McMaster University
DigitalCommons@McMaster

Open Access Dissertations and Theses | Open Dissertations and Theses

10-1-2012

Gaussian Robust Sequential and Predictive Coding

Lin Song
McMaster University, songlin@grads.ece.mcmaster.ca

Follow this and additional works at: http://digitalcommons.mcmaster.ca/opendissertations

Part of the Electrical and Electronics Commons

Recommended Citation
Song, Lin, "Gaussian Robust Sequential and Predictive Coding" (2012). *Open Access Dissertations and Theses.* Paper 7288.

This Thesis is brought to you for free and open access by the Open Dissertations and Theses at DigitalCommons@McMaster. It has been accepted for inclusion in Open Access Dissertations and Theses by an authorized administrator of DigitalCommons@McMaster. For more information, please contact scom@mcmaster.ca.

Gaussian Robust Sequential and Predictive Coding

GAUSSIAN ROBUST SEQUENTIAL AND PREDICTIVE CODING

BY

LIN SONG

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

© Copyright by Lin Song, August 2012

All Rights Reserved

Doctor of Philosophy (2012) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE: Gaussian Robust Sequential and Predictive Coding

AUTHOR: Lin Song

M. Sc., (Electrical Engineering)

Harbin Institute of Technology, Heilongjiang, China

SUPERVISOR: Dr. Jun Chen

NUMBER OF PAGES: ix, 77

To the past four years, to the happiness and sharing with my family and friends.

Abstract

Video coding schemes designed based on sequential or predictive coding models are vulnerable to the loss of encoded frames at the decoder end. Motivated by this observation, in this thesis we propose two new coding models: robust sequential coding and robust predictive coding. For the Gauss-Markov source with the mean squared error distortion measure, we characterize certain supporting hyperplanes of the rate region of these two coding problems. The proof is divided into three steps: 1) it is shown that each supporting hyperplane of the rate region of Gaussian robust sequential coding admits a max-min lower bound; 2) the corresponding min-max upper bound is shown to be achievable by a robust predictive coding scheme; 3) a saddle point analysis proves that the max-min lower bound coincides with the min-max upper bound. Furthermore, it is shown that the proposed robust predictive coding scheme can be implemented using a successive quantization system. Theoretical and experimental results indicate that this scheme has a desirable "self-recovery" property. Our investigation also reveals an information-theoretic minimax theorem and the associated extremal inequalities.

Acknowledgements

This work would not have been possible without the guidance, support and contributions of many individuals. I would like to express my immense gratitude to them.

First of all, I would like to express my heartfelt gratitude to my advisor, Dr. Jun Chen, for his consistent guidance, support, encouragement and insightful discussions. His insight in information theory and logical thinking led me on the correct way to overcome all the difficulties. Dr. Chen is no doubt the best supervisor that I have met. He is nice and humorous. He told me many interesting stories about information-theory people and illustrated a vivid information-theory world in front of me, not just formal mathematics. He also shows his incredible intuition in discovering and solving problems during discussions. I have learned a lot from him.

Next, I would like to sincerely thank my committee members, Dr. Jiankang Zhang, Dr. Sorina Dumitrescuand and Dr. Lizhong Zheng for their valuable comments and suggestions. Many thanks to all the staff members in the ECE department, especially to Cheryl who helped me a lot. I want to thank all my friends at McMaster University who provide a supporting and exciting atmosphere.

Finally, I would like to thank my parents for their support and encourage. I would also like to thank my husband, Jihai, for his love and tolerance. I hope we will have a happy and sweet life together.

Notation and abbreviations

X	Random variable
$\mathcal{X}$	Alphabet of random variable X
X_i	Source variable at time i or the ith video frame
X^n	$1 \times n$ random vector $(X_1, X_2, \cdots X_n)$
σ_X^2	Variance of X
$p(x)$	Distribution of X
$\overline{R}$	Rate vector $(R_1, R_2, \cdots, R_L)$
$\mathbb{E}[\cdot]$	Expectation
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
MMSE	Minimum mean square error

Contents

Abstract **iv**

Acknowledgements **v**

Notation and abbreviations **vi**

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Robust Sequential and Predictive Coding 2

1.3 Main Results 7

1.4 Thesis Outline 11

2 Lower Bound **12**

2.1 Extremal Inequality 12

2.2 Proof of the Lower Bound 19

3 Upper Bound **25**

3.1 An Inner Bound of $\mathcal{R}_P(\overline{d}, \overline{\delta})$ 27

3.2 Proof of the Upper Bound 30

4 Saddle Point Analysis **34**

4.1 Proof of Theorem 3 37

5 Minimax Theorem **42**

5.1 Extremal Inequality 43

5.2 Proof of the Minimax Theorem 46

6 Miscellaneous Results **53**

6.1 The Minimum Sum Rate: A Special Case 53

6.2 Robust Predictive Coding System 56

6.3 Reconstruction Based on an Arbitrary Subset of Encoder Outputs . . 59

7 Conclusion and Future Work **64**

A Mathematical Elementary **65**

A.1 Jensen's Inequality 65

A.2 Mean Square Error Estimation 65

B Information Theory Elementary **67**

B.1 Entropy and Differential Entropy 67

B.2 Lossy Source Coding 68

B.3 Multiple Descriptions 69

C Proof of Lemma 1 **72**

D The Continuity of $\tilde{\gamma}_{i^*}(\gamma^*_{i^*})$ **73**

List of Figures

1.1 Sequential coding with hierarchical distortion constraints. 2

1.2 Predictive coding with hierarchical distortion constraints. 3

1.3 Robust sequential coding with hierarchical and individual distortion constraints. 5

1.4 Robust predictive coding with hierarchical and individual distortion constraints. 6

1.5 Multiple descriptions. 7

6.1 Robust predictive coding via successive quantization. Here $\tilde{X}_i^n$ can be viewed as a multi-letter version of $\tilde{U}_i$, $i = 1, \cdots, L$, and $\hat{X}_i^n$ can be viewed as a multi-letter version of U_i, $i = 2, \cdots, L$. 60

6.2 Reconstruction based on a subset of encoder outputs. 63

Chapter 1

Introduction

1.1 Background and Motivation

The sequential coding problem was first introduced by Viswanathan and Berger in [1]. Due to its potential relevance to video coding applications, this problem has received renewed interests in recent years [2; 3]. In a sequential coding system, shown in Fig. 1.1, L sources $X_1, \cdots, X_L$, each representing a video frame, are encoded and decoded in a causal manner, where Encoder i has access to $X_1, \cdots, X_i$, $i = 1, \cdots, L$, and the decoder reconstructs X_i based on the outputs from the first i encoders, $i = 1, \cdots, L$. If Encoder i is only allowed to have access to X_i as well as the outputs from the first $i-1$ encoders (if $i \geq 2$), then the resulting problem is known as predictive coding (see Fig. 1.2). It is shown in [4] that the rate regions of these two coding problems are identical if $X_1 \leftrightarrow X_2 \leftrightarrow \cdots \leftrightarrow X_L$ form a Markov chain. Note that this Markov chain condition is trivially satisfied when $L = 2$.

The existing schemes for sequential coding and predictive coding rely critically on the assumption that the decoder has access to the first i encoded frames (i.e.,

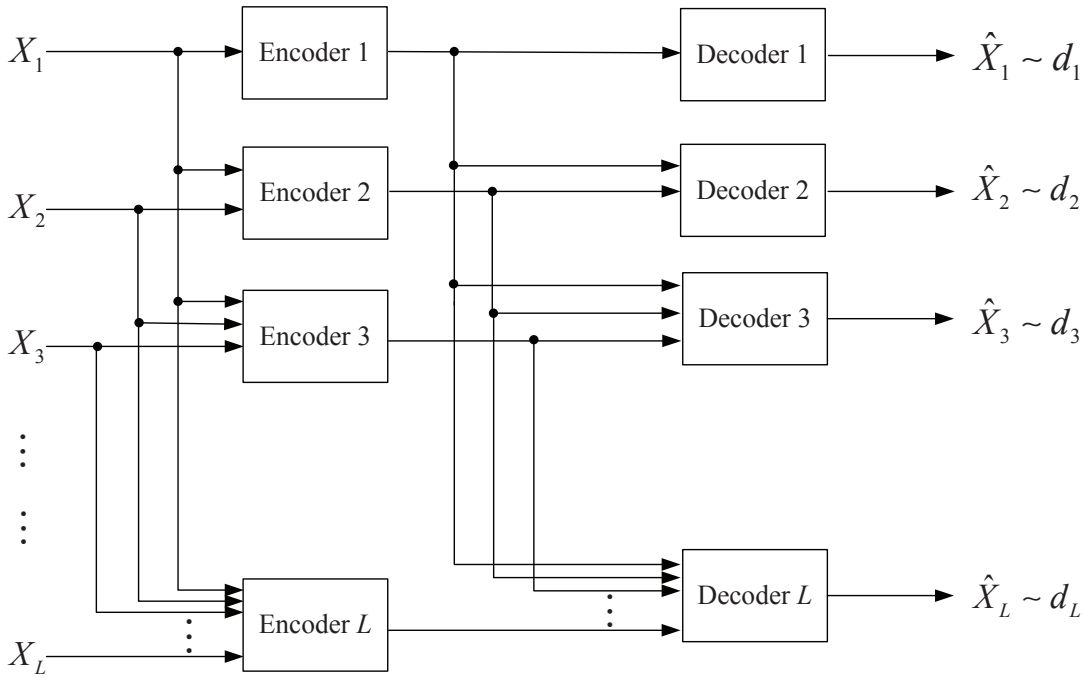

Figure 1.1: Sequential coding with hierarchical distortion constraints.

the outputs from the first i encoders) when reconstructing the ith frame (i.e., X_i). As a consequence, these schemes are vulnerable to the loss of encoded frames at the decoder end. Motivated by this observation, in this thesis we introduce a robust version of these two coding problems.

1.2 Robust Sequential and Predictive Coding

For the robust sequential coding problem and the robust predictive coding problem, it is required that the reconstruction of the ith frame has to meet a certain fidelity constraint even when the decoder only has access to the output from the ith encoder. This formulation is also applicable to the scenario where the encoded frames are to be decoded by two types of decoders: one has the capability of using multiple encoded

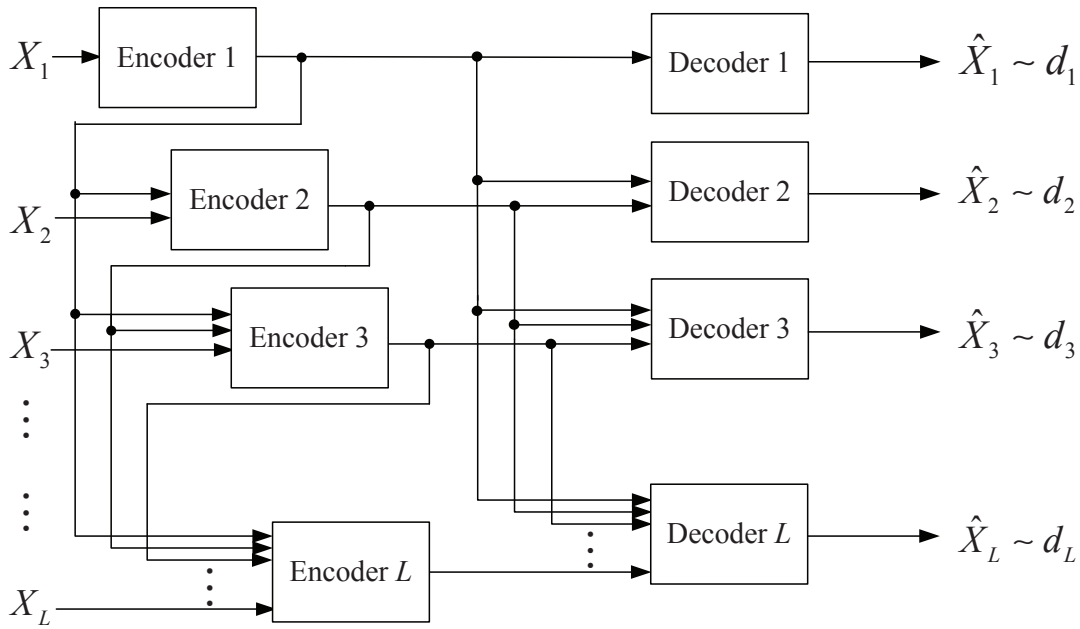

Figure 1.2: Predictive coding with hierarchical distortion constraints.

frames to reconstruct a target frame while the other can only perform the reconstruction operation based on a single encoded frame (due to storage or complexity constraints).

Consider L sources $X_1, \cdots, X_L$ with joint distribution $p(x_1, \cdots, x_L)$. Let $\{(X_{1j}, \cdots, X_{Lj})\}_{j=1}^{\infty}$ be i.i.d. copies of $(X_1, \cdots, X_L)$. Let $w_i : \mathcal{X}_i \times \hat{\mathcal{X}}_i \to [0, \infty)$ be a distortion measure, where $\mathcal{X}_i$ and $\hat{\mathcal{X}}_i$ are respectively the source alphabet (of X_i) and reconstruction alphabet, $i = 1, \cdots, L$.

Definition 1. *A rate vector $\overline{R} \triangleq (R_1, \cdots, R_L)$ is said to be achievable with a sequential coding system subject to hierarchical distortion constraint $\overline{d} \triangleq (d_1, \cdots, d_L)$ and individual distortion constraint $\overline{\delta} \triangleq (\delta_2, \cdots, \delta_L)$ if for every $\epsilon > 0$ there exist encoding functions $f_i^{(n)} : \mathcal{X}_1^n \times \cdots \times \mathcal{X}_i^n \to \mathcal{C}_i$, $i = 1, \cdots, L$, and decoding functions*

$\tilde{g}_i^{(n)} : \mathcal{C}_1 \times \cdots \times \mathcal{C}_i \to \hat{\mathcal{X}}_i$, $i = 1, \cdots, L$, *and* $\hat{g}_i^{(n)} : \mathcal{C}_i \to \hat{\mathcal{X}}_i$, $i = 2, \cdots, L$, *such that*

$$\frac{1}{n} \log |\mathcal{C}_i| \leq R_i + \epsilon, \quad i = 1, \cdots, L,$$

$$\mathbb{E}\Big[\frac{1}{n} \sum_{j=1}^{n} w_i(X_{ij}, \tilde{X}_{ij})\Big] \leq d_i + \epsilon, \quad i = 1, \cdots, L, \tag{1.1}$$

$$\mathbb{E}\Big[\frac{1}{n} \sum_{j=1}^{n} w_i(X_{ij}, \hat{X}_{ij})\Big] \leq \delta_i + \epsilon, \quad i = 2, \cdots, L, \tag{1.2}$$

where $\tilde{X}_i^n = \tilde{g}_i^{(n)}(C_1, \cdots, C_i)$, $i = 1, \cdots, L$, *and* $\hat{X}_i^n = \hat{g}_i^{(n)}(C_i)$, $i = 2, \cdots, L$, *with* $C_i = f_i^{(n)}(X_1^n, \cdots, X_i^n)$, $i = 1, \cdots, L$.

Definition 2. *The rate region* $\mathcal{R}_S(\overline{d}, \overline{\delta})$ *is the set of all the rate vectors achievable with a sequential coding system subject to hierarchical distortion constraint* $\overline{d}$ *and individual distortion constraint* $\overline{\delta}$.

A system diagram of robust sequential coding with hierarchical and individual distortion constraints can be found in Fig. 1.3.

Definition 3. *A rate vector* $\overline{R} \triangleq (R_1, \cdots, R_L)$ *is said to be achievable with a predictive coding system subject to hierarchical distortion constraint* $\overline{d} \triangleq (d_1, \cdots, d_L)$ *and individual distortion constraint* $\overline{\delta} \triangleq (\delta_2, \cdots, \delta_L)$ *if for every* $\epsilon > 0$ *there exist encoding functions* $f_1^{(n)} : \mathcal{X}_1^n \to \mathcal{C}_1$ *and* $f_i^{(n)} : \mathcal{C}_1 \times \cdots \times \mathcal{C}_{i-1} \times \mathcal{X}_i^n \to \mathcal{C}_i$, $i = 2, \cdots, L$, *and decoding functions* $\tilde{g}_i^{(n)} : \mathcal{C}_1 \times \cdots \times \mathcal{C}_i \to \hat{\mathcal{X}}_i$, $i = 1, \cdots, L$, *and* $\hat{g}_i^{(n)} : \mathcal{C}_i \to \hat{\mathcal{X}}_i$,

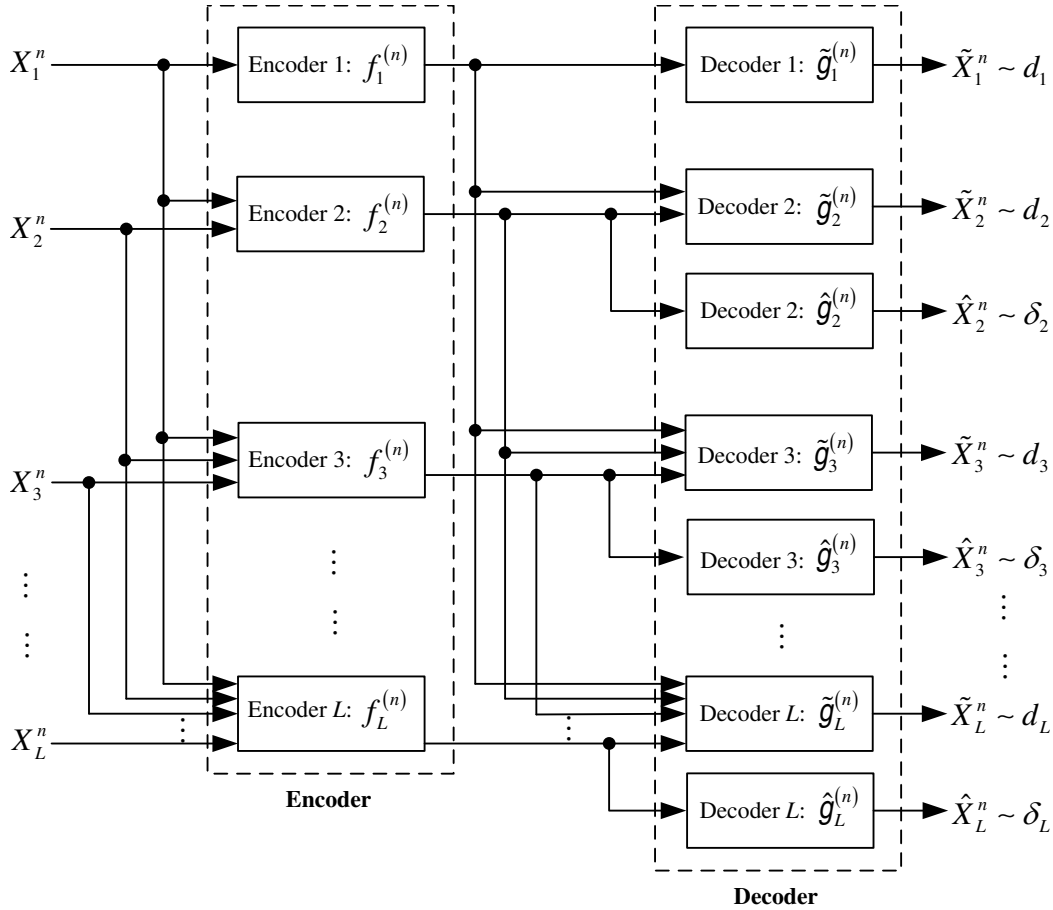

Figure 1.3: Robust sequential coding with hierarchical and individual distortion constraints.

$i = 2, \cdots, L$, *such that*

$$\frac{1}{n}\log|\mathcal{C}_i| \leq R_i + \epsilon, \quad i = 1, \cdots, L,$$

$$\mathbb{E}\Big[\frac{1}{n}\sum_{j=1}^{n} w_i(X_{ij}, \tilde{X}_{ij})\Big] \leq d_i + \epsilon, \quad i = 1, \cdots, L, \tag{1.3}$$

$$\mathbb{E}\Big[\frac{1}{n}\sum_{j=1}^{n} w_i(X_{ij}, \hat{X}_{ij})\Big] \leq \delta_i + \epsilon, \quad i = 2, \cdots, L, \tag{1.4}$$

where $\tilde{X}_i^n = \tilde{g}_i^{(n)}(C_1, \cdots, C_i)$, $i = 1, \cdots, L$, *and* $\hat{X}_i^n = \hat{g}_i^{(n)}(C_i)$, $i = 2, \cdots, L$, *with*

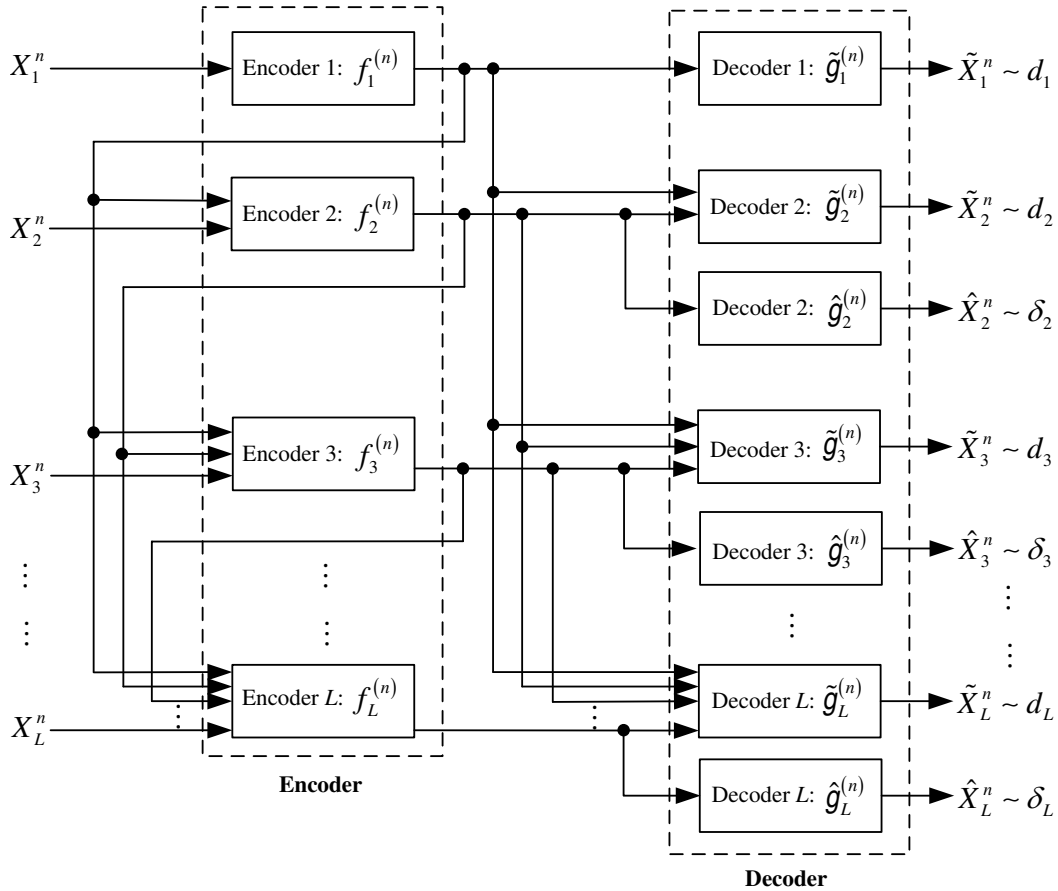

Figure 1.4: Robust predictive coding with hierarchical and individual distortion constraints.

$C_1 = f_1^{(n)}(X_1^n)$ *and* $C_i = f_i^{(n)}(C_1, \cdots, C_{i-1}, X_i^n)$, $i = 2, \cdots, L$.

Definition 4. *The rate region* $\mathcal{R}_P(\overline{d}, \overline{\delta})$ *is the set of all the rate vectors achievable with a predictive coding system subject to hierarchical distortion constraint* $\overline{d}$ *and individual distortion constraint* $\overline{\delta}$.

A system diagram of robust predictive coding with hierarchical and individual distortion constraints can be found in Fig. 1.4.

Note that the case $L = 1$ corresponds to the conventional lossy source coding

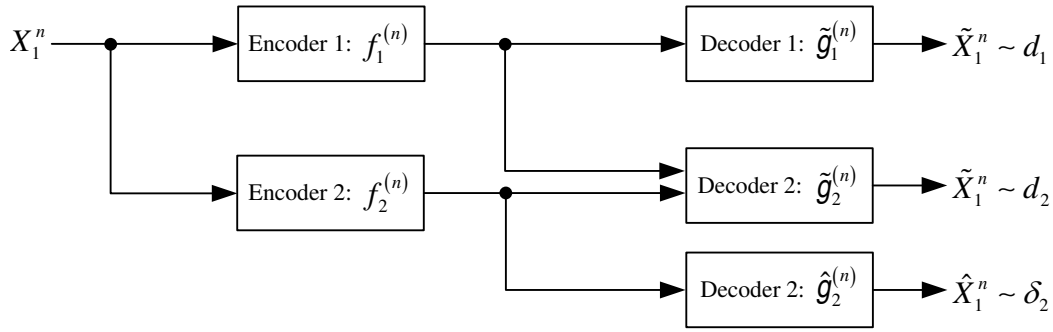

Figure 1.5: Multiple descriptions.

problem, and the tradeoff between rate and distortion is characterized by the rate-distortion function (see Appendix B.2).

If $L = 2$ and $X_1 = X_2$, then we recover the classic multiple description problem (see Fig. 1.5). A general inner bound of the rate region for this problem was derived by El Gamal and Cover [7] (see Appendix B.3). A tighter inner bound was found by Zhang and Berger [8] (see Appendix B.3). For the quadratic Gaussian case, Ozarow [9] showed that El Gamal-Cover inner bound is tight (see Appendix B.3).

Throughout this thesis, for any random object W and $1 \times n$ random vector X^n we define $\sigma^2_{X^n} = \frac{1}{n}\mathbb{E}[X^n(X^n)^T]$ and $\sigma^2_{X^n|W} = \sigma^2_{X^n - \mathbb{E}[X^n|W]}$; the logarithm function is to base e unless specified otherwise.

1.3 Main Results

In this thesis we focus on the special case where $X_1 \leftrightarrow X_2 \leftrightarrow \cdots \leftrightarrow X_L$ form a Gauss-Markov chain. With no essential loss of generality, we assume $X_{i+1} = X_i + \Delta_i$, $i = 1, \cdots, L-1$, where $X_1, \Delta_1, \cdots, \Delta_{L-1}$ are mutually independent zero-mean Gaussian random variables with $\sigma^2_{X_1} > 0$ and $\sigma^2_{\Delta_i} > 0$, $i = 1, \cdots, L-1$. Furthermore, we use the mean squared error as the distortion measure, i.e., $w_i(x_i, \hat{x}_i) = (x_i - \hat{x}_i)^2$

for all $x_i \in \mathbb{R}$ and $\hat{x}_i \in \mathbb{R}$, $i = 1, \cdots, L$. Note that in this setting there is no loss of optimality in assuming $\tilde{g}_i^{(n)}(C_1, \cdots, C_i) = \mathbb{E}[X_i^n | C_1, \cdots, C_i]$, $i = 1, \cdots, L$, and $\hat{g}_i^{(n)}(C_i) = \mathbb{E}[X_i^n | C_i]$, $i = 2, \cdots, L$. As a consequence, (1.1) and (1.3) can be rewritten as

$$\sigma^2_{X_i^n | C_1, \cdots, C_i} \leq d_i + \epsilon, \quad i = 1, \cdots, L;$$

for the same reason, (1.2) and (1.4) can be rewritten as

$$\sigma^2_{X_i^n | C_i} \leq \delta_i + \epsilon, \quad i = 2, \cdots, L.$$

Without loss of generality, we assume $0 < d_i \leq \sigma^2_{X_i}$, $i = 1, \cdots, L$, and $0 < \delta_i \leq \sigma^2_{X_i}$, $i = 2, \cdots, L$. Since both $\mathcal{R}_S(\overline{d}, \overline{\delta})$ and $\mathcal{R}_P(\overline{d}, \overline{\delta})$ are closed convex sets, it suffices to characterize their supporting hyperplanes, i.e., to solve the following optimization problems

$$\inf_{\overline{R} \in \mathcal{R}_S(\overline{d}, \overline{\delta})} \overline{\mu} \overline{R}^T \quad \text{and} \quad \inf_{\overline{R} \in \mathcal{R}_P(\overline{d}, \overline{\delta})} \overline{\mu} \overline{R}^T,$$

where $\overline{\mu} = (\mu_1, \cdots, \mu_L)$ with $\mu_i \geq 0$, $i = 1, \cdots, L$. In view of the fact that $\mathcal{R}_P(\overline{d}, \overline{\delta}) \subseteq \mathcal{R}_S(\overline{d}, \overline{\delta})$, we must have

$$\inf_{\overline{R} \in \mathcal{R}_S(\overline{d}, \overline{\delta})} \overline{\mu} \overline{R}^T \leq \inf_{\overline{R} \in \mathcal{R}_P(\overline{d}, \overline{\delta})} \overline{\mu} \overline{R}^T. \tag{1.5}$$

To state the main results of this thesis, we need to define the following function:

$$
\begin{aligned}
&\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}) \\
&= \frac{\mu_1}{2} \log\Big(\frac{\sigma_{X_1}^2}{\gamma_1}\Big) + \sum_{i=2}^{L} \frac{\mu_i}{2} \log\Big(\frac{\sigma_{X_i}^4(\gamma_{i-1} + \sigma_{\Delta_{i-1}}^2)}{(\sigma_{X_i}^2 - \theta_{i-1})(\gamma_{i-1} + \sigma_{\Delta_{i-1}}^2) + \sigma_{X_i}^2 \theta_{i-1}}\Big) \\
&\quad + \sum_{i=2}^{L} \frac{\mu_i}{2} \log\Big(\frac{(\sigma_{X_i}^2 - \theta_{i-1})\gamma_i + \sigma_{X_i}^2 \theta_{i-1}}{((\sigma_{X_i}^2 - \theta_{i-1})\delta_i + \sigma_{X_i}^2 \theta_{i-1})\gamma_i}\Big),
\end{aligned}
$$

where $\overline{\gamma} = (\gamma_1, \cdots, \gamma_L)$ and $\overline{\theta} = (\theta_1, \cdots, \theta_{L-1})$. Furthermore, let

$$
\begin{aligned}
\kappa_l(\overline{\mu}, \overline{d}, \overline{\delta}) &= \sup_{\theta_i \in (0, \sigma_{X_{i+1}}^2), i=1,\cdots,L-1} \; \min_{\gamma_i \in [0, d_i], i=1,\cdots,L} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}), \\
\kappa_u(\overline{\mu}, \overline{d}, \overline{\delta}) &= \inf_{\gamma_i \in (0, d_i), i=1,\cdots,L} \; \max_{\theta_i \in [0, \sigma_{X_{i+1}}^2], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}).
\end{aligned}
$$

Theorem 1. *For $\overline{\mu}$ with $\mu_1 \geq \cdots \geq \mu_L \geq 0$,*

$$
\inf_{\overline{R} \in \mathcal{R}_S(\overline{d}, \overline{\delta})} \overline{\mu} \overline{R}^T \geq \kappa_l(\overline{\mu}, \overline{d}, \overline{\delta}).
$$

Theorem 2. *For $\overline{\mu}$ with $\mu_i \geq 0$, $i = 1, \cdots, L$,*

$$
\inf_{\overline{R} \in \mathcal{R}_P(\overline{d}, \overline{\delta})} \overline{\mu} \overline{R}^T \leq \kappa_u(\overline{\mu}, \overline{d}, \overline{\delta}).
$$

Theorem 3. *For $\overline{\mu}$ with $\mu_1 \geq \cdots \geq \mu_L \geq 0$,*

$$
\kappa_l(\overline{\mu}, \overline{d}, \overline{\delta}) = \kappa_u(\overline{\mu}, \overline{d}, \overline{\delta}).
$$

The proofs of Theorems 1, 2, and 3 are given in Chapters 2, 3, and 4, respectively.

These theorems together with (1.5) lead to the following result, which provides a characterization of certain supporting hyperplanes of $\mathcal{R}_S(\overline{d},\overline{\delta})$ and $\mathcal{R}_P(\overline{d},\overline{\delta})$; in particular, setting $\mu_1 = \cdots = \mu_L = 1$ gives the minimum sum rate of these two rate regions.

Theorem 4. *For $\overline{\mu}$ with $\mu_1 \geq \cdots \geq \mu_L \geq 0$,*

$$\inf_{\overline{R}\in\mathcal{R}_S(\overline{d},\overline{\delta})} \overline{\mu}\overline{R}^T = \inf_{\overline{R}\in\mathcal{R}_P(\overline{d},\overline{\delta})} \overline{\mu}\overline{R}^T = \kappa_l(\overline{\mu},\overline{d},\overline{\delta}) = \kappa_u(\overline{\mu},\overline{d},\overline{\delta}).$$

For the special case $L = 2$, it can be verified that $\kappa_l(\overline{\mu},\overline{d},\overline{\delta})$ and $\kappa_u(\overline{\mu},\overline{d},\overline{\delta})$ have the following explicit expression (when $\mu_1 \geq \mu_2 \geq 0$):

$$\begin{aligned}
&\kappa_l(\overline{\mu},\overline{d},\overline{\delta}) = \kappa_u(\overline{\mu},\overline{d},\overline{\delta}) \\
&= \begin{cases} \frac{\mu_1}{2}\log\left(\frac{\sigma^2_{X_1}}{d_1}\right) + \frac{\mu_2}{2}\log\left(\frac{d_1+\sigma^2_{\Delta_1}}{d_2}\right), & d_2 \leq d_1 + \sigma^2_{\Delta_1} + \delta_2 - \sigma^2_{X_2} \\ \frac{\mu_1}{2}\log\left(\frac{\sigma^2_{X_1}}{d_1}\right) + \frac{\mu_2}{2}\log\left(\frac{\sigma^2_{X_2}}{\delta_2}\right), & d_2 \geq (\frac{1}{d_1+\sigma^2_{\Delta_1}} + \frac{1}{\delta_2} - \frac{1}{\sigma^2_{X_2}})^{-1} \\ \frac{\mu_1}{2}\log\left(\frac{\sigma^2_{X_1}}{d_1}\right) + \frac{\mu_2}{2}\log\left(\frac{(d_1+\sigma^2_{\Delta_1})\eta}{d_2}\right), & \text{otherwise} \end{cases},
\end{aligned}$$

where

$$\eta = \frac{(\sigma^2_{X_2} - d_2)^2}{(\sigma^2_{X_2} - d_2)^2 - (\sqrt{(\sigma^2_{X_2} - d_1 - \sigma^2_{\Delta_1})(\sigma^2_{X_2} - \delta_2)} - \sqrt{(d_1 + \sigma^2_{\Delta_1} - d_2)(\delta_2 - d_2)})^2}.$$

Our formulation of robust sequential coding and predictive coding is partly inspired by the classic multiple description problem (see, e.g., [7; 10]). In fact, the problem treated in [10] can be viewed as a degenerate case of our setup with $\sigma^2_{\Delta_i} = 0$, $i = 1, \cdots, L-1$. It will be seen that such a connection allows us to leverage the techniques developed for the multiple description problem and, albeit somewhat implicitly, provides conceptual

guidelines for our analysis. However, a straightforward application of the existing techniques turns out to be insufficient for handling these new problems. Indeed, we need to establish a new extremal inequality for the converse argument (see Chapter 2); the achievability scheme (see Chapter 3) and the saddle point analysis (see Chapter 4) are also more delicate than their counterparts in [10]. Moreover, the new coding problems possess certain features not found in the multiple description problem; for example, the special case studied in Chapter 6 has no natural counterpart in multiple description coding. Finally and most importantly, the analysis of the new coding problems enables us to extract an information-theoretic minimax theorem which is of interest in its own right (see Chapter 5).

1.4 Thesis Outline

The remainder of this thesis is organized as follows. We state our main results in Chapter 1; these results provide a partial characterization of the rate region of robust sequential coding and robust predictive coding for the Gauss-Markov source model under the mean squared error distortion constraint. The proofs of these results are given in Chapters 2, 3, and 4. It is shown in Chapter 5 that our main results can be viewed as a manifestation of an information-theoretic minimax theorem. Chapter 6 contains an explicit characterization of the minimum sum rate for a special class of sources and distortion constraints and provides a detailed discussion of the proposed robust predictive coding scheme. We conclude this thesis in Chapter 7.

Chapter 2

Lower Bound

2.1 Extremal Inequality

The following extremal inequality is the main technical ingredient in the proof of Theorem 1. It can be viewed as a generalization of [10, Lemma 1].

Theorem 5. *Let N_i^n be a zero-mean Gaussian random vector with i.i.d. entries of positive variance $\sigma^2_{N_i}$, $i = 1, 2, 3$, where $\sigma^2_{N_2} \leq \sigma^2_{N_3}$. Let ν_1, ν_2, ρ_1, ρ_2, and d be arbitrary real numbers satisfying $\nu_1 \geq \nu_2 \geq 0$ and $d > 0$. Then for any random vector S^n and random object W, jointly independent of (N_1^n, N_2^n, N_3^n), such that $\sigma^2_{S^n|W} \leq d$,*

$$\begin{aligned}&\nu_1(h(\rho_1 S^n + N_1^n|W) - h(S^n|W)) - \nu_2(h(\rho_2 S^n + N_3^n|W) - h(\rho_2 S^n + N_2^n|W))\\ &\geq \min_{\gamma\in[0,d]} \frac{\nu_1 n}{2}\log\Big(\frac{\rho_1^2\gamma + \sigma^2_{N_1}}{\gamma}\Big) - \frac{\nu_2 n}{2}\Big(\frac{\rho_2^2\gamma + \sigma^2_{N_3}}{\rho_2^2\gamma + \sigma^2_{N_2}}\Big).\end{aligned}$$

The following lemmas are the special cases of Theorem 5.

Lemma 1. *For any random vector S^n and random object W such that $\sigma^2_{S^n|W} \leq d$,*

$$h(S^n|W) \leq \frac{n}{2}\log(2\pi e d).$$

For completeness we give a proof in Appendix C.

The following result is a variant of the worst additive noise lemma by Ihara [18] as well as Diggavi and Cover [19, Lemma II.2]. Its proof can be found in [11, Appendix B].

Lemma 2. *Let Z^n be a zero-mean Gaussian random vector with i.i.d. entries of positive variance σ^2_Z. For any random vector S^n and random object W, jointly independent of Z^n, such that $\sigma^2_{S^n|W} \leq d$,*

$$h(S^n + Z^n|W) - h(S^n|W) \geq \frac{n}{2}\log\Big(\frac{d+\sigma^2_Z}{d}\Big).$$

Lemma 3. *Let Z^n_i be a zero-mean Gaussian random vector with i.i.d. entries of positive variance $\sigma^2_{Z_i}$, $i = 1, 2$. Let ν_1 and ν_2 be arbitrary real numbers satisfying $\nu_1 \geq \nu_2 \geq 0$. Then for any random vector S^n and random object W, jointly independent of (Z^n_1, Z^n_2), such that $\sigma^2_{S^n|W} \leq d$,*

$$\begin{aligned}&-\nu_1 h(S^n|W) - \nu_2(h(S^n + Z^n_2|W) - h(S^n + Z^n_1|W))\\ &\geq -\frac{\nu_1 n}{2}\log(2\pi e d) - \frac{\nu_2 n}{2}\log\Big(\frac{d+\sigma^2_{Z_2}}{d+\sigma^2_{Z_1}}\Big).\end{aligned}$$

Proof: Note that

$$\begin{aligned}&-\nu_1 h(S^n|W)-\nu_2(h(S^n+Z_2^n|W)-h(S^n+Z_1^n|W))\\&=\nu_1(h(S^n+Z_1^n|W)-h(S^n|W))-\nu_2 h(S^n+Z_2^n|W)-(\nu_1-\nu_2)h(S^n+Z_1^n|W).\end{aligned} \tag{2.1}$$

Since $\sigma^2_{S^n+Z_i^n|W}=\sigma^2_{S^n|W}+\sigma^2_{Z_i}\le d+\sigma^2_{Z_i}$, $i=1,2$, it follows from Lemma 1 that

$$\begin{aligned}&-\nu_2 h(S^n+Z_2^n|W)-(\nu_1-\nu_2)h(S^n+Z_1^n|W)\\&\ge-\frac{\nu_2 n}{2}\log(2\pi e(d+\sigma^2_{Z_2}))-\frac{(\nu_1-\nu_2)n}{2}\log(2\pi e(d+\sigma^2_{Z_1})).\end{aligned} \tag{2.2}$$

Furthermore, by Lemma 2, we have

$$\nu_1(h(S^n+Z_1^n|W)-h(S^n|W))\ge\frac{\nu_1 n}{2}\log\Big(\frac{d+\sigma^2_{Z_1}}{d}\Big). \tag{2.3}$$

Substituting (2.2) and (2.3) into (2.1) completes the proof of Lemma 3. □

Now we are ready to prove Theorem 5. It can be verified that

- if $\rho_1=\rho_2=0$, then Theorem 5 is implied by Lemma 1;
- if $\rho_1\ne\rho_2=0$ or $\sigma^2_{N_2}=\sigma^2_{N_3}$, then Theorem 5 is implied by Lemma 2;
- if $\rho_2\ne\rho_1=0$, then Theorem 5 is implied by Lemma 3.

Therefore, it suffices to consider the case where $\sigma^2_{N_2}<\sigma^2_{N_3}$ and $\rho_i\ne 0$, $i=1,2$. With no loss of generality, we shall assume $\rho_1=\rho_2=1$.

First consider the case $\sigma_{N_1}^2 \geq \sigma_{N_3}^2$. Note that

$$\begin{aligned}
&\nu_1(h(S^n + N_1^n|W) - h(S^n|W)) - \nu_2(h(S^n + N_3^n|W) - h(S^n + N_2^n|W)) \\
&= \nu_1(h(S^n + N_1^n|W) - h(S^n + N_3^n|W)) + \nu_2(h(S^n + N_2^n|W) - h(S^n|W)) \\
&\quad + (\nu_1 - \nu_2)(h(S^n + N_3^n|W) - h(S^n|W)). \qquad (2.4)
\end{aligned}$$

By Lemma 2,

$$\begin{aligned}
&\nu_2(h(S^n + N_2^n|W) - h(S^n|W)) + (\nu_1 - \nu_2)(h(S^n + N_3^n|W) - h(S^n|W)) \\
&\geq \frac{\nu_2 n}{2} \log\Big(\frac{d + \sigma_{N_2}^2}{d}\Big) + \frac{(\nu_1 - \nu_2)n}{2} \log\Big(\frac{d + \sigma_{N_3}^2}{d}\Big). \qquad (2.5)
\end{aligned}$$

If $\sigma_{N_1}^2 > \sigma_{N_3}^2$, then without loss of generality we can assume $N_1^n = N_3^n + \Theta^n$, where Θ^n is independent of (N_3^n, S^n, W), and the entries of Θ^n are i.i.d. Gaussian random variables with mean zero and variance $\sigma_{N_1}^2 - \sigma_{N_3}^2$. Hence,

$$\begin{aligned}
&\nu_1(h(S^n + N_1^n|W) - h(S^n + N_3^n|W)) \\
&= \nu_1(h(S^n + N_3^n + \Theta^n|W) - h(S^n + N_3^n|W)) \\
&\geq \frac{\nu_1 n}{2} \log\Big(\frac{d + \sigma_{N_1}^2}{d + \sigma_{N_3}^2}\Big), \qquad (2.6)
\end{aligned}$$

where (2.6) follows from Lemma 2 and the fact that $\sigma_{S^n+N_3^n|W}^2 = \sigma_{S^n|W}^2 + \sigma_{N_3}^2 \leq d + \sigma_{N_3}^2$. It is clear that (2.6) also holds when $\sigma_{N_1}^2 = \sigma_{N_3}^2$. Substituting (2.5) and (2.6) into

(2.4) gives

$$\begin{aligned}&\nu_1(h(S^n+N_1^n|W)-h(S^n|W))-\nu_2(h(S^n+N_3^n|W)-h(S^n+N_2^n|W))\\&\geq\frac{\nu_1 n}{2}\log\Big(\frac{d+\sigma_{N_1}^2}{d}\Big)-\frac{\nu_2 n}{2}\log\Big(\frac{d+\sigma_{N_3}^2}{d+\sigma_{N_2}^2}\Big),\end{aligned}$$

which is the desired result.

Now it suffices to prove Theorem 5 for the case $\sigma_{N_1}^2<\sigma_{N_3}^2$. To this end, we use a reduction method inspired by [20]. Without loss of generality, we assume $N_3^n=N_i^n+\Theta_i^n$, where Θ_i^n is independent of (N_i^n,S^n,W), and the entries of Θ_i^n are i.i.d. Gaussian random variables with mean zero and variance $\sigma_{N_3}^2-\sigma_{N_i}^2$, $i=1,2$. Note that

$$\begin{aligned}&\nu_1(h(S^n+N_1^n|W)-h(S^n|W))-\nu_2(h(S^n+N_3^n|W)-h(S^n+N_2^n|W))\\&=\nu_1(h(S^n+N_3^n|W)+h(S^n+N_1^n|S^n+N_3^n,W)-h(S^n+N_3^n|S^n+N_1^n,W))\\&\quad-\nu_1(h(S^n+N_3^n|W)+h(S^n|S^n+N_3^n,W)-h(S^n+N_3^n|S^n,W))-\nu_2h(S^n+N_3^n|W)\\&\quad+\nu_2(h(S^n+N_3^n|W)+h(S^n+N_2^n|S^n+N_3^n,W)-h(S^n+N_3^n|S^n+N_2^n,W))\\&=\nu_1(h(S^n+N_1^n|S^n+N_3^n,W)-h(\Theta_1^n))-\nu_1(h(S^n|S^n+N_3^n,W)-h(N_3^n))\\&\quad+\nu_2(h(S^n+N_2^n|S^n+N_3^n,W)-h(\Theta_2^n))\\&=\nu_1\Big(h(S^n+N_1^n|S^n+N_3^n,W)-\frac{n}{2}\log(2\pi e(\sigma_{N_3}^2-\sigma_{N_1}^2))\Big)\\&\quad-\nu_1\Big(h(S^n|S^n+N_3^n,W)-\frac{n}{2}\log(2\pi e\sigma_{N_3}^2)\Big)\\&\quad+\nu_2\Big(h(S^n+N_2^n|S^n+N_3^n,W)-\frac{n}{2}\log(2\pi e(\sigma_{N_3}^2-\sigma_{N_2}^2))\Big).\end{aligned}\tag{2.7}$$

Let $Q_i^n=N_i^n-\mathbb{E}[N_i^n|N_3^n]$, $i=1,2$. It can be verified that $Q_i^n=N_i^n-\sigma_{N_i}^2\sigma_{N_3}^{-2}N_3^n$, $i=1,2$. Moreover, it is clear that Q_i^n is independent of (N_3^n,S^n,W), and the entries

of Q_i^n are i.i.d. Gaussian random variables with mean zero and variance $\sigma_{N_i}^2 - \sigma_{N_i}^4 \sigma_{N_3}^{-2}$, $i = 1, 2$. Note that

$$\begin{aligned}
&h(S^n + N_i^n | S^n + N_3^n, W) \\
&= h(S^n + \sigma_{N_i}^2 \sigma_{N_3}^{-2} N_3^n + Q_i^n | S^n + N_3^n, W) \\
&= h((1 - \sigma_{N_i}^2 \sigma_{N_3}^{-2}) S^n + Q_i^n | S^n + N_3^n, W) \\
&\geq \frac{n}{2} \log \left(e^{\frac{2}{n} h((1-\sigma_{N_i}^2 \sigma_{N_3}^{-2}) S^n | S^n + N_3^n, W)} + e^{\frac{2}{n} h(Q_i^n)} \right) && (2.8) \\
&= \frac{n}{2} \log \left((1 - \sigma_{N_i}^2 \sigma_{N_3}^{-2})^2 e^{\frac{2}{n} h(S^n | S^n + N_3^n, W)} + 2\pi e (\sigma_{N_i}^2 - \sigma_{N_i}^4 \sigma_{N_3}^{-2}) \right), \quad i = 1, 2, && (2.9)
\end{aligned}$$

where (2.8) follows by the entropy power inequality. Substituting (2.9) into (2.7), we obtain

$$\begin{aligned}
&\nu_1 (h(S^n + N_1^n | W) - h(S^n | W)) - \nu_2 (h(S^n + N_3^n | W) - h(S^n + N_2^n | W)) \\
&\geq \nu_1 \Big(\frac{n}{2} \log \left((1 - \sigma_{N_1}^2 \sigma_{N_3}^{-2})^2 e^{\frac{2}{n} h(S^n | S^n + N_3^n, W)} + 2\pi e (\sigma_{N_1}^2 - \sigma_{N_1}^4 \sigma_{N_3}^{-2}) \right) \\
&\quad - \frac{n}{2} \log (2\pi e (\sigma_{N_3}^2 - \sigma_{N_1}^2)) \Big) \\
&\quad - \nu_1 \Big(h(S^n | S^n + N_3^n, W) - \frac{n}{2} \log (2\pi e \sigma_{N_3}^2) \Big) \\
&\quad + \nu_2 \Big(\frac{n}{2} \log \left((1 - \sigma_{N_2}^2 \sigma_{N_3}^{-2})^2 e^{\frac{2}{n} h(S^n | S^n + N_3^n, W)} + 2\pi e (\sigma_{N_2}^2 - \sigma_{N_2}^4 \sigma_{N_3}^{-2}) \right) \\
&\quad - \frac{n}{2} \log (2\pi e (\sigma_{N_3}^2 - \sigma_{N_2}^2)) \Big). && (2.10)
\end{aligned}$$

Now we proceed to bound $h(S^n | S^n + N_3^n, W)$. Let $\hat{S}^n$ be an estimate of S^n based on $(S^n + N_3^n, W)$, where

$$\hat{S}^n = \mathbb{E}[S^n | W] + \sigma_{S^n|W}^2 (\sigma_{S^n|W}^2 + \sigma_{N_3}^2)^{-1} (S^n - \mathbb{E}[S^n | W] + N_3^n).$$

It can be verified that

$$\frac{1}{n}\mathbb{E}[(S^n-\hat{S}^n)(S^n-\hat{S}^n)^T]=\frac{\sigma^2_{S^n|W}\sigma^2_{N_3}}{\sigma^2_{S^n|W}+\sigma^2_{N_3}}.$$

Since $\sigma^2_{S^n|S^n+N_3^n,W}\leq\frac{1}{n}\mathbb{E}[(S^n-\hat{S}^n)(S^n-\hat{S}^n)^T]$ and $\sigma^2_{S^n|W}\leq d$, we have

$$\sigma^2_{S^n|S^n+N_3^n,W}\leq(d^{-1}+\sigma^{-2}_{N_3})^{-1},$$

which, by Lemma 1, implies that

$$h(S^n|S^n+N_3^n,W)\leq\frac{n}{2}\log(2\pi e(d^{-1}+\sigma^{-2}_{N_3})^{-1}). \tag{2.11}$$

In view of (2.10) and (2.11), we have

$$\begin{aligned}
&\nu_1(h(S^n+N_1^n|W)-h(S^n|W))-\nu_2(h(S^n+N_3^n|W)-h(S^n+N_2^n|W))\\
&\geq\min_{\tilde{\gamma}\in[0,(d^{-1}+\sigma^{-2}_{N_3})^{-1}]}\nu_1\Big(\frac{n}{2}\log\Big(2\pi e(1-\sigma^2_{N_1}\sigma^{-2}_{N_3})^2\tilde{\gamma}+2\pi e(\sigma^2_{N_1}-\sigma^4_{N_1}\sigma^{-2}_{N_3})\Big)\\
&\quad-\frac{n}{2}\log(2\pi e(\sigma^2_{N_3}-\sigma^2_{N_1}))\Big)\\
&\quad-\nu_1\Big(\frac{n}{2}\log(2\pi e\tilde{\gamma})-\frac{n}{2}\log(2\pi e\sigma^2_{N_3})\Big)\\
&\quad+\nu_2\Big(\frac{n}{2}\log\Big(2\pi e(1-\sigma^2_{N_2}\sigma^{-2}_{N_3})^2\tilde{\gamma}+2\pi e(\sigma^2_{N_2}-\sigma^4_{N_2}\sigma^{-2}_{N_3})\Big)-\frac{n}{2}\log(2\pi e(\sigma^2_{N_3}-\sigma^2_{N_2}))\Big).
\end{aligned}$$

Let $\gamma=\frac{\tilde{\gamma}\sigma^2_{N_3}}{\sigma^2_{N_3}-\tilde{\gamma}}$. Note that there is a one-to-one correspondence between $\gamma\in[0,d]$ and

$\tilde{\gamma} \in [0, (d^{-1} + \sigma_{N_3}^{-2})^{-1}]$. Moreover, it can be verified that

$$\begin{aligned}
&\nu_1\Big(\frac{n}{2}\log\Big(2\pi e(1-\sigma_{N_1}^2\sigma_{N_3}^{-2})^2\tilde{\gamma} + 2\pi e(\sigma_{N_1}^2 - \sigma_{N_1}^4\sigma_{N_3}^{-2})\Big) - \frac{n}{2}\log(2\pi e(\sigma_{N_3}^2 - \sigma_{N_1}^2))\Big)\\
&-\nu_1\Big(\frac{n}{2}\log(2\pi e\tilde{\gamma}) - \frac{n}{2}\log(2\pi e\sigma_{N_3}^2)\Big)\\
&+\nu_2\Big(\frac{n}{2}\log\Big(2\pi e(1-\sigma_{N_2}^2\sigma_{N_3}^{-2})^2\tilde{\gamma} + 2\pi e(\sigma_{N_2}^2 - \sigma_{N_2}^4\sigma_{N_3}^{-2})\Big) - \frac{n}{2}\log(2\pi e(\sigma_{N_3}^2 - \sigma_{N_2}^2))\Big)\\
&= \frac{\nu_1 n}{2}\log\Big(\frac{\gamma+\sigma_{N_1}^2}{\gamma}\Big) - \frac{\nu_2 n}{2}\Big(\frac{\gamma+\sigma_{N_3}^2}{\gamma+\sigma_{N_2}^2}\Big),
\end{aligned}$$

which completes the proof.

2.2 Proof of the Lower Bound

Now we proceed to prove Theorem 1. The proof relies on Theorem 5 as well as the techniques developed in [9; 10; 11]. Given $\overline{R} \in \mathcal{R}_S(\overline{d}, \overline{\delta})$, it suffices to show that

$$\overline{\mu}\overline{R}^T \geq \min_{\gamma_i \in [0,d_i], i=1,\cdots,L} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}) \tag{2.12}$$

for all $\overline{\theta}$ with $\theta_i \in (0, \sigma_{X_{i+1}}^2)$, $i = 1, \cdots, L-1$.

By Definition 1, for every $\epsilon > 0$ there exist L encoding functions $f_i^{(n)} : \mathbb{R}^{i\times n} \to \mathcal{C}_i$, $i = 1, \cdots, L$, such that

$$\begin{aligned}
&\frac{1}{n}\log|\mathcal{C}_i| \leq R_i + \epsilon, \quad i = 1, \cdots, L,\\
&\sigma_{X_i^n|C_1,\cdots,C_i}^2 \leq d_i + \epsilon, \quad i = 1, \cdots, L,\\
&\sigma_{X_i^n|C_i}^2 \leq \delta_i + \epsilon, \quad i = 2, \cdots, L,
\end{aligned}$$

where $C_i = f_i^{(n)}(X_1^n, \cdots, X_i^n)$, $i = 1, \cdots, L$. One can readily verify that

$$\sum_{i=1}^{L} \mu_i (R_i + \epsilon) \tag{2.13}$$

$$\geq \sum_{i=1}^{L} \frac{\mu_i}{n} H(C_i)$$

$$\geq \frac{\mu_1}{n} I(X_1^n; C_1) + \sum_{i=2}^{L} \frac{\mu_i}{n} I(C_1, \cdots, C_{i-1}, X_i^n; C_i)$$

$$= \frac{\mu_1}{n} I(X_1^n; C_1)$$

$$+ \sum_{i=2}^{L} \frac{\mu_i}{n} (I(C_1, \cdots, C_{i-1}; C_i) + I(X_i^n; C_1, \cdots, C_i) - I(X_i^n; C_1, \cdots, C_{i-1})). \tag{2.14}$$

Let Z_i^n be a zero-mean Gaussian random vector with i.i.d. entries of positive variance $\sigma_{Z_i}^2$, $i = 1, \cdots, L-1$; moreover, we assume Z_i^n is independent of $(X_{i+1}^n, C_1, \cdots, C_{i+1})$, $i = 1, \cdots, L-1$. Note that

$$I(C_1, \cdots, C_{i-1}; C_i)$$

$$= I(X_i^n + Z_{i-1}^n; C_1, \cdots, C_{i-1}) + I(X_i^n + Z_{i-1}^n; C_i)$$

$$+ I(C_1, \cdots, C_{i-1}; C_i | X_i^n + Z_{i-1}^n) - I(X_i^n + Z_{i-1}^n; C_1, \cdots, C_i)$$

$$\geq I(X_i^n + Z_{i-1}^n; C_1, \cdots, C_{i-1}) + I(X_i^n + Z_{i-1}^n; C_i)$$

$$- I(X_i^n + Z_{i-1}^n; C_1, \cdots, C_i), \quad i = 2, \cdots, L. \tag{2.15}$$

Continuing from (2.14),

$$
\begin{aligned}
&\sum_{i=1}^{L} \mu_i (R_i + \epsilon) \\
&\geq \frac{\mu_1}{n} I(X_1^n; C_1) + \sum_{i=2}^{L} \frac{\mu_i}{n} (I(X_i^n + Z_{i-1}^n; C_1, \cdots, C_{i-1}) + I(X_i^n + Z_{i-1}^n; C_i) \\
&\quad - I(X_i^n + Z_{i-1}^n; C_1, \cdots, C_i) + I(X_i^n; C_1, \cdots, C_i) - I(X_i^n; C_1, \cdots, C_{i-1})) \qquad (2.16) \\
&= \frac{\mu_1}{n} I(X_1^n, C_1) - \frac{\mu_2}{n} (I(X_2^n; C_1) - I(X_2^n + Z_1^n; C_1)) \\
&\quad + \sum_{i=2}^{L-1} \Big(\frac{\mu_i}{n} (I(X_i^n; C_1, \cdots, C_i) - I(X_i^n + Z_{i-1}^n; C_1, \cdots, C_i)) \\
&\quad - \frac{\mu_{i+1}}{n} (I(X_{i+1}^n; C_1, \cdots, C_i) - I(X_{i+1}^n + Z_i^n; C_1, \cdots, C_i)) \Big) \\
&\quad + \frac{\mu_L}{n} (I(X_L^n; C_1, \cdots, C_L) - I(X_L^n + Z_{L-1}^n; C_1, \cdots, C_L)) + \sum_{i=2}^{L} \frac{\mu_i}{n} I(X_i^n + Z_{i-1}^n; C_i) \\
&= \frac{\mu_1}{2} \log(2\pi e \sigma_{X_1}^2) - \frac{\mu_2}{2} \log \Big(\frac{\sigma_{X_2}^2}{\sigma_{X_2}^2 + \sigma_{Z_1}^2} \Big) - \frac{\mu_1}{n} h(X_1^n | C_1) \\
&\quad - \frac{\mu_2}{n} (h(X_2^n + Z_1^n | C_1) - h(X_2^n | C_1)) \\
&\quad + \sum_{i=2}^{L-1} \Big(\frac{\mu_i}{2} \log \Big(\frac{\sigma_{X_i}^2}{\sigma_{X_i}^2 + \sigma_{Z_{i-1}}^2} \Big) - \frac{\mu_{i+1}}{2} \log \Big(\frac{\sigma_{X_{i+1}}^2}{\sigma_{X_{i+1}}^2 + \sigma_{Z_i}^2} \Big) \\
&\quad + \frac{\mu_i}{n} (h(X_i^n + Z_{i-1}^n | C_1, \cdots, C_i) - h(X_i^n | C_1, \cdots, C_i)) \\
&\quad - \frac{\mu_{i+1}}{n} (h(X_{i+1}^n + Z_i^n | C_1, \cdots, C_i) - h(X_{i+1}^n | C_1, \cdots, C_i)) \Big) \\
&\quad + \frac{\mu_L}{2} \log \Big(\frac{\sigma_{X_L}^2}{\sigma_{X_L}^2 + \sigma_{Z_{L-1}}^2} \Big) + \frac{\mu_L}{n} (h(X_L^n + Z_{L-1}^n | C_1, \cdots, C_L) - h(X_L^n | C_1, \cdots, C_L)) \\
&\quad + \sum_{i=2}^{L} \Big(\frac{\mu_i}{2} \log(2\pi e (\sigma_{X_i}^2 + \sigma_{Z_{i-1}}^2)) - \frac{\mu_i}{n} h(X_i^n + Z_{i-1}^n | C_i) \Big), \qquad (2.17)
\end{aligned}
$$

where (2.16) is due to (2.15). Note that

$$
\begin{aligned}
&-\mu_1 h(X_1^n|C_1) - \mu_2(h(X_2^n + Z_1^n|C_1) - h(X_2^n|C_1)) \\
&= -\mu_1 h(X_1^n|C_1) - \mu_2(h(X_1^n + \Delta_1^n + Z_1^n|C_1) - h(X_1^n + \Delta_1^n|C_1)) \\
&\geq \frac{-\mu_1 n}{2}\log(2\pi e(d_1+\epsilon)) - \frac{\mu_2 n}{2}\log\Big(\frac{d_1+\epsilon+\sigma^2_{\Delta_1}+\sigma^2_{Z_1}}{d_1+\epsilon+\sigma^2_{\Delta_1}}\Big) \qquad (2.18) \\
&= \min_{\gamma_1\in[0,d_1+\epsilon]} \frac{-\mu_1 n}{2}\log(2\pi e\gamma_1) - \frac{\mu_2 n}{2}\log\Big(\frac{\gamma_1+\sigma^2_{\Delta_1}+\sigma^2_{Z_1}}{\gamma_1+\sigma^2_{\Delta_1}}\Big), \qquad (2.19)
\end{aligned}
$$

where (2.18) follows from Lemma 1 and 2 in Section 2.1. Moreover, we have

$$
\begin{aligned}
&\mu_i(h(X_i^n + Z_{i-1}^n|C_1,\cdots,C_i) - h(X_i^n|C_1,\cdots,C_i)) \\
&-\mu_{i+1}(h(X_{i+1}^n + Z_i^n|C_1,\cdots,C_i) - h(X_{i+1}^n|C_1,\cdots,C_i)) \\
&= \mu_i(h(X_i^n + Z_{i-1}^n|C_1,\cdots,C_i) - h(X_i^n|C_1,\cdots,C_i)) \\
&\quad -\mu_{i+1}(h(X_i^n + \Delta_i^n + Z_i^n|C_1,\cdots,C_i) - h(X_i^n + \Delta_i^n|C_1,\cdots,C_i)) \\
&\geq \min_{\gamma_i\in[0,d_i+\epsilon]} \frac{\mu_i n}{2}\log\Big(\frac{\gamma_i+\sigma^2_{Z_{i-1}}}{\gamma_i}\Big) - \frac{\mu_{i+1} n}{2}\log\Big(\frac{\gamma_i+\sigma^2_{\Delta_i}+\sigma^2_{Z_i}}{\gamma_i+\sigma^2_{\Delta_i}}\Big), \quad i=2,\cdots,L-1,
\end{aligned}
\qquad (2.20)
$$

where (2.20) follows from Theorem 5. Note that

$$
\begin{aligned}
&h(X_L^n + Z_{L-1}^n|C_1,\cdots,C_L) - h(X_L^n|C_1,\cdots,C_L) \\
&\geq \frac{n}{2}\log\Big(\frac{d_L+\epsilon+\sigma^2_{Z_{L-1}}}{d_L+\epsilon}\Big) \qquad (2.21) \\
&= \min_{\gamma_L\in[0,d_L+\epsilon]} \frac{n}{2}\log\Big(\frac{\gamma_L+\sigma^2_{Z_{L-1}}}{\gamma_L}\Big), \qquad (2.22)
\end{aligned}
$$

where (2.21) is due to Lemma 2. In view of Lemma 1 and the fact that $\sigma^2_{X_i^n+Z_{i-1}^n|C_i} =$

$\sigma^2_{X_i^n|C_i} + \sigma^2_{Z_{i-1}} \leq \delta_i + \epsilon + \sigma^2_{Z_{i-1}}$, we have

$$-h(X_i^n + Z_{i-1}^n|C_i) \geq -\frac{n}{2}\log(2\pi e(\delta_i + \epsilon + \sigma^2_{Z_{i-1}})), \quad i = 2, \cdots, L. \tag{2.23}$$

Substituting (2.19), (2.20), (2.22), (2.23) into (2.17) yields

$$\begin{aligned}&\sum_{i=1}^{L}\mu_i(R_i+\epsilon)\\ &\geq \min_{\gamma_i\in[0,d_i+\epsilon],i=1,\cdots,L}\frac{\mu_1}{2}\log\left(\frac{\sigma^2_{X_1}}{\gamma_1}\right)+\sum_{i=2}^{L}\frac{\mu_i}{2}\log\frac{(\sigma^2_{X_i}+\sigma^2_{Z_{i-1}})(\gamma_{i-1}+\sigma^2_{\Delta_{i-1}})(\gamma_i+\sigma^2_{Z_{i-1}})}{(\gamma_{i-1}+\sigma^2_{\Delta_{i-1}}+\sigma^2_{Z_{i-1}})(\delta_i+\epsilon+\sigma^2_{Z_{i-1}})\gamma_i}.\end{aligned} \tag{2.24}$$

Replacing $\sigma^2_{Z_i}$ with $(\theta_i^{-1} - \sigma^{-2}_{X_{i+1}})^{-1}$, $i = 1, \cdots, L-1$, in (2.24) gives

$$\sum_{i=1}^{L}\mu_i(R_i+\epsilon) \geq \min_{\gamma_i\in[0,d_i+\epsilon],i=1,\cdots,L}\psi(\overline{\mu},\overline{\gamma},\overline{\delta^{(\epsilon)}},\overline{\theta}),$$

where $\overline{\delta^{(\epsilon)}} = (\delta_2 + \epsilon, \cdots, \delta_L + \epsilon)$. Note that there is a one-to-one correspondence between $\sigma^2_{Z_i} \in (0, \infty)$ and $\theta_i \in (0, \sigma^2_{X_{i+1}})$, $i = 1, \cdots, L-1$. Let $\overline{\gamma^{(\epsilon_m)}}$ be a minimizer to

$$\min_{\gamma_i\in[0,d_i+\epsilon_m],i=1,\cdots,L}\psi(\overline{\mu},\overline{\gamma},\overline{\delta^{(\epsilon_m)}},\overline{\theta}),$$

where $\epsilon_m > 0$, and ϵ_m tends to zero as $m \to \infty$. Without loss of generality, we assume that $\overline{\gamma^{(\epsilon_m)}}$ converges to some $\overline{\gamma^*}$ as $m \to \infty$ (otherwise, one can take a converging

subsequence of $\overline{\gamma^{(\epsilon_m)}}$, $m \geq 1$). We have

$$\begin{aligned}
\overline{\mu}\overline{R}^T &\geq \lim_{m\to\infty} \sum_{i=1}^{L} \mu_i (R_i + \epsilon_m) \\
&\geq \lim_{m\to\infty} \psi(\overline{\mu}, \overline{\gamma^{(\epsilon_m)}}, \overline{\delta^{(\epsilon_m)}}, \overline{\theta}) \\
&= \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta}) \\
&\geq \min_{\gamma_i \in [0, d_i], i=1,\cdots,L} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}).
\end{aligned}$$

This completes the proof of (2.12).

Chapter 3

Upper Bound

In order to prove Theorem 2, it suffices to prove that

$$\inf_{\overline{R}\in\mathcal{R}_P(\overline{d},\overline{\delta})} \overline{\mu}\overline{R}^T \leq \max_{\theta_i\in[0,\sigma^2_{X_{i+1}}],i=1,\cdots,L-1} \psi(\overline{\mu},\overline{\gamma},\overline{\delta},\overline{\theta}) \tag{3.1}$$

for all $\overline{\gamma}$ with $\gamma_i \in (0, d_i)$, $i = 1, \cdots, L$.

The maximization problem in (3.1) can be decomposed into

$$\max_{\theta_i\in[0,\sigma^2_{X_{i+1}}]} \frac{(\sigma^2_{X_{i+1}}-\theta_i)\gamma_{i+1}+\sigma^2_{X_{i+1}}\theta_i}{((\sigma^2_{X_{i+1}}-\theta_i)(\gamma_i+\sigma^2_{\Delta_i})+\sigma^2_{X_{i+1}}\theta_i)((\sigma^2_{X_{i+1}}-\theta_i)\delta_{i+1}+\sigma^2_{X_{i+1}}\theta_i)}, \tag{3.2}$$

for $i = 1, \cdots, L-1$.

The maximizers to (3.2) are characterized by the following result [10, Lemma 2].

Lemma 4. *For $i = 1, \cdots, L-1$, define*

$$\gamma^{\diamond}_{i+1} = \Big(\frac{1}{\gamma_i+\sigma^2_{\Delta_i}} + \frac{1}{\delta_{i+1}} - \frac{1}{\sigma^2_{X_{i+1}}}\Big)^{-1},$$

$$\gamma^{\star}_{i+1} = \gamma_i + \sigma^2_{\Delta_i} + \delta_{i+1} - \sigma^2_{X_{i+1}}.$$

1. *If* $\max\{\gamma_i + \sigma^2_{\Delta_i}, \delta_{i+1}\} < \sigma^2_{X_{i+1}}$, *then* $\gamma^\diamond_{i+1} > \gamma^\star_{i+1}$ *and the maximizers to (3.2) are given by*

$$\theta_i = \begin{cases} 0, & \gamma_{i+1} \geq \gamma^\diamond_{i+1} \\ \sigma^2_{X_{i+1}}, & \gamma_{i+1} \leq \gamma^\star_{i+1} \\ \hat{\theta}_i, & \textit{otherwise} \end{cases}, \quad i = 1, \cdots, L-1, \tag{3.3}$$

where

$$\hat{\theta}_i \triangleq \sqrt{\Big(\frac{\sigma^2_{X_{i+1}}(\gamma_i + \sigma^2_{\Delta_i})}{\sigma^2_{X_{i+1}} - \gamma_i - \sigma^2_{\Delta_i}} - \frac{\sigma^2_{X_{i+1}}\gamma_{i+1}}{\sigma^2_{X_{i+1}} - \gamma_{i+1}}\Big)\Big(\frac{\sigma^2_{X_{i+1}}\delta_{i+1}}{\sigma^2_{X_{i+1}} - \delta_{i+1}} - \frac{\sigma^2_{X_{i+1}}\gamma_{i+1}}{\sigma^2_{X_{i+1}} - \gamma_{i+1}}\Big)} - \frac{\sigma^2_{X_{i+1}}\gamma_{i+1}}{\sigma^2_{X_{i+1}} - \gamma_{i+1}}$$

is the unique solution to the following equation

$$\Big(\frac{\sigma^2_{X_{i+1}}\gamma_{i+1}}{\sigma^2_{X_{i+1}} - \gamma_{i+1}} + \hat{\theta}_i\Big)^{-1} = \Big(\frac{\sigma^2_{X_{i+1}}(\gamma_i + \sigma^2_{\Delta_i})}{\sigma^2_{X_{i+1}} - \gamma_i - \sigma^2_{\Delta_i}} + \hat{\theta}_i\Big)^{-1} + \Big(\frac{\sigma^2_{X_{i+1}}\delta_{i+1}}{\sigma^2_{X_{i+1}} - \delta_{i+1}} + \hat{\theta}_i\Big)^{-1}$$

for $\hat{\theta}_i \in (0, \sigma^2_{X_{i+1}})$, $i = 1, \cdots, L-1$. *The maximizers* θ_i, $i = 1, \cdots, L-1$, *given in (3.3) are monotonically increasing continuous functions of* γ_i *and monotonically decreasing continuous functions of* γ_{i+1}*; furthermore, the monotonicity is strict when* $\theta_i \in (0, \sigma^2_{X_{i+1}})$.

2. *If* $\max\{\gamma_i + \sigma^2_{\Delta_i}, \delta_{i+1}\} = \sigma^2_{X_{i+1}}$, *then* $\gamma^\diamond_{i+1} = \gamma^\star_{i+1} = \min\{\gamma_i + \sigma^2_{\Delta_i}, \delta_{i+1}\}$ *and the*

maximizers to (3.2) are given by

$$\theta_i = \begin{cases} 0, & \gamma_{i+1} > \min\{\gamma_i + \sigma^2_{\Delta_i}, \delta_{i+1}\} \\ \sigma^2_{X_{i+1}}, & \gamma_{i+1} < \min\{\gamma_i + \sigma^2_{\Delta_i}, \delta_{i+1}\} \\ \textit{any number in } [0, \sigma^2_{X_{i+1}}], & \textit{otherwise} \end{cases}, \quad i = 1, \cdots, L-1. \tag{3.4}$$

The proof of the following lemma is essentially the same as that of [10, Lemma 3] and thus is omitted.

Lemma 5. *There exist* $\overline{\gamma'} \triangleq (\gamma'_1, \cdots, \gamma'_L)$ *and* $\overline{\delta'} \triangleq (\delta'_2, \cdots, \delta'_L)$ *with*

$$0 < \gamma'_i \leq \gamma_i, \quad i = 1, \cdots, L,$$

$$0 < \delta'_i \leq \delta_i, \quad i = 2, \cdots, L,$$

$$\gamma'_i + \sigma^2_{\Delta_i} + \delta'_{i+1} - \sigma^2_{X_{i+1}} \leq \gamma'_{i+1} \leq \left(\frac{1}{\gamma'_i + \sigma^2_{\Delta_i}} + \frac{1}{\delta'_{i+1}} - \frac{1}{\sigma^2_{X_{i+1}}}\right)^{-1}, \quad i = 1, \cdots, L-1,$$

such that

$$\max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma'}, \overline{\delta'}, \overline{\theta}) \leq \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}).$$

3.1 An Inner Bound of $\mathcal{R}_P(\overline{d}, \overline{\delta})$

The following lemma provides an inner bound of $\mathcal{R}_P(\overline{d}, \overline{\delta})$.

Lemma 6. *Let* $(U_1, \cdots, U_L)$ *be jointly Gaussian with* $(X_1, \cdots, X_L)$ *such that*

1. $(X_2, \cdots, X_L) \leftrightarrow X_1 \leftrightarrow U_1$ *form a Markov chain,*

   and $(X_j)_{j \neq i} \leftrightarrow (X_i, U_1, \cdots, U_{i-1}) \leftrightarrow U_i$ *form a Markov chain,* $i = 1, \cdots, L$,

2. $\sigma^2_{X_i|U_1,\cdots,U_i} \leq d_i$, $i = 1, \cdots, L$, *and* $\sigma^2_{X_i|U_i} \leq \delta_i$, $i = 2, \cdots, L$.

Then $\overline{R} \in \mathcal{R}_P(\overline{d}, \overline{\delta})$ *for any* $\overline{R}$ *satisfying*

$$R_1 \geq I(X_1; U_1),$$
$$R_i \geq I(X_i, U_1, \cdots, U_{i-1}; U_i), \quad i = 2, \cdots, L.$$

Proof: Consider L discrete memoryless sources $X_1, \cdots, X_L$ with joint probability mass function $p(x_1, \cdots, x_L)$. Let $w_i(\cdot, \cdot)$ be a bounded distortion measure on $\mathcal{X}_i \times \hat{\mathcal{X}}_i$, where both $\mathcal{X}_i$ and $\hat{\mathcal{X}}_i$ are finite, $i = 1, \cdots, L$. We shall show that if there exist auxiliary random variables U_i (over finite alphabet $\mathcal{U}_i$), $i = 1, \cdots, L$, and functions $\tilde{g}_i : \mathcal{U}_1 \times \cdots \times \mathcal{U}_i \to \hat{\mathcal{X}}_i$, $i = 1, \cdots, L$, and $\hat{g}_i : \mathcal{U}_i \to \hat{\mathcal{X}}_i$, $i = 2, \cdots, L$, such that

P1) $(X_2, \cdots, X_L) \leftrightarrow X_1 \leftrightarrow U_1$ form a Markov chain,
and $(X_j)_{j \neq i} \leftrightarrow (X_i, U_1, \cdots, U_{i-1}) \leftrightarrow U_i$ form a Markov chain, $i = 1, \cdots, L$,

P2) $\mathbb{E}[w_i(X_i, \tilde{g}_i(U_1, \cdots, U_i))] \leq d_i$, $i = 1, \cdots, L$, and $\mathbb{E}[w_i(X_i, \hat{g}_i(U_i))] \leq \delta_i$, $i = 2, \cdots, L$,

then $\overline{R} \in \mathcal{R}_P(\overline{d}, \overline{\delta})$ for any $\overline{R}$ satisfying

$$R_1 \geq I(X_1; U_1),$$
$$R_i \geq I(X_i, U_1, \cdots, U_{i-1}; U_i), \quad i = 2, \cdots, L.$$

One can readily extend this result to the quadratic Gaussian case via a discretization procedure and certain limiting arguments [21].

As the proof is based on the standard techniques in network information theory, we only give a sketch here. We adopt the notation in [21].

Codebook generation:

Fix a conditional probability mass function $p(u_1, \cdots, u_L|x_1, \cdots, x_L)$ and functions $\tilde{g}_i : \mathcal{U}_1 \times \cdots \times \mathcal{U}_i \to \hat{\mathcal{X}}_i$, $i = 1, \cdots, L$, and $\hat{g}_i : \mathcal{U}_i \to \hat{\mathcal{X}}_i$, $i = 2, \cdots, L$, such that P1) and P2) are satisfied. Note that P1) is satisfied if $p(u_1, \cdots, u_L|x_1, \cdots, x_L)$ factors as

$$p(u_1, \cdots, u_L|x_1, \cdots, x_L) = p(u_1|x_1) \prod_{i=2}^{L} p(u_i|x_i, u_1, \cdots, u_{i-1}).$$

For $i = 1, \cdots, L$, randomly and independently generate e^{nR_i} sequences $u_i^n(c_i)$, $c_i \in [1 : e^{nR_i}]$, each according to $\prod_{j=1}^{n} p_{U_i}(u_{ij})$. The codebook is revealed to the encoders and the decoder.

Encoding:

Given x_1^n, Encoder 1 finds an index $c_1 \in [1 : e^{nR_1}]$ such that $(x_1^n, u_1^n(c_1)) \in \mathcal{T}_{\epsilon_1}^{(n)}$; if there is more than one such index, it picks the smallest one among them; if there is no such index, it sets $c_1 = 1$. For $i = 2, \cdots, L$, given $(x_i^n, c_1, \cdots, c_{i-1})$, Encoder i finds an index $c_i \in [1 : e^{nR_i}]$ such that $(x_i^n, u_1^n(c_1), \cdots, u_i^n(c_i)) \in \mathcal{T}_{\epsilon_i}^{(n)}$; if there is more than one such index, it picks the smallest one among them; if there is no such index, it sets $c_i = 1$. Here we assume that $\epsilon_L > \cdots > \epsilon_1 > 0$. The indices $c_1, \cdots, c_L$ are then sent to the decoder.

Decoding:

For $i = 1, \cdots, L$, given $(c_1, \cdots, c_i)$, the decoder computes $\tilde{x}_{ij} \triangleq \tilde{g}_i(u_{1j}(c_1), \cdots, u_{ij}(c_i))$, $j = 1, \cdots, n$, and uses $\tilde{x}_i^n$ as the reconstruction of x_i^n. For $i = 2, \cdots, L$, given c_i, the decoder computes $\hat{x}_{ij} \triangleq \hat{g}_i(u_{ij}(c_i))$, $j = 1, \cdots, n$, and uses $\hat{x}_i^n$ as the reconstruction of x_i^n.

Error analysis:

Let C_i denote the output of Encoder i, $i = 1, \cdots, L$. By the covering lemma [21,

Lemma 3.3, p. 62], $\mathbb{P}\{(X_1^n, U_1^n(C_1)) \in \mathcal{T}_{\epsilon_1}^{(n)}\}$ tends to one as $n \to \infty$ if $R_1 > I(X_1; U_1) + \delta_1(\epsilon_1)$, where $\delta_1(\epsilon_1)$ tends to zero as $\epsilon_1 \to 0$. Then it follows from the conditional typicality lemma [21, p. 27] that $\mathbb{P}\{(X_1^n, \cdots, X_L^n, U_1^n(C_1)) \in \mathcal{T}_{\epsilon_2}^{(n)}\}$ tends to one as $n \to \infty$. For $i = 2, \cdots, L$, by the covering lemma, $\mathbb{P}\{(X_i^n, U_1^n(C_1), \cdots, U_i^n(C_i)) \in \mathcal{T}_{\epsilon_i}^{(n)}\}$ tends to one as $n \to \infty$ if $R_i > I(X_i, U_1, \cdots, U_{i-1}; U_i) + \delta_i(\epsilon_i)$, where $\delta_i(\epsilon_i)$ tends to zero as $\epsilon_i \to 0$; furthermore, it follows from [21, Lemma 12.3, p. 299] and the Markov lemma [21, Lemma 12.1, p. 296] that $\mathbb{P}\{(X_1^n, \cdots, X_L^n, U_1^n(C_1), \cdots, U_i^n(C_i)) \in \mathcal{T}_{\epsilon_{i+1}}^{(n)}\}$ tends to one as $n \to \infty$ if ϵ_i is sufficiently small compared to ϵ_{i+1}. Therefore, for every $\epsilon > 0$ and every $n \geq n(\epsilon)$ (with $n(\epsilon)$ determined by ϵ), there exists a deterministic codebook conditioned on which the probability of $(X_1^n, \cdots, X_L^n, U_1^n(C_1), \cdots, U_L^n(C_L)) \notin \mathcal{T}_{\epsilon_{L+1}}^{(n)}$ is less than ϵ. Now one can readily complete the proof by invoking the typical average lemma [21, p. 26]. □

3.2 Proof of the Upper Bound

For the purpose of proving Theorem 2, it suffices to construct $(U_1, \cdots, U_L)$ satisfying the conditions in Lemma 6 such that

$$\mu_1 I(X_1; U_1) + \sum_{i=2}^{L} \mu_i I(X_i, U_1, \cdots, U_{i-1}; U_i) = \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma'}, \overline{\delta'}, \overline{\theta}), \quad (3.5)$$

where $\overline{\gamma'}$ and $\overline{\delta'}$ are specified in Lemma 5.

Now, we are ready to construct $(U_1, \cdots, U_L)$.

Define

$$\xi_i = \begin{cases} 0, & \max\{\gamma_i' + \sigma^2_{\Delta_i}, \delta_{i+1}'\} = \sigma^2_{X_{i+1}} \\ \frac{\sigma^4_{X_{i+1}}(\sigma^2_{X_{i+1}}+\varsigma_i-\sqrt{(\tau_i-\varsigma_i)(\lambda_i-\varsigma_i)})}{(\sigma^2_{X_{i+1}}+\tau_i)(\sigma^2_{X_{i+1}}+\lambda_i)}, & \text{otherwise} \end{cases}, \tag{3.6}$$

$i = 1, \cdots, L-1,$

$$a_i = \begin{cases} 0, & \gamma_i' + \sigma^2_{\Delta_i} = \sigma^2_{X_{i+1}} \\ \frac{(\sigma^2_{X_{i+1}}+\tau_i)\sqrt{\lambda_i-\varsigma_i}}{(\sigma^2_{X_{i+1}}+\varsigma_i)(\sqrt{\tau_i-\varsigma_i}+\sqrt{\lambda_i-\varsigma_i})}, & \text{otherwise} \end{cases}, \quad i = 1, \cdots, L-1, \tag{3.7}$$

$$b_i = \begin{cases} 0, & \delta_{i+1}' = \sigma^2_{X_{i+1}} \\ \frac{(\sigma^2_{X_{i+1}}+\lambda_i)\sqrt{\tau_i-\varsigma_i}}{(\sigma^2_{X_{i+1}}+\varsigma_i)(\sqrt{\tau_i-\varsigma_i}+\sqrt{\lambda_i-\varsigma_i})}, & \text{otherwise} \end{cases}, \quad i = 1, \cdots, L-1, \tag{3.8}$$

$$\theta_i' = \begin{cases} \text{any number in } [0, \sigma^2_{X_{i+1}}], & \max\{\gamma_i' + \sigma^2_{\Delta_i}, \delta_{i+1}'\} = \sigma^2_{X_{i+1}} \\ \sqrt{(\tau_i-\varsigma_i)(\lambda_i-\varsigma_i)} - \varsigma_i, & \text{otherwise} \end{cases}, \quad i = 1, \cdots, L-1, \tag{3.9}$$

where

$$\tau_i = \frac{\sigma^2_{X_{i+1}}(\gamma_i' + \sigma^2_{\Delta_i})}{\sigma^2_{X_{i+1}} - \gamma_i' - \sigma^2_{\Delta_i}}, \quad i = 1, \cdots, L-1,$$
$$\lambda_i = \frac{\sigma^2_{X_{i+1}}\delta_{i+1}'}{\sigma^2_{X_{i+1}} - \delta_{i+1}'}, \quad i = 1, \cdots, L-1,$$
$$\varsigma_i = \frac{\sigma^2_{X_{i+1}}\gamma_{i+1}'}{\sigma^2_{X_{i+1}} - \gamma_{i+1}'}, \quad i = 1, \cdots, L-1.$$

We assume $\mathbb{E}[U_i] = 0$, $i = 1, \cdots, L$. Let U_1 be jointly Gaussian with $(X_1, \cdots, X_L)$ such that

$$\mathbb{E}[U_1^2] = \mathbb{E}[X_1 U_1] = \sigma^2_{X_1} - \gamma_1',$$

and $(X_2, \cdots, X_L) \leftrightarrow X_1 \leftrightarrow U_1$ form a Markov chain. Now let U_i be jointly Gaussian with $(X_1, \cdots, X_L, U_1, \cdots, U_{i-1})$ such that

$$\mathbb{E}[U_i^2] = \mathbb{E}[X_i U_i] = \sigma_{X_i}^2 - \delta_i',$$
$$\mathbb{E}[\tilde{U}_{i-1} U_i] = \xi_{i-1},$$

and $((X_j)_{j \neq i}, U_1, \cdots, U_{i-1}) \leftrightarrow (X_i, \tilde{U}_{i-1}) \leftrightarrow U_i$ form a Markov chain, $i = 2, \cdots, L$, where

$$\tilde{U}_1 = U_1,$$
$$\tilde{U}_i = a_{i-1}\tilde{U}_{i-1} + b_{i-1}U_i, \quad i = 2, \cdots, L.$$

It is clear that the covariance matrix of (X_1, U_1) is positive semidefinite; moreover, one can readily verify that the covariance matrix of $(X_i, \tilde{U}_{i-1}, U_i)$ is positive semidefinite, $i = 2, \cdots, L$. As a consequence, the joint distribution of $(X_1, \cdots, X_L)$ and the constructed $(U_1, \cdots, U_L)$ (as well as the induced $(\tilde{U}_1, \cdots, \tilde{U}_L)$) is well defined. It can be verified that

$$\tilde{U}_1 = \mathbb{E}[X_1|U_1],$$
$$\tilde{U}_i = \mathbb{E}[X_i|U_1, \cdots, U_i] = \mathbb{E}[X_i|\tilde{U}_{i-1}, U_i], \quad i = 2, \cdots, L,$$
$$\mathbb{E}[(X_i - \tilde{U}_i)^2] = \gamma_i', \quad i = 1, \cdots, L,$$
$$\mathbb{E}[(X_i - U_i)^2] = \delta_i', \quad i = 2, \cdots, L.$$

Therefore, the constructed $(U_1, \cdots, U_L)$ satisfies the conditions in Lemma 6. Note

that

$$I(X_1;U_1) = \frac{1}{2}\log\Big(\frac{\sigma_{X_1}^2}{\gamma_1'}\Big);$$

moreover, we have

$$\begin{aligned}
&I(X_i, U_1, \cdots, U_{i-1}; U_i) \\
&= I(X_i, \tilde{U}_{i-1}; U_i) \qquad (3.10)\\
&= I(\tilde{U}_{i-1}; U_i) + I(X_i; U_i|\tilde{U}_{i-1}) \\
&= \frac{1}{2}\log\Big(\frac{\sigma_{X_i}^4(\gamma_{i-1}' + \sigma_{\Delta_{i-1}}^2)((\sigma_{X_i}^2 - \theta_{i-1}')\gamma_i' + \sigma_{X_i}^2\theta_{i-1}')}{((\sigma_{X_i}^2 - \theta_{i-1}')(\gamma_{i-1} + \sigma_{\Delta_{i-1}}^2) + \sigma_{X_i}^2\theta_{i-1}')((\sigma_{X_i}^2 - \theta_{i-1}')\delta_i' + \sigma_{X_i}^2\theta_{i-1}')\gamma_i'}\Big), \qquad (3.11)
\end{aligned}$$

where (3.10) is due to the fact that $(X_i, U_1, \cdots, U_{i-1}) \leftrightarrow (X_i, \tilde{U}_i) \leftrightarrow U_i$ form a Markov chain and that $\tilde{U}_i$ is a (linear) function of $(U_1, \cdots, U_{i-1})$, and (3.11) is by direct evaluation (see (3.9) for the definition of θ_i', $i = 1, \cdots, L-1$). Now one can readily prove (3.5) by invoking Lemma 4. This completes the proof of Theorem 2.

Chapter 4

Saddle Point Analysis

Without loss of generality, we assume $\mu_1 \geq \cdots \geq \mu_L > 0$ throughout this chapter. Note that the minimization problem

$$\min_{\gamma_i \in [0,d_i], i=1,\cdots,L} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta})$$

can be decomposed into

$$\min_{\gamma_1 \in [0,d_1]} \frac{\mu_1}{2} \log\Big(\frac{\sigma_{X_1}^2}{\gamma_1}\Big) + \frac{\mu_2}{2} \log\Big(\frac{\gamma_1 + \sigma_{\Delta_1}^2}{(\sigma_{X_2}^2 - \theta_1)(\gamma_1 + \sigma_{\Delta_1}^2) + \sigma_{X_2}^2 \theta_1}\Big), \tag{4.1}$$

$$\begin{aligned} &\min_{\gamma_i \in [0,d_i]} \frac{\mu_i}{2} \log\Big(\frac{(\sigma_{X_i}^2 - \theta_{i-1})\gamma_i + \sigma_{X_i}^2 \theta_{i-1}}{\gamma_i}\Big) \\ &\qquad + \frac{\mu_{i+1}}{2} \log\Big(\frac{\gamma_i + \sigma_{\Delta_i}^2}{(\sigma_{X_{i+1}}^2 - \theta_i)(\gamma_i + \sigma_{\Delta_i}^2) + \sigma_{X_{i+1}}^2 \theta_i}\Big), \quad i = 2, \cdots, L-1, \end{aligned} \tag{4.2}$$

$$\min_{\gamma_L \in [0,d_L]} \frac{\mu_L}{2} \log\Big(\frac{(\sigma_{X_L}^2 - \theta_{L-1})\gamma_L + \sigma_{X_L}^2 \theta_{L-1}}{\gamma_L}\Big). \tag{4.3}$$

The minimizers to (4.1), (4.2), and (4.3) are characterized by the following lemmas. Only the proof of Lemma 8 is provided. The proofs of Lemma 7 and Lemma 9 are

straightforward and thus omitted.

Lemma 7. *The minimizer to (4.1) is given by* $\gamma_1 = d_1$.

Lemma 8. *For* $i = 2, \cdots, L-1$, *define*

$$\tilde{a}_i = \mu_{i+1}(\sigma^2_{X_i} - \theta_{i-1})\sigma^2_{X_{i+1}}\theta_i - \mu_i(\sigma^2_{X_{i+1}} - \theta_i)\sigma^2_{X_i}\theta_{i-1}, \tag{4.4}$$

$$\tilde{b}_i = (\mu_{i+1} - \mu_i)\sigma^2_{X_i}\sigma^2_{X_{i+1}}\theta_{i-1}\theta_i - 2\mu_i\sigma^2_{X_i}\theta_{i-1}\sigma^2_{\Delta_i}(\sigma^2_{X_{i+1}} - \theta_i), \tag{4.5}$$

$$\tilde{c}_i = -\mu_i\sigma^2_{X_i}\theta_{i-1}\sigma^2_{\Delta_i}(\sigma^2_{X_{i+1}}\sigma^2_{\Delta_i} + \sigma^2_{X_i}\theta_i). \tag{4.6}$$

The minimizers to (4.2) are given by

$$\gamma_i = \begin{cases} \min\{\hat{\gamma}_i, d_i\} & \tilde{a}_i > 0 \\ d_i, & \tilde{a}_i \leq 0, \theta_{i-1} \in (0, \sigma^2_{X_i}] \\ \textit{any number in } [0, d_i], & \theta_{i-1} = \theta_i = 0 \end{cases}, \quad i = 2, \cdots, L-1, \tag{4.7}$$

where

$$\hat{\gamma}_i = \frac{-\tilde{b}_i + \sqrt{\tilde{b}_i^2 - 4\tilde{a}_i\tilde{c}_i}}{2\tilde{a}_i}, \quad i = 2, \cdots, L-1. \tag{4.8}$$

Lemma 9. *The minimizer to (4.3) is given by*

$$\gamma_L = \begin{cases} d_L, & \theta_{L-1} \in (0, \sigma^2_{X_L}] \\ \textit{any number in } [0, d_L], & \theta_{L-1} = 0 \end{cases}. \tag{4.9}$$

Proof: Consider the following minimization problem

$$\min_{\gamma_i \in [0, d_i]} \phi(\gamma_i), \tag{4.10}$$

where

$$\phi(\gamma_i) = \frac{\mu_i}{2}\log\Big(\frac{(\sigma^2_{X_i}-\theta_{i-1})\gamma_i+\sigma^2_{X_i}\theta_{i-1}}{\gamma_i}\Big)+\frac{\mu_{i+1}}{2}\log\Big(\frac{\gamma_i+\sigma^2_{\Delta_i}}{(\sigma^2_{X_{i+1}}-\theta_i)(\gamma_i+\sigma^2_{\Delta_i})+\sigma^2_{X_{i+1}}\theta_i}\Big).$$

It is easy to verify that the objective function is a constant if $\theta_{i-1}=\theta_i=0$; moreover, the minimum in (4.10) is achieved at $\gamma_i=0$ if $\theta_{i-1}=0$ and $\theta_i\in(0,\sigma^2_{X_{i+1}}]$.

In the rest of the proof we shall assume $\theta_{i-1}\in(0,\sigma^2_{X_i}]$ (which implies that the minimum in (4.10) is not achieved at $\gamma_i=0$). Note that

$$\frac{\partial\phi(\gamma_i)}{\partial\gamma_i}=\frac{1}{\hbar}(\tilde{a}_i\gamma_i^2+\tilde{b}_i\gamma_i+\tilde{c}_i),$$

where $\tilde{a}_i$, $\tilde{b}_i$, and $\tilde{c}_i$ are defined in (4.4), (4.5), and (4.6), respectively, and

$$\hbar=2((\sigma^2_{X_i}-\theta_{i-1})\gamma_i+\sigma^2_{X_i}\theta_{i-1})\gamma_i(\gamma_i+\sigma^2_{\Delta_i})((\sigma^2_{X_{i+1}}-\theta_i)(\gamma_i+\sigma^2_{\Delta_i})+\sigma^2_{X_{i+1}}\theta_i).$$

It is clear that $\tilde{b}_i\le 0$, $\tilde{c}_i<0$, and $\hbar>0$ for $\gamma_i\in(0,d_i]$.

Now consider the following cases.

1. If $\tilde{a}_i>0$, then the equation

$$\tilde{a}_i\gamma_i^2+\tilde{b}_i\gamma_i+\tilde{c}_i=0$$

has a unique positive root at $\gamma_i=\hat{\gamma}_i$, where $\hat{\gamma}_i$ is defined in (4.8). We have $\frac{\partial\phi(\gamma_i)}{\partial\gamma_i}<0$ for $\gamma_i\in(0,\hat{\gamma}_i)$ and $\frac{\partial\phi(\gamma_i)}{\partial\gamma_i}>0$ for $\gamma_i>\hat{\gamma}_i$. As a consequence, the minimum in (4.10) is achieved at $\min\{\hat{\gamma}_i,d_i\}$.

2. If $\tilde{a}_i\le 0$, we have $\frac{\partial\phi(\gamma_i)}{\partial\gamma_i}<0$ for $\gamma_i>0$. As a consequence, the minimum in

(4.10) is achieved at d_i.

This completes the proof of Lemma 8. □

4.1 Proof of Theorem 3

Now we proceed to prove Theorem 3. The key step is to show the existence of a saddle point $(\overline{\gamma^*}, \overline{\theta^*})$ with the property that

$$\psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}) = \min_{\gamma_i \in [0, d_i], i=1,\cdots,L} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta^*}), \tag{4.11}$$

$$\psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}) = \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta}). \tag{4.12}$$

First consider the case where $d_i < \sigma^2_{X_i}$, $i = 1, \cdots, L$, and $\delta_i < \sigma^2_{X_i}$, $i = 2, \cdots, L$. Let $\gamma^*_1 = d_1$ and $\gamma^*_L = d_L$. Define

$$\omega(\gamma_2, \cdots, \gamma_{L-1}) = \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta})\big|_{\gamma_1 = \gamma^*_1, \gamma_L = \gamma^*_L}.$$

Note that $\omega(\gamma_2, \cdots, \gamma_{L-1})$ is a continuous function of $(\gamma_2, \cdots, \gamma_{L-1})$ for $\gamma_i \in (0, \sigma^2_{X_i}]$, $i = 2, \cdots, L-1$, and $\omega(\gamma_2, \cdots, \gamma_{L-1}) = \infty$ if $\gamma_i = 0$ for some i; moreover, it is clear from the proof of Theorem 2 that

$$\omega(\gamma_2, \cdots, \gamma_{L-1}) \geq \frac{\mu_i}{2} \log\left(\frac{\sigma^2_{X_i}}{\gamma_i}\right), \quad i = 2, \cdots, L-1.$$

Therefore, the minimum of $\omega(\gamma_2, \cdots, \gamma_{L-1})$ over $(\gamma_2, \cdots, \gamma_{L-1})$ with $\gamma_i \in [0, \sigma^2_{X_i}]$, $i = 2, \cdots, L-1$, is achieved at some $(\gamma^*_2, \cdots, \gamma^*_{L-1})$ satisfying $\gamma^*_i \in (0, \sigma^2_{X_i}]$, $i =$

$2, \cdots, L-1$. Let $\overline{\theta^*} \triangleq (\theta_1^*, \cdots, \theta_L^*)$ be a maximizer to

$$\max_{\theta_i \in [0, \sigma_{X_i}^2], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta}),$$

where $\overline{\gamma^*} = (\gamma_1^*, \cdots, \gamma_L^*)$. We shall prove that $(\overline{\gamma^*}, \overline{\theta^*})$ satisfies (4.11) (note that (4.12) is automatically satisfied).

In view of Lemma 7, Lemma 9, and our choice of (γ_1^*, γ_L^*), we just need to show that $(\gamma_2^*, \cdots, \gamma_{L-1}^*)$ satisfies the optimality condition (4.7) (with $\theta_i = \theta_i^*$, $i = 1, \cdots, L-1$) in Lemma 8. Let us assume that (4.7) is violated by $\gamma_{i^*}^*$ for some i^*. Note that $\theta_{i^*-1}^*$ and $\theta_{i^*}^*$ are determined by $\gamma_{i^*}^*$ according to (3.3). To stress this dependence, we denote $\theta_{i^*-1}^*$ and $\theta_{i^*}^*$ by $\theta_{i^*-1}^*(\gamma_{i^*}^*)$ and $\theta_{i^*}^*(\gamma_{i^*}^*)$, respectively. It is clear that at least one of $\theta_{i^*-1}^*(\gamma_{i^*}^*)$ and $\theta_{i^*}^*(\gamma_{i^*}^*)$ is not zero since otherwise (4.7) is satisfied by $\gamma_{i^*}^*$. Now let $\tilde{\gamma}_{i^*}(\gamma_{i^*}^*)$ be the minimizer determined by $\theta_{i^*-1}^*(\gamma_{i^*}^*)$ and $\theta_{i^*}^*(\gamma_{i^*}^*)$ according to (4.7). We shall move $\gamma_{i^*}^*$ toward $\tilde{\gamma}_{i^*}(\gamma_{i^*}^*)$ (and change $\theta_{i^*-1}^*(\gamma_{i^*}^*)$, $\theta_{i^*}^*(\gamma_{i^*}^*)$, and $\tilde{\gamma}_{i^*}(\gamma_{i^*}^*)$ correspondingly) while keeping γ_i^* $(i \neq i^*)$ and θ_i^* $(i \neq i^*-1$ and $i \neq i^*)$ fixed. It is shown in Appendix D that $\tilde{\gamma}_{i^*}(\gamma_{i^*}^*)$ varies continuously with $\gamma_{i^*}^*$ if at least one of $\theta_{i^*-1}^*(\gamma_{i^*}^*)$ and $\theta_{i^*}^*(\gamma_{i^*}^*)$ is not zero. As a consequence, we can keep moving $\gamma_{i^*}^*$ until $\gamma_{i^*}^* = \tilde{\gamma}_{i^*}^*$ at which one of the following cases happens:

1. $\tilde{\gamma}_{i^*}^* = \tilde{\gamma}_{i^*}(\tilde{\gamma}_{i^*}^*)$;

2. $\theta_{i^*-1}^*(\tilde{\gamma}_{i^*}^*) = \theta_{i^*}^*(\tilde{\gamma}_{i^*}^*) = 0$.

Clearly, $\tilde{\gamma}_{i^*}^*$ satisfies (4.7) with $\theta_{i^*-1} = \theta_{i^*-1}^*(\tilde{\gamma}_{i^*}^*)$ and $\theta_{i^*} = \theta_{i^*}^*(\tilde{\gamma}_{i^*}^*)$. Define $\overline{\gamma^\circ} = (\gamma_1^\circ, \cdots, \gamma_L^\circ)$ with $\gamma_{i^*}^\circ = \tilde{\gamma}_{i^*}^*$ and $\gamma_i^\circ = \gamma_i^*$ for $i \neq i^*$. Moreover, define $\overline{\theta^\circ} = (\theta_1^\circ, \cdots, \theta_L^\circ)$ with $\theta_{i^*-1}^\circ = \theta_{i^*-1}^*(\tilde{\gamma}_{i^*}^*)$, $\theta_{i^*}^\circ = \theta_{i^*}^*(\tilde{\gamma}_{i^*}^*)$, and $\theta_i^\circ = \theta_i^*$ for $i \neq i^*-1$ and $i \neq i^*$. If $\overline{\theta^\circ} \neq \overline{\theta^*}$,

then

$$\begin{aligned}
&\psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}) \\
&> \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^\circ}) \\
&\geq \psi(\overline{\mu}, \overline{\gamma^\circ}, \overline{\delta}, \overline{\theta^\circ}) \\
&= \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma^\circ}, \overline{\delta}, \overline{\theta}); \qquad (4.13)
\end{aligned}$$

if $\overline{\theta^\circ} = \overline{\theta^*}$, then

$$\begin{aligned}
&\psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}) \\
&> \psi(\overline{\mu}, \overline{\gamma^\circ}, \overline{\delta}, \overline{\theta^*}) \\
&= \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma^\circ}, \overline{\delta}, \overline{\theta}). \qquad (4.14)
\end{aligned}$$

Note that both (4.13) and (4.14) contradict with the fact that $(\gamma_2^*, \cdots, \gamma_{L-1}^*)$ achieves the minimum of $\omega(\gamma_2, \cdots, \gamma_{L-1})$ over $(\gamma_2, \cdots, \gamma_{L-1})$ with $\gamma_i \in [0, \sigma^2_{X_i}]$, $i = 2, \cdots, L-1$. Therefore, $(\overline{\gamma^*}, \overline{\theta^*})$ indeed satisfies (4.11) and thus is a saddle point.

Now consider the general case where $d_i \in (0, \sigma^2_{X_i}]$, $i = 1, \cdots, L$, and $\delta_i \in (0, \sigma^2_{X_i}]$, $i = 2, \cdots, L$. The preceding argument shows the existence of $(\overline{\gamma^{(k)}}, \overline{\theta^{(k)}})$ such that

$$\begin{aligned}
\psi(\overline{\mu}, \overline{\gamma^{(k)}}, \overline{\delta}, \overline{\theta^{(k)}}) &= \min_{\gamma_i \in [0, \rho_k d_i], i=1,\cdots,L} \psi(\overline{\mu}, \overline{\gamma}, \rho_k \overline{\delta}, \overline{\theta^{(k)}}), \\
\psi(\overline{\mu}, \overline{\gamma^{(k)}}, \overline{\delta}, \overline{\theta^{(k)}}) &= \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma^{(k)}}, \rho_k \overline{\delta}, \overline{\theta}),
\end{aligned}$$

where $\rho_k \in (0, 1)$, and ρ_k tends to one as $k \to \infty$. By taking a converging subsequence of $(\overline{\gamma^{(k)}}, \overline{\theta^{(k)}})$, $k \geq 1$, with its limit denoted by $(\overline{\gamma^*}, \overline{\theta^*})$, one can readily verify that

$(\overline{\gamma^*}, \overline{\theta^*})$ satisfies (4.11) and (4.12).

Let $(\overline{\gamma^*}, \overline{\theta^*})$ be an arbitrary saddle point satisfying (4.11) and (4.12). It can be shown[1] that

S1) $\gamma_i^* > 0$, $i = 1, \cdots, L$;

S2) there exists some i^* such that $\theta_i^* > 0$ for $i < i^*$ and $\theta_i^* = 0$ for $i \geq i^*$ (we set $i^* = 1$ if all the entries of $\overline{\theta^*}$ are zero, and set $i^* = L$ if all the entries of $\overline{\theta^*}$ are positive).

To complete the proof of Theorem 3, it suffices to show that

$$\kappa_u(\overline{\mu}, \overline{d}, \overline{\delta}) \leq \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}), \tag{4.15}$$

$$\kappa_l(\overline{\mu}, \overline{d}, \overline{\delta}) \geq \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}). \tag{4.16}$$

Clearly,

$$\kappa_u(\overline{\mu}, \overline{d}, \overline{\delta}) \leq \psi(\overline{\mu}, \overline{\gamma^{(m)}}, \overline{\delta}, \overline{\theta^{(m)}}), \tag{4.17}$$

where $\overline{\gamma^{(m)}} \triangleq (\gamma_1^{(m)}, \cdots, \gamma_L^{(m)})$ (with $\gamma_i^{(m)} \in (0, d_i)$, $i = 1, \cdots, L$) tends to $\overline{\gamma^*}$ as $m \to \infty$, and $\overline{\theta^{(m)}}$ is a maximizer to

$$\max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma^{(m)}}, \overline{\delta}, \overline{\theta}).$$

[1]Note that we must have $\gamma_1^* = d_1$. If $\gamma_i^* = 0$ for some $i \geq 2$, then it follows by (3.3) and (3.4) that $\theta_{i-1}^* > 0$; on the other hand, according to (4.7) and (4.9), we must have $\theta_{i-1}^* = 0$, which leads to a contradiction. In view of (4.7) and the fact that $\gamma_i^* > 0$, we must have $\theta_i^* = 0$ if $\theta_{i-1}^* = 0$. One can verify that $\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta})$, as a function of $(\overline{\gamma}, \overline{\theta})$, is continuous at $(\overline{\gamma}, \overline{\theta})$ if $\gamma_i > 0$ for all i, but not necessarily so if $\gamma_i = 0$ for some i. It will be seen that S1) and S2) allow us to circumvent such points of discontinuity.

Without loss of generality, we assume that $\overline{\theta^{(m)}}$ converges to some $\overline{\theta^{\circ}}$ as $m \to \infty$ (otherwise one can take a converging subsequence of $\overline{\theta^{(m)}}$, $m \geq 1$). Note that

$$\lim_{m\to\infty} \psi(\overline{\mu}, \overline{\gamma^{(m)}}, \overline{\delta}, \overline{\theta^{(m)}}) = \psi(\overline{\mu}, \overline{\gamma^{*}}, \overline{\delta}, \overline{\theta^{\circ}}) \leq \psi(\overline{\mu}, \overline{\gamma^{*}}, \overline{\delta}, \overline{\theta^{*}}), \tag{4.18}$$

where the first equality is due to S1) and the fact that $\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta})$, as a function of $(\overline{\gamma}, \overline{\theta})$, is continuous at $(\overline{\gamma}, \overline{\theta})$ if $\gamma_i > 0$ for all i. Combining (4.17) and (4.18) proves (4.15). Now construct $\overline{\theta^{(n)}} \triangleq (\theta_1^{(n)}, \cdots, \theta_L^{(n)})$ with $\theta_i^{(n)} \in (0, \sigma^2_{X_{i+1}})$, $i = 1, \cdots, L-1$, such that $\overline{\theta^{(n)}}$ converges to $\overline{\theta^{*}}$ as $n \to \infty$, and

$$\mu_{i+1}(\sigma^2_{X_i} - \theta_{i-1}^{(n)})\sigma^2_{X_{i+1}}\theta_i^{(n)} - \mu_i(\sigma^2_{X_{i+1}} - \theta_i^{(n)})\sigma^2_{X_i}\theta_{i-1}^{(n)} \leq 0, \quad i = i^* + 1, \cdots, L-1, \tag{4.19}$$

where (4.19) is void if $i^* \geq L-1$. Let $\overline{\gamma^{(n)}}$ be the minimizer to

$$\min_{\gamma_i \in [0, d_i], i=1,\cdots,L} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta^{(n)}}).$$

It is easy to verify that $\overline{\gamma^{(n)}}$ converges to $\overline{\gamma^{\circ}} \triangleq (\gamma_1^{\circ}, \cdots, \gamma_L^{\circ})$ as $n \to \infty$, where $\gamma_i^{\circ} = \gamma^*$ for $i < i^*$ and $\gamma_i^{\circ} = d_i$ for $i \geq i^*$. Clearly,

$$\kappa_l(\overline{\mu}, \overline{d}, \overline{\delta}) \geq \lim_{n\to\infty} \psi(\overline{\mu}, \overline{\gamma^{(n)}}, \overline{\delta}, \overline{\theta^{(n)}}) = \psi(\overline{\mu}, \overline{\gamma^{\circ}}, \overline{\delta}, \overline{\theta^{*}}) \geq \psi(\overline{\mu}, \overline{\gamma^{*}}, \overline{\delta}, \overline{\theta^{*}}),$$

which proves (4.16).

Chapter 5

Minimax Theorem

We shall show that our main results in Section 1.3 can be viewed as a manifestation of a certain information-theoretic minimax theorem. It will be seen that this minimax theorem can be used to explain why there is no loss of optimality in choosing the auxiliary random vectors $Z_1^n, \cdots, Z_{L-1}^n$ to be Gaussian in the proof of Theorem 1.

Let $X_1^n, \cdots, X_L^n$ be defined as in Section 1.2. Define

$$\Phi = \mu_1 I(X_1^n; W_1) + \sum_{i=2}^{L} \mu_i (I(V_{i-1}; W_i) + I(X_i^n; W_i | W_1, \cdots, W_{i-1}, V_{i-1})),$$

where $\mu_1 \geq \cdots \geq \mu_L \geq 0$. We assume that $V_i \leftrightarrow X_{i+1}^n \leftrightarrow (W_1, \cdots, W_{i+1})$ form a Markov chain, $i = 1, \cdots, L-1$. As a consequence, in order to determine Φ, it suffices to specify the conditional distribution of V_i given X_{i+1}^n, $i = 1, \cdots, L-1$, as well as the conditional distribution of $(W_1, \cdots, W_L)$ given $(X_1^n, \cdots, X_L^n)$. Let $\mathcal{P}$ denote the set of conditional distributions $(p_{V_1|X_2^n}, \cdots, p_{V_{L-1}|X_L^n})$. Moreover, let $\mathcal{Q}$ denote the set of conditional distributions of $(W_1, \cdots, W_L)$ given $(X_1^n, \cdots, X_L^n)$ such that $\sigma^2_{X_i^n|W_1,\cdots,W_i} \leq d_i$, $i = 1, \cdots, L$, $\sigma^2_{X_i^n|W_i} \leq \delta_i$, $i = 2, \cdots, L$, and $X_{i+1}^n \leftrightarrow X_i^n \leftrightarrow$

$(W_1, \cdots, W_i)$ form a Markov chain, $i = 1, \cdots, L-1$, where $d_i \in (0, \sigma^2_{X_i}]$, $i = 1, \cdots, L$, and $\delta_i \in (0, \sigma^2_{X_i}]$, $i = 2, \cdots, L$.

Theorem 6.

$$\sup_{\mathcal{P}} \inf_{\mathcal{Q}} \Phi = \inf_{\mathcal{Q}} \sup_{\mathcal{P}} \Phi.$$

5.1 Extremal Inequality

The following extremal inequality is needed for the proof of Theorem 6.

Theorem 7. *Let N_i^n be a zero-mean Gaussian random vector with i.i.d. entries of positive variance $\sigma^2_{N_i}$, $i = 1, 2, 3$. Let d be an arbitrary positive real number. Then for any random vector S^n and random object W, jointly independent of (N_1^n, N_2^n, N_3^n), such that $\sigma^2_{S^n|W} \leq d$, we have*

$$\begin{aligned} &h(S^n + N_3^n|W) - h(S^n + N_1^n|W) - h(S^n + N_2^n|W) \\ &\leq \max_{\gamma \in [0,d]} \frac{n}{2} \log(2\pi e(\gamma + \sigma^2_{N_3})) - \frac{n}{2} \log(2\pi e(\gamma + \sigma^2_{N_1})) - \frac{n}{2} \log(2\pi e(\gamma + \sigma^2_{N_2})). \end{aligned}$$

Proof: First consider the case $\sigma^2_{N_3} \geq \sigma^2_{N_1}$. Without loss of generality, we can assume $N_3^n = N_1^n + \Theta^n$, where Θ^n is independent of (N_1^n, S^n, W), and the entries of Θ^n are

i.i.d. Gaussian random variables with mean zero and variance $\sigma^2_{N_3} - \sigma^2_{N_1}$. Note that

$$\begin{aligned}
&h(S^n + N_3^n|W) - h(S^n + N_1^n|W) \\
&= h(S^n + N_1^n + \Theta^n|W) - h(S^n + N_1^n|W) \\
&= I(\Theta^n; S^n + N_1^n + \Theta^n|W) \\
&\leq I(\Theta^n; S^n + N_1^n + \Theta^n, W) \\
&\leq I(\Theta^n; N_1^n + \Theta^n) \qquad (5.1) \\
&= \frac{n}{2}\log\left(\frac{\sigma^2_{N_3}}{\sigma^2_{N_1}}\right),
\end{aligned}$$

where (5.1) is due to the fact that $\Theta^n \leftrightarrow (N_1^n + \Theta^n) \leftrightarrow (S^n + N_1^n + \Theta^n, W)$ form a Markov chain. Moreover, we have

$$h(S^n + N_2^n|W) \geq h(S^n + N_2^n|S^n, W) = h(N_2^n) = \frac{n}{2}\log(2\pi e\sigma^2_{N_2}).$$

As a consequence,

$$\begin{aligned}
&h(S^n + N_3^n|W) - h(S^n + N_1^n|W) - h(S^n + N_2^n|W) \\
&\leq \frac{n}{2}\log(2\pi e\sigma^2_{N_3}) - \frac{n}{2}\log(2\pi e\sigma^2_{N_1}) - \frac{n}{2}\log(2\pi e\sigma^2_{N_2}),
\end{aligned}$$

which is the desired result. By symmetry, this upper bound also holds when $\sigma^2_{N_3} \geq \sigma^2_{N_2}$.

Now consider the case $\sigma^2_{N_3} < \min\{\sigma^2_{N_1}, \sigma^2_{N_2}\}$. Without loss of generality, we assume $N_i^n = N_3^n + \Theta_i^n$, where Θ_i^n is independent of (N_3^n, S^n, W), and the entries of Θ_i^n are i.i.d. Gaussian random variables with mean zero and variance $\sigma^2_{N_i} - \sigma^2_{N_3}$, $i = 1, 2$.

Note that

$$
\begin{aligned}
&h(S^n + N_i^n|W) \\
&= h(S^n + N_3^n + \Theta_i^n|W) \\
&\geq \frac{n}{2}\log\left(e^{\frac{2}{n}h(S^n+N_3^n|W)} + e^{\frac{2}{n}h(\Theta_i^n)}\right) && (5.2)\\
&= \frac{n}{2}\log\left(e^{\frac{2}{n}h(S^n+N_3^n|W)} + 2\pi e(\sigma_{N_i}^2 - \sigma_{N_3}^2)\right), \quad i = 1, 2, && (5.3)
\end{aligned}
$$

where (5.2) is due to the entropy power inequality. Note that $\sigma^2_{S^n+N_i^n|W} = \sigma^2_{S^n|W} + \sigma^2_{N_i} \leq d + \sigma^2_{N_i}$, $i = 1, 2$. Therefore, it follows from Lemma 1 in Section 2.1 that

$$
h(S^n + N_i^n|W) \leq \frac{n}{2}\log(2\pi e(d + \sigma_{N_i}^2)), \quad i = 1, 2. \tag{5.4}
$$

In view of (5.3) and (5.4), we have

$$
\begin{aligned}
&h(S^n + N_3^n|W) - h(S^n + N_1^n|W) - h(S^n + N_2^n|W) \\
&\leq \min_{\gamma\in[0,d]} \frac{n}{2}\log(2\pi e(\gamma + \sigma_{N_3}^2)) - \frac{n}{2}\log(2\pi e(\gamma + \sigma_{N_3}^2) + 2\pi e(\sigma_{N_1}^2 - \sigma_{N_3}^2)) \\
&\quad - \frac{n}{2}\log(2\pi e(\gamma + \sigma_{N_3}^2) + 2\pi e(\sigma_{N_2}^2 - \sigma_{N_3}^2)) \\
&= \min_{\gamma\in[0,d]} \frac{n}{2}\log(2\pi e(\gamma + \sigma_{N_3}^2)) - \frac{n}{2}\log(2\pi e(\gamma + \sigma_{N_1}^2)) - \frac{n}{2}\log(2\pi e(\gamma + \sigma_{N_2}^2)),
\end{aligned}
$$

which completes the proof. □

5.2 Proof of the Minimax Theorem

Now we proceed to prove Theorem 6. It suffices to show that

$$\sup_{\mathcal{P}} \inf_{\mathcal{Q}} \Phi \geq \inf_{\mathcal{Q}} \sup_{\mathcal{P}} \Phi. \tag{5.5}$$

Let Z_i^n be a zero-mean Gaussian random vector with i.i.d. entries of positive variance $\sigma^2_{Z_i}$, $i = 1, \cdots, L-1$; moreover, we assume Z_i^n is independent of $(X_{i+1}^n, W_1, \cdots, W_{i+1})$, $i = 1, \cdots, L-1$. Let $V_i = X_{i+1}^n + Z_i^n$, $i = 1, \cdots, L-1$. Note that

$$\begin{aligned}
&I(X_i^n; W_i | W_1, \cdots, W_{i-1}, X_i^n + Z_{i-1}^n) \\
&= I(X_i^n + Z_{i-1}^n, X_i^n; W_i | W_1, \cdots, W_{i-1}) - I(X_i^n + Z_{i-1}^n; W_i | W_1, \cdots, W_{i-1}) \\
&= I(X_i^n; W_i | W_1, \cdots, W_{i-1}) - I(X_i^n + Z_{i-1}^n; W_i | W_1, \cdots, W_{i-1}), \quad i = 2, \cdots, L.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\Phi &= \mu_1 I(X_1^n; W_1) + \sum_{i=2}^{L} \mu_i (I(X_i^n + Z_{i-1}^n; W_i) + I(X_i^n; W_i | W_1, \cdots, W_{i-1}) \\
&\quad - I(X_i^n + Z_{i-1}^n; W_i | W_1, \cdots, W_{i-1})) \\
&= \frac{\mu_1 n}{2} \log(2\pi e \sigma^2_{X_1}) - \mu_1 h(X_1^n | W_1) + \sum_{i=2}^{L} \mu_i \Big(\frac{n}{2} \log(2\pi e (\sigma^2_{X_i} + \sigma^2_{Z_{i-1}})) \\
&\quad - h(X_i^n + Z_{i-1}^n | W_i) + h(X_i^n | W_1, \cdots, W_{i-1}) - h(X_i^n | W_1, \cdots, W_i) \\
&\quad - h(X_i^n + Z_{i-1}^n | W_1, \cdots, W_{i-1}) + h(X_i^n + Z_{i-1}^n | W_1, \cdots, W_i) \Big)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mu_1 n}{2}\log(2\pi e\sigma_{X_1}^2) - \mu_1 h(X_1^n|W_1) - \mu_2(h(X_2^n+Z_1^n|W_1) - h(X_2^n|W_1)) \\
&\quad + \sum_{i=2}^{L-1}\Big(\mu_i(h(X_i^n+Z_{i-1}^n|W_1,\cdots,W_i) - h(X_i^n|W_1,\cdots,W_i)) \\
&\quad - \mu_{i+1}(h(X_{i+1}^n+Z_i^n|W_1,\cdots,W_i) - h(X_{i+1}^n|W_1,\cdots,W_i))\Big) \\
&\quad + \frac{\mu_L}{n}(h(X_L^n+Z_{L-1}^n|W_1,\cdots,W_L) - h(X_L^n|W_1,\cdots,W_L)) \\
&\quad + \sum_{i=2}^{L}\mu_i\Big(\frac{n}{2}\log(2\pi e(\sigma_{X_i}^2+\sigma_{Z_{i-1}}^2)) - h(X_i^n+Z_{i-1}^n|W_i)\Big). \qquad (5.6)
\end{aligned}
$$

It can be shown (cf. (2.19), (2.20), (2.22), and (2.23)) that

$$
\begin{aligned}
&- \mu_1 h(X_1^n|W_1) - \mu_2(h(X_2^n+Z_1^n|W_1) - h(X_2^n|W_1)) \\
&\geq \min_{\gamma_1\in[0,d_1]} \frac{-\mu_1 n}{2}\log(2\pi e\gamma_1) - \frac{\mu_2 n}{2}\log\Big(\frac{\gamma_1+\sigma_{\Delta_1}^2+\sigma_{Z_1}^2}{\gamma_1+\sigma_{\Delta_1}^2}\Big), \qquad (5.7)
\end{aligned}
$$

$$
\begin{aligned}
&\mu_i(h(X_i^n+Z_{i-1}^n|W_1,\cdots,W_i) - h(X_i^n|W_1,\cdots,W_i)) \\
&- \mu_{i+1}(h(X_{i+1}^n+Z_i^n|W_1,\cdots,W_i) - h(X_{i+1}^n|W_1,\cdots,W_i)) \\
&\geq \min_{\gamma_i\in[0,d_i]} \frac{\mu_i n}{2}\log\Big(\frac{\gamma_i+\sigma_{Z_{i-1}}^2}{\gamma_i}\Big) - \frac{\mu_{i+1} n}{2}\log\Big(\frac{\gamma_i+\sigma_{\Delta_i}^2+\sigma_{Z_{i-1}}^2}{\gamma_i+\sigma_{\Delta_i}^2}\Big), \quad i=2,\cdots,L-1, \qquad (5.8)
\end{aligned}
$$

$$
h(X_L^n+Z_{L-1}^n|W_1,\cdots,W_L) - h(X_L^n|W_1,\cdots,W_L) \geq \min_{\gamma_L\in[0,d_L]} \frac{n}{2}\log\Big(\frac{\gamma_L+\sigma_{Z_{L-1}}^2}{\gamma_L}\Big), \qquad (5.9)
$$

$$
- h(X_i^n+Z_{i-1}^n|W_i) \geq -\frac{n}{2}\log(2\pi e(\delta_i+\sigma_{Z_{i-1}}^2)), \quad i=2,\cdots,L. \qquad (5.10)
$$

Substituting (5.7), (5.8), (5.9), (5.10) into (5.6) and setting $\sigma^2_{Z_i} = (\theta_i^{-1} - \sigma^{-2}_{X_{i+1}})^{-1}$, $i = 1, \cdots, L-1$, yields

$$\Phi \geq \min_{\gamma_i \in [0,d_i], i=1,\cdots,L} n\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}),$$

which further implies

$$\inf_{\mathcal{Q}} \Phi \geq \min_{\gamma_i \in [0,d_i], i=1,\cdots,L} n\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}).$$

Therefore, we have

$$\begin{aligned}\sup_{\mathcal{P}} \inf_{\mathcal{Q}} \Phi &\geq \sup_{\sigma^2_{Z_i}>0, i=1,\cdots,L-1} \min_{\gamma_i \in [0,d_i], i=1,\cdots,L} n\ \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta})\big|_{\theta_i = (\sigma^{-2}_{X_{i+1}} + \sigma^{-2}_{Z_i})^{-1}, i=1,\cdots,L-1} \\ &= \sup_{\theta_i \in (0, \sigma^2_{X_{i+1}}), i=1,\cdots,L-1} \min_{\gamma_i \in [0,d_i], i=1,\cdots,L} n\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}). \qquad (5.11)\end{aligned}$$

In view of (5.11) and Theorem 3, for the purpose of establishing (5.5), it suffices to prove that

$$\inf_{\mathcal{Q}} \sup_{\mathcal{P}} \Phi \leq \inf_{\gamma_i \in (0,d_i), i=1,\cdots,L} \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} n\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}).$$

To this end, we shall show that given any $\overline{\gamma}$ with $\gamma_i \in (0, d_i)$, $i = 1, \cdots, L$, there exists $(W_1, \cdots, W_L)$ such that

A1) $\sigma^2_{X_i^n | W_1, \cdots, W_i} \leq \gamma_i$, $i = 1, \cdots, L$,

A2) $\sigma^2_{X_i^n | W_i} \leq \delta_i$, $i = 2, \cdots, L$,

A3) $X_{i+1}^n \leftrightarrow X_i^n \leftrightarrow (W_1, \cdots, W_i)$ form a Markov chain, $i = 1, \cdots, L-1$;

moreover,

$$\sup_{\mathcal{P}} \Phi \leq \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} n\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}). \tag{5.12}$$

Let $\overline{\gamma'}$ and $\overline{\delta'}$ be as specified in Lemma 5. Note that $\gamma'_i \in (0, \gamma_i]$, $i = 1, \cdots, L$, $\delta'_i \in (0, \delta_i]$, $i = 2, \cdots, L$, and

$$\max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma'}, \overline{\delta'}, \overline{\theta}) \leq \max_{\theta_i \in [0, \sigma^2_{X_{i+1}}], i=1,\cdots,L-1} \psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}). \tag{5.13}$$

As shown in Chapter 3, one can construct a zero-mean random vector $(U_1, \cdots, U_L, \tilde{U}_1, \cdots, \tilde{U}_L)$ jointly Gaussian with $(X_1, \cdots, X_L)$ such that

B1) $\tilde{U}_i = \mathbb{E}[X_i | U_1, \cdots, U_i]$ and $\mathbb{E}[(X_i - \tilde{U}_i)^2] = \gamma'_i$, $i = 1, \cdots, L$,

B2) $\mathbb{E}[U_i^2] = \mathbb{E}[X_i U_i] = \sigma^2_{X_i} - \delta'_i$, $i = 2, \cdots, L$,

B3) $(X_2, \cdots, X_L) \leftrightarrow X_1 \leftrightarrow U_1$ form a Markov chain, and $(X_j)_{j \neq i} \leftrightarrow (X_i, U_1, \cdots, U_{i-1}) \leftrightarrow U_i$ form a Markov chain, $i = 1, \cdots, L$.

First assume that $\delta'_i < \sigma^2_{X_i}$, $i = 2, \cdots, L$. Define

$$N_i = \frac{\sigma^2_{X_i}}{\sigma^2_{X_i} - \gamma'_{i-1} - \sigma^2_{\Delta_{i-1}}} \tilde{U}_{i-1} - X_i, \quad i = 2, \cdots, L,$$
$$\hat{N}_i = \frac{\sigma^2_{X_i}}{\sigma^2_{X_i} - \delta'_i} U_i - X_i, \quad i = 2, \cdots, L,$$
$$\tilde{N}_i = \frac{\sigma^2_{X_i}}{\sigma^2_{X_i} - \gamma'_i} \tilde{U}_i - X_i, \quad i = 2, \cdots, L.$$

It can be verified that

$$\sigma_{N_i}^2 = \frac{\sigma_{X_i}^2(\gamma_{i-1}' + \sigma_{\Delta_{i-1}}^2)}{\sigma_{X_i}^2 - \gamma_{i-1}' - \sigma_{\Delta_{i-1}}^2}, \quad i = 2, \cdots, L, \tag{5.14}$$

$$\sigma_{\hat{N}_i}^2 = \frac{\sigma_{X_i}^2 \delta_i'}{\sigma_{X_i}^2 - \delta_i'}, \quad i = 2, \cdots, L, \tag{5.15}$$

$$\sigma_{\tilde{N}_i}^2 = \frac{\sigma_{X_i}^2 \gamma_i'}{\sigma_{X_i}^2 - \gamma_i'}, \quad i = 2, \cdots, L. \tag{5.16}$$

Let $(X_{ij}, U_{ij}, \tilde{U}_{ij}, N_{ij}, \hat{N}_{ij}, \tilde{N}_{ij}, i = 1, \cdots, L)_{j=1}^n$ be n i.i.d. copies of $(X_i, U_i, \tilde{U}_i, N_i, \hat{N}_i, \tilde{N}_i, i = 1, \cdots, L)$, and let $W_i = U_i^n$, $i = 1, \cdots, L$. It is easy to see that A1), A2), and A3) are implied by B1), B2), and B3), respectively. Moreover, one can readily verify that $(N_{i,1}^n, N_{i,2}^n, N_{i,3}^n)$ is independent of (X_i^n, V_{i-1}), $i = 2, \cdots, L$. Note that

$$\begin{aligned}
\Phi &= \mu_1 I(X_1^n; U_1^n) + \sum_{i=2}^{L} \mu_i (I(V_{i-1}; U_i^n) + I(X_i^n; U_i^n | U_1^n, \cdots, U_{i-1}^n, V_{i-1})) \\
&= \mu_1 I(X_1^n; U_1^n) + \sum_{i=2}^{L} \mu_i (I(V_{i-1}; U_i^n) + I(X_i^n; U_i^n | U_1^n, \cdots, U_{i-1}^n) \\
&\quad - I(V_{i-1}; U_i^n | U_1^n, \cdots, U_{i-1}^n)) \\
&= \mu_1 I(X_1^n; U_1^n) + \sum_{i=2}^{L} \mu_i (I(V_{i-1}; U_i^n) + I(X_i^n; U_1^n, \cdots, U_i^n) - I(X_i^n; U_1^n, \cdots, U_{i-1}^n) \\
&\quad - I(V_{i-1}; U_1^n, \cdots, U_i^n) + I(V_{i-1}; U_1^n, \cdots, U_{i-1}^n)) \\
&= \mu_1 I(X_1^n; U_1^n) + \sum_{i=2}^{L} \mu_i (I(V_{i-1}; U_i^n) + I(X_i^n; \tilde{U}_i^n) - I(X_i^n; \tilde{U}_{i-1}^n) \\
&\quad - I(V_{i-1}; \tilde{U}_i^n) + I(V_{i-1}; \tilde{U}_{i-1}^n)) \\
&= \frac{\mu_1 n}{2} \log\left(\frac{\sigma_{X_1}^2}{\gamma_1'}\right) + \sum_{i=2}^{L} \mu_i \Big(\frac{1}{2} \log\left(\frac{\gamma_{i-1}' + \sigma_{\Delta_{i-1}}^2}{\gamma_i'}\right) + I(V_{i-1}; U_i^n) \\
&\quad - I(V_{i-1}; \tilde{U}_i^n) + I(V_{i-1}; \tilde{U}_{i-1}^n)\Big).
\end{aligned} \tag{5.17}$$

We have

$$\begin{aligned}
&I(V_{i-1};U_i^n) - I(V_{i-1};\tilde{U}_i^n) + I(V_{i-1};\tilde{U}_{i-1}^n) \\
&= I(V_{i-1};X_i^n + \hat{N}_i^n) - I(V_{i-1};X_i^n + \tilde{N}_i^n) + I(V_{i-1};X_i^n + N_i^n) \\
&= \frac{n}{2}\log(2\pi e(\sigma_{X_i}^2 + \sigma_{N_i}^2)) + \frac{n}{2}\log(2\pi e(\sigma_{X_i}^2 + \sigma_{\hat{N}_i}^2)) - \frac{n}{2}\log(2\pi e(\sigma_{X_i}^2 + \sigma_{\tilde{N}_i}^2)) \\
&\quad + h(X_i^n + \tilde{N}_i^n|V_{i-1}) - h(X_i^n + N_i^n|V_{i-1}) - h(X_i^n + \hat{N}_i^n|V_{i-1}), \quad i = 2, \cdots, L.
\end{aligned}$$

Since $\sigma_{X_i^n|V_{i-1}}^2 \leq \sigma_{X_i}^2$, it follows from Theorem 7 that

$$\begin{aligned}
&h(X_i^n + \tilde{N}_i^n|V_{i-1}) - h(X_i^n + N_i^n|V_{i-1}) - h(X_i^n + \hat{N}_i^n|V_{i-1}) \\
&\leq \max_{\theta_{i-1}\in[0,\sigma_{X_i}^2]} \frac{n}{2}\log(2\pi e(\theta_{i-1} + \sigma_{\tilde{N}_i}^2)) - \frac{n}{2}\log(2\pi e(\theta_{i-1} + \sigma_{N_i}^2)) \\
&\qquad - \frac{n}{2}\log(2\pi e(\theta_{i-1} + \sigma_{\hat{N}_i}^2)), \quad i = 2, \cdots, L.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&I(V_{i-1};U_i^n) - I(V_{i-1};\tilde{U}_i^n) + I(V_{i-1};\tilde{U}_{i-1}^n) \\
&\leq \frac{n}{2}\log\Big(\frac{(\sigma_{X_i}^2 + \sigma_{N_i}^2)(\sigma_{X_i}^2 + \sigma_{\hat{N}_i}^2)(\theta_{i-1} + \sigma_{\tilde{N}_i}^2)}{(\theta_{i-1} + \sigma_{N_i}^2)(\theta_{i-1} + \sigma_{\hat{N}_i}^2)(\sigma_{X_i}^2 + \sigma_{\tilde{N}_i}^2)}\Big), \quad i = 2, \cdots, L. \qquad (5.18)
\end{aligned}$$

Substituting (5.18) into (5.17) and invoking (5.14)-(5.16) yields

$$\Phi \leq \max_{\theta_i\in[0,\sigma_{X_{i+1}}^2], i=1,\cdots,L-1} n\psi(\overline{\mu}, \overline{\gamma'}, \overline{\delta'}, \overline{\theta}). \qquad (5.19)$$

If $\delta_i' = \sigma_{X_i}^2$ for some i, then

$$I(V_{i-1}; U_i^n) - I(V_{i-1}; \tilde{U}_i^n) + I(V_{i-1}; \tilde{U}_{i-1}^n) = 0.$$

As a consequence, one can readily verify that (5.19) continues to hold. Combining (5.13) and (5.19) gives

$$\Phi \leq \max_{\theta_i \in [0, \sigma_{X_{i+1}}^2], i=1,\cdots,L-1} n\psi(\overline{\mu}, \overline{\gamma}, \overline{\delta}, \overline{\theta}),$$

which further implies (5.12). This completes the proof of Theorem 6.

Chapter 6

Miscellaneous Results

6.1 The Minimum Sum Rate: A Special Case

In this section we focus on the case $\overline{\mu} = (1, \cdots, 1)$ (which corresponds to the sum rate). Let $\sigma^2_{X_i} = \rho^{i-1}$, $i = 1, \cdots, L$, and $\sigma^2_{\Delta_i} = \rho^{i-1}(\rho - 1)$, $i = 1, \cdots, L-1$, where $\rho > 1$. Moreover, let $\overline{d} = (\rho^0 d, \cdots, \rho^{L-1} d)$ and $\overline{\delta} = (\rho\delta, \cdots, \rho^{L-1}\delta)$, where $d \in (0, 1]$ and $\delta \in (0, 1]$. By Theorem 3, $\kappa_l(\overline{\mu}, \overline{d}, \overline{\delta})$ and $\kappa_u(\overline{\mu}, \overline{d}, \overline{\delta})$ coincide in this special case; therefore, we shall denote them by $\kappa(\overline{\mu}, \overline{d}, \overline{\delta})$. The main result of this section is an explicit characterization of $\kappa(\overline{\mu}, \overline{d}, \overline{\delta})$.

Theorem 8. *1. If $d \geq (\frac{\rho}{d+\rho-1} + \frac{1}{\delta} - 1)^{-1}$, then*

$$\kappa(\overline{\mu}, \overline{d}, \overline{\delta}) = \frac{1}{2} \log\left(\frac{1}{d}\right) + \frac{L-1}{2} \log\left(\frac{1}{\delta}\right).$$

2. If $d \leq \frac{\rho\delta-1}{\rho-1}$, then

$$\kappa(\overline{\mu}, \overline{d}, \overline{\delta}) = \frac{1}{2}\log\Big(\frac{1}{d}\Big) + \frac{L-1}{2}\log\Big(\frac{d+\rho-1}{\rho d}\Big).$$

3. If $\frac{\rho\delta-1}{\rho-1} < d < (\frac{\rho}{d+\rho-1} + \frac{1}{\delta} - 1)^{-1}$, then

$$\kappa(\overline{\mu}, \overline{d}, \overline{\delta}) = \frac{1}{2}\log\Big(\frac{1}{d}\Big) + \frac{L-1}{2}\log\Big(\frac{(d+\rho-1)((1-\theta)d+\theta)}{((1-\theta)(d+\rho-1)+\rho\theta)((1-\theta)\delta+\theta)d}\Big),$$

where

$$\theta = \sqrt{\frac{\rho-1}{1-d}\Big(\frac{\delta}{1-\delta} - \frac{d}{1-d}\Big)} - \frac{d}{1-d}.$$

Proof: 1. Let $\overline{\gamma^*} = \overline{d}$ and $\overline{\theta^*} = (0, \cdots, 0)$. It can be verified that $(\overline{\gamma^*}, \overline{\theta^*})$ satisfies (4.11) and (4.12) (see the optimality conditions in Lemmas 4, 7, 8, and 9). Therefore, it follows by the proof of Theorem 3 that

$$\kappa(\overline{\mu}, \overline{d}, \overline{\delta}) = \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}) = \frac{1}{2}\log\Big(\frac{1}{d}\Big) + \frac{L-1}{2}\log\Big(\frac{1}{\delta}\Big).$$

Note that in this case we can decrease the hierarchical distortion constraint $\overline{d}$ to $\overline{d'} \triangleq (d'_1, \cdots, d'_L)$ without affecting the minimum sum rate, where

$$d'_1 = d,$$
$$d'_i = \rho^{i-1}\Big(\frac{\rho^{i-1}}{d'_{i-1} + \rho^{i-2}(\rho-1)} + \frac{1}{\delta} - 1\Big)^{-1}, \quad i = 2, \cdots, L.$$

2. Let $\overline{\gamma^*} = \overline{d}$ and $\overline{\theta^*} = (\rho, \cdots, \rho^{L-1})$. It can be verified that $(\overline{\gamma^*}, \overline{\theta^*})$ satisfies (4.11) and (4.12) (see the optimality conditions in Lemmas 4, 7, 8, and 9). Therefore,

it follows by the proof of Theorem 3 that

$$\kappa(\overline{\mu}, \overline{d}, \overline{\delta}) = \psi(\overline{\mu}, \overline{\gamma^*}, \overline{\delta}, \overline{\theta^*}) = \frac{1}{2}\log\Big(\frac{1}{d}\Big) + \frac{L-1}{2}\log\Big(\frac{d+\rho-1}{\rho d}\Big).$$

Note that in this case we can decrease the individual distortion constraint $\overline{\delta}$ to $\overline{\delta'} \triangleq (\delta'_2, \cdots, \delta'_L)$ without affecting the minimum sum rate, where

$$\delta'_i = \rho^{i-2}(\rho-1)d + \rho^{i-2}, \quad i = 2, \cdots, L.$$

3. Let $\overline{\gamma^*} = \overline{d}$ and $\overline{\theta^*} = (\rho\theta, \cdots, \rho^{L-1}\theta)$. It is easy to verify that $(\overline{\gamma^*}, \overline{\theta^*})$ satisfies (4.12) (see the optimality condition in Lemma 4). Note that the optimality conditions in Lemma 7 and Lemma 9 are clearly satisfied; therefore, to verify (4.11), it suffices to show that $(\overline{\gamma^*}, \overline{\theta^*})$ satisfies the optimality condition in Lemma 8. In view of the fact that $\theta \in (0, 1)$ and that

$$\tilde{a}_i = \rho^{3i-2}(\rho-1)(1-\theta)\theta > 0,$$

we just need to show that $\hat{\gamma}_i \geq \rho^{i-1}d$, where

$$\hat{\gamma}_i = \rho^{i-1}\Big(1 + \sqrt{\frac{\rho}{1-\theta}}\Big), \quad i = 2, \cdots, L-1.$$

This is indeed true since $1 + \sqrt{\frac{\rho}{1-\theta}} > 1 \geq d$. Therefore, it follows by the proof

of Theorem 3 that

$$\begin{aligned}&\kappa(\overline{\mu},\overline{d},\overline{\delta})=\psi(\overline{\mu},\overline{\gamma^*},\overline{\delta},\overline{\theta^*})\\=&\frac{1}{2}\log\Big(\frac{1}{d}\Big)+\frac{L-1}{2}\log\Big(\frac{(d+\rho-1)((1-\theta)d+\theta)}{((1-\theta)(d+\rho-1)+\rho\theta)((1-\theta)\delta+\theta)d}\Big).\end{aligned}$$

□

6.2 Robust Predictive Coding System

In this section we propose an efficient implementation of the robust predictive coding scheme associated with Lemma 6. For simplicity, throughout this section we describe the scheme in the form of single-letter operations; however, it should be understood that in fact such a scheme has to be implemented over long blocks in order to approach the information-theoretic limits.

As shown in Chapter 3, to minimize the weighted sum rate $\overline{\mu}\overline{R}^T$ of the robust predictive coding scheme associated with Lemma 6, there is no loss of optimality in considering zero-mean random vector $(U_1,\cdots,U_L,\tilde{U}_1,\cdots,\tilde{U}_L)$ jointly Gaussian with $(X_1,\cdots,X_L)$ such that

$$\begin{aligned}&\mathbb{E}[U_1^2]=\mathbb{E}[X_1U_1]=\sigma_{X_1}^2-\gamma_1',\\&\mathbb{E}[U_i^2]=\mathbb{E}[X_iU_i]=\sigma_{X_i}^2-\delta_i',\quad i=2,\cdots,L,\\&\mathbb{E}[\tilde{U}_{i-1}U_i]=\xi_{i-1},\quad i=2,\cdots,L,\\&\tilde{U}_1=\mathbb{E}[X_1|U_1]=U_1,\end{aligned}$$

$$\tilde{U}_i = \mathbb{E}[X_i|U_1, \cdots, U_i] = \mathbb{E}[X_i|\tilde{U}_{i-1}, U_i] = a_{i-1}\tilde{U}_{i-1} + b_{i-1}U_i, \quad i = 2, \cdots, L,$$
$$\mathbb{E}[(X_i - \tilde{U}_i)^2] = \gamma_i', \quad i = 1, \cdots, L,$$
$$\mathbb{E}[(X_i - U_i)^2] = \delta_i', \quad i = 2, \cdots, L,$$

where $\overline{\gamma'}$ and $\overline{\delta'}$ satisfy

$$0 < \gamma_i' \leq d_i, \quad i = 1, \cdots, L,$$
$$0 < \delta_i' \leq \delta_i, \quad i = 2, \cdots, L,$$
$$\gamma_i' + \sigma^2_{\Delta_i} + \delta_{i+1}' - \sigma^2_{X_{i+1}} \leq \gamma_{i+1}' \leq \left(\frac{1}{\gamma_i' + \sigma^2_{\Delta_i}} + \frac{1}{\delta_{i+1}'} - \frac{1}{\sigma^2_{X_{i+1}}}\right)^{-1}, \quad i = 1, \cdots, L-1,$$

and the parameters ξ_i, a_i, and b_i, $i = 1, \cdots, L-1$ are defined in (3.6), (3.7), and (3.8), respectively; moreover, $(X_2, \cdots, X_L) \leftrightarrow X_1 \leftrightarrow U_1$ form a Markov chain, and $((X_j)_{j \neq i}, U_1, \cdots, U_{i-1}) \leftrightarrow (X_i, \tilde{U}_{i-1}) \leftrightarrow U_i$ form a Markov chain, $i = 2, \cdots, L$. Let

$$R_1 = I(X_1; U_1),$$
$$R_i = I(X_i, U_1, \cdots, U_{i-1}; U_i), \quad i = 2, \cdots, L,$$

where R_i is the rate of Encoder i. One can interpret X_1 and U_1, respectively, as the input and the output of Encoder 1; similarly, $(X_i, U_1, \cdots, U_{i-1})$ and U_i can be interpreted, respectively, as the input and the output of Encoder i, $i = 2, \cdots, L$. Given the outputs from the first i encoders, the decoder can compute $\tilde{U}_i$ and use it as the reconstruction of X_i, and the resulting distortion is γ_i', $i = 1, \cdots, L$. If the decoder only receives the output from Encoder i, then it simply uses U_i as the reconstruction of X_i, and the resulting distortion is δ_i', $i = 2, \cdots, L$. Moreover, in

view of the fact that

$$I(X_i, U_1, \cdots, U_{i-1}; U_i) = I(X_i, \tilde{U}_{i-1}; U_i),$$

it suffices to provide Encoder i with $(X_i, \tilde{U}_{i-1})$ as the input, $i = 2, \cdots, L$. Note that we can write

$$\begin{aligned}
&U_1 = \mathbb{E}[U_1|X_1] + N_1 = \alpha_1 X_1 + N_1, \\
&U_i = \mathbb{E}[U_i|X_i, \tilde{U}_{i-1}] + N_i = \alpha_i X_i + \beta_{i-1}\tilde{U}_{i-1} + N_i, \quad i = 2, \cdots, L,
\end{aligned}$$

where

$$\begin{aligned}
&\alpha_1 = \frac{\sigma^2_{X_1} - \gamma'_1}{\sigma^2_{X_1}}, \\
&\alpha_i = \frac{\sigma^2_{X_i} - \delta'_i - \xi_{i-1}}{\gamma'_{i-1} + \sigma^2_{\Delta_{i-1}}}, \quad i = 2, \cdots, L, \\
&\beta_i = \begin{cases} 0, & \gamma'_i = \sigma^2_{X_i} \\ \frac{\xi_i}{\sigma^2_{X_i} - \gamma'_i} - \frac{\sigma^2_{X_{i+1}} - \delta'_{i+1} - \xi_i}{\gamma'_i + \sigma^2_{\Delta_i}}, & \text{otherwise} \end{cases}, \quad i = 1, \cdots, L-1, \\
&\mathbb{E}[N_1^2] = \frac{(\sigma^2_{X_1} - \gamma'_1)\gamma'_1}{\sigma^2_{X_1}}, \\
&\mathbb{E}[N_i^2] = \begin{cases} \frac{(\sigma^2_{X_i} - \delta'_i)\delta'_i}{\sigma^2_{X_i}}, & \gamma'_{i-1} = \sigma^2_{X_{i-1}} \\ \sigma^2_{X_i} - \delta'_i - \frac{(\sigma^2_{X_i} - \delta'_i - \xi_{i-1})^2}{\gamma'_{i-1} + \sigma^2_{\Delta_{i-1}}} - \frac{\xi^2_{i-1}}{\sigma^2_{X_{i-1}} - \gamma'_{i-1}}, & \text{otherwise} \end{cases}, \quad i = 2, \cdots, L.
\end{aligned}$$

It is clear that

$$\begin{aligned}
&I(X_1; U_1) = I(\alpha_1 X_1; \alpha_1 X_1 + N_1), \\
&I(X_i, \tilde{U}_{i-1}; U_i) = I(\alpha_i X_i + \beta_{i-1}\tilde{U}_{i-1}; \alpha_i X_i + \beta_{i-1}\tilde{U}_{i-1} + N_i), \quad i = 2, \cdots, L.
\end{aligned}$$

As a consequence, we can interpret Encoder 1 as a quantizer with $\alpha_1 X_1$, U_1, and N_1 respectively as the input, the output, and the quantization error; similarly, we can interpret Encoder i as a quantizer with $\alpha_i X_i + \beta_{i-1}\tilde{U}_{i-1}$, U_i, and N_i respectively as the input, the output, and the quantization error, $i = 2, \cdots, L$. Furthermore, in view of the fact that

$$\tilde{U}_i = a_{i-1}\tilde{U}_{i-1} + b_{i-1}U_i, \quad i = 2, \cdots, L,$$

the calculation of $(\tilde{U}_1, \cdots, \tilde{U}_L)$ at the encoders and the decoder can be performed iteratively. A robust predictive coding system based on this interpretation is depicted in Fig. 6.1. It is worth mentioning that one can implement the quantization operation in such a system by using entropy-coded dithered lattice quantizers (see, e.g., [12; 13; 14]).

6.3 Reconstruction Based on an Arbitrary Subset of Encoder Outputs

As pointed out in Section 6.2, one can interpret U_i as the output of Encoder i, $i = 1, \cdots, L$, for the robust predictive coding scheme associated with Lemma 6. Although it is developed for the scenario where only the hierarchical distortion constraint and the individual distortion constraint are imposed, this scheme has a desirable property that every subset of the encoder outputs is decodable. For example, if at the time of reconstructing X_5, the decoder only receives the outputs from a subset of the first 5 encoders (say, (U_1, U_3, U_4)), then it can still decode these outputs and further use

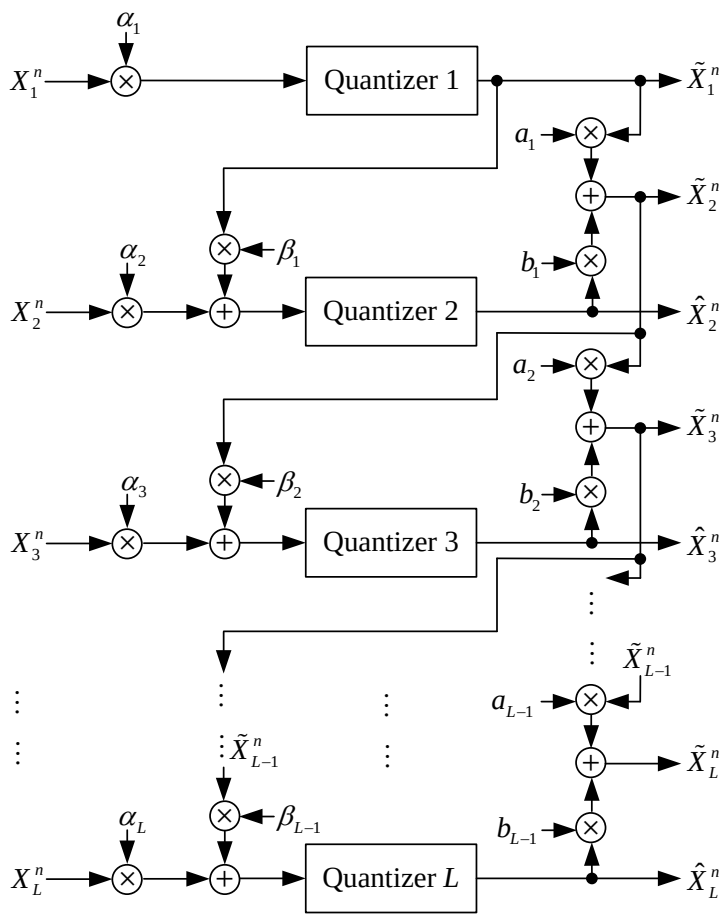

Figure 6.1: Robust predictive coding via successive quantization. Here $\tilde{X}_i^n$ can be viewed as a multi-letter version of $\tilde{U}_i$, $i = 1, \cdots, L$, and $\hat{X}_i^n$ can be viewed as a multi-letter version of U_i, $i = 2, \cdots, L$.

$\mathbb{E}[X_5|U_1, U_3, U_4]$ as the reconstruction of X_5 (with the resulting distortion equal to $\sigma^2_{X_5|U_1,U_3,U_4}$).

Now we proceed to give a detailed analysis for this kind of scenario. Again we shall focus on the case where $(U_1, \cdots, U_L)$ satisfies the conditions listed in Section 6.2. Assume that the decoder receives U_j, $j \in \mathcal{A}$ for some non-empty set $\mathcal{A} \subseteq \{1, \cdots, L\}$; moreover, at the time of reconstructing X_i, the decoder is only allowed to use $(U_j)_{j\in\mathcal{A}, j\leq i}$. With no loss of generality, we shall assume $\mathbb{E}[(U_i^2)] \neq 0$ (which

implies $\delta_i' < \sigma_{X_i}^2$) for all $i \in \mathcal{A}$.

Define

$$\hat{U}_i = \mathbb{E}[X_i|(U_j)_{j\in\mathcal{A},j\leq i}], \quad i = 1, \cdots, L,$$
$$\hat{d}_i = \mathbb{E}[(X_i - \hat{U}_i)^2], \quad i = 1, \cdots, L.$$

Note that $\hat{U}_i = 0$ and $\hat{d}_i = \sigma_{X_i}^2$ if $j > i$ for all $j \in \mathcal{A}$. We shall show that $(\hat{U}_1, \cdots, \hat{U}_L)$ and $(\hat{d}_1, \cdots, \hat{d}_L)$ can be computed iteratively.

It is clear that

$$\hat{U}_1 = \begin{cases} 0, & 1 \notin \mathcal{A} \\ U_1, & \text{otherwise} \end{cases},$$
$$\hat{d}_1 = \begin{cases} \sigma_{X_1}^2, & 1 \notin \mathcal{A} \\ \gamma_1', & \text{otherwise} \end{cases}.$$

For $i = 2, \cdots, L$, we have $\hat{U}_i = \hat{U}_{i-1}$ and $\hat{d}_i = \hat{d}_{i-1} + \sigma_{\Delta_{i-1}}^2$ if $i \notin \mathcal{A}$; moreover, we have $\hat{U}_i = U_i$ and $\hat{d}_i = \delta_i'$ if $\hat{d}_{i-1} = \sigma_{X_{i-1}}^2$. Therefore, it suffices to consider the case where $i \in \mathcal{A}$ and $\hat{d}_{i-1} < \sigma_{X_{i-1}}^2$ (which implies $\gamma_{i-1}' < \sigma_{X_{i-1}}^2$). Since $(U_1, \cdots, U_{i-1}) \leftrightarrow \tilde{U}_{i-1} \leftrightarrow (X_i, \tilde{U}_{i-1}) \leftrightarrow U_i$ form a Markov chain, it follows that $(U_j)_{j\in\mathcal{A},j<i} \leftrightarrow \hat{U}_{i-1} \leftrightarrow \tilde{U}_{i-1} \leftrightarrow (X, \tilde{U}_i) \leftrightarrow U_i$ form a Markov chain. As a consequence,

$$\hat{U}_i = \mathbb{E}[X_i|\hat{U}_{i-1}, U_i].$$

In view of the fact that

$$\mathbb{E}[\hat{U}_{i-1}^2] = \mathbb{E}[\tilde{U}_{i-1}\hat{U}_{i-1}] = \sigma_{X_{i-1}}^2 - \hat{d}_{i-1},$$

we can write

$$\hat{U}_{i-1} = \mathbb{E}[\hat{U}_{i-1}|\tilde{U}_{i-1}] + \hat{N}_{i-1} = \frac{\sigma_{X_{i-1}}^2 - \hat{d}_{i-1}}{\sigma_{X_{i-1}}^2 - \gamma'_{i-1}}\tilde{U}_{i-1} + \hat{N}_{i-1},$$

where $\hat{N}_{i-1}$ is independent of $(X_i, \tilde{U}_{i-1}, U_i)$ (recall that $\hat{U}_{i-1} \leftrightarrow \tilde{U}_{i-1} \leftrightarrow (X_i, U_i)$ form a Markov chain). Therefore,

$$\mathbb{E}[\hat{U}_{i-1}U_i] = \frac{\sigma_{X_{i-1}}^2 - \hat{d}_{i-1}}{\sigma_{X_{i-1}}^2 - \gamma'_{i-1}}\mathbb{E}[\tilde{U}_{i-1}U_i] = \frac{(\sigma_{X_{i-1}}^2 - \hat{d}_{i-1})\xi_{i-1}}{\sigma_{X_{i-1}}^2 - \gamma'_{i-1}}.$$

It is also easy to see that

$$\mathbb{E}[X_i\hat{U}_{i-1}] = \sigma_{X_{i-1}}^2 - \hat{d}_{i-1}.$$

Now one can readily verify that

$$\begin{aligned}
\hat{U}_i &= \hat{\alpha}_{i-1}\hat{U}_{i-1} + \hat{\beta}_{i-1}U_i, \\
\hat{d}_i &= \sigma_{X_i}^2 - \hat{\alpha}_{i-1}(\sigma_{X_{i-1}}^2 - \hat{d}_{i-1}) - \hat{\beta}_{i-1}(\sigma_{X_i}^2 - \delta'_i),
\end{aligned}$$

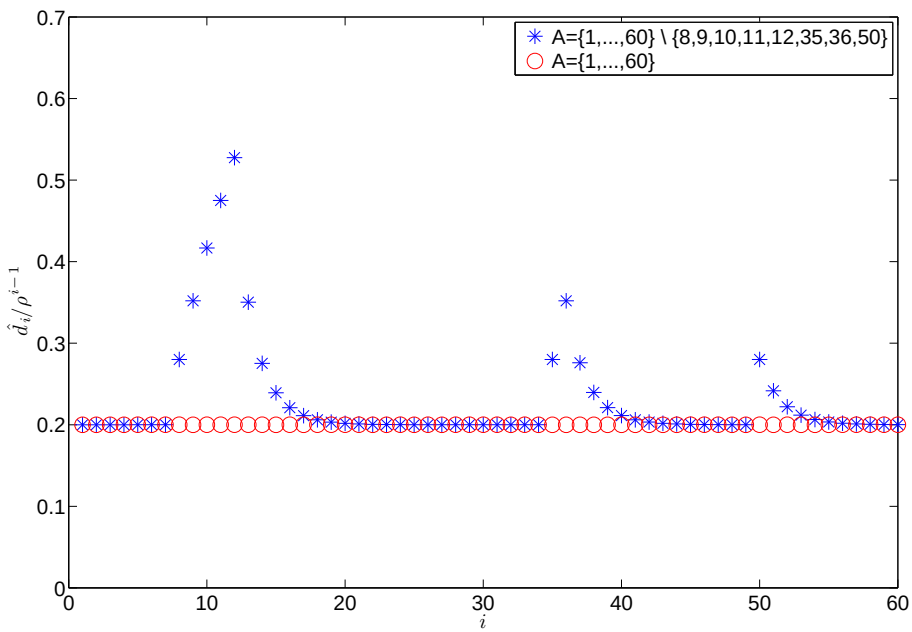

Figure 6.2: Reconstruction based on a subset of encoder outputs.

where

$$\hat{\alpha}_{i-1} = \frac{(\sigma^2_{X_{i-1}} - \gamma'_{i-1})^2(\sigma^2_{X_{i-1}} - \hat{d}_{i-1})(\sigma^2_{X_i} - \delta'_i) - (\sigma^2_{X_{i-1}} - \gamma'_{i-1})(\sigma^2_{X_{i-1}} - \hat{d}_{i-1})\xi_{i-1}(\sigma^2_{X_i} - \delta'_i)}{(\sigma^2_{X_{i-1}} - \gamma'_{i-1})^2(\sigma^2_{X_{i-1}} - \hat{d}_{i-1})(\sigma^2_{X_i} - \delta'_i) - (\sigma^2_{X_{i-1}} - \hat{d}_{i-1})^2\xi^2_{i-1}},$$

$$\hat{\beta}_{i-1} = \frac{-(\sigma^2_{X_{i-1}} - \gamma'_{i-1})(\sigma^2_{X_{i-1}} - \hat{d}_{i-1})^2\xi_{i-1} + (\sigma^2_{X_{i-1}} - \gamma'_{i-1})^2(\sigma^2_{X_{i-1}} - \hat{d}_{i-1})(\sigma^2_{X_i} - \delta'_i)}{(\sigma^2_{X_{i-1}} - \gamma'_{i-1})^2(\sigma^2_{X_{i-1}} - \hat{d}_{i-1})(\sigma^2_{X_i} - \delta'_i) - (\sigma^2_{X_{i-1}} - \hat{d}_{i-1})^2\xi^2_{i-1}}.$$

An illustrative example is given in Fig. 6.2. In this example, we choose $\sigma^2_{X_i} = \rho^{i-1}$, $i = 1, \cdots, L$, and $\sigma^2_{\Delta_i} = \rho^{i-1}(\rho - 1)$, $i = 1, \cdots, L-1$, where $L = 60$ and $\rho = \frac{10}{9}$; moreover, we set $\overline{\gamma'} = (\rho^0\gamma', \cdots, \rho^{L-1}\gamma')$ and $\overline{\delta'} = (\rho\delta', \cdots, \rho^{L-1}\delta')$, where $\gamma' = 0.2$ and $\delta' = 0.5$. We plot $\hat{d}_i/\rho^{i-1}$, $i = 1, \cdots, L$, for the scenario where $\mathcal{A} = \{1, \cdots, 60\}\backslash\{8, 9, 10, 11, 12, 35, 36, 50\}$. A comparison with the ideal scenario (i.e., $\mathcal{A} = \{1, \cdots, 60\}$) shows that the proposed scheme has a desirable "self-reovery" property.

Chapter 7

Conclusion and Future Work

We have partially characterized the rate region of robust sequential coding and robust predictive coding for the Gauss-Markov source model under the mean squared error distortion constraint. More fundamentally, our investigation reveals an information-theoretic minimax theorem, which can be obtained by coupling two extremal inequalities. It is worth noting that most of the results in this thesis can be extended to the vector source setting in a relatively straightforward manner. In particular, one can establish the vector version of Theorem 5 and Theorem 7 by leveraging techniques developed in [15; 16; 17].

Appendix A

Mathematical Elementary

A.1 Jensen's Inequality

If X is a random variable and ϕ is a convex function, then

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)].$$

If X is a random variable and ϕ is a concave function, then

$$\phi(\mathbb{E}[X]) \geq \mathbb{E}[\phi(X)].$$

A.2 Mean Square Error Estimation

Let X and Y be two random variables. The minimum mean squared error (MMSE) estimate of X given Y is a function $\hat{x}(Y)$ of Y that minimizes the mean squared error

$\mathbb{E}[(X - \hat{X})^2]$ and is given by

$$\hat{x}_{MMSE}(Y) = \mathbb{E}[X|Y].$$

If X and Y are zero-mean and jointly Gaussian, then

$$\hat{x}_{MMSE}(Y) = aY,$$

where $a = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$.

Appendix B

Information Theory Elementary

B.1 Entropy and Differential Entropy

Definition 5. *The entropy of a discrete random variable X with distribution $p(x)$ is defined as*

$$H(X) \triangleq -\sum_{x \in \mathcal{X}} p(x) \log p(x),$$

where $\mathcal{X}$ is alphabet of X.

Definition 6. *The differential entropy of a continuous random variable X with probability density function $f(x)$ is defined as*

$$h(X) \triangleq -\int f(x) \log f(x) dx.$$

If X is a Gaussian random variable with variance σ_X^2, then

$$h(X) = \frac{1}{2} \log(2\pi e \sigma_X^2).$$

Theorem 9. *The maximum differential entropy under the average power constraint is achieved by the Gaussian distribution*

$$\max_{f(x):\mathbb{E}[X^2]\leq P} h(X) = \frac{1}{2}\log(2\pi e P).$$

Theorem 10. *Entropy Power Inequality: Let X^n and Z^n be conditionally independent random vectors given a random variable U and let $Y^n = X^n + Z^n$, then*

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

B.2 Lossy Source Coding

Let X be a discrete memoryless source with distribution $p(x)$, and $w : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ be a distortion measure, where $\mathcal{X}$ and $\hat{\mathcal{X}}$ are the source alphabet and the reconstruction alphabet, respectively.

Definition 7. *A rate R is said to be achievable subject to distortion constraint D if for every $\epsilon > 0$ there exist an encoding function $f^{(n)} : \mathcal{X}^n \to \mathcal{C}$ and a decoding function $g^{(n)} : \mathcal{C} \to \hat{\mathcal{X}}^n$ such that*

$$\frac{1}{n}\log|\mathcal{C}| \leq R + \epsilon,$$
$$\mathbb{E}\Big[\frac{1}{n}\sum_{j=1}^{n} w(X_j, \hat{X}_j)\Big] \leq D + \epsilon,$$

where $\hat{X}^n = g^{(n)}(f^{(n)}(X^n))$.

Let $R(D)$ denote the minimum achievable rate subject distortion constraint D.

Theorem 11. *Shannon's Lossy Source Coding Theorem [6]:*

$$R(D) = \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X})$$

for $D \geq D_{\min} \triangleq \mathbb{E}[\min_{\hat{x}} w(X,\hat{x})]$.

Theorem 12. *The rate-distortion function for a zero-mean Gaussian source with variance* P *and the mean squared error distortion measure is*

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{P}{D} & 0 \leq D < P \\ 0 & D \geq P \end{cases}$$

B.3 Multiple Descriptions

Consider a discrete memoryless source X with distribution $p(x)$ and distortion measures $w_i : \mathcal{X} \times \hat{\mathcal{X}}_i \to [0,\infty)$, where $\mathcal{X}$ and $\hat{\mathcal{X}}_i$ are respectively the source alphabet and the reconstruction alphabet, $i = 0, 1, 2$.

Definition 8. *A pair of rates* (R_1, R_2) *is said to be achievable subject to distortion constraint* (D_0, D_1, D_2) *if for every* $\epsilon > 0$ *there exist encoding functions* $f_i^{(n)} : \mathcal{X}^n \to \mathcal{C}_i$, $i = 1, 2$, *decoding functions* $g_0^{(n)} : \mathcal{C}_1 \times \mathcal{C}_2 \to \hat{\mathcal{X}}_0^n$, *and* $g_i^{(n)} : \mathcal{C}_i \to \hat{\mathcal{X}}_i^n$, $i = 1, 2$ *such that*

$$\begin{aligned} &\frac{1}{n}\log|\mathcal{C}_i| \leq R_i + \epsilon, \quad i = 1, 2 \\ &\mathbb{E}\Big[\frac{1}{n}\sum_{j=1}^{n} w_i(X_j, \hat{X}_{ij})\Big] \leq D_i + \epsilon, \quad i = 0, 1, 2, \end{aligned}$$

where $\hat{X}_0^n = g_0^{(n)}(f_1^{(n)}(X^n), f_2^{(n)}(X^n))$ *and* $\hat{X}_i^n = g_i^{(n)}(f_i^{(n)}(X^n))$, $i = 1, 2$.

The rate region $\mathcal{R}(D_0, D_1, D_2)$ is the set of all rate vectors achievable subject to distortion constraints (D_0, D_1, D_2).

Theorem 13. *El Gamal-Cover Inner Bound [7]: A rate pair (R_1, R_2) is achievable subject to distortion constraint (D_0, D_1, D_2) if*

$$\begin{aligned} R_1 &> I(X, \hat{X}_1|Q), \\ R_2 &> I(X, \hat{X}_2|Q), \\ R_1 + R_2 &> I(X, \hat{X}_0, \hat{X}_1, \hat{X}_2|Q) + I(\hat{X}_1; \hat{X}_2|Q), \end{aligned}$$

for some $p(q)p(\hat{x}_0, \hat{x}_1, \hat{x}_2|x, q)$ such that $\mathbb{E}[w_i(X, \hat{X}_i)] \leq D_i$, $i = 0, 1, 2$.

Theorem 14. *Zhang-Berger Inner Bound [8]: A rate pair (R_1, R_2) is achievable subject to distortion constraint (D_0, D_1, D_2) if*

$$\begin{aligned} R_1 &> I(X, \hat{X}_1, U), \\ R_2 &> I(X, \hat{X}_2, U), \\ R_1 + R_2 &> I(X, \hat{X}_0, \hat{X}_1, \hat{X}_2|U) + 2I(U; X) + I(\hat{X}_1; \hat{X}_2|U), \end{aligned}$$

for some $p(u, \hat{x}_0, \hat{x}_1, \hat{x}_2|x)$ such that $\mathbb{E}[w_i(X, \hat{X}_i)] \leq D_i$, $i = 0, 1, 2$.

Ozarow [9] showed that the El Gamal-Cover inner bound is tight for the quadratic Gaussian case.

Theorem 15. *The multiple description rate region $\mathcal{R}(D_0, D_1, D_2)$ for a zero-mean Gaussian source X with variance P and mean squared error distortion measures is*

the set of rate pairs (R_1, R_2) *satisfying*

$$\begin{aligned} R_1 &\geq \frac{1}{2}\log\frac{P}{D_1}, \\ R_2 &\geq \frac{1}{2}\log\frac{P}{D_2}, \\ R_1 + R_2 &\geq \frac{1}{2}\log\frac{P}{D_0} + \Delta, \end{aligned}$$

where

$$\Delta = \frac{1}{2}\log\frac{(P-D_0)^2}{(P-D_0)^2 - (\sqrt{(P-D_1)(P-D_2)} - \sqrt{(D_1-D_0)(D_2-D_0)})^2}$$

if $D_1 + D_2 \leq P + D_0$*, and* $\Delta = 0$ *otherwise.*

Appendix C

Proof of Lemma 1

Proof: Note that

$$\begin{aligned}
h(S^n|W) &= \sum_{i=1}^{n} h(S_i|W, S^{i-1}) \\
&\leq \sum_{i=1}^{n} h(S_i|W) \\
&\leq \sum_{i=1}^{n} \frac{1}{2}\log(2\pi e \sigma^2_{S_i|W}) \\
&\leq \frac{n}{2}\log(2\pi e \sigma^2_{S^n|W}) \\
&\leq \frac{n}{2}\log(2\pi e d),
\end{aligned}$$

where the third inequality follows by Jensen's inequality. □

Appendix D

The Continuity of $\tilde{\gamma}_{i^*}(\gamma^*_{i^*})$

To stress their dependence on $(\theta_{i^*-1}, \theta_{i^*})$, we shall denote $\tilde{a}_{i^*}$, $\tilde{b}_{i^*}$, $\tilde{c}_{i^*}$, and $\hat{\gamma}_{i^*}$ by $\tilde{a}_{i^*}(\theta_{i^*-1}, \theta_{i^*})$, $\tilde{b}_{i^*}(\theta_{i^*-1}, \theta_{i^*})$, $\tilde{c}_{i^*}(\theta_{i^*-1}, \theta_{i^*})$, and $\hat{\gamma}_{i^*}(\theta_{i^*-1}, \theta_{i^*})$ respectively. Define regions $\mathcal{R}_1$ and $\mathcal{R}_2$ as follows:

$$\mathcal{R}_1 = \{(\theta_{i^*-1}, \theta_{i^*}) : \tilde{a}_{i^*}(\theta_{i^*-1}, \theta_{i^*}) > 0\},$$
$$\mathcal{R}_2 = \{(\theta_{i^*-1}, \theta_{i^*}) : \tilde{a}_{i^*}(\theta_{i^*-1}, \theta_{i^*}) \leq 0, \theta_{i^*-1} \in (0, \sigma^2_{X_{i^*}}]\}.$$

It is clear that $\tilde{\gamma}_{i^*}(\gamma^*_{i^*})$ varies continuously with $\gamma^*_{i^*}$ if $(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*}))$ moves inside one of these two regions. Therefore, we only need to consider the case where $(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*}))$ traverses through the boundary between $\mathcal{R}_1$ and $\mathcal{R}_2$.

Let $(\theta_{i^*-1}, \theta_{i^*})$ be a boundary point between $\mathcal{R}_1$ and $\mathcal{R}_2$. It is clear that $\tilde{a}_{i^*}(\theta_{i^*-1}, \theta_{i^*}) = 0$; moreover, it suffices to consider the case $\theta_{i^*-1} > 0$ since we have $\theta_{i^*} = 0$ if both $\tilde{a}_{i^*}(\theta_{i^*-1}, \theta_{i^*})$ and θ_{i^*-1} are zero. As a consequence, we have $(\theta_{i^*-1}, \theta_{i^*}) \in \mathcal{R}_2$. Note that $\tilde{\gamma}_{i^*}(\gamma^*_{i^*}) = d_{i^*}$ if $(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*})) \in \mathcal{R}_2$. On the other hand, as

$(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*}))$ moves toward $(\theta_{i^*-1}, \theta_{i^*})$ from the $\mathcal{R}_1$ side, we have

$$\tilde{a}_{i^*}(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*})) \to 0,$$
$$\tilde{c}_{i^*}(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*})) \to \tilde{c}_{i^*}(\theta_{i^*-1}, \theta_{i^*}) > 0;$$

moreover, since

$$\hat{\gamma}_{i^*}(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*})) \geq \sqrt{\frac{-\tilde{c}_{i^*}(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*}))}{\tilde{a}_{i^*}(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*}))}},$$

it follows that $\hat{\gamma}_{i^*}(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*})) \to \infty$, which further implies that

$$\tilde{\gamma}_{i^*}(\gamma^*_{i^*}) \triangleq \min\{\hat{\gamma}_{i^*}(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*})), d_{i^*}\} = d_{i^*}$$

when $(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*}))$ is sufficiently close to $(\theta_{i^*-1}, \theta_{i^*})$. Therefore, $\tilde{\gamma}_{i^*}(\gamma^*_{i^*})$ varies continuously with $\gamma^*_{i^*}$ when $(\theta^*_{i^*-1}(\gamma^*_{i^*}), \theta^*_{i^*}(\gamma^*_{i^*}))$ traverses through the boundary between $\mathcal{R}_1$ and $\mathcal{R}_2$.

Bibliography

[1] H. Viswanathan and T. Berger, "Sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 236-246, Jan. 2000.

[2] N. Ma and P. Ishwar, "On delayed sequential coding of correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3763-3782, Jun. 2011.

[3] E.-H. Yang, L. Zheng, D.-K. He, and Z. Zhang, "Rate distortion theory for causal video coding: characterization, computation algorithm, and comparison," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5258-5280, Aug. 2011.

[4] J. Wang and X. Wu, "Information flows in video coding," in *Proc. IEEE Data Comp. Conf.*, Snowbird, UT, USA, Mar. 24-26, 2010, pp. 149-158.

[5] I. E. Richardson, *The H. 264 Advanced Video Compression Standard*. Chichester, UK: John Wiley & Sons Ltb, 2010.

[6] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Int. Conv. Rec.*, part 4, 1959, vol. 7, pp. 142-163, reprinted with changes in *Information and Decision Processes*, R. E. Manchol, Ed. New York: McGraw-Hill, 1960, pp. 93-126.

[7] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851-857, Nov. 1982.

[8] Z. Zhang and T. Berger, "New results in binary multiple descriptios,", *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 502-521, 1987.

[9] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909-1921, Dec. 1980.

[10] J. Chen, "Rate region of Gaussian multiple description coding with individual and central distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3991-4005, Sep. 2009.

[11] H. Wang and P. Viswanath, "Vector Gaussian multiple description with individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2133-2153, Jun. 2007.

[12] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizer," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 428-436, Mar. 1992.

[13] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152-1159, Jul. 1996.

[14] R. Zamir and M. Feder, "Information rates of pre/post-filtered dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1340-1353, Sep. 1996.

[15] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839-1851, May 2007.

[16] H. Weingarten, T. Liu, S. Shamai (Shitz), Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple-input multiple-output compound broadcast channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5011-5023, Nov. 2009.

[17] R. Liu, T. Liu, H. V. Poor, and S. Shamai (Shitz), "A vector generalization of Costa's entropy-power inequality with applications," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1865-1879, Apr. 2010.

[18] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," *Inform. Contr.*, vol. 37, no. 1, pp. 34-39, Apr. 1978.

[19] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072-3081, Nov. 2001.

[20] S. Watanabe and Y. Oohama, "Secret key agreement from vector Gaussian sources by rate limited public communication," [Online]. Available: http://arxiv.org/abs/1009.5760.

[21] A. El Gamal and Y.-H Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.