ON THE WAGNER-ANANTHARAM OUTER BOUND
AND ACHIEVABLE GAUSSIAN SOURCE CODING
EXPONENTS

A Thesis
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Master of Science

by
Aggelos Vamvatsikos
January 2007
ABSTRACT

Tightness of the Wagner-Anantharam (W-A) outer bound [1], for the quadratic Gaussian two-terminal source coding problem, is examined. The proof of the sum rate constraint for the rate region of this problem provides some hints on possible looseness of the bound [2]. We prove tightness to the rate region for this setup, by first proving tightness for the many-help-one problem with conditional independence [6]. We also look at the performance of the W-A bound and find the worst choice of the auxiliary random variable $X$, appearing in the expression of the bound, for the sum rate constraint.

In the second part of this work, the Gaussian point-to-point source coding problem is considered. The error exponent for this problem was presented by Ihara and Kubo [11]. We generalize the Gaussian method of types, introduced by Arikan and Merhav [10], and use Marton’s approach [9] to retrieve the best achievable error exponent for this setup. Our method is readily extendable to more complex Gaussian source coding problems.
Biographical Sketch

Aggelos Vamvatsikos was born in Nea Ionia Volos, Greece in 1981. He received his Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens in 2004. Later that year, he joined the graduate program at Cornell University, where he studied Information Theory.
Acknowledgements

I would like to thank the chairman of my committee, Prof. Aaron Wagner, for his advice and patience during the preparation of this thesis. I am also grateful to Prof. Toby Berger for introducing me to the beautiful world of Information Theory and to Prof. Sergio Servetto for sitting on my thesis committee.

Finally, I would like to thank my family for supporting my decisions and helping me overcome all difficulties throughout the last two years. I also wish to thank Maria for encouraging me and reminding me that life is a lot more than studying.
# Table of Contents

1 Introduction ................................. 1

2 Results on the W-A Outer Bound .......... 4
   2.1 The Quadratic Gaussian Two-Terminal Source Coding Problem .... 4
       2.1.1 Definitions, Problem Statement and Result .................. 6
       2.1.2 Proof of Result ........................................ 11
   2.2 The Worst Choice of X for the Sum Rate Constraint ............. 23

3 Results on Achievable Gaussian Error Exponents ... 25
   3.1 The Point-to-Point Lossy Source Coding Problem ............... 25
       3.1.1 Problem Statement and Result .......................... 26
       3.1.2 The Gaussian Method of Types and Other Lemmas .......... 28
       3.1.3 Proof of Achievable Exponent .......................... 39
   3.2 Overview on Source Coding Exponents ........................ 42

Bibliography .................................. 45
## List of Figures

2.1 The two-terminal source coding problem. .......................... 4  
2.2 The multiterminal source coding problem. .......................... 7  
2.3 Communication system with $L$ helpers. ........................... 12  
3.1 The point-to-point source coding problem. .......................... 25  
3.2 Error exponent $\theta(R, d, P_Y)$ for $d = 0.5$. ...................... 29
Chapter 1

Introduction

Multiterminal source coding theory has been launched by an award-winning paper by Slepian and Wolf [13]. They considered the system, where two correlated information sources are separately encoded and sent to a single decoder over rate-constrained, noiseless channels. The decoder attempts to reconstruct the original sources with an arbitrarily small probability of error. For this system, Slepian and Wolf determined the set of admissible rate pairs, which in the sequel will be called the rate region. An immediate generalization of this setup is the rate-distortion problem, where the source outputs are reconstructed within average distortions smaller than a prescribed amount. We seek to determine the rate region that allows us to meet a given pair of target distortions. Such a situation suggests multiterminal rate-distortion theory. This problem is open.

Many special cases have been solved, however. For some of these, the reader is referred to papers of Wyner and Ziv [16], Berger and Yeung [21], Oohama [6] and references therein. In an effort to approach the problem in its full generality, Berger and Tung developed inner and outer bounds on the set of admissible rates [4, 5]. By examining instances of the problem that can be solved from first principles,
both bounds were proved to be loose to the true rate region. In 2005, Wagner and Anantharam provided an improved outer bound for this problem [1]. They proved that their bound subsumes the Berger-Tung (B-T) outer bound and gave examples for which the containment is strict. They also showed that the Wagner-Anantharam (W-A) outer bound is tight to all known solutions in multiterminal rate-distortion theory.

A question that naturally rises from this work is whether this improved bound solves the multiterminal rate-distortion problem. Is there an example, where the W-A bound can be proved to be loose? Recently, using a novel approach, Wagner, Tavildar, and Viswanath [2] improved Oohama’s work [7] and derived the rate region for the quadratic Gaussian two-terminal source coding problem. The construction of their proof provides some hints on possible looseness of the W-A outer bound for the sum rate constraint. Examining tightness of the latter bound to the rate region of this system is an interesting problem to work on. Part of this thesis addresses this issue.

A natural criterion for assessing the performance of a block code is the probability of excess distortion. Since this probability will typically converge to zero exponentially rapidly, it is the rate of this exponential convergence, commonly referred to as the “error exponent”, which is of significance. What is the form of this exponent for the quadratic Gaussian two-terminal source coding problem? In order to answer this question one has to examine more fundamental problems in this area. The Gaussian point-to-point problem is the easiest setup to start. A direct extension of Marton’s exponent for discrete sources [9] to the Gaussian case is the simplest way to follow. The main drawback, however, is the fact that for continuous alphabet sources the method of types no longer applies. We deal with
this issue in the second part of our work.

The organization of this thesis is as follows. In Chapter 2, we prove tightness of the W-A outer bound for the quadratic Gaussian two-terminal source coding problem. We also present the worst choice of the auxiliary variable X, appearing in the expression of the W-A bound, for the case of the sum rate constraint. Chapter 3 is devoted to extending the Gaussian method of types, originally introduced in [10], and developing a straightforward proof for the best achievable error exponent in the Gaussian point-to-point lossy source coding setup [10, 11]. An overview of existing results in the area of error exponents for lossy source coding is also provided.
Chapter 2

Results on the W-A Outer Bound

2.1 The Quadratic Gaussian Two-Terminal Source Coding Problem

So far there have been found several examples for which the W-A outer bound [1] has been proved to be tight. However, there are no examples reported where the bound fails to provide the actual rate region. A natural candidate is the quadratic Gaussian two-terminal source coding problem, the setup of which is depicted in Figure 2.1. Two encoders each observe one component of a memoryless Gaussian vector-valued source. The encoders separately communicate with a decoder, which attempts to reproduce the vector-valued source subject to a mean squared error.

![Diagram](image)

Figure 2.1: The two-terminal source coding problem.
distortion constraint for each component. The rate region for this problem was recently found by Wagner, Tavildar, and Viswanath [2].

The reason why this problem is a candidate for proving looseness of the W-A outer bound can easily be explained. Take the expression of the bound for the sum rate constraint

\[ R_1 + R_2 \geq \max_X \min_{U_1, U_2, W, T} I(X; U_1, U_2|T) + \sum_{i=1}^{2} I(Y_i; U_i|X, W, T). \]

In order to evaluate this expression, we first have to fix the random variable \( X \) and then specify a coding scheme for the chosen setup. On the contrary, the proof of Wagner, Tavildar, and Viswanath [2] is built in the exact opposite way. They derive a lower bound for the sum rate by first fixing a coding scheme and then evaluating the sum rate for two different setups. Taking the maximum rate of the two setups yields the desired sum rate bound, which is proved to be tight to the true sum rate. This method points towards an improved expression for the sum rate constraint, where the coding is performed before the specification of \( X \), namely

\[ R_1 + R_2 \geq \min_{U_1, U_2, T} \max_X \min_W I(X; U_1, U_2|T) + \sum_{i=1}^{2} I(Y_i; U_i|X, W, T). \]

This expression gives a better lower bound for the sum rate and shows why the W-A bound is likely to be loose. In this section, however, we prove that this is not the case, i.e that the W-A outer bound is tight to the actual rate region of the problem. Our proof makes use of the techniques of Oohama [6] and Wagner, Tavildar, and Viswanath [2].

We use boldface letters to denote vectors \((\mu, U)\) and matrices \((D)\). Lightface letters denote scalars \((\rho, R)\). Whether a boldface letter is a vector or a matrix should be clear from the context.
2.1.1 Definitions, Problem Statement and Result

We start by giving the definitions of the W-A and B-T outer bounds for discrete random variables. We then specialize the W-A bound in the case of the quadratic Gaussian two-terminal source coding problem and give the statement of our result.

The W-A and B-T Outer Bounds

Let \( L \) denote the number of encoders and \( \Lambda = \{1, \ldots, L\} \). In addition, let \( \{Y^n_0(t), Y^n_1(t), \ldots, Y^n_L(t)\}_{t=1}^n \) be a vector-valued, discrete memoryless source. For \( A \subset \Lambda \), we denote \( (Y^n_l(t))_{t \in A} \) by \( Y^n_A(t) \). If \( A = \Lambda \), we write this simply as \( Y^n(t) \).

In this context, the set \( A^c \) should be interpreted as \( \{1, \ldots, L\} - A \) rather than \( \{0, \ldots, L\} - A \). When \( A = \{l\} \), we shall write \( Y^n_l(t) \) and \( Y^n_{l^c}(t) \) in place of \( Y^n_{\{l\}}(t) \) and \( Y^n_{\{l^c\}}(t) \), respectively. We use \( Y^n_l \) to denote \( \{Y^n_l(t)\}_{t=1}^n \). Similar notation will be used for other vectors that appear later.

For each \( l \in \Lambda \), encoder \( l \) observes \( Y^n_l \) and then employs a mapping

\[
f^{(n)}_l : Y^n_l \mapsto \{1, \ldots, M^{(n)}_l\}
\]

to convey information about it to the decoder (see Figure 2.2). The decoder uses the received messages to estimate \( K \) functions of the vector-valued source according to the mappings

\[
\phi^{(n)}_k : \prod_{l=1}^L \{1, \ldots, M^{(n)}_l\} \mapsto \hat{Y}^{(n)}_k \text{ for } k = 1, \ldots, K.
\]

We assume that \( K \) distortion measures \( D_k : \prod_{l=0}^L Y_l \times \hat{Y}_k \mapsto \mathbb{R}_+ \) are given.

Definition 1. The rate-distortion vector

\[
(R, d) = (R_1, R_2, \ldots, R_L, d_1, d_2, \ldots, d_K)
\]
Figure 2.2: The multiterminal source coding problem.

is achievable if there exists a block length $n$, encoders $f^{(n)}_l$, and a decoder

$$
\left( \phi^{(n)}_1, \ldots, \phi^{(n)}_K \right)
$$

such that

$$
R_l \geq \frac{1}{n} \log M_l^{(n)} \text{ for all } l, \text{ and}
$$

$$
d_k \geq \mathbb{E} \left[ \frac{1}{n} \sum_{t=1}^{n} D_k(Y_0^n(t), Y^n(t), \hat{Y}_k^n(t)) \right] \text{ for all } k.
$$

Let $\mathcal{RD}^*$ be the set of achievable rate-distortion vectors. Its closure, $\overline{\mathcal{RD}^*}$, is called the rate-distortion region.

Now let $Y_0, \ldots, Y_L$ be generic random variables with the distribution of the source at a single time. We denote by $\chi$ the set of discrete random variables $X$ with the property that $Y_1, \ldots, Y_L$ are conditionally independent given $X$. We also let $\Gamma_o$ be the set of discrete random variables

$$
\gamma = (U_1, \ldots, U_L, \hat{Y}_1, \ldots, \hat{Y}_K, W, T)
$$

satisfying

\footnote{All logarithms in this chapter are base two.}
1. \((W, T)\) is independent of \((Y_0, Y)\),

2. \(U_t \leftrightarrow (Y_t, W, T) \leftrightarrow (Y_0, Y_{t'}, U_{t'})\),

3. \((Y_0, Y, W) \leftrightarrow (U, T) \leftrightarrow \hat{Y}\).

In addition, we assume that \(X \in \chi\) and \(\gamma \in \Gamma_o\) are Markov coupled, i.e. \(X \leftrightarrow (Y_0, Y) \leftrightarrow \gamma\). The definition of the W-A bound follows.

**Definition 2.** Let

\[
RD_o(X, \gamma) = \left\{ (R, d) : \sum_{t \in A} R_t \geq I(X; U_A|U_{A-A}, T) + \sum_{l \in A} I(Y_l; U_l|X, W, T) \right. \\
\left. \text{for all } A \subset \Lambda \text{ and } d_k \geq \mathbb{E}[D_k(Y_0, Y, \hat{Y}_k)] \text{ for all } k \right\}.
\]

Then define

\[
RD_o = \bigcap_{X \in \chi} \bigcup_{\gamma \in \Gamma_o} RD_o(X, \gamma).
\]

We also give the definition of the B-T outer bound. Note that this definition is more general that the one originally introduced by Berger and Tung [4, 5].

**Definition 3.** Let \(\Gamma_{o}^{BT}\) denote the set of finite-alphabet random variables

\[\gamma = (U_1, \ldots, U_L, \hat{Y}_1, \ldots, \hat{Y}_K, T)\]

satisfying

1. \(T\) is independent of \((Y_0, Y)\),

2. \(U_t \leftrightarrow (Y_t, T) \leftrightarrow (Y_0, Y_{t'})\),

3. \((Y_0, Y) \leftrightarrow (U, T) \leftrightarrow \hat{Y}\).
Then let
\[
\mathcal{RD}_o^{BT}(\gamma) = \left\{ (\mathbf{R}, \mathbf{d}) : \sum_{l \in A} R_l \geq I(\mathbf{Y}; \mathbf{U}_A | \mathbf{U}_{\Lambda - A}, T) \right\}.
\]

for all \( A \subset \Lambda \) and \( d_k \geq \mathbb{E}[D_k(Y_0, \hat{Y}_k)] \) for all \( k \).

Finally, let
\[
\mathcal{RD}_o^{HT} = \bigcup_{\gamma \in \Gamma_o^{HT}} \mathcal{RD}_o^{HT}(\gamma).
\]

Problem Formulation and Statement of Result

Let \( \{(Y_1^n(t), Y_2^n(t))\}_{t=1}^n \) be a sequence of independent and identically distributed (i.i.d.) Gaussian zero-mean random vectors and let
\[
\mathbf{K}_y = \begin{bmatrix}
1 & \rho \\
\rho & 1
\end{bmatrix}
\]
denote the covariance matrix of \((Y_1^n(1), Y_2^n(1))\). We also assume that
\[
\sigma_{Y_1}^2 = \sigma_{Y_2}^2 = 1.
\]

Note that there is no loss of generality in this assumption, because the observations and estimates can be scaled to reduce the general case to this one. In what follows, we assume that \( 0 < \rho < 1 \), since the extreme cases \( \rho = 0 \) and \( \rho = 1 \) can be handled using classical results, in which the W-A outer bound is known to be tight.

Let us redefine \( \chi \) to be the set of real-valued random variables \( X \) such that \( Y_1, Y_2 \) are conditionally independent given \( X \). Moreover, let us redefine \( \Gamma_o \) to be the set of random variables \((U_1, U_2, \hat{Y}_1, \hat{Y}_2, W, T)\) such that each takes values in a finite-dimensional Euclidean space, and collectively they satisfy the Markov conditions defining the original \( \Gamma_o \).
1. \((W, T)\) is independent of \((Y_1, Y_2)\),

2. \(U_1 \leftrightarrow (Y_1, W, T) \leftrightarrow (Y_2, U_2), \quad U_2 \leftrightarrow (Y_2, W, T) \leftrightarrow (Y_1, U_1)\),

3. \((Y_1, Y_2, W) \leftrightarrow (U_1, U_2, T) \leftrightarrow (\hat{Y}_1, \hat{Y}_2)\), and

4. the conditional distribution of \(U_l\) given \(W\) and \(T\) is discrete for each \(l\).

Note that the last condition is an extra technical condition that we have introduced for this problem. We again assume that \(X \in \chi\) and \(\gamma \in \Gamma_o\) are Markov coupled, i.e. \(X \leftrightarrow (Y_1, Y_2) \leftrightarrow \gamma\). The definition of the W-A bound for this problem follows.

**Definition 4.** Let

\[
\mathcal{RD}_o(X, \gamma) = \left\{ (R, d) : \sum_{l \in A} R_l \geq I(X; U_A|U_{\{1,2\} \setminus A}, T) + \sum_{l \in A} I(Y_l; U_l|X, W, T) \right\}
\]

for all \(A \subset \{1, 2\}\) and \(d_k \geq \mathbb{E} \left[ (Y_k - \hat{Y}_k)^2 \right] \) for all \(k \in \{1, 2\}\).

Then define

\[
\mathcal{RD}_o = \bigcap_{X \in \chi} \bigcup_{\gamma \in \Gamma_o} \mathcal{RD}_o(X, \gamma).
\]

We also define the following three sets of rate-distortion vectors that we will be using for the proof of our result. Let

\[
R_1^* = \left\{ (R_1, R_2, d_1, d_2) \in \mathbb{R}_+^4 : R_1 \geq \frac{1}{2} \log^+ \left[ \frac{1}{d_1} \left( 1 - \rho^2 + \rho^2 2^{-2R_2} \right) \right] \right\},
\]

where \(\log^+ = \max(\log x, 0)\). In addition,

\[
R_2^* = \left\{ (R_1, R_2, d_1, d_2) \in \mathbb{R}_+^4 : R_2 \geq \frac{1}{2} \log^+ \left[ \frac{1}{d_2} \left( 1 - \rho^2 + \rho^2 2^{-2R_1} \right) \right] \right\}
\]

and finally

\[
R_{12}^* = \left\{ (R_1, R_2, d_1, d_2) \in \mathbb{R}_+^4 : R_1 + R_2 \geq \frac{1}{2} \log^+ \left[ \frac{(1 - \rho^2)\beta(d_1, d_2)}{2d_1d_2} \right] \right\},
\]
where
\[ \beta(d_1, d_2) = 1 + \sqrt{1 + \frac{4\rho^2d_1d_2}{(1 - \rho^2)^2}}. \]

From previous results we know that

**Theorem 1** (Wagner, Tavildar, and Viswanath [2]). For the quadratic Gaussian two-terminal source coding problem

\[ \mathcal{RD}^* = R_1^* \cap R_2^* \cap R_{12}^*. \]

We will prove the following result.

**Theorem 2.** For the quadratic Gaussian two-terminal source coding problem

\[ \mathcal{RD}_o = R_1^* \cap R_2^* \cap R_{12}^*. \]

### 2.1.2 Proof of Result

Using a standard coding method that combines vector quantization with the binning technique [13], it is trivial to show that [4, 5, 7]

\[ \mathcal{RD}_o \supset R_1^* \cap R_2^* \cap R_{12}^*. \]  \( \text{(2.1)} \)

In the sequel, we focus on proving the second half of Theorem 2. Our proof makes use of the rate region characterization for the Gaussian many-help-one problem, when the auxiliary sources are conditionally independent given the primary source. The setup for this problem is depicted in Figure 2.3. Let \( Y_0, \ldots, Y_L \) be correlated zero mean Gaussian random variables and \( Y_1, \ldots, Y_L \) be conditionally independent given \( Y_0 \). Moreover, let \( \Lambda = \{1, \ldots, L\} \) and

\[ Y_l = a_lY_0 + N_l, \quad l \in \Lambda, \]
Figure 2.3: Communication system with $L$ helpers.

where $a_l$ are nonzero normalizing constants and $N_l$ are zero-mean Gaussian random variables. Finally, let $Y_0, N_1, \ldots, N_L$ be mutually independent. Without loss of generality, the variances of $Y_0, Y_l$ can be set to $\sigma^2_{Y_0} = \sigma^2_{Y_l} = 1$. The decoder estimates $Y_0$ subject to a mean squared error distortion constraint

$$\mathbb{E}[(Y_0 - \hat{Y}_0)^2] \leq d.$$ 

The rate region for this problem was found by Oohama [6].

**Lemma 1.** If $\gamma$ is in $\Gamma_0$, then for all $S \subset \Lambda$,

$$I(Y_0; U_S|W, T) \leq \frac{1}{2} \log \left[ 1 + \sum_{l \in S} \frac{a_l^2}{1 - a_l^2} \left( 1 - 2^{-2I(Y_l; U_l|Y_0, W, T)} \right) \right].$$

**Proof.** For any realization of $(W, T)$, it follows from Lemma 3 in Oohama [6] that

$$2^{2I(Y_0; U_S|W=w, T=t)} \leq 1 + \sum_{l \in S} \frac{a_l^2}{1 - a_l^2} \left( 1 - 2^{-2I(Y_l; U_l|Y_0, W=w, T=t)} \right).$$

The conclusion follows by averaging over $(w, t)$ and invoking the convexity of $\exp_2(.)$ twice, once on each side.

Let us now prove tightness of the W-A outer bound for the Gaussian many-help-one problem with conditional independence. The rate region for this setup is given on the right-hand side of (2.2).
Lemma 2. For the Gaussian many-help-one problem with conditional independence and allowable distortion $d$, the $W$-$A$ outer bound for $X = Y_0$ satisfies

$$\bigcup_{\gamma \in \Gamma_o} \mathcal{RD}_o(Y_0, \gamma) \subset \left\{ (R_0, \ldots, R_L, d) \in \mathbb{R}^{L+2}_+: \text{there exist } r_l \geq 0, \ l \in \Lambda : \right. \left. \begin{array}{c}
R_l \geq r_l \text{ for all } l \in \Lambda, \text{ and } \\
R_0 + \sum_{l \in S} R_l \geq \frac{1}{2} \log^+ \left[ \frac{1}{d} \left( 1 + \sum_{l \in \Lambda - S} \frac{a_l^2}{1 - a_l^2} (1 - 2^{-2r_l}) \right) \right]^{-1} \\
+ \sum_{l \in S} r_l \text{ for all } S \subset \Lambda \right\}. \quad (2.2)$$

Proof. Let $\Lambda_0 = \{0, 1, \ldots, L\}$. If $(R_0, \ldots, R_L, d)$ is in $\bigcup_{\gamma \in \Gamma_o} \mathcal{RD}_o(Y_0, \gamma)$, then there exists $\gamma$ in $\Gamma_o$ such that $\mathbb{E}[(Y_0 - \hat{Y}_0)^2] \leq d$ and for all $A \subset \Lambda_0$

$$\sum_{l \in A} R_l \geq I(Y_0; U_A | U_{\Lambda_0 - A}, T) + \sum_{l \in A} I(Y_l; U_l | Y_0, W, T),$$

which implies that for all $S \subset \Lambda$

$$R_0 + \sum_{l \in S} R_l \geq I(Y_0; U_0, U_S | U_{\Lambda_0 - S}, T) + \sum_{l \in S} I(Y_l; U_l | Y_0, W, T) \quad = [I(Y_0; U | T) - I(Y_0; U_{\Lambda_0 - S} | T)]^+ + \sum_{l \in S} I(Y_l; U_l | Y_0, W, T) \quad = [I(Y_0; U, T) - I(Y_0; U_{\Lambda_0 - S} | T)]^+ + \sum_{l \in S} I(Y_l; U_l | Y_0, W, T). \quad (2.3)$$

Since $Y_0 \leftrightarrow (U, T) \leftrightarrow \hat{Y}_0$, it follows that

$$I(Y_0; U, T) \geq I(Y_0; \hat{Y}_0) \geq \frac{1}{2} \log \frac{\sigma_{\hat{Y}_0}^2}{\mathbb{E}[(Y_0 - \hat{Y}_0)^2]} \geq \frac{1}{2} \log \frac{1}{d}. \quad (2.4)$$
Moreover it holds that

\[ I(Y_0; U_{\Lambda-S}|T) + I(Y_0; W|U_{\Lambda-S}, T) = I(Y_0; W|T) + I(Y_0; U_{\Lambda-S}|W, T) \]

and \( I(Y_0; W|T) = 0 \), so

\[ I(Y_0; U_{\Lambda-S}|T) \leq I(Y_0; U_{\Lambda-S}|W, T). \]

Making use of Lemma 1 and defining \( r_l = I(Y_l; U_l|Y_0, W, T) \), we get that

\[ I(Y_0; U_{\Lambda-S}|T) \leq \frac{1}{2} \log \left[ 1 + \sum_{l \in \Lambda-S} \frac{a_l^2}{1-a_l^2} (1 - 2^{-2r_l}) \right]. \quad (2.5) \]

Moreover, for every \( l \in \Lambda \) the W-A bound requires

\[ R_l \geq I(Y_0; U_l|U_{\Lambda_0-l}, T) + I(Y_l; U_l|Y_0, W, T) \]

\[ \geq I(Y_l; U_l|Y_0, W, T). \]

Summarizing the results above, when \((R_0, \ldots, R_L, d)\) is in \( \bigcup_{\gamma \in \Gamma_o} \mathcal{RD}_o(Y_0, \gamma) \)

\[ R_l \geq r_l \geq 0 \quad (2.6) \]

for \( l \in \Lambda \), and

\[ R_0 + \sum_{l \in S} R_l \geq \frac{1}{2} \log^+ \left[ \frac{1}{d} \left( 1 + \sum_{l \in \Lambda-S} \frac{a_l^2}{1-a_l^2} (1 - 2^{-2r_l}) \right)^{-1} \right] + \sum_{l \in S} r_l, \quad (2.7) \]

for \( S \subset \Lambda \).

We break the proof of Theorem 2 in two parts. First, we prove tightness of the W-A bound for the marginal rate constraints. This is done by reducing Lemma 2 to the case of the one-helps-one problem \((L = 1)\). Second, we prove tightness for the sum rate constraint. Our work in the second part makes use of the techniques introduced by Wagner, Tavildar, and Viswanath [2].
Tightness for the Marginal Rate Constraints

Making use of the symmetry of the problem, we just need to prove the following lemma in order to show tightness for the marginal constraints.

**Lemma 3.**

$$\mathcal{R}D_o \subset R^*_1.$$  

**Proof.** If \((R_1, R_2, d_1, d_2)\) is in \(\mathcal{R}D_o\), then there exists \(\gamma\) in \(\Gamma_o\) such that 

$$\mathbb{E}[(Y_l - \hat{Y}_l)^2] \leq d_l,$$  

\(l = \{1, 2\}\) and for \(A \subset \{1, 2\}\) with \(X = Y_1\)

$$\sum_{l \in A} R_l \geq I(Y_1; U_A | U_{\{1,2\} \setminus A}, T) + \sum_{l \in A} I(Y_l; U_l | Y_1, W, T),$$

which implies that

$$R_1 \geq I(Y_1; U_1 | U_2, T)$$

$$R_2 \geq I(Y_2; U_2 | Y_1, W, T)$$

$$R_1 + R_2 \geq I(Y_1; U_1, U_2 | T) + I(Y_2; U_2 | Y_1, W, T).$$

This set of rates can be weakened to the set

$$R_1 \geq \frac{1}{2} \log^+ \left[ \frac{1}{d_1} \left( 1 + \frac{\rho^2}{1 - \rho^2} (1 - 2^{-2R_2}) \right)^{-1} \right]$$

$$R_1 + R_2 \geq \frac{1}{2} \log^+ \frac{1}{d_1} + r_2,$$

where \(R_2 \geq r_2 \geq 0\).

The above is just an application of Lemma 2 for the case of the Gaussian one-helps-one problem with conditional independence \((L = 1)\). Note that we have switched enc 0 \(\rightarrow\) enc 1 and enc 1 \(\rightarrow\) enc 2 and that we have set \(a_2 = \rho\), so that we match the notation in our original problem. From existing results [6, 7], we know that the latter set of rates coincides with the set

$$\left\{ (R_1, R_2, d_1, d_2) \in \mathbb{R}_+^4 : R_1 \geq \frac{1}{2} \log^+ \left[ \frac{1}{d_1} \left( 1 - \rho^2 + \rho^2 e^{-2R_2} \right) \right] \right\}.$$
Hence, the lemma is proved.

Tightness for the Sum Rate Constraint

Let $\mathcal{D}_G$ be the set of matrices $\mathbf{D}$ such that

$$\mathbf{D}^{-1} = \mathbf{K}_y^{-1} + \mathbf{\Delta}^{-1}$$

for some diagonal and positive definite matrix $\mathbf{\Delta}$. We call $\mathcal{D}_G$ the set of distributed Gaussian distortion matrices. Moreover, let $\text{diag}(\mathcal{D}_G)$ denote the set of distortion pairs $(d_1, d_2)$ such that there exists a distributed Gaussian distortion matrix $\mathbf{D}$ with top-left entry $d_1$ and bottom-right entry $d_2$. Following the original proof of Wagner, Tavildar, and Viswanath [2], we separately examine the cases $(d_1, d_2) \in \text{diag}(\mathcal{D}_G)$ and $(d_1, d_2) \notin \text{diag}(\mathcal{D}_G)$.

Case that $(d_1, d_2) \notin \text{diag}(\mathcal{D}_G)$: We have already proved that the W-A outer bound is tight for the marginal rate constraints. The latter result is sufficient for also proving tightness of the W-A outer bound for the sum rate constraint when $(d_1, d_2) \notin \text{diag}(\mathcal{D}_G)$. The following lemma holds.

**Lemma 4.** For $(d_1, d_2) \notin \text{diag}(\mathcal{D}_G)$,

$$\mathcal{RD}_o \subset \mathcal{R}_{12}^*.$$

**Proof.** We focus on the case that $\min(d_1, d_2) < 1$ and consider $d_1 = \min(d_1, d_2)$. The proof for $\min(d_1, d_2) \geq 1$ is trivial. Since $(d_1, d_2) \notin \text{diag}(\mathcal{D}_G)$, it follows that

$$\rho^2 d_1 + 1 - \rho^2 \leq d_2.$$

Making use of the inequality above and defining distortion $\tilde{d}_2 = \rho^2 d_1 + 1 - \rho^2$, we get

$$\tilde{R}_{12}^* \subset \mathcal{R}_{12}^*.$$
where
\[
\tilde R_{12}^* = \left\{ (R_1, R_2, d_1, \tilde d_2) \in \mathbb{R}_+^4 : R_1 + R_2 \geq \frac{1}{2} \log^+ \left[ \frac{(1 - \rho^2)\beta(d_1, \tilde d_2)}{2d_1 \tilde d_2} \right] \right\}.
\]

It can be verified that
\[
\tilde R^*_{12} = \left\{ (R_1, R_2, d_1, \tilde d_2) \in \mathbb{R}_+^4 : R_1 + R_2 \geq \frac{1}{2} \log \frac{1}{d_1} \right\}.
\]

On the other hand, if \((R_1, R_2, d_1, d_2)\) \(\in R_1^*\) then \(R_1\) and \(R_2\) must satisfy
\[
R_1 + R_2 \geq \frac{1}{2} \log \frac{1}{d_1}.
\]

Hence, it follows that
\[
R_1^* \subset \tilde R^*_{12},
\]

which by using Lemma 3 gives
\[
\mathcal{R} \mathcal{D}_o \subset R^*_1.
\]

\[\square\]

Case that \((d_1, d_2) \in \text{diag}(\mathcal{D}_G)\): Let \(D^*\) denote the element of \(\mathcal{D}_G\) whose top-left and bottom-right elements are \(d_1\) and \(d_2\), respectively [2].

**Definition 5.** For \(\theta \in (-1, 1)\), let
\[
D_\theta = \begin{bmatrix}
d_1 & \theta \sqrt{d_1 d_2} \\
\theta \sqrt{d_1 d_2} & d_2
\end{bmatrix},
\]

and define
\[
R_{coop}(\theta) = \frac{1}{2} \log^+ \frac{|K_y|}{|D_\theta|} = \frac{1}{2} \log^+ \frac{1 - \rho^2}{(1 - \theta^2)d_1 d_2}.
\]

Let \(\mu = (\mu_1, \mu_2)\) denote a vector with \(\mu_1 \cdot \mu_2 > 0\) such that [2, Lemma 4]
\[
\frac{1}{2} \log \frac{|K_y|}{|D^*|} = \inf \left\{ \frac{1}{2} \log \frac{|K_y|}{|D|} : D \in \mathcal{D}_G \text{ and } \mu^T D \mu \leq \mu^T D^* \mu \right\}
\]
and
\[ \mu_1 \mu_2 = \frac{\lambda^2}{\rho} \left( \frac{\rho}{1 - \rho^2} \right)^2 \left( \frac{\mu_1}{\mu_2} + \rho \right) \left( \frac{\mu_2}{\mu_1} + \rho \right), \]
with
\[ \lambda^{-1} = 1 + \frac{\rho}{1 - \rho^2} \left( \frac{\mu_1}{\mu_2} + \frac{\mu_2}{\mu_1} + 2 \rho \right). \]
Then define
\[ R_{\text{sum}}(\theta) = \inf \left\{ \frac{1}{2} \log \frac{|K_y|}{|D|} : D \in \mathcal{D}_G \text{ and } \mu^T D \mu \leq \mu^T D_\theta \mu \right\}. \]

Lemma 5. For \((d_1, d_2) \in \text{diag}(\mathcal{D}_G),\)
\[ \mathcal{R}D_o \subset R^{*}. \]

Proof. Let
\[ a_1 = \left( \frac{1 - \rho^2}{\rho \left( \frac{\mu_1}{\mu_2} + \rho \right)} + 1 \right)^{-\frac{1}{2}}, \]
\[ a_2 = \left( \frac{1 - \rho^2}{\rho \left( \frac{\mu_2}{\mu_1} + \rho \right)} + 1 \right)^{-\frac{1}{2}}. \]
One can verify that \(a_1, a_2\) are contained in \((\rho, 1)\) and \(a_1 a_2 = \rho\). Let \(Y_1, Y_2\) be of the form
\[ Y_1 = a_1 Y_0 + N_1 \]
\[ Y_2 = a_2 Y_0 + N_2, \]
where \(Y_0, N_1, N_2\) are independent Gaussian random variables. Assume that \(N_1\) and \(N_2\) have positive variances and \(\sigma_{Y_0}^2 = 1.\)

Since \(Y_1 \perp Y_2|Y_0, \) it follows that \(Y_0 \in \chi.\) So if \((R_1, R_2, d_1, d_2)\) is in \(\mathcal{R}D_o,\) then there exists \(\gamma\) in \(\Gamma_o\) such that \(\mathbb{E}[(Y_l - \hat{Y}_l)^2] \leq d_l\) for \(l = \{1, 2\}\) and
\[ R_1 + R_2 \geq I(Y_0; U_1, U_2|T) + I(Y_1; U_1|Y_0, W, T) + I(Y_2; U_2|Y_0, W, T) \]
\[ = I(Y_0; U_1, U_2|T) + I(Y_1, Y_2; U_1, U_2|Y_0, W, T). \quad (2.8) \]
But

\[
I(Y_1, Y_2; W|Y_0, T) + I(Y_1, Y_2; U_1, U_2|Y_0, W, T)
= I(Y_1, Y_2; U_1, U_2|Y_0, T) + I(Y_1, Y_2; W|U_1, U_2, Y_0, T)
\]

and the first term on the left-hand side is zero. Thus

\[
I(Y_1, Y_2; U_1, U_2|Y_0, W, T) \geq I(Y_1, Y_2; U_1, U_2|Y_0, T).
\]

Substituting this result in (2.8), we get

\[
R_1 + R_2 \geq I(Y_0; U_1, U_2|T) + I(Y_1, Y_2; U_1, U_2|Y_0, T)
= I(Y_0, Y_1, Y_2; U_1, U_2|T)
= I(Y_1, Y_2; U_1, U_2|T) + I(Y_0; U_1, U_2|Y_1, Y_2, T)
= I(Y_1, Y_2; U_1, U_2|T)
= h(Y_1, Y_2|T) - h(Y_1, Y_2|U_1, U_2, T)
= \frac{1}{2} \log [(2\pi e)^2 |K_y|] - h(Y_1, Y_2|U_1, U_2, T).
\]

(2.9)

In addition

\[
h(Y_1, Y_2|U_1, U_2, T) = h(Y_1 - \hat{Y}_1, Y_2 - \hat{Y}_2|U_1, U_2, T)
\leq h(Y_1 - \hat{Y}_1, Y_2 - \hat{Y}_2).
\]

(2.10)

Let \(\hat{D}\) denote the distortion matrix

\[
\hat{D} = \mathbb{E} \left[ (Y - \hat{Y})(Y - \hat{Y})^T \right],
\]

where \(Y = (Y_1, Y_2)\). We may assume that \(\hat{Y}_1, \hat{Y}_2\) are MMSE estimators, i.e.

\[
\hat{Y}_1 = \mathbb{E}[Y_1|U_1, U_2, T]
\]
\[
\hat{Y}_2 = \mathbb{E}[Y_2|U_1, U_2, T],
\]
in which case [8, Theorem 9.6.5] implies that
\[ h(Y_1 - \hat{Y}_1, Y_2 - \hat{Y}_2) \leq \frac{1}{2} \log \left( (2\pi e)^2 |\tilde{D}| \right). \]

However, \( h(Y_1, Y_2|U_1, U_2, T) > -\infty \) by (2.9), so \( \tilde{D} \) must be nonsingular and positive definite. Let us write
\[ \tilde{D} = \begin{bmatrix} \tilde{d}_1 & \tilde{\theta} \sqrt{\tilde{d}_1 \tilde{d}_2} \\ \tilde{\theta} \sqrt{\tilde{d}_1 \tilde{d}_2} & \tilde{d}_2 \end{bmatrix}, \]
where \( \tilde{d}_1 \leq d_1, \tilde{d}_2 \leq d_2 \) and \( \tilde{\theta} \in (-1, 1) \). Now define
\[ \theta = \frac{\tilde{\theta} \sqrt{\tilde{d}_1 \tilde{d}_2}}{\sqrt{d_1 d_2}}. \]

Then \( \tilde{D} \preceq D_\theta \) and \( |\tilde{D}| \leq |D_\theta| \), which implies
\[ h(Y_1 - \hat{Y}_1, Y_2 - \hat{Y}_2) \leq \frac{1}{2} \log \left( (2\pi e)^2 |D_\theta| \right). \] (2.11)

So by substituting (2.10) and (2.11) in (2.9), we get
\[ R_1 + R_2 \geq \frac{1}{2} \log \frac{|K_y|}{|D_\theta|} \]
\[ = R_{coop}(\theta). \] (2.12)

Next observe that
\[ \mathbb{E} \left[ \left( \mu^T Y - \mu^T \hat{Y} \right)^2 \right] = \mu^T \tilde{D} \mu, \]
and in particular
\[ \mathbb{E} \left[ \left( \mu^T Y - \mu^T \hat{Y} \right)^2 \right] \leq \mu^T D_\theta \mu, \]
so the distortion for estimating \( \mu^T Y \) is at most \( \mu^T D_\theta \mu \). Then define
\[ d = \mu^T D_\theta \mu \]
and let
\[ \hat{Y}_0 = \mathbb{E}[Y_0|U_1, U_2, T]. \]
A standard calculation shows that
\[ E[Y_0|Y] = \mu^T Y \]
and \( Y_0 \) can be written as
\[ Y_0 = \mu^T Y + \tilde{N}, \]
where \( \tilde{N} \) is a zero mean Gaussian random variable with variance \( \lambda \) and is independent of \( Y \). So
\[
\mathbb{E} \left[ \left( Y_0 - \hat{Y}_0 \right)^2 \right] = \mathbb{E} \left[ \left( Y_0 - \mathbb{E}[Y_0|U_1, U_2, T] \right)^2 \right] \\
= \mathbb{E} \left[ \left( Y_0 - \mu^T Y + \mu^T Y - \mathbb{E}[\mu^T Y|U_1, U_2, T] \right)^2 \right] \\
= \mathbb{E} \left[ \left( \tilde{N} + \mu^T Y - \mathbb{E}[\mu^T Y|U_1, U_2, T] \right)^2 \right] \\
= \lambda + \mathbb{E} \left[ \left( \mu^T Y - \mathbb{E}[\mu^T Y|U_1, U_2, T] \right)^2 \right] \\
\leq \lambda + d.
\]
The above imply that for the already chosen \( \gamma = (U_1, U_2, \hat{Y}_1, \hat{Y}_2, W, T) \) the rate-distortion vector \((R_1, R_2, d)\) satisfies
\[
\sum_{l \in A} R_l \geq I(Y_0; U_A|U_{\{1,2\} - A}, T) + \sum_{l \in A} I(Y_l; U_l|Y_0, W, T)
\]
for all \( A \subset \{1, 2\} \) and \( \lambda + d \geq \mathbb{E}[\left( Y_0 - \hat{Y}_0 \right)^2] \). Let us now construct a
\[
\gamma' = (U_0, U_1, U_2, \hat{Y}_0, W, T)
\]
with \( U_1, U_2, W, T \) the same as in \( \gamma \) and \( U_0 = \text{const} \). The random variable \( \hat{Y}_0 \) is completely determined by the specified \( U_1, U_2, T \). It can be easily verified that \( \gamma' \in \Gamma_o \) for the many-help-one problem with conditional independence. Furthermore,
the rate-distortion vector \((0, R_1, R_2, \lambda + d)\) is contained in

\[
\mathcal{R}D_o^{mho}(Y_0, \gamma') = \left\{ (R, \lambda + d) : \sum_{l \in A} R_l \geq I(Y_0; U_A|U_{\Lambda_0-A}, T) + \sum_{l \in A} I(Y_l; U_l|Y_0, W, T) \right\}
\]

for all \(A \subset \Lambda_0\) and \(\lambda + d \geq \mathbb{E} \left[ \left(Y_0 - \hat{Y}_0\right)^2 \right]\),

where \(\mathcal{R}D_o^{mho}(-, -)\) is the outer bound to the many-help-one problem with conditional independence and two helpers \((L = 2)\). Thus, applying Lemma 2, the sum rate constraint can be weakened to the set

\[
R_1 + R_2 \geq \inf \left\{ \frac{1}{2} \log \frac{1}{\lambda + d} + r_1 + r_2 : r_1, r_2 \geq 0 \quad \text{and} \quad \frac{1}{\lambda + d} \leq 1 + \sum_{l=1}^{2} \frac{a_l^2}{1 - a_l^2} (1 - 2^{-2r_l}) \right\},
\]

or equivalently to [2, Lemma 8]

\[
R_1 + R_2 \geq \inf \left\{ \frac{1}{2} \log \frac{|K_y|}{|D|} : D \in \mathcal{D}_G \text{ and } \mu^T D \mu \leq d \right\}.
\]

Hence, the minimum sum rate will be given by

\[
R_1 + R_2 \geq \inf_{\theta \in (-1,1)} \max \left( R_{coop}(\theta), R_{sum}(\theta) \right).
\]

Combining (2.13) with (2.12) gives

\[
R_1 + R_2 \geq \max \left( R_{coop}(\theta), R_{sum}(\theta) \right),
\]

and by taking the infimum over \(\theta\) in \((-1,1)\)

\[
R_1 + R_2 \geq \inf_{\theta \in (-1,1)} \max \left( R_{coop}(\theta), R_{sum}(\theta) \right).
\]
The latter result and [2, Lemma 6] imply that if \((R_1, R_2, d_1, d_2)\) is in \(\mathcal{RD}_o\), then the sum rate satisfies
\[
R_1 + R_2 \geq \frac{1}{2} \log^+ \left[ \frac{(1 - \rho^2)\beta(d_1, d_2)}{2d_1d_2} \right].
\]

Hence, it follows that
\[
\mathcal{RD}_o \subset R^*_12.
\]

\[
\mathcal{RD}_o = R^*_1 \cap R^*_2 \cap R^*_12.
\]

Proof of Theorem 2. Lemmas 3, 4 and 5 combined with (2.1) and Theorem 1, yield

2.2 The Worst Choice of X for the Sum Rate Constraint

Wagner and Anantharam proved that the W-A outer bound (Def. 2) improves upon the B-T outer bound (Def. 3) in two ways [1]. The first is that \(\mathcal{RD}_o\) allows for optimization over \(X\) while \(\mathcal{RD}_o^{BT}\) effectively requires the choice of \(X = Y\). The second is that \(\Gamma_o\) is “smaller” than \(\Gamma_o^{BT}\) in the sense that if \((U, \hat{Y}, W, T)\) is in \(\Gamma_o\) then \((U, \hat{Y}, T)\) is in \(\Gamma_o^{BT}\). A natural question that arises from the definition of the W-A outer bound is whether there exists a good or bad choice for the auxiliary random variable X. Is it possible to know beforehand which choices of X yield optimum performance of the bound? In this part of the thesis we try to shed some light to this question, by finding the choice of X that forces the W-A sum rate constraint to its worst possible performance. It will turn out that the B-T sum rate constraint is the worst performance that the W-A sum rate constraint can have.
Theorem 3. For any instance,

\[
\min_{X \in \chi} \min_{\gamma \in \Gamma_o} I(X; U|T) + \sum_{l \in \Lambda} I(Y_l; U_l|X, W, T) = \min_{\gamma \in \Gamma_o} I(Y; U|T),
\]

where minimum over \(X \in \chi\) is achieved for \(X = Y\).

Proof. Observe that

\[
I(X; U|T) + \sum_{l \in \Lambda} I(Y_l; U_l|X, W, T) = I(X; U|T) + I(Y; U|X, W, T). \quad (2.14)
\]

In addition,

\[
I(Y; U|X, T) + I(Y; W|U, X, T) = I(Y; W|X, T) + I(Y; U|W, X, T) \quad (2.15)
\]

and the first term on the right-hand side is zero. Thus

\[
I(Y; U|W, X, T) \geq I(Y; U|X, T).
\]

Substituting this in (2.14), we get

\[
I(X; U|T) + \sum_{l \in \Lambda} I(Y_l; U_l|X, W, T) \geq I(X; U|T) + I(Y; U|X, T)
\]

\[
= I(X, Y; U|T)
\]

\[
= I(Y; U|T) + I(X; U|Y, T)
\]

\[
= I(Y; U|T),
\]

where the last equality follows from the Markov coupling. Minimizing both sides over all \(X \in \chi\) and \(\gamma \in \Gamma_o\) yields

\[
\min_{X \in \chi} \min_{\gamma \in \Gamma_o} I(X; U|T) + \sum_{l \in \Lambda} I(Y_l; U_l|X, W, T) \geq \min_{\gamma \in \Gamma_o} I(Y; U|T). \quad (2.16)
\]

Noting that equality in (2.16) is satisfied when choosing \(X = Y\), completes the proof of the theorem. \(\square\)
Chapter 3

Results on Achievable Gaussian Error Exponents

3.1 The Point-to-Point Lossy Source Coding Problem

The first major contribution on error exponents for lossy compression systems was made by Marton, who derived an exponent for the finite alphabet point-to-point lossy source coding problem (Figure 3.1) [9]. As a byproduct of their work on guessing exponents, Arikan and Merhav extended this result in the case of a Gaussian source [10]. Two years later, their error exponent was verified by Ihara and Kubo using a different approach [11]. To the best of our knowledge, these two papers constitute the only work so far on error exponents for Gaussian distributed sources. None of them, however, seems readily extendable to multiterminal source

Figure 3.1: The point-to-point source coding problem.
coding problems. On one hand, guessing exponents is not a direct way of dealing with lossy source coding exponents. On the other hand, Ihara and Kubo were unaware of the method of types for Gaussian distributions, introduced in [10, Theorem 5], and did not use it in their proof.

In this chapter, we focus on the point-to-point Gaussian source coding problem, combine the existing techniques and obtain the best achievable error exponent. Once the Gaussian type-classes are introduced, our system can be viewed as a direct extension of its discrete analog. This simplifies the problem and enables us to build a proof closely resembling Marton’s work [9]. Most results on source coding error exponents are reported for discrete alphabet sources. The Gaussian method of types couples a Gaussian problem to its discrete equivalent. It, therefore, gives the chance to use some of the existing proofs on discrete systems, for obtaining new results on Gaussian source coding exponents.

3.1.1 Problem Statement and Result

Let $Y$ be a zero-mean Gaussian memoryless source with variance $\sigma_Y^2$. We denote its probability distribution by $P_Y$ and write $n$ independent copies of $Y$ as

$$ Y^n \equiv (Y_1, Y_2, \ldots, Y_n). $$

We use $y$ to denote the sequence $(y_1, y_2, \ldots, y_n)$. The point-to-point lossy source coding problem for a continuous alphabet source can be stated as follows. The encoder observes $Y^n$ and then sends a message to the decoder using the map

$$ f^{(n)} : \mathbb{R}^n \to \{1, 2, \ldots, M^{(n)}\}. $$

The decoder uses the received message to estimate $Y^n$ according to the map

$$ \phi^{(n)} : \{1, 2, \ldots, M^{(n)}\} \to \mathbb{R}^n. $$
Definition 6. A rate-distortion vector \((R, d)\) is achievable if there exists a block length \(n\), encoder \(f^{(n)}\) and a decoder \(\phi^{(n)}\) such that

\[
R \geq \frac{1}{n} \log M^{(n)}
\]

and

\[
d \geq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(y_i - \hat{y}_i)^2],
\]

where

\[
\hat{y} = \phi^{(n)}(f^{(n)}(y)).
\]

Let \(\mathcal{RD}^*\) denote the set of achievable rate-distortion vectors. Let

\[
\mathcal{R}^*(d) = \left\{ R : (R, d) \in \overline{\mathcal{RD}^*} \right\},
\]

where \(\overline{\mathcal{RD}^*}\) denotes the closure of \(\mathcal{RD}^*\). We call \(\mathcal{R}^*(\cdot)\) the rate region for the problem.

We note that there is no loss of generality in assuming that \(\sigma_Y^2 = 1\), since the observations and the estimates can be scaled to reduce the general case to this one. We may, therefore, assume that \(P_Y \sim \mathcal{N}(0, 1)\).

For a given distortion \(d > 0\), we are interested in the error event

\[
\mathcal{E}(f^{(n)}, \phi^{(n)}) = \left\{ y \in \mathcal{Y}^n : \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 > d \right\}.
\]

Our purpose is to examine the asymptotic behavior of the error probability \(\Pr(\mathcal{E}(f^{(n)}, \phi^{(n)}))\) as the blocklength \(n\) becomes large. Let us now define the error exponent

\[
\theta(R, d, P_Y) = \liminf_{n \to \infty} -\frac{1}{n} \log \left[ \min_{f^{(n)}, \phi^{(n)}} \Pr(\mathcal{E}(f^{(n)}, \phi^{(n)})) \right],
\]

\(^1\)All logarithms in this chapter are base \(e\).
where the minimization ranges over all possible \((f^{(n)}, \phi^{(n)})\) satisfying (3.1).

Consider an auxiliary random variable \(\bar{Y}\) taking values in \(\mathcal{Y}\) and let \(Q_{\bar{Y}}\) be its probability distribution. We denote by \(D(\bar{Y} \| Y)\) the Kullback-Leibler distance between \(Q_{\bar{Y}}\) and \(P_Y\) and by \(R(Y, d)\) the rate-distortion function of the source \(P_Y\). Under the current setup,

\[
R(Y, d) = \begin{cases} 
\frac{1}{2} \log \frac{1}{d} & \text{if } 0 < d \leq 1 \\
0 & \text{if } d > 1
\end{cases}
\]

We are now in position to state our main result, an instance of which is illustrated in Figure 3.2.

**Theorem 4.** For any \(R > 0\), distortion \(d > 0\) and distribution \(P_Y \sim \mathcal{N}(0, 1)\)

\[
\theta(R, d, P_Y) \geq \begin{cases} 
\frac{1}{2} \left(d e^{2R} - \log d - 2R - 1\right) & \text{if } R \geq R(Y, d) \\
0 & \text{if } R < R(Y, d)
\end{cases}
\]

### 3.1.2 The Gaussian Method of Types and Other Lemmas

Our work heavily relies on the concepts of Gaussian types that were first introduced by Arikan and Merhav [10]. In this section, we extend their definitions and provide some useful lemmas.

**Definition 7.** For a given \(0 < \epsilon < 1\) and \(\sigma_Y^2 > 0\), a Gaussian type-class \(T_{\sigma_Y^2}^\epsilon\) is defined as the set of \(n\)-sequences

\[
T_{\sigma_Y^2}^\epsilon = \{y \in \mathcal{Y}^n : |y' y - n \sigma_Y^2| \leq n \epsilon \sigma_Y^2\}.
\]

Similarly, for a given \(0 < \epsilon < 1\) and covariance matrix

\[
K = \begin{bmatrix} 
\sigma_Y^2 & \rho \sigma_Y \sigma_{Y_2} \\
\rho \sigma_Y \sigma_{Y_2} & \sigma_{Y_2}^2
\end{bmatrix},
\]
with non-zero variances, a joint Gaussian type-class \( T_K^\epsilon \) is defined as the set of pairs of \( n \)-sequences

\[
T_K^\epsilon = \{ (y_1, y_2) \in Y_1^n \times Y_2^n : \quad |y_1^i y_1 - n\sigma_{Y_1}^2| \leq n\epsilon\sigma_{Y_1}^2 \}
\]

Furthermore, for a given \( y_1 \in T_{\sigma_{Y_1}^2}^\epsilon \), we define the conditional Gaussian type-class \( T_W^\epsilon (y_1) \) as the set of \( n \)-sequences

\[
T_W^\epsilon (y_1) = \{ y_2 \in Y_2^n : (y_1, y_2) \in T_K^\epsilon \},
\]

where \( W : Y_1 \rightarrow Y_2 \) denotes the conditional distribution (or “channel”) \( P_{Y_2|Y_1} \) induced by \( P_{Y_1Y_2} \sim \mathcal{N}(0, K) \).
From the definition above it is clear that, if \((y_1, y_2) \in T^\epsilon_K\) then \(y_1 \in T^\epsilon_{\sigma^2_{Y_1}}\) and \(y_2 \in T^\epsilon_{\sigma^2_{Y_2}}\). In what follows, we assume that \(\rho \geq 0\), i.e. that \(Y_1\) and \(Y_2\) are non-negatively correlated.

**Lemma 6.** For a given \(0 < \epsilon < 1\), covariance matrix \(K\) as in Definition 7 and \(y_1 \in T^\epsilon_{\sigma^2_{Y_1}}\), the conditional Gaussian type-class satisfies

\[
T^\epsilon_W(y_1) \subset \left\{ y_2 \in \mathcal{Y}^n_2 : y_2 = \left( \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right) y_1 + v \text{ with } |v^t y_1| \leq 2n\epsilon \rho \sigma_{Y_1} \sigma_{Y_2} \right.
\]

\[
\text{and } |v^t v - n \sigma^2_{Y_2} (1 - \rho^2)| \leq n \epsilon \sigma^2_{Y_2} (1 + 3 \rho^2) \right\}.
\]

**Proof.** Consider the Gaussian channel \(W : \mathcal{Y}_1 \rightarrow \mathcal{Y}_2\) with

\[
Y_2 = aY_1 + V,
\]

where \(R_{Y_1} \sim \mathcal{N}(0, \sigma^2_{Y_1})\), \(R_V \sim \mathcal{N}(0, \sigma^2_{Y_2})\) and \(V \perp Y_1\). Trivially

\[
\mathbb{E}(Y_2^2) = \mathbb{E}(a^2 Y_1^2 + 2aY_1V + V^2)
\]

\[
= a^2 \mathbb{E}(Y_1^2) + \mathbb{E}(V^2),
\]

which gives \(\sigma^2_{Y_2} = \sigma^2_{Y_2} - a^2 \sigma^2_{Y_1}\) and

\[
\mathbb{E}(Y_1Y_2) = \mathbb{E}(aY_1^2 + Y_1V)
\]

\[
= a \mathbb{E}(Y_1^2)
\]

that yields \(a = \frac{\sigma_{Y_2}}{\sigma_{Y_1}}\). Hence, the Gaussian channel takes the form

\[
Y_2 = \left( \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right) Y_1 + V,
\]

where \(R_{Y_1} \sim \mathcal{N}(0, \sigma^2_{Y_1})\), \(R_V \sim \mathcal{N}(0, \sigma^2_{Y_2}(1 - \rho^2))\) and \(V \perp Y_1\). From the definition
of the conditional Gaussian type-class, we have

\[ v^t y_1 = y^t_2 y_1 - \left( \rho \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right) y^t_1 y_1 \]

\[ \leq n \rho \sigma_{Y_1} \sigma_{Y_2} (1 + \epsilon) - \left( \rho \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right) n \sigma_{Y_1}^2 (1 - \epsilon) \]

\[ = 2 n \epsilon \rho \sigma_{Y_1} \sigma_{Y_2}. \]

In the same manner, one can prove a lower bound for \( v^t y_1 \). Consequently

\[ |v^t y_1| \leq 2 n \epsilon \rho \sigma_{Y_1} \sigma_{Y_2}. \] (3.5)

In addition, it holds that

\[ v^t v = y^t_2 y_2 + \left( \rho \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right)^2 y^t_1 y_1 - \left( \rho \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right) (y^t_1 y_2 + y^t_2 y_1) \]

\[ \leq n \sigma_{Y_2}^2 (1 + \epsilon) + \left( \rho \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right)^2 n \sigma_{Y_1}^2 (1 + \epsilon) - \left( \rho \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \right) 2 n \rho \sigma_{Y_1} \sigma_{Y_2} (1 - \epsilon) \]

\[ = n \sigma_{Y_2}^2 (1 - \rho^2) + n \epsilon \sigma_{Y_2}^2 (1 + 3 \rho^2). \]

A lower bound can be proved by a small modification of the above, yielding

\[ |v^t v - n \sigma_{Y_2}^2 (1 - \rho^2)| \leq n \epsilon \sigma_{Y_2}^2 (1 + 3 \rho^2), \] (3.6)

which completes the proof of the lemma. \( \square \)

**Lemma 7.** For two Gaussian memoryless sources \( P_Y \sim \mathcal{N}(0, \sigma_Y^2) \) and \( Q_Y \sim \mathcal{N}(\mu, \bar{\sigma}_Y^2) \), their Kullback-Leibler distance is given by

\[ D(\bar{Y} \| Y) = \frac{1}{2} \left( \frac{\mu^2 + \bar{\sigma}_Y^2}{\sigma_Y^2} - \log \frac{\bar{\sigma}_Y^2}{\sigma_Y^2} - 1 \right). \]
Proof. From the definition of relative entropy

\[
D(\bar{Y} \parallel Y) = \int Q_{\bar{Y}} \log \frac{Q_{\bar{Y}}}{P_Y} dy
\]

\[
= -h(Q_{\bar{Y}}) + \int Q_{\bar{Y}} \log (2\pi \sigma_{\bar{Y}}^2)^{\frac{1}{2}} + \frac{1}{2} \log(2\pi \sigma_{\bar{Y}}^2) dy
\]

\[
= -h(Q_{\bar{Y}}) + \frac{1}{2} \log(2\pi \sigma_{\bar{Y}}^2) + \frac{1}{2} \log(2\pi \sigma_{\bar{Y}}^2) + \frac{1}{2} \log(\text{Var}[Y])
\]

\[
= \frac{1}{2} \left( \frac{\mu^2 + \sigma_{\bar{Y}}^2}{\sigma_Y^2} - \log \frac{\sigma_{\bar{Y}}^2}{\sigma_Y^2} - 1 \right).
\]

We now present some useful lemmas on the properties of Gaussian type-classes.

**Lemma 8.** For any \( n, 0 < \epsilon < 1, \) and distribution \( P_Y \sim \mathcal{N}(0, \sigma_Y^2) \) on \( \mathcal{Y} \)

\[
\left(1 - \frac{2}{n\epsilon^2}\right) e^{n(h(Y) - \frac{\epsilon^2}{2})} \leq \text{Vol}(T_{\sigma_Y^2}^n) \leq e^{n(h(Y) + \frac{\epsilon^2}{2})}.
\]

Moreover, for any \( \sigma_{\bar{Y}}^2 > 0 \) and \( y \in T_{\sigma_{\bar{Y}}^2}^n \)

\[
e^{-n\left(D(\bar{Y} \parallel Y) + h(Y) + \frac{\sigma_{\bar{Y}}^2}{2\sigma_Y^2}\right)} \leq P_Y^n(y) \leq e^{-n\left(D(\bar{Y} \parallel Y) + h(Y) - \frac{\sigma_{\bar{Y}}^2}{2\sigma_Y^2}\right)},
\]

where the distribution \( Q_{\bar{Y}} \sim \mathcal{N}(0, \sigma_{\bar{Y}}^2) \).

Proof. Consider a zero-mean Gaussian memoryless source with variance \( \sigma_Y^2 \). Then

\[
1 \geq \int_{T_{\sigma_Y^2}^n} (2\pi \sigma_{\bar{Y}}^2)^{-\frac{n}{2}} \exp \left\{ -\frac{y^t y}{2\sigma_Y^2} \right\} dy
\]

\[
\geq \int_{T_{\sigma_Y^2}^n} (2\pi \sigma_{\bar{Y}}^2)^{-\frac{n}{2}} \exp \left\{ -\frac{n(1 + \epsilon)}{2} \right\} dy
\]

\[
= \text{Vol}(T_{\sigma_Y^2}^n) \exp \left\{ -n \left( h(Y) + \frac{\epsilon}{2} \right) \right\},
\]

which obviously gives

\[
\text{Vol}(T_{\sigma_Y^2}^n) \leq e^{n(h(Y) + \frac{\epsilon^2}{2})}.
\]

(3.7)
On the other hand, using Chebyshev’s inequality

\[
1 - \Pr \left( \frac{\epsilon^2 Y - n\sigma_Y^2}{\sigma_Y^2} > n\epsilon \sigma_Y^2 \right) = \Pr \left( |\epsilon^2 y - n\sigma_Y^2| > n\epsilon \sigma_Y^2 \right)
\leq \frac{\mathbb{E}[(\epsilon^2 y - n\sigma_Y^2)^2]}{n^2 \epsilon^2 \sigma_Y^4}
= \frac{2n\sigma_Y^4}{n^2 \epsilon^2 \sigma_Y^4}
= \frac{2}{n\epsilon^2}.
\]

Moreover,

\[
\Pr \left( \frac{\epsilon^2 Y - n\sigma_Y^2}{\sigma_Y^2} \right) = \int_{\frac{\epsilon^2 Y - n\sigma_Y^2}{\sigma_Y^2}}^{\infty} \left(2\pi\sigma_Y^2\right)^{-\frac{n}{2}} \exp \left\{-\frac{y^2}{2\sigma_Y^2}\right\} dy
\leq \int_{\frac{\epsilon^2 Y - n\sigma_Y^2}{\sigma_Y^2}}^{\infty} \left(2\pi\sigma_Y^2\right)^{-\frac{n}{2}} \exp \left\{-\frac{n(1 - \epsilon)}{2}\right\} dy
= \text{Vol} \left( \frac{\epsilon^2 Y - n\sigma_Y^2}{\sigma_Y^2} \right) \exp \left\{-n \left(h(Y) - \frac{\epsilon^2}{2}\right)\right\}.
\]

Combining the last two relations, we get

\[
\text{Vol} \left( \frac{\epsilon^2 Y - n\sigma_Y^2}{\sigma_Y^2} \right) \geq \left(1 - \frac{2}{n\epsilon^2}\right) e^{n(h(Y) - \frac{\epsilon^2}{2})}.
\]

(3.8)

Let us now consider an arbitrary \( y \in T_{\sigma_Y^2}^e \) and a zero-mean Gaussian memoryless source with variance \( \sigma_Y^2 \). It then holds that

\[
P_Y^n = \left(2\pi\sigma_Y^2\right)^{-\frac{n}{2}} \exp \left\{-\frac{y^2}{2\sigma_Y^2}\right\},
\]

which can be lower bounded as follows

\[
P_Y^n \geq \left(2\pi\sigma_Y^2\right)^{-\frac{n}{2}} \exp \left\{-\frac{n\sigma_Y^2(1 + \epsilon)}{2\sigma_Y^2}\right\}
= \exp \left\{-n \left[\frac{1}{2} \log(2\pi\sigma_Y^2) + (1 + \epsilon) \frac{\bar{\sigma}_Y^2}{2\sigma_Y^2}\right]\right\}
= \exp \left\{-n \left[\frac{1}{2} \left(\frac{\bar{\sigma}_Y^2}{\sigma_Y^2} - \log \frac{\bar{\sigma}_Y^2}{\sigma_Y^2} - 1\right) + \frac{1}{2} \log(2\pi e\sigma_Y^2) + \epsilon \frac{\sigma_Y^2}{2\sigma_Y^2}\right]\right\}
= \exp \left\{-n \left(D(\bar{Y}\|Y) + h(\bar{Y}) + \frac{\epsilon}{2} \frac{\sigma_Y^2}{2\sigma_Y^2}\right)\right\}.
\]

(3.9)
The latter equality is a direct application of Lemma 7 for a distribution \( Q_{\bar{Y}} \sim \mathcal{N}(0, \bar{\sigma}^2_Y) \). An upper bound of \( P_{Y}^n \) can be derived by a small modification of the above, hence, completing the proof of the lemma. \( \square \)

As an immediate consequence, the probability of a Gaussian type-class can be upper bounded by

\[
P_Y^n \left(T^{\epsilon}_{\bar{\sigma}^2_Y}\right) \leq e^{-n \left(D(\bar{Y}||Y) - \frac{\epsilon^2}{2\bar{\sigma}^2_Y} + 1\right)}.
\]

In a similar manner, we can extend Lemma 8 to the case of joint Gaussian type-classes. In the sequel, we use \( o_\epsilon(1) \) to denote a quantity \( g(\epsilon) > 0 \) with

\[
\lim_{\epsilon \to 0} g(\epsilon) = 0.
\]

**Lemma 9.** For any \( n, 0 < \epsilon < 1 \), and distribution \( P_{Y_1Y_2} \sim \mathcal{N}(0, K) \) on \( Y_1 \times Y_2 \)

\[
\left(1 - \frac{5}{nc^2} - \frac{1}{nc^2\rho^2}\right) e^{n(h(Y_1Y_2) - \frac{\epsilon}{2})} \leq \text{Vol} \left(T_K^c\right) \leq e^{n(h(Y_1Y_2) + \frac{\epsilon}{2})}.
\]

Moreover, for any \( K \) with non-zero variances and \((y_1, y_2) \in T_K^c\)

\[
e^{-n(D(Y_1Y_2||Y_1Y_2)+h(Y_1Y_2)+o(1))} \leq P_{Y_1Y_2}^n(y_1, y_2) \leq e^{-n(D(Y_1Y_2||Y_1Y_2)+h(Y_1Y_2)-o(1))},
\]

where the distribution \( Q_{\bar{Y}_1\bar{Y}_2} \sim \mathcal{N}(0, \bar{K}) \).

The proof is analogous to the proof of Lemma 8 and so is omitted. The probability of a joint Gaussian type-class can be upper bounded by

\[
P_{Y_1Y_2}^n \left(T_K^c\right) \leq e^{-n(D(Y_1Y_2||Y_1Y_2)-o(1))}.
\]

Furthermore, by similar arguments to the proofs of Lemma 8 and Lemma 9, one can prove that for a given \( y_1 \in T_{\sigma^2_{Y_1}}^c \) and Gaussian channel \( W : Y_1 \to Y_2 \) the volume of a conditional Gaussian type-class can be bounded by

\[
\left(1 - \frac{2}{nc^2} - \frac{1}{nc^2\rho^2}\right) e^{n(h(Y_2|Y_1) - \epsilon)} \leq \text{Vol} \left(T_W^c(y_1)\right) \leq e^{n(h(Y_2|Y_1) + \epsilon)},
\] (3.10)
where the marginal $P_{Y_1} \sim \mathcal{N}(0, \sigma_{Y_1}^2)$.

The following type covering assertion is of key importance to the construction of our coding scheme.

**Lemma 10.** For any sufficiently large $n$, distribution $P_{Y_1} \sim \mathcal{N}(0, \sigma_{Y_1}^2)$ on $Y_1$ and numbers $d > 0$, $0 < \epsilon < 1$, there exists a set $C_{Y_1} \subset Y_1^n$ satisfying both

$$\min_{y_2 \in C_{Y_1}} \sum_{i=1}^{n} (y_1(i) - y_2(i))^2 \leq nd(1 + o(1)) \quad \text{for every } y_1 \in T_{\sigma_{Y_1}^2}^\epsilon$$

and

$$|C_{Y_1}| \leq e^{n(R(Y_1, d) + 6\epsilon)}.$$

**Proof.** The proof that follows is a modification of the original proof given by Arikan and Merhav [10, Theorem 5, p. 1095]. It is an extension of Berger’s type covering lemma to the case of two jointly Gaussian random variables with a mean square error distortion measure [12, Lemma 2.4.1].

We want to prove that $T_{\sigma_{Y_1}^2}^\epsilon$ can be covered by exponentially $e^{nR(Y_1, d)}$ code vectors $\{y_2(i)\}$ within Euclidean distance essentially as small as $\sqrt{nd}$. For $\sigma_{Y_1}^2 \leq d$, this is trivial as the vector $y_2 = 0$ represents any $y_1 \in T_{\sigma_{Y_1}^2}^\epsilon$ within distortion $d(1+\epsilon)$.

Assume next that $\sigma_{Y_1}^2 > d$ and let $0 < \epsilon < 1$. Let us construct a grid $S$ of all vectors in the Euclidean space $\mathbb{R}^n$, whose components are integer multiples of $2\delta$ for some small $0 < \delta \ll \sqrt{d}$. Consider the $n$-dimensional cubes of size $\delta$, centered at the grid points. For a given code $C_{Y_1} = \{y_2(1), \ldots, y_2(M)\}$, let $U(d)$ denote the subset of cubes intersecting $T_{\sigma_{Y_1}^2}^\epsilon$ for which the cube center $y_1^*$ satisfies

$$\|y_1^* - y_2(i)\|_2^2 > n(d + \mu), \quad \text{for all } i = 1, \ldots, M,$$

where $\mu$ is a small positive real, which will be specified later. This means that $U(d)$ is the set of cubes intersecting $T_{\sigma_{Y_1}^2}^\epsilon$, whose centers are not covered by $C_{Y_1}$ within distortion $d + \mu$. 
Consider a pair of jointly Gaussian random variables \((Y_1, Y_2)\) such that \(\mathbb{E}(Y_1 - Y_2)^2 \leq d\) and \(P_{Y_1} \sim \mathcal{N}(0, \sigma^2_{Y_1})\). If we denote by \(W : \mathcal{Y}_1 \rightarrow \mathcal{Y}_2\) the conditional distribution \(P_{Y_2|Y_1}\), then

\[
d \geq \mathbb{E}(Y_1 - Y_2)^2
= \sigma_{Y_2}^2 - 2\rho \sigma_{Y_1} \sigma_{Y_2} + \sigma_{Y_1}^2,
\]
giving \(\rho \frac{\sigma_{Y_2}}{\sigma_{Y_1}} \geq \frac{\sigma_{Y_1}^2 + \sigma_{Y_2}^2 - d}{2\sigma_{Y_1}^2}\). Assuming equality in the last relation and making use of Lemma 6, we get the Gaussian channel

\[
Y_2 = \left(\frac{\sigma_{Y_1}^2 + \sigma_{Y_2}^2 - d}{2\sigma_{Y_1}^2}\right) Y_1 + V
\]
with \(P_V \sim \mathcal{N}\left(0, \sigma^2_{Y_2} - \frac{(\sigma_{Y_1}^2 + \sigma_{Y_2}^2 - d)^2}{4\sigma_{Y_1}^2}\right)\) and \(Y_1 \perp V\). Let us now define a positive number \(\epsilon'\) satisfying

\[
\sigma_{Y_1}^2 (1 + \epsilon') = \left(\sqrt{\sigma_{Y_1}^2 (1 + \epsilon) + 2\delta}\right)^2
\]
and choose \(\delta\) such that \(\epsilon' = 2\epsilon\). In addition, let \(Y_2(1), \ldots, Y_2(M)\) denote i.i.d. vectors drawn uniformly in \(T_{\sigma_{Y_2}^2}^\epsilon\). If we show that \(\mathbb{E}|U(d)| < 1\), then there must exist a code for which \(U(d)\) is empty, which means that all cube centers are covered within distortion \(d + \mu\) and, therefore, by the triangle inequality, \(T_{\sigma_{Y_1}^2}^\epsilon\) is entirely covered by \(M\) spheres within distortion \((\sqrt{d + \mu} + \delta)^2\). We focus on the set of grid points

\[
S' = \left\{ y_1^* \in S : \sup_{x : \|x - y_1^*\|_\infty \leq \delta} x^* x \leq n\sigma_{Y_1}^2 (1 + \epsilon') \right\}
\]
and evaluate

\[
\mathbb{E}|U(d)| \leq \mathbb{E} \left\{ \sum_{y_1^* \in S'} \prod_{i=1}^M 1 \{ \|y_1^* - Y_2(i)\|_2^2 > n(d + \mu) \} \right\}
= \sum_{y_1^* \in S'} \left[ 1 - \Pr (\|y_1^* - Y_2(i)\|_2^2 \leq n(d + \mu)) \right]^M.
\]
By definition, $T'_W(y^*_1)$ is a subset of $T'_{\sigma^2_{Y_2}}$ for $y^*_1 \in T'_{\sigma^2_{Y_1}}$. In addition, it is easy to check that for a given $y^*_1$, the set $T'_W(y^*_1)$ has only $y_2$-vectors with $\|y^*_1 - y_2(i)\|_2^2 \leq n(d + \mu)$, where

$$\mu = \epsilon'(d + 2(\sigma^2_{Y_1} + \sigma^2_{Y_2} - d)).$$

Since the codewords are selected randomly with respect to a uniform distribution within $T'_{\sigma^2_{Y_2}}$, then by Lemma 8 and the lower bound in (3.10)

$$\Pr \left( \|y^*_1 - Y_2(i)\|_2^2 \leq n(d + \mu) \right) \geq \frac{\text{Vol} \left( T'_W(y^*_1) \right)}{\text{Vol} \left( T'_{\sigma^2_{Y_2}} \right)} \geq \left( 1 - \frac{2}{ne^2} - \frac{1}{ne^2 \rho^2} \right) e^{-n(I(Y_1;Y_2) + \frac{n}{2})}.$$ 

Thus

$$\mathbb{E} |U(d)| \leq |S'| \left[ 1 - \left( 1 - \frac{2}{ne^2} - \frac{1}{ne^2 \rho^2} \right) e^{-n(I(Y_1;Y_2) + \frac{n}{2})} \right]^M \leq e^{n(h(Y_1) + \frac{\epsilon'}{2} - \log 2\delta)} \exp \left[ -M \left( 1 - \frac{2}{ne^2} - \frac{1}{ne^2 \rho^2} \right) e^{-n(I(Y_1;Y_2) + \frac{n}{2})} \right],$$

where we have used the facts that $1 - x \leq e^{-x}$ and that the number of cubes in $T'_{\sigma^2_{Y_1}}$ cannot exceed the ratio between the volume of $T'_{\sigma^2_{Y_1}}$ and the volume of a cube $(2\delta)^n$. It is readily seen that for $\mathbb{E} |U(d)| \to 0$ as $n \to \infty$, it is sufficient for $M$ to satisfy

$$M \geq e^{n(I(Y_1;Y_2) + 2\epsilon')}.$$ 

This proves the existence of a set $C_{Y_1} \subset Y_2^n$ with $U(d) = \emptyset$ and

$$|C_{Y_1}| \leq M \leq e^{n(I(Y_1;Y_2) + 3\epsilon')}$$

for sufficiently large $n$. Noting that $R(Y_1,d) = I(Y_1;Y_2)$, completes the proof of the lemma. \qed

From the theory of large deviations we get the following lemma.
Lemma 11. For any distribution $P_Y \sim \mathcal{N}(0, \sigma_Y^2)$ on $\mathcal{Y}$ and $M_U > \sigma_Y^2$

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr \left( \frac{1}{n} y^t y > M_U \right) = -\frac{1}{2} \left( \frac{M_U}{\sigma_Y^2} - \log \frac{M_U}{\sigma_Y^2} - 1 \right)$$

and for any $0 < M_L < \sigma_Y^2$

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr \left( \frac{1}{n} y^t y < M_L \right) = -\frac{1}{2} \left( \frac{M_L}{\sigma_Y^2} - \log \frac{M_L}{\sigma_Y^2} - 1 \right).$$

Proof. Consider the set of probability distributions

$$E = \left\{ Q_Y : \int y^2 Q_Y dy \geq M_U \right\}.$$ 

By Sanov’s theorem [8, Theorem 12.4.1] it holds that

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr \left( \frac{1}{n} y^t y \geq M_U \right) = -D(Q_Y^* \| P_Y),$$

where

$$Q_Y^* = \arg \min_{Q_Y \in E} D(Q_Y \| P_Y).$$

Let us first argue that given $P_Y \sim \mathcal{N}(0, \sigma_Y^2)$ and any distribution $Q_Y$ with mean $\mu$ and variance $\sigma_Y^2$, the normal distribution $Q_Y' \sim \mathcal{N}(\mu, \sigma_Y^2)$ minimizes the Kullback-Leibler distance between $Q_Y$ and $P_Y$. From the definition of the relative entropy

$$D(Q_Y^* \| P_Y) = \int Q_Y \log \frac{Q_Y}{P_Y}$$

$$= \int Q_Y \log Q_Y - \int Q_Y \log P_Y$$

$$= -h(Q_Y) - \int Q_Y' \log P_Y$$

$$\geq -h(Q_Y') - \int Q_Y' \log P_Y$$

$$= D(Q_Y' \| P_Y),$$
where we have used the facts that the normal distribution maximizes entropy and that the distributions $Q_Y$, $Q'_Y$ yield the same moments of the quadratic form $\log P_Y$.

Making use of Lemma 7 and choosing $\mu = 0$ gives

$$D(Q_Y \parallel P_Y) \geq \frac{1}{2} \left( \frac{\bar{\sigma}_Y^2}{\sigma_Y^2} - \log \frac{\bar{\sigma}_Y^2}{\sigma_Y^2} - 1 \right).$$

Hence, our original optimization problem takes the form

$$D(Q^*_Y \parallel P_Y) = \min_{\bar{\sigma}_Y^2 \geq M_U} \left\{ \frac{1}{2} \left( \frac{\bar{\sigma}_Y^2}{\sigma_Y^2} - \log \frac{\bar{\sigma}_Y^2}{\sigma_Y^2} - 1 \right) \right\}.$$

By the monotonicity of $x - \log x - 1$ and the fact that $M_U > \sigma_Y^2$, we can easily conclude that the minimum is achieved when $\bar{\sigma}_Y^2 = M_U$, i.e.

$$D(Q^*_Y \parallel P_Y) = \frac{1}{2} \left( \frac{M_U}{\sigma_Y^2} - \log \frac{M_U}{\sigma_Y^2} - 1 \right).$$

The proof of the second part of the lemma is analogous and so is omitted.

### 3.1.3 Proof of Achievable Exponent

Having defined all necessary quantities, we are in position to present a coding scheme for the Gaussian point-to-point lossy compression system.

**Proof of Theorem 4.** For given $0 < M_L < 1$ and $0 < \Delta < M_L$, consider the grid

$$\bar{\sigma}_Y^2(i) = M_L + \Delta i \quad i = 1, 2, \ldots$$

Letting $\epsilon = \Delta/M_L$, the sphere $\{y : y^t y \leq nM_L\}$ together with the sets

$$T_{Y_i} = T_{\bar{\sigma}_Y^2(i)}^\epsilon \quad i = 1, 2, \ldots$$

entirely cover the space $Y^n \equiv \mathcal{R}^n$. Consider, now, the set of Gaussian type-classes

$$\mathcal{P}_Y = \{T_{Y_i} : \exists y \in T_{Y_i} \text{ with } y^t y \leq nM_U\},$$
where \( M_U \) is a number to be specified later and define

\[
\delta_n = \frac{1}{n} \log (|\mathcal{P}_Y|) \to 0 \quad (\text{as } n \to +\infty).
\]

Let \( R' < R, d' < d \) and choose \( \epsilon \) such that

\[
R' + 6\epsilon + \delta_n \leq R
\]

\[
d'(1 + o_\epsilon(1)) \leq d,
\]

where \( o_\epsilon(1) \) is the function from Lemma 10. By this lemma, if \( n \) is sufficiently large, then for every \( T_{\bar{Y}_i} \in \mathcal{P}_Y \) satisfying the condition \( R(\bar{Y}, d') \leq R' \), there exists a code \( C_{\bar{Y}_i} \subset \hat{\mathcal{Y}}^n \) such that

\[
|C_{\bar{Y}_i}| \leq e^{n(R(\bar{Y}, d') + 6\epsilon)} \leq e^{n(R' + 6\epsilon)}
\]

and

\[
\min_{\hat{y} \in C_{\bar{Y}_i}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \leq nd'(1 + o_\epsilon(1)) \leq nd
\]

for every \( y \in T_{\bar{Y}_i} \). Defining the code

\[
C_{\bar{Y}_i} = \bigcup_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') \leq R'} C_{\bar{Y}_i},
\]

we get that

\[
|C_{\bar{Y}_i}| \leq \sum_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') \leq R'} |C_{\bar{Y}_i}|
\]

\[
\leq |\mathcal{P}_Y| e^{n(R' + 6\epsilon)}
\]

\[
= e^{n(R' + 6\epsilon + \delta_n)}
\]

\[
\leq e^{nR}.
\]

Letting

\[
\mathcal{S}_{LU} = \{ y \in \mathcal{Y}^n : nM_L \leq y'y \leq nM_U \},
\]

\[
40
\]
the error event satisfies

\[ \mathcal{E}(f^{(n)}, \phi^{(n)}) \cap S_{LU} \subset \bigcup_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') > R'} T_{\bar{Y}_i}. \]

It, therefore, follows that

\[
\Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) \cap S_{LU} \right) \leq P^n_Y \left( \bigcup_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') > R'} T_{\bar{Y}_i} \right)
\leq \sum_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') > R'} P^n_Y (T_{\bar{Y}_i})
\leq \sum_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') > R'} e^{-n(D(\bar{Y}\|Y) - \alpha(1))}
\leq |\mathcal{P}_Y| e^{-n \left( \min_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') > R'} D(\bar{Y}\|Y) - \alpha(1) \right)}
= e^{-n \left( \min_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') > R'} D(\bar{Y}\|Y) - \alpha(1) - \delta_n \right)}. \tag{3.13}
\]

For the probability of error, we have

\[
\Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) \right) = \Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) \cap S_{LU} \right)
+ \Pr (S_{LU}^c) \Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) | S_{LU}^c \right)
\leq \Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) \cap S_{LU} \right) + \Pr (S_{LU}^c). \tag{3.14}
\]

In addition, making use of Lemma 11

\[
\Pr (S_{LU}^c) \leq \Pr \left( \mathbf{y}^t \mathbf{y} < nM_L \right) + \Pr \left( \mathbf{y}^t \mathbf{y} > nM_U \right)
\leq e^{-\frac{n}{2} (M_L - \log M_L - 1 - \alpha(1))} + e^{-\frac{n}{2} (M_U - \log M_U - 1 - \alpha(1))}. \tag{3.15}
\]

Combining (3.14) with (3.13) and (3.15) yields

\[
\Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) \right) \leq e^{-n \left( \min_{T_{\bar{Y}_i} \in \mathcal{P}_Y: R(\bar{Y}, d') > R'} D(\bar{Y}\|Y) - \alpha(1) - \delta_n \right)}
+ e^{-\frac{n}{2} (M_L - \log M_L - 1 - \alpha(1))} + e^{-\frac{n}{2} (M_U - \log M_U - 1 - \alpha(1))}. \tag{3.16}
\]
Choosing $0 < M_L \ll 1$ and $M_U \gg 1$, the first term in (3.16) will dominate in the limit, which implies

$$
\lim_{n \to \infty} -\frac{1}{n} \log \left[ \min_{f^{(n)}, \phi^{(n)}} \Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) \right) \right] \geq \min_{D(\tilde{Y}||Y) > R'} D(\tilde{Y}||Y),
$$

(3.17)

where the last inequality follows from the fact that the second minimization is over a larger set of distributions. Using the same arguments as in the proof of Lemma 11

$$
\min_{Q_Y:R(Y,d') > R'} D(\tilde{Y}||Y) \geq \min_{\sigma^2 > d' e^{2R'}} \left\{ \frac{1}{2} \left( \sigma^2_Y - \log \sigma^2_Y - 1 \right) \right\}
$$

where the last equality is due to the monotonicity of $x - \log x - 1$. We, therefore, have

$$
\lim_{n \to \infty} -\frac{1}{n} \log \left[ \min_{f^{(n)}, \phi^{(n)}} \Pr \left( \mathcal{E}(f^{(n)}, \phi^{(n)}) \right) \right] \geq \begin{cases} 
\frac{1}{2} \left( d' e^{2R'} - \log d' e^{2R'} - 1 \right) & \text{if } R' \geq R(Y, d') \\
0 & \text{if } R' < R(Y, d') \end{cases}
$$

Letting $R' \to R$, $d' \to d$ and $\epsilon \to 0$, the theorem follows.

### 3.2 Overview on Source Coding Exponents

In this part of the thesis we give an overview of the main results in the field of error exponents in lossy source coding. As mentioned earlier, Marton presented
the best error exponent for the point-to-point lossy compression system [9]. The Slepian-Wolf setup [13] was the next problem to be examined. The solution was given by Csiszár, Körner, and Marton, who derived tight exponential error bounds in a neighborhood of the boundary of the admissible rate region [14]. An improvement was later presented by Oohama and Han [15]. The problem of attainable error exponents for the Wyner-Ziv setup [16] was investigated by Haroutunian and Mekoush [17]. In 1995, Jayaraman also attempted to give a solution to this problem [18]. He derived upper and lower bounds on the probability of error, which, however, were proved to be loose. The gap between these two bounds was a result of a suboptimal coding scheme that failed to come up with a “good” encoding for the atypical source sequences. One year later, Haroutunian published a paper on error exponents for the two terminal source coding problem [19]. He presented exponents for systems considered by Kaspi and Berger [20], as well as by Berger and Yeung [21]. His paper, however, was proved to be partially erroneous [22]. Haroutunian’s mistake was to feed the output sequences of the binning process to the decoder output, without first using the side information to further improve these reconstructions. This way he failed to use auxiliary random variables in his expressions, ending up with suboptimal achievable error exponents and an error in his converse result. Guessing exponents and the establishment of a fundamental relation to Marton’s exponent was the next contribution in the field [10]. The work of Arikan and Merhav resulted also in a error exponent for the Gaussian point-to-point source coding problem and the idea of the Gaussian method of types. The latter exponent was verified by Ihara and Kubo, who used a different approach in their proof [11]. In 2005, Iriyama further extended this work and proved an error exponent for the point-to-point problem, when a general
continuous alphabet source is used [23]. In the meantime, the problem of noisy source coding exponents was introduced by Weissman and Merhav [24]. Their work is a generalization of Marton’s result not only because they introduce a lossy channel between the source and the encoder, but also because they allow the use of variable-length codes. Recently, Eswaran and Gastpar derived a bound on the convergence rate of the Markov lemma [25]. They used this result for getting achievable error exponents in lossy multiterminal source coding problems. In conclusion, finding a tight error exponent for even the Wyner-Ziv compression system still remains an open problem.

In the course of conducting this MS research, we attempted to derive an exponent for the Wyner-Ziv problem and its Gaussian extension. Using tools introduced by Arikan and Merhav [10], we presented an alternative proof for the best achievable error exponent in the Gaussian point-to-point setup. An extension of our work to the Gaussian noisy source coding problem would be the next step to take. Improving the current Gaussian method of types, by introducing disjoint type-classes, would also be an interesting problem to examine. A refinement of the already defined type-classes, by removing all common sequences from one of the adjacent type-classes, seems like a natural way of improvement.
Bibliography


