)] \left[ \frac{P t \exp(-2(R_1 + R_2)) + \lambda (P + \lambda)}{(D_1 + \lambda)(D_2 + \lambda)} \right] \]

or

\[ \exp[-2(R_1 + R_2)] \leq \frac{D_0 \lambda^2 + (D_1 + D_2)\lambda + D_1 D_2}{k \lambda^2 + (k + D_0)\lambda + k D_0} \]  

(4.11)

This inequality holds for any \( \lambda \geq 0 \). In particular, we can choose that \( \lambda \) which minimizes the RHS of (4.11). Taking derivative and setting to zero, it is shown that the minimizing \( \lambda \) is given by

\[ \lambda = \frac{D_1 D_2 - PD_0 + \sqrt{(P - D_1)(P - D_2)(D_1 - D_0)(D_2 - D_0)}}{P + D_0 - D_1 - D_2}. \]

Substituting this \( \lambda \) in (4.11), we have

\[ \exp[-2(R_1 + R_2)] \leq \frac{D_0 [(P + D_0)(D_1 + D_2) - 2(P D_0 + D_1 D_2)]}{P (P - D_0)^2} \]

\[ + \frac{2D_0 \sqrt{(P - D_1)(k - D_2)(D_1 - D_0)(D_2 - D_0)}}{P (P - D_0)^2} \]

or

\[ R_1 + R_2 \geq \frac{1}{2} \log \left[ \frac{P (P - D_0)^2}{D_0 [(P + D_0)(D_1 + D_2) - 2(P D_0 + D_1 D_2)] + P D_0 \sqrt{(P - D_1)(D_1 - D_0)(D_2 - D_0)}} \right] \]

which is the same as (4.3). Combining the above inequality with (4.5) and substituting the maximizing \( t \) in (4.10), the third inequality of Theorem 11 can be proved.

Theorem 11 characterized all achievable points of MD problem for the case of Gaussian source with mean-square distance function. In the following we state the Ozarow’s proof for forward part of his characterization which shows that EGC theorem is tight in this particular case.
4.1 Achievability

Let $X$ be a vector of i.i.d. zero mean Gaussian source with variance $P$. Let also $U$ and $V$ are two encoded description of $S$ sent to the decoders. We construct $U$ and $V$ by adding some Gaussian noises to the source sequence.

$$U = S + N_1$$
$$V = S + N_2$$

where $N_1$ and $N_2$ are jointly zero-mean Gaussian noises with covariance matrix

$$\Phi = \begin{pmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{pmatrix}$$

where $\sigma_1$, $\sigma_2$, and $\rho$ will be determined later such that the distortion constraints are satisfied. As $N_1$ and $N_2$ are independent from $X$, it is easy to see that $U$ and $V$ are jointly Gaussian random variables with the covariance matrix

$$Q = \begin{pmatrix} P + \sigma_x^2 & P + \rho \sigma_x \sigma_y \\ P + \rho \sigma_x \sigma_y & P + \sigma_y^2 \end{pmatrix}.$$ 

Clearly, the best decoding functions $f_0(u, v)$, $f_1(u)$, and $f_2(v)$ are the minimum mean-squared error estimates (MMSE) of $s$, given the respective argument. These are given by

$$\hat{s}_1 = f_1(u) = \frac{P}{P + \sigma_1^2} u$$
$$\hat{s}_2 = f_2(v) = \frac{P}{P + \sigma_2^2} v$$
$$\hat{s}_0 = f_0(u, v) = \frac{P(\sigma_2^2 - \rho \sigma_1 \sigma_2)}{\sigma_1^2 \sigma_2^2 (1 - \rho^2) + P(\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2)} u + \frac{P(\sigma_1^2 - \rho \sigma_1 \sigma_2)}{\sigma_1^2 \sigma_2^2 (1 - \rho^2) + P(\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2)} v$$
As $\hat{S}_1$, $\hat{S}_2$, and $\hat{S}_0$ are deterministic functions of $U$ and $V$, we can calculate the terms in EGC theorem as the following.

\[
I(S; \hat{S}_1) = I(S; U) = h(U) - h(V | S) \\
= \frac{1}{2} \log \frac{P + \sigma_1^2}{\sigma_1^2},
\]

\[
I(S; \hat{S}_2) = I(S; V) = h(V) - h(V | S) \\
= \frac{1}{2} \log \frac{P + \sigma_2^2}{\sigma_2^2},
\]

and

\[
I(S; \hat{S}_0, \hat{S}_1, \hat{S}_2) + I(\hat{S}_1, \hat{S}_2) = I(S; U, V) + I(U, V) \\
= h(U, V) - h(U, V | S) + h(U) + h(V) - h(U, V) \\
= \frac{1}{2} \log \frac{(P + \sigma_1^2)(P + \sigma_2^2)}{\sigma_1^2 \sigma_2^2(1 - \rho^2)}. 
\]

So, EGC Theorem is translated to

\[
R_1 \geq \frac{1}{2} \log \frac{P + \sigma_1^2}{\sigma_1^2} \\
R_2 \geq \frac{1}{2} \log \frac{P + \sigma_2^2}{\sigma_2^2} \\
R_1 + R_2 \geq \frac{1}{2} \log \frac{(P + \sigma_1^2)(P + \sigma_2^2)}{\sigma_1^2 \sigma_2^2(1 - \rho^2)} \tag{4.12}
\]

where $\sigma_1$, $\sigma_2$, and $\rho$ should satisfy the distortion constraints

\[
\mathbb{E}\left[ (S - \hat{S}_1)^2 \right] = \frac{P \sigma_1^2}{P + \sigma_1^2} \leq D_1, \tag{4.13}
\]

\[
\mathbb{E}\left[ (S - \hat{S}_2)^2 \right] = \frac{P \sigma_2^2}{P + \sigma_2^2} \leq D_2, \tag{4.14}
\]

and

\[
\mathbb{E}\{(S - \hat{S}_0)^2\} = \frac{P \sigma_1^2 \sigma_2^2(1 - \rho^2)}{\sigma_1^2 \sigma_2^2(1 - \rho^2) + P \sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2} \leq D_0. \tag{4.15}
\]
σ₁ and σ₂ are determined by setting (4.13) and (4.14) to equality, respectively. Combining the distortion constraints and (4.12), we have

\[ R_1 \geq \frac{1}{2} \log \frac{P}{D_1} \]
\[ R_2 \geq \frac{1}{2} \log \frac{P}{D_2} \]
\[ R_1 + R_2 \geq \frac{1}{2} \log \frac{P^2}{D_1 D_2 (1 - \rho^2)} \]

We can therefore choose ρ arbitrary, so long as

\[ D_1 D_2 (1 - \rho^2) \geq P^2 \exp[-2(R_1 + R_2)] \]

or

\[ \rho^2 \leq \frac{D_1 D_2 - P^2 \exp[-2(R_1 + R_2)]}{D_1 D_2} \]

Choose

\[ \rho = -\sqrt{\frac{D_1 D_2 - P^2 \exp[-2(R_1 + R_2)]}{D_1 D_2}} \]

Substituting the values of σ₁, σ₂, and ρ in (4.15), we have

\[ D_0 = \frac{P \exp[-2(R_1 + R_2)]}{1 - (\sqrt{\Pi} - \sqrt{\Delta})^2} \]

which is the same result as we have seen in (4.1).

### 4.2 Analysis of the Proof

There are two important inequalities in the converse part proof of Ozarow’s Theorem which should appear as equality in order to the outer bound being tight. The first one is the conditional entropy power inequality , (2.17). Since S and Z are independent Gaussian random variables, the equality not surprisingly holds.
The second inequality which should be tight is

\[ I(\hat{S}_1; \hat{S}_2 \mid \tilde{S}) \geq 0. \]

The RHS equals zero iff \( \hat{S}_1 \) and \( \hat{S}_2 \) are independent conditioned on \( \tilde{S} \). The other expression for that is \( \tilde{S} \) includes all common information of \( \hat{S}_1 \) and \( \hat{S}_2 \), or

\[ \hat{S}_1 \leftrightarrow \tilde{S} \leftrightarrow \hat{S}_2. \quad (4.17) \]

In the following we will characterize the role of \( \tilde{S} \) and its relation with \( \hat{S}_1 \) and \( \hat{S}_2 \). Since we are using Gaussian encoding, the best estimation of \( \hat{S}_1 \) and \( \hat{S}_2 \) in terms of \( \tilde{S} \) is linear estimation

\[ \begin{align*}
\hat{S}_1 &= \alpha \tilde{S} + \tilde{N}_1, \\
\hat{S}_2 &= \beta \tilde{S} + \tilde{N}_2,
\end{align*} \quad (4.18) \]

where the Markov chain property in (4.17), guarantees that \( \tilde{N}_1 \) and \( \tilde{N}_2 \) are zero-mean Gaussian independent noises with variances \( \tilde{\sigma}_1^2 \) and \( \tilde{\sigma}_2^2 \), respectively. They are also independent of \( \tilde{S} \). We have

\[ \mathbb{E}[\hat{S}_1 \mid \tilde{S}] = \mathbb{E}[\hat{S}_1 \tilde{S}] \mathbb{E}[\tilde{S}^2]^{-1} \tilde{S} \]

So,

\[ \alpha \tilde{S}_1 = \frac{P^2}{P + \sigma_1^2} \frac{1}{P + \lambda} \tilde{S}_1. \]

Therefore

\[ \alpha = \frac{P^2}{(P + \sigma_1^2)(P + \lambda)} \]

and similarly we have

\[ \beta = \frac{P^2}{(P + \sigma_2^2)(P + \lambda)}. \]
By calculating $\mathbb{E}[\hat{S}_1^2]$ and $\mathbb{E}[\hat{S}_2^2]$, we can also show

$$\hat{\sigma}^2_1 = \frac{P^2[(P + \sigma^2_1)(P + \lambda) - P^2]}{(P + \sigma^2_2)^2(P + \lambda)}$$

Finally, by comparing the above results, and calculating $\mathbb{E}[\hat{S}_1\hat{S}_2]$ we can show that the optimal $\lambda$ in the converse proof satisfies

$$(P + \lambda)(P + \rho\sigma_1\sigma_2) = P^2. \quad (4.19)$$

We will refer to this results later during the proof of theorem for vector multiple descriptions problem.
Chapter 5

Achievable Rates for Vector Multiple Descriptions

As we stated before in Chapter 3, El Gamal and Cover[4] have characterized an inner bound for the achievable rates for the scalar multiple descriptions problem. This result determines some lower bounds for each of the marginal rates as well as the sum rate in terms of the mutual information between the original source and each of the estimations in the decoders. In this chapter we will extend this result for the vector case and find an inner bound for the achievable region in the rates plane.

5.1 Problem Definition and the Main Theorem

In following we will consider the problem with two correlated sources rather than one.

Let $S = [X, Y]^t$, be a vector of zero-mean Gaussian random variables where $X$ and $Y$ are zero-mean random variables with covariance matrix

$$K = \begin{pmatrix} \sigma_x^2 & \mu \sigma_x \sigma_y \\ \mu \sigma_x \sigma_y & \sigma_y^2 \end{pmatrix}.$$ 

and $s(1), s(2), \ldots, s(n)$ be a sequence of i.i.d. source symbols drawn from the alphabet
Figure 5.1: Multiple Descriptions problem for two Source.

set $(\mathcal{X} \times \mathcal{Y})^n$. Given this sequence of the source, the encoders 1 and 2, produce two description $U$ and $V$ of rates $R_1$ and $R_2$, respectively.

$$
U = g_1(S) \\
V = g_2(S)
$$

At each of the marginal decoders, given any of the descriptions, we need to estimate $S$. The central decoder has access to both of the descriptions and has to estimate $S$ using them. Figure 5.1 shows a schematic structure of the problem. Denoting these estimations by $\hat{S}_1, \hat{S}_2$, and $\hat{S}_0$, we have

$$
\hat{S}_1 = f_1(U) \\
\hat{S}_2 = f_2(V) \\
\hat{S}_0 = f_0(U, V)
$$

where the estimations are in some reproducing alphabets, $\hat{\mathcal{X}}_1 \times \hat{\mathcal{Y}}_1$, $\hat{\mathcal{X}}_2 \times \hat{\mathcal{Y}}_2$, and $\hat{\mathcal{X}}_0 \times \hat{\mathcal{Y}}_0$ which may not coincide with each other or with the source alphabet in general. The coordinates of any reconstructed sequences have to satisfy in the distortion constraints as the following
The following theorem is an extension of the EGC theorem for the correlated sources.

Theorem 12. Let $S = [X, Y]^\dagger$ denotes the pair of random variables $X$ and $Y$ and $S_1, S_2, \ldots$ be a sequence of i.i.d. random variables drawn randomly according to a probability mass function $p(x, y)$ from a finite size alphabet $\mathcal{X} \times \mathcal{Y}$. Let $d_{i,x}$ and $d_{i,y}$, $i = 0, 1, 2$, are bounded distortion functions for $X$ and $Y$, respectively. An achievable rate region for distortion tuple $D = (D_{1,x}, D_{1,y}, D_{2,x}, D_{2,y}, D_{0,x}, D_{0,y})$ is given by the convex hull of all $(R_1, R_2)$ such that

\[
R_1 \geq I(X, Y; \hat{X}_1, \hat{Y}_1)
\]
\[
R_2 \geq I(X, Y; \hat{X}_2, \hat{Y}_2)
\]
\[
R_1 + R_2 \geq I(X, Y; \hat{X}_0, \hat{Y}_0, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2) + I(\hat{X}_1, \hat{Y}_1; \hat{X}_2, \hat{Y}_2)
\]  

(5.2)

for some probability mass function

\[
p(x, y, \hat{x}_0, \hat{y}_0, \hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2) = p(x, y)p(\hat{x}_0, \hat{y}_0, \hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2 \mid x, y)
\]  

(5.3)

such that the distortion constraints in (5.14) are satisfied.

5.2 Proof of the vector EGC Theorem

Actually, the stated theorem is a straight extension of the EGC theorem. Although may be not easy to show the convexity the region characterized by this theorem,
but convexity can be proved by time sharing arguments. The interpretation of the theorem is each of the marginal decoders needs a minimum amount of information about the source in order to find an estimation of it not worse than some given distortion. But for the central decoder, it is not true to say that having a sum rate equal to some similar expression in terms of the mutual information between \((X, Y)\) and \((\hat{X}_0, \hat{Y}_0)\) suffices to reconstruct \(\hat{X}_0\) and \(\hat{Y}_0\) with the given distortion. The reason is the information coded by rates \(R_1\) and \(R_2\) have some overlap in order to make the marginal decoders able to estimate the source. This common information cannot help the central decoder and should be .... by some additional new information with the same rate, \(I(\hat{X}_1, \hat{Y}_1; \hat{X}_2, \hat{Y}_2)\). In the next section we will explain a method of encoding and decoding using the strong typicality and proof the theorem by analyzing the corresponding error.

### 5.2.1 Random Coding

Fix a joint probability mass function of the form

\[
p(x, y, \hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \hat{x}_0, \hat{y}_0) = p(x, y)p(\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \hat{x}_0, \hat{y}_0 | x, y)
\]

on \(X \times Y \times \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{X}_0, \hat{Y}_0\) such that the constraints in (5.14) are satisfied, i. e.,

\[
\mathbb{E}\{d_{i,x}(X, \hat{X}_i)\} < D_{i,x}
\]
\[
\mathbb{E}\{d_{i,y}(Y, \hat{Y}_i)\} < D_{i,y}
\]

for \(i = 0, 1, 2\). Choose arbitrary real numbers \(R'_1, R'_2, \Delta \geq 0\). Let

\[
(\hat{X}_1(1), \hat{Y}_1(1)), (\hat{X}_1(2), \hat{Y}_1(2)), \ldots, (\hat{X}_1(2^{nR'_1}), \hat{Y}_1(2^{nR'_1}))
\]
are pairs of jointly typical \((\hat{X}_1, \hat{Y}_1)\) drawn independently according to uniform distribution over the set \(T_\epsilon(\hat{X}_1, \hat{Y}_1)\). Similarly,

\[(\hat{X}_2(1), \hat{Y}_2(1)), (\hat{X}_2(2), \hat{Y}_2(2)), \ldots, (\hat{X}_2(2^{nR_2'}), \hat{Y}_2(2^{nR_2'}))\]

are pairs of jointly typical sequences drawn independently and uniformly from \(T_\epsilon(\hat{X}_2, \hat{Y}_2)\).

For every jointly typical 4-tuple of \((\hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j))\), let

\[(\hat{X}_0(1), \hat{Y}_0(1)), (\hat{X}_0(2), \hat{Y}_0(2)), \ldots, (\hat{X}_0(2^n\Delta), \hat{Y}_0(2^n\Delta))\]

be drawn i.i.d. according to uniform distribution over the set

\[T_\epsilon(\hat{X}_0, \hat{Y}_0 | \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j))\]

of conditionally \(\epsilon\)-typical \((\hat{x}_0, \hat{y}_0)\)’s, conditioned on \((\hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j))\).

### 5.2.2 Encoding

For a given sequence of the source, \((x^n, y^n) \in \mathcal{A}^n \times \mathcal{B}^n\), find a tuple \((i, j, k)\) such that

\[(x^n, y^n, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j), \hat{x}_{12}(k), \hat{y}_{12}(k))\]

is typical. If no such \((i, j, k)\) exists, simply set \((i, j, k) = (0, 0, 0)\). We can simply divide \(k\) into two parts \((k_1, k_2)\), \(k_1 \in \{1, 2, \ldots, 2^{n\Delta_1}\}\), \(k_2 \in \{1, 2, \ldots, 2^{n\Delta_2}\}\), where \(\Delta_1, \Delta_2 \geq 0\) and \(\Delta_1 + \Delta_2 = \Delta\). Now we send \((i, k_1)\) and \((j, k_2)\) to the marginal decoders 1 and 2, respectively. Therefore, the resulting rates of transmission becomes

\[R_1 = R_1' + \Delta_1\]
\[R_2 = R_2' + \Delta_2\]

\[(5.4)\]
We shall prove that the above encoding will result in a successful reconstruction of \((X, Y)\) with high probability for large enough \(n\) and if

\[
R'_1 > I(X, Y; \hat{X}_1, \hat{Y}_1),
\]

\[
R'_2 > I(X, Y; \hat{X}_2, \hat{Y}_2),
\]

\[
R'_1 + R'_2 - I(\hat{X}_1, \hat{Y}_1; \hat{X}_2, \hat{Y}_2) > I(X, Y; \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2),
\]

\[
\Delta_1 \geq 0,
\]

\[
\Delta_2 \geq 0,
\]

\[
\Delta_1 + \Delta_2 > I(X, Y; \hat{X}_0, \hat{Y}_0 | \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2)
\]

5.2.3 Reconstruction

Decoders 1 and 2 respectively receive \((i, k_1)\) and \((j, k_2)\) where \(0 \leq i \leq 2^{nR'_1}, \ 0 \leq k_1 \leq 2^{n\Delta_1}, \ 0 \leq j \leq 2^{nR'_2}, \ \text{and} \ 0 \leq k_2 \leq 2^{n\Delta_2}\). Given \((i, k_1)\), decoder 1, announces \((\hat{x}_1(i), \hat{Y}_1(i))\) as its estimation of the sources sequence. Similarly, having \((j, k_2)\) announces \((\hat{x}_2(j), \hat{y}_2(j))\) as its reconstruction of \((x^n, y^n)\) and decoder 0, announces \((\hat{x}_0(k), \hat{y}_0(k))\) as the estimation of the sources, where \(k = (k_1, k_2)\) obtains by combining both of the description.

If there exist \((i, j, k)\) in the encoding process such that

\[
(X, Y, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{X}_0, \hat{Y}_0)
\]

are jointly typical, the joint distribution of the above random variables guarantees that distortion constraints are satisfied. Nonetheless, we will have an error in the encoding process and therefore, decoders may be not able to reconstruct the sources sequence. In the next subsection we will show that the probability of such event is small and will be negligible as \(n\) goes to zero.
5.2.4 Analysis of Error

As we stated in the last subsection, the only possibility of occurring an error, is that there does not exist a tuple \((i, j, k)\) such that

\[
(X, Y, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{X}_0, \hat{Y}_0) \in T_\epsilon.
\]

More precisely speaking, an error will occur if one or more of the following events occurs.

\[
E_0 : (x, y) \notin T_\epsilon, \\
E_1 : (x, y, \hat{x}_1(i), \hat{y}_1(i)) \notin T_\epsilon, \quad \forall i \in \{1, 2, \ldots, 2^{nR_1'}\}, \\
E_2 : (x, y, \hat{x}_2(j), \hat{y}_2(j)) \notin T_\epsilon, \quad \forall j \in \{1, 2, \ldots, 2^{nR_2'}\}, \\
E_3 : (x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j)) \notin T_\epsilon, \\
\forall (i, j) \in \{1, 2, \ldots, 2^{nR_1'}\} \times \{1, 2, \ldots, 2^{nR_2'}\}, \\
E_4 : \exists (i, j) \in \{1, 2, \ldots, 2^{nR_1'}\} \times \{1, 2, \ldots, 2^{nR_2'}\}, \\
such that \((x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j)) \in T_\epsilon, \\
but (x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j), \hat{x}_0(k), \hat{y}_0(k)) \notin T_\epsilon \quad \forall k \in \{1, 2, \ldots, 2^\Delta\}. 
\]

We can use union bound to obtain an upper bound for the error probability.

\[
Pr(\text{Error}) = Pr(E_0) + Pr \left( \bigcup_{i=1}^{4} E_0^c \right) \leq Pr(E_0) + \sum_{i=1}^{4} Pr(E_i \cap E_0^c) \quad (5.5)
\]

Using the properties of typicality, it is easy to see that the probability of \(E_0\) goes to 0, as \(n \to \infty\). These properties also say that for a fix \(i\), we have

\[
Pr(x, y, \hat{x}_1(i), \hat{y}_1(i)) \notin T_\epsilon) = 2^{-n[I(X;Y;\hat{X}_1, \hat{Y}_1)\pm o(\epsilon)]}
\]
Thus,

\[
Pr(E_1 \cap E_0^c) = \left(1 - 2^{-n[I(X,Y;\hat{X}_1,\hat{Y}_1)]}\right)^{2^{nR'_1}} \simeq \exp\left[-2^{n(R'_1 - I(X,Y;\hat{X}_1,\hat{Y}_1))}\right]
\]  

(5.6)

So, \( R'_1 > I(X,Y;\hat{X}_1,\hat{Y}_1) \) implies \( Pr(E_1 \cap E_0^c) \to 0 \) by increasing of \( n \). Using similar argument, we can show that if \( R'_2 > I(X,Y;\hat{X}_2,\hat{Y}_2) \), then \( Pr(E_2 \cap E_0^c) \to 0 \).

We will use the below argument to calculate the probability of the fourth event which is similar to the argument used in [7]. For each \((x, y) \in T_\epsilon\), we define the following set

\[ C_{x,y} = \left\{ (i, j) \in \{1, 2, \ldots, 2^{nR'_1}\} \times \{1, 2, \ldots, 2^{nR'_2}\} : \right. \\
(x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j)) \in T_\epsilon(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2 | x, y) \left. \right\} \]

As \((x, y)\) is drawn randomly, \( \| C_{x,y} \| \) is a non-negative random variable. By definition, we have

\[
Pr(E_3 \cap E_0^c) = Pr(\| C_{x,y} \| = 0, (x, y) \in T_\epsilon) \\
\leq \max_{(x,y) \in T_\epsilon} Pr(\| C_{x,y} \| = 0)
\]

For all \((x, y) \in T_\epsilon\), and \( 0 < \alpha < 1 \), we have

\[
Pr(\| C_{x,y} \| = 0) \leq \Pr\left\{ \| C_{x,y} \| - \mathbb{E}\| C_{x,y} \| \geq \alpha \mathbb{E}\| C_{x,y} \| \right\} \\
\leq \frac{\text{Var}\| C_{x,y} \|}{\alpha^2 (\mathbb{E}\| C_{x,y} \|)^2}
\]

where we have used Chebyshev’s inequality in the last inequality. We will use the idea of [20] for calculating the expected value and the variance of size of \( \| C_{x,y} \| \). By definition of \( \| C_{x,y} \| \), we have

\[
\| C_{x,y} \| = \sum_{i \in I, j \in J} 1[(i, j) \in C_{x,y}]
\]
where $\mathbb{1}(\cdot)$ is the indicator function defined as

$$
\mathbb{1}(x) = \begin{cases} 
1, & \text{if } x \text{ is true} \\
0, & \text{otherwise}
\end{cases}
$$

and $I = \{1, 2, \ldots, 2^n R'_1\}$ and $J = \{1, 2, \ldots, 2^n R'_2\}$. But since

$$
\mathbb{E}\{\mathbb{1}((i, j) \in C_{x,y})\} \geq \frac{2^n[H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2|X,Y) - o(1)]}{2^n H(\hat{X}_1, \hat{Y}_1) 2^n H(\hat{X}_2, \hat{Y}_2)} = 2^{-n[H(\hat{X}_1, \hat{Y}_1) + H(\hat{X}_2, \hat{Y}_2) - H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2|X,Y)]},
$$

we have

$$
\mathbb{E}\{\| C_{x,y} \|\} \geq 2^{n[R'_1 + R'_2 - H(\hat{X}_1, \hat{Y}_1) - H(\hat{X}_2, \hat{Y}_2) + H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2|X,Y) - o(1)]} \quad (5.7)
$$

On the other hand, we can calculate the variance of $\| C_{x,y} \|$ as

$$
\| C_{x,y} \|^2 = \left( \sum_{i \in I, j \in J} \mathbb{1}[(i, j) \in C_{x,y}] \right)^2
= \sum_{i \in I, j \in J} \sum_{i' \in I, j' \in J} \mathbb{1}[(i, j) \in C_{x,y}, (i', j') \in C_{x,y}]
= \sum_{i \in I, j \in J} \mathbb{1}[(i, j) \in C_{x,y}]
+ \sum_{i \notin I, j \notin J} \mathbb{1}[(i, j) \in C_{x,y}, (i', j') \in C_{x,y}]
+ \sum_{i \in I, j \notin J} \mathbb{1}[(i, j) \in C_{x,y}, (i', j') \in C_{x,y}]
+ \sum_{i \notin I, j \in J} \mathbb{1}[(i, j) \in C_{x,y}, (i', j') \in C_{x,y}].
$$
Taking expectation, we obtain

\[
E[\|C_{x,y}\|^2] = 2^{nR'_1}2^{nR'_2} \Pr[(i, j) \in C_{x,y}]
\]

\[
+ (2^{nR'_1} - 2^{nR'_1})2^{nR'_2} \Pr[(i, j) \in C_{x,y}, (i', j) \in C_{x,y}]
\]

\[
+ 2^{nR'_1}(2^{nR'_2} - 2^{nR'_2}) \Pr[(i, j) \in C_{x,y}, (i', j) \in C_{x,y}]
\]

\[
+ (2^{nR'_1} - 2^{nR'_1})(2^{nR'_2} - 2^{nR'_2}) \Pr[(i, j) \in C_{x,y}, (i', j') \in C_{x,y}]
\]

(5.8)

Using properties of strong typicality, it is easily seen that

- For arbitrary \(i\) and \(j\),

\[
\Pr[(i, j) \in C_{x,y}] \leq 2^{-n[H(\hat{X}_1, \hat{Y}_1) + H(\hat{X}_2, \hat{Y}_2) - H(\hat{X}_1, \hat{X}_2, \hat{Y}_1, \hat{Y}_2 | X, Y) - o(1)]}
\]

- For \(i \neq i'\) and arbitrary \(j\),

\[
\Pr[(i, j) \in C_{x,y}, (i', j) \in C_{x,y}] \leq 2^{-2n[H(\hat{X}_1, \hat{Y}_1) + H(\hat{X}_2, \hat{Y}_2) - H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2 | X, Y) - o(1)]}
\]

- For arbitrary \(i\), and \(j \neq j'\)

\[
\Pr[(i, j) \in C_{x,y}, (i', j) \in C_{x,y}] \leq 2^{-2n[H(\hat{X}_1, \hat{Y}_1) + H(\hat{X}_2, \hat{Y}_2) - H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2 | X, Y) - o(1)]}
\]

- For \(i \neq i'\) and \(j \neq j'\),

\[
\Pr[(i, j) \in C_{x,y}, (i', j') \in C_{x,y}] \leq 2^{-2n[H(\hat{X}_1, \hat{Y}_1) + H(\hat{X}_2, \hat{Y}_2) - H(\hat{X}_1, \hat{X}_2, \hat{Y}_1, \hat{Y}_2 | X, Y) - o(1)]}
\]
Substituting the above upper bounds in (5.8), results in an upper bound for variance of $\| C_{x,y} \|$ as

$$\text{Var} \| C_{x,y} \| \leq 2^{n(R'_1 + R'_2 - H(\hat{X}_1, \hat{Y}_1) - H(\hat{X}_2, \hat{Y}_2) + H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2 | X, Y) + o(1))}. \quad (5.9)$$

Therefore,

$$\Pr(E_3 \cap E_6^c) \leq \frac{1}{\alpha^2} 2^{-n(R'_1 + R'_2 - H(\hat{X}_1, \hat{Y}_1) - H(\hat{X}_2, \hat{Y}_2) + H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2 | X, Y) + o(1))}$$

which the RHS tends to 0 if

$$R'_1 + R'_2 > H(\hat{X}_1, \hat{Y}_1) + H(\hat{X}_2, \hat{Y}_2) - H(\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2 | X, Y)$$

$$= I(X, Y; \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2) + I(\hat{X}_1, \hat{Y}_1; \hat{X}_2, \hat{Y}_2). \quad (5.10)$$

Now we just have to bound the probability of last error event. For each $(i, j)$, the probability of $(x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j), \hat{x}_0(k), \hat{y}_0(k))$ being jointly typical, conditioned on $(x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j))$ are jointly typical, is given by

$$\Pr\left[ (x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j), \hat{x}_0(k), \hat{y}_0(k)) \in T_{\epsilon} \bigg| (x, y, \hat{x}_1(i), \hat{y}_1(i), \hat{x}_2(j), \hat{y}_2(j)) \in T_{\epsilon} \right]$$

$$= \frac{2^{nH(X,Y;\hat{X}_1,\hat{Y}_1,\hat{X}_2,\hat{Y}_2,\hat{X}_0,\hat{Y}_0 - o(1))}}{2^{nH(X,Y;\hat{X}_0,\hat{Y}_0|\hat{X}_1,\hat{Y}_1,\hat{X}_2,\hat{Y}_2)}}$$

$$= 2^{-nI(X,Y;\hat{X}_0,\hat{Y}_0|\hat{X}_1,\hat{Y}_1,\hat{X}_2,\hat{Y}_2) - o(1)} \quad (5.11)$$

Thus

$$\Pr(E_4 \cap E_6^c) = (1 - 2^{-nI(X,Y;\hat{X}_0,\hat{Y}_0|\hat{X}_1,\hat{Y}_1,\hat{X}_2,\hat{Y}_2)}) 2^{n\Delta}$$

$$\simeq \exp[-2^n(\Delta - I(X,Y;\hat{X}_0,\hat{Y}_0|\hat{X}_1,\hat{Y}_1,\hat{X}_2,\hat{Y}_2))] \quad (5.12)$$

which goes to zero by increasing of $n$, if

$$\Delta > I(X,Y;\hat{X}_0,\hat{Y}_0|\hat{X}_1,\hat{Y}_1,\hat{X}_2,\hat{Y}_2) \quad (5.13)$$
Therefore, the necessary and sufficient conditions for successful reconstruction, can be stated as

\[ R_1 = R'_1 + \Delta_1 > I(X, Y; \hat{X}_1, \hat{Y}_1), \]
\[ R_2 = R'_2 + \Delta_2 > I(X, Y; \hat{X}_2, \hat{Y}_2), \]
\[ R_1 + R_2 = R'_1 + R'_2 + \Delta \]
\[ > I(X, Y; \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2) + I(\hat{X}_1, \hat{Y}_1; \hat{X}_2, \hat{Y}_2) \]
\[ + I(X, Y; \hat{X}_0, \hat{Y}_0 | \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2) \]
\[ = I(X, Y; \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2, \hat{X}_0, \hat{Y}_0) + I(\hat{X}_1, \hat{Y}_1; \hat{X}_2, \hat{Y}_2) \]

5.3 Extension of Zhang-Berger Theorem for Vector Multiple Descriptions

We stated the theorem by Zhang and Berger which introduces an inner bound for the achievable rates of multiple descriptions problem in chapter 3. This Theorem is also extendable for the case of vector multiple descriptions.

Theorem 13. Let \( S_1, S_2, \ldots \) be a sequence of i.i.d. random variables drawn randomly according to a probability mass function \( p(x, y) \) from a finite size alphabet \( \mathcal{X} \times \mathcal{Y} \) where \( S = [X, Y]^{\dagger} \) denotes the pair of random variables \( X \) and \( Y \). Let \( d_{i,x} \) and \( d_{i,y}, i = 0, 1, 2, \) are bounded distortion functions for \( X \) and \( Y \), respectively. A tuple \( (R_1, R_2, D_{1,x}, D_{1,y}, D_{2,x}, D_{2,y}, D_{0,x}, D_{0,y}) \) is achievable if there exist random variables \( \hat{X}_0, \hat{Y}_0, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \) and \( \hat{Y}_2 \) jointly distributed with the source such that

\[ R_1 \geq I(X, Y; \hat{X}_0, \hat{Y}_0, \hat{X}_1, \hat{Y}_1) \]
\[ R_2 \geq I(X, Y; \hat{X}_0, \hat{Y}_0, \hat{X}_2, \hat{Y}_2) \]
\[ R_1 + R_2 \geq 2I(X, Y; \hat{X}_0, \hat{Y}_0) + I(\hat{X}_1, \hat{Y}_1; \hat{X}_2, \hat{Y}_2 | \hat{X}_0, \hat{Y}_0) \]
\[ + I(X, Y; \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2 | \hat{X}_0, \hat{Y}_0) \]
and there exist functions $\phi_{0,x}$, $\phi_{0,y}$, $\phi_{1,x}$, $\phi_{1,y}$, $\phi_{2,x}$, and $\phi_{2,y}$, which satisfy

\begin{align*}
E\left[ d_{1,x}(X, \phi_{1,x}(\hat{X}_0, \hat{Y}_0, \hat{X}_1, \hat{Y}_1)) \right] &\leq D_{1,x} \\
E\left[ d_{1,y}(Y, \phi_{1,y}(\hat{X}_0, \hat{Y}_0, \hat{X}_1, \hat{Y}_1)) \right] &\leq D_{1,y} \\
E\left[ d_{2,x}(X, \phi_{2,x}(\hat{X}_0, \hat{Y}_0, \hat{X}_2, \hat{Y}_2)) \right] &\leq D_{2,x} \\
E\left[ d_{2,y}(Y, \phi_{2,y}(\hat{X}_0, \hat{Y}_0, \hat{X}_2, \hat{Y}_2)) \right] &\leq D_{2,y} \\
E\left[ d_{0,x}(X, \phi_{0,x}(\hat{X}_0, \hat{Y}_0, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2)) \right] &\leq D_{0,x} \\
E\left[ d_{0,y}(Y, \phi_{0,y}(\hat{X}_0, \hat{Y}_0, \hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2)) \right] &\leq D_{0,y}
\end{align*}

(5.14)

Here we omit the proof for brevity. The whole idea of the proof is similar to the original theorem for the scalar case.
Chapter 6

Vector Multiple Descriptions for Jointly Gaussian Sources

In this chapter we are going to characterize the achievable rates for the MD problem for jointly Gaussian sources.

6.1 The Outer Bound

The upper bound on each of $R_1$ and $R_2$ are obtained by the mutual information required for reconstruction of $\hat{S}_1$ and $\hat{S}_2$. In order to bounding the sum rate, $R_1 + R_2$, we need to bound the mutual information between $S$ and $\hat{S}_0$.

\[
I(S; \hat{S}_3) \leq I(S; f_1(S), f_2(S)) \\
\leq h(f_1(S), f_2(S)) \\
= h(f_1(S)) + h(f_2(S)) - I(f_1(S); f_2(S)) \\
\leq n(R_1 + R_2) - I(f_1(S); f_2(S)) \\
\leq n(R_1 + R_2) - I(\hat{S}_1; \hat{S}_2) \tag{6.1}
\]
where the steps are the same as in scalar case proof. Using the same techniques used in Ozarow’s proof, we define an auxiliary random variable \( \tilde{S} \) to bounding the last of in (6.1) which is formed by adding to \( S \) a zero-mean Gaussian vector \( Z \) with some covariance matrix \( \mathbf{\Lambda}, Z \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}) \). We have

\[
I(\hat{S}_1; \hat{S}_2) = I(\hat{S}_1; \hat{S}_2 | \tilde{S}) + I(\hat{S}_1; \tilde{S}) - I(\hat{S}_1; \hat{S} | \hat{S}_2)
\]

\[
\geq I(\hat{S}_1; \tilde{S}) - I(\hat{S}_1; \hat{S} | \hat{S}_2)
\]

\[
= I(\hat{S}_1; \tilde{S}) + I(\hat{S}_1; \hat{S}) - I(\hat{S}_1, \hat{S}_2; \hat{S})
\]

(6.2)

where the inequality is tight if and only if \( S_1 \) and \( S_2 \) be independent conditioned on \( \tilde{S} \). On the other hand we know that

\[
\mathbb{E}[(\hat{S}_1 - \tilde{S})(\hat{S}_1 - \tilde{S})^\dagger] = \mathbb{E}[(\hat{S}_1 - S + S - \tilde{S})(\hat{S}_1 - S + S - \tilde{S})^\dagger]
\]

\[
= \mathbb{E}[(\hat{S}_1 - \tilde{S})(\hat{S}_1 - S)^\dagger] + \mathbb{E}[(S - \tilde{S})(S - \tilde{S})^\dagger]
\]

\[
= \mathbb{E}[(\hat{S}_1 - S)(\hat{S}_1 - S)^\dagger] + \mathbb{E}[ZZ^\dagger]
\]

\[
= D_1 + \mathbf{\Lambda}
\]

\[
:= D_1'
\]

Similarly,

\[
\mathbb{E}[(\hat{S}_2 - \tilde{S})(\hat{S}_2 - \tilde{S})^\dagger] = D_2 + \mathbf{\Lambda}
\]

\[
:= D_2'
\]

Let \( \tilde{R}(D_1') \) be the minimum rate required in order to reconstructing a noisy version of the source, \( \tilde{S} \), as \( \hat{S}_1 \) such that the matrix distortion constraint \( D_1' \) is satisfied. In the same way we can define \( \tilde{R}(D_2') \) for reconstructing \( \tilde{S} \) as \( \hat{S}_2 \) and matrix distortion constraint \( D_2' \). So,

\[
\frac{1}{n} I(\hat{S}_1; \tilde{S}) \geq \tilde{R}_y(D_1') = \frac{1}{2} \log \frac{\det(\mathbf{K} + \mathbf{\Lambda})}{D_1'}
\]
and
\[
\frac{1}{n} I(\hat{S}_2; \bar{S}) \geq \bar{R}(D'_2) = \frac{1}{2} \log \frac{\det(K + \Lambda)}{D'_2}
\]

We have also
\[
I(\hat{S}_1, \hat{S}_2; \bar{S}) = h(\bar{S}) - h(\bar{S}|\hat{S}_1, \hat{S}_2)
\]
\[
= \frac{n}{2} \log \det(2\pi e(K + \Lambda)) - h(\bar{S}|\hat{S}_1, \hat{S}_2)
\]

We can use vector entropy power inequality to come up with a lower bound for the second term of the above expression.

\[
\exp \left( \frac{2}{2n} h(\bar{S}|\hat{S}_1, \hat{S}_2) \right) \geq \exp \left( \frac{2}{2n} h(S|\hat{S}_1, \hat{S}_2) \right) + \exp \left( \frac{2}{2n} h(Z) \right)
\]

where
\[
h(S|\hat{S}_1, \hat{S}_2) = h(S) - I(S; \hat{S}_1, \hat{S}_2)
\]
\[
= h(S) - H(\hat{S}_1, \hat{S}_2) + H(\hat{S}_1, \hat{S}_2 | S)
\]
\[(a)\]
\[
h(S) - H(\hat{S}_1, \hat{S}_2)
\]
\[
= h(S) - H(\hat{S}_1) - H(\hat{S}_2) + I(\hat{S}_1; \hat{S}_2)
\]
\[
\geq \frac{n}{2} \log \det(2\pi eK) - nR_1 + nR_2 + I(\hat{S}_1; \hat{S}_2)
\]

where (a) follows from the fact that \(\hat{S}_1\) and \(\hat{S}_2\) are deterministic functions of \(S\).

Therefore,
\[
\exp \left( \frac{2}{2n} h(\bar{S}|\hat{S}_1, \hat{S}_2) \right) \geq \exp \left( \frac{2}{2n} \left[ \frac{n}{2} \log \det(2\pi eK) - nR_1 + nR_2 + I(\hat{S}_1; \hat{S}_2) \right] \right)
\]
\[
+ \exp \left( \frac{2}{2n} \left[ \frac{n}{2} \log \det(2\pi e\Lambda) \right] \right)
\]
\[
= \det(2\pi eK) \exp[-(R_1 + R_2)] \exp(\frac{1}{n} I(\hat{S}_1; \hat{S}_2)) + \det(2\pi e\Lambda)
\]
\[
= \alpha t \det(2\pi eK)^{\frac{1}{2}} + \det(2\pi e\Lambda)^{\frac{1}{2}}
\]
where \( \alpha = \exp[-(R_1 + R_2)] \) and \( t = \exp\left(\frac{1}{n} I(\hat{S}_1; \hat{S}_2)\right) \). So,

\[
I(\hat{S}_1, \hat{S}_2; \hat{S}) \leq \frac{n}{2} \log \det (2\pi e (K + \Lambda)) - n \log \left( \alpha t \det(2\pi e K)^{\frac{1}{2}} + \det(2\pi e \Lambda)^{\frac{1}{2}} \right)
\]

\[
= n \log \frac{\det(K + \Lambda)^{\frac{1}{2}}}{\alpha t \det(K)^{\frac{1}{2}} + \det(\Lambda)^{\frac{1}{2}}}
\]

and

\[
\log t = \frac{1}{n} I(\hat{S}_1; \hat{S}_2)
\]

\[
\geq \tilde{R}(D'_1) + \tilde{R}(D'_2) - \log \frac{\det(K + \Lambda)^{\frac{1}{2}}}{\alpha t \det(K)^{\frac{1}{2}} + \det(\Lambda)^{\frac{1}{2}}}
\]

Finally we will have

\[
t \geq \frac{\det(\Lambda)^{\frac{1}{2}}}{\det(K + \Lambda)^{\frac{1}{2}} \exp[-\tilde{R}(D'_1) - \tilde{R}(D'_2)] - \det(K)^{\frac{1}{2}} \exp[-R_1 - R_2]}
\]

\[
= \frac{\det(K + \Lambda)^{\frac{1}{2}} \det(D_0 + \Lambda)^{\frac{1}{2}} \det(D_1 + \Lambda)^{\frac{1}{2}} - \det(K)^{\frac{1}{2}} \exp[-R_1 - R_2]}
\]

Thus, since \( \det(D_0)^{\frac{1}{2}} / \det(K)^{\frac{1}{2}} \geq \alpha t \), we have

\[
\exp[-R_1 - R_2] \leq \frac{\det(D_0)^{\frac{1}{2}} \det(D_1 + \Lambda)^{\frac{1}{2}} \det(D_2 + \Lambda)^{\frac{1}{2}}}{\det(K)^{\frac{1}{2}} \det(K + \Lambda)^{\frac{1}{2}} \left( \det(\Lambda)^{\frac{1}{2}} + \det(D_0)^{\frac{1}{2}} \right)}
\]

or

\[
R_1 + R_2 \geq \log \left[ \frac{\det(D_0)^{\frac{1}{2}} \det(D_1 + \Lambda)^{\frac{1}{2}} \det(D_2 + \Lambda)^{\frac{1}{2}}}{\det(K)^{\frac{1}{2}} \det(K + \Lambda)^{\frac{1}{2}} \left( \det(\Lambda)^{\frac{1}{2}} + \det(D_0)^{\frac{1}{2}} \right)} \right]
\]

and therefore we have to solve the following maximization problem

\[
\max_{\Lambda \succeq 0} \frac{\det(D_0)^{\frac{1}{2}} \det(D_1 + \Lambda)^{\frac{1}{2}} \det(D_2 + \Lambda)^{\frac{1}{2}}}{\det(K)^{\frac{1}{2}} \det(K + \Lambda)^{\frac{1}{2}} \left( \det(\Lambda)^{\frac{1}{2}} + \det(D_0)^{\frac{1}{2}} \right)}
\]

(6.4)

to find the optimal \( \Lambda \).
6.2 Gaussian Encoding

In this Section we are going to characterize rates which are achievable by using Gaussian encoding for the vector MD problem. As before, assume the source is zero-mean Gaussian with covariance matrix

$$K = \begin{pmatrix} \sigma_x^2 & \mu \sigma_x \sigma_y \\ \mu \sigma_x \sigma_y & \sigma_y^2 \end{pmatrix}.$$  \hspace{1cm} (6.5)

We denote the two descriptions of the source sequence by $U$ and $V$ where

$$U = S + N_1,$$
$$V = S + N_2.$$

In the above equations, $N_i$, $i = 1, 2$ is a zero-mean Gaussian noise with covariance matrix $\Sigma_i^2$. Note that $\Sigma_i^2$’s are symmetric, and so we can write them in this form instead of $\Sigma_i \Sigma_i^\dagger$. They are also independent from the source. There may be some cross correlation between $N_1$ and $N_2$ in the optimal case. So in general we can define the cross correlation matrix of $N = [N_1 N_2]^\dagger$ as

$$\Phi = \mathbb{E} \left[ [N_{11}, N_{12}, N_{21}, N_{22}] [N_{11}, N_{12}, N_{21}, N_{22}]^\dagger \right]$$

$$= \begin{pmatrix} \Sigma_1^2 & \Sigma_1 R \Sigma_2 \\ \Sigma_2 R^\dagger \Sigma_1 & \Sigma_2^2 \end{pmatrix}.$$

where $\Sigma_1^2, \Sigma_2^2$, and $\Phi$ are semi-positive definite matrices. It is easy to show that the covariance matrix of $U$ and $V$ is

$$Q = \begin{pmatrix} K + \Sigma_1^2 & K + \Sigma_1^R \Sigma_2 \\ K + \Sigma_2 R^\dagger \Sigma_1 & K + \Sigma_2^2 \end{pmatrix}.$$

Using MMSE decoder, it is easy to show that the best estimations for $S$ in the
marginal decoders are

\[ \hat{S}_1 = K(K + \Sigma_1^2)^{-1}U \]
\[ \hat{S}_2 = K(K + \Sigma_2^2)^{-1}V \]

and the best estimation for the source in the central decoder by using both the
descriptions is

\[ \hat{S}_0 = \left( \begin{array}{cc} K & K \\ K & K + \Sigma_1 R \Sigma_2 \end{array} \right)^{-1} \left( \begin{array}{c} U \\ V \end{array} \right). \]

Using the above notation and the fact that the original source and the noises
are independent and all of them are Gaussian random variables, we can write the
following expressions.

\[ I(S; U) = h(U) - h(U|S) \]
\[ = h(U) - h(U - S|S) \]
\[ \overset{(a)}{=} h(U) - h(N_1) \]
\[ \overset{(b)}{=} \frac{1}{2} \log \det(K + \Sigma_1^2) - \frac{1}{2} \log \det(\Sigma_1^2) \]
\[ = \frac{1}{2} \log \det((K + \Sigma_1^2)\Sigma_1^{-2}) \]

where in (a) we have used the independence of \( S \) and \( N_1 \) and (b) holds because both
of the random vectors are Gaussian. Similarly, we have

\[ I(S; V) = \frac{1}{2} \log \det((K + \Sigma_2^2)\Sigma_2^{-2}) \]

and

\[ I(S; U, V) = h(U, V) - h(U, V|S) \]
\[ = \frac{1}{2} \log \det Q - \frac{1}{2} \log \det \Phi. \]
We have also

\[
I(U; V) = h(U) + h(V) - h(U, V)
\]

\[
= \frac{1}{2} \log \det(K + \Sigma^2_{1}) + \frac{1}{2} \log \det(K + \Sigma^2_{2}) - \frac{1}{2} \log \det Q.
\]

Therefore

\[
I(S; U, V) + I(U; V) = \frac{1}{2} \log \det(K + \Sigma^2_{1})(K + \Sigma^2_{2}) - \frac{1}{2} \log \det \Phi.
\]

Now we can calculate the distortion between the estimated versions of the source and the original version, and find the necessary constraints on the covariance of the noise.

\[
\mathbb{E}[(S - \hat{S}_1)(S - \hat{S}_1)^\dagger] \overset{(c)}{=} \mathbb{E}[(S - \hat{S}_1)S^\dagger]
\]

\[
= \mathbb{E}[SS^\dagger] - \mathbb{E}[\hat{S}_1S^\dagger]
\]

\[
= \mathbb{E}[SS^\dagger] - \mathbb{E}[K(K + \Sigma^2_{1})^{-1}US^\dagger]
\]

\[
= \mathbb{E}[SS^\dagger] - K(K + \Sigma^2_{1})^{-1}\mathbb{E}[(S + N_1)S^\dagger]
\]

\[
= K - K(K + \Sigma^2_{1})^{-1}K
\]

\[
= (K^{-1} + \Sigma^{-2}_{1})^{-1}
\]

\[
:= D_1
\]

where in (c) we have used the orthogonality of error to the estimator in MMSE.

Similarly, we have

\[
\mathbb{E}[(S - \hat{S}_2)(S - \hat{S}_2)^\dagger] = K - K(K + \Sigma^2_{2})^{-1}K
\]

\[
= (K^{-1} + \Sigma^{-2}_{2})^{-1}
\]

\[
:= D_2
\]
where the diagonal elements of $D_1$ and $D_2$ should satisfy in the given distortion conditions. Finally, for central decoder’s estimation we have

$$
\mathbb{E} \left[ (S - \hat{S}_0)(S - \hat{S}_0)^\dagger \right] = \mathbb{E} \left[ (S - \hat{S}_0)S^\dagger \right] = \mathbb{E}[SS^\dagger] - \mathbb{E}\left[ \begin{pmatrix} K & K \end{pmatrix} Q^{-1} \begin{pmatrix} S + N_1 \\ S + N_1 \end{pmatrix} S^\dagger \right] = K - \begin{pmatrix} K & K \end{pmatrix} Q^{-1} \begin{pmatrix} K \\ K \end{pmatrix} \quad (6.7)
$$

It is easy to show that

$$
\mathbb{E} \left[ (S - \hat{S}_0)(S - \hat{S}_0)^\dagger \right] = \left[ K^{-1} + \begin{pmatrix} I & I \end{pmatrix} \Phi^{-1} \begin{pmatrix} I \\ I \end{pmatrix} \right]^{-1} \quad (6.8)
$$

In order to satisfaction of the marginal distortion constraints, we have

$$
R_1 \geq I(S; U) = \frac{1}{2} \log \det \left( (K + \Sigma_1^2)\Sigma_1^{-2} \right) = \frac{1}{2} \log \det(KD_1^{-1})
$$

and

$$
R_2 \geq I(S; V) = \frac{1}{2} \log \det \left( (K + \Sigma_2^2)\Sigma_2^{-2} \right) = \frac{1}{2} \log \det(KD_2^{-1})
$$

and for the central decoder, we

$$
R_1 + R_2 \geq I(S; U, V) + I(U; V) = \frac{1}{2} \log \det(K + \Sigma_1^2)(K + \Sigma_2^2) - \frac{1}{2} \log \det \Phi = \frac{1}{2} \log \det(KD_1^{-1}) + \frac{1}{2} \log \det(KD_2^{-1}) - \frac{1}{2} \log \det(I - RR^\dagger)
$$
So, in order to find a tight lower bound for the summation of the rates, we have to maximize \( \det(\Phi) \) subject to some conditions, i.e.,

\[
\max \begin{bmatrix} K^{-1} + \mathbf{I} & \mathbf{I} \\ \Phi^{-1} \left( \begin{array}{c} \mathbf{I} \\ \mathbf{I} \end{array} \right) \end{bmatrix}^{-1} \preceq D_0
\]

We also know that in order for \( \Phi \) to be positive definite, it is necessary and sufficient that one of the diagonal blocks and its Schur complement being positive definite. Since \( \Sigma_1^2 \succeq 0 \) and \( \Sigma_2^2 \succeq 0 \), the condition reduces to \( \mathbf{I} - RR^\dagger \succeq 0 \). Therefore, the optimization problem is translated to

\[
\max_{\mathbf{I} - RR^\dagger \succeq 0} \det(\mathbf{I} - RR^\dagger) \quad (6.10)
\]

Although it is not too hard to solve the similar optimization problems (6.4) and (6.10) in scalar case, solving them in matrix case may be difficult and messy. On the other hand we don’t need to come up with the exact solution of the optimization problem, but just need to show that they have the same solution. By the way, in order to the EGC region being tight in this case, all the inequalities we used, should appear as equalities. These give us some more conditions which may help us to deal with the optimization problem. In the next two sections, we will address the conditions.

### 6.3 Entropy Power Inequality Constraint

The conditional entropy power inequality is one of the inequalities which has to appear as equality for the inner bound coincides to the outer bound. We saw before in Section 4.2 that holding the equality in the scalar case is guaranteed by some simple conditions.
about the independency of the added noise and the source. In the case of vector MD, that condition is not sufficient and we need more. Using the Gaussian encoding in (6.6), and the fact that $\hat{S}_1$ and $\hat{S}_2$ are deterministic and one-to-one functions of $U$ and $V$, we can write the terms in (6.3) as

$$h(\hat{S}|\hat{S}_1, \hat{S}_2) = h(\hat{S} | U, V)$$

$$= h(\hat{S}, U, V) - h(U, V)$$

$$= \frac{n}{2} \log \det \left( \begin{bmatrix} E[\hat{S}\hat{S}^T] & E[\hat{S}U^T] & E[\hat{S}V^T] \\ E[U\hat{S}^T] & UU^T & E[UV^T] \\ E[V\hat{S}^T] & VU^T & E[VV^T] \end{bmatrix} \right)$$

$$- \frac{n}{2} \log \det \left( \begin{bmatrix} E[UU^T] & E[UV^T] \\ E[VU^T] & E[VV^T] \end{bmatrix} \right).$$

According to the definition of $\hat{S}$, $U$, and $V$ we have

$$E[\hat{S}\hat{S}^T] = K + \Lambda$$

where $\Lambda = E[ZZ^T]$, and

$$E[\hat{S}U^T] = E[(S + Z)(S^T + N_1^T)] = K.$$

Similarly, we have

$$E[\hat{S}V^T] = K$$

$$E[U\hat{S}^T] = K$$

$$E[UU^T] = (K + \Sigma_2^2)$$

$$E[UV^T] = K + \Sigma_1 R \Sigma_1$$

$$E[V\hat{S}^T] = K$$

$$E[VU^T] = K + \Sigma_2 R^\dagger \Sigma_1$$

$$E[VV^T] = K + \Sigma_2^2.$$
So
\[
 h(\tilde{S}|U, V) = \frac{n}{2} \log \det \left( \begin{array}{ccc}
 K + \Lambda & K & K \\
 K & K + \Sigma_1^2 & K + \Sigma_1 R \Sigma_2 \\
 K & K + \Sigma_2 R \Sigma_1 & K + \Sigma_2^2
 \end{array} \right) -
\]
\[
 \frac{n}{2} \log \det \left( \begin{array}{cc}
 K + \Sigma_1^2 & K + \Sigma_1 R \Sigma_2 \\
 K + \Sigma_2 R \Sigma_1 & K + \Sigma_2^2
 \end{array} \right)
\]

Using Schur complements for calculating the determinants,
\[
 \det \left( \begin{array}{cc}
 A & B \\
 C & D
 \end{array} \right) = \det D \cdot \det(A - BD^{-1}C), \quad (6.11)
\]
we have
\[
 \det \left( \begin{array}{ccc}
 K + \Lambda & K & K \\
 K & K + \Sigma_1^2 & K + \Sigma_1 R \Sigma_2 \\
 K & K + \Sigma_2 R \Sigma_1 & K + \Sigma_2^2
 \end{array} \right) =
\]
\[
 \det Q \cdot \det \left( K + \Lambda - \left( \begin{array}{cc}
 K & K \\
 K & K
 \end{array} \right) Q^{-1} \left( \begin{array}{c}
 K \\
 K
 \end{array} \right) \right)
\]

Thus,
\[
 h(\tilde{S}|U, V) = \frac{n}{2} \log \det \left[ K + \Lambda - \left( \begin{array}{cc}
 K & K \\
 K & K
 \end{array} \right) Q^{-1} \left( \begin{array}{c}
 K \\
 K
 \end{array} \right) \right]
\]

Therefore, we have
\[
 \exp \left( \frac{2}{2n} h(\tilde{S}|U, V) \right) = \det \left[ K + \Lambda - \left( \begin{array}{cc}
 K & K \\
 K & K
 \end{array} \right) Q^{-1} \left( \begin{array}{c}
 K \\
 K
 \end{array} \right) \right]^{\frac{1}{2}}
\]

Using the same argument we can show
\[
 \exp \left( \frac{2}{2n} h(S|\hat{S}_1, \hat{S}_2) \right) = \det \left[ K - \left( \begin{array}{cc}
 K & K \\
 K & K
 \end{array} \right) Q^{-1} \left( \begin{array}{c}
 K \\
 K
 \end{array} \right) \right]^{\frac{1}{2}}
\]
and
\[
\exp \left( \frac{2}{2n} h(Z) \right) = \det(\Lambda)^{\frac{1}{2}}
\]

According to (6.3) we have
\[
\det \left[ K + \Lambda - \left( \begin{array}{cc} K & K \\ K & K \end{array} \right) Q^{-1} \left( \begin{array}{c} K \\ K \end{array} \right) \right]^{\frac{1}{2}} \geq \det \left[ K - \left( \begin{array}{cc} K & K \\ K & K \end{array} \right) Q^{-1} \left( \begin{array}{c} K \\ K \end{array} \right) \right]^{\frac{1}{2}} + \det(\Lambda)^{\frac{1}{2}}
\]
where according to Minkowski inequality, the equality holds if and only if all the above matrices are proportional, i. e.,
\[
K - \left( \begin{array}{cc} K & K \\ K & K \end{array} \right) Q^{-1} \left( \begin{array}{c} K \\ K \end{array} \right) = \alpha \Lambda
\]  
(6.12)

### 6.4 Conditional Markov Chain

The other inequality in the Ozarow’s proof which has to be equality in order to the outer bound meets the inner bound, is

\[
I(\hat{S}_1, \hat{S}_2 | \tilde{S}) \geq 0.
\]

In the case of equality, we have

\[
h(\hat{S}_1, \tilde{S}) + h(\hat{S}_2, \tilde{S}) = h(\hat{S}_1, \hat{S}_2, \tilde{S}) + h(\tilde{S})
\]

which means
\[
\frac{1}{2} \log \det 2\pi e \left( \begin{array}{cc} K + \Sigma_1^2 & K \\ K & K + \Lambda \end{array} \right) + \frac{1}{2} \log \det 2\pi e \left( \begin{array}{cc} K + \Sigma_2^2 & K \\ K & K + \Lambda \end{array} \right) = \\
\frac{1}{2} \log \det 2\pi e \left( \begin{array}{ccc} K + \Sigma_1^2 & K + \Sigma_1 R \Sigma_2 & K \\ K + \Sigma_2 R^T \Sigma_1 & K + \Sigma_1^2 & K \\ K & K & K + \Lambda \end{array} \right) + \frac{1}{2} \log \det 2\pi e (K + \Lambda)
\]
or

\[
\det \begin{pmatrix}
\mathbf{K} + \Sigma_1^2 & \mathbf{K} \\
\mathbf{K} & \mathbf{K} + \Lambda
\end{pmatrix} \cdot \det \begin{pmatrix}
\mathbf{K} + \Sigma_2^2 & \mathbf{K} \\
\mathbf{K} & \mathbf{K} + \Lambda
\end{pmatrix} =
\]

\[
\det \begin{pmatrix}
\mathbf{K} + \Sigma_1^2 & \mathbf{K} + \Sigma_1 \Sigma \Sigma_1 & \mathbf{K} \\
\mathbf{K} + \Sigma_2 \Sigma \Sigma_1^1 & \mathbf{K} + \Sigma_1^2 & \mathbf{K} \\
\mathbf{K} & \mathbf{K} & \mathbf{K} + \Lambda
\end{pmatrix} \cdot \det (\mathbf{K} + \Lambda).
\]

By expanding all determinants in terms of \( \mathbf{K} + \Lambda \), and its Schur complement, we have

\[
\det (\mathbf{K} + \Sigma_1^2 - \mathbf{K} (\mathbf{K} + \Lambda)^{-1} \mathbf{K}) \cdot \det (\mathbf{K} + \Sigma_2^2 - \mathbf{K} (\mathbf{K} + \Lambda)^{-1} \mathbf{K}) =
\]

\[
\det \left[ \mathbf{Q} - \begin{pmatrix} \mathbf{K} \\ \mathbf{K} \end{pmatrix} (\mathbf{K} + \Lambda)^{-1} \begin{pmatrix} \mathbf{K} & \mathbf{K} \end{pmatrix} \right]
\]

Let \( \Psi = \mathbf{K} (\mathbf{K} + \Lambda)^{-1} \mathbf{K} - \mathbf{K} \). By Substituting \( \Psi \) in the above expression, we have

\[
\det (\Sigma_1^2 - \Psi) \cdot \det (\Sigma_2^2 - \Psi) = \det \begin{pmatrix}
\Sigma_1^2 - \Psi & \Sigma_1 \Sigma \Sigma_2 - \Psi \\
\Sigma_2 \Sigma \Sigma_1^1 - \Psi & \Sigma_2^2 - \Psi
\end{pmatrix}
\]

Note that the RHS is determinant of a semi-positive definite matrix and the above equality holds if and only if \( \Psi = \Sigma_1 \Sigma \Sigma_2 \) (or \( \Psi = \Sigma_2 \Sigma^\dagger \Sigma_1 \)). This gives us that the optimal covariance of the noise of the auxiliary variable, can be found from the equation

\[
\Lambda = \mathbf{K} (\mathbf{K} + \Sigma_1 \Sigma \Sigma_2)^{-1} \mathbf{K} - \mathbf{K}.
\]

(6.13)
Chapter 7

Conclusion and Future Work

In this work we considered the problem of multiple description source coding for a vector of Gaussian sources where there is some correlation between the coordinates. Assume the setting of the system as the following:

Let $S = [X, Y]^\dagger$, be a zero-mean jointly Gaussian source and $s(1), s(2), \ldots$ be a sequence of i.i.d. source symbols drawn from the alphabet set ($\mathcal{X} \times \mathcal{Y}$)$^n$. $X$ and $Y$ are zero-mean Gaussian random variables with covariance matrix

$$K = \begin{pmatrix} \sigma_x^2 & \mu \sigma_x \sigma_y \\ \mu \sigma_x \sigma_y & \sigma_y^2 \end{pmatrix}. $$

Given the sequence of the source, each of the encoders produces a description of the sequence of some rate, and sends its description to the decoders. At the marginal decoders, given any of the descriptions, we need to reconstruct $S$ with quality not worse than some given distortion. The central decoder which has access to both of the descriptions, has to find a better estimation of the source, $S$. Denoting the marginal estimations by $\hat{X}_1, \hat{Y}_1, \hat{X}_2, \hat{Y}_2$ and the central estimations by $\hat{X}_0$ and $\hat{Y}_0$, they should satisfy in the distortion constraints as the following
\[ \mathbb{E}\left[d_{1,x}(X, \hat{X}_1)\right] \leq D_{1,x} \]
\[ \mathbb{E}\left[d_{1,y}(Y, \hat{Y}_1)\right] \leq D_{1,y} \]
\[ \mathbb{E}\left[d_{2,x}(X, \hat{X}_2)\right] \leq D_{2,x} \]
\[ \mathbb{E}\left[d_{2,y}(Y, \hat{Y}_2)\right] \leq D_{2,y} \]
\[ \mathbb{E}\left[d_{0,x}(X, \hat{X}_0)\right] \leq D_{0,x} \]
\[ \mathbb{E}\left[d_{0,y}(Y, \hat{Y}_0)\right] \leq D_{0,y} \] (7.1)

where \( d_{i,x}(\cdot, \cdot) \) and \( d_{i,y}(\cdot, \cdot) \) are some single letter distortion measures.

The problem is to characterize all achievable \((R_1, R_2, D_{0,x}, D_{0,y}, D_{1,x}, D_{1,y}, D_{2,x}, D_{2,y})\).

We stated the problem for the scalar source case and reviewed some important results. In particular we investigated the technique which used by Ozarow for the Gaussian source and tried to apply it for the new problem.

We extended the El Gamal and Cover Theorem for the vector source multiple descriptions which results in an inner bound for the achievable points. In the converse part, although we could not completely fit the Ozarow’s proof for our problem, we found some important conditions which will lead us to solve the problem.

A very good idea to complete this solution is to deal with these conditions and to show that our optimization problems have the same solution. We showed that in order to the marginal estimations be independent conditioned on the auxiliary random variable, its covariance matrix should satisfy in some equation. On the other hand, it should be proportional to some other matrix. In order to make both the conditions satisfied, we can use a notion of enhanced vector multiple description, in which we relax some of non-strict distortion constraints.
This problem also can be considered for the case of more than two correlated sources. In such problem, we have more non-strict distortion constraint. The other interesting problem is to characterize achievable rates for correlated source of other distributions and in particular binary sources.

Another interesting problem is considering the same question for vector of correlated sources with correlated side information which can be available only at the decoders or at both of the encoders and decoders.

This is also interesting to focus on the vector multiple descriptions problem for the no excess rate case and try to extend Ahlswede results.

Finally, we propose the problem of successive refinement of information for correlated vector sources.
Bibliography


Index

Ahlswede R., 22
Berger T., 21
Blachman N., 16
Chebychev inequality, 42
convexity, 17
covariance matrix, 50, 53
Cover T. M., 19
date-processing theorem, 26
Diggavi, S. N., 22
distortion-rate function, 22
distribution
  Gaussian, 25, 27
  Normal, 20
  uniform, 39
diversity, 1
  multiple routes, 1
EGC theorem, 19
El Gamal, A. A., 19
entropy power inequality, 16, 32
  conditional, 28
  vector, 51
excess rate, 22
Fu, Fang-Wei, 22
Gaussian r.v.
  jointly, 49
Hamming distance function, 21
Jensen’s inequality, 17
Markov chain, 16, 33
mean-square measure, 29
minimum mean-squared error, 30
Minkowski inequality, 60
MMSE, 53
multiple descriptions, 1, 19
  vector, 25, 34
mutual information, 26
optimum diversity, 19
Ozarow L., 20, 25
rate-distortion theorem, 27
rate-distortion theory, 3, 26
Schur complement, 57, 59
semi-positive definite, 53, 61
side information, 22
source coding, 19
typicality
  strong, 14, 44
Vaishampayan V. A., 22
Witsenhausen H. S., 20
Wyner, A. D., 22
Yeung, Raymond W., 22
Zhang Z., 21
Ziv, J., 22