Finite Blocklength Lossy Source Coding for Discrete Memoryless Sources

Other titles in Foundations and Trends® in Communications and Information Theory

Topics and Techniques in Distribution Testing: A Biased but Representative Sample
Clément L. Canonne
ISBN: 978-1-63828-100-9

Codes in the Sum-Rank Metric: Fundamentals and Applications
Umberto Martínez-Peñas, Mohannad Shehadeh and Frank R. Kschischang
ISBN: 978-1-63828-030-9

Codes for Distributed Storage
Vinayak Ramkumar, S. B. Balaji, Birenjith Sasidharan, Myna Vajha, M. Nikhil Krishnan and P. Vijay Kumar
ISBN: 978-1-63828-024-8

Rank-Metric Codes and Their Applications
Hannes Bartz, Lukas Holzbaur, Hedongliang Liu, Sven Puchinger, Julian Renner and Antonia Wachter-Zeh
ISBN: 978-1-63828-000-2

Common Information, Noise Stability, and Their Extensions
Lei Yu and Vincent Y. F. Tan
ISBN: 978-1-63828-014-9

Information-Theoretic Foundations of DNA Data Storage
Ilan Shomorony and Reinhard Heckel
ISBN: 978-1-68083-956-2

Finite Blocklength Lossy Source Coding for Discrete Memoryless Sources

Lin Zhou
School of Cyber Science and Technology
Beihang University
lzhou@buaa.edu.cn

Mehul Motani
Department of Electrical and Computer Engineering
National University of Singapore
motani@nus.edu.sg

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

L. Zhou and M. Motani. *Finite Blocklength Lossy Source Coding for Discrete Memoryless Sources*. Foundations and Trends® in Communications and Information Theory, vol. 20, no. 3, pp. 157–389, 2023.

ISBN: 978-1-63828-183-2

Foundations and Trends® in Communications and Information Theory
Volume 20, Issue 3, 2023

Editorial Scope

Topics

Foundations and Trends® in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2023, Volume 20, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328 . Also available as a combined paper and online subscription.

Contents

Finite Blocklength Lossy Source Coding for Discrete Memoryless Sources

Lin Zhou[1] and Mehul Motani[2]

[1] *School of Cyber Science and Technology, Beihang University, China; lzhou@buaa.edu.cn*

[2] *Department of Electrical and Computer Engineering, National University of Singapore, Singapore; motani@nus.edu.sg*

ABSTRACT

Shannon propounded a theoretical framework (collectively called information theory) that uses mathematical tools to understand, model and analyze modern mobile wireless communication systems. A key component of such a system is source coding, which compresses the data to be transmitted by eliminating redundancy and allows reliable recovery of the information from the compressed version. In modern 5G networks and beyond, finite blocklength lossy source coding is essential to provide ultra-reliable and low-latency communications. The analysis of point-to-point and multi-terminal settings from the perspective of finite blocklength lossy source coding is therefore of great interest to 5G system designers and is also related to other long-standing problems in information theory.

In this monograph, we review recent advances in second-order asymptotics for lossy source coding, which provides approximations to the finite blocklength performance of optimal codes. The monograph is divided into three parts. In

Lin Zhou and Mehul Motani (2023), "Finite Blocklength Lossy Source Coding for Discrete Memoryless Sources", Foundations and Trends® in Communications and Information Theory: Vol. 20, No. 3, pp. 157–389. DOI: 10.1561/0100000134.

Part I, we motivate the monograph, present basic definitions, introduce mathematical tools and illustrate the motivation of non-asymptotic and second-order asymptotics via the example of lossless source coding. In Part II, we first present existing results for the rate-distortion problem with proof sketches. Subsequently, we present five generations of the rate-distortion problem to tackle various aspects of practical quantization tasks: noisy source, noisy channel, mismatched code, Gauss-Markov source and fixed-to-variable length compression. By presenting theoretical bounds for these settings, we illustrate the effect of noisy observation of the source, the influence of noisy transmission of the compressed information, the effect of using a fixed coding scheme for an arbitrary source and the roles of source memory and variable rate. In Part III, we present four multiterminal generalizations of the rate-distortion problem to consider multiple encoders, decoders or source sequences: the Kaspi problem, the successive refinement problem, the Fu-Yeung problem and the Gray-Wyner problem. By presenting theoretical bounds for these multiterminal problems, we illustrate the role of side information, the optimality of stop and transmit, the effect of simultaneous lossless and lossy compression, and the tradeoff between encoders' rates in compressing correlated sources. Finally, we conclude the monograph, mention related results and discuss future directions.

Part I

Basics

1

Introduction

1.1 Motivation

Shannon [104] developed a theoretical framework (collectively called information theory) that uses mathematical tools to understand, model and analyze digital communication systems over noisy channels. A basic digital communication system includes blocks for source and channel encoding at the transmitter and blocks for source and channel decoding at the receiver. Source coding, also known as data compression, aims to remove the redundancy of information and allows reliable recovery of the information from its compressed version. In contrast, channel coding aims to counter the noise in the transmission channel between the transmitter and the receiver and allows reliable recovery of a message.

For discrete memoryless sources (DMS), Shannon showed that the asymptotic minimal compression rate that ensures accurate recovery with vanishing error probability is the entropy of the source, provided that the blocklength of the source sequence to be compressed tends to infinity. However, lossless source coding does not apply to continuous sources since it requires an infinite number of bits to describe a real number. Furthermore, practical image and video compression systems usually tolerate some imperfection. To resolve these issues, Shannon studied the lossy source coding problem [105] (also known as the rate-distortion problem) and derived the asymptotic minimal achievable rate.

For discrete memoryless channels (DMC), Shannon showed that the maximal asymptotic message rate to ensure reliable recovery with vanishing error probability at the receiver is the capacity of the noisy channel, provided that the blocklength (the number of channel uses) tends to infinity. In other words, Shannon showed that, at rates below the channel capacity, there exist good channel coding strategies with arbitrarily low probability of error. The above results for source coding and channel coding are collectively known as Shannon's coding theorems [19]. These results are very insightful and set benchmarks for practical code design in the last seventy years.

In practical communication systems, especially 5G and beyond, low-latency is desired and dictates the use of short blocklength codes. However, Shannon's coding theorems *cannot* provide exact theoretical benchmarks for low-latency communication since these theorems hold under the assumption that the blocklength tends to infinity, which leads to undesired arbitrarily large latency. To tackle this problem, information theorists have developed the theory of finite blocklength analyses and second-order asymptotics. This line of research dates back to Strassen [109] in 1962, who derived the dominant backoff term of the coding rate of an optimal code from the maximum asymptotic transmission rate—the capacity $C(P_{Y|X})$ of a noisy channel while tolerating a non-vanishing error probability when the blocklength increases. The result was recently rediscovered by Hayashi [50] and by Polyanskiy *et al.* [92]. In particular, Hayashi [50] named the result as second-order asymptotics since the dominant backoff term is exactly the *second* largest term in the expression of the maximal achievable rate of any code for the *asymptotic* case when the blocklength becomes large. In addition to the second-order asymptotics, the authors of [92] derived finite blocklength upper and lower bounds and showed that the bounds match the second-order asymptotics for blocklengths of hundreds for various types of point-to-point channels. Therefore, the results in [50], [92], [109], and especially [92], establish the critical role of second-order asymptotics in characterizing the finite blocklength performance of optimal codes and have been generalized to various channel models. Readers can refer to [112] for a systematic review of such advances.

Finite blocklength analyses and second-order asymptotics have also been derived for source coding. The simplest such example is the lossless source coding problem. In this problem, one aims to recover a random source sequence X^n exactly from its compressed version that takes values in a finite set of M elements. The performance metric is the error probability in reproducing the source sequence and the rate is defined as $R_n := \frac{\log M}{n}$, where the unit is bits per source symbol when the logarithm is base 2. In second-order asymptotics for lossless source coding, one is interested in characterizing the dominant backoff term of the coding rate R_n from the minimum asymptotic compression rate—the entropy of the source $H(P_X)$, while tolerating a non-vanishing error probability when the blocklength increases. Second-order asymptotics for lossless source coding was first shown by Yushkevich for sources with Markovian memory [142] and rediscovered by Strassen [109] and later Hayashi [49].

As noted by Shannon [105], lossless source coding is not possible for continuous sources and lossy source coding with imperfect recovery is thus important. Shannon's rate-distortion theory [105] forms a core part of modern quantization theory and is usually known as vector quantization. For a complete survey of various aspects of quantization, readers may refer to the seminal paper by Gray and Neuhoff [44]. For the rate-distortion problem that deals with point-to-point lossy data compression, the second-order asymptotics for DMS were derived by Ingber and Kochman [54], and both finite blocklength bounds and second-order asymptotics were derived by Kostina and Verdú [70] for DMS and Gaussian memoryless sources (GMS). The results in [54], [70] were further generalized to various scenarios in the point-to-point case [67], [71], [72], [117], [126], [152] and to problems in network information theory [82], [147]–[151], usually for DMS.

However, despite the undeniable importance of lossy source coding and its diverse applications beyond low-latency communications in various domains including privacy utility tradeoff [101], machine learning [39] and image/video compression [46], [86], [110], there is no publication that systematically summarizes recent advances for finite blocklength analyses and second-order asymptotics of lossy source coding problems, especially the multiterminal cases. One might argue that [112] covers these topics. Specifically, [112, Chapter 3] focuses on the point-to-point setting by presenting non-asymptotic and refined asymptotics bounds for both lossless and lossy source coding problems, [112, Chapter 4.5] briefly presents the results for joint source-channel coding without proof sketches while [112, Chapter 6] studies a lossless multiterminal source coding problem named the Slepian-Wolf problem [107]. It is important to note that recent advances of lossy source coding (e.g., [67], [72], [150], [152]) and the multiterminal cases [82], [147]–[151] are not included in [112]. Our monograph aims to fill the missing piece of finite blocklength analyses by summarizing recent theoretical advances for finite blocklength lossy source coding problems. Furthermore, for point-to-point lossless and lossy source coding problems, we present proof techniques different from those covered in [112, Chapter 3].

1.2 Organization

The rest of this monograph is organized as follows. In the rest of this section, we present the notation used throughout the monograph and recall critical mathematical theorems on sums of i.i.d. random variables including the Berry-Esseen theorem [12], [32]. In Section 2, we illustrate the meaning of finite blocklength analysis, first-order asymptotics, and second-order asymptotics via the example of lossless source coding. We also recall other refined asymptotics including large and moderate deviations and explain why we focus on second-order asymptotics.

Part II of this monograph is devoted to the rate-distortion problem and its five generalizations to consider various aspects of practical quantization tasks. Based on [54], [70], Section 3 reviews existing results on the rate-distortion problem. Specifically, we formulate the problem of finite blocklength analysis of the rate-distortion problem, define the distortion-tilted information density, present non-asymptotic and second-order asymptotic theorems, and finally provide detailed proof sketches.

Based on [72], Section 4 deals with the noisy lossy source coding problem, where the encoder can only access a noisy version of the source sequence. This problem is also known as quantizing noisy sources and is motivated by practical compression of speech signals distorted by environmental noise or images corrupted by camera imperfections. The non-asymptotic and second-order asymptotic results for this problem reveal the role of noisy observations in the finite blocklength regime, which is not apparent in asymptotic analyses [29], [100], [130].

Based on [71], [126], Section 5 deals with the lossy joint source-channel coding problem, where the output of the encoder is passed though a noisy channel and then provided to the decoder. This problem is also known as quantization for a noisy channel. The classical separation theorem of Shannon establishes that it is asymptotically optimal to separate lossy source coding and channel coding. However, non-asymptotic and second-order asymptotic results suggest that, at finite blocklengths, separate source-channel coding is strictly suboptimal.

Based on [152], Section 6 deals with mismatched compression of Lapidoth [76, Theorem 3], where a fixed code with an i.i.d. Gaussian codebook and minimum Euclidean distance encoding is used to compress an arbitrary memoryless source. This problem is motivated by the fact that the distribution of the source to be compressed is usually unknown and thus the matched coding scheme where the source distribution is assumed perfectly known is impractical. Theoretical results demonstrate that both i.i.d. Gaussian and spherical codebooks achieve the same finite blocklength performance.

Based on [117], Section 7 deals with the Gauss-Markov source, where the source sequence forms a first-order Markov chain and thus has memory. This problem is motivated by practical applications where the source sequence, such as sensor data, is usually not memoryless. The non-asymptotic and second-order results for the Gauss-Markov source is the first for a source with memory and reveal the role of memory on the finite blocklength performance of optimal codes.

Based on [67], Section 8 deals with fixed-to-variable length compression, where the encoder's output to each source sequence is a binary string with potentially different lengths. The motivation is to further reduce the average coding rate based on the intuition that more frequent symbols should be assigned codewords with fewer bits, an idea captured in the Huffman code. The theoretical results reveal the role of flexible rates on the finite blocklength performance and demonstrate a stark difference with the fixed-length counterpart.

Part III deals with four multiterminal extensions of the rate-distortion problem with increasing complexity and also includes a conclusion section. Based on the first part of [147], Section 9 deals with the Kaspi problem [56], which is a lossy source coding problem with one encoder and two decoders. This problem generalizes the rate-distortion problem by providing side information at the encoder and adding one additional decoder that accesses the same side information. Both decoders share the same compressed information of the source sequence and the decoder with side information is required to produce a finer estimate of the source sequence. Through the lens of this problem, we reveal the impact of side information on the finite blocklength performance of optimal codes.

Based on [82], [151], Section 10 deals with the successive refinement problem [95]. This problem generalizes the rate-distortion problem by having one additional encoder and decoder pair. The additional encoder further compresses the source sequence and the additional decoder uses compressed information from both encoders to produce a finer estimate of the source sequence than the other decoder that only has access to the original encoder. We present results under two performance criteria: the joint excess-distortion probability (JEP) and the separate excess-distortion probabilities (SEP). Under JEP, we reveal the tradeoff between the coding rate of the two encoders and, under SEP, we revisit the successively refinability property, from a second-order asymptotic perspective. A key message from this section is that considering a joint excess-distortion probability enables us to characterize the tradeoff of rates of different encoders in second-order asymptotics.

Based on the second part of [147], Section 11 deals with the multiple description problem with one deterministic decoder [34]. In this problem, two encoders compress the source sequence and three decoders aim to recover the source sequence with different criteria: two decoders aim to recover the source sequence in a lossy manner with different distortion levels and the other decoder aims to perfectly reproduce a function of the source sequence. This problem generalizes the successive refinement problem by having one additional lossless decoder. Under

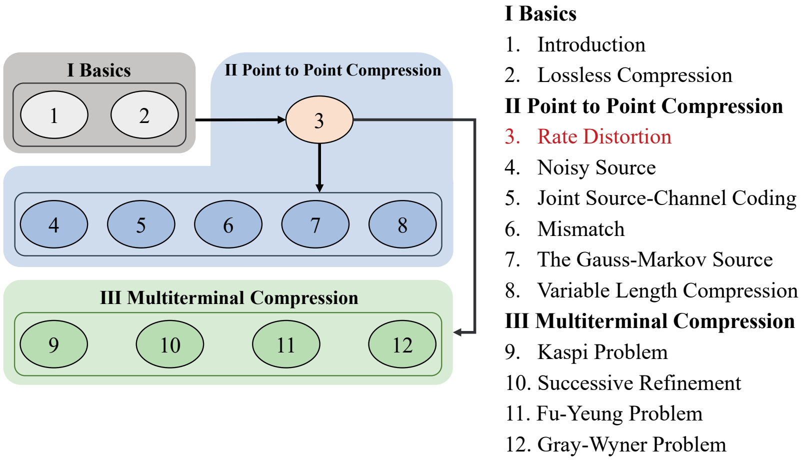

Figure 1.1: Relationship among the sections of this monograph.

the joint excess-distortion and error probability criterion, we reveal the tradeoff among encoders and decoders in simultaneous lossless and lossy compression in second-order asymptotics.

Based on [150], Section 12 deals with the lossy Gray-Wyner problem [43]. In this problem, three encoders compress two correlated source sequences and each of the two decoders aims to recover one source sequence. This is a fully multiterminal lossy compression problem with multiple encoders, multiple decoders and multiple correlated source sequences. It significantly generalizes the rate-distortion problem by having one more source sequence, two more encoders and one more decoder. Under the joint excess-distortion probability criterion, we reveal the tradeoff among the coding rates of the three encoders in second-order asymptotics.

Finally, in Section 13, we conclude the monograph and discuss future research directions. The relationship among sections of this monograph is illustrated in Figure 1.1.

1.3 Preliminaries

In this section, we set up the mathematical notation used throughout the monograph and review definitions of basic information theoretical quantities, key properties in method of types and mathematical theorems central to our analyses.

1.3.1 Notation

The set of real numbers, non-negative real numbers, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{N}$, respectively. For any two natural numbers $(a,b)\in\mathbb{N}^2$, we use $[a:b]$ to denote the set of all natural numbers between a and b (inclusive) and use $[a]$ to denote $[1:a]$. For any $(m_1,m_2)\in\mathbb{N}^2$, we use $\mathbf{0}_{m_1}$ to denote the length-m_1 vector of all zeroes and use $\mathbf{1}_{m_1,m_2}$ to denote the $m_1\times m_2$ matrix of all ones. For any real number $a\in\mathbb{R}$, we use $|a|^+$ to denote $\max\{a,0\}$.

Random variables and their realizations are in capital (e.g., X) and lower case (e.g., x) respectively. All sets (e.g., alphabets of random variables) are denoted in calligraphic font (e.g., $\mathcal{X}$). We use $\mathcal{X}^{\mathrm{c}}$ to denote the complement of $\mathcal{X}$. Let $X^n:=(X_1,\ldots,X_n)$ be a random vector of length-n and $x^n=(x_1,\ldots,x_n)$ be a particular realization. We use $\|x^n\|=\sqrt{\sum_{i\in[n]}x_i^2}$ to denote the ℓ_2 norm of a vector $x^n\in\mathbb{R}^n$. Given two sequences x^n and y^n, the quadratic distortion measure (squared Euclidean norm) is defined as $d(x^n,y^n):=\frac{1}{n}\|x^n-y^n\|^2=\frac{1}{n}\sum_{i\in[n]}(x_i-y_i)^2$.

The set of all probability distributions on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$ and the set of all conditional probability distribution from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. Given $P\in\mathcal{P}(\mathcal{X})$, we use $\mathrm{supp}(P)$ to denote the support of distribution P, i.e., $\mathrm{supp}(P)=\{x\in\mathcal{X}:P(x)>0\}$. Given a conditional distribution $P_{Y|X}\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $x\in\mathcal{X}$, we use $P_{Y|x}$ to denote the conditional distribution $P_{Y|X}(\cdot|x)$. Given $P\in\mathcal{P}(\mathcal{X})$ and $V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, we use $P\times V$ to denote the joint distribution induced by P and V. Given a joint probability distribution $P_{XY}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, let $m=|\mathrm{supp}(P_{XY})|$ and let $\Gamma(P_{XY})$ be the sorted distribution such that for each $i\in[m]$, $\Gamma_i(P_{XY})=P_{XY}(x_i,y_i)$ is the i-th largest value of $\{P_{XY}(x,y):\ (x,y)\in\mathcal{X}\times\mathcal{Y}\}$.

We use standard asymptotic notations such as $\Theta(\cdot)$, $O(\cdot)$ and $o(\cdot)$ (cf. [18]). We use $\mathbb{1}(\cdot)$ as the indicator function and we use $\log(\cdot)$ with base e unless otherwise stated. We let $\mathrm{Q}(t):=\int_t^\infty\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,\mathrm{d}u$ be the complementary cumulative distribution function of the standard Gaussian. Let Q^{-1} be the inverse of Q. We use $\Psi_k(x_1,\ldots,x_k;\boldsymbol{\mu},\boldsymbol{\Sigma})$ to denote the multivariate generalization of the Gaussian cumulative distribution function (cdf), i.e., $\Psi(x_1,\ldots,x_k;\boldsymbol{\mu},\boldsymbol{\Sigma}) = \int_{-\infty}^{x_1}\ldots\int_{-\infty}^{x_k}\mathcal{N}(\mathbf{x};\boldsymbol{\mu};\boldsymbol{\Sigma})\,\mathrm{d}\mathbf{x}$, where $\mathcal{N}(\mathbf{x};\boldsymbol{\mu};\boldsymbol{\Sigma})$ is the probability density function (PDF) of a k-variate Gaussian with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

1.3.2 Basic Definitions

To smoothly present the results in this monograph, we recall necessary information theoretical definitions. Given any distribution $P_X \in \mathcal{P}(\mathcal{X})$ defined on a finite alphabet $\mathcal{X}$, the entropy is defined as

$$H(X) = H(P_X) := \sum_{x\in\mathrm{supp}(P_X)} -P_X(x)\log P_X(x). \tag{1.1}$$

Note that the notation $H(X)$ is used in classical textbooks as [19] and the notation $H(P_X)$ that clarifies the dependence of the entropy on the distribution is used in [22]. We use both notations for the entropy and other information theoretical quantities interchangeably. Specifically, when we need to specify the distribution of a random variable, we use the distribution dependence version $H(P_X)$; when the distribution of the random variable is clear, we use $H(X)$ for its simplicity. Analogously, given a joint probability mass function (PMF) $P_{XY} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ defined on a finite alphabet $\mathcal{X}\times\mathcal{Y}$, the joint entropy is defined as

$$H(X,Y) = H(P_{XY}) = \sum_{(x,y)\in\mathrm{supp}(P_{XY})} -P_{XY}(x,y)\log P_{XY}(x,y), \tag{1.2}$$

and the conditional entropy of Y given X is defined as

$$H(Y|X) = H(P_{Y|X}|P_X) = \sum_{(x,y)\in\mathrm{supp}(P_{XY})} -P_{XY}(x,y)\log P_{Y|X}(x,y), \tag{1.3}$$

where $(P_{Y|X}, P_X)$ are the induced conditional and marginal distributions of P_{XY}. The conditional entropy $H(P_{Y|X}|P_X)$ of X given Y is defined similarly.

Furthermore, the mutual information that measures dependence of two random variables (X,Y) with distribution P_{XY} is defined as

$$I(X;Y) = I(P_X, P_{X|Y}) = H(P_X) - H(P_{X|Y}|P_Y), \tag{1.4}$$

where $P_{X|Y}$ is also induced by P_{XY}. Note that mutual information $I(X;Y)$ is symmetric so that $I(X;Y) = I(Y;X)$. Similar to the definition of entropy, we use $I(X;Y)$ and the distribution dependence version $I(P_X, P_{X|Y})$ interchangeably. Analogously, given the joint distribution P_{XYZ} of three random variables (X, Y, Z) defined on a finite alphabet $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, define the conditional mutual information $I(X;Y|Z)$ as

$$I(X;Y|Z) = I(P_{X|Z}, P_{X|YZ}|P_Z) = H(P_{X|Z}|P_Z) - H(P_{X|YZ}|P_{YZ}), \tag{1.5}$$

where all distributions are induced by the joint distribution P_{XYZ}.

Another critical quantity that we use frequency is the Kullback-Leibler (KL) divergence, also known as the relative entropy. Given any two distributions (P_X, Q_X) defined on the finite alphabet $\mathcal{X}$, the KL divergence $D(P_X\|Q_X)$ is defined as

$$D(P_X\|Q_X) = \sum_{x \in \mathrm{supp}(P_X)} P_X(x) \log \frac{P_X(x)}{Q_X(x)}. \tag{1.6}$$

Note that $D(P_X\|Q_X)$ measures closeness of two distributions P_X and Q_X and equals zero if and only if $P_X = Q_X$. For any two distributions P_{XY} and Q_{XY} defined on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, the KL divergence $D(P_{XY}\|Q_{XY})$ is defined similarly; when the marginal distributions $P_X = Q_X$, the conditional KL divergence is defined as

$$D(P_{Y|X}\|Q_{Y|X}|P_X) = \sum_{x \in \mathrm{supp}(P_X)} P_X(x) D(P_{Y|X}(\cdot|x)\|Q_{Y|X}(\cdot|x)). \tag{1.7}$$

1.3.3 The Method of Types

Since we focus on DMS, the method of types plays a critical role in our analyses. Thus, we also recall definitions and results in this domain [21] (see also [19, Chapter 11] and [22, Chapter 2]). Given a length-n discrete sequence $x^n \in \mathcal{X}^n$, the empirical distribution $\hat{T}_{x^n}$ is defined as

$$\hat{T}_{x^n}(a) = \frac{1}{n} \sum_{i \in [n]} 1\{x_i = a\}, \ \forall\ a \in \mathcal{X}. \tag{1.8}$$

The set of types formed from length-n sequences in $\mathcal{X}$ is denoted by $\mathcal{P}_n(\mathcal{X})$. Given a type $P_X \in \mathcal{P}_n(\mathcal{X})$, the set of all sequences of length-n

with type P_X is the type class denoted by $\mathcal{T}_{P_X}$. For any $n \in \mathbb{N}$, the number of types satisfies

$$|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}. \tag{1.9}$$

For any type $P_X \in \mathcal{P}_n(\mathcal{X})$, the size of type class $\mathcal{T}_P^n$ satisfies

$$(n+1)^{-|\mathcal{X}|} \exp(nH(P_X)) \leq |\mathcal{T}_{P_X}^n| \leq \exp(nH(P_X)). \tag{1.10}$$

For any sequence x^n that is generated i.i.d. from a distribution $P_X \in \mathcal{P}(\mathcal{X})$, its probability satisfies

$$P_X^n(x^n) = \exp(-n(D(\hat{T}_{x^n} \| P_X) + H(\hat{T}_{x^n}))). \tag{1.11}$$

Thus, for any type $Q_X \in \mathcal{P}_n(\mathcal{X})$, the probability of the type class $\mathcal{T}_{Q_X}^n$ satisfies

$$(n+1)^{-|\mathcal{X}|} \leq \frac{P_X^n(\mathcal{T}_{Q_X}^n)}{\exp(-nD(Q_X \| P_X))} \leq 1. \tag{1.12}$$

Given any two sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ for any $(a,b) \in \mathcal{X} \times \mathcal{Y}$, the joint empirical distribution $\hat{T}_{x^n y^n}$ is defined as

$$\hat{T}_{x^n y^n}(a,b) = \frac{1}{n} \sum_{i \in [n]} 1\{(x_i, y_i) = (a,b)\}. \tag{1.13}$$

Given any $x^n \in \mathcal{X}^n$ and conditional distribution $V_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, the set of all sequences $y^n \in \mathcal{Y}^n$ such that $\hat{T}_{x^n y^n} = \mathcal{T}_{x^n} \times V_{Y|X}$ is the conditional type class denoted by $\mathcal{T}_{V_{Y|X}}(x^n)$. For any $x^n \in \mathcal{T}_{P_X}^n$, the set of all conditional distributions $V_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ such that the conditional type class $\mathcal{T}_{V_{Y|X}}(x^n)$ is not empty is the set of conditional types given the marginal type P_X and is denoted by $\mathcal{V}^n(\mathcal{Y}; P_X)$.

The following results hold. For any $P_X \in \mathcal{P}_n(\mathcal{X})$, the number of conditional types is upper bounded by

$$|\mathcal{V}^n(\mathcal{Y}; P_X)| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|}. \tag{1.14}$$

For any $x^n \in \mathcal{T}_{P_X}^n$, the size of the conditional type class $\mathcal{T}_{V_{Y|X}}(x^n)$ satisfies

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp(nH(V_{Y|X})|P_X) \leq |\mathcal{T}_{V_{Y|X}}(x^n)| \leq \exp(nH(V_{Y|X})|P_X). \tag{1.15}$$

Given any $x^n \in \mathcal{T}_{P_X}^n$, $W_{Y|X} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $V_{Y|X} \in \mathcal{V}^n(\mathcal{Y}; P_X)$, for any $y^n \in \mathcal{T}_{V_{Y|X}}(x^n)$,

$$W_{Y|X}^n(y^n|x^n) = \exp(-n(H(V_{Y|X}) + D(V_{Y|X}\|W_{Y|X})|P_X)). \quad (1.16)$$

Thus, it follows from (1.15) and (1.16) that

$$(n+1)^{-|\mathcal{X}||\mathcal{Y}|} \le \frac{W_{Y|X}^n(\mathcal{T}_{V_{Y|X}}(x^n))}{\exp(-nD(V_{Y|X}\|W_{Y|X})|P_X))} \le 1 \quad (1.17)$$

1.3.4 Mathematical Tools

In this section, we present the mathematical tools used to prove second-order asymptotics, which are essentially the generalization of central limit theorems. Let $X^n = (X_1, \ldots, X_n)$ be a collection of n i.i.d. random variables with zero mean and variance σ^2 and let the normalized sum of these n random variables be

$$S_n := \frac{1}{n} \sum_{i \in [n]} X_i. \quad (1.18)$$

We first recall the weak law of large numbers [33], which states that the normalized sum S_n converges in probability to its mean.

Theorem 1.1 (The Weak Law of Large Numbers)**.** For any positive real number $\delta \in \mathbb{R}_+$,

$$\lim_{n\to\infty} \Pr\{S_n > \delta\} = 0. \quad (1.19)$$

In the proofs of many theorems, the Markov inequality is used.

Theorem 1.2 (The Markov Inequality)**.** For any non-negative real number $\theta \in \mathbb{R}_+$ and any positive real number t,

$$\Pr\{S_n > t\} \le \frac{\mathsf{E}[\exp(\theta S_n)]}{\exp(t\theta)}. \quad (1.20)$$

The Berry-Esseen Theorem for i.i.d. random variables [12], [32] is critical in deriving second-order asymptotics.

Theorem 1.3 (The Berry-Esseen Theorem)**.** Assume that the third absolute moment of X_1 is finite, i.e., $T := \mathbb{E}|X_1|^3 < \infty$. For each $n \in \mathbb{N}$,

$$\sup_{t\in\mathbb{R}} \left| \Pr\left\{ S_n \geq t\sqrt{\frac{\sigma^2}{n}} \right\} - \mathrm{Q}(t) \right| \leq \frac{T}{\sigma^3\sqrt{n}}. \tag{1.21}$$

The Berry-Esseen theorem states that the probability that the normalized sum S_n deviates from its mean by a sequence which scales as $\Theta\left(\frac{1}{\sqrt{n}}\right)$ is well approximated by the same probability for a standard normal variable, with the difference in the order of $O\left(\frac{1}{\sqrt{n}}\right)$ that depends on the variance σ^2 and the third absolute moment T. The assumption that T is finite is satisfied for DMS. It is the mathematical theorem that one applies in the analysis of second-order asymptotics for source and channel coding problems that involve a single encoder.

To tackle certain problems, we need to consider independent but not identically distributed (i.n.i.d.) random variables. Let $X^n = (X_1, \ldots, X_n)$ be a sequence of random variables, where each random variable X_i has zero mean, variance $\sigma_i^2 := \mathsf{E}[X_i^2] > 0$ and finite third-absolute moment $T_i := \mathsf{E}[|X_i|^3]$. Define the average variance and third-absolute moment as follows:

$$\sigma^2 := \frac{1}{n}\sum_{i\in[n]} \sigma_i^2, \tag{1.22}$$

$$T := \frac{1}{n}\sum_{i\in[n]} T_i \tag{1.23}$$

The Berry-Esseen theorem for i.n.i.d. random variables states as follows.

Theorem 1.4. For each $n \in \mathbb{N}$,

$$\sup_{t\in\mathbb{R}} \left| \Pr\left\{ S_n \geq t\sqrt{\frac{\sigma^2}{n}} \right\} - \mathrm{Q}(t) \right| \leq \frac{6T}{\sigma^3\sqrt{n}}. \tag{1.24}$$

To derive results for multiterminal lossy source coding problems with multiple encoders, we need the following multivariate generalization of the Berry-Esseen theorem [40]. Given $d \in \mathbb{N}$, for each $i \in [n]$, let $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,k})$ be a k-dimensional random vector with zero mean vector and covariance matrix $\boldsymbol{\Sigma}$. Let the normalized sum vector be $\mathbf{S}_n := \frac{1}{\sqrt{n}}\mathbf{X}_i$.

Theorem 1.5 (Vector Version of the Berry-Esseen Theorem)**.** Let the third absolute moment of $\mathbf{X}_1$ be $T := \mathbb{E}[\|\mathbf{X}_1\|^3]$. For each $n \in \mathbb{N}$, we have

$$\sup_{(t_1,\dots,t_d)\in\mathbb{R}^d} \Big| \Pr\{\mathbf{S}_n \le \mathbf{t}\} - \Psi_k(t_1,\dots,t_k;\mathbf{0}_k,\mathbf{\Sigma}) \Big| \le \frac{K(d)T}{\sqrt{n}}, \tag{1.25}$$

where $>$ refers to elementwise comparison and $K(d)$ is a constant depending on the dimension d only (see [7], [94] for explicit bounds).

2

Lossless Compression

This section focuses on lossless source coding, the notably simplest problem in vector quantization. In his seminal 1948 paper [104], Shannon showed that the minimal compression rate for reliable lossless source coding is the entropy of the discrete memoryless source, assuming that the blocklength of the source to be compressed tends to infinity. Inspired by the low-latency requirement of practical communications systems, one wonders what the performance degradation is if one operates at a finite blocklength. This question was answered by Yushkevich [142] and by Strassen [109] who derived the second-order asymptotic approximation to the finite blocklength performance, revived by Hayashi [50] who rediscovered the result using the information spectrum method and further refined by Kontoyiannis and Verdú [59] and by Chen *et al.* [16] who improved the previous bounds.

In this section, we present finite blocklength and second-order asymptotic bounds for lossless source coding, demonstrate the tightness of the second-order asymptotics and discuss the relationship of second-order asymptotics and other refined asymptotic analyses. This section is largely based on [50], [109].

2.1 Problem Formulation and Shannon's Result

Consider any length-n source sequence X^n that is generated i.i.d. from a probability mass function (PMF) $P_X \in \mathcal{P}(\mathcal{X})$. In lossless source coding, one is interested in perfectly recovering the source sequence X^n from its compressed version. Formally, a code is defined as follows.

Definition 2.1. Given any $(n, M) \in \mathbb{N}^2$, an (n, M)-code for source coding consists of

- an encoder $f : \mathcal{X}^n \to \mathcal{M} := [1 : M]$,
- a decoder $\phi : \mathcal{M} \to \mathcal{X}^n$.

For simplicity, we use $\hat{X}^n$ to denote the reproduced source sequence at the decoder, i.e., $\hat{X}^n = \phi(f(X^n))$. The performance metric for lossless source coding is the error probability, i.e.,

$$\mathrm{P}_{\mathrm{e},n} := \Pr\{\phi(f(X^n)) \neq X^n\} \tag{2.1}$$

$$= \Pr\{\hat{X}^n \neq X^n\}. \tag{2.2}$$

In the above definition, n is the blocklength of the source sequence and M is the number of codewords that encoder can use.

To achieve zero error, M should be chosen such that $M \geq |\mathcal{X}|^n$ to allow one to one mapping. However, this means no compression is done. Thus, to compress the source, we need to tolerate a non-zero error probability. For efficient compression, one hopes M is as small as possible given any blocklength n and error probability $\mathrm{P}_{\mathrm{e},n}$. To capture the fundamental limit of lossless source coding, for any $n \in \mathbb{N}$, let $M^*(n, \varepsilon)$ be the minimum number of codewords such that there exists an (n, M)-code satisfying $\mathrm{P}_{\mathrm{e},n} \leq \varepsilon$, i.e.,

$$M^*(n, \varepsilon) := \inf\left\{M : \ \exists \text{ an } (n, M)-\text{code s.t. } \mathrm{P}_{\mathrm{e},n} \leq \varepsilon\right\}. \tag{2.3}$$

Ideally, one would like to exactly characterize $M^*(n, \varepsilon)$ for each finite $n \in \mathbb{N}$ and any tolerable error probability $\varepsilon \in (0, 1)$. But this is very challenging and information theorists instead derived approximations to $M^*(n, \varepsilon)$.

The most famous such approximation for lossless source coding was provided by Shannon [104], which states that

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log M^*(n,\varepsilon)=H(P_X). \tag{2.4}$$

The above result means that to achieve vanishing error probability with respect to the blocklength n, the average minimal number of bits that one should use to compress a source symbol equals the entropy $H(P_X)$ of the source. The above result is also known as the first-order asymptotics since it characterizes the first dominant term in the expansion of the non-asymptotic rate $R(n,\varepsilon):=\frac{1}{n}\log M^*(n,\varepsilon)$ of an optimal code when $\varepsilon\to 0$. In fact, the above result holds for any $\varepsilon\in(0,1)$, which is known as strong converse and implied by second-order asymptotics.

2.2 Non-Asymptotic Bounds

Second-order asymptotics provides approximation to the finite blocklength performance $M^*(n,\varepsilon)$, which demonstrates a deeper understanding for the interplay among the blocklength, the error probability and the coding rate. Usually, to obtain second-order asymptotics, one first derives non-asymptotic achievability and converse bounds for any finite blocklength n and next apply the Berry-Esseen theorem to the derived bounds appropriately.

In [112, Sections 3.1-3.2], the non-asymptotic and second-order asymptotic bounds by Strassen [109] were presented and in [112, Section 3.3], an alternative proof of second-order asymptotic using the method of types [21], [22] was given. In this section, we present the non-asymptotic bounds of Han [48] based on the information spectrum method and provide an alternative proof of second-order asymptotics using Han's results.

Given any $x\in\mathcal{X}$, define the entropy density $\imath(x|P_X)$ as

$$\imath(x|P_X):=-\log P_X(x). \tag{2.5}$$

We first recall a finite blocklength achievability bound [48, Lemma 1.3.1].

Theorem 2.1. For any $(n, M) \in \mathbb{N}^2$, there exists an (n, M)-code whose error probability is upper bounded by

$$\mathrm{P}_{\mathrm{e},n} \le \Pr\Big\{ \sum_{i\in[n]} \imath(X_i|P_X) \ge \log M \Big\}. \tag{2.6}$$

The proof of Theorem 2.1 is simple and elegant. For completeness, we present the proof here.

Proof. For any $n \in \mathbb{N}$, define a set

$$\mathcal{A}_n := \Big\{ x^n \in \mathcal{X}^n : \sum_{i\in[n]} \imath(x_i|P_X) < \log M \Big\}. \tag{2.7}$$

Note that if $x^n \in \mathcal{A}^n$, we have

$$P_X^n(x^n) = \prod_{i\in[n]} P_X(x_i) \tag{2.8}$$

$$= \prod_{i\in[n]} \exp(-\imath(x_i|P_X)) \tag{2.9}$$

$$= \exp\big(- \sum_{i\in[n]} \imath(x_i|P_X)\big) \tag{2.10}$$

$$> \frac{1}{M}. \tag{2.11}$$

It follows that

$$1 \ge \sum_{x^n\in\mathcal{A}_n} P_X^n(x^n) \tag{2.12}$$

$$\ge \sum_{x^n\in\mathcal{A}^n} \frac{1}{M} \tag{2.13}$$

$$= \frac{|\mathcal{A}_n|}{M}. \tag{2.14}$$

Thus, $|\mathcal{A}_n| \le M$. Then we can construct an (n, M)-code where the encoder f encodes each element of $\mathcal{A}_n$ to a unique number in $[|\mathcal{A}_n|]$ and declares an error otherwise. This way, the number of codewords required is $|\mathcal{A}_n| \le M$ and the error probability satisfies (2.6). □

We next recall the finite blocklength converse bound [48, Lemma 1.3.2], which presents a lower bound on the error probability of any (n, M)-code.

Theorem 2.2. For any $(n, M) \in \mathbb{N}^2$ and $\gamma \in \mathbb{R}_+$, any (n, M)-code satisfies

$$\mathrm{P}_{\mathrm{e},n} \geq \Pr\Big\{ \sum_{i\in[n]} \imath(X_i|P_X) \geq \log M + n\gamma \Big\} - \exp(-n\gamma). \tag{2.15}$$

The proof of Theorem 2.2 is similar to that of Theorem 2.1 and is also recalled here.

Proof. Analogously to $\mathcal{A}^n$ in (2.7), for any $\gamma \in \mathbb{R}$, define a set

$$\mathcal{B}_n(\gamma) := \Big\{ x^n \in \mathcal{X}^n : \sum_{i\in[n]} \imath(x_i|P_X) \geq \log M + n\gamma \Big\}. \tag{2.16}$$

Furthermore, define the set of correctly decoded source sequences as

$$\mathcal{C}_n := \{x^n \in \mathcal{X}^n : \phi(f(x^n)) = x^n\}. \tag{2.17}$$

Then,

$$\begin{aligned}
&\Pr\{X^n \in \mathcal{B}_n(\gamma)\} \\
&= \Pr\{X^n \in (\mathcal{B}_n(\gamma) \cap \mathcal{C}_n^{\mathrm{c}})\} + \Pr\{X^n \in (\mathcal{B}_n(\gamma) \cap \mathcal{C}_n)\} && (2.18)\\
&\leq \Pr\{X^n \in \mathcal{C}_n^{\mathrm{c}}\} + \Pr\{X^n \in (\mathcal{B}_n(\gamma) \cap \mathcal{C}_n)\} && (2.19)\\
&\leq \mathrm{P}_{\mathrm{e},n} + \Pr\{X^n \in (\mathcal{B}_n(\gamma) \cap \mathcal{C}_n), && (2.20)
\end{aligned}$$

where (2.20) follows from the definition of the error probability $\mathrm{P}_{\mathrm{e},n}$. Similar to (2.11), if $x^n \in \mathcal{B}_n(\gamma)$,

$$\begin{aligned}
P_X(x^n) &= \exp\Big(-\sum_{i\in[n]} \imath(x_i|P_X) \Big) && (2.21)\\
&\leq \frac{\exp(-n\gamma)}{M}. && (2.22)
\end{aligned}$$

It follows that

$$\begin{aligned}
\Pr\{X^n \in (\mathcal{B}_n(\gamma) \cap \mathcal{C}_n)\} &= \sum_{x^n \in (\mathcal{B}_n(\gamma)\cap\mathcal{C}_n)} P_X^n(x^n) && (2.23)\\
&\leq \sum_{x^n \in (\mathcal{B}_n(\gamma)\cap\mathcal{C}_n)} \frac{\exp(-n\gamma)}{M} && (2.24)\\
&\leq \frac{|\mathcal{C}_n| \exp(-n\gamma)}{M} && (2.25)\\
&\leq \exp(-n\gamma), && (2.26)
\end{aligned}$$

where (2.26) follows since for any (n, M)-code, the number of corrected decoded source sequences is no greater than M. □

In subsequent sections, the proofs of Theorems 2.1 and 2.2 are generalized to obtain finite blocklength bounds for lossy source coding problems, which are also known as lossy vector quantization [44].

2.3 Second-Order Asymptotics

Applying the Berry-Esseen theorem to the finite blocklength bounds in Theorems 2.1 and 2.2, one can obtain the second-order asymptotics, which provides a finer characterization of $M^*(n, \varepsilon)$ in (2.3) beyond Shannon's classical first-order asymptotic result. To present the result, define the dispersion of the source P_X as

$$\mathrm{V}(P_X) := \mathrm{Var}[-\log P_X(X)]. \tag{2.27}$$

Theorem 2.3. For any $\varepsilon \in (0, 1)$,

$$\log M^*(n, \varepsilon) = nH(P_X) + \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon) + O(\log n). \tag{2.28}$$

We remark that Theorem 2.3 was first obtained by Yushkevich [142] for a Markov source and by Strassen [109] for DMS. Hayashi [49] rediscovered Theorem 2.3. The $O(\log n)$ term was found to be $-\frac{1}{2}\log n + O(1)$ by Kontoyiannis and Verdú [59] and was recently further refined by Chen *et al.* [16, Theorem 5] with explicit lower and upper bounds on the $O(1)$ term.

Furthermore, the achievability part of Theorem 2.3 can also be proved using the method of types [19, Chapter 11], as demonstrated in [112, Chapter 3.3]. The achievability proof of second-order asymptotics based on the method of types finds applications in many other problems, including the point-to-point and multiterminal settings of lossy source coding problems to be discussed in this monograph.

To illustrate the tightness of the second-order asymptotic bound in Theorem 2.3, in Figure 2.1, we plot the second-order asymptotic approximation in Theorem 2.3 and compare the approximation with finite blocklength bounds in Theorems 2.1 and 2.2 for a Bernoulli source with parameter 0.2 with the target error probability of $\varepsilon = 0.01$. As

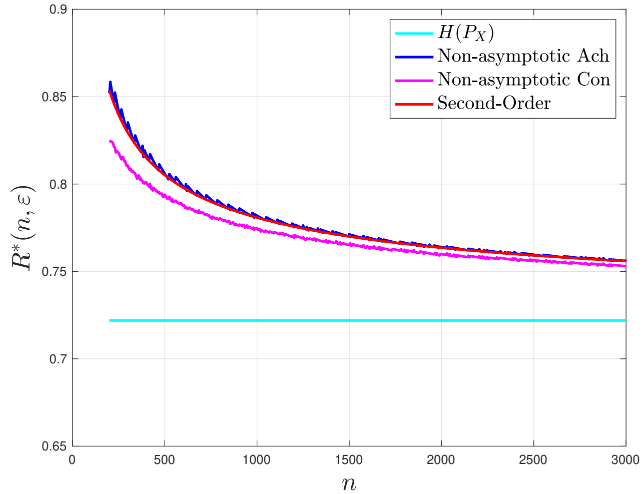

Figure 2.1: The second-order asymptotic bound in Theorem 2.3 and non-asymptotic bounds in Theorems 2.1 and 2.2 for a Bernoulli source with parameter 0.2 and a target error probability of $\varepsilon = 0.01$.

observed from Figure 2.1, for n moderately large, the second-order asymptotic bound provides rather tight approximation to the finite blocklength performance. Furthermore, the gap between the second-order asymptotic result and the first-order asymptotic result of Shannon is significant unless $n \to \infty$.

Note that Theorem 2.3 is known as the second-order asymptotic result because it characterizes the second dominant term in the expansion of $\log M^*(n, \varepsilon)$. An equivalent presentation of Theorem 2.3 is to characterize the so called second-order coding rate coined by Hayashi [49], [50]. For lossless source coding, the second-order coding rate is defined as follows.

Definition 2.2. Given any $\varepsilon \in [0, 1)$, a real number $L \in \mathbb{R}$ is said to be a second-order achievable rate if there exists a sequence of (n, M)-codes such that

$$\limsup_{n \to \infty} \frac{1}{\sqrt{n}} (\log M - nH(P_X)) \leq L, \tag{2.29}$$

$$\limsup_{n \to \infty} \mathrm{P}_{\mathrm{e},n} \leq \varepsilon. \tag{2.30}$$

For any $\varepsilon \in [0,1)$, the infimum of all second-order achievable rates is called the optimal second-order coding rate and denoted by $L^*(\varepsilon)$.

We remark that $L^*(\varepsilon)$ has the unit of nats per square root number of source symbols. With this definition, Theorem 2.3 is equivalent to the following statement.

Theorem 2.4. For any $\varepsilon \in [0,1)$, the optimal second-order rate coding is

$$L^*(\varepsilon) = \sqrt{\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon). \tag{2.31}$$

In second-order asymptotics, by allowing a non-vanishing error probability $\varepsilon \in (0,1)$, we observe that the backoff of the non-asymptotic coding rate $R^*(n,\varepsilon) := \frac{1}{n}\log M^*(n,\varepsilon)$ from Shannon's asymptotic rate $H(P_X)$ is in the order of $\Theta\left(\frac{1}{\sqrt{n}}\right)$ with the coefficient determined by a function of the tolerable error probability and the source dispersion.

2.4 Proof of Second-Order Asymptotics

We next present the proof of Theorem 2.3 by illustrating how one can apply the Berry-Esseen theorem (cf. Theorem 1.3) to the non-asymptotic bounds in Theorems 2.1 and 2.2,

For the smooth presentation of the proof steps, let

$$T(P_X) := \mathsf{E}[|\imath(X|P_X) - H(P_X)|^3]. \tag{2.32}$$

2.4.1 Achievability

Given any $\varepsilon \in (0,1)$, let

$$\varepsilon_n = \varepsilon - \frac{T(P_X)}{\sqrt{n(\mathrm{V}(P_X))^3}}, \tag{2.33}$$

$$\log M = nH(P_X) + \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon_n). \tag{2.34}$$

It follows from Theorem 2.1 that the error probability of the code satisfies

$$\mathrm{P}_{\mathrm{e},n} \le \Pr\Big\{\sum_{i\in[n]}(\imath(X_i|P_X)-H(P_X)) \ge \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon)\Big\} \tag{2.35}$$

$$= \Pr\Big\{\frac{1}{n}\sum_{i\in[n]}(\imath(X_i|P_X)-H(P_X)) \ge \sqrt{\frac{\mathrm{V}(P_X)}{n}}\mathrm{Q}^{-1}(\varepsilon)\Big\} \tag{2.36}$$

$$\le \varepsilon_n + \frac{T(P_X)}{\sqrt{n(\mathrm{V}(P_X))^3}} \tag{2.37}$$

$$= \varepsilon, \tag{2.38}$$

where (2.37) follows from the Berry-Esseen theorem for i.i.d. random variables in Theorem 1.3 since the random variables $\{\imath(X_i|P_X) - H(P_X)\}_{i\in[n]}$ are a sequence of i.i.d. random variables with mean 0 and the identical variance $\mathrm{V}(P_X)$.

Thus, using the Taylor expansion of $\mathrm{Q}^{-1}(\varepsilon_n)$ around ε that states $\mathrm{Q}^{-1}(\varepsilon_n) = \mathrm{Q}^{-1}(\varepsilon) + O(\varepsilon-\varepsilon_n)$, we have

$$\log M^*(n,\varepsilon) \le nH(P_X) + \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon) + O(1). \tag{2.39}$$

2.4.2 Converse

For any $\varepsilon \in (0,1)$, let

$$\varepsilon_n' = \varepsilon + \frac{T(P_X)}{\sqrt{n(\mathrm{V}(P_X))^3}} + \frac{1}{n}, \tag{2.40}$$

$$\log M = nH(P_X) + \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon_n') - \log n. \tag{2.41}$$

Invoking Theorem 2.2 with $\gamma = \log n$ and using the Berry-Esseen theorem, the error probability of any (n, M)-code satisfies

$$\mathrm{P}_{\mathrm{e},n}$$

$$\ge \Pr\Big\{\sum_{i\in[n]}(\imath(X_i|P_X)-H(P_X)) \ge \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon_n')\Big\} - \frac{1}{n} \tag{2.42}$$

$$\ge \varepsilon. \tag{2.43}$$

Therefore,

$$\log M^*(n,\varepsilon) \geq nH(P_X) + \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon_n') - \frac{1}{2}\log n + O(1) \quad (2.44)$$

$$= nH(P_X) + \sqrt{n\mathrm{V}(P_X)}\mathrm{Q}^{-1}(\varepsilon) + O(\log n). \quad (2.45)$$

The converse proof is now completed.

2.5 Other Refined Asymptotics

Besides second-order asymptotics, there are also other refined asymptotics beyond Shannon's source coding theorem. Two examples are the large and moderate deviations analyses.

In large deviations, one characterizes the decay rate of the error probability $\mathrm{P}_{\mathrm{e},n}$ for any asymptotic rate greater than $H(P_X)$.

Definition 2.3. A non-negative number E is said to be a rate-R achievable error exponent if there exists a sequence of (n, M)-codes such that

$$\limsup_{n\to\infty} \frac{1}{n}\log M \leq R, \quad (2.46)$$

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathrm{P}_{\mathrm{e},n} \geq E. \quad (2.47)$$

The supremum of all rate-R achievable error exponents is called the optimal error exponent and denoted by $E^*(R)$.

The exact characterization of $E^*(R)$ was given by Gallager [36] and by Csiszár and Longo [23].

Theorem 2.5. The optimal error exponent for the lossless source coding problem is

$$E^*(R) = \max_{\rho\geq 0}\left(\rho R - (1+\rho)\log\Big(\sum_{x\in\mathrm{supp}(P_X)} P_X^{\frac{1}{1+\rho}}(x)\Big)\right) \quad (2.48)$$

$$= \min_{Q_X: H(Q_X)\geq R} D(Q_X \| P_X). \quad (2.49)$$

As a result of Theorem 2.5, we conclude that the error probability decays exponentially fast for any rate above the first-order coding rate, i.e., $R > H(P_X)$. The characterization in (2.48) was proved by Gallager using the maximum likelihood decoding with the ρ trick and the characterization in (2.49) was proved by Csiszár and Longo [23] using the method of types. The equivalence of the two characterizations is hinted in [22, Problem 2.14].

The moderate deviations regime interpolates between the large deviations and second-order asymptotic regimes. In this regime, one is interested in a sequence of (n, M)-codes whose rates approach $H(P_X)$ and whose error probabilities decay to zero simultaneously.

Definition 2.4. Consider any sequence $\{\xi_n\}_{n\in\mathbb{N}}$ such that $\xi_n \to 0$ and $\sqrt{n\xi_n} \to \infty$ as $n \to \infty$. A non-negative number ν is said to be an achievable moderate deviations constant if there exists a sequence of (n, M)-codes such that

$$\limsup_{n\to\infty} \frac{\log M - nH(P_X)}{n\xi_n} \leq 1, \tag{2.50}$$

$$\liminf_{n\to\infty} -\frac{1}{n\xi_n^2} \log \mathrm{P}_{\mathrm{e},n} \geq \nu. \tag{2.51}$$

The supremum of all moderate deviations constants is called the optimal moderate deviations constant and is denoted by ν^*.

Note that in moderate deviations, the speed of the rate approaching $H(P_X)$ is in the order of ξ_n, which is slower than $O(\frac{1}{\sqrt{n}})$ in second-order asymptotics and the decay rate of the error probability is subexponential, which is slower than the exponential decay in large deviations. This is precisely the reason why moderate deviations is said to interpolate second-order and large deviations asymptotics.

The optimal moderate deviations constant for the lossless source coding problem was obtained by Altŭg, Wagner and Kontoyiannis in [4].

Theorem 2.6. The optimal moderate deviations constant is

$$\nu^* = \frac{1}{2\mathrm{V}(P_X)}. \tag{2.52}$$

Theorem 2.6 states that the sequence of optimal codes approaches $H(P_X)$ at the speed of ξ_n with the error probability decaying subexponentially fast, which can be proved by applying the moderate deviations

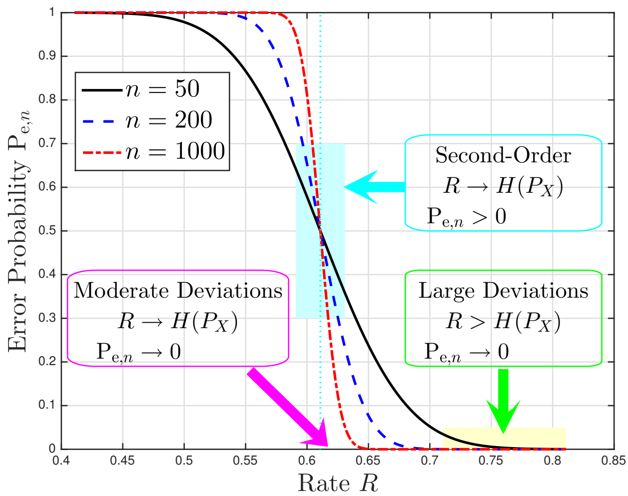

Figure 2.2: Illustration of refined asymptotics for lossless source coding of a binary memoryless source. Note that both large and moderate deviations asymptotics provide tight characterization when n is sufficiently large and violates the low latency requirement of practical communication systems. In contrast, second-order asymptotics provides approximations to the performance of optimal codes with finite blocklength.

theorem [28, p. 3.7.1] to the non-asymptotic bounds in Theorems 2.1 and 2.2.

To illustrate the relationship between second-order, large and moderate deviations to the non-asymptotic bounds, we plot the relationship between the error probability and coding rate for different blocklengths for a binary memoryless source distributed according to a Bernoulli distribution with parameter 0.3 in Figure 2.2, using the second-order asymptotic bound in Theorem 2.3 as the approximation. Note that both large and moderate deviations theorems are tight for sufficiently large blocklength and thus violate the low-latency requirement of practical communication systems. In this monograph, for all lossy source coding problems to be covered, we focus on the second-order asymptotics that provide good approximations to the performance of optimal codes at finite blocklengths (cf. [16, Fig. 1]), and we also present non-asymptotic bounds from which the second-order asymptotics are derived.

Part III

Multiterminal Compression

9

Kaspi Problem

In this section, we study the lossy source coding problem with one encoder and two decoders, where side information is available at the encoder and one of the two decoders. We term the problem as the Kaspi problem since this problem was first introduced by Kaspi, who derived the asymptotically optimal achievable rate to ensure reliable lossy reconstruction at both decoders [56, Theorem 1]. Analogous to the rate-distortion problem, we term the asymptotic optimal achievable rate as the Kaspi rate-distortion function. The Kaspi problem generalizes the rate-distortion problem by adding one additional decoder and allowing the encoder and the additional decoder to access to some correlated side information.

Kaspi's asymptotic results were recently refined by Zhou and Motani in [147], [148], in which the authors derived non-asymptotic and second-order asymptotics bounds for the Kaspi problem. In this section, we present the results in [147], [148] and illustrate the role of side information on lossy data compression in the finite blocklength regime. Specifically, we first present a parametric representation for the Kaspi rate-distortion function. Subsequently, we generalize the notion of the distortion-tilted information density for the rate-distortion problem in

Section 3 to the Kaspi problem and present a non-asymptotic converse bound. Finally, for DMS under bounded distortion measures, we present second-order asymptotics and illustrate the results via two numerical examples.

Since the Kaspi problem generalizes the rate-distortion problem, the results for the Kaspi problem generalize those in Section 3. Furthermore, another special case of the Kaspi problem is the conditional rate-distortion problem where side information is available to both the encoder and decoder in the rate-distortion problem. Thus, the results for the Kaspi problem generalize those for the conditional rate-distortion problem [77] as well.

9.1 Problem Formulation and Asymptotic Result

The setting of the Kaspi problem is shown in Figure 9.1. There are one encoder f and two decoders ϕ_1, ϕ_2. The side information Y^n is available to the encoder f and the decoder ϕ_2 but *not* to the decoder ϕ_1. The encoder f compresses the source X^n into a message S given the side information Y^n. Decoder ϕ_1 aims to recover source sequence X^n within distortion level D_1 under distortion measure d_1 using the message S. Decoder ϕ_2 aims to recover X^n within distortion level D_2 under distortion measure d_2 using the message S and the side information Y^n. Consider a correlated memoryless source with distribution P_{XY} defined on the alphabet $\mathcal{X} \times \mathcal{Y}$. Assume that the source sequence and side information (X^n, Y^n) is generated i.i.d. from P_{XY}. Furthermore, assume that the reproduction alphabets for decoders ϕ_1 and ϕ_2 are $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$ respectively.

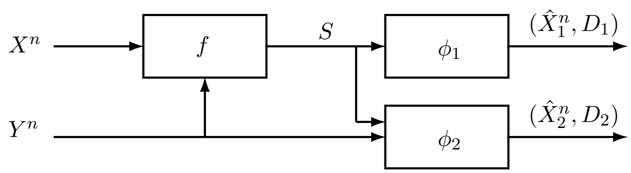

Figure 9.1: System model for the Kaspi problem of lossy source with side information at the encoder and one of the two decoders [56, Theorem 1].

Definition 9.1. An (n, M)-code for the Kaspi problem consists of one encoder

$$f : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{M} = [M], \tag{9.1}$$

and two decoders

$$\phi_1 : \mathcal{M} \to \hat{\mathcal{X}}_1^n, \tag{9.2}$$

$$\phi_2 : \mathcal{M} \times \mathcal{Y}^n \to \hat{\mathcal{X}}_2^n. \tag{9.3}$$

For simplicity, let $\hat{X}_1^n = \phi_1(f(X^n, Y^n))$ and $\hat{X}_2^n = \phi_2(f(X^n, Y^n), Y^n)$. For $i \in [2]$, let $d_i : \mathcal{X} \times \hat{\mathcal{X}}_i \to [0, \infty]$ be two distortion measures. For any $x^n \in \mathcal{X}^n$ and $\hat{x}_i^n \in \hat{\mathcal{X}}_i^n$, let the distortion between x^n and $\hat{x}_i^n$ be additive and defined as $d_i(x^n, \hat{x}_i^n) := \frac{1}{n}\sum_{j\in[n]} d_i(x_j, \hat{x}_{i,j})$.

Following [56], the rate-distortion function of the Kaspi problem is defined as follows, which characterizes the asymptotically minimal rate to ensure reliable lossy compression at both decoders as the blocklength tends to infinity.

Definition 9.2. A rate R is said to be (D_1, D_2)-achievable for the Kaspi problem if there exists a sequence of (n, M)-codes such that

$$\limsup_{n\to\infty} \frac{\log M}{n} \le R, \tag{9.4}$$

and

$$\limsup_{n\to\infty} \mathbb{E}[d_i(X^n, \hat{X}_i^n)] \le D_i, \;\; i \in [2]. \tag{9.5}$$

The minimum (D_1, D_2)-achievable rate is called the Kaspi rate-distortion function and denoted as $R^*(D_1, D_2)$.

Define

$$\begin{aligned} &R(P_{XY}, D_1, D_2) \\ &:= \min_{\substack{P_{\hat{X}_1|XY},\, P_{\hat{X}_2|XY\hat{X}_1}: \\ \mathbb{E}[d_1(X,\hat{X}_1)]\le D_1, \\ \mathbb{E}[d_2(X^n,\hat{X}_2^n)]\le D_2}} I(P_{XY}, P_{XY|\hat{X}_1}) + I(P_{X|Y\hat{X}_1}, P_{X|\hat{X}_1\hat{X}_2Y}|P_{Y\hat{X}_1}), \end{aligned} \tag{9.6}$$

where the distributions $(P_{XY|\hat{X}_1}, P_{X|Y\hat{X}_1}, P_{X|\hat{X}_1\hat{X}_2Y})$ are induced by $(P_{XY}, P_{\hat{X}_1|XY},\; P_{\hat{X}_2|XY\hat{X}_1})$.

Kaspi [56, Theorem 1] derived the following result.

Theorem 9.1. The minimum (D_1, D_2)-achievable rate for the Kaspi problem satisfies

$$R^*(D_1, D_2) = R(P_{XY}, D_1, D_2). \tag{9.7}$$

We refer to $R(P_{XY}, D_1, D_2)$ as the Kaspi rate-distortion function. Note that $R(P_{XY}, D_1, D_2)$ is convex and non-increasing in both D_1 and D_2. We remark that the explicit formulas of the Kaspi rate-distortion function was derived by Perron, Diggavi and Telatar for GMS under quadratic distortion measures [89] and a binary memoryless erasure source under Hamming distortion measures [90].

To derive non-asymptotic and second-order asymptotic bounds, instead of using the average distortion criterion, we adopt the following joint excess-distortion probability as the performance criterion:

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) := \Pr\left\{d_1(X^n, \hat{X}_1^n) > D_1 \text{ or } d_2(X^n, \hat{X}_2^n) > D_2\right\}. \tag{9.8}$$

Note that the probability in (9.8) is calculated with respect to the distribution of the source sequences for a fixed (n, M)-code. For bounded distortion measures, the asymptotically minimal rate to ensure vanishing joint excess-distortion probability $\mathrm{P}_{\mathrm{e},n}(D_1, D_2)$ is also $R(P_{XY}, D_1, D_2)$. The justification is similar to the case of the rate-distortion problem below Theorem 3.1.

The Kaspi rate-distortion function $R(P_{XY}, D_1, D_2)$ equals the rate-distortion function $R(P_X, D_2)$ if D_1 is large enough, and equals the conditional rate-distortion function $R(P_{X|Y}, D_1|P_Y)$ if D_2 is large enough where

$$R(P_{X|Y}, D_1|P_Y) := \min_{P_{\hat{X}_1|XY}:\mathsf{E}[d_1(X,\hat{X}_1)]\le D_1} I(P_{X|Y}, P_{X|\hat{X}_1Y}|P_Y), \tag{9.9}$$

with $P_{X|\hat{X}_1Y}$ induced by $(P_Y, P_{X|Y}, P_{\hat{X}_1|XY})$.

Note that the conditional rate-distortion function $R(P_{X|Y}, D_1|P_Y)$ is the minimal achievable rate of lossy compression when side information is available at both the encoder and the decoder, which is also known as the conditional rate-distortion problem. Similarly, for second-order asymptotics, the results for the Kaspi problem specialize to either the rate-distortion problem or the conditional rate-distortion problem.

9.2 Properties of the Rate-Distortion Function

We first present the properties of the Kaspi rate-distortion function, which allows us to define the distortions-tilted information density for the Kaspi problem and derive a non-asymptotic converse bound that generalizes the non-asymptotic converse bound for the rate-distortion problem in Theorem 3.4 of the rate-distortion problem.

Given any (conditional) distributions $(P_{\hat{X}_1|XY}, P_{\hat{X}_2|XY\hat{X}_1})$, let $P_{\hat{X}_1}$, $P_{X\hat{X}_1}$, $P_{X\hat{X}_2}$ and $P_{Y\hat{X}_1}$, $P_{\hat{X}_2|Y\hat{X}_1}$, $P_{Y\hat{X}_1\hat{X}_2}$ be induced by P_{XY}, $P_{\hat{X}_1|XY}$ and $P_{\hat{X}_2|XY\hat{X}_1}$. Consider the distortion levels (D_1, D_2) such that $R(P_{XY}, D_1, D_2)$ is finite and there exists test channels $(P^*_{\hat{X}_1|XY}, P^*_{\hat{X}_2|XY\hat{X}_1})$ that achieve $R(P_{XY}, D_1, D_2)$. Note that $R(P_{XY}, D_1, D_2)$ (see (9.6)) corresponds to a convex optimization problem and the dual problem is given by

$$\sup_{(\lambda_1,\lambda_2)\in\mathbb{R}_+^2} \min_{P_{\hat{X}_1|XY}, P_{\hat{X}_2|XY\hat{X}_1}}$$

$$\Big(I(P_{XY}, P_{XY|\hat{X}_1}) + I(P_{X|Y\hat{X}_1}, P_{X|\hat{X}_1\hat{X}_2Y}|P_{Y\hat{X}_1}) \tag{9.10}$$

$$+ \lambda_1(\mathbb{E}[d_1(X,\hat{X}_1) - D_1]) + \lambda_2(\mathbb{E}[d_2(X^n,\hat{X}_2^n) - D_2])\Big). \tag{9.11}$$

For any given distortion levels (D_1, D_2), the optimal solutions to the dual problem of $R(P_{XY}, D_1, D_2)$ are

$$\lambda_1^* := \frac{\partial R(P_{XY}, D, D_2)}{\partial D}\Big|_{D=D_1}, \tag{9.12}$$

$$\lambda_2^* := \frac{\partial R(P_{XY}, D_1, D)}{\partial D}\Big|_{D=D_2}. \tag{9.13}$$

Given any $(x, y, \hat{x}_1) \in \mathcal{X} \times \mathcal{Y} \times \hat{\mathcal{X}}_1$ and distributions $(Q_{\hat{X}_1}, Q_{\hat{X}_2|Y\hat{X}_1}) \in \mathcal{P}(\hat{\mathcal{X}}_1) \times \mathcal{P}(\hat{\mathcal{X}}_2|\mathcal{Y}, \hat{\mathcal{X}}_1)$, let

$$\alpha_2(x, y, \hat{x}_1|Q_{\hat{X}_2|Y\hat{X}_1})$$

$$:= \Big\{\mathbb{E}_{Q_{\hat{X}_2|Y\hat{X}_1}}\Big[\exp(-\lambda_2^* d_2(X^n, \hat{X}_2^n))\Big|Y = y, \hat{X}_1 = \hat{x}_1\Big]\Big\}^{-1}, \tag{9.14}$$

$$\alpha(x, y|Q_{\hat{X}_1}, Q_{\hat{X}_2|Y\hat{X}_1})$$

$$:= \left\{\mathbb{E}_{Q_{\hat{X}_1}}\left[\frac{\exp\big(-\lambda_1^* d_1(x,\hat{X}_1)\big)}{\alpha_2(x, y, \hat{X}_1|Q_{\hat{X}_2|Y\hat{X}_1})}\right]\right\}^{-1}. \tag{9.15}$$

Lemma 9.2. A pair of conditional distributions $(P^*_{\hat{X}_1|XY}, P^*_{\hat{X}_2|XY\hat{X}_1})$ achieves $R(P_{XY}, D_1, D_2)$ if and only if

- For all $(x, y, \hat{x}_1)$,

$$P^*_{\hat{X}_1|XY}(\hat{x}_1|x,y) = \frac{\alpha(x,y|P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1})P^*_{\hat{X}_1}(\hat{x}_1)\exp(-\lambda_1^* d_1(x,\hat{x}_1))}{\alpha_2(x,y,\hat{x}_1|P^*_{\hat{X}_2|Y\hat{X}_1})}, \tag{9.16}$$

- For all $(x, y, \hat{x}_1, \hat{x}_2)$ such that $P^*_{\hat{X}_1}(\hat{x}_1) > 0$,

$$P^*_{\hat{X}_2|XY\hat{X}_1}(\hat{x}_2|x,y,\hat{x}_1) = \alpha_2(x,y,\hat{x}_1|P^*_{\hat{X}_2|Y\hat{X}_1})P^*_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|y,\hat{x}_1) \times \exp(-\lambda_2^* d_2(X^n, \hat{X}_2^n)). \tag{9.17}$$

Furthermore, if the pair of distributions $(P^*_{\hat{X}_1|XY}, P^*_{\hat{X}_2|XY\hat{X}_1})$ achieves $R(P_{XY}, D_1, D_2)$,

$$R(P_{XY}, D_1, D_2) = \mathbb{E}[\log \alpha(x,y|P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1})] - \lambda_1^* D_1 - \lambda_2^* D_2. \tag{9.18}$$

The proof of Lemma 9.2 is similar to [69, Properties 1-3] for the rate-distortion problem that mainly uses the KKT conditions for convex optimization problems. Lemma 9.2 paves the way for the definition of the distortions-tilted information density for the Kaspi problem and also implies critical properties for the Kaspi distortions-tilted information density that parallel Lemma 3.2 for the rate-distortion problem.

We remark that for any pair of optimal test channels $(P^*_{\hat{X}_1|XY}, P^*_{\hat{X}_2|XY\hat{X}_1})$, similar to [127, Lemma 2], one can verify that the values of $\alpha_2(x,y,\hat{x}_1|P^*_{\hat{X}_2|Y\hat{X}_1})$ and $\alpha(x,y|P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1})$ remain the same. Hence, for simplicity, we define

$$\alpha_2(x,y,\hat{x}_1) := \alpha_2(x,y,\hat{x}_1|P^*_{\hat{X}_2|Y\hat{X}_1}), \tag{9.19}$$

$$\alpha(x,y) := \alpha(x,y|P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1}). \tag{9.20}$$

Furthermore, for any $\hat{x}_1 \in \hat{\mathcal{X}}_1$ and distribution $Q_{\hat{X}_2|Y\hat{X}_1} \in \mathcal{P}(\hat{\mathcal{X}}_2|\mathcal{Y}, \hat{\mathcal{X}}_1)$, define the following function:

$$\nu(\hat{x}_1, Q_{\hat{X}_2|Y\hat{X}_1}) := \mathbb{E}_{P_{XY}\times Q_{\hat{X}_2|Y\hat{X}_1}}\Big[\alpha(X,Y) \times \exp(-\lambda_1^* d_1(X,\hat{X}_1) - \lambda_2^* d_2(X,\hat{X}_2))\big|\hat{X}_1 = \hat{x}_1\Big]. \quad (9.21)$$

The following lemma holds.

Lemma 9.3. For any $\hat{x}_1$ and arbitrary distribution $Q_{\hat{X}_2|Y\hat{X}_1}$, we have

$$\nu(\hat{x}_1, Q_{\hat{X}_2|Y\hat{X}_1}) \leq 1. \quad (9.22)$$

We remark that Lemma 9.3 holds for both discrete and continuous memoryless sources. The proof of Lemma 9.3 is inspired by [20, Lemma 1.4], [120, Lemma 5] and [68] and available in [147, Appendix A]. Invoking Lemma 9.3, we prove a non-asymptotic converse bound for the Kaspi problem in Theorem 9.5.

9.3 Distortions-Tilted Information Density

Now we introduce the distortions-tilted information density for the Kaspi problem that generalizes distortion-tilted information density for the lossy source coding problem [54], [70]. Recall the definition of $\alpha(\cdot)$ in (9.20).

Definition 9.3. For any $(x,y) \in \mathcal{X}\times\mathcal{Y}$, the (D_1,D_2)-tilted information density for the Kaspi problem is defined as

$$\jmath(x,y|D_1,D_2,P_{XY}) := \log\alpha(x,y) - \lambda_1^* D_1 - \lambda_2^* D_2. \quad (9.23)$$

The properties of the (D_1,D_2)-tilted information density follows from Lemma 9.2. For example, invoking (9.16) and (9.17), we conclude that for all $(x,y,\hat{x}_1,\hat{x}_2)$ such that $P^*_{\hat{X}_1}(\hat{x}_1)P^*_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|y,\hat{x}_1) > 0$,

$$\begin{aligned}\jmath(x,y|D_1,D_2,P_{XY}) &= \log\frac{P^*_{\hat{X}_1|XY}(\hat{x}_1|x,y)}{P^*_{\hat{X}_1}(\hat{x}_1)} + \log\frac{P^*_{\hat{X}_2|XY\hat{X}_1}(\hat{x}_2|x,y,\hat{x}_1)}{P^*_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|y,\hat{x}_1)} \\ &\quad + \lambda_1^*(d_1(x,\hat{x}_1) - D_1) + \lambda_2^*(d_2(X^n,\hat{X}_2^n) - D_2). \quad (9.24)\end{aligned}$$

Furthermore, it follows from (9.18) that

$$R(P_{XY}, D_1, D_2) = \mathsf{E}[\jmath(X, Y|D_1, D_2, P_{XY})]. \tag{9.25}$$

Finally, we have the following lemma that further relates the distortions-tilted information density with the derivative of the Kaspi rate-distortion function with respect to the distribution P_{XY}. Given a joint probability mass function P_{XY}, recall that $m = |\mathrm{supp}(P_{XY})|$ and $\Gamma(P_{XY})$ is the sorted distribution such that for each $i \in [m]$, $\Gamma_i(P_{XY}) = P_{XY}(x_i, y_i)$ is the i-th largest value of $\{P_{XY}(x, y) : (x, y) \in \mathcal{X} \times \mathcal{Y}\}$.

Lemma 9.4. Suppose that for all Q_{XY} in the neighborhood of P_{XY}, $\mathrm{supp}(Q^*_{\hat{X}_1\hat{X}_2}) = \mathrm{supp}(P^*_{\hat{X}_1\hat{X}_2})$. Then, for each $i \in [m-1]$,

$$\begin{aligned}&\frac{\partial R(Q_{XY}, D_1, D_2)}{\partial \Gamma_i(Q_{XY})}\Big|_{Q_{XY}=P_{XY}}\\&= \jmath(x_i, y_i|D_1, D_2, P_{XY}) - \jmath(x_m, y_m|D_1, D_2, P_{XY}).\end{aligned} \tag{9.26}$$

The proof of Lemma 9.4 is available in [147, Appendix I]. Lemma 9.4 parallels Claim (iv) in Lemma 3.2 for the rate-distortion problem and is critical in the achievability proof of second-order asymptotics.

9.4 A Non-Asymptotic Converse Bound

Invoking Lemma 9.3, we obtain the following non-asymptotic converse bound for the Kaspi problem that generalizes Theorem 3.4 for the rate-distortion problem.

Theorem 9.5. Given any $\gamma > 0$, the joint excess-distortion probability of any (n, M)-code for the Kaspi problem satisfies

$$\begin{aligned}\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \geq \Pr\Big\{\sum_{i\in[n]} \jmath(X_i, Y_i|D_1, D_2, P_{XY}) \geq \log M + n\gamma\Big\}\\ - \exp(-n\gamma).\end{aligned} \tag{9.27}$$

We remark that Theorem 9.5 plays a central role in the converse proof the second-order asymptotics and holds for any memoryless sources.

Proof. The proof of Theorem 9.5 is similar to that of Theorem 3.4. Given any (n,M)-code with encoder f and decoders (ϕ_1,ϕ_2), let $S=f(X^n)$ be the compressed index that takes values in $\mathcal{M}$, let $P_{S|X^n}$ be the conditional distribution induced by the encoder f and let $P_{\hat{X}_1^n|S}$ and let the conditional distributions $P_{\hat{X}_1^n|S}$ and $P_{\hat{X}_2^n|S,Y^n}$ be induced by the decoders ϕ_1 and ϕ_2, respectively. Furthermore, let Q_S be the uniform distribution over $\mathcal{M}$ and let

$$Q_{\hat{X}_1^n}(\hat{x}_1^n) := \sum_{s\in\mathcal{M}} Q_S(s)P_{\hat{X}_1^n|S}(\hat{x}_1^n|s), \tag{9.28}$$

$$Q_{\hat{X}_2^n|Y^n}(\hat{x}_2^n|y^n) := \frac{\sum_s Q_S(s)P_{\hat{X}_1^n|S}(\hat{x}_1^n|s)P_{\hat{X}_2^n|SY^n}(\hat{x}_2^n|s,y^n)}{Q_{\hat{X}_1^n}(\hat{x}_1^n)}. \tag{9.29}$$

For ease of notation, we use $\mathcal{C}(D_1,D_2)$ to denote the non-excess-distortion event, i.e., the event that $\{d_1(X^n,\hat{X}_1^n)\le D_1,\ d_2(X^n,\hat{X}_2^n)\le D_2\}$ and use $\mathcal{E}(D_1,D_2)$ to denote the excess-distortion event $\{d_1(X^n,\hat{X}_1^n) > D_1 \text{ or } d_2(X^n,\hat{X}_2^n) > D_2\}$. For any $\gamma>0$, it follows that

$$\begin{aligned}
&\Pr\Big\{\sum_{i\in[n]} \jmath(X_i,Y_i|D_1,D_2,P_{XY}) \ge \log M + n\gamma\Big\}\\
&\le \Pr\Big\{\sum_{i\in[n]} \jmath(X_i,Y_i|D_1,D_2,P_{XY}) \ge \log M + n\gamma \text{ and } \mathcal{C}(D_1,D_2)\Big\}\\
&\quad + \Pr\{\mathcal{E}(D_1,D_2)\},
\end{aligned} \tag{9.30}$$

where the second term in (9.30) is exactly the joint excess-distortion probability $\mathrm{P}_{\mathrm{e},n}(D_1,D_2)$.

The first term in (9.30) can be upper bounded as follows:

$$\begin{aligned}
&\Pr\Big\{\sum_{i\in[n]} \jmath(X_i,Y_i|D_1,D_2,P_{XY}) \ge \log M + n\gamma,\ \mathcal{C}(D_1,D_2)\Big\}\\
&= \Pr\Big\{M \le \exp\Big(\sum_{i\in[n]} \jmath(X_i,Y_i|D_1,D_2,P_{XY}) - n\gamma\Big)\mathbb{1}(\mathcal{C}(D_1,D_2))\Big\}
\end{aligned} \tag{9.31}$$

$$\le \frac{\exp(-\gamma)}{M}\mathbb{E}\Big[\exp\Big(\sum_{i\in[n]} \jmath(X_i,Y_i|D_1,D_2,P_{XY})\Big)\mathbb{1}(\mathcal{C}(D_1,D_2))\Big] \tag{9.32}$$

$$\leq \frac{\exp(-n\gamma)}{M}\mathbb{E}\Big[\exp\Big(\sum_{i\in[n]}\jmath(X_i,Y_i|D_1,D_2,P_{XY}) + \sum_{i\in[2]}\lambda_i^*(D_i-d_i(X^n,\hat{X}_i^n))\Big)\Big], \tag{9.33}$$

$$= \exp(-n\gamma)\sum_{s}\sum_{(x^n,y^n)}\sum_{\hat{x}_1^n,\hat{x}_2^n} Q_S(s)P_{XY}^n(x^n,y^n)P_{S|X^n}(s|x^n)P_{\hat{X}_1^n|S}(\hat{x}_1^n|s) \times P_{\hat{X}_2^n|Y^n,S}(\hat{x}_2^n|y^n,s)\prod_{i\in[n]}\alpha(x_i,y_i) \times \exp(-\lambda_1^* d(x_i,\hat{x}_{1,i})-\lambda_2^*(d(x_i,\hat{x}_{2,i}))) \tag{9.34}$$

$$\leq \exp(-n\gamma)\sum_{(x^n,y^n)}\sum_{\hat{x}_1^n,\hat{x}_2^n} Q_{\hat{X}_1^n}(\hat{x}_1^n)Q_{\hat{X}_2^n|\hat{X}_1^n}(\hat{x}_1^n|\hat{x}_1^n)P_{XY}^n(x^n,y^n) \times \prod_{i\in[n]}\alpha(x_i,y_i)\exp(-\lambda_1^* d(x_i,\hat{x}_{1,i})-\lambda_2^*(d(x_i,\hat{x}_{2,i}))) \tag{9.35}$$

$$= \exp(-n\gamma)\sum_{\hat{x}_1^n}Q_{\hat{X}_1^n}(\hat{x}_1^n)\Big[\prod_{i\in[n]}\nu(\hat{x}_{1,i},Q_{\hat{X}_{2,i}|Y_i\hat{X}_{1,i}})\Big] \tag{9.36}$$

$$\leq \exp(-n\gamma), \tag{9.37}$$

where (9.32) follows from the Markov inequality and (9.33) follows since $\lambda_i^* \geq 0$ for $i \in [2]$, (9.34) follows from the definitions of $\alpha(\cdot)$ in (9.20) and $\jmath(\cdot)$ in (9.23), (9.35) follows from the fact $P_{S|X^n}(s|x^n) \leq 1$ and the definitions of distributions $(Q_{\hat{X}_1^n}, Q_{\hat{X}_2^n|\hat{X}_1^n}, Q_{\hat{X}_2^n|Y^n})$, (9.36) since we define $Q_{\hat{X}_{2,i}|Y_i\hat{X}_{1,i}}$ as the marginal distribution $Q_{\hat{X}_{2,i}|Y_i}$ of $Q_{\hat{X}_2^n|Y^n}$ and use the definition of $\nu(\cdot)$ in (9.21) and (9.37) follows from the result in (9.22).

The proof of Theorem 9.5 is completed by combining (9.30) and (9.37). □

9.5 Second-Order Asymptotics

In this section, we define and present second-order asymptotics of the Kaspi problem for DMS under bounded distortion measures. In other words, we assume that $\mathcal{X}$, $\mathcal{Y}$, $\hat{\mathcal{X}}_1$, $\hat{\mathcal{X}}_2$ are all finite sets and $\max_{x,\hat{x}_i} d_i(x,\hat{x}_i)$, $i \in [2]$ is finite.

9.5.1 Definition, Main Result and Discussions

Let $\varepsilon \in (0,1)$ be fixed.

Definition 9.4. A rate L is said to be second-order (D_1, D_2, ε)-achievable for the Kaspi problem if there exists a sequence of (n, M)-codes such that

$$\limsup_{n\to\infty} \frac{\log M - nR(P_{XY}, D_1, D_2)}{\sqrt{n}} \leq L, \tag{9.38}$$

and

$$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \varepsilon. \tag{9.39}$$

The infimum second-order (D_1, D_2, ε)-achievable rate is called the optimal second-order coding rate and denoted as $L^*(D_1, D_2, \varepsilon)$.

Note that in Definition 9.2 of the rate-distortion region, the average distortion criterion is used, while in Definition 9.4, the excess-distortion probability is considered. The reason is that for second-order asymptotics, second-order asymptotics always companies with the probability of a certain event. To be specific, the excess-distortion probability plays a similar role as error probability for the lossless source coding problem [49] or channel coding problems [50], [91]. Let $\mathrm{V}(D_1, D_2, P_{XY})$ be the distortions-dispersion function for the Kaspi problem, i.e.,

$$\mathrm{V}(D_1, D_2, P_{XY}) := \mathrm{Var}\big[\jmath(X, Y|D_1, D_2, P_{XY})\big]. \tag{9.40}$$

We impose following conditions:

1. The distortion levels are chosen such that $R(P_{XY}, D_1, D_2) > 0$ is finite;

2. $Q_{XY} \to R(Q_{XY}, D_1, D_2)$ is twice differentiable in the neighborhood of P_{XY} and the derivatives are bounded.

Theorem 9.6. Under conditions (1) and (2), the optimal second-order coding rate for the Kaspi problem is

$$L^*(D_1, D_2, \varepsilon) = \sqrt{\mathrm{V}(D_1, D_2, P_{XY})}\,\mathrm{Q}^{-1}(\varepsilon). \tag{9.41}$$

The converse proof of Theorem 9.6 follows by applying the Berry-Esseen Theorem to the non-asymptotic bound in Theorem 9.5. In the achievability proof, we first prove a type-covering lemma tailored for the Kaspi problem. Subsequently, we make use of the properties of $\jmath(x, y|D_1, D_2, P_{XY})$ in Lemma 9.2 and appropriate Taylor expansions.

We remark that the distortions-tilted information density for the Kaspi problem $\jmath(x, y|D_1, D_2, P_{XY})$ reduces to the distortion-tilted information density for the lossy source coding problem [70], or the distortion-tilted information density for the lossy source coding problem with encoder and decoder side information [77] for particular choices of distortion levels (D_1, D_2). Hence, our result in Theorem 9.6 is a strict generalization of the second-order coding rate for the lossy source coding problem [70] and the conditional lossy source coding problem [77] for DMS under bounded distortion measures. We also illustrate this point in Section 9.5.2 via a numerical example for the doubly symmetric binary source.

In the next two subsections, we illustrate Theorem 9.6 via two numerical examples by calculating the second-order coding rate $L^*(D_1, D_2, \varepsilon)$ in close form.

9.5.2 Numerical Examples

Asymmetric Correlated Source

In order to illustrate our results in Lemma 9.2 and Theorem 9.6, we consider the following source. Let $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, \mathrm{e}\}$ and $P_X(0) = P_X(1) = \frac{1}{2}$. Let Y be the output of passing X through a Binary Erasure Channel (BEC) with erasure probability p, i.e., $P_{Y|X}(y|x) = 1 - p$ if $x = y$ and $P_{Y|X}(\mathrm{e}|x) = p$. The explicit formula of the Kaspi rate-distortion function for the above correlated source under Hamming distortion measures was derived by Perron, Diggavi and Telatar in [90]. Here we only recall the *non-degenerate* result, i.e., the case where the distortion levels (D_1, D_2) are chosen such that $\lambda_1^* > 0$ and $\lambda_2^* > 0$.

Define the set

$$\mathcal{D}_{\mathrm{bec}} := \Big\{(D_1, D_2) \in \mathbb{R}_+^2 : D_1 \leq \frac{1}{2}, \\ D_1 - \frac{1-p}{2} \leq D_2 \leq pD_1\Big\}. \quad (9.42)$$

Lemma 9.7. If $(D_1, D_2) \in \mathcal{D}_{\mathrm{bec}}$, then the Kaspi rate-distortion function for the above asymmetric correlated source under Hamming distortion measures is

$$R(P_{XY}, D_1, D_2) = \log 2 - (1-p)H_b\left(\frac{D_1 - D_2}{1-p}\right) \\ - pH_b\left(\frac{D_2}{p}\right). \quad (9.43)$$

Hence, for $(D_1, D_2) \in \mathcal{D}_{\mathrm{bec}}$, using the definitions of λ_1^* in (9.12) and λ_2^* in (9.13), we obtain

$$\lambda_1^* = \log \frac{(1-p) - (D_1 - D_2)}{1-p} - \log \frac{D_1 - D_2}{1-p} \quad (9.44)$$

$$= \log \frac{(1-p) - (D_1 - D_2)}{D_1 - D_2}, \quad (9.45)$$

$$\lambda_2^* = \log \frac{p - D_2}{p} + \log \frac{D_1 - D_2}{1-p} \\ - \log \frac{(1-p) - (D_1 - D_2)}{1-p} - \log \frac{D_2}{p} \quad (9.46)$$

$$= -\lambda_1^* + \log \frac{p - D_2}{D_2}. \quad (9.47)$$

Then, using the definitions of $\alpha_2(\cdot)$ in (9.19) and $\alpha(\cdot)$ in (9.20), we have

$$\alpha_2(0,0,0) = \alpha_2(0,0,1) = \alpha_2(1,1,0) = \alpha_2(1,1,1) \\ = \alpha_2(0,\mathrm{e},0) = \alpha_2(1,\mathrm{e},1) = 1, \quad (9.48)$$

$$\alpha_2(1,0,0) = \alpha_2(1,0,1) = \alpha_2(0,1,0) = \alpha_2(0,1,1) \\ = \alpha_2(1,\mathrm{e},0) = \alpha_2(0,\mathrm{e},1) = \exp(\lambda_2^*), \quad (9.49)$$

and

$$\alpha(0,0) = \alpha(1,1) = \frac{2}{1 + \exp(-\lambda_1^*)}, \quad (9.50)$$

$$\alpha(0,\mathrm{e}) = \alpha(1,\mathrm{e}) = \frac{2}{1 + \exp(-\lambda_1^* - \lambda_2^*)}. \quad (9.51)$$

It can be verified easily that (9.16), (9.17), (9.18) hold. In the following, we will verify that (9.22) holds for arbitrary $Q_{\hat{X}_2|Y\hat{X}_1}$ and $\hat{x}_1$. As a first step, we can verify that for any $(y, \hat{x}_1, \hat{x}_2)$, we have

$$\begin{aligned}&\sum_x P_{XY}(x,y)\alpha(x,y)\exp(-\lambda_1^* d_1(x,\hat{x}_1) - \lambda_2^* d_2(X^n,\hat{X}_2^n))\\ &\le \sum_x P_{XY}(x,y)\frac{\alpha(x,y)}{\alpha_2(x,y,\hat{x}_1)}\exp(-\lambda_1^* d_1(x,\hat{x}_1)). \quad (9.52)\end{aligned}$$

Then, for any distribution $Q_{\hat{X}_2|Y\hat{X}_1}$, using the definition of $\nu(\cdot)$ in (9.21), multiplying $Q_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|y,\hat{x}_1)$ over both sides of (9.52), and summing over $(y, \hat{x}_2)$, we obtain that

$$\nu(\hat{x}_1, Q_{\hat{X}_2|Y\hat{X}_1}) \le 1. \quad (9.53)$$

Using the definition of $\jmath(\cdot)$ in (9.23), we have

$$\begin{aligned}\jmath(0,0|D_1,D_2,P_{XY}) &= \jmath(1,1|D_1,D_2,P_{XY}) && (9.54)\\ &= \log\alpha(0,0) - \lambda_1^* D_1 - \lambda_2^* D_2, && (9.55)\end{aligned}$$

and

$$\begin{aligned}\jmath(0,\mathrm{e}|D_1,D_2,P_{XY}) &= \jmath(1,\mathrm{e}|D_1,D_2,P_{XY}) && (9.56)\\ &= \log\alpha(0,\mathrm{e}) - \lambda_1^* D_1 - \lambda_2^* D_2. && (9.57)\end{aligned}$$

Furthermore, using the definition of the distortion-dispersion function $\mathrm{V}(D_1, D_2, P_{XY})$ in (9.40), we have

$$\begin{aligned}&\mathrm{V}(D_1,D_2,P_{XY})\\ &= \mathrm{Var}[\jmath(X,Y|D_1,D_2,P_{XY})] && (9.58)\\ &= p(1-p)\left(\log\frac{p-D_2}{p} - \log\frac{(1-p)-(D_1-D_2)}{1-p}\right)^2. && (9.59)\end{aligned}$$

Thus,

$$L^*(D_1,D_2,\varepsilon) = \sqrt{\mathrm{V}(D_1,D_2,P_{XY})}\,\mathrm{Q}^{-1}(\varepsilon). \quad (9.60)$$

Doubly Symmetric Binary Source (DSBS)

In this example, we show that under certain distortion levels, the Kaspi rate-distortion function reduces to the rate-distortion function [105] (see also [38, Theorem 3.5]) and the conditional rate-distortion function [38, Eq. (11.2)]. We consider the DSBS where $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $P_{XY}(0,0) = P_{XY}(1,1) = \frac{1-p}{2}$ and $P_{XY}(0,1) = P_{XY}(1,0) = \frac{p}{2}$ for some $p \in [0, \frac{1}{2}]$.

Lemma 9.8. Depending on the distortion levels (D_1, D_2), the Kaspi rate-distortion function for the DSBS with Hamming distortion measures satisfies

- $D_1 \geq \frac{1}{2}$ and $D_2 \geq p$
$$R(P_{XY}, D_1, D_2) = 0. \tag{9.61}$$

- $D_1 < \frac{1}{2}$ and $D_2 \geq \min\{p, D_1\}$
$$R(P_{XY}, D_1, D_2) = \log 2 - H_b(D_1), \tag{9.62}$$
where $H_b(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function.

- $D_1 \geq D_2 + \frac{1-2p}{2}$ and $D_2 < p$
$$R(P_{XY}, D_1, D_2) = H_b(p) - H_b(D_2). \tag{9.63}$$

When $D_1 < \frac{1}{2}$ and $D_2 < pD_1$, the Kaspi rate-distortion function reduces to the rate-distortion function for the lossy source coding problem. Thus, the distortion-tilted information density for the Kaspi problem reduces to the D_1-tilted information density in (9.64), i.e.,

$$\jmath(x, y|D_1, D_2, P_{XY}) = \log 2 - H_b(D_1). \tag{9.64}$$

Hence, $L^*(D_1, D_1|P_{XY}) = 0$. When $D_1 \geq D_2 + \frac{1-2p}{2}$ and $D_2 < p$, the Kaspi rate-distortion function reduces to the conditional rate-distortion function. Under the optimal test channel, we have $\hat{X}_1 = 0/1$ and $X \to \hat{X}_2 \to Y$ forms a Markov chain. In this case, the distortion-tilted information density for the Kaspi problem reduces to the conditional distortion-tilted information density [69, Definition 5] (see also [77]), i.e.,

$$\jmath(x, y|D_1, D_2, P_{XY}) = -\log P_{X|Y}(x|y) - H_b(D_2). \tag{9.65}$$

Hence,

$$\begin{aligned}\mathrm{V}(D_1, D_2, P_{XY}) &= \mathrm{Var}[-\log P_{X|Y}(X|Y)] & (9.66)\\ &= (1-p)(-\log(1-p) - H_b(p))^2 \\ &\quad + p(-\log p - H_b(p))^2 & (9.67)\\ &:= \mathrm{V}(p), & (9.68)\end{aligned}$$

and

$$L^*(D_1, D_2, \varepsilon) = \sqrt{\mathrm{V}(p)}\mathrm{Q}^{-1}(\varepsilon). \tag{9.69}$$

9.6 Proof of Second-Order Asymptotics

9.6.1 Achievability

We first prove a type covering lemma for the Kaspi problem, based on which we derive an upper bound on the excess-distortion probability. Subsequently, using the Berry-Esseen theorem together with proper Taylor expansions, we manage to prove the desired achievable second-order coding rate.

To present our type covering lemma, define the following constant:

$$c = \Big(8|\mathcal{X}| \cdot |\mathcal{Y}| \cdot |\hat{\mathcal{X}}_1| \cdot |\hat{\mathcal{X}}_2| + 6\Big). \tag{9.70}$$

Lemma 9.9. There exists a set $\mathcal{B} \subset \hat{\mathcal{X}}_1^n$ such that for each $(x^n, y^n) \in \mathcal{T}_{Q_{XY}}$, if

$$(z^n)^* = \arg\min_{\hat{x}_1^n \in \mathcal{B}} d_1(x^n, \hat{x}_1^n), \tag{9.71}$$

then the following conclusion hold.

1. The distortion between x^n and $(z^n)^*$ is upper bounded by D_1, i.e.,

$$d_1(x^n, (z^n)^*) \le D_1, \tag{9.72}$$

2. there exists a set $\mathcal{B}((z^n)^*, y^n) \subset \hat{\mathcal{X}}_2^n$ such that

$$\min_{\hat{x}_2^n \in \mathcal{B}((z^n)^*, y^n)} d_2(x^n, \hat{x}_2^n) \le D_2, \tag{9.73}$$

3. and the size of the set $\mathcal{B} \cup \mathcal{B}((z^n)^*, y^n)$ satisfies

$$\log \left|\mathcal{B} \cup \mathcal{B}((z^n)^*, y^n)\right| \leq nR(Q_{XY}, D_1, D_2) + c \log(n+1). \tag{9.74}$$

The proof of Lemma 9.9 is similar to the proof of type covering lemmas for rate-distortion problem.

Invoking Lemma 9.9, we can upper bound the excess-distortion probability of an (n, M)-code. To do so, for any $(n, M) \in \mathbb{N}^2$, define

$$R_n := \frac{1}{n} \log M - (c + |\mathcal{X}| \cdot |\mathcal{Y}|) \frac{\log(n+1)}{n}. \tag{9.75}$$

Lemma 9.10. There exists an (n, M)-code whose excess-distortion probability satisfies

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \Pr\left\{R_n < R(\hat{T}_{X^n Y^n}, D_1, D_2)\right\}. \tag{9.76}$$

Proof. Consider the following coding scheme. Given source sequence pair (x^n, y^n), the encoder first calculates the joint type $\hat{T}_{x^n y^n}$, which can be transmitted reliably using at most $|\mathcal{X}| \cdot |\mathcal{Y}| \log(n+1)$ nats. Then the encoder calculates $R(\hat{T}_{x^n y^n}, D_1, D_2)$ and declares an error if $nR(\hat{T}_{x^n y^n}, D_1, D_2) + c \log(n+1) + |\mathcal{X}| \cdot |\mathcal{Y}| \log(n+1) > \log M$. Otherwise, the encoder chooses a set $\mathcal{B}$ satisfying the properties specified in Lemma 9.9 and sends the index of $(z^n)^* = \arg\min_{\hat{x}_1^n \in \mathcal{B}} d_1(x^n, \hat{x}_1^n)$. Subsequently, the decoder chooses a set $\mathcal{B}((z^n)^*, y^n)$ satisfying the properties specified in Lemma 9.9 and sends the index of $\arg\min_{\hat{x}_2^n \in \mathcal{B}((z^n)^*, y^n)} d_2(x^n, \hat{x}_2^n)$. Lemma 9.9 implies that the decoding is error free if $nR(\hat{T}_{x^n y^n}, D_1, D_2) + c \log(n+1) + |\mathcal{X}| \cdot |\mathcal{Y}| \log(n+1) \leq \log M$. The proof of Lemma 9.10 is now completed. □

Given any distribution P_{XY} on the finite set $\mathcal{X} \times \mathcal{Y}$, define the typical set

$$\mathcal{A}_n(P_{XY}) := \left\{Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) : \|Q_{XY} - P_{XY}\|_\infty \leq \sqrt{\frac{\log n}{n}}\right\}. \tag{9.77}$$

It follows from [119, Lemma 22] that

$$\Pr\left\{\hat{T}_{X^nY^n} \notin \mathcal{A}_n(P_{XY})\right\} \leq \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2}. \tag{9.78}$$

If we choose

$$\frac{1}{n}\log M = R(P_{XY}, D_1, D_2) + \frac{L}{\sqrt{n}} + \Big(c + |\mathcal{X}|\cdot|\mathcal{Y}|\Big)\frac{\log(n+1)}{n}, \tag{9.79}$$

then

$$R_n = R(P_{XY}, D_1, D_2) + \frac{L}{\sqrt{n}}. \tag{9.80}$$

For any (x^n, y^n) such that $\hat{T}_{x^ny^n} \in \mathcal{A}_n(P_{XY})$, since the mapping $Q_{XY} \to R(Q_{XY}, D_1, D_2)$ is twice differentiable in the neighborhood of P_{XY} and the derivative is bounded, applying Taylor expansion of $R(\hat{T}_{x^ny^n}, D_1, D_2)$ around $\hat{T}_{x^ny^n} = P_{XY}$ and using Lemma 9.4, we have

$$R(\hat{T}_{x^ny^n}, D_1, D_2) = \frac{1}{n}\sum_{i\in[n]} \jmath(x_i, y_i|D_1, D_2, P_{XY}) + O\Big(\frac{\log n}{n}\Big). \tag{9.81}$$

Define $\xi_n = \frac{\log n}{n}$. It follows from Lemma 9.10 that

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\leq \Pr\left\{R_n < R(\hat{T}_{X^nY^n}, D_1, D_2), \hat{T}_{X^nY^n} \in \mathcal{A}_n(P_{XY})\right\} \\
&\qquad + \Pr\left\{\hat{T}_{X^nY^n} \notin \mathcal{A}_n(P_{XY})\right\} &(9.82)\\
&\leq \Pr\Bigg\{R(P_{XY}, D_1, D_2) + \frac{L}{\sqrt{n}} \\
&\qquad\quad < \frac{1}{n}\sum_{i\in[n]} \jmath(X_i, Y_i|D_1, D_2, P_{XY}) + O(\xi_n)\Bigg\} + \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2} &(9.83)\\
&\leq \Pr\Bigg\{\frac{1}{\sqrt{n}}\sum_{i\in[n]} \big(\jmath(X_i, Y_i|D_1, D_2, P_{XY}) \\
&\qquad\quad - R(P_{XY}, D_1, D_2)\big) > L + O(\xi_n\sqrt{n})\Bigg\} + \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2} &(9.84)\\
&\leq \mathrm{Q}\left(\frac{L + O(\xi_n\sqrt{n})}{\sqrt{\mathrm{V}(D_1, D_2|P_{XY})}}\right) + \frac{6\mathrm{T}(D_1, D_2|P_{XY})}{\sqrt{n}\mathrm{V}^{3/2}(D_1, D_2|P_{XY})} + \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2}, &(9.85)
\end{aligned}$$

where (9.83) follows from the results in (9.78) and Lemma 9.81 and (9.85) follows from Berry-Esseen theorem, where $\mathrm{T}(D_1, D_2|P_{XY})$ is the third absolute moment of $\jmath(X, Y|D_1, D_2, P_{XY})$, which is finite for DMS.

Therefore, if L satisfies

$$L \geq \sqrt{\mathrm{V}(D_1, D_2|P_{XY})}\mathrm{Q}^{-1}(\varepsilon), \tag{9.86}$$

by noting that $O(\xi_n\sqrt{n}) = O(\log n/\sqrt{n})$, it follows that

$$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \varepsilon. \tag{9.87}$$

Thus, the optimal second-order coding rate satisfies that $L^*(\varepsilon, D_1, D_2) \leq \sqrt{\mathrm{V}(D_1, D_2|P_{XY})}Q^{-1}(\varepsilon)$.

9.6.2 Converse

The converse part follows by applying the Berry-Esseen theorem to the non-asymptotic converse bound in Theorem 9.5. Let

$$\log M := nR(P_{XY}, D_1, D_2) + L\sqrt{n} - \frac{1}{2}\log n. \tag{9.88}$$

Invoking (9.27) with $\varepsilon = \frac{\log n}{2n}$, we obtain

$$\begin{aligned} &\mathrm{P}_{\mathrm{e},n}(D_1, D_2) + \frac{1}{\sqrt{n}} \\ &\geq \Pr\Big\{\sum_{i\in[n]} \jmath(x, y|D_1, D_2, P_{XY}) \geq nR(P_{XY}, D_1, D_2) + L\sqrt{n}\Big) && (9.89) \\ &\geq \mathrm{Q}\left(\frac{L}{\sqrt{\mathrm{V}(D_1, D_2|P_{XY})}}\right) - \frac{6\mathrm{T}(D_1, D_2|P_{XY})}{\sqrt{n}\mathrm{V}^{3/2}(D_1, D_2|P_{XY})}, && (9.90) \end{aligned}$$

where (9.90) follows from the Berry-Esseen theorem. If

$$L < \sqrt{\mathrm{V}(D_1, D_2|P_{XY})}Q^{-1}(\varepsilon), \tag{9.91}$$

then

$$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) > \varepsilon. \tag{9.92}$$

Hence, the optimal second-order coding rate satisfies $L^*(\varepsilon, D_1, D_2) \geq \sqrt{\mathrm{V}(D_1, D_2|P_{XY})}Q^{-1}(\varepsilon)$.

10

Successive Refinement

In this section, we study the successive refinement problem with two encoders and two decoders, which generalizes the rate-distortion problem by introducing an additional pair of encoders and decoders. Based on the encoding process of the original encoder, the additional encoder further compresses the source sequence and the additional decoder uses compressed information from both encoders to produce a finer estimate of the source sequence than the first decoder that only accesses the original encoder. The optimal rate-distortion region for DMS under bounded distortion measures was derived by Rimoldi in [95], which collects rate pairs of encoders with vanishing joint excess-distortion probabilities.

Successive refinement is the first lossy source coding problem with multiple encoders studied in this monograph. The successive refinement problem is an information-theoretic formulation of whether it is possible to interrupt a transmission to provide a finer reconstruction of the source sequence without any loss of optimality for lossy compression. For such a problem, in order to derive the second-order asymptotics, we need to study the backoff of the encoders' rates from a boundary rate-point on the rate-distortion region, analogous to the study of the backoff of the encoder's rate from the rate-distortion function in second-order asymptotics for the rate-distortion problem. For DMS under bounded distortion measures, we derive the optimal second-order coding region under a joint excess-distortion probability (JEP) criterion [151]. We also recall the second-order asymptotics under the separate excess-distortion probabilities (SEP) criteria by No, Ingber and Weissman [82]. For successively refinable discrete memoryless source-distortion measure triplets [31], [64], under SEP, the second-order region is significantly simplified and the notion of successive refinability [31], [64] is generalized to the second-order asymptotic regime. This section is largely based on [82], [151].

There are several new insights on the second-order coding region that we can glean when we consider the joint excess-distortion probability (cf. Section 10.3.4). For example, under the joint excess-distortion probability criterion, the second-order region is curved for successively refinable source-distortion triplets, which implies that if one second-order coding rate is small, the other is necessarily large. This reveals a fundamental tradeoff that cannot be observed if one adopts the separate excess-distortion probability criterion. Therefore, in subsequent sections that involve more complicated multiterminal lossy source coding problems, we only consider the joint excess-distortion probability criterion that better captures the rate tradeoff of multiple encoders.

10.1 Problem Formulation and Asymptotic Result

10.1.1 Problem Formulation

The successive refinement source coding problem [31], [63]–[65], [95] is shown in Figure 10.1. There are two encoders and two decoders. For each $i \in [2]$, encoder f_i has access to a source sequence X^n and compresses it into a message S_i. Decoder ϕ_1 aims to recover source sequence X^n under distortion measure d_1 and distortion level D_1 with the encoded message S_1 from encoder f_1. The decoder ϕ_2 aims to recover X^n under distortion measure d_2 and distortion level D_2 with messages S_1 and S_2.

We consider a memoryless source with distribution P_X supported on a finite alphabet $\mathcal{X}$. Thus, X^n is an i.i.d. sequence where each X_i

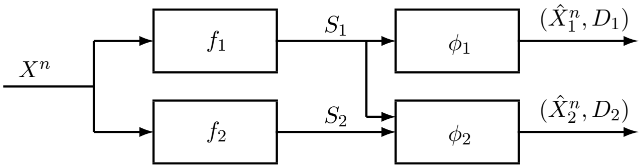

Figure 10.1: System model for the successive refinement problem [95].

is generated according to P_X. We assume the reproduction alphabets for decoder ϕ_1, ϕ_2 are respectively alphabets $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$. We follow the definitions in [95] for codes and the achievable rate region.

Definition 10.1. An (n, M_1, M_2)-code for successive refinement source coding consists of two encoders:

$$f_1 : \mathcal{X}^n \to \mathcal{M}_1 = [M_1], \tag{10.1}$$

$$f_2 : \mathcal{X}^n \to \mathcal{M}_2 = [M_2], \tag{10.2}$$

and two decoders:

$$\phi_1 : \mathcal{M}_1 \to \hat{\mathcal{X}}_1^n, \tag{10.3}$$

$$\phi_2 : \mathcal{M}_1 \times \mathcal{M}_2 \to \hat{\mathcal{X}}_2^n. \tag{10.4}$$

For each $i \in [2]$, define a distortion measure $d_i : \mathcal{X} \times \hat{\mathcal{X}}_i \to [0, \infty)$ and let the distortion between x^n and $\hat{x}_i^n$ be defined as $d_i(x^n, \hat{x}_i^n) := \frac{1}{n} \sum_{j \in [n]} d_i(x_j, \hat{x}_{j,i})$. Define the joint excess-distortion probability as

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) := \Pr\left\{ d_1(X^n, \hat{X}_1^n) > D_1 \text{ or } d_2(X^n, \hat{X}_2^n) > D_2 \right\}, \tag{10.5}$$

where $\hat{X}_1^n = \phi_1(f_1(X^n))$ and $\hat{X}_2^n = \phi_2(f_1(X^n), f_2(X^n))$ are the reconstructed sequences.

Definition 10.2. A rate pair (R_1, R_2) is said to be (D_1, D_2)-achievable for the successive refinement problem if there exists a sequence of (n, M_1, M_2)-codes such that

$$\limsup_{n \to \infty} \frac{1}{n} \log M_1 \le R_1, \tag{10.6}$$

$$\limsup_{n \to \infty} \frac{1}{n} \log(M_1 M_2) \le R_1 + R_2, \tag{10.7}$$

and

$$\lim_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) = 0. \tag{10.8}$$

The closure of the set of all (D_1, D_2)-achievable rate pairs is called optimal (D_1, D_2)-achievable rate region and denoted as $\mathcal{R}(P_X, D_1, D_2)$.

Note that in the original work by Rimoldi [95], the rate R_2 corresponds to the *sum* rate $R_1 + R_2$. In this monograph, to be consistent with other sections, we use R_2 to denote the rate of message S_2 in Figure 10.1.

10.1.2 Rimoldi's Rate-Distortion Region

The optimal rate region for DMS with arbitrary distortion measures was characterized in [95]. Let $\mathcal{P}(P_X, D_1, D_2)$ be the set of joint distributions $P_{X\hat{X}_1\hat{X}_2}$ such that the $\mathcal{X}$-marginal is P_X, $\mathbb{E}[d_1(X, \hat{X}_1)] \le D_1$ and $\mathbb{E}[d_2(X, \hat{X}_2)] \le D_2$. Given $P_{X\hat{X}_1\hat{X}_2} \in \mathcal{P}(P_X, D_1, D_2)$, let

$$\begin{aligned}&\mathcal{R}(P_{X\hat{X}_1\hat{X}_2})\\&:= \Big\{(R_1, R_2) : R_1 \ge I(P_X, P_{X|\hat{X}_1}),\ R_1 + R_2 \ge I(P_X, P_{X|\hat{X}_1,\hat{X}_2})\Big\}.\end{aligned} \tag{10.9}$$

Theorem 10.1. The optimal (D_1, D_2)-achievable rate region for DMS with arbitrary distortion measures under successive refinement source coding is

$$\mathcal{R}(P_X, D_1, D_2) = \bigcup_{P_{X\hat{X}_1\hat{X}_2} \in \mathcal{P}(P_X, D_1, D_2)} \mathcal{R}(P_{X\hat{X}_1\hat{X}_2}). \tag{10.10}$$

Now we introduce an important quantity for subsequent analyses for DMS. Given a rate R_1 and distortion pair (D_1, D_2), let the minimal sum rate $R_1 + R_2$ such that $(R_1, R_2) \in \mathcal{R}(P_X, D_1, D_2)$ be $\mathsf{R}(R_1|P_X, D_1, D_2)$, i.e.,

$$\begin{aligned}&\mathsf{R}(R_1|P_X, D_1, D_2)\\&:= \min\{R_1 + R_2 : (R_1, R_2) \in \mathcal{R}(P_X, D_1, D_2)\} &(10.11)\\&= \inf_{\substack{P_{\hat{X}_1\hat{X}_2|X} : \mathbb{E}[d_1(X,\hat{X}_1)] \le D_1\\ \mathbb{E}[d_2(X,\hat{X}_2)] \le D_2, I(P_X, P_{X|\hat{X}_1}) \le R_1}} I(P_X, P_{X|\hat{X}_1,\hat{X}_2}), &(10.12)\end{aligned}$$

where (10.12) follows from [55, Corollary 1].

Let $R(P_X, D_1)$ and $R(P_X, D_2)$ be the rate-distortion functions [38, Chapter 3] (see also (3.7)) when the reproduction alphabets are $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$ respectively, i.e., for each $i \in [2]$,

$$R(P_X, D_i) := \inf_{P_{\hat{X}_i|X}:\mathbb{E}[d_i(X,\hat{X}_i)]\le D_i} I(P_X, P_{X|\hat{X}_i}). \tag{10.13}$$

Note that if $R_1 < R(P_X, D_1)$, then the convex optimization in (10.12) is infeasible. Otherwise, since $\mathsf{R}(R_1|P_X, D_1, D_2)$ is a convex optimization problem, the minimization in (10.12) is attained for some test channel $P_{\hat{X}_1\hat{X}_2|X}$ satisfying

$$\sum_{x,y,z} P_X(x)P_{\hat{X}_1\hat{X}_2|X}(\hat{x}_1, \hat{x}_2|x)d_1(x, \hat{x}_1) = D_1, \tag{10.14}$$

$$\sum_{x,y,z} P_X(x)P_{\hat{X}_1\hat{X}_2|X}(\hat{x}_1, \hat{x}_2|x)d_2(x, \hat{x}_2) = D_2, \tag{10.15}$$

$$I(P_X, P_{\hat{X}_1|X}) = R_1. \tag{10.16}$$

Therefore, a rate pair (R_1^*, R_2^*) lies on the boundary of the rate-distortion region $\mathcal{R}(P_X, D_1, D_2)$ if and only if $R_1^* = R(P_X, D_1)$ and $R_1^* + R_2^* \ge \mathsf{R}(R_1^*|P_X, D_1, D_2)$ or $R_1^* > R(P_X, D_1)$ and $R_1^* + R_2^* = \mathsf{R}(R_1^*|P_X, D_1, D_2)$.

10.1.3 Successive Refinability

Next we introduce the notion of a *successively refinable source-distortion measure triplet* [31], [64]. We recall the definitions with a slight generalization in accordance to [82, Definition 2].

Definition 10.3. Given distortion measures d_1, d_2 and a source X with distribution P_X, the source-distortion measure triplet (X, d_1, d_2) is said to be (D_1, D_2)-successively refinable if the rate pair $(R(P_X, D_1), R(P_X, D_2))$ is (D_1, D_2)-achievable. If the source-distortion measure triplet is (D_1, D_2)-successively refinable for all (D_1, D_2) such that $R(P_X, D_1) < R(P_X, D_2)$, then it is said to be successively refinable.

For a successively refinable source-distortion measure triplet, the minimal sum rate $R_1 + R_2$ given R_1 in a certain interval is exactly the rate-distortion function (see (10.27) to follow). This reduces the computation of the optimal rate region in (10.10).

Koshelev [64] presented a sufficient condition for a source-distortion measure triplet to be successively refinable while Equitz and Cover [31, Theorem 2] presented a necessary and sufficient condition which we reproduce below.

Theorem 10.2. A memoryless source-distortion measure triplet is successively refinable if and only if there exists a conditional distribution $P^*_{\hat{X}_1\hat{X}_2|X}$ such that

$$R(P_X, D_1) = I(P_X, P^*_{\hat{X}_1|X}),\ \mathbb{E}_{P_X \times P^*_{\hat{X}_1|X}}[d_1(X, \hat{X}_1)] \leq D_1, \quad (10.17)$$

$$R(P_X, D_2) = I(P_X, P^*_{\hat{X}_2|X}),\ \mathbb{E}_{P_X \times P^*_{\hat{X}_2|X}}[d_2(X, \hat{X}_2)] \leq D_2, \quad (10.18)$$

and

$$P^*_{\hat{X}_1\hat{X}_2|X} = P^*_{\hat{X}_1|X} P^*_{\hat{X}_2|X}. \quad (10.19)$$

In [31], it was shown that DMS with Hamming distortion measures, GMS with quadratic distortion measures, and Laplacian sources with absolute distortion measures are successively refinable. Note that in the original paper of Equitz and Cover [31], the authors only considered the case where both decoders use the same distortion measure, i.e., $d_1 = d_2 = d$. Interestingly, as pointed out in [82, Theorem 4], the result still holds even when $d_1 \neq d_2$. This can be verified easily for DMS by invoking [95, Theorem 1].

10.2 Rate-Distortions-Tilted Information Density

Throughout the section, we assume that $R(P_X, D_1) \leq R_1^* < R(P_X, D_2)$ and $\mathcal{R}(P_X, D_1, D_2)$ is smooth on a boundary rate pair (R_1^*, R_2^*) of our interest, i.e.,

$$\xi^* := -\frac{\mathsf{R}(R_1|P_X, D_1, D_2)}{\partial R}\bigg|_{R=R_1^*}, \quad (10.20)$$

is well-defined. Note that $\xi^* \geq 0$ since $\mathsf{R}(R_1, D_2, D_2)$ is convex and non-increasing in R_1. Further, for a positive distortion pair (D_1, D_2), define

$$\nu_1^* := -\frac{\mathsf{R}(P_X, R_1, D, D_2)}{\partial D}\bigg|_{D=D_1}, \tag{10.21}$$

$$\nu_2^* := -\frac{\mathsf{R}(P_X, R_1, D_1, D)}{\partial D}\bigg|_{D=D_2}. \tag{10.22}$$

Note that for a successively refinable discrete memoryless source-distortion measure triplet, from (10.27), we obtain $\xi^* = 0$ and $\nu_1^* = 0$. Let $P^*_{\hat{X}_1\hat{X}_2|X}$ be the optimal test channel achieving $\mathsf{R}(R_1|P_X, D_1, D_2)$ in (10.11) (assuming it is unique)[1]. Let $P^*_{X\hat{X}_1}, P^*_{X\hat{X}_2}, P^*_{\hat{X}_1\hat{X}_2}$, $P^*_{\hat{X}_1}$, and $P^*_{\hat{X}_1|X}$ be the induced (conditional) marginal distributions. We are now ready to define the tilted information density for successive refinement source coding problem.

Let (R_1^*, R_2^*) be any boundary rate pair of the rate-distortion region $\mathcal{R}(P_X, D_1, D_2)$.

Definition 10.4. For any $x \in \mathcal{X}$, the rate-distortions tilted information density for the successive refinement problem is defined as

$$\begin{aligned}
&\jmath(x, R_1^*|D_1, D_2, P_X) \\
&:= -\log \mathbb{E}_{P^*_{\hat{X}_1\hat{X}_2}}\Bigg(\exp\bigg\{-\xi^*\bigg(\log\frac{P^*_{\hat{X}_1|X}(\hat{X}_1|x)}{P^*_{\hat{X}_1}(\hat{X}_1)} - R_1^*\bigg) \\
&\qquad - \nu_1^*(d_1(x,\hat{X}_1) - D_1) - \nu_2^*(d_2(x,\hat{X}_2) - D_2))\bigg\}\Bigg).
\end{aligned} \tag{10.23}$$

The properties of $\jmath(x, R_1^*|D_1, D_2, P_X)$ are summarized in the following lemma.

Lemma 10.3. The following claims hold.

1. For any $(\hat{x}_1, \hat{x}_2)$ such that $P^*_{\hat{X}_1\hat{X}_2}(\hat{x}_1, \hat{x}_2) > 0$,

$$\begin{aligned}
&\jmath(x, R_1^*|D_1, D_2, P_X) \\
&= \log\frac{P^*_{\hat{X}_1\hat{X}_2|X}(\hat{x}_1,\hat{x}_2|x)}{P^*_{\hat{X}_1\hat{X}_2}(\hat{x}_1,\hat{x}_2)} + \xi^*\left(\log\frac{P^*_{\hat{X}_1|X}(\hat{x}_1|x)}{P^*_{\hat{X}_1}(\hat{x}_1)} - R_1^*\right) \\
&\qquad - \nu_1^*(d_1(x,\hat{x}_1) - D_1) - \nu_2^*(d_2(x,\hat{x}_2) - D_2).
\end{aligned} \tag{10.24}$$

[1]If optimal test channels are not unique, then following the proof of [127, Lemma 2], we can argue that the tilted information density is still well defined.

2. The minimal sum rate $\mathsf{R}(R_1^*|P_X, D_1, D_2)$ equals the expectation of the rate-distortions-tilted information density, i.e.,

$$\mathsf{R}(R_1^*|P_X, D_1, D_2) = \mathbb{E}_{P_X}\left[\jmath(X, R_1^*|D_1, D_2, P_X)\right]. \tag{10.25}$$

3. Suppose that for all Q_X in the neighborhood of P_X, $\mathrm{supp}(Q^*_{\hat{X}_1\hat{X}_2}) = \mathrm{supp}(P^*_{\hat{X}_1\hat{X}_2})$. Then for all $a \in \mathcal{X}$,

$$\left.\frac{\partial \mathsf{R}(R_1^*|Q_X, D_1, D_2)}{\partial Q_X(a)}\right|_{Q_X=P_X} = \jmath(a, R_1^*|D_1, D_2, P_X) - (1+\xi^*). \tag{10.26}$$

Lemma 10.3 generalizes the properties of the distortion-tilted information density for the rate-distortion problem in Lemma 3.2, which are also available in [69, Properties 1-3] and [66, Theorems 2.1-2.2].

For a successively refinable discrete memoryless source-distortion measure triplet, it follows from Definition 10.3 that if $R(P_X, D_1) \leq R_1 < R(P_X, D_2)$,

$$\mathsf{R}(R_1|P_X, D_1, D_2) = R(P_X, D_2). \tag{10.27}$$

In this case, $\xi^* = 0$, $\nu_1^* = 0$. The rate-distortions-tilted information density $\jmath(x, R_1^*|D_1, D_2, P_X)$ reduces to the distortion-tilted information density $\jmath(x|D_2, P_X)$ in (3.17) for the rate-distortion problem, where

$$\jmath(x|D, P_X) = -\log \mathbb{E}_{P^*_{\hat{X}}}[\exp(-\lambda_1^*(d(x,\hat{X}) - D))], \tag{10.28}$$

$$\lambda^* = -\left.\frac{\partial R(P_X, D')}{\partial D'}\right|_{D'=D}. \tag{10.29}$$

10.3 Second-Order Asymptotics

10.3.1 Definitions and Discussions

Recall that for the rate-distortion problem with only one encoder, the second-order coding rate is defined as the backoff from the minimal achievable rate, i.e., the rate-distortion function $R(P_X, D)$ (cf. Definition 3.6). Analogously, for a multiterminal lossy source coding problem such as successive refinement, in order to derive the second-order asymptotics, we need to study the backoff of the rates of encoders from a boundary point on the rate-distortion region, which is a minimal achievable rate pair and takes role of the rate-distortion function for the rate-distortion problem.

Formally, let (R_1^*, R_2^*) be a rate pair on the boundary of the rate-distortion region $\mathcal{R}(P_X, D_1, D_2)$. The second-order coding region for the successive refinement problem is defined as follows.

Definition 10.5. Given any $\varepsilon \in (0,1)$, a pair (L_1, L_2) is said to be second-order $(R_1^*, R_2^*, D_1, D_2, \varepsilon)$-achievable if there exists a sequence of (n, M_1, M_2)-codes such that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} (\log M_1 - nR_1^*) \le L_1, \tag{10.30}$$

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} (\log(M_1 M_2) - n(R_1^* + R_2^*)) \le L_2, \tag{10.31}$$

and

$$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \le \varepsilon. \tag{10.32}$$

The closure of the set of all second-order $(R_1^*, R_2^*, D_1, D_2, \varepsilon)$-achievable pairs is called the second-order $(R_1^*, R_2^*, D_1, D_2, \varepsilon)$ coding region and denoted as $\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)$.

We emphasize that the JEP criterion (10.32) is consistent with original setting of successive refinement in Rimoldi's work [95] and the error exponent analysis of Kanlis and Narayan [55]. In contrast, Tuncel and Rose [121] considered the separate excess-distortion events and probabilities and derived the tradeoff between exponents of two excess-distortion probabilities. Note that the rate-distortion region remains the same [31], [95] regardless whether we consider vanishing joint or the separate excess-distortion probabilities. In the study of second-order asymptotics, the second-order coding region can also be defined under the SEP criterion [82]. Specifically, the second-order coding region $\mathcal{L}_{\mathrm{sep}}(R_1^*, R_2^*, D_1, D_2, \varepsilon_1, \varepsilon_2)$ is defined similar to Definition 10.5, except that (10.32) is replaced by

$$\limsup_{n\to\infty} \Pr\left\{d_1(X^n, \hat{X}_1^n) > D_1\right\} \le \varepsilon_1, \tag{10.33}$$

$$\limsup_{n\to\infty} \Pr\left\{d_2(X^n, \hat{X}_2^n) > D_2\right\} \le \varepsilon_2, \tag{10.34}$$

for some fixed $(\varepsilon_1, \varepsilon_2) \in (0,1)^2$ and the boundary rate-pair (R_1^*, R_2^*) is fixed as $R_1^* = R(P_X, D_1)$ and $R_1^* + R_2^* = \mathsf{R}(R(P_X, D)|P_X, D_1, D_2)$, which corresponds to the case where both encoders respectively use their own optimal (i.e., minimum possible) asymptotic rates.

The main content of this section is the characterization of $\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)$ and $\mathcal{L}_{\mathrm{sep}}(R_1^*, R_2^*, D_1, D_2, \varepsilon_1, \varepsilon_2)$ for DMS under bounded distortion measures, e.g., a binary source with Hamming distortion measures. We note that $\mathcal{L}(R_1, R_2, D_1, D_2, \varepsilon)$ can, in principle, be evaluated for rate pairs that are not on the boundary of the first-order region $\mathcal{R}(P_X, D_1, D_2)$. However, this would lead to degenerate solutions.

We next explain some advantages of using the JEP criterion over the SEP criterion in second-order asymptotics.

1. The JEP criterion is consistent with recent works in the second-order literature [78], [113], [127]. For example, in [78], Le, *et al.* established the second-order asymptotics for the Gaussian interference channel in the strictly very strong interference regime under the joint error probability criterion. If in [78], one adopts the separate error probabilities criterion, one would *not* be able to observe the performance tradeoff between the two decoders.

2. In Section 10.3.4, we show, via different proof techniques compared to existing works, that the second-order region is curved for successively refinable source-distortion triplets. This shows that if one second-order coding rate is small, the other is necessarily large. This reveals a fundamental tradeoff that cannot be observed if one adopts the separate excess-distortion probability criterion.

10.3.2 General DMS

Recall that $\Psi(x_1, x_2; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the bivariate generalization of the Gaussian cdf. Given each $i \in [2]$, let $\mathrm{V}(D_i|P_X) := \mathrm{Var}[\jmath(X|D_i, P_X)]$ be the *rate-dispersion function* (cf. (3.57)). Given a rate pair (R_1^*, R_2^*) on the boundary of $\mathcal{R}(P_X, D_1, D_2)$, also define another rate-dispersion function $\mathrm{V}(R_1^*|P_X, D_1, D_2) := \mathrm{Var}\left[\jmath(X, R_1^*|D_1, D_2, P_X)\right]$. Let $\mathbf{V}(R_1^*|P_X, D_1, D_2) \succeq 0$ be the covariance matrix of the two-dimensional random vector $[\jmath(X|D_1, P_X), \jmath(X, R_1^*|D_1, D_2, P_X)]^\top$, i.e., the *rate-dispersion matrix*.

We impose the following conditions on the rate pair (R_1^*, R_2^*), the distortion measures (d_1, d_2), the distortion levels (D_1, D_2) and the source distribution P_X:

1. $\mathsf{R}(R_1^*|P_X, D_1, D_2)$ is finite;
2. $\xi^* \geq 0$ in (10.20) and $\nu_i^*,\ i = 1, 2$ in (10.21), (10.22) are well-defined;
3. $(Q_X, D_1') \mapsto R(Q_X, D_1')$ is twice differentiable in the neighborhood of (P_X, D_1) and the derivatives are bounded (i.e., the spectral norm of the Hessian matrix is bounded);
4. $(R_1, D_1', D_2', Q_X) \mapsto \mathsf{R}(R_1|Q_X, D_1', D_2')$ is twice differentiable in the neighborhood of (R_1^*, D_1, D_2, P_X) and the derivatives are bounded;

Note that similar regularity assumptions were made on second-order asymptotics for the rate-distortion and Kaspi problems.

We first present the second-order asymptotics under the JEP criterion.

Theorem 10.4. Under conditions (1) to (4), depending on the values of (R_1^*, R_2^*), for any $\varepsilon \in (0, 1)$, the second-order coding region satisfies:

- Case (i): $R(P_X, D_1) < R_1^* < \mathsf{R}(R_1^*|P_X, D_1, D_2)$ and $R_1^* + R_2^* = \mathsf{R}(R_1^*|P_X, D_1, D_2)$

$$\begin{aligned}&\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)\\&= \left\{(L_1, L_2) : \xi^* L_1 + L_2 \geq \sqrt{\mathrm{V}(R_1^*|P_X, D_1, D_2)}\mathrm{Q}^{-1}(\varepsilon)\right\}.\end{aligned} \tag{10.35}$$

- Case (ii): $R_1^* = R(P_X, D_1)$ and $R_1^* + R_2^* > \mathsf{R}(R_1^*|P_X, D_1, D_2)$

$$\begin{aligned}&\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)\\&= \left\{(L_1, L_2) : L_1 \geq \sqrt{\mathrm{V}(P_X, D_1)}\mathrm{Q}^{-1}(\varepsilon)\right\}.\end{aligned} \tag{10.36}$$

- Case (iii): $R_1^* = R(P_X, D_1)$, $R_1^* + R_2^* = \mathsf{R}(R_1^*|P_X, D_1, D_2)$ and $\mathrm{rank}(\mathbf{V}(R_1^*|P_X, D_1, D_2)) \geq 1$,

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)$$
$$= \Big\{(L_1, L_2) : \Psi(L_1, \xi^* L_1 + L_2; \mathbf{0}, \mathbf{V}(R_1^*|P_X, D_1, D_2)) \geq 1 - \varepsilon\Big\}. \tag{10.37}$$

The proof of Theorem 10.4 is provided in Section 10.4. In the achievability part, we leverage the type covering lemma [82, Lemma 8]. In the converse part, we follow the perturbation approach proposed by Gu and Effros in their proof for the strong converse of Gray-Wyner problem [45], leading to a type-based strong converse. In the proofs of both directions, we leverage the properties of appropriately defined rate-distortions-tilted information densities and use the (multi-variate) Berry-Esseen theorem. An alternative converse proof of Theorem 10.4 is possible by applying the Berry-Esseen theorem to the non-asymptotic converse bound in [68, Corollary 2] (see also Lemma 11.6 from our analysis of the Fu-Yeung problem), analogous to the converse proof of second-order asymptotics for the rate-distortion and Kaspi problems. We omit the alternative converse proof of Theorem 10.4.

In both Cases (i) and (ii), the code is operating at a rate pair bounded away from one of the first-order fundamental limits. Hence, a univariate Gaussian suffices to characterize the second-order behavior. In contrast, for Case (iii), the code is operating at precisely the two first-order fundamental limits. Hence, in general, we need a bivariate Gaussian to characterize the second-order behavior. Using an argument by Tan and Kosut [113, Theorem 6], we note that this result holds for both positive definite and rank deficient rate-dispersion matrices $\mathbf{V}(R_1^*|P_X, D_1, D_2)$. However, we exclude the degenerate case in which $\mathrm{rank}(\mathbf{V}(R_1^*|P_X, D_1, D_2)) = 0$. Note that if the rank of $\mathbf{V}(R_1^*|P_X, D_1, D_2)$ is 0, it means that the dispersion matrix is all zeros matrix, i.e., $\mathrm{Cov}[\jmath(X|D_1, P_X), \jmath(X, R_1^*|D_1, D_2, P_X)] = 0$, $\mathrm{V}(P_X, D_1) = 0$, and $\mathrm{V}(R_1^*|P_X, D_1, D_2) = 0$. This implies that $\jmath(x|D_1, P_X)$ and $\jmath(x, R_1^*|D_1, D_2, P_X)$ are both deterministic. In this case, the second-order term (dispersion) vanishes, and if one seeks refined asymptotic estimates for the optimal finite blocklength coding rates, one would then be interested to analyze the *third-order* or $\Theta(\log n)$ asymptotics (cf. [70, Theorem 18]).

We next present inner (achievability) and outer (converse) bounds on the second-order coding region under the SEP criterion.

Theorem 10.5. Under conditions (1) to (4), for any $(\varepsilon_1, \varepsilon_2) \in (0,1)^2$, the second-order coding region $\mathcal{L}_{\rm sep}(R_1^*, R_2^*, D_1, D_2, \varepsilon_1, \varepsilon_2)$ satisfies that when $R_1^* = R(P_X, D_1)$ and $R_1^* + R_2^* = \mathrm{R}(R(P_X, D_1)|P_X, D_1, D_2)$,

$$\begin{aligned}
&\Big\{(L_1, L_2):\ L_1 \geq \sqrt{\mathrm{V}(P_X, D_1)}\mathrm{Q}^{-1}(\min\{\varepsilon_1, \varepsilon_2\}),\\
&\qquad\qquad L_2 \geq \sqrt{\mathrm{V}(R_1^*|P_X, D_1, D_2)}\mathrm{Q}^{-1}(\min\{\varepsilon_1, \varepsilon_2\})\Big\}\\
&\subseteq \mathcal{L}_{\rm sep}(R_1^*, R_2^*, D_1, D_2, \varepsilon_1, \varepsilon_2)\\
&\subseteq \Big\{(L_1, L_2):\ L_1 \geq \sqrt{\mathrm{V}(P_X, D_1)}\mathrm{Q}^{-1}(\varepsilon_1),\\
&\qquad\qquad L_2 \geq \sqrt{\mathrm{V}(R_1^*|P_X, D_1, D_2)}\mathrm{Q}^{-1}(\varepsilon_2)\Big\}. \qquad (10.38)
\end{aligned}$$

The achievability proof of Theorem 10.5 was proved by No *et al.* using the type covering lemma for the successive refinement problem [82, Section V] and the converse part follows by applying the Berry-Esseen theorem to the non-asymptotic converse bound by Kostina and Tuncel [68, Theorem 3]. The inner bound could also be obtained similar to the proof Case (iii) of Theorem 10.4 with ε replaced by $\min\{\varepsilon_1, \varepsilon_2\}$.

The inner and outer bounds match when $\varepsilon_1 = \varepsilon_2$. It was claimed by No *et al.* [82] that the outer bound was achievable for any $(\varepsilon_1, \varepsilon_2) \in (0,1)^2$. However, a careful check suggests that it is impossible. This is because, in order not to incur an excess-distortion event at decoder ϕ_2 for a sequence x^n, decoder ϕ_1 should not incur an excess-distortion event since otherwise, "correct" decoding of decoder ϕ_2 is not guaranteed.

10.3.3 Successively Refinable DMS

In this subsection, we specialize the results in Theorem 10.4 to successively refinable discrete memoryless source-distortion measure triplets. Note that for such source-distortion measure triplets, $\mathsf{R}(R_1^*|P_X, D_1, D_2) = R(P_X, D_2)$ if $R(P_X, D_1) \leq R_1^* < R(P_X, D_2)$. Hence, $\xi^* = 0$, $\nu_1^* = 0$ and $\jmath(X, R_1^*|D_1, D_2, P_X) = \jmath(X|D_2, P_X)$. The covariance matrix

$\mathbf{V}(R_1^*|P_X, D_1, D_2)$ is also simplified to $\mathbf{V}(P_X, D_1, D_2)$ with diagonal elements being $\mathrm{V}(P_X, D_1)$ and $\mathrm{V}(P_X, D_2)$ and off-diagonal element being the covariance $\mathrm{Cov}[\jmath(X|D_1, P_X), \jmath(X|D_2, P_X)]$. The conditions in Theorem 10.4 are also now simplified to: $(Q_X, D_1') \mapsto R(Q_X, D_1')$ and $(Q_X, D_2') \mapsto R(Q_X, D_2')$ are twice differentiable in the neighborhood of (P_X, D_1, D_2) and the derivatives are bounded.

Corollary 10.6. Under the conditions stated above, depending on (R_1^*, R_2^*), the optimal second-order $(R_1^*, R_2^*, D_1, D_2, \varepsilon)$ coding region for a successively refinable discrete memoryless source-distortion measure triplet is as follows:

- Case (i): $R(P_X, D_1) < R_1^* < R(P_X, D_2)$ and $R_1^* + R_2^* = R(P_X, D_2)$

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \left\{(L_1, L_2) : L_2 \geq \sqrt{\mathrm{V}(P_X, D_2)}\mathrm{Q}^{-1}(\varepsilon)\right\}. \tag{10.39}$$

- Case (ii): $R_1^* = R(P_X, D_2)$ and $R_1^* + R_2^* > R(P_X, D_2)$

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \left\{(L_1, L_2) : L_1 \geq \sqrt{\mathrm{V}(P_X, D_1)}\mathrm{Q}^{-1}(\varepsilon)\right\}. \tag{10.40}$$

- Case (iii): $R_1^* = R(P_X, D_2)$ and $R_1^* + R_2^* = R(P_X, D_2)$ and $\mathrm{rank}(\mathbf{V}(P_X, D_1, D_2)) \geq 1$,

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \left\{(L_1, L_2) : \Psi(L_1, L_2; \mathbf{0}, \mathbf{V}(P_X, D_1, D_2)) \geq 1 - \varepsilon\right\}. \tag{10.41}$$

  Specifically, if $\mathbf{V}(P_X, D_1, D_2) = \mathrm{V}(P_X, D_1) \cdot \mathrm{ones}(2, 2)$, or equivalently $\jmath(X|D_1, P_X) - R_1^* = \jmath(X|D_2, P_X) - R_2^*$ almost surely,

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \left\{(L_1, L_2) : \min\{L_1, L_2\} \geq \sqrt{\mathrm{V}(P_X, D_1)}\mathrm{Q}^{-1}(\varepsilon)\right\}. \tag{10.42}$$

Corollary 10.6 results from specializations of Theorem 10.4. The special case in (10.42) is proved in Section 10.4.3. We notice that the expressions in the second-order regions are simplified for successively refinable discrete memoryless source-distortion measure triplets. In particular, the optimization to compute the optimal test channel $P^*_{\hat{X}_1\hat{X}_2|X}$ in $\mathsf{R}(R_1, D_1, D_2|P_X)$, defined in (10.11)–(10.12), is no longer necessary since the Markov chain $X - Z - Y$ holds for $P^*_{\hat{X}_1\hat{X}_2|X}$ [31].

Furthermore, in Section 10.4.4, we provide an alternative converse proof of Corollary 10.6 by generalizing the one-shot converse bound of Kostina and Verdú in [69, Theorem 1]. We remark that the alternative converse proof is also applicable to successively refinable continuous memoryless source-distortion measure triplets such as GMS with quadratic distortion measures.

The case in (10.42) pertains, for example, to a binary source with Hamming distortion measures. For such a source-distortion measure triplet, $\mathbf{V}(P_X, D_1, D_2)$ is rank 1 and proportional to the all ones matrix. See Section 10.3.4. The result in (10.42) implies that both excess-distortion events in (10.5) are perfectly correlated so that the one consisting of the *smaller* second-order rate L_i, $i = 1, 2$ dominates, since the first-order rates are fixed at the first-order fundamental limits $(R(P_X, D_1), R(P_X, D_2))$. In fact, our result in (10.42) specializes to the scenario where one considers the *separate* excess-distortion criterion [82] in (10.33)–(10.34) with $\varepsilon_1 = \varepsilon_2 = \varepsilon$ and $\mathrm{V}(P_X, D_1) = \mathrm{V}(P_X, D_2)$. More importantly, the case in (10.41) when $\mathbf{V}(P_X, D_1, D_2)$ is full rank pertains to a source-distortion measure triplets with more "degrees-of-freedom". See Section 10.3.4 for a concrete example. Thus our work is a strict generalization of that in [82].

The result under the SEP criterion follows from Theorem 10.5.

Corollary 10.7. Under conditions (1) to (4), for any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$, the second-order coding region $\mathcal{L}_{\mathrm{sep}}(R_1^*, R_2^*, D_1, D_2, \varepsilon_1, \varepsilon_2)$ satisfies that when $R_1^* = R(P_X, D_1)$ and $R_1^* + R_2^* = R(P_X, D_2)$,

$$\begin{aligned}
&\Big\{(L_1,L_2):\ L_1 \ge \sqrt{\mathrm{V}(P_X,D_1)}\mathrm{Q}^{-1}(\min\{\varepsilon_1,\varepsilon_2\}),\\
&\qquad L_2 \ge \sqrt{\mathrm{V}(P_X,D_2)}\mathrm{Q}^{-1}(\min\{\varepsilon_1,\varepsilon_2\})\Big\}\\
&\subseteq \mathcal{L}_{\mathrm{sep}}(R_1^*,R_2^*,D_1,D_2,\varepsilon_1,\varepsilon_2)\\
&\subseteq \Big\{(L_1,L_2):\ L_1 \ge \sqrt{\mathrm{V}(P_X,D_1)}\mathrm{Q}^{-1}(\varepsilon_1),\\
&\qquad L_2 \ge \sqrt{\mathrm{V}(P_X,D_2)}\mathrm{Q}^{-1}(\varepsilon_2)\Big\}. \qquad (10.43)
\end{aligned}$$

The converse part also follows from the converse proof of second-order asymptotics for the rate-distortion problem in Theorem 3.5. Corollary 10.7 implies that when $\varepsilon_1 = \varepsilon_2$, for successively refinable DMS, under the SEP criterion, the second-order coding rates are also successively refinable since the pair $L_1 = \sqrt{\mathrm{V}(P_X,D_1)}\mathrm{Q}^{-1}(\varepsilon_1)$ and $L_2 = \sqrt{\mathrm{V}(P_X,D_2)}\mathrm{Q}^{-1}(\varepsilon_2)$ is second-order achievable for the boundary rate pair $(R_1^*,R_2^*) = (R(P_X,D_1),R(P_X,D_2))$. Such a result implies that it is optimal to interrupt a transmission to provide a finer reconstruction of the source sequence without any loss in terms of second-order asymptotics, which is stronger than the original definition of successively refinability in terms of first-order asymptotics and coined "strong successive refinability" in [82].

10.3.4 Numerical Examples

Recall that any discrete memoryless source with Hamming distortion measures is successively refinable [31]. In this subsection, we consider two such numerical examples originated in [70] to illustrate Corollary 10.6. We use the logarithm with base 2 in this subsection.

A Binary Memoryless Source with Hamming Distortion Measures

Fix $p \in [0,1]$. We consider a binary source with $P_X(0) = p$. For any distortion levels $D_2 < D_1 < p$, it follows from (3.28) that for each $i \in [2]$,

$$\jmath(x|D_i,P_X) = \imath(x|P_X) - H_{\mathrm{b}}(D_i). \qquad (10.44)$$

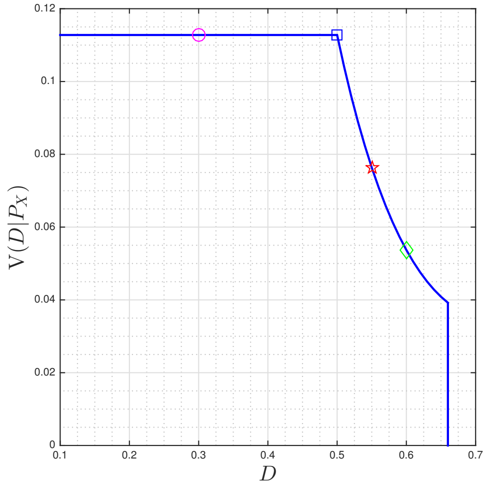

Figure 10.2: Rate-dispersion function $\mathrm{V}(P_X, D)$ for the source $P_X = [1/3, 1/4, 1/4, 1/6]$ [70, Section VII.B] as a function of the distortion D.

Hence,

$$\mathrm{V}(P_X, D_1) = \mathrm{V}(P_X, D_2) = p(1-p)\log^2\left(\frac{1-p}{p}\right), \tag{10.45}$$

and the rate-dispersion matrix is

$$\mathbf{V}(P_X, D_1, D_2) = \mathrm{V}(P_X, D_1) \cdot \mathrm{ones}(2,2) \tag{10.46}$$

$$= p(1-p)\log^2\left(\frac{1-p}{p}\right) \cdot \mathrm{ones}(2,2), \tag{10.47}$$

which does not depend on (D_1, D_2). From the above considerations, we see that a binary source with Hamming distortion measures is an example that falls under (10.42) in Corollary 10.6.

A Quaternary Memoryless Source with Hamming Distortion Measures

We next consider a more interesting source with the joint excess-distortion probability upper bounded by $\varepsilon = 0.005$. In particular, we consider a quaternary memoryless source with distribution $P_X = [1/3, 1/4, 1/4, 1/6]$. This example illustrates Case (iii) of Corollary 10.6

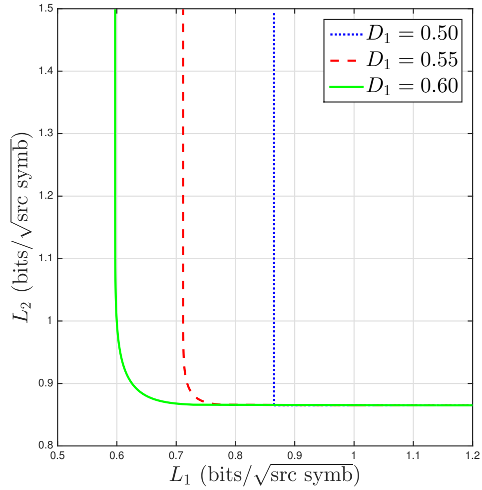

Figure 10.3: Boundaries of the second-order coding region $\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)$ for Case (iii) in Corollary 10.6. The regions are to the top right of the boundaries.

and is adopted from [70, Section VII.B]. The expressions for the rate-distortion function and the distortion-tilted information density are given in [70, Section VII.B] (and will not be reproduced here as they are not important for our discussion). Since $\jmath(x|D_1, P_X) = \jmath(x|D_2, P_X)$ when $D_1 = D_2 = D$, we use $\jmath(x|D, P_X)$ to denote the common value of the distortion-tilted information density. Similarly, let $\mathrm{V}(P_X, D)$ be the common value of $\mathrm{V}(P_X, D_1)$ and $\mathrm{V}(P_X, D_2)$ when $D_1 = D_2 = D$. As shown in Figure 10.2 (reproduced from [70, Section VII.B, Figure 4]), the rate-dispersion function $\mathrm{V}(P_X, D)$ is dependent on the distortion level D, unlike the binary example in Section 10.3.4.

In this numerical example, we fix $D_2 = 0.3$, which is denoted by the circle in Figure 10.2. Then we decrease D_1 from 0.6 to 0.55 and finally to 0.5. These points are denoted respectively by the diamond, the pentagram and the square in Figure 10.2. Given these values of (D_1, D_2), we plot the second-order coding rate for Case (iii) of Corollary 10.6 in Figure 10.3.

From Figure 10.3, we make the following observations and conclusions.

- The minimum L_1 converges to $\sqrt{V(P_X, D_1)}\mathrm{Q}^{-1}(\varepsilon)$ as $L_2 \uparrow \infty$. This is because as L_2 increases, the bivariate Gaussian cdf asymptotically degenerates to the univariate Gaussian cdf with mean 0 and variance $\mathrm{V}(P_X, D_1)$. A similar observation was made for the Slepian-Wolf problem in [113].

- As we decrease the value of D_1, the second-order coding region shrinks. We remark that there is a transition from (10.41) with $\mathrm{rank}(\mathbf{V}(D_1|P_X, D_2)) = 2$ to (10.42) (where $\mathrm{rank}(\mathbf{V}(D_1|P_X, D_2)) = 1$) as we decrease D_1 with the critical value of D_1 being 0.5.

- When $D_2 < D_1 \leq 0.5$, the rate-dispersion matrix $\mathbf{V}(P_X, D_1, D_2)$ is rank 1 (and proportional to the all ones matrix). Correspondingly, the result in (10.42) applies. Here, the second-order region is a (unbounded) rectangle with a sharp corner at the left bottom since the smaller $L_i,\ i = 1, 2$ dominates. The second-order region remains unchanged as we decrease D_1 towards D_2 for fixed $D_2 = 0.3$.

- When $0.5 < D_1 < 2/3$, the result in (10.41) applies. In this case, neither L_1 nor L_2 dominates. The second-order coding rates (L_1, L_2) are coupled together by the full rank rate-dispersion matrix $\mathbf{V}(P_X, D_1, D_2)$, resulting the smooth boundary at the left bottom.

We conclude that depending on the value of the distortion levels, the rate-dispersion matrix is either rank 1 or rank 2, illustrating Case (iii) of Corollary 10.6. These interesting observations cannot be gleaned from the work of No *et al.* [82] in which the separate excess-distortion criteria are employed for the successive refinement problem. When $\mathbf{V}(D_1|P_X, D_2)$ is rank 1, exactly one excess-distortion event dominates the probability in (10.5) entirely; when $\mathbf{V}(D_1|P_X, D_2)$ is rank 2, both excess-distortion events contribute non-trivially to the probability and a bivariate Gaussian is required to characterize the second-order fundamental limit.

10.4 Proof of Second-Order Asymptotics

10.4.1 Achievability

We make use of the type covering lemma [82, Lemma 8], which is modified from [55, Lemma 1]. Leveraging the type covering lemma, we can then upper bound the excess-distortion probability. Finally, we Taylor expand appropriate terms and invoke the Berry-Essen theorem to obtain an achievable second-order coding region.

Define two constants

$$c_1 := 4|\mathcal{X}||\hat{\mathcal{X}}_1| + 9, \tag{10.48}$$

$$c_2 := 6|\mathcal{X}||\hat{\mathcal{X}}_1||\hat{\mathcal{X}}_2| + 2|\mathcal{X}||\hat{\mathcal{X}}_1| + 17. \tag{10.49}$$

We are now ready to recall the *discrete* type covering lemma for successive refinement.

Lemma 10.8. Given type $Q_X \in \mathcal{P}_n(\mathcal{X})$, for all $R_1 \geq R(Q_X, D_1)$, the following holds:

- There exists a set $\mathcal{B}_1 \subset \hat{\mathcal{X}}_1^n$ such that

$$\frac{1}{n}\log|\mathcal{B}_1| \leq R_1 + c_1\frac{\log n}{n} \tag{10.50}$$

  and the type class is D_1-covered by the set $\mathcal{B}_1$, i.e.,

$$\mathcal{T}_{Q_X} \subset \bigcup_{\hat{x}_1^n \in \mathcal{B}_1} \{x^n : d_1(x^n, \hat{x}_1^n) \leq D_1\}. \tag{10.51}$$

- For each $x^n \in \mathcal{T}_Q$ and each $\hat{x}_1^n \in \mathcal{B}_1$, there exists a set $\mathcal{B}_2(\hat{x}_1^n) \subset \hat{\mathcal{X}}_2^n$ such that

$$\frac{1}{n}\log\left(\sum_{\hat{x}_1^n \in \mathcal{B}} |\mathcal{B}_2(\hat{x}_1^n)|\right) \leq \mathsf{R}(R_1|Q_X, D_1, D_2) + c_2\frac{\log n}{n} \tag{10.52}$$

  and the D_1-distortion ball $\mathcal{N}_1(\hat{x}_1^n, D_1) := \{x^n : d_1(x^n, \hat{x}_1^n) \leq D_1\}$ is D_2-covered by the set $\mathcal{B}_2(\hat{x}_1^n)$ i.e.,

$$\mathcal{N}_1(\hat{x}_1^n, D_1) \subset \bigcup_{\hat{x}_2^n \in \mathcal{B}_2(\hat{x}_1^n)} \{x^n : d_2(x^n, \hat{x}_2^n) \leq D_2\}. \tag{10.53}$$

Invoking Lemma 10.8, we can then upper bound the excess-distortion probability for some (n, M_1, M_2)-code. Given any (n, M_1, M_2)-code, define

$$R_{1,n} := \frac{1}{n}\bigg(\log M_1 - c_1 \log n - |\mathcal{X}| \log(n+1)\bigg), \tag{10.54}$$

$$R_{2,n} := \frac{1}{n}\bigg(\log(M_1 M_2) - c_2 \log n\bigg) - R_{1,n}. \tag{10.55}$$

Lemma 10.9. There exists an (n, M_1, M_2)-code such that

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \le \Pr\Big\{R_{1,n} < R(\hat{T}_{X^n}, D_1) \text{ or } R_{1,n} + R_{2,n} < \mathsf{R}(R_{1,n}, D_1, D_2|\hat{T}_{X^n})\Big\}. \tag{10.56}$$

The proof of Lemma 10.9 is similar to [127, Lemma 5] and available in [151, Appendix D].

Recall the definition of the typical set in (3.87) and the result in (3.88) that

$$\Pr\Big\{\hat{T}_{X^n} \notin \mathcal{A}_n(P_X)\Big\} \le \frac{2|\mathcal{X}|}{n^2}. \tag{10.57}$$

For a rate pair (R_1^*, R_2^*) satisfying the conditions in Theorem 10.4, we choose

$$\frac{1}{n}\log M_1 = R_1^* + \frac{L_1}{\sqrt{n}} + \frac{c_1 \log n + |\mathcal{X}| \log(n+1)}{n}, \tag{10.58}$$

$$\frac{1}{n}\log(M_1 M_2) = R_1^* + R_2^* + \frac{L_2}{\sqrt{n}} + c_2\frac{\log n}{n}. \tag{10.59}$$

Hence,

$$R_{1,n} = R_1^* + \frac{L_1}{\sqrt{n}}, \tag{10.60}$$

$$R_{1,n} + R_{2,n} = R_1^* + R_2^* + \frac{L_2}{\sqrt{n}}. \tag{10.61}$$

From the conditions in Theorem 10.4, we know that the second derivative of $R(Q_X, D_1)$ is bounded in the neighborhood of P_X, and that the second derivative of $\mathsf{R}(R_1|Q_X, D_1, D_2)$ with respect to (R_1, R_2, Q_X) is bounded around a neighborhood of (R_1^*, P_X). Hence, for any x^n such that $\hat{T}_{x^n} \in \mathcal{A}_n(P_X)$, applying Taylor's expansion and invoking Lemmas 3.2 and 10.3, we obtain

$$\begin{aligned}
&R(\hat{T}_{x^n}, D_1) \\
&= R(P_X, D_1) + \sum_x \left(\hat{T}_{x^n}(x) - P_X(x)\right) \jmath(x|D_1, P_X) + O\left(\frac{\log n}{n}\right) && (10.62)\\
&= \frac{1}{n}\sum_{i\in[n]} \jmath(x_i|D_1, P_X) + O\left(\frac{\log n}{n}\right), && (10.63)
\end{aligned}$$

and

$$\begin{aligned}
&\mathsf{R}(R_{1,n}, D_1, D_2|\hat{T}_{x^n}) \\
&= \mathsf{R}(R_1^*, D_1, D_2|P_{XY}) - \xi^* \frac{L_1}{\sqrt{n}} + O\left(\frac{\log n}{n}\right) \\
&\quad + \sum_x \left(\hat{T}_{x^n}(x) - P_X(x)\right) \jmath(x, R_1^*|D_1, D_2, P_X) && (10.64)\\
&= \frac{1}{n}\sum_{i\in[n]} \jmath(x_i, R_1^*|D_1, D_2, P_X) - \xi^* \frac{L_1}{\sqrt{n}} + O\left(\frac{\log n}{n}\right). && (10.65)
\end{aligned}$$

Define $\eta_n = \frac{\log n}{n}$.

In subsequent analyses, for ease of notation, we use $\jmath(x, R_1^*)$ and $\jmath(x, R_1^*|D_1, D_2, P_X)$ interchangeably. It follows from Lemma 10.9 that

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\le \Pr\left\{R_{1,n} < R(\hat{T}_{X^n}, D_1) \text{ or } R_{1,n} + R_{2,n} < \mathsf{R}(R_{1,n}, D_1, D_2|\hat{T}_{X^n})\right\} && (10.66)\\
&\le \Pr\Big\{R_{1,n} < R(\hat{T}_{X^n}, D_1) \text{ or } R_{1,n} + R_{2,n} < \mathsf{R}(R_{1,n}, D_1, D_2|\hat{T}_{X^n}), \\
&\qquad \text{and } \hat{T}_{X^n} \in \mathcal{A}_n(P_X)\Big\} + \Pr\left\{\hat{T}_{X^n} \notin \mathcal{A}_n(P_X)\right\} && (10.67)
\end{aligned}$$

$$\leq \Pr\left\{R_1^* + \frac{L_1}{\sqrt{n}} < \frac{1}{n}\sum_{i\in[n]} \jmath(X_i|D_1,P_X) + O(\eta_n) \text{ or } R_1^*+R_2^*+\frac{L_2}{\sqrt{n}} < \frac{1}{n}\sum_{i\in[n]}\jmath(X_i,R_1^*) - \xi^*\frac{L_1}{\sqrt{n}} + O(\eta_n)\right\} + \frac{2|\mathcal{X}|}{n^2} \tag{10.68}$$

$$= \Pr\left\{R_1^* + \frac{L_1}{\sqrt{n}} < \frac{1}{n}\sum_{i\in[n]} \jmath(X_i|D_1,P_X) + O(\eta_n) \text{ or } R_1^*+R_2^*+\xi^*\frac{L_1}{\sqrt{n}}+\frac{L_2}{\sqrt{n}} < \frac{1}{n}\sum_{i\in[n]}\jmath(X_i,R_1^*) + O(\eta_n)\right\} + \frac{2|\mathcal{X}|}{n^2}. \tag{10.69}$$

Thus,

$$1-\mathrm{P}_{\mathrm{e},n}(D_1,D_2) \geq \Pr\left\{\frac{1}{n}\sum_{i\in[n]}\jmath(X_i|D_1,P_X) \leq R_1^* + \frac{L_1}{\sqrt{n}} + O(\eta_n), \frac{1}{n}\sum_{i\in[n]}\jmath(X_i,R_1^*) \leq R_1^*+R_2^*+\xi^*\frac{L_1}{\sqrt{n}}+\frac{L_2}{\sqrt{n}}+O(\eta_n)\right\} - \frac{2|\mathcal{X}|}{n^2}. \tag{10.70}$$

We first consider Case (i) where $R(P_X,D_1) < R_1^* < \mathsf{R}(R_1^*|P_X,D_1,D_2)$ and $R_1^*+R_2^* = \mathsf{R}(R_1^*|P_X,D_1,D_2)$. Using the weak law of large numbers in Theorem 1.1, we obtain

$$\Pr\left\{\frac{1}{n}\sum_{i\in[n]}\jmath(X_i|D_1,P_X) \leq R_1^* + \frac{L_1}{\sqrt{n}} + O(\eta_n)\right\} \to 1. \tag{10.71}$$

Invoking the Berry-Esseen Theorem in Theorem 1.3, we obtain

$$\Pr\left\{\frac{1}{n}\sum_{i\in[n]}\jmath(X_i,R_1^*) \leq R_1^*+R_2^*+\xi^*\frac{L_1}{\sqrt{n}}+\frac{L_2}{\sqrt{n}}+O(\eta_n)\right\} \geq 1-\mathrm{Q}\left(\frac{\xi^*L_1+L_2+O(\sqrt{n}\eta_n)}{\sqrt{\mathrm{V}(R_1^*|P_X,D_1,D_2)}}\right) - \frac{6\mathrm{T}(R_1^*|P_X,D_1,D_2)}{\sqrt{n}\mathrm{V}^{3/2}(R_1^*|P_X,D_1,D_2)}, \tag{10.72}$$

where $\mathrm{T}(R_1^*|P_X, D_1, D_2)$ is the third absolute moment of $\jmath(X, R_1^*|D_1, D_2, P_X)$, which is finite for DMS. Hence,

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1, D_2)\\
&\le \mathrm{Q}\left(\frac{\xi^* L_1 + L_2 + O(\sqrt{n}\eta_n)}{\sqrt{\mathrm{V}(R_1^*|P_X, D_1, D_2)}}\right) + \frac{6\mathrm{T}(R_1^*|P_X, D_1, D_2)}{\sqrt{n}\mathrm{V}^{3/2}(R_1^*|P_X, D_1, D_2)}\\
&\quad + \frac{2|\mathcal{X}|}{n^2}.
\end{aligned} \tag{10.73}$$

Hence, if (L_1, L_2) satisfies

$$\xi^* L_1 + L_2 \ge \sqrt{\mathrm{V}(R_1^*|P_X, D_1, D_2)}\mathrm{Q}^{-1}(\varepsilon), \tag{10.74}$$

then $\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \le \varepsilon$. The proof of Case (ii) is omitted since it is similar to Case (i).

The most interesting case is Case (iii) where $R_1^* = R(P_X, D_1)$ and $R_1^* + R_2^* = \mathsf{R}(R_1^*|P_X, D_1, D_2)$. If $\mathbf{V}(R_1^*|P_X, D_1, D_2)$ is positive definite we invoke the multi-variate Berry-Esseen Theorem in Theorem 1.5 to obtain

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1, D_2)\\
&\le 1 - \Psi\left(L_1 + O(\eta_n), \xi^* L_1 + L_2 + O(\eta_n); \mathbf{0}, \mathbf{V}(R_1^*|P_X, D_1, D_2)\right)\\
&\quad + O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned} \tag{10.75}$$

Note that if $\mathbf{V}(R_1^*|P_X, D_1, D_2)$ is rank 1, we can use the argument (projection onto a lower-dimensional subspace) in [113, Proof of Theorem 6] to conclude that (10.75) also holds. Now if we choose (L_1, L_2) such that

$$\Psi\left(L_1, \xi^* L_1 + L_2; \mathbf{0}, \mathbf{V}(R_1^*|P_X, D_1, D_2)\right) \ge 1 - \varepsilon, \tag{10.76}$$

then $\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \le \varepsilon$. The achievability proof is now completed.

10.4.2 Converse

We first prove a type-based strong converse. Define $\overline{d}_i := \max_{x,y} d_1(x, \hat{x}_i)$ for each $i \in [2]$. Given a type $Q_X \in \mathcal{P}_n(\mathcal{X})$, define

$$g(Q_X) := \Pr\left\{d_1(X^n, \hat{X}_1^n) \leq D_1, \text{ and } d_2(X^n, \hat{X}_2^n) \leq D_2 \,\middle|\, X^n \in \mathcal{T}_{Q_X}\right\}. \tag{10.77}$$

Lemma 10.10. Fix $\alpha > 0$ and a type $Q_X \in \mathcal{P}_n(\mathcal{X})$. If the excess-distortion probability satisfies

$$g(Q_X) \geq \exp(-n\alpha), \tag{10.78}$$

then there exists a conditional distribution $Q_{\hat{X}_1\hat{X}_2|X}$ such that

$$\log M_1 \geq nI(Q_X, Q_{\hat{X}_1|X}) - \vartheta_n, \tag{10.79}$$

$$\log(M_1M_2) \geq nI(Q_X, Q_{\hat{X}_1\hat{X}_2|X}) - \vartheta_n, \tag{10.80}$$

where $\vartheta_n := |\mathcal{X}|\log(n+1) + \log n + n\alpha$, and the expected distortions are bounded as

$$\mathbb{E}_{Q_X \times Q_{\hat{X}_1\hat{X}_2|X}}[d_1(X, \hat{X}_1)] \leq D_1 + \frac{\overline{d}_1}{n} =: D_{1,n}, \tag{10.81}$$

$$\mathbb{E}_{Q_X \times Q_{\hat{X}_1\hat{X}_2|X}}[d_2(X, \hat{X}_2)] \leq D_2 + \frac{\overline{d}_2}{n} =: D_{2,n}. \tag{10.82}$$

The proof of Lemma 10.10 is inspired by [45], which generalizes Lemma 3.8 for the rate-distortion problem and is available in [151, Appendix E].

Invoking Lemma 10.10 with $\alpha = \frac{\log n}{n}$, we can lower bound the excess-distortion probability for any (n, M_1, M_2)-code. Define $\beta_n = |\mathcal{X}|\log(n+1) + 2\log n$. Define

$$R'_{1,n} := \frac{1}{n}\log M_1 + \beta_n, \tag{10.83}$$

$$R'_{2,n} := \frac{1}{n}\log(M_1M_2) + \beta_n - R'_{1,n}. \tag{10.84}$$

Lemma 10.11. For any (n, M_1, M_2)-code, we have

$$\begin{aligned}&\mathrm{P}_{\mathrm{e},n}(D_1, D_2)\\&\geq \Pr\left\{R'_{1,n} < R(\hat{T}_{X^n}, D_{1,n}) \text{ or } R'_{1,n} + R'_{2,n} < \mathsf{R}(R_{1,n}, D_{1,n}, D_{2,n}|\hat{T}_{X^n})\right\}\\&\quad -\frac{1}{n}.\end{aligned} \tag{10.85}$$

Choose $\log M_1 = nR_1^* + L_1\sqrt{n} + \beta_n$ and $\log(M_1M_2) = n(R_1^* + R_2^*) + L_2\sqrt{n} + \beta_n$. Recall the shorthand notation $\eta_n := \frac{\log n}{n}$. Now for x^n such that $\hat{T}_{x^n} \in \mathcal{A}_n(P_X)$, applying Taylor's expansion in a similar manner as (10.63) and (10.65), invoking Lemma 10.11 and noting that $\Pr\{\mathcal{F} \cap \mathcal{G}\} \geq \Pr\{\mathcal{F}\} - \Pr\{\mathcal{G}^{\mathrm{c}}\}$, we obtain

$$\begin{aligned}&1 - \mathrm{P}_{\mathrm{e},n}(D_1, D_2)\\&\leq \Pr\left\{\frac{1}{n}\sum_{i\in[n]} \jmath(X_i|D_1, P_X) \leq R_1^* + \frac{L_1}{\sqrt{n}} + O(\eta_n),\right.\\&\qquad\left.\frac{1}{n}\sum_{i\in[n]} \jmath(X_i, R_1^*) \leq R_1^* + R_2^* + \xi^*\frac{L_1}{\sqrt{n}} + \frac{L_2}{\sqrt{n}} + O(\eta_n)\right\}\\&\quad + \frac{1}{n} + \frac{2|\mathcal{X}|}{n^2}.\end{aligned} \tag{10.86}$$

Note that in (10.86), we Taylor expand $R(\hat{T}_{X^n}, D_{1,n})$ around the source distribution P_X and distortion level D_1. We also Taylor expand the minimal sum rate function $\mathsf{R}(R_{1,n}, D_1, D_2|\hat{T}_{X^n})$ at (P_X, D_1, D_2). The residual terms when we Taylor expand with respect to the distortion levels are of the order $O(\frac{1}{n})$, which can be absorbed into $O(\eta_n)$. Furthermore, recall that we use $\jmath(x, R_1^*)$ and $\jmath(x, R_1^*|D_1, D_2, P_X)$ interchangeably.

The rest of converse proof can be done similarly as the achievability part in Section 10.4.1 by using the uni- or multi-variate Berry-Esseen Theorem for Cases (i), (ii) and (iii).

10.4.3 Proof of a Special Case

We now present a proof for the special case where the source-distortion measure triplet is successively refinable. Recall that for this case, $\xi^* = 0$, $\nu_1^* = 0$, and $\jmath(x, R_1^*|D_1, D_2, P_X) = \jmath(x|D_2, P_X)$ for $R(P_X, D_1) \leq R_1^* < R(P_X, D_2)$. For the achievability part, invoking (10.70), we obtain

$$
\begin{aligned}
&1 - \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\geq \Pr\left\{\frac{1}{n}\sum_{i\in[n]}(\jmath(X_i|D_1,P_X) - R_1^*) \leq \frac{L_1}{\sqrt{n}} + O(\eta_n),\right. \\
&\qquad \left.\frac{1}{n}\sum_{i\in[n]}(\jmath(X_i|D_2,P_X) - (R_1^* + R_2^*)) \leq \frac{L_2}{\sqrt{n}} + O(\eta_n)\right\} \\
&\quad - \frac{2|\mathcal{X}|}{n^2}. \qquad (10.87)
\end{aligned}
$$

According to the assumption in (10.42) of Corollary 10.6, we have $\jmath(X_i|D_1,P_X) - R_1^* = \jmath(X_i|D_2,P_X) - (R_1^* + R_2^*)$. Given a random variable X and two real numbers $a < b$, we obtain $\Pr\{X < a \text{ and } X < b\} = \Pr\{X < a\}$. Hence,

$$
\begin{aligned}
&1 - \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\geq \Pr\left\{\frac{1}{n}\sum_{i\in[n]}(\jmath(X_i|D_1,P_X) - R_1^*) \leq \frac{\min\{L_1, L_2\}}{\sqrt{n}} + O(\eta_n)\right\}.
\end{aligned} \qquad (10.88)
$$

The rest of the proof is similar to Case (i) in Section 10.4.1.

Using (10.86), similar to the achievability part, we complete the proof of converse part.

10.4.4 Alternative Converse Proof

We next present an alternative converse proof of Corollary 10.6 using the finite blocklength converse bound in [151, Lemma 15] that generalizes Theorem 3.4 for the rate-distortion problem.

Lemma 10.12. Given any $(\gamma_1, \gamma_2) \in \mathbb{R}_+^2$, any (n, M_1, M_2)-code for the successive refinement problem satisfies

$$
\begin{aligned}
\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \geq \Pr\Big\{&\sum_{i\in[n]}\jmath(X_i|D_1,P_X) \geq \log M_1 + \gamma_1 \text{ or} \\
&\sum_{i\in[n]}\jmath(X_i|D_2,P_X) \geq \log(M_1M_2) + \gamma_2\Big\} \\
&- \exp(-n\gamma_1) - \exp(-n\gamma_2). \qquad (10.89)
\end{aligned}
$$

Choose $\gamma_1 = \gamma_2 = \frac{\log n}{2n}$. Let $\log M_1 = nR_1^* + L_1\sqrt{n} - \frac{1}{2}\log n$ and $\log(M_1 M_2) = n(R_1^* + R_2^*) + L_2\sqrt{n} - \frac{1}{2}\log n$. Invoking Lemma 10.12, we obtain

$$\begin{aligned}
&1 - \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\leq \frac{2}{\sqrt{n}} + \Pr\Big\{ \sum_{i\in[n]} \jmath(X_i|D_1, P_X) < nR_1^* + L_1\sqrt{n} \\
&\qquad \text{and} \sum_{i\in[n]} \jmath(X_i|D_2, P_X) < n(R_1^* + R_2^*) + L_2\sqrt{n} \Big\}. \qquad (10.90)
\end{aligned}$$

The rest of the proof is similar to the converse proof of Corollary 10.6 in Section 10.4.3. We remark that this alternative converse proof also applies to continuous memoryless sources, such as GMS under quadratic distortion measures and a Laplacian source with absolute distortion measures [146].

A stronger non-asymptotic converse bound is provided in [68, Corollary 2], which holds for any memoryless source and yields an alternative converse proof of Theorem 10.4. The same bound is also presented in Lemma 11.6 in the next section, which is obtained as a special case of the non-asymptotic converse bound in Theorem 11.5 for the Fu-Yeung problem.

11

Fu-Yeung Problem

In this section, we study a special case of the multiple descriptions problem [1], [37], [122], [123], [132], [143], [144] with two encoders and three decoders proposed by Fu and Yeung [34] and thus we term the problem as the Fu-Yeung problem. The Fu-Yeung problem generalizes the successive refinement problem by adding an additional decoder that aims to recover a deterministic function of the source sequence losslessly. The rate-distortion region was characterized by Fu and Yeung [34, Theorem 1], which collects rate pairs to ensure reliable lossy compression at two decoders and reliable lossless data compression at the other decoder. For this special case of multiple descriptions, the El Gamal-Cover inner bound [37] was proved optimal.

Through the lens of the Fu-Yeung problem, this section reveals the tradeoff between encoders for simultaneous lossless and lossy compression. We will present a non-asymptotic converse bound and second-order asymptotics for the Fu-Yeung problem. Specifically, we first present properties of the minimal sum rate function given the rate of one encoder. Subsequently, we generalize the rate-distortions-tilted information for the successive refinement problem to the Fu-Yeung problem and present a non-asymptotic converse bound. This non-asymptotic bound, when specialized to the case where $|\mathcal{Y}| = 1$, gives a stronger non-asymptotic converse bound for the successive refinement problem than Lemma 10.12. Finally, we present the second-order asymptotics for DMS under bounded distortion measures and illustrate the results with numerical examples. This section is largely based on [149] and the second part of [147].

11.1 Problem Formulation and Asymptotic Result

11.1.1 Problem Formulation

The setting for the Fu-Yeung problem is shown in Figure 11.1. There are two encoders and three decoders. Each encoder $f_i,\ i = 1,2$ has access to the source sequence X^n and compresses it into a message $S_i,\ i = 1,2$. Decoder ϕ_1 aims to recover X^n with distortion level D_1 using the encoded message S_1 from encoder f_1. Decoder ϕ_2 aims to recover X^n with distortion level D_2 using encoded messages S_1 and S_2. Decoder ϕ_3 aims to recover Y^n, which is a symbolwise deterministic function of the source sequence X^n.

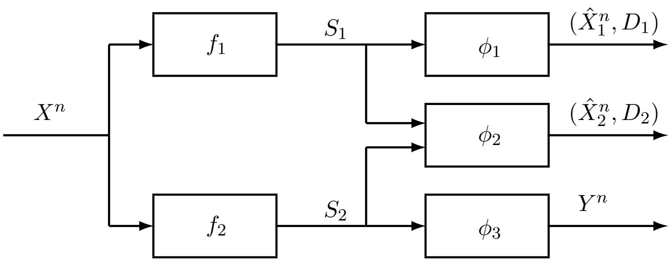

Figure 11.1: System model for the Fu-Yeung problem of multiple descriptions with one Semi-deterministic decoder [34].

Consider a memoryless source X^n generated i.i.d. from a probability mass function P_X supported on a finite alphabet $\mathcal{X}$. Let reproduction alphabets for decoders ϕ_1, ϕ_2 be $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$ respectively. Fix a finite set $\mathcal{Y}$ and define a deterministic function $g : \mathcal{X} \to \mathcal{Y}$. Let $Y_i = g(X_i),\ i \in [1:n]$. Note that P_Y is induced by the source distribution P_X and the deterministic function g, i.e., for $y \in \mathcal{Y}$, $P_Y(y) = \sum_{x:g(x)=y} P_X(x)$. We assume that for each y, $P_Y(y) > 0$. Decoder ϕ_3 is required to recover $Y^n = g(X^n) = (g(X_1), \ldots, g(X_n))$ losslessly and the decoded sequence is denoted as $\hat{Y}^n$. We follow the definitions of codes and the rate-distortion region in [34].

Definition 11.1. An (n, M_1, M_2)-code for the Fu-Yeung problem consists of two encoders:

$$f_1 : \mathcal{X}^n \to \mathcal{M}_1 = [M_1], \tag{11.1}$$
$$f_2 : \mathcal{X}^n \to \mathcal{M}_2 = [M_2], \tag{11.2}$$

and three decoders:

$$\phi_1 : \mathcal{M}_1 \to \hat{\mathcal{X}}_1^n, \tag{11.3}$$
$$\phi_2 : \mathcal{M}_1 \times \mathcal{M}_2 \to \hat{\mathcal{X}}_2^n, \tag{11.4}$$
$$\phi_3 : \mathcal{M}_2 \to \mathcal{Y}^n. \tag{11.5}$$

Using the encoding and decoding functions, we have $\hat{X}_1^n = \phi_1(f_1(X^n))$, $\hat{X}_2^n = \phi_2(f_1(X^n), f_2(X^n))$ and $\hat{Y}^n = \phi_3(f_2(X^n))$. Let d_{H} denote the Hamming distortion measure in (3.1) and let the average distortion between y^n and its reproduced version $\hat{y}^n$ be defined as $d_{\mathrm{H}}(Y^n, \hat{Y}^n) := \frac{1}{n}\sum_{i\in[n]} d_{\mathrm{H}}(Y_i, \hat{Y}_i)$. For each $i \in [2]$, let the distortion function $d_i : \mathcal{X} \times \hat{\mathcal{X}}_i \to [0, \infty)$ be a bounded distortion measure and let $d_i(x^n, \hat{x}_i^n) = \frac{1}{n}\sum_{j\in[n]} d_i(x_j, \hat{x}_{j,i})$. The rate-distortion region for the Fu-Yeung problem is defined as follows.

Definition 11.2. A rate pair (R_1, R_2) is said to be (D_1, D_2)-achievable for the Fu-Yeung problem if there exists a sequence of (n, M_1, M_2)-codes such that

$$\limsup_{n\to\infty} \frac{\log M_i}{n} \le R_i, \ i = 1, 2, \tag{11.6}$$

and

$$\limsup_{n\to\infty} \mathsf{E}[d_i(X^n, \hat{X}_i^n)] \le D_i, \ i = 1, 2, \tag{11.7}$$
$$\lim_{n\to\infty} \mathsf{E}[d_{\mathrm{H}}(Y^n, \hat{Y}^n)] = 0. \tag{11.8}$$

The closure of the set of all (D_1, D_2)-achievable rate pairs is called the first-order (D_1, D_2)-coding region and denoted as $\mathcal{R}(D_1, D_2|P_X)$.

11.1.2 Rate-Distortion Region

The first-order coding region $\mathcal{R}(D_1, D_2|P_X)$ was characterized by Fu and Yeung in [34] for DMS. In particular, Fu and Yeung [34] showed that the El-Gamal-Cover inner bound [37] for the multiple description coding problem is tight.

To present the result, let $\mathcal{P}(P_X, D_1, D_2)$ be the set of all pairs of conditional distributions $(P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1}) \in \mathcal{P}(\hat{\mathcal{X}}_1|X) \times \mathcal{P}(\hat{\mathcal{X}}_2|\mathcal{X}\hat{\mathcal{X}}_1)$ such that $\mathsf{E}[d_1(X, \hat{X}_1)] \leq D_1$ and $\mathsf{E}[d_2(X, \hat{X}_2)] \leq D_2$. Given a pair of conditional distributions $(P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1})$, let $\mathcal{R}(P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1})$ be the collection of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ such that

$$R_1 \geq I(P_X, P_{X|\hat{X}_1}), \tag{11.9}$$

$$R_2 \geq H(P_Y), \tag{11.10}$$

$$R_1 + R_2 \geq H(P_Y) + I(P_{\hat{X}_1}, P_{\hat{X}_1|Y}) + I(P_{X|Y}, P_{X|Y\hat{X}_1\hat{X}_2}|P_Y). \tag{11.11}$$

Theorem 11.1. The rate-distortion region for the Fu-Yeung problem satisfies

$$\mathcal{R}(D_1, D_2|P_X) = \bigcup_{(P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1}) \in \mathcal{P}(P_X, D_1, D_2)} \mathcal{R}(P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1}). \tag{11.12}$$

When the Y is a constant, i.e., $|\mathcal{Y}| = 1$, the rate-distortion region in Theorem 11.1 reduced to the rate-distortion region of the successive refinement problem. The rate-distortion function of the Kaspi problem can also be recovered from Theorem 11.1 as the minimal rate R_1 by setting $R_2 = H(Y)$ and choosing the source as $X = (S_1, S_2)$ and the side information as $Y = S_2$ for correlated discrete random variables (S_1, S_2).

Although Theorem 11.1 was derived under the average distortion criterion, the same rate-distortion region holds when one considers a vanishing joint excess-distortion and error probability $\mathrm{P}_{\mathrm{e},n}(D_1, D_2)$ defined as follows:

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) := \Pr\Big\{d_1(X^n, \hat{X}_1^n) > D_1 \text{ or } d_2(X^n, \hat{X}_2^n) > D_2 \text{ or } \hat{Y}^n \neq Y^n\Big\}. \tag{11.13}$$

The reason is analogous to why Theorem 3.1 derived under the average distortion criterion still holds under the excess-distortion probability criterion for the rate-distortion problem.

11.1.3 Boundary Rate Pairs

We next discuss conditions for a rate pair (R_1^*, R_2^*) to be on the boundary of the rate-distortion region $\mathcal{R}(D_1, D_2|P_X)$, which enables our definition and analyses of second-order asymptotics.

Given any distributions $(P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1})$, let P_{XY}, $P_{X|Y}$, $P_{\hat{X}_1}$, $P_{Y\hat{X}_1}$, $P_{X\hat{X}_1}$, $P_{X\hat{X}_2}$, $P_{XY\hat{X}_1}$, $P_{\hat{X}_1|XY}$ and $P_{\hat{X}_2|Y\hat{X}_1}$ be induced by P_X, $P_{\hat{X}_1|X}$, $P_{\hat{X}_2|X\hat{X}_1}$ and the deterministic function $g : \mathcal{X} \to \mathcal{Y}$. Recall the definition of $\mathcal{P}(P_X, D_1, D_2)$ above Theorem 11.1. Given any rate R_1 of encoder f_1, define the following function:

$$\begin{aligned}&\mathsf{R}(R_1|P_X, D_1, D_2)\\ &:= \min_{\substack{(P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1})\\ \in\mathcal{P}(P_X,D_1,D_2): R_1 \geq I(P_X, P_{X|\hat{X}_1})}} I(P_{\hat{X}_1}, P_{\hat{X}_1|Y}) + I(P_{X|Y}, P_{X|Y\hat{X}_1\hat{X}_2}|P_Y).\end{aligned} \tag{11.14}$$

It follows from the rate-distortion region in Theorem 11.1 that given a rate R_1 of encoder f_1, the minimal achievable sum rate is $\mathsf{R}(R_1|P_X, D_1, D_2) + H(P_Y)$. Furthermore, the minimal achievable rate R_1 for encoder f_1 is the rate-distortion function $R(P_X, D_1)$ [19] and the minimal achievable rate R_2 for encoder f_2 is the entropy $H(P_Y)$. When $R_1 = R(P_X, D_1)$, the minimal achievable rate R_2 is

$$\begin{aligned}&\mathsf{R}_2^*(P_X, D_1, D_2)\\ &:= H(P_Y) + \mathsf{R}(R(P_X, D_1), D_1, D_2|P_X) - R(P_X, D_1),\end{aligned} \tag{11.15}$$

and when $R_2 = H(P_Y)$, the minimal achievable rate R_1 is

$$\begin{aligned}&\mathsf{R}_1^*(P_X, D_1, D_2)\\ &:= \min_{P_{\hat{X}_1|X}, P_{\hat{X}_2|X\hat{X}_1} \in \mathcal{P}(P_X,D_1,D_2)} I(P_{\hat{X}_1}, P_{\hat{X}_1|Y}) + I(P_{X|Y}, P_{X|Y\hat{X}_1\hat{X}_2}|P_Y),\end{aligned} \tag{11.16}$$

since $\mathsf{R}_1^*(P_X, D_1, D_2)$ is the solution to $R_1 = \mathsf{R}(R_1|P_X, D_1, D_2)$. With these observations, we find all cases of boundary rate pairs and illustrate

Full text available at: http://dx.doi.org/10.1561/0100000134

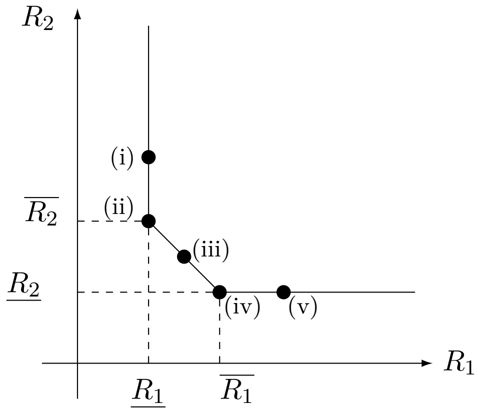

Figure 11.2: Illustration of boundary rate pairs on the rate-distortion region of the Fu-Yeung problem, where $\underline{R_1} = R(P_X, D)$, $\overline{R_1} = \mathsf{R}_1^*(P_X, D_1, D_2)$, $\underline{R_2} = H(P_Y)$ and $\overline{R_2} = \mathsf{R}_2^*(P_X, D_1, D_2)$.

it in Figure 11.2. Note that the Curve from case (ii) to Case (iv) is drawn as a line segment for ease of plot. In fact, it should be a convex curve.

11.2 Minimal Sum Rate Function and Its Properties

11.2.1 Definitions

Note that (11.14) is a convex optimization problem. Assume that (R_1, D_1, D_2) is chosen such that $\mathsf{R}(R_1|P_X, D_1, D_2)$ is finite. Therefore, there exist test channels achieving $\mathsf{R}(R_1|P_X, D_1, D_2)$. Let $(\xi^*, \lambda_1^*, \lambda_2^*)$ be the optimal solutions to the dual problem of $\mathsf{R}(R_1|P_X, D_1, D_2)$, i.e.,

$$\xi^* := -\frac{\partial \mathsf{R}(R, D_1, D_2|P_X)}{\partial R}\bigg|_{R=R_1}, \tag{11.17}$$

$$\lambda_1^* := -\frac{\partial \mathsf{R}(R_1, D, D_2|P_X)}{\partial D}\bigg|_{D=D_1}, \tag{11.18}$$

$$\lambda_2^* := -\frac{\partial \mathsf{R}(R_1, D_1, D|P_X)}{\partial D}\bigg|_{D=D_2}. \tag{11.19}$$

Given distributions $(Q_{\hat{X}_1}, Q_{\hat{X}_2|Y\hat{X}_1})$ and $(x, y, \hat{x}_1)$, define the following two functions:

$$\begin{aligned}
&\beta_2(x, y, \hat{x}_1|Q_{\hat{X}_2|Y\hat{X}_1}) \\
&:= \left\{\mathsf{E}_{Q_{\hat{X}_2|Y\hat{X}_1}}\left[\exp(-\lambda_2^* d_2(x, \hat{X}_2))|Y = y, \hat{X}_1 = \hat{x}_1\right]\right\}^{-1}, \qquad (11.20)\\
&\beta(x, y|Q_{\hat{X}_1}, Q_{\hat{X}_2|Y\hat{X}_1}) \\
&:= \left\{\mathsf{E}_{Q_{\hat{X}_1}}\left[\exp\left(-\frac{\lambda_1^* d_1(x, \hat{X}_1)}{1+\xi^*} - \frac{\log \beta_2(x, y, \hat{X}_1|Q_{\hat{X}_2|Y\hat{X}_1})}{1+\xi^*}\right)\right]\right\}^{-1}. \qquad (11.21)
\end{aligned}$$

11.2.2 Properties

We first present the properties of the optimal test channels that achieve (11.14).

Lemma 11.2. A pair of test channels $(P^*_{\hat{X}_1|X}, P^*_{\hat{X}_2|X\hat{X}_1})$ achieves $\mathsf{R}(R_1|P_X, D_1, D_2)$ if and only if

- For all $(x, y, \hat{x}_1, \hat{x}_2)$ such that $y = g(x)$,

$$\begin{aligned}
P^*_{\hat{X}_1|X}(\hat{x}_1|x) &= \beta(x, y|P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1})P^*_{\hat{X}_1}(\hat{x}_1) \\
&\quad \times \exp\left(-\frac{\lambda_1^* d_1(x, \hat{x}_1) + \log \beta_2(x, y, \hat{x}_1|P^*_{\hat{X}_2|Y\hat{X}_1})}{1+\xi^*}\right), \qquad (11.22)
\end{aligned}$$

- For all $(x, y, \hat{x}_1, \hat{x}_2)$ such that $y = g(x)$ and $P^*_{\hat{X}_1|X}(\hat{x}_1|x) > 0$

$$\begin{aligned}
P^*_{\hat{X}_2|X\hat{X}_1}(\hat{x}_2|x, \hat{x}_1) &= \beta_2(x, y, \hat{x}_1|P^*_{\hat{X}_2|Y\hat{X}_1}) \\
&\quad \times P^*_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|y, \hat{x}_1)\exp(-\lambda_2^* d_2(x, \hat{x}_2)). \qquad (11.23)
\end{aligned}$$

- For all $(x, \hat{x}_1, \hat{x}_2)$ such that $P^*_{\hat{X}_1|X}(\hat{x}_1|x) = 0$, $P^*_{\hat{X}_2|X\hat{X}_1}(\cdot|x, \hat{x}_1)$ can be arbitrary distribution.

Furthermore, if a pair of channels $(P^*_{\hat{X}_1|X}, P^*_{\hat{X}_2|X\hat{X}_1})$ achieves $\mathsf{R}(R_1, D_1, D_2)$, the following claims hold.

- The parametric representation of $\mathsf{R}(R_1|P_X, D_1, D_2)$ is

$$\begin{aligned}
&\mathsf{R}(R_1|P_X, D_1, D_2) \\
&= (1+\xi^*)\mathsf{E}_{P_{XY}}[\log \beta(X, Y|P^*_{\hat{X}_1}, P_{\hat{X}_2|Y\hat{X}^*_1})] \\
&\quad - \xi^* R_1 - \lambda_1^* D_1 - \lambda_2^* D_2.
\end{aligned} \tag{11.24}$$

- For $(x, y, \hat{x}_1, \hat{x}_2)$ such that $y = g(x)$ and $P^*_{\hat{X}_1}(\hat{x}_1) P^*_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|g(x), \hat{x}_1) > 0$,

$$\begin{aligned}
&(1+\xi^*) \log \beta(x, y|P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1}) \\
&= (1+\xi^*) \log \frac{P^*_{\hat{X}_1|X}(\hat{x}_1|x)}{P^*_{\hat{X}_1}(\hat{x}_1)} + \lambda_1^* d_1(x, \hat{x}_1) \\
&\quad + \log \frac{P^*_{\hat{X}_2|X\hat{X}_1}(\hat{x}_2|x, \hat{x}_1)}{P^*_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|y, \hat{x}_1)} + \lambda_2^* d_2(x, \hat{x}_2).
\end{aligned} \tag{11.25}$$

The proof of Lemma 11.2 is similar to [20, Lemma 1.4], [127, Lemma 3], Lemma 9.3 for the Kaspi problem and Lemma 10.3 for the successive refinement problem.

Similar to [127], we can show that, for any pair of optimal test channels $(P^*_{\hat{X}_1|X}, P^*_{\hat{X}_2|X\hat{X}_1})$, the value of $\beta(x, y|P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1})$ and $\beta_2(x, y, \hat{x}_1|P^*_{\hat{X}_2|Y\hat{X}_1})$ remain the same. From now on, fix a pair of test channels $(P^*_{\hat{X}_1|X}, P^*_{\hat{X}_2|X\hat{X}_1})$ such that that i) (11.22), (11.23) hold; ii) for any $(y, \hat{x}_1)$ such that $P^*_{Y\hat{X}_1}(y, \hat{x}_1) = 0$, the induced distribution defined as $P^*_{\hat{X}_2|Y\hat{X}_1}(\hat{x}_2|y, \hat{x}_1) := \sum_x P_X(x)\mathbb{1}(y = g(x)) P^*_{\hat{X}_2|X\hat{X}_1}(\hat{x}_2|x, \hat{x}_1)$ satisfies

$$\begin{aligned}
P^*_{\hat{X}_2|Y\hat{X}_1} = \operatorname*{arg\,sup}_{Q_{\hat{X}_2|Y\hat{X}_1}} \mathsf{E}_{P_{X|y}}\Big[&\beta(X, y)\beta_2^{-\frac{1}{1+\xi^*}}(X, y, \hat{x}_1|Q_{\hat{X}_2|Y\hat{X}_1}) \\
&\times \exp\Big(-\frac{\lambda_1^*}{1+\xi^*} d_1(X, \hat{x}_1)\Big)\Big].
\end{aligned} \tag{11.26}$$

Note that the choice of $P^*_{\hat{X}_2|X\hat{X}_1}$ satisfying (11.26) is possible since the set $\{x : g(x) = y\}$ is disjoint for each $y \in \mathcal{Y}$.

For simplicity, given any $(x, y, \hat{x}_1)$, let

$$\beta_2(x, y, \hat{x}_1) := \beta_2(x, y, \hat{x}_1 | P^*_{\hat{X}_2|Y\hat{X}_1}), \tag{11.27}$$

$$\beta(x, y) := \beta(x, y | P^*_{\hat{X}_1}, P^*_{\hat{X}_2|Y\hat{X}_1}), \tag{11.28}$$

$$\imath_1(x, y, \hat{x}_1) := \log \beta(x, y) - \frac{1}{1+\xi^*} \log \beta_2(x, y, \hat{x}_1), \tag{11.29}$$

$$\imath_2(x, y, \hat{x}_1) := \log \beta(x, y) + \frac{\xi^*}{1+\xi^*} \log \beta_2(x, y, \hat{x}_1). \tag{11.30}$$

Furthermore, given any $\hat{x}_1$ and arbitrary conditional distribution $Q_{\hat{X}_2|Y\hat{X}_1}$, define

$$w_1(\hat{x}_1) := \mathsf{E}_{P_{XY}}\left[\exp\left(\imath_1(X, Y, \hat{x}_1) - \frac{\lambda_1^* d_1(X, \hat{x}_1)}{1+\xi^*}\right)\right], \tag{11.31}$$

$$\begin{aligned} &w_2(\hat{x}_1, Q_{\hat{X}_2|Y\hat{X}_1}) \\ &:= \mathsf{E}_{P_{XY}\times Q_{\hat{X}_2|Y\hat{X}_1}}\Big[\exp\big(\imath_2(X, Y, \hat{x}_1) - \frac{\lambda_1^*}{1+\xi^*} d_1(X, \hat{x}_1) \\ &\qquad - \lambda_2^* d_2(X, \hat{X}_2)\big)\Big|\hat{X}_1 = \hat{x}_1\Big]. \end{aligned} \tag{11.32}$$

In the following, we present an important property of the quantities in (11.31) and (11.32).

Lemma 11.3. Given any $(P^*_{\hat{X}_1|XY}, P^*_{\hat{X}_2|X\hat{X}_1})$ satisfying (11.22), (11.23), and (11.26), for any $\hat{x}_1 \in \hat{\mathcal{X}}_1$ and arbitrary distribution $Q_{\hat{X}_2|Y\hat{X}_1}$,

$$w_2(\hat{x}_1, Q_{\hat{X}_2|Y\hat{X}_1}) \le w_1(\hat{x}_1) \le 1. \tag{11.33}$$

The proof of Lemma 11.3 is inspired by [120, Lemma 5], [68, Theorem 2] and omitted due to similarity to Lemma 9.3 for the Kaspi problem. We remark that Lemmas 11.2 and 11.3 hold for any memoryless source, not restricted to DMS. As we shall show, the result in Lemma 11.3 leads to a non-asymptotic converse bound for the Fu-Yeung problem.

11.3 Rate-Distortions-Tilted Information Density

Recall that P_{XY} is induced by P_X and the deterministic function $g: \mathcal{X} \to \mathcal{Y}$.

Definition 11.3. For any $(x,y) \in \mathcal{X} \times \mathcal{Y}$ such that $y = g(x)$, the rate-distortions-tilted information density for the Fu-Yeung problem is defined as

$$\jmath(x,y|R_1,D_1,D_2,P_X) := (1+\xi^*)\log\beta(x,y) - \xi^* R_1 \\ - \lambda_1^* D_1 - \lambda_2^* D_2, \tag{11.34}$$

where $\beta(\cdot)$ was defined in (11.28).

The properties of $\jmath(x,y|R_1,D_1,D_2,P_X)$ follow from Lemma 11.2. For example, it follows from (11.24) that

$$\begin{aligned} &\mathsf{R}(R_1|P_X,D_1,D_2) \\ &= \mathsf{E}_{P_{XY}}[\jmath(X,Y|R_1,D_1,D_2,P_X)] &(11.35)\\ &= \mathsf{E}_{P_X}[\jmath(X,g(X)|R_1,D_1,D_2,P_X)]. &(11.36)\end{aligned}$$

Let $\jmath(x|D_1,P_X)$ be the D_1-tilted information density in (3.17), i.e.,

$$\jmath(x|D_1,P_X) \\ := -\log\Big(\sum_{\hat{x}_1} P^*_{\hat{X}_1}(\hat{x}_1)\exp(-\lambda^*(d_1(x,\hat{x}_1)-D_1))\Big), \tag{11.37}$$

where $P^*_{\hat{X}_1}$ is induced by the source distribution P_X and the optimal test channel $P^*_{\hat{X}_1|X}$ for the rate-distortion function $R(P_X,D_1)$ (cf. (3.7)) and $\lambda^* = -\frac{\partial R(P_X,D)}{\partial D}|_{D=D_1}$.

Furthermore, similar to the proofs Lemma 9.4 for the Kaspi problem and Claim (iii) in Lemma 10.3, we have the following lemma that further relates the rate-distortions-tilted information density with the derivative of the minimum sum rate function with respect to the distribution P_X for the Fu-Yeung problem.

Lemma 11.4. Suppose that for all Q_X in the neighborhood of P_X, $\mathrm{supp}(Q^*_{\hat{X}_1\hat{X}_2}) = \mathrm{supp}(P^*_{\hat{X}_1\hat{X}_2})$. Then for any $a \in \mathrm{supp}(P_X)$,

$$\frac{\partial \mathsf{R}(R_1|Q_X,D_1,D_2)}{\partial Q_X(a)}\Bigg|_{Q_X=P_X} \\ = \jmath(x,g(x)|R_1,D_1,D_2,P_X) - (1+s^*). \tag{11.38}$$

11.4 A Non-Asymptotic Converse Bound

We next present a non-asymptotic converse bound for the Fu-Yeung problem. Given any $\gamma \in \mathbb{R}_+$, define the following three sets:

$$\mathcal{A}_1^n := \Big\{(x^n, y^n) : \sum_{i\in[n]} \jmath(x_i|D_1, P_X) \geq \log M_1 + n\gamma\Big\}, \tag{11.39}$$

$$\mathcal{A}_2^n := \Big\{(x^n, y^n) : -\sum_{i\in[n]} \log P_Y(y_i) \geq \log M_2 + n\gamma\Big\}, \tag{11.40}$$

$$\mathcal{A}_3^n := \Big\{(x^n, y^n) : \sum_{i\in[n]} \jmath(x_i, y_i|R_1, D_1, D_2, P_X) \geq \log M_1 M_2 \\ + \xi^* \log M_1 + (1+\xi^*)n\gamma\Big\}. \tag{11.41}$$

Lemma 11.5. Any (n, M_1, M_2)-code for the Fu-Yeung problem satisfies that for any $\gamma \geq 0$,

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \geq \Pr\Big\{(X^n, Y^n) \in \bigcup_{i\in[3]} \mathcal{A}_i^n\Big\} - 4\exp(-\gamma). \tag{11.42}$$

We remark that Lemma 11.3 plays an important role in the proof of Lemma 11.5. This can be made clear by the following definitions. Given $(x, y, \hat{x}_1, \hat{x}_2)$, using the definitions of $\imath_1(\cdot)$ in (11.29) and $\imath_2(\cdot)$ in (11.30), we define

$$\jmath_1(x, y, \hat{x}_1, D_1) := \imath_1(x, y, \hat{x}_1) - \frac{\lambda_1^* D_1}{1+\xi^*}, \tag{11.43}$$

$$\jmath_2(x, y, \hat{x}_1, D_1, D_2) := \imath_2(x, y, \hat{x}_1) - \frac{\lambda_1^* D_1}{1+\xi^*} - \lambda_2^* D_2. \tag{11.44}$$

Using the definition of the rate-distortions-tilted information density in (11.34), we conclude that

$$\jmath(x, y|R_1, D_1, D_2, P_X) = \xi^* \jmath_1(x, y, \hat{x}_1, D_1) \\ + \jmath_2(x, y, \hat{x}_1, D_1, D_2) - \xi^* R_1. \tag{11.45}$$

In the proof of Lemma 11.5, we make use of (11.45) and the fact that $\Pr\{A + B \geq c + d\} \leq \Pr\{A \geq c\} + \Pr\{B \geq d\}$ for any variables (A, B) and constants (c, d).

Recall the setting of the Fu-Yeung problem in Figure 11.1. Note that when Y is a constant, i.e. $|\mathcal{Y}| = 1$, we recover the setting of the successive refinement problem [95]. Recall the definitions of an (n, M_1, M_2)-code for the successive refinement problem Definition 10.1, the definition of the joint excess-distortion probability $\mathrm{P}_{\mathrm{e},n}^{\mathrm{SR}}(D_1, D_2)$ in (10.5), the definition of the minimal sum rate $\mathsf{R}_{\mathrm{SR}}(R_1|P_X, D_1, D_2|)$ in (10.11) and the definition of the rate-distortions tilted information density $\jmath_{\mathrm{SR}}(x, R_1|D_1, D_2, P_X)$ in (10.23). When $|\mathcal{Y}| = 1$, it follows that

$$\mathsf{R}_{\mathrm{SR}}(R_1|P_X, D_1, D_2|) = \mathsf{R}(R_1|P_X, D_1, D_2), \tag{11.46}$$

$$\jmath_{\mathrm{SR}}(x, R_1|D_1, D_2, P_X) = \jmath(x, g(x)|R_1, D_1, D_2, P_X). \tag{11.47}$$

We remark that although the definition of the rate-distortions-tilted information density for the successive refinement problem in the right hand side of (11.47) appears different from (10.23), the two quantities share same properties (cf. [151, Lemma 3]) and are thus essentially the same. Invoking Lemma 11.5 with $\mathcal{Y} = \{1\}$, we obtain the following non-asymptotic converse bound for the successive refinement problem.

Lemma 11.6. Any (n, M_1, M_2)-code for the successive refinement problem satisfies that for any $\gamma \geq 0$,

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}^{\mathrm{SR}}(D_1, D_2) \\
&\geq \Pr\Big\{ \sum_{i\in[n]} \jmath(X_i|D_1, P_X) \geq \log M_1 + n\gamma \text{ or} \\
&\qquad \sum_{i\in[n]} \jmath_{\mathrm{SR}}(X_i, R_1|D_1, D_2, P_X) \geq \log M_1 M_2 \\
&\qquad\qquad + \xi^* \log M_1 + (1+\xi^*)n\gamma \Big\} - 4\exp(-n\gamma).
\end{aligned} \tag{11.48}$$

Lemma 11.6 was also derived by Kostina and Tuncel [68, Corollary 2]. We remark that the non-asymptotic converse bound in (11.48) can be used to establish converse results for second-order asymptotics for any memoryless source, including the results in Theorem 10.4 for DMS. Invoking Lemma 11.6, for the successive refinement problem, we have the potential to establish tight second-order asymptotics for non-successively refinable continuous memoryless sources, e.g., symmetric GMS under quadratic distortion measures [17].

11.5 Second-Order Asymptotics

11.5.1 Preliminaries

Let $\varepsilon \in (0,1)$ be fixed and let (R_1^*, R_2^*) be a boundary rate pair on the rate-distortion region $\mathcal{D}(D_1, D_2|P_X)$ of the Fu-Yeung problem.

Definition 11.4. Given any $\varepsilon \in (0,1)$, a pair (L_1, L_2) is said to be second-order $(R_1^*, R_2^*, D_1, D_2, \varepsilon)$-achievable for the Fu-Yeung problem if there exists a sequence of (n, M_1, M_2)-codes such that

$$\limsup_{n\to\infty} \frac{\log M_i - nR_i}{\sqrt{n}} \le L_i,\ \ i = 1, 2, \tag{11.49}$$

and

$$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \le \varepsilon. \tag{11.50}$$

The closure of the set of all second-order $(R_1^*, R_2^*, D_1, D_2, \varepsilon)$-achievable pairs is called the second-order $(R_1^*, R_2^*, D_1, D_2, \varepsilon)$ coding region and denoted as $\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)$.

To present characterization of $\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon)$, we need several definitions. Recall that P_{XY} and P_Y are induced by P_X and the deterministic function $g : \mathcal{X} \to \mathcal{Y}$ and the definition of the source dispersion function (cf. (2.27)), i.e.,

$$\begin{aligned}\mathrm{V}(P_Y) &= \sum_y P_Y(y)\big(-\log P_Y(y) - H(P_Y)\big)^2 &(11.51)\\ &= \sum_x P_X(x)\big(-\log P_Y(g(x)) - H(P_Y)\big)^2. &(11.52)\end{aligned}$$

Recall that $\mathrm{V}(P_X, D_1) = \mathrm{Var}[\jmath(X|D_1, P_X)]$ is the distortion-dispersion function (cf. (3.57)). Let the rate-distortion-dispersion function be

$$\begin{aligned}&\mathrm{V}(R_1|P_X, D_1, D_2)\\ &:= \mathrm{Var}\Big[\jmath(X, g(X)|R_1, D_1, D_2, P_X) - \log P_Y(Y)\Big]. &(11.53)\end{aligned}$$

Define two covariance matrices:

$$\begin{aligned}&\mathbf{V}_1(R_1|P_X,D_1,D_2)\\&:=\mathrm{Cov}\Big([\jmath(X,g(X)|R_1,D_1,D_2,P_X)-\log P_Y(g(X))]^\top,\\&\qquad\qquad \jmath(X|D_1,P_X)\Big),\end{aligned}\tag{11.54}$$

$$\begin{aligned}&\mathbf{V}_2(R_1|P_X,D_1,D_2)\\&:=\mathrm{Cov}\Big([\jmath(X,g(X)|R_1,D_1,D_2,P_X)-\log P_Y(g(X)),\\&\qquad\qquad -\log P_Y(g(X))]^T\Big).\end{aligned}\tag{11.55}$$

Finally, recall that $\Psi(x_1,x_2;\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the bivariate generalization of the Gaussian cdf.

11.5.2 Main Result and Discussions

Suppose the following conditions hold:

1. $(Q_X,D_1')\to R(Q_X,D_1')$ is twice differentiable in the neighborhood of (P_X,D_1) and the derivatives are bounded;
2. $(Q_X,R_1',D_1',D_2')\to \mathsf{R}(R_1'|Q_X,D_1',D_2')$ is twice differentiable in the neighborhood of (P_X,R_1,D_1,D_2) and the derivatives are bounded;
3. The functions $R(P_X,D_1)$, $R_1^*(D_1,D_2|P_X)$, $R_2^*(D_1,D_2|P_X)$ are positive and finite;
4. The dispersion functions $\mathrm{V}(P_X,D_1)$ and $\mathrm{V}(P_Y)$ are positive and the dispersion function $\mathrm{V}(R_1^*|P_X,D_1,D_2)$ is positive for $R(P_X,D_1)<R_1^*<\mathsf{R}_1^*(P_X,D_1,D_2)$;
5. The covariance matrices $\mathbf{V}_1(R(P_X,D_1)|P_X,D_1,D_2)$ and $\mathbf{V}_2(\mathsf{R}_1^*(P_X,D_1,D_2)P_X,D_1,D_2|)$ are positive semi-definite.

Conditions (i) and (ii) concern the differentiability of rate-distortion functions and have been discussed in detail by Ingber and Kochman in [54, Section III.A]. Condition (iii) can easily verified by calculating the values of rate-distortion functions using convex optimization tools such as [14]. In order to verify conditions (iv) and (v), in general, one needs to develop specialized Blahut-Arimoto-type algorithms [22, Chapter 8] to solve for the optimal test channels.

Theorem 11.7. Under conditions (1) to (5), depending on (R_1^*, R_2^*), for any $\varepsilon \in (0,1)$, the second-order coding region satisfies

- Case (i): $R_1^* = R(P_X, D_1)$ and $R_2^* > \mathsf{R}_2^*(P_X, D_1, D_2)$

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \{(L_1, L_2) : L_1 \geq \sqrt{\mathrm{V}(P_X, D_1)}\mathrm{Q}^{-1}(\varepsilon)\}. \tag{11.56}$$

- Case (ii): $R_1^* = R(P_X, D_1)$ and $R_2^* = \mathsf{R}_2^*(P_X, D_1, D_2)$

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \{(L_1, L_2) : \Psi(L_1, (1+\xi^*)L_1 + L_2; \mathbf{0}_2; \mathbf{V}_1(R_1^*|P_X, D_1, D_2)) \geq 1 - \varepsilon\}. \tag{11.57}$$

- Case (iii): $R(P_X, D_1) < R_1^* < \mathsf{R}_1^*(P_X, D_1, D_2)$ and $R_2^* = \mathsf{R}^*(R_1^*|P_X, D_1, D_2) + H(P_Y) - R_1^*$,

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \{(L_1, L_2) : (1+\xi^*)L_1 + L_2 \geq \sqrt{\mathrm{V}(R_1^*|P_X, D_1, D_2)}\mathrm{Q}^{-1}(\varepsilon)\}. \tag{11.58}$$

- Case (iv) $R_1^* = \mathsf{R}_1^*(P_X, D_1, D_2)$ and $R_2^* = H(P_Y)$

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \{(L_1, L_2) : \Psi(L_1 + L_2, L_2; \mathbf{0}_2, \mathbf{V}_2(R_1^*|P_X, D_1, D_2)) \geq 1 - \varepsilon\}. \tag{11.59}$$

- Case (v) $R_1^* > \mathsf{R}_1^*(P_X, D_1, D_2)$ and $R_2^* = H(P_Y)$

$$\mathcal{L}(R_1^*, R_2^*, D_1, D_2, \varepsilon) = \{(L_1, L_2) : L_2 \geq \sqrt{\mathrm{V}(P_Y)}\mathrm{Q}^{-1}(\varepsilon)\}. \tag{11.60}$$

The proof of Theorem 11.7 is provided in Section 11.6. The achievability part follows by the method of types, where we first prove a type-covering lemma tailored to the Fu-Yeung problem, and subsequently apply Taylor expansions of the rate-distortion function and the minimal sum rate function of empirical distributions around the source distribution P_X, and finally apply the Berry-Esseen theorem for each case. The converse part follows by deriving a type-based strong converse analogously to the converse proof the successive refinement problem in Theorem 10.4 and proceeding similarly to the achievability proof.

Since the successive refinement problem is special case of the Fu-Yeung problem when $Y = g(X)$ is a constant, the second-order asymptotics for the successive refinement problem for DMS under bounded distortion measures in Theorem 10.4 is recovered by cases (i)-(iii) in Theorem 11.7.

11.5.3 An Numerical Example

We consider the numerical example inspired by [90] and calculate the dispersion function for cases (iii) and (iv) in Theorem 11.7. Let $\mathcal{S}_1 = \{0,1\}$ and $\mathcal{S}_2 = \{0,1,\mathrm{e}\}$. Let S_1 take values in $\mathcal{S}_1$ with equal probability and let $P_{S_2|S_1}(s_2|s_1) = (1-p)\mathbb{1}(s_1 = s_2) + p\mathbb{1}(s_2 = \mathrm{e})$. Let the source be $X = (S_1, S_2)$ and the deterministic function be $Y = g(X) = g(S_1, S_2) = S_2$. Let $\hat{\mathcal{X}}_1 = \hat{\mathcal{X}}_2 = \{0,1\}$ and the distortion measures be $d_1(x, \hat{x}_1) = \mathbb{1}(s_1 = \hat{x}_1)$ and $d_2(x, \hat{x}_2) = \mathbb{1}(s_2 = \hat{x}_2)$. Choose (p, D_1, D_2) such that $D_1 \leq \frac{1}{2}$ and $D_1 - \frac{1-p}{2} \leq D_2 \leq pD_1$. For this case, using the definitions of ξ^* in (11.17), λ_1^* in (11.18) and λ_2^* in (11.19), we have

$$\xi^* = 0, \tag{11.61}$$

$$\lambda_1^* = \log\left((1-p)/(D_1 - D_2) - 1\right), \tag{11.62}$$

$$\lambda_2^* = -\lambda_1^* + \log\left(p/D_2 - 1\right). \tag{11.63}$$

Recall that $H_\mathrm{b}(\cdot)$ is the binary entropy function. Let

$$\alpha_0 := \log\left(2/(1 + \exp(-\lambda_1^*))\right) - \lambda_1^* D_1 - \lambda_2^* D_2, \tag{11.64}$$

$$\alpha := \log\left(2/(1 + \exp(-\lambda_1^* - \lambda_2^*))\right) - \lambda_1^* D_1 - \lambda_2^* D_2, \tag{11.65}$$

$$g_1(p, D_1, D_2) := \log 2 - (1-p)H_\mathrm{b}((D_1 - D_2)/(1-p)) - pH_\mathrm{b}(D_2/p), \tag{11.66}$$

$$g_2(p, D_1, D_2) := p(1-p)\Big\{\log(1 - D_2/p) - \log(1 - (D_1 - D_2)/(1-p))\Big\}^2, \tag{11.67}$$

$$g_3(p, D_1, D_2) := (1-p)\alpha_0 \log\frac{2}{1-p} + p\alpha \log\frac{1}{p}. \tag{11.68}$$

Then, it can be verified that

$$H(P_Y) = (1-p)\log 2 + H_{\mathrm{b}}(p), \tag{11.69}$$

$$\mathrm{V}(P_Y) = p(1-p)\Big(\log\frac{2p}{1-p}\Big)^2. \tag{11.70}$$

Thus,

$$\begin{aligned}\mathrm{V}(R_1|P_X,D_1,D_2) &= g_2(p,D_1,D_2) + \mathrm{V}(P_Y)\\ &\quad + 2\Big(g_3(p,D_1,D_2) - H(P_Y)g_1(p,D_1,D_2)\Big).\end{aligned} \tag{11.71}$$

11.6 Proof of Second-Order Asymptotics

11.6.1 Achievability

In this subsection, we first present a type covering lemma tailored to the Fu-Yeung problem, using which we derive an upper bound on the joint excess-distortion and error probability. Finally, invoking Taylor expansions and the Berry-Esseen Theorem, we derive an achievable second-order coding region.

Define

$$c_1 = |\mathcal{X}|\cdot|\mathcal{Y}|\cdot|\hat{\mathcal{X}}_1| + 2, \tag{11.72}$$

$$c_2 = 7|\mathcal{X}|\cdot|\mathcal{Y}|\cdot|\hat{\mathcal{X}}_1|\cdot|\hat{\mathcal{X}}_2| + 4. \tag{11.73}$$

We are now ready to present the type covering lemma.

Lemma 11.8. Consider any type $Q_X \in \mathcal{P}_n(\mathcal{X})$. Let Q_Y be induced by Q_X and the deterministic function $g : \mathcal{X} \to \mathcal{Y}$, and let $R_1 \geq R(Q_X, D_1)$. The following conclusions hold.

1. There exists a set $\mathcal{B} \in \mathcal{X}_1^n$ such that for each $x^n \in \mathcal{T}_{Q_X}$, $d_1(x^n,(z^n)^*) \leq D_1$ where $(z^n)^* := \arg\min_{z\in\mathcal{B}} d_1(x^n, z)$.
2. Given $(z^n)^*$, there exists a set $\mathcal{B}((z^n)^*) \in \mathcal{X}_2^n$ such that

$$\min_{\hat{x}_2^n\in\mathcal{B}((z^n)^*)} d_2(x^n,\hat{x}_2^n) \leq D_2. \tag{11.74}$$

3. There exists a set $\mathcal{B}_Y \in \hat{\mathcal{Y}}^n$ satisfying that $\frac{1}{n}\log|\mathcal{B}_Y| \leq H(Q_Y)$ and there exists $\hat{y}^n \in \mathcal{B}_Y$ such that $\hat{y}^n = g(x^n)$.

4. The sizes of sets $\mathcal{B}$ and $\mathcal{B}((z^n)^*)$ satisfy

$$\frac{1}{n}\log|\mathcal{B}| \le R_1 + c_1\log(n+1), \tag{11.75}$$

$$\frac{1}{n}\log(|\mathcal{B}|\cdot|\mathcal{B}((z^n)^*)|) \le \mathsf{R}(R_1|Q_X, D_1, D_2) + (c_1+c_2)\log(n+1). \tag{11.76}$$

The proof of Lemma 11.8 is similar to the type covering lemma for the successive refinement problem [82].

Let

$$R_{1,n} := \frac{1}{n}\big(\log M_1 - (c_1+|\mathcal{X}|)\log(n+1)\big), \tag{11.77}$$

$$R_{2,n} := \frac{1}{n}\big(\log M_2 - (c_2+|\mathcal{Y}|)\log(n+1)\big). \tag{11.78}$$

Invoking Lemma 11.8, we can upper bound the joint excess-distortion and error probability for an (n, M_1, M_2)-code.

Lemma 11.9. There exists an (n, M_1, M_2)-code such that

$$\begin{aligned}&\mathrm{P}_{\mathrm{e},n}(D_1,D_2)\\ &\le \Pr\bigg\{R_{1,n} < R(\hat{T}_{X^n},D_1) \text{ or } R_{2,n} + \frac{c_2\log(n+1)}{n} < H(\hat{T}_{g(X^n)})\\ &\qquad \text{or } R_{1,n}+R_{2,n} < \mathsf{R}(R_{1,n}|\hat{T}_{X^n},D_1,D_2) + H(\hat{T}_{g(X^n)})\bigg\}.\end{aligned} \tag{11.79}$$

Proof. Set $(R_1, R_2) = (R_{1,n}, R_{2,n})$. Consider the following coding scheme. Given a source x^n, the encoder f_2 calculates its type $\hat{T}_{x^n}$. Then, the encoder f_2 obtain y^n using the deterministic function $y_i = g(x_i)$ and its type $\hat{T}_{y^n}$. Now encoder f_2 calculates $R(\hat{T}_{x^n}, D_1)$ and $\mathsf{R}(R_{1,n}|\hat{T}_{x^n}, D_1, D_2)$. If $\log M_1 < nR(\hat{T}_{x^n}, D_1) + (c_1 + |\mathcal{X}|)\log(n+1)$ or $\log M_2 < nH(\hat{T}_{y^n}) + |\mathcal{Y}|\log(n+1)$ or $\log M_1M_2 < n\mathsf{R}(R_{1,n}|\hat{T}_{x^n}, D_1, D_2) + nH(\hat{T}_{y^n}) + (c_1+c_2+|\mathcal{X}|+|\mathcal{Y}|)\log(n+1)$, then the system declares an error. Otherwise, the encoder f_1 sends the type of x^n with at most $|\mathcal{X}|\log(n+1)$ nats and the encoder f_2 sends the type of y^n using at most $|\mathcal{Y}|\log(n+1)$ nats. Furthermore, the encoder f_2 sends the index of $y^n = g(x^n)$ in the type class $\mathcal{T}_{\hat{T}_{y^n}}$. Now, choose $\mathcal{B} \in \mathcal{X}_1^n$ in Lemma 11.8 and let $(z^n)^* = \arg\min_{z\in\mathcal{B}} d_1(x^n, z)$. Given $(z^n)^*$, choose

$\mathcal{B}((z^n)^*)$ in Lemma 11.8 and let $z_2^* = \arg\min_{z_2 \in \mathcal{B}((z^n)^*)} d_2(x^n, z_2)$. Finally, we use the encoder f_1 to send the index of z_1^* and use either f_1 or f_2 to send out the index of z_2^*. Invoking Lemma 11.8, we conclude that no error will be made if $\log M_1 \geq nR(\hat{T}_{x^n}, D_1) + (c_1 + |\mathcal{X}|)\log(n+1)$, $\log M_2 \geq nH(\hat{T}_{y^n}) + |\mathcal{Y}|\log(n+1)$ and $\log M_1M_2 \geq n\mathsf{R}(R_{1,n}|\hat{T}_{x^n}, D_1, D_2) + nH(\hat{T}_{y^n}) + (c_1 + c_2 + |\mathcal{X}| + |\mathcal{Y}|)\log(n+1)$. The proof is now complete.

□

Recall that (R_1^*, R_2^*) is a boundary rate-pair on the rate-distortion region of the Fu-Yeung problem. Choose (M_1, M_2) such that

$$\log M_1 = nR_1^* + L_1\sqrt{n} + (c_1 + |\mathcal{X}|)\log(n+1), \tag{11.80}$$
$$\log M_2 = nR_2^* + L_2\sqrt{n} + (c_2 + |\mathcal{Y}|)\log(n+1). \tag{11.81}$$

It follows from (11.77) and (11.78) that

$$R_{i,n} = R_i^* + \frac{L_i}{\sqrt{n}}, \ i = 1, 2. \tag{11.82}$$

Recall the definition of the typical set $\mathcal{A}_n(P_X)$ in (3.87). The result in (3.88) states that

$$\Pr\left\{\hat{T}_{X^n} \notin \mathcal{A}_n(P_X)\right\} \leq \frac{2|\mathcal{X}|}{n^2}. \tag{11.83}$$

Recall that P_Y is induced by the source distribution P_X and the deterministic function $g : \mathcal{X} \to \mathcal{Y}$. Thus, given any x^n, for each $y \in \mathcal{Y}$,

$$\hat{T}_{y^n}(y) - P_Y(y) = \hat{T}_{g(x^n)}(y) - P_Y(y) \tag{11.84}$$
$$= \sum_{x:g(x)=y} \Big(\hat{T}_{x^n}(x) - P_X(x)\Big). \tag{11.85}$$

Thus, if $\hat{T}_{X^n} \in \mathcal{A}_n(P_X)$,

$$\|\hat{T}_{Y^n} - P_Y\|_\infty \leq |\mathcal{X}|\sqrt{\frac{\log n}{n}}. \tag{11.86}$$

For x^n such that $\hat{T}_{x^n} \in \mathcal{A}_n(P_X)$, applying Taylor's expansions and noting that $y^n = g(x^n)$, we obtain

$$H(\hat{T}_{g(x^n)}) = H(\hat{T}_{y^n}) \tag{11.87}$$

$$= H(P_Y) + \sum_y \Big(\hat{T}_{y^n}(y) - P_Y(y)\Big)(-\log P_Y(y)) + O\Big(\|\hat{T}_{y^n} - P_Y\|^2\Big) \tag{11.88}$$

$$= \sum_y -\hat{T}_{y^n}(y)\log P_Y(y) + O\left(\frac{\log n}{n}\right) \tag{11.89}$$

$$= \frac{1}{n}\sum_{i\in[n]} -\log P_Y(y_i) + O\left(\frac{\log n}{n}\right), \tag{11.90}$$

and

$$\mathsf{R}(R_{1,n}|\hat{T}_{x^n}, D_1, D_2)$$

$$= \mathsf{R}(R_1^*|P_X, D_1, D_2) - s^*\frac{L_1}{\sqrt{n}} + O(|R_{1,n} - R_1^*|^2) + \sum_x \Big(\hat{T}_{x^n} - P_X(x)\Big)\jmath(x, g(x)|R_1^*, D_1, D_2, P_X) + O\Big(\|\hat{T}_{x^n} - P_X\|^2\Big) \tag{11.91}$$

$$= \frac{1}{n}\sum_{i\in[n]} \jmath(x_i, g(x_i)|R_1^*, D_1, D_2, P_X) - \frac{s^* L_1}{\sqrt{n}} + O\Big(\frac{\log n}{n}\Big), \tag{11.92}$$

where (11.91) follows from Lemma 11.4. Furthermore, for x^n such that $\hat{T}_{x^n} \in \mathcal{A}_n(P_X)$, it follows from (3.90)that

$$R(\hat{T}_{x^n}, D) = \frac{1}{n}\sum_{i\in[n]} \jmath(x_i|D, P_X) + O\left(\frac{\log n}{n}\right). \tag{11.93}$$

Recall that $\xi_n = \frac{\log n}{n}$. Therefore, invoking Lemma 11.9, we obtain

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\leq \Pr\Big\{R_{1,n} < R(\hat{T}_{X^n}, D_1) \text{ or } R_{2,n} + \frac{c_2 \log(n+1)}{n} < H(\hat{T}_{g(X^n)}) \\
&\qquad \text{or } R_{1,n} + R_{2,n} < H(\hat{T}_{g(X^n)}) + \mathsf{R}(R_{1,n}|\hat{T}_{x^n}, D_1, D_2) \\
&\qquad \text{and } \hat{T}_{X^n} \in \mathcal{A}_n(P_{XY})\Big\} \\
&\quad + \Pr\Big\{\hat{T}_{X^n} \notin \mathcal{A}_n(P_{XY})\Big\} \qquad (11.94) \\
&\leq \Pr\Big\{R_1 + \frac{L_1}{\sqrt{n}} < \frac{1}{n}\sum_{i\in[n]} \jmath(X_i|D_1, P_X) + O(\xi_n) \\
&\qquad \text{or } R_2 < \frac{1}{n}\sum_{i\in[n]} \log\frac{1}{P_Y(Y_i)} + O(\xi_n) \\
&\qquad \text{or } R_1 + R_2 + \frac{(1+s^*)L_1 + L_2}{n} \\
&\qquad\quad < \frac{1}{n}\sum_{i\in[n]} \Big(\jmath(X_i, g(X_i)|R_1, D_1, D_2, P_X) - \log P_Y(Y_i)\Big) \\
&\qquad\qquad + O(\xi_n)\Big\} + \frac{2|\mathcal{X}|}{n^2}. \qquad (11.95)
\end{aligned}$$

Subsequently, we upper bound (11.95) for different cases of boundary rate pairs (R_1^*, R_2^*) in Theorem 11.7. For simplicity, let $\jmath(x|R_1^*)$ denote $\jmath(x^*, g(x^*)|R_1^*, D_1, D_2, P_X)$ for each $x \in \mathcal{X}$.

- Case (i) $R_1^* = R(P_X, D_1)$ and $R_2^* > \mathsf{R}_2^*(P_X, D_1, D_2)$

  In this case $R_2^* > H(P_Y)$. Thus, it follows from the weak law of large numbers in Theorem 1.1 that

$$\kappa_{1,n} := \Pr\Big\{R_2^* < \frac{1}{n}\sum_{i\in[n]} \log\frac{1}{P_Y(Y_i)} + O(\xi_n)\Big\} \to 0 \qquad (11.96)$$

and

$$\kappa_{2,n} := \Pr\Big\{R_1^* + R_2^* + \frac{(1+s^*)L_1 + L_2}{n} < \frac{1}{n}\sum_{i\in[n]}\Big(\jmath(X_i|R_1^*) - \log P_Y(Y_i)\Big) + O(\xi_n)\Big\} \to 0. \tag{11.97}$$

It follows from (11.95) that

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1, D_2)\\
&\le \Pr\Big\{R_1^* + \frac{L_1}{\sqrt{n}} < \frac{1}{n}\sum_{i\in[n]}\jmath(X_i|D_1,P_X) + O(\xi_n)\Big\}\\
&\quad + \frac{2|\mathcal{X}|}{n^2} + \kappa_{1,n} + \kappa_{2,n} \qquad (11.98)\\
&\le \mathrm{Q}\left(\frac{L_1 + O(\sqrt{n}\xi_n)}{\sqrt{\mathrm{V}(P_X,D_1)}}\right) + \frac{6\mathrm{T}(P_X,D_1)}{\sqrt{n}\mathrm{V}(P_X,D_1)}\\
&\quad + \frac{2|\mathcal{X}|}{n^2} + \kappa_{1,n} + \kappa_{2,n}, \qquad (11.99)
\end{aligned}$$

where $\mathrm{T}(P_X, D_1)$ is the third absolute moment of $\jmath(X|D_1, P_X)$ (which is finite for DMS) and (11.99) follows by applying the Berry-Esseen theorem to the first term in (11.98). If we choose (L_1, L_2) such that

$$L_1 \ge \sqrt{\mathrm{V}(P_X, D_1)}\mathrm{Q}^{-1}(\varepsilon), \tag{11.100}$$

then $\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \le \varepsilon$ as desired.

- Case (ii) $R_1^* = R(P_X, D_1)$ and $R_2^* = \mathsf{R}_2^*(P_X, D_1, D_2)$

  In this case, $R_2^* > H(P_Y)$ still holds. Hence, invoking (11.95), we obtain

$$
\begin{aligned}
&1 - \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\geq \Pr\Bigg\{ R_1^* + \frac{L_1}{\sqrt{n}} \geq \frac{1}{n}\sum_{i\in[n]} \jmath(X_i|D_1, P_X) + O(\xi_n), \\
&\qquad R_1^* + R_2^* + \frac{(1+s^*)L_1 + L_2}{\sqrt{n}} \geq \\
&\qquad\qquad \frac{1}{n}\sum_{i\in[n]} \Big(-\log P_Y(Y_i) + \jmath(X_i|R_1^*) + O(\xi_n) \Big\} \\
&\qquad - \frac{2|\mathcal{X}|}{n^2} - \kappa_{1,n} \qquad (11.101)\\
&\geq 1 - \Psi\big(L_1 + O(\xi_n), (1+s^*)L_1 + L_2 + O(\xi_n); \\
&\qquad\qquad \mathbf{0}_2; \mathbf{V}_1(R_1^*|P_X, D_1, D_2)\big) \qquad (11.102)\\
&\qquad - \frac{2|\mathcal{X}|}{n^2} - \kappa_{1,n} + O\Big(\frac{1}{\sqrt{n}}\Big). \qquad (11.103)
\end{aligned}
$$

Hence, if we choose (L_1, L_2) such that

$$\Psi\big(L_1, (1+s^*)L_1 + L_2; \mathbf{0}_2; \mathbf{V}_1(R_1^*|P_X, D_1, D_2)\big) \geq 1 - \varepsilon, \qquad (11.104)$$

$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \varepsilon$.

- Case (iii) $R(P_X, D_1) < R_1^* < \mathsf{R}_1^*(P_X, D_1, D_2)$, and $R_2^* = \mathsf{R}_1^*(P_X, D_1, D_2) + H(P_Y) - R_1^*$

  In this case, $R_2^* > H(P_Y)$ holds again. The analysis is similar to Case (i). It can be verified that if we choose (L_1, L_2) such that

  $$(1+s^*)L_1 + L_2 \geq \sqrt{\mathrm{V}(R_1^*|P_X, D_1, D_2)}\mathrm{Q}^{-1}(\varepsilon), \qquad (11.105)$$

  $\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \varepsilon$.

- Case (iv) $R_1^* = \mathsf{R}_1^*(P_X, D_1, D_2)$ and $R_2^* = H(P_Y)$

  The analysis is similar to Case (ii). It can be verified that if

  $$\Psi((1+s^*)L_1 + L_2, L_2; \mathbf{0}_2; \mathbf{V}_2(R_1^*, D_1, D_2|P_X)) \geq 1 - \varepsilon, \qquad (11.106)$$

  $\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \varepsilon$.

- Case (v) $R_1 > \mathsf{R}_1^*(P_X, D_1, D_2)$ and $R_2^* = H(P_Y)$

  The analysis is similar to Case (i). It can be verified that if

$$L_2 \geq \sqrt{\mathrm{V}(P_Y)}\mathrm{Q}^{-1}(\varepsilon), \tag{11.107}$$

  we have $\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \varepsilon$.

The achievability proof of Theorem 11.7 is now completed.

11.6.2 Converse

The following type-based strong converse lemma is critical in the converse proof.

Lemma 11.10. Fix $c > 0$ and a type $Q_X \in \mathcal{P}_n(P_X)$. For any (n, M_1, M_2)-code such that

$$\Pr\left\{d_1(X^n, \hat{X}_1^n) \leq D_1,\ d_2(X^n, \hat{X}_2^n) \leq D_2,\ \hat{Y}^n = Y^n | X^n \in \mathcal{T}_{Q_X}\right\} \\ \geq \exp(-nc), \tag{11.108}$$

there exists a conditional distribution $Q_{\hat{X}_1\hat{X}_2|X}$ such that

$$\frac{1}{n}\log M_1 \geq I(Q_X, Q_{\hat{X}_1|X}) - \xi_{1,n}, \tag{11.109}$$

$$\frac{1}{n}\log M_2 \geq H(Q_Y) - \xi_{2,n}, \tag{11.110}$$

$$\frac{1}{n}\log M_1M_2 \geq H(Q_Y) + I(Q_Y, Q_{\hat{X}_1|Y}) \\ + I(Q_{X|Y}, Q_{\hat{X}_1\hat{X}_2|XY}|Q_Y) - \xi_{1,n} - \xi_{2,n}, \tag{11.111}$$

where

$$\xi_{1,n} = \frac{|\mathcal{X}|\log(n+1) + \log n + nc}{n}, \tag{11.112}$$

$$\xi_{2,n} = 2\xi_{1,n} + \frac{2(\log n + nc) + |\mathcal{X}| \cdot |\mathcal{Y}| \log(n+1)}{n} \\ + \frac{\log|\mathcal{Y}| + h_b(1/n)}{n}. \tag{11.113}$$

and $Q_{X|Y}$, $Q_{\hat{X}_1|Y}$, $Q_{\hat{X}_1\hat{X}_2|XY}$ are induced by Q_X, $Q_{\hat{X}_1\hat{X}_2|X}$ and the deterministic function $y = g(x)$.

Furthermore, the expected distortions are bounded as

$$\mathbb{E}_{Q_X \times Q_{\hat{X}_1\hat{X}_2|X}}[d_1(X,\hat{X}_1)] \le D_1 + \frac{\overline{d}_1}{n} := D_{1,n}, \tag{11.114}$$

$$\mathbb{E}_{Q_X \times Q_{\hat{X}_1\hat{X}_2|X}}[d_2(X,\hat{X}_2)] \le D_2 + \frac{\overline{d}_2}{n} := D_{2,n}. \tag{11.115}$$

The proof of Lemma 11.10 is similar to Lemma 3.8 for the rate-distortion problem and Lemma 10.10 for the successive refinement problem. The main technique is the perturbation approach by Gu and Effros [45] and the generalization with method of types [127].

Let $c = \frac{\log n}{n}$, then we have

$$\xi_{1,n} = \frac{|\mathcal{X}|\log(n+1) + 2\log n}{n}, \tag{11.116}$$

$$\xi_{2,n} = \frac{8\log n + (|\mathcal{X}|\cdot|\mathcal{Y}| + 2|\mathcal{X}|)\log(n+1)}{n} + \frac{\log|\mathcal{Y}| + h_b(1/n)}{n}. \tag{11.117}$$

Define

$$R_{i,n} = \frac{1}{n}(\log M_i + n\xi_{i,n}),\ i \in [2]. \tag{11.118}$$

Invoking Lemma 11.10, we can prove the following lower bound on the joint excess-distortion and error probability for any (n, M_1, M_2)-code.

Lemma 11.11. Any (n, M_1, M_2)-code satisfies that

$$\begin{aligned}&\mathrm{P}_{\mathrm{e},n}(D_1, D_2)\\&\ge \Pr\Big\{R_{1,n} < R(\hat{T}_{X^n}, D_{1,n}) \text{ or } R_{2,n} < H(\hat{T}_{g(X^n)}) \text{ or}\\&\qquad R_{1,n} + R_{2,n} < \mathsf{R}(R_{1,n}|\hat{T}_{X^n}, D_{1,n}, D_{2,n}) + H(\hat{T}_{g(X^n)})\Big\}.\end{aligned} \tag{11.119}$$

The rest of the converse proof is omitted since it is analogous to the achievability proof where we use Taylor expansions similar to (11.90) to (11.93) and apply (multi-variate) Berry-Esseen theorems for each case of boundary rate pairs (R_1^*, R_2^*) in Theorem 11.7.

12

Gray-Wyner Problem

This section studies the lossy Gray-Wyner problem where three encoders cooperatively compress two correlated source sequences so that each of the two decoders could recover a source sequence reliably in a lossy manner. The lossy Gray-Wyner problem is a paradigm of the multiterminal lossy source coding problem where there exist multiple source sequences, multiple encoders and multiple decoders. The problem significantly generalizes the rate-distortion problem by introducing one more source sequence, two more encoders and one more decoder.

The rate-distortion region for the problem was derived by Gray and Wyner [43] and this is why the problem is so named. An auxiliary random variable is needed to characterize the rate-distortion region of the lossy Gray-Wyner problem, which makes it significantly different from all problems discussed in previous sections. The second-order asymptotics for the lossless version of the Gray-Wyner problem was derived by Watanabe [127]. This section presents the generalization of [127] to the lossy case, analogous to the generalization of second-order asymptotics from lossless source coding in Section 2 (cf. [49], [109]) to the rate-distortion problem in Section 3 (cf. [54], [70]).

The Gray-Wyner problem is interesting beyond data compression. In the Gray-Wyner problem, there is an encoder who transmits messages to both decoders and its rate is known as the common rate. Given rates of the other two encoders, the minimal common rate equals a measure of common information of two correlated random variables [125]. Leveraging results on lossy common information by Viswanatha *et al.* [125] and considering rate triples on the Pangloss plane where the sum rate is constrained, the second-order asymptotic result is simplified and numerically illustrated. This section is largely based on [150].

12.1 Problem Formulation and Asymptotic Result

12.1.1 Problem Formulation

The lossy Gray-Wyner source coding problem [43] is shown in Figure 12.1. There are three encoders and two decoders. Encoder f_i has access to a source sequence pair (X^n, Y^n) and compresses it into a message S_i. Decoder ϕ_1 aims to recover source sequence X^n under fidelity criterion d_1 and distortion level D_1 with the encoded message S_0 from encoder f_0 and S_1 from encoder f_1. Similarly, the decoder ϕ_2 aims to recover Y^n with messages S_0 and S_2. We consider a correlated memoryless source (X^n, Y^n) generated i.i.d. from a joint distribution P_{XY} defined on a finite alphabet $\mathcal{X} \times \mathcal{Y}$.

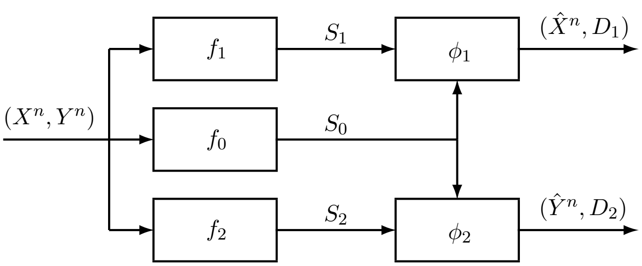

Figure 12.1: System model for the lossy Gray-Wyner source coding problem [43].

Definition 12.1. An (n, M_0, M_1, M_2)-code for lossy Gray-Wyner source coding consists of three encoders:

$$f_0 : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{M}_0 := [M_0], \tag{12.1}$$

$$f_1 : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{M}_1 := [M_1], \tag{12.2}$$

$$f_2 : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{M}_2 := [M_2], \tag{12.3}$$

and two decoders:

$$\phi_1 : \mathcal{M}_0 \times \mathcal{M}_1 \to \hat{\mathcal{X}}^n, \tag{12.4}$$

$$\phi_2 : \mathcal{M}_0 \times \mathcal{M}_2 \to \hat{\mathcal{Y}}^n. \tag{12.5}$$

Let $d_1 : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ and $d_2 : \mathcal{Y} \times \hat{\mathcal{Y}} \to [0, \infty)$ be two bounded distortion measures. Let $\overline{d}_1 := \max_{x,\hat{x}} d_1(x, \hat{x})$ and $\underline{d}_1 := \min_{x,\hat{x}:d_1(x,\hat{x})>0} d_1(x, \hat{x})$ denote the maximal and minimal distortion, respectively. Similarly, we define $\overline{d}_2$ and $\underline{d}_2$. Furthermore, let the average distortion between x^n and $\hat{x}^n$ be defined as $d_1(x^n, \hat{x}^n) := \frac{1}{n}\sum_{i=1}^n d_1(x_i, \hat{x}_i)$ and the average distortion $d_2(y^n, \hat{y}^n)$ be defined similarly.

12.1.2 Rate-Distortion Region

The rate-distortion region of the lossy Gray-Wyner problem is defined as follows.

Definition 12.2. A rate triplet (R_0, R_1, R_2) is said to be (D_1, D_2)-achievable if there exists a sequence of (n, M_0, M_1, M_2)-codes such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_0 \le R_0, \tag{12.6}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log M_1 \le R_1, \tag{12.7}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log M_2 \le R_2, \tag{12.8}$$

and

$$\limsup_{n\to\infty} \mathbb{E}\left[d_1(X^n, \hat{X}^n)\right] \le D_1, \tag{12.9}$$

$$\limsup_{n\to\infty} \mathbb{E}\left[d_2(Y^n, \hat{Y}^n)\right] \le D_2. \tag{12.10}$$

The closure of the set of all (D_1, D_2)-achievable rate triplets is the (D_1, D_2)-optimal rate region and denoted as $\mathcal{R}(P_{XY}, D_1, D_2)$.

Gray and Wyner characterized the (D_1, D_2)-achievable rate region in [43]. Let $\mathcal{P}(P_{XY})$ be the set of all joint distributions $P_{XYW} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{W})$ such that the $\mathcal{X} \times \mathcal{Y}$-marginal of P_{XYW} is the source distribution P_{XY} and $|\mathcal{W}| \leq |\mathcal{X}||\mathcal{Y}| + 2$. Denote the $\mathcal{X} \times \mathcal{W}$ marginal distribution as P_{XW} and the $\mathcal{Y} \times \mathcal{W}$ marginal distribution as P_{YW}.

Theorem 12.1. The (D_1, D_2)-achievable rate region for lossy Gray-Wyner source coding is

$$\begin{aligned}\mathcal{R}(P_{XY}, D_1, D_2) \\ = \bigcup_{P_{XYW} \in \mathcal{P}(P_{XY})} \Big\{ (R_0, R_1, R_2) : R_0 \geq I(P_{XY}, P_{XY|W}) \\ R_1 \geq R_{X|W}(P_{XW}, D_1), R_2 \geq R_{Y|W}(P_{YW}, D_2) \Big\},\end{aligned} \quad (12.11)$$

where $R_{X|W}(P_{XW}, D_1)$ and $R_{Y|W}(P_{YW}, D_2)$ are conditional rate-distortion functions [38, pp. 275, Chapter 11], i.e.,

$$R_{X|W}(P_{XW}, D_1) = \min_{P_{\hat{X}|XW} : \mathbb{E}[d_1(X,\hat{X})] \leq D_1} I(P_{X|W}, P_{X|W\hat{X}} | P_W), \quad (12.12)$$

and $R_{Y|W}(P_{YW}, D_2)$ is defined similarly.

Similar to the rate-distortion and the Kaspi problems, the rate-distortion region in Theorem 12.1 still hold under the vanishing joint excess-distortion probability criterion, i.e., when $\lim_{n \to \infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) = 0$, where

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) := \Pr\Big\{ d_1(X^n, \hat{X}^n) > D_1 \text{ or } d_2(Y^n, \hat{Y}^n) > D_2 \Big\}. \quad (12.13)$$

An equivalent form of the first-order coding region for Gray-Wyner problem was given in [38, Exercise 14.9] and states that

$$\begin{aligned}\mathcal{R}(P_{XY}, D_1, D_2) = \bigcup_{\substack{P_{W|XY}, P_{\hat{X}_1|XW}, P_{\hat{Y}|YW} : \\ \mathbb{E}[d_1(X,\hat{X})] \leq D_1,\ \mathbb{E}[d_2(Y,\hat{Y})] \leq D_2}} \\ \Big\{ (R_0, R_1, R_2) : R_0 \geq I(P_{XY}, P_{XY|W}), \\ R_1 \geq I(P_{X|W}, P_{X|W\hat{X}} | P_W), R_2 \geq I(P_{Y|W}, P_{Y|W\hat{Y}} | P_W) \Big\}.\end{aligned} \quad (12.14)$$

Given any rates (R_1, R_2), let the minimal common rate be defined as

$$\begin{aligned}
&\mathsf{R}_0(R_1, R_2|P_{XY}, D_1, D_2) \\
&:= \min\{R_0 : (R_0, R_1, R_2) \in \mathcal{R}(P_{XY}, D_1, D_2)\} \quad (12.15)\\
&= \min_{\substack{P_{XYW} \in \mathcal{P}(P_{XY}):\\ R_1 \geq R_{X|W}(P_{XW}, D_1)\\ R_2 \geq R_{Y|W}(P_{YW}, D_2)}} I(P_{XY}, P_{XY|W}) \quad (12.16)\\
&= \min_{\substack{P_{W|XW} P_{\hat{X}_1|XW} P_{\hat{Y}|YW}:\\ \mathbb{E}[d_1(X,\hat{X})] \leq D_1,\ \mathbb{E}[d_2(Y,\hat{Y})] \leq D_2\\ I(P_{X|W}, P_{X|W\hat{X}}|P_W) \leq R_1\\ I(P_{Y|W}, P_{Y|W\hat{Y}}|P_W) \leq R_2}} I(P_{XY}, P_{XY|W}), \quad (12.17)
\end{aligned}$$

where (12.16) follows from Theorem 12.1 and (12.17) follows from (12.14). Given distortion levels (D_1, D_2), a rate triple (R_0^*, R_1^*, R_2^*) lies on the boundary of the rate-distortion region if and only if $R_0^* = \mathsf{R}_0(R_1^*, R_2^*, D_1, D_2)$, which is of interest in the study of second-order asymptotics.

12.2 Rates-Distortions-Tilted Information Density

Analogous to the derivation of second-order asymptotics for the rate-distortion problem, the definition of a tilted information density is critical and is usually related to the rate-distortion function (region). For the lossy Gray-Wyner problem, a slight obstacle is encountered on whether to define the rates-distortions-tilted information density using the formula of the minimal common rate in (12.16) that follows from Theorem 12.1 or the formula in (12.17) that follows from the equivalent form of the rate-distortion region in (12.14). This section shows that the latter is more amenable since it does not involve optimization in the conditional rate-distortion function in (12.12).

We now introduce the rates-distortions-tilted information density for the lossy Gray-Wyner problem. Since $\mathcal{R}(P_{XY}, D_1, D_2)$ is a convex set [43], the minimization in (12.16) is attained when $R_1 = R_{X|W}(P_{XW}, D_1)$ and $R_2 = R_{Y|W}(P_{YW}, D_2)$ for some optimal test channel $P_{W|XY}$ unless $\mathsf{R}_0(R_1, R_2|P_{XY}, D_1, D_2) = 0$ or ∞. To avoid degenerate cases, assume that $\mathsf{R}_0(R_1, R_2|P_{XY}, D_1, D_2) > 0$ is finite and $\mathcal{R}(P_{XY}, D_1, D_2)$ is smooth at a boundary rate triplet (R_0^*, R_1^*, R_2^*) of our interest, i.e.,

$$\xi_i^* := -\frac{\partial \mathsf{R}_0(R_1, R_2|P_{XY}, D_1, D_2)}{\partial R_i}\Big|_{(R_1,R_2)=(R_1^*,R_2^*)}, \tag{12.18}$$

$$\lambda_i^* := -\frac{\partial \mathsf{R}_0(R_1, R_2|P_{XY}, D_1', D_2')}{\partial D_i'}\Big|_{(D_1',D_2')=(D_1,D_2)}, \tag{12.19}$$

are well-defined for $i \in [2]$[1]. Note that $\xi_i^*,\ \lambda_i^* \geq 0$ since $\mathsf{R}_0(R_1, R_2|P_{XY}, D_1, D_2)$ is non-increasing in (R_1, R_2, D_1, D_2). Assume that all derivatives $(\xi_1^*, \xi_2^*, \lambda_1^*, \lambda_2^*)$ are strictly positive, which holds for all rate triplets (R_0^*, R_1^*, R_2^*) such that $R_0^* = \mathsf{R}_0(R_1^*, R_2^*|P_{XY}, D_1, D_2)$ is positive and finite.

Let $(P_{W|XY}^* P_{\hat{X}|XW}^* P_{\hat{Y}|YW}^*)$ be a tuple of optimal test channels[2] that achieves $\mathsf{R}_0(R_1^*, R_2^*|P_{XY}, D_1, D_2)$ in (12.17). Let $P_{\hat{X}|W}^*, P_{\hat{Y}|W}^*, P_W^*$ be the induced (conditional) distributions. Given any $(x, y, w) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{W}$, define the following two conditional distortion-tilted information densities:

$$\jmath(x, D_1|w) := \log \frac{1}{\sum_{\hat{x}} P_{\hat{X}|W}^*(\hat{x}|w) \exp\left(\frac{\lambda_1^*}{\xi_1^*}(D_1 - d_1(x, \hat{x}))\right)}, \tag{12.20}$$

$$\jmath(y, D_2|w) := \log \frac{1}{\sum_{\hat{y}} P_{\hat{Y}|W}^*(\hat{y}|w) \exp\left(\frac{\lambda_2^*}{\xi_2^*}(D_2 - d_2(y, \hat{y}))\right)}. \tag{12.21}$$

The rates-distortions-tilted information density for the lossy Gray-Wyner problem is defined as follows.

Definition 12.3. For a boundary rate triplet (R_0^*, R_1^*, R_2^*), given any (D_1, D_2), the rates-distortions-tilted information density for lossy Gray-Wyner source coding is defined as

[1] Due to these regularity conditions, our result in Section 2 does not hold for some singular points (e.g., where the derivatives do not exist) of the rate-distortion region, as in the lossless case by Watanabe in [127].

[2] The following tilted information density is still well-defined even if the optimal test channel is not unique due to similar arguments as [127, Lemma 2].

$$\jmath(x,y|R_1^*,R_2^*,D_1,D_2) := -\log\bigg(\sum_w P_W^*(w)\exp\Big(\xi_1^*(R_1^* - \jmath(x,D_1|w)) + \xi_2^*(R_2^* - \jmath(y,D_2|w))\Big)\bigg). \quad (12.22)$$

Recall that there are two equivalent characterizations of the Gray-Wyner region, one defined in terms of conditional rate-distortion functions in Theorem 12.1 and the other defined solely in terms of (conditional) mutual information quantities in (12.14). For the lossless Gray-Wyner problem [127], the two regions are exactly the same. The tilted information densities derived based on these two regions are subtly different. We find that the tilted information density derived from the second region in (12.14) is more amenable to subsequent second-order analyses on the Pangloss plane (Lemma 12.6). Thus the "correct" non-asymptotic fundamental quantity for the lossy Gray-Wyner problem is the rates-distortions-tilted information density in (12.22).

The rates-distortions-tilted information density for lossy Gray-Wyner source coding has the following properties.

Lemma 12.2. The following properties hold.

1. The minimal common rate function equals the following expectation of the rate-distortions-tilted information density, i.e.,

$$\begin{aligned}&\mathsf{R}_0(R_1^*,R_2^*|P_{XY},D_1,D_2)\\&=\mathbb{E}_{P_{XY}}\left[\jmath(X,Y|R_1^*,R_2^*,D_1,D_2,P_{XY})\right],\end{aligned} \quad (12.23)$$

2. For $(w,\hat{x},\hat{y})$ such that $P_W^*(w)P_{\hat{X}|W}^*(\hat{x}|w)P_{\hat{Y}|W}^*(\hat{y}|w) > 0$,

$$\begin{aligned}&\jmath(x,y|R_1^*,R_2^*,D_1,D_2,P_{XY})\\&=\log\frac{P_{W|XY}^*(w|x,y)}{P_W^*(w)}+\xi_1^*\log\frac{P_{\hat{X}|XW}^*(\hat{x}|x,w)}{P_{\hat{X}|W}^*(\hat{x}|w)}-\xi_1^*R_1^*\\&\quad+\xi_2^*\log\frac{P_{\hat{Y}|YW}^*(\hat{y}|y,w)}{P_{\hat{Y}|W}^*(\hat{y}|w)}-\xi_2^*R_2^*\\&\quad+\lambda_1^*(d_1(x,\hat{x})-D_1)+\lambda_2^*(d_2(y,\hat{y})-D_2).\end{aligned} \quad (12.24)$$

Lemma 12.2 generalizes [127, Lemma 1] for the lossless Gray-Wyner problem and [20, Lemma 1.4] for the rate-distortion problem.

In the following lemma, we relate the derivative of the minimum common rate function with the rates-distortions-tilted information density. Recall that given a joint probability distribution $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, $m = |\operatorname{supp}(P_{XY})|$ and $\Gamma(P_{XY})$ be the sorted distribution such that for each $i \in [m]$, $\Gamma_i(P_{XY}) = P_{XY}(x_i, y_i)$ is the i-th largest value of $\{P_{XY}(x,y) : (x,y) \in \mathcal{X} \times \mathcal{Y}\}$. For any Q_{XY}, let $Q^*_{W|XY} Q^*_{\hat{X}|XW} Q^*_{\hat{Y}|YW}$ be the optimal test channel for $\mathsf{R}_0(R_1^*, R_2^* | \Gamma(Q_{XY}), D_1, D_2)$ in (12.17). Let $Q^*_W, Q^*_{\hat{X}|W}, Q^*_{\hat{Y}|W}$ be the corresponding induced distributions.

Lemma 12.3. Suppose that for all Q_{XY} in some neighborhood of P_{XY}, $\operatorname{supp}(Q^*_W) \subset \operatorname{supp}(P^*_W)$, $\operatorname{supp}(Q^*_{\hat{X}|W}) \subset \operatorname{supp}(P^*_{\hat{X}|W})$ and $\operatorname{supp}(Q^*_{\hat{Y}|W}) \subset \operatorname{supp}(P^*_{\hat{X}|W})$. Then for $i \in [1 : m-1]$,

$$\begin{aligned}
&\frac{\partial \mathsf{R}_0(R_1^*, R_2^* | \Gamma(Q_{XY}), D_1, D_2)}{\partial \Gamma_i(Q_{XY})}\bigg|_{Q_{XY} = P_{XY}} \\
&= \jmath(x_i, y_i | R_1^*, R_2^*, D_1, D_2, \Gamma(P_{XY})) \\
&\quad - \jmath(x_m, y_m | R_1^*, R_2^*, D_1, D_2, \Gamma(P_{XY})). \qquad (12.25)
\end{aligned}$$

Lemma 12.3 generalizes [127, Lemma 3] for the lossless Gray-Wyner problem and [66, Theorem 2.2] for the rate-distortion problem.

12.3 Second-Order Asymptotics

12.3.1 Result

Let (R_0^*, R_1^*, R_2^*) be a boundary rate triplet on the rate-distortion region of the lossy Gray-Wyner problem.

Definition 12.4. Given any $\varepsilon \in (0,1)$, a triplet (L_0, L_1, L_2) is said to be second-order $(R_0^*, R_1^*, R_2^*, D_1, D_2, \varepsilon)$-achievable if there exists a sequence of (n, M_0, M_1, M_2)-codes such that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} (\log M_0 - nR_0) \le L_0, \tag{12.26}$$
$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} (\log M_1 - nR_1) \le L_1, \tag{12.27}$$
$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} (\log M_2 - nR_2) \le L_2, \tag{12.28}$$

and

$$\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1, D_2) \le \varepsilon. \tag{12.29}$$

The closure of the set of all second-order $(R_0^*, R_1^*, R_2^*, D_1, D_2, \varepsilon)$-achievable triplets is called the second-order coding region and denoted as $\mathcal{L}(R_0^*, R_1^*, R_2^*, D_1, D_2, \varepsilon)$.

Note that in Definition 12.2 of the rate-distortion region, the expected distortion measure is considered, whereas in Definition 12.4, the excess-distortion probability is considered. This is consistent with other lossy source coding problems studied in previous sections and the joint-excess-distortion probability allows us to derive second-order asymptotics that provides deeper understanding of the tradeoff among encoders beyond the rate-distortion region.

Let the rates-distortions-dispersion function be

$$\begin{aligned}&\mathrm{V}(R_1^*, R_2^*|P_{XY}, D_1, D_2)\\ &:= \mathrm{Var}\left[\jmath(X, Y|R_1^*, R_2^*, D_1, D_2, P_{XY})\right].\end{aligned} \tag{12.30}$$

For any boundary rate triplet $(R_0^*, R_1^*, R_2^*) \in \mathcal{R}(P_{XY}, D_1, D_2)$, we impose the following conditions:

1. $R_0^* = \mathsf{R}_0(R_1^*, R_2^*|P_{XY}, D_1, D_2)$ is positive and finite;
2. For $i \in [2]$, the derivatives ξ_i in (12.18) and λ_i^* in (12.19) are well-defined and positive;
3. $(R_1, R_2, Q_{XY}) \mapsto \mathsf{R}_0(R_1, R_2|Q_{XY}, D_1, D_2)$ is twice differentiable in the neighborhood of (R_1^*, R_2^*, P_{XY}) and the derivatives are bounded;
4. The dispersion function $\mathrm{V}(R_1^*, R_2^*|P_{XY}, D_1, D_2)$ is finite.

Theorem 12.4. Under conditions (1) to (3), given any $\varepsilon \in (0,1)$, the second-order coding region satisfies

$$\begin{aligned}&\mathcal{L}(R_0^*, R_1^*, R_2^*, D_1, D_2, \varepsilon)\\&= \Big\{(L_0, L_1, L_2) : L_0 + \xi_1^* L_1 + \xi_2^* L_2\\&\qquad\qquad \geq \sqrt{\mathrm{V}(R_1^*, R_2^* | P_{XY}, D_1, D_2)}\mathrm{Q}^{-1}(\varepsilon)\Big\}. \quad (12.31)\end{aligned}$$

Theorem 12.4 is proved in Section 12.4. In the achievability proofs, we derive a type covering lemma (cf. Lemma 12.7) designed specifically for the lossy Gray-Wyner source coding problem. While the proof of this type covering lemma itself hinges on various other works, e.g., [79], [82], [127], piecing the ingredients together and ensuring that the resultant asymptotic results are tight is non-trivial. One of the main challenges here in proving the type covering lemma is the requirement to establish the uniform continuity of the conditional rate-distortion function in *both* the source distribution and distortion level. The converse proof is done similarly to the successive refinement or the Fu-Yeung problem where we first derive a type-based strong converse, then use Taylor expansions of the minimal common rate function of empirical distributions and finally apply the Berry-Esseen theorem (cf. Theorem 1.3).

12.3.2 Specialization to the Pangloss Plane

In general, it is not easy to calculate $\mathcal{L}(R_0^*, R_1^*, R_2^*, D_1, D_2, \varepsilon)$. Here we consider calculating $\mathcal{L}(R_0^*, R_1^*, R_2^*, D_1, D_2, \varepsilon)$ for a rate triplet (R_0^*, R_1^*, R_2^*) on the Pangloss plane [43]. It is shown in Theorem 6 in [43] that (R_0, R_1, R_2) is (D_1, D_2)-achievable if

$$R_0 + R_1 + R_2 \geq R(P_{XY}, D_1, D_2), \quad (12.32)$$

$$R_0 + R_1 \geq R(P_X, D_1), \quad (12.33)$$

$$R_0 + R_2 \geq R(P_Y, D_2), \quad (12.34)$$

where $R(P_X, D_1)$, $R(P_Y, D_2)$ are rate-distortion functions (cf. (3.7)) and $R(P_{XY}, D_1, D_2)$ is the following joint rate-distortion function

$$\begin{aligned}&R(P_{XY}, D_1, D_2)\\&:= \min_{P_{\hat{X}\hat{Y}|XY} : \mathbb{E}[d_1(X,\hat{X})] \leq D_1,\ \mathbb{E}[d_2(Y,\hat{Y})] \leq D_2} I(P_{XY}, P_{XY|\hat{X}\hat{Y}}). \quad (12.35)\end{aligned}$$

The set of (D_1,D_2)-achievable rate triplets (R_0,R_1,R_2) satisfying $R_0+R_1+R_2=R(P_{XY},D_1,D_2)$ is called the Pangloss plane, denoted as $\mathcal{R}_{\rm pgp}(P_{XY},D_1,D_2)$, i.e.,

$$\begin{aligned}&\mathcal{R}_{\rm pgp}(P_{XY},D_1,D_2)\\&:=\Big\{(R_0,R_1,R_2):(R_0,R_1,R_2)\in\mathcal{R}(P_{XY},D_1,D_2),\\&\qquad\qquad R_0+R_1+R_2=R(P_{XY},D_1,D_2)\Big\}.\end{aligned}\tag{12.36}$$

Let $P^*_{\hat{X}\hat{Y}|XY}$ be an optimal conditional distribution that achieves $R(P_{XY},D_1,D_2)$. Let $P^*_{\hat{X}\hat{Y}}$ be induced by $P^*_{\hat{X}\hat{Y}|XY}$ and P_{XY}. Define the following distortions-tilted information density:

$$\begin{aligned}&\imath_{XY}(x,y|P_{XY},D_1,D_2)\\&:=-\log\mathbb{E}_{P^*_{\hat{X}\hat{Y}}}\left[\exp\left(\nu_1^*(D_1-d_1(x,\hat{X}))+\nu_2^*(D_2-d_2(y,\hat{Y}))\right)\right],\end{aligned}\tag{12.37}$$

where

$$\nu_1^*:=-\frac{\partial R(P_{XY},D,D_2)}{\partial D}\bigg|_{D=D_1},\tag{12.38}$$

$$\nu_2^*:=-\frac{\partial R(P_{XY},D_1,D)}{\partial D}\bigg|_{D=D_2}.\tag{12.39}$$

Lemma 12.5. The properties of $\imath_{XY}(\cdot|P_{XY},D_1,D_2)$ include

- The joint rate-distortion function is the expectation of the joint tilted information density, i.e.,

$$R(P_{XY},D_1,D_2)=\mathbb{E}_{P_{XY}}\left[\imath_{XY}(X,Y|P_{XY},D_1,D_2)\right].\tag{12.40}$$

- For each $(\hat{x},\hat{y})\in\mathrm{supp}(P_{\hat{X}\hat{Y}})^*$,

$$\begin{aligned}\imath_{XY}(x,y|P_{XY},D_1,D_2)&=\log\frac{P^*_{\hat{X}\hat{Y}|XY}(\hat{x},\hat{y}|x,y)}{P^*_{\hat{X}\hat{Y}}(\hat{x},\hat{y})}\\&\quad+\nu_1^*(d_1(x,\hat{x})-D_1)+\nu_2^*(d_2(y,\hat{y})-D_2).\end{aligned}\tag{12.41}$$

Lemma 12.5 can be proved similarly to [127, Lemma 1] for the lossless Gray-Wyner problem and [20, Lemma 1.4] for the rate-distortion problem. By considering a fixed rate triplet on the Pangloss plane, we can relate $\jmath(x, y|R_1^*, R_2^*, D_1, D_2, P_{XY})$ to $\imath_{XY}(x, y|P_{XY}, D_1, D_2)$.

Lemma 12.6. When a boundary rate-triple lies in the Pangloss plane, i.e., $(R_0^*, R_1^*, R_2^*) \in \mathcal{R}_{\mathrm{pgp}}(P_{XY}, D_1, D_2)$ and the common rate $R_0^* > 0$,

$$\begin{aligned}&\jmath(x, y|R_1^*, R_2^*, D_1, D_2, P_{XY})\\&= \imath(x, y|D_1, D_2, P_{XY}) - R_1^* - R_2^*.\end{aligned} \tag{12.42}$$

The proof of Lemma 12.6 invokes Lemma 12.2. Besides, we use an idea from [125] in which it was shown that the following Markov chains hold for the optimal test channels $P_{W|XY}^*$ achieving $\mathcal{R}(R_1^*, R_2^*|P_{XY}, D_1, D_2)$ and $P_{\hat{X}|XW}^*$ as well as $P_{\hat{Y}|YW}^*$ achieving conditional rate-distortion functions $R_{X|W}(P_{XW}^*, D_1)$ and $R_{Y|W}(P_{YW}^*, D_2)$:

$$\hat{X} \to W \to \hat{Y}, \tag{12.43}$$

$$(X, Y) \to (\hat{X}, \hat{Y}) \to W, \tag{12.44}$$

$$\hat{X} \to (X, Y, W) \to \hat{Y}, \tag{12.45}$$

$$\hat{X} \to (X, W) \to Y, \tag{12.46}$$

$$\hat{Y} \to (Y, W) \to X. \tag{12.47}$$

Invoking Lemma 12.6, for a rate triplet (R_0^*, R_1^*, R_2^*) on the Pangloss plane, the expression of the second-order coding region is simplified as follows.

Proposition 12.1. When $(R_0^*, R_1^*, R_2^*) \in \mathcal{R}_{\mathrm{pgp}}(P_{XY}, D_1, D_2)$ and the conditions in Theorem 12.4 are satisfied, we have

$$\begin{aligned}&\mathcal{L}(R_0^*, R_1^*, R_2^*, D_1, D_2, \varepsilon)\\&= \Big\{(L_0, L_1, L_2) : L_0 + L_1 + L_2\\&\qquad \geq \sqrt{\mathrm{V}(R_1^*, R_2^*|P_{XY}, D_1, D_2)}\mathrm{Q}^{-1}(\varepsilon)\Big\},\end{aligned} \tag{12.48}$$

where the rate-dispersion function [70] is

$$\begin{aligned}&\mathrm{V}(R_1^*, R_2^*|P_{XY}, D_1, D_2)\\&= \mathrm{Var}[\jmath(X, Y|R_1^*, R_2^*, D_1, D_2, P_{XY})] \qquad (12.49)\\&= \mathrm{Var}[\imath(X, Y|D_1, D_2, P_{XY})]. \qquad (12.50)\end{aligned}$$

12.3.3 A Numerical Example for the Pangloss Plane

Consider a doubly symmetric binary source (DSBS), where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $P_{XY}(0,0) = P_{XY}(1,1) = \frac{1-p}{2}$ and $P_{XY}(0,1) = P_{XY}(1,0) = \frac{p}{2}$ for $p \in [0, \frac{1}{2}]$. We consider $\hat{\mathcal{X}} = \hat{\mathcal{Y}} = \{0, 1\}$ and Hamming distortion for both sources, i.e., $d_1(x, \hat{x}) = \mathbb{1}(x = \hat{x})$ and $d_2(y, \hat{y}) = \mathbb{1}(y = \hat{y})$. Furthermore, let $R_1 = R_2 = R$ and $D_1 = D_2 = D$. Recall that $H_{\mathrm{b}}(\delta) = -\delta \log(\delta) - (1-\delta)\log(1-\delta)$ is the binary entropy function. Define $f(x) := -x \log x$. Let $p_1 := \frac{1}{2} - \frac{1}{2}\sqrt{1-2p}$. It follows from [9, Exercise 2.7.2] that

$$R(P_{XY}, D, D) = \begin{cases} 1 + H_{\mathrm{b}}(p) - 2H_{\mathrm{b}}(D) & 0 \leq D \leq p_1, \\ f(1-p) - \frac{1}{2}\left(f(2D-p) + f(2(1-D)-p)\right) & p_1 \leq D \leq \frac{1}{2}. \end{cases} \tag{12.51}$$

It was shown in [43, Example 2.5(A)] that for $0 \leq D \leq \Delta \leq p_1$, if $R_0 = R(P_{XY}, \Delta, \Delta)$, $R_1 = R_2 = H_{\mathrm{b}}(\Delta) - H_{\mathrm{b}}(D)$, then $(R_0, R_1, R_2) \in \mathcal{R}_{\mathrm{pgp}}(P_{XY}, D, D)$. When $D \leq p_1$, the joint (D, D)-tilted information density satisfies

$$\jmath_{XY}(0, 0|P_{XY}, D, D) = \jmath_{XY}(1, 1|P_{XY}, D, D) \tag{12.52}$$

$$= \log \frac{1}{(2p-1)D - (2p-1)D^2 + \frac{1}{2}(1-p)} - 2H_{\mathrm{b}}(D), \tag{12.53}$$

$$\jmath_{XY}(0, 1|P_{XY}, D, D) = \jmath_{XY}(1, 0|P_{XY}, D, D) \tag{12.54}$$

$$= \log \frac{1}{(2p-1)D^2 - (2p-1)D + \frac{1}{2}p} - 2H_{\mathrm{b}}(D). \tag{12.55}$$

Hence, the joint dispersion function satisfies

$$\begin{aligned}
&\mathrm{Var}[\imath_{XY}(X,Y|P_{XY},D,D)] \\
&= \sum_{x,y} P_{XY}(x,y)\left(\imath_{XY}(x,y|P_{XY},D,D) - R(P_{XY},D,D)\right)^2 && (12.56) \\
&= (1-p)\left(\log\frac{1}{(2p-1)D-(2p-1)D^2+\frac{1}{2}(1-p)} - 1 - H_{\mathrm{b}}(p)\right)^2 \\
&\quad + p\left(\log\frac{1}{(2p-1)D^2-(2p-1)D+\frac{1}{2}p} - 1 - H_{\mathrm{b}}(p)\right)^2. && (12.57)
\end{aligned}$$

12.4 Proof of Second-Order Asymptotics

12.4.1 Achievability

We first prove that for any given joint type $Q_{XY} \in \mathcal{P}_n(\mathcal{X}\times\mathcal{Y})$, there exists an (n, M_0, M_1, M_2)-code such that the excess-distortion probability is mainly due to the incorrect decoding of side information W. To do so, we present a novel type covering lemma for the lossy Gray-Wyner problem. Using this result, we then prove an upper bound of the excess-distortion probability for the (n, M_0, M_1, M_2)-code. Finally, we establish the achievable second-order coding region by estimating this probability.

Define four constants

$$c_0 = (3|\mathcal{X}||\mathcal{Y}||\mathcal{W}| + 4), \tag{12.58}$$

$$c_0' = c_0 + |\mathcal{X}||\mathcal{Y}|, \tag{12.59}$$

$$c_1 = \left(\frac{11\overline{d}_1}{\underline{d}_1}|\mathcal{X}||\mathcal{Y}||\mathcal{W}| + 3|\mathcal{X}||\mathcal{W}||\hat{\mathcal{X}}| + 5\right), \tag{12.60}$$

$$c_2 = \left(\frac{11\overline{d}_2}{\underline{d}_2}|\mathcal{X}||\mathcal{Y}||\mathcal{W}| + 3|\mathcal{Y}||\mathcal{W}||\hat{\mathcal{Y}}| + 5\right). \tag{12.61}$$

The following type covering lemma is critical for second-order analysis for the lossy Gray-Wyner problem.

Lemma 12.7. Let n satisfy $(n+1)^4 > n\log|\mathcal{X}||\mathcal{Y}|$, $\log n \geq \frac{|\mathcal{X}||\mathcal{W}||\hat{\mathcal{X}}|\log|\mathcal{X}|\overline{d}_1}{D_1}$, $\log n \geq \frac{|\mathcal{Y}||\mathcal{W}||\hat{\mathcal{Y}}|\log|\mathcal{Y}|\overline{d}_2}{D_2}$, and $\log n \geq \log\frac{|\hat{\mathcal{X}}|}{|\mathcal{Y}|}$. Given a joint type $Q_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$, for any rate pair $(R_1, R_2) \in \mathbb{R}_+^2$ such that $\mathsf{R}_0(R_1, R_2|Q_{XY}, D_1, D_2)$ is achievable by some test channel, there exists a conditional type $Q_{W|XY} \in \mathcal{V}_n(\mathcal{W}, Q_{XY})$ such that the following holds:

- There exists a set $\mathcal{C}_n \subset \mathcal{T}_{Q_W}$ (Q_W is induced by Q_{XY} and $Q_{W|XY}$) such that
  - For any $(x^n, y^n) \in \mathcal{T}_{Q_{XY}}$, there exists a $w^n \in \mathcal{C}_n$ whose joint type with (x^n, y^n) is Q_{XYW}, i.e., $(x^n, y^n, w^n) \in \mathcal{T}_{Q_{XYW}}$.
  - The size of $\mathcal{C}_n$ is upper bounded by

$$\frac{1}{n} \log |\mathcal{C}_n| \leq \mathsf{R}_0(R_1, R_2|Q_{XY}, D_1, D_2) + c_0 \frac{\log(n+1)}{n}. \tag{12.62}$$

- For each $w^n \in \mathcal{T}_{Q_{W|XY}}(x^n, y^n)$, there exist sets $\mathcal{B}_{\hat{X}}(w^n) \in \hat{\mathcal{X}}^n$ and $\mathcal{B}_{\hat{Y}}(w^n) \in \hat{\mathcal{Y}}^n$ satisfying
  - For each $(x^n, y^n) \in \mathcal{T}_{Q_{XY|W}}(w^n)$, there exists $\hat{x}^n \in \mathcal{B}_{\hat{X}}(w^n)$ and $\hat{y}^n \in \mathcal{B}_{\hat{Y}}(w^n)$ such that $d_1(x^n, \hat{x}^n) \leq D_1$ and $d_2(y^n, \hat{y}^n) \leq D_2$,
  - The sizes of $\mathcal{B}_{\hat{X}}(w^n)$ and $\mathcal{B}_{\hat{Y}}(w^n)$ are upper bounded as

$$\frac{1}{n} \log |\mathcal{B}_{\hat{X}}(w^n)| \leq R_1 + c_1 \frac{\log n}{n}, \tag{12.63}$$

$$\frac{1}{n} \log |\mathcal{B}_{\hat{Y}}(w^n)| \leq R_2 + c_2 \frac{\log n}{n}. \tag{12.64}$$

Lemma 12.7 is proved by combining a few ideas from the literature: a type covering lemma for the conditional rate-distortion problem (modified from Lemma 4.1 in [22] for the standard rate-distortion problem and Lemma 8 in [82] for the successive refinement problem), a type covering lemma for the common side information for the Gray-Wyner problem (Lemma 4 in [127]) and finally, a uniform continuity lemma for the conditional rate-distortion function (modified from [82], [88]). The proof of Lemma 12.7 adopts similar ideas as the proof of the first-order coding region [43] and is available in [150, Appendix F]. The main idea is that we first send the common information via the common link carrying S_0 and then we consider two conditional rate-distortion problems on the two private links carrying S_1, S_2 using the common information as the side information.

Invoking Lemma 12.7, we show that there exists an (n, M_0, M_1, M_2)-code whose excess-distortion probability can be upper bounded as follows. Recall the definitions of c_0' in (12.59), c_1 in (12.60) and c_2 in (12.61). Define three rates

$$R_{0,n} = \frac{1}{n}\log M_0 - c_0' \frac{\log(n+1)}{n}, \tag{12.65}$$

$$R_{1,n} = \frac{1}{n}\log M_1 - c_1 \frac{\log n}{n}, \tag{12.66}$$

$$R_{2,n} = \frac{1}{n}\log M_2 - c_2 \frac{\log n}{n}. \tag{12.67}$$

Lemma 12.8. There exists an (n, M_0, M_1, M_2)-code such that

$$\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \leq \Pr\left\{R_{0,n} < \mathsf{R}_0(R_{1,n}, R_{2,n}|\hat{T}_{X^nY^n}, D_1, D_2)\right\}. \tag{12.68}$$

The proof of Lemma 12.8 is similar to [127, Lemma 5] and available in [150, Appendix J].

Recall the definition of the typical set $\mathcal{A}_n(P_{XY})$ in (9.77) and the result in (9.78) that

$$\Pr\left\{\hat{T}_{X^nY^n} \notin \mathcal{A}_n(P_{XY})\right\} \leq \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2}. \tag{12.69}$$

For a rate triplet (R_0^*, R_1^*, R_2^*) satisfying conditions in Theorem 12.4, let

$$\frac{1}{n}\log M_0 = \mathsf{R}_0(R_1^*, R_2^*|P_{XY}, D_1, D_2) + \frac{L_0}{\sqrt{n}} + c_0' \frac{\log(n+1)}{n}, \tag{12.70}$$

$$\frac{1}{n}\log M_1 = R_1^* + \frac{L_1}{\sqrt{n}} + c_1 \frac{\log n}{n}, \tag{12.71}$$

$$\frac{1}{n}\log M_2 = R_2^* + \frac{L_2}{\sqrt{n}} + c_2 \frac{\log n}{n}. \tag{12.72}$$

It follows that

$$R_{i,n} = R_i^* + \frac{L_i}{\sqrt{n}},\ \ i = 0, 1, 2. \tag{12.73}$$

In subsequent analyses, for ease of notation, we use $\jmath(X_i, Y_i)$ to denote $\jmath(X_i, Y_i|R_1^*, R_2^*, D_1, D_2, P_{XY})$. From the conditions in Theorem 12.4, the second derivatives of the minimal sum rate function

$\mathsf{R}_0(R_1,R_2|P_{XY},D_1,D_2)$ with respect to (R_1,R_2,P_{XY}) are bounded around a neighborhood of (R_1^*,R_2^*,P_{XY}). Hence, for any $\hat{T}_{x^ny^n}\in\mathcal{A}_n(P_{XY})$, for large n, invoking Lemma 12.3 and applying Taylor's expansion for $\mathsf{R}_0(R_{1,n},R_{2,n}|\hat{T}_{x^ny^n},D_1,D_2)$, we obtain:

$$\begin{aligned}
&\mathsf{R}_0(R_{1,n},R_{2,n}|\hat{T}_{x^ny^n},D_1,D_2)\\
&=\mathsf{R}_0(R_1^*,R_2^*|P_{XY},D_1,D_2)-\xi_1^*\frac{L_1}{\sqrt{n}}-\xi_2^*\frac{L_2}{\sqrt{n}}\\
&\qquad+\sum_{i=1}^{m}\Big(\lambda_i(\hat{T}_{x^ny^n})-\lambda_i(P_{XY})\Big)\Big(\jmath(x_i,y_i)-\jmath(x_m,y_m)\Big)\\
&\qquad+O\left(\|\lambda(\hat{T}_{x^ny^n})-\Gamma(P_{XY})\|^2\right)\\
&\qquad+O\left((R_{1,n}-R_1^*)^2+(R_{2,n}-R_2^*)^2\right) &(12.74)\\
&=\mathsf{R}_0(R_1^*,R_2^*|P_{XY},D_1,D_2)-\xi_1^*\frac{L_1}{\sqrt{n}}-\xi_2^*\frac{L_2}{\sqrt{n}}+O\left(\frac{\log n}{n}\right)\\
&\qquad+\sum_{x,y}\Big(\hat{T}_{x^ny^n}(x,y)-P_{XY}(x,y)\Big)\jmath(x,y) &(12.75)\\
&\le\sum_{x,y}Q_{XY}(x,y)\jmath(x,y)-\xi_1^*\frac{L_1}{\sqrt{n}}-\xi_2^*\frac{L_2}{\sqrt{n}}+O\left(\frac{\log n}{n}\right) &(12.76)\\
&=\frac{1}{n}\sum_{i=1}^{n}\jmath(x_i,y_i)-\xi_1^*\frac{L_1}{\sqrt{n}}-\xi_2^*\frac{L_2}{\sqrt{n}}+O\left(\frac{\log n}{n}\right), &(12.77)
\end{aligned}$$

where (12.76) follows from Lemma 12.2 and the definition of the typical set $\mathcal{A}_n(P_{XY})$ in (9.77).

Define $\eta_n=\frac{\log n}{n}$. Invoking Lemma 12.8, we can upper bound the excess-distortion probability as follows:

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1,D_2)\\
&\le\Pr\left\{R_{0,n}<\mathsf{R}_0(R_{1,n},R_{2,n}|\hat{T}_{X^nY^n},D_1,D_2)\right\} &(12.78)\\
&\le\Pr\left\{\hat{T}_{X^nY^n}\in\mathcal{A}_n(P_{XY}),R_{0,n}<\mathsf{R}_0(R_{1,n},R_{2,n}|\hat{T}_{X^nY^n},D_1,D_2)\right\}\\
&\qquad+\Pr\left\{\hat{T}_{X^nY^n}\notin\mathcal{A}_n(P_{XY})\right\} &(12.79)\\
&\le\Pr\left\{R_{0,n}<\frac{1}{n}\sum_{i=1}^{n}\jmath(X_i,Y_i)-\xi_1^*\frac{L_1}{\sqrt{n}}-\xi_2^*\frac{L_2}{\sqrt{n}}+O(\eta_n)\right\}\\
&\qquad+\frac{2|\mathcal{X}||\mathcal{Y}|}{n^2} &(12.80)
\end{aligned}$$

$$= \Pr\left\{\frac{L_0}{\sqrt{n}} + \xi_1^* \frac{L_1}{\sqrt{n}} + \xi_2^* \frac{L_2}{\sqrt{n}} + O(\eta_n) < \frac{1}{n}\sum_{i=1}^{n}\Big(\jmath(X_i,Y_i) - \mathsf{R}_0(R_1^*,R_2^*|P_{XY},D_1,D_2)\Big)\right\} + \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2} \tag{12.81}$$

$$\leq \mathrm{Q}\left(\frac{L_0 + \xi_1^* L_1 + \xi_2^* L_2 + O\left(\sqrt{n}\eta_n\right)}{\sqrt{\mathrm{V}(R_1^*,R_2^*|P_{XY},D_1,D_2)}}\right) + \frac{6\mathrm{T}(R_1^*,R_2^*,D_1,D_2)}{\sqrt{n}\mathrm{V}^{3/2}(R_1^*,R_2^*,D_1,D_2)} + \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2}, \tag{12.82}$$

where (12.82) follows from the Berry-Esseen Theorem and $\mathrm{T}(R_1^*, R_2^*, D_1, D_2)$ is third absolute moment of the rates-distortions-tilted information density $\jmath(X,Y)$, which is finite for DMS from the conditions in Theorem 12.4. Therefore, if (L_0, L_1, L_2) satisfies

$$L_0 + \xi_1^* L_1 + \xi_2^* L_2 \geq \sqrt{\mathrm{V}(R_1^*,R_2^*|P_{XY},D_1,D_2)}\mathrm{Q}^{-1}(\varepsilon), \tag{12.83}$$

then $\limsup_{n\to\infty} \mathrm{P}_{\mathrm{e},n}(D_1,D_2) \leq \varepsilon$.

12.4.2 Converse

We follow the method of types, similar to the proof of the lossless case in [127] and to the converse proof of the successive refinement and Fu-Yeung problem in previous sections. We first establish a type-based strong converse and use it to derive a lower bound on excess-distortion probability $\mathrm{P}_{\mathrm{e},n}(D_1,D_2)$. Subsequently, we use a Taylor expansion and apply the Berry-Esseen Theorem to obtain an outer region expressed essentially using $\mathrm{V}(R_1^*,R_2^*|P_{XY},D_1,D_2)$.

We now consider an (n, M_0, M_1, M_2)-code for the correlated source (X^n, Y^n) with joint distribution $U_{\mathcal{T}_{Q_{XY}}}(x^n,y^n) = |\mathcal{T}_{Q_{XY}}|^{-1}$, the uniform distribution over the type class $\mathcal{T}_{Q_{XY}}$.

Lemma 12.9. If the non-excess-distortion probability satisfies

$$\Pr\left\{d_1(X^n,\hat{X}^n) \leq D_1, d_2(Y^n,\hat{Y}^n) \leq D_2 | (X^n,Y^n) \in \mathcal{T}_{Q_{XY}}\right\} \geq \exp(-n\alpha) \tag{12.84}$$

for some positive number α, then for n large enough such that $\log n \geq \max\{\overline{d}_1, \overline{d}_2\} \log |\mathcal{X}|$, there exists a conditional distribution $Q_{W|XY}$ with $|\mathcal{W}| \leq |\mathcal{X}||\mathcal{Y}| + 2$ such that

$$\frac{1}{n}\log M_0 \geq I(X,Y;W) - \frac{\Big(|\mathcal{X}||\mathcal{Y}|+1\Big)\log(n+1)}{n} - \alpha, \tag{12.85}$$

$$\frac{1}{n}\log M_1 \geq R_{X|W}(Q_{XW}, D_1) - \frac{\log n}{n}, \tag{12.86}$$

$$\frac{1}{n}\log M_2 \geq R_{Y|W}(Q_{YW}, D_2) - \frac{\log n}{n}. \tag{12.87}$$

where $(X,Y,W) \sim Q_{XY} \times Q_{W|XY}$.

The proof of Lemma 12.9 is similar to the lossless Gray-Wyner problem [127, Lemma 6] but we need to also combine this with the (weak) converse proof for lossy Gray-Wyner problem under the expected distortion criterion in [43]. Readers could refer to [150, Appendix K] for details.

We then prove a lower bound on the excess-distortion probability $\mathrm{P}_{\mathrm{e},n}(D_1, D_2)$ in (12.13). Define the constant $c = \frac{|\mathcal{X}||\mathcal{Y}|+2}{n}$ and the three quantities

$$R_{0,n} := \frac{1}{n}\log M_0 + c\frac{\log(n+1)}{n}, \tag{12.88}$$

$$R_{1,n} := \frac{1}{n}\log M_1 + \frac{\log n}{n}, \tag{12.89}$$

$$R_{2,n} := \frac{1}{n}\log M_2 + \frac{\log n}{n}. \tag{12.90}$$

Lemma 12.10. Consider any $n \in \mathbb{N}$ such that $\log n \geq \max\{\overline{d}_1, \overline{d}_2\} \log |\mathcal{X}|$. Any (n, M_0, M_1, M_2)-code satisfies

$$\begin{aligned} &\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\ &\geq \Pr\Big\{R_{0,n} < \mathsf{R}_0(R_{1,n}, R_{2,n}|\hat{T}_{X^nY^n}, D_1, D_2)\Big\} - \frac{1}{n}. \end{aligned} \tag{12.91}$$

The proof of Lemma 12.10 is similar to [127, Lemma 7] and available in [150, Appendix L].

Choose (M_0, M_1, M_2) such that

$$\frac{1}{n}\log M_0 = R_0^* + \frac{L_0}{\sqrt{n}} - c\frac{\log(n+1)}{n}, \tag{12.92}$$

$$\frac{1}{n}\log M_1 = R_1^* + \frac{L_1}{\sqrt{n}} - \frac{\log n}{n}, \tag{12.93}$$

$$\frac{1}{n}\log M_2 = R_2^* + \frac{L_2}{\sqrt{n}} - \frac{\log n}{n}. \tag{12.94}$$

Hence, according to (12.88) to (12.90) in Lemma 12.10, for $i \in [0:2]$,

$$R_{i,n} = R_i^* + \frac{L_i}{\sqrt{n}}. \tag{12.95}$$

Recall that we use $\jmath(X_i, Y_i)$ to denote $\jmath(X_i, Y_i|R_1^*, R_2^*, D_1, D_2, P_{XY})$. Invoking Lemma 12.10, similar to the achievability proof,

$$\begin{aligned}
&\mathrm{P}_{\mathrm{e},n}(D_1, D_2) \\
&\geq \Pr\left\{R_{0,n} < \mathsf{R}_0(R_{1,n}, R_{2,n}|\hat{T}_{X^nY^n}, D_1, D_2)\right\} - \frac{1}{n} &(12.96)\\
&\geq \Pr\left\{R_{0,n} < \mathsf{R}_0(R_{1,n}, R_{2,n}|\hat{T}_{X^nY^n}, D_1, D_2),\ \hat{T}_{X^nY^n} \in \mathcal{A}_n(P_{XY})\right\} \\
&\qquad - \frac{1}{n} &(12.97)\\
&\geq \Pr\left\{R_0^* + \frac{L_0}{\sqrt{n}} < \frac{1}{n}\sum_{i=1}^n \jmath(X_i, Y_i) - \xi_1^*\frac{L_1}{\sqrt{n}} - \xi_2^*\frac{L_2}{\sqrt{n}} + O(\eta_n)\right. \\
&\qquad\qquad\qquad \left.\text{and } \hat{T}_{X^nY^n} \in \mathcal{A}_n(P_{XY})\right\} - \frac{1}{n} &(12.98)\\
&\geq \Pr\left\{R_0^* + \frac{L_0}{\sqrt{n}} < \frac{1}{n}\sum_{i=1}^n \jmath(X_i, Y_i) - \xi_1^*\frac{L_1}{\sqrt{n}} - \xi_2^*\frac{L_2}{\sqrt{n}} + O(\eta_n)\right\} \\
&\qquad - \Pr\left\{\hat{T}_{X^nY^n} \notin \mathcal{A}_n(P_{XY})\right\} - \frac{1}{n} &(12.99)\\
&= \Pr\left\{\frac{L_0}{\sqrt{n}} + \xi_1^*\frac{L_1}{\sqrt{n}} + \xi_2^*\frac{L_2}{\sqrt{n}} + O(\eta_n) < \frac{1}{n}\sum_{i=1}^n \jmath(X_i, Y_i)\right. \\
&\qquad\qquad \left. - \mathsf{R}_0(R_1^*, R_2^*|P_{XY}, D_1, D_2)\right\} - \frac{2|\mathcal{X}||\mathcal{Y}|}{n^2} - \frac{1}{n} &(12.100)
\end{aligned}$$

$$\geq \mathrm{Q}\left(\frac{L_0+\xi_1^*L_1+\xi_2^*L_2+O\left(\sqrt{n}\eta_n\right)}{\sqrt{\mathrm{V}(R_1^*,R_2^*|P_{XY},D_1,D_2)}}\right)-\frac{6\mathrm{T}(R_1^*,R_2^*,D_1,D_2)}{\sqrt{n}\mathrm{V}^{3/2}(R_1^*,R_2^*,D_1,D_2)}$$
$$-\frac{2|\mathcal{X}||\mathcal{Y}|}{n^2}-\frac{1}{n}, \tag{12.101}$$

where (12.99) follows from the fact that $\Pr\{\mathcal{E}\cap\mathcal{F}\}\geq\Pr\{\mathcal{E}\}-\Pr\{\mathcal{F}^c\}$. Hence, if (L_0,L_1,L_2) satisfies

$$L_0+\xi_1^*L_1+\xi_2^*L_2<\sqrt{\mathrm{V}(R_1^*,R_2^*|P_{XY},D_1,D_2)}\mathrm{Q}^{-1}(\varepsilon), \tag{12.102}$$

then $\liminf_{n\to\infty}\mathrm{P}_{\mathrm{e},n}(D_1,D_2)>\varepsilon$. Therefore, for sufficiently large n, any second-order $(R_0^*,R_1^*,R_2^*,D_1,D_2,\varepsilon)$-achievable triplet (L_0,L_1,L_2) must satisfy

$$L_0+\xi_1^*L_1+\xi_2^*L_2\geq\sqrt{\mathrm{V}(R_1^*,R_2^*|P_{XY},D_1,D_2)}\mathrm{Q}^{-1}(\varepsilon). \tag{12.103}$$

13

Reflections, Other Results and Future Directions

13.1 Reflections

In this monograph, we reviewed recent advances in the second-order asymptotics for lossy source coding, which provides approximation to the finite blocklength performance of optimal codes. Specifically, in Section 1, we introduced the notation and critical mathematical background. In Section 2, we illustrated non-asymptotic and second-order asymptotic analyses via lossless source coding. Subsequently, in Section 3 of Part II, we presented the generalization of the results from lossless source coding to the rate-distortion problem of lossy source coding, highlighted the role of the distortion-tilted information density and introduced two proof sketches. One proof method to yield second-order asymptotics is applying the Berry-Esseen theorem to carefully derive non-asymptotic achievability and converse bounds, where the achievability part uses random coding and minimal distortion encoding while the converse part relies on the properties of the distortion-tilted information density. Although this method is simple and elegant, it is not always possible to derive the desired non-asymptotic bounds for multiterminal lossy source coding problems. Thus, we also introduced another proof technique using the method of types, where the achievability part uses the type covering lemma tailored to the rate-distortion problem and the converse part depends on a type-based strong converse analysis. The first proof sketch using the non-asymptotic bounds usually applies to any memoryless source while the method of types is valid only for DMS. In the rest of Part II, the results and proofs for the rate-distortion problem are generalized to account for noisy sources, noisy channels, mismatched compression, sources with memory and variable length compression in Sections 4 to 8.

In Part III, the two proof methods for the rate-distortion problem are generalized in combination to derive non-asymptotic and second-order asymptotic bounds for four multiterminal lossy source coding problems in the increasingly complicated order: the Kaspi problem in Section 9; the successive refinement problem in Section 10; the Fu-Yeung problem in Section 11; and the Gray-Wyner problem in Section 12. For the Kaspi problem, we introduced the distortions-tilted information density, illustrated the role of side information and showed that the conditional rate-distortion problem is a special case of the Kaspi problem. For the successive refinement problem, we defined a rate-distortions-tilted information density, showed its connection to the minimal sum rate subject to the rate of one encoder, demonstrated the tradeoff between second-order coding rates of two encoders, and validated the joint excess-distortion probability as the "correct" performance criterion. For the Fu-Yeung problem, we presented a non-asymptotic converse bound which yielded tight second-order converse result when specializing to the successive refinement problem and presented tight second-order asymptotics for simultaneous lossless and lossy compression. Finally, for the Gray-Wyner problem in which an auxiliary random variable is required in the characterization of the rate-distortion region, we presented a second-order asymptotic result, where the achievability part follows by deriving a type covering lemma tailored to the problem which uses the continuity of conditional rate-distortion function with respect to the distortion level and the distribution.

13.2 Other Results

This monograph mainly focused on fixed-length compression of DMS under bounded distortion measures with the excess-distortion probability as the performance criterion. For GMS under quadratic distortion measures, the second-order asymptotics for the rate-distortion problem was derived by Ingber and Kochman [54, Theorem 2] and by Kostina and Verdú [70, Theorem 40], and the second-order asymptotics for the successive refinement problem was derived by No, *et al.* [82, Theorem 7] and by Zhou *et al.* [151, Theorem 20], and the second-order asymptotics for a Laplacian source under the magnitude-error distortion measure could be derived using the type-covering lemma in [146] for the achievability result and using the non-asymptotic converse bound in [68, Corollary 2]. When the distortion measure is the logarithm loss, the non-asymptotic analysis for the rate-distortion and the multiple descriptions problem was derived by Shkel and Verdú [106] and the successive refinement problem was studied by No [81]. When the excess-distortion probability is replaced by the average distortion, a non-asymptotic analysis of the rate-distortion problem was done by Moulin [80] and by Elkayam and Feder [30].

Besides second-order asymptotics, the large and moderate deviations asymptotic analyses also provide deeper understanding beyond Shannon theory analyses, as illustrated in Figure 2.2 for lossless source coding. For simplicity, we call the rate-distortion function or the rate-distortion region the Shannon limit. Large deviations, also known as the error exponent analysis, focuses on deriving the exponential decay rate of excess-distortion probabilities for rates beyond the Shannon limit in lossy source coding problems. For the rate-distortion problem, the error exponent was derived by Marton for DMS [79], by Ihara and Kubo [53] for GMS under the quadratic distortion measure and by Zhong *et al.* [146] for a Laplacian memoryless source under the magnitude-error distortion measure. For the successive refinement problem with DMS, the error exponent region was derived by Tuncel and Rose [121] under the separate excess-distortion probabilities criterion and by Kanlis and Narayan under the joint excess-distortion probability criterion [55]. For DMS, the error exponent (region) for the Kaspi problem was derived in [147, Theorem 7], for the Fu-Yeung problem was derived in [147, Theorem 16] and for the Gray-Wyner problem was derived in [150, Theorem 12].

Moderate deviations asymptotics [3], [15], [51] compromise between large deviations and second-order asymptotics by deriving the subexponential decay rates, also known as the moderate deviations constants, of excess-distortion probabilities while allowing rates to approach the Shannon limit. The moderate deviations constant for the rate-distortion problem was derived by Tan [111] for DMS. For the successive refinement problem, the moderate deviations constants were derived by Zhou, Tan and Motaini for both DMS and GMS [151, Theorems 6 and 15]. For DMS, the moderate deviations asymptotics was derived for the Kaspi problem in [147, Theorem 8], for the Fu-Yeung problem was derived in [147, Theorem 17] and for the Gray-Wyner problem was derived in [150, Theorem 13].

13.3 Future Directions

We briefly discuss possible future research directions for lossy source coding beyond the results covered in this monograph.

13.3.1 Higher-Order Asymptotics

For the rate-distortion problem and its five generalizations in Part II of this monograph, we present a second-order asymptotic approximation to the finite blocklength performance. It was recently shown by Yavas *et al.* [140] that for channel coding, the third-order asymptotic approximation in the moderate deviations regime could provide a rather accurate approximation to the performance of an optimal code for blocklengths as small as $n = 100$ with error probabilities as small as 10^{-10}. This high-order approximation is of great interest for beyond 5G communication networks where enhanced ultra-reliable and low-latency communication is required. However, to the best of our knowledge, in general, no tight third-order asymptotic results have been established for the rate-distortion problem. It would be worthwhile to derive higher-order asymptotic results to complement the second-order asymptotics for the problems presented in this monograph.

13.3.2 Multiterminal Compression of GMS

Although the second-order asymptotics results of the rate-distortion and the successive refinement problems have been established for GMS under quadratic distortion measures, the second-order asymptotics of many other multiterminal lossy source coding for GMS is generally unknown. For the Kaspi problem, the non-asymptotic converse bound in Section 3 is valid for GMS, but the achievability analysis is non-trivial despite the rate-distortion function was derived by Perron *et al.* [89]. For the multiple descriptions problem [132], the rate-distortion region for GMS was derived by Ozarow [87]. Both achievability and converse analyses of non-asymptotic and second-order asymptotic bounds require novel ideas. For the Gray-Wyner problem, although the rate-distortion region is known [43] and the exact formula for GMS is recently derived by Yu [141], the second-order asymptotics are challenging.

13.3.3 Mismatched Multiterminal Compression

Most contents in this monograph concerned matched compression, where the distribution of the source sequence is assumed perfectly known. Such an assumption is invalid in practice because one is not able to know the exact distribution of a source to be compressed. Thus, it is important to use mismatched coding schemes ignorant of the exact source distribution to compress any memoryless sources. In Section 6, we presented the second-order asymptotics by Zhou, *et al.* [152], who analyzed the mismatched compression scheme proposed by Lapidoth [76, Theorem 3], where the minimum Euclidean distance encoding with the i.i.d. Gaussian codebook is used to compress an arbitrary memoryless source. However, the non-asymptotic and second-order asymptotic analyses for more complicated multiterminal lossy source coding remain largely unexplored. Some attempts have been made very recently by Bai *et al.* [5], Bai *et al.* [6], and Wu *et al.* [133] in the achievability analysis of the mismatched successive refinement problem.

13.3.4 Variable-Length Multiterminal Compression

This monograph focused on fixed-length lossy source coding. Motivated by the need to reduce the codeword length of frequently appeared symbols, fixed-to-variable length (FVL) source coding has also been widely studied for the point-to-point case [47], [57], [67], [96], [97], [99]. In particular, Kostina and Verdú derived the second-order asymptotics for average codeword length of the FVL rate-distortion problem subject to a non-vanishing excess-distortion probability, which was presented in Section 8. Saito *et al.* [96], [97] studied the FVL rate-distortion problem under constraints on both the excess-distortion probability and the excess-length probability. However, no results have been established for FVL multiterminal lossy source coding. It would be worthwhile to derive non-asymptotic and second-order asymptotic bounds on the average codeword length for a multiterminal lossy source coding problem such as the successive refinement problem.

13.3.5 Decoder Side Information Problems

Although we have presented results for several multiterminal lossy source coding problems, many more remain open, such as the Wyner-Ziv problem [136], the Kaspi-Heegard-Berger problem [56, Theorem 2], [52], and the Berger-Tung problem [10]. A common feature of these problems is that in the asymptotic rate-distortion function (region), there exists an auxiliary random variable that forms a Markov chain with the source sequences and/or the side information. For the Wyner-Ziv problem, some attempts in characterizing the second-order asymptotics have been made in the achievability part by Watanabe *et al.* [128] and by Yassaee *et al.* [139] and the converse part by Oohama [85]. However, the achievability and converse bounds do not match even in the sign of the second-order term. Novel ideas and mathematical tools are required to establish second-order asymptotics.

13.3.6 Rate-Distortion-Perception Tradeoff

As evidenced in many applications of image compression, optimal schemes achieving the rate-distortion function lead to low performance due to the ignorance of the distribution of the reproduced sequences. The perceptual quality of an image is shown to be determined by the distribution of the reproduced sequences. However, this information is omitted in the design of codes described in this monograph. To solve this problem, recent studies on rate-distortion-perception tradeoff [13], [116] revisit the rate-distortion problem by constraining that the distribution of the output of the decoder is either identical or approximately identical to the distribution of the source sequence. All these results are asymptotic Shannon theoretical analysis on the rate-distortion function for the point-to-point case. It would be of interest to conduct a non-asymptotic and second-order asymptotic analysis of the rate-distortion-perception problem and also generalize it to multiterminal lossy source coding problems.

Acknowledgements

We are grateful to Lei Yu, Qiaosheng (Eric) Zhang, Zhuangfei Wu and Yiming Wang for proofreading parts of this monograph. We thank Prof. Ioannis Kontoyiannis for comments concerning related works [26], [27], [60], [61] on rate-distortion and mismatched rate-distortion discussed in Sections 3 and 6, respectively. We thank Prof. Masahito Hayashi for drawing our attention to [137], where the assumption of unique variance for the discrete memoryless channel in Section 5 is relaxed.

We thank Prof. Alexander Barg for handling our submission and thank anonymous reviewers for constructive comments and helpful suggestions. We thank the editorial board and staff of the Foundations and Trends in Communications and Information Theory for their help and support.

L. Zhou was supported in part by the National Natural Science Foundation of China under Grant 62201022, in part by the Beijing Natural Science Foundation under Grant 4232007 and in part by a start up grant of Beihang University. M. Motani was supported in part by the Singapore Ministry of Education under Grant MOE2019-T2-2-171.

References

[1] R. Ahlswede, "On multiple descriptions and team guessing," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, 1986, pp. 543–549.

[2] N. Alon and A. Orlitsky, "A lower bound on the expected length of one-to-one codes," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, 1994, pp. 1670–1672.

[3] Y. Altuğ and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, 2014, pp. 4417–4426.

[4] Y. Altuğ, A. B. Wagner, and I. Kontoyiannis, "Lossless compression with moderate error probability," in *IEEE ISIT*, pp. 1744–1748, 2013.

[5] L. Bai, Z. Wu, and L. Zhou, "Achievable second-order asymptotics for successive refinement using Gaussian codebooks," in *IEEE ISIT*, pp. 2888–2893, 2021.

[6] L. Bai, Z. Wu, and L. Zhou, "Achievable refined asymptotics for successive refinement using Gaussian codebooks," *IEEE Trans. Inf. Theory*, vol. 69, no. 6, 2023, pp. 3525–3543.

[7] V. Bentkus, "On the dependence of the Berry-Esseen bound on dimension," *J. Stat. Planning and Inference*, vol. 113, 2003, pp. 385–402.

[8] T. Berger, "Information rates of Wiener processes," *IEEE Trans. Inf. Theory*, vol. 16, no. 2, 1970, pp. 134–139.

[9] T. Berger, *Rate-Distortion Theory*. Wiley Online Library, 1971.

[10] T. Berger, "Multiterminal source coding," *The information theory approach to communications*, 1978.

[11] T. Berger and J. D. Gibson, "Lossy source coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998, pp. 2693–2723.

[12] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Am. Math. Soc.*, vol. 49, no. 1, 1941, pp. 122–136.

[13] Y. Blau and T. Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *International Conference on Machine Learning*, pp. 675–685, 2019.

[14] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[15] J. Chen, D. K. He, A. Jagmohan, and L. A. Lastras-Montano, "On the redundancy-error tradeoff in Slepian-Wolf coding and channel coding," in *IEEE ISIT*, pp. 1326–1330, 2007.

[16] S. Chen, M. Effros, and V. Kostina, "Lossless source coding in the point-to-point, multiple access, and random access scenarios," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, 2020, pp. 6688–6722.

[17] J. Chow and T. Berger, "Failure of successive refinement for symmetric Gaussian mixtures," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, 1997, pp. 350–352.

[18] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd. McGraw-Hill Science/Engineering/Math, 2003.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[20] I. Csiszár, "On an extremum problem of information theory," *Studia Sci. Math. Hungarica*, vol. 9, no. 1, 1974, pp. 57–72.

[21] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998, pp. 2505–2523.

[22] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[23] I. Csiszár and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses," *Studia Sci. Math. Hungarica*, vol. 6, 1971, p. 181.

[24] I. Csiszzár, "Joint source-channel error exponent," *Prob. Contr. & Info. Theory*, vol. 9, no. 5, 1980, pp. 315–328.

[25] L. D. Davisson, "Rate-distortion theory and application," *Proceedings of the IEEE*, vol. 60, no. 7, 1972, pp. 800–808.

[26] A. Dembo and I. Kontoyiannis, "Critical behavior in lossy source coding," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, 2001, pp. 1230–1236.

[27] A. Dembo and I. Kontoyiannis, "Source coding, large deviations, and approximate pattern matching," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, 2002, pp. 1590–1615.

[28] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, vol. 38. Springer, 2009.

[29] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Trans. Inf. Theory*, vol. 8, no. 5, 1962, pp. 293–304.

[30] N. Elkayam and M. Feder, "One shot approach to lossy source coding under average distortion constraints," in *IEEE ISIT*, pp. 2389–2393, 2020.

[31] W. H. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, 1991, pp. 269–275.

[32] C. G. Esseen, *On the Liapounoff limit of error in the theory of probability*. Almqvist & Wiksell, 1942.

[33] W. Feller, "Law of large numbers for identically distributed variables," *An introduction to probability theory and its applications*, vol. 2, 1971, pp. 231–234.

[34] F. W. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, 2002, pp. 2012–2021.

[35] R. Gallager, "The random coding bound is tight for the average code (corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, 1973, pp. 244–246.

[36] R. G. Gallager, "Source coding with side information and universal coding," LIDS, MIT, Tech. Rep., 1976.

[37] A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, 1982, pp. 851–857.

[38] A. E. Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[39] W. Gao, Y. H. Liu, C. Wang, and S. Oh, "Rate distortion for model compression: From theory to practice," in *International Conference on Machine Learning*, pp. 2102–2111, 2019.

[40] F. Gotze, "On the rate of convergence in the multivariate CLT," *Ann. Probab.*, 1991, pp. 724–739.

[41] R. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, 1970, pp. 412–421.

[42] R. Gray, "Rate distortion functions for finite-state finite-alphabet Markov sources," *IEEE Trans. Inf. Theory*, vol. 17, no. 2, 1971, pp. 127–134.

[43] R. Gray and A. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, 1974, pp. 1681–1721.

[44] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998, pp. 2325–2383.

[45] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *IEEE ISIT*, pp. 2316–2320, 2009.

[46] A. Habibian, T. v. Rozendaal, J. M. Tomczak, and T. S. Cohen, "Video compression with rate-distortion autoencoders," in *International Conference on Computer Vision*, pp. 7033–7042, 2019.

[47] T. S. Han, "Weak variable-length source coding," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, 2000, pp. 1217–1226.

[48] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, 2003.

[49] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, 2008, pp. 4619–4637.

[50] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, 2009, pp. 4947–4966.

[51] D. K. He, L. A. Lastras-Montaňo, E. H. Yang, A. Jagmohan, and J. Chen, "On the redundancy of Slepian–Wolf coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, 2009, pp. 5607–5627.

[52] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, 1985, pp. 727–734.

[53] S. Ihara and M. Kubo, "Error exponent for coding of memoryless Gaussian sources with a fidelity criterion," *IEICE Trans. Fundamentals*, vol. 83, no. 10, 2000, pp. 1891–1897.

[54] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *IEEE DCC*, pp. 53–62, 2011.

[55] A. Kanlis and P. Narayan, "Error exponents for successive refinement by partitioning," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, 1996, pp. 275–282.

[56] A. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, 1994, pp. 2031–2034.

[57] H. Koga and H. Yamamoto, "Asymptotic properties on codeword lengths of an optimal FV code for general sources," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, 2005, pp. 1546–1555.

[58] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf. Theory*, vol. 2, no. 4, 1956, pp. 102–108.

[59] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, 2014, pp. 777–795.

[60] I. Kontoyiannis and R. Zamir, "Mismatched codebooks and the role of entropy coding in lossy data compression," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, 2006, pp. 1922–1938.

[61] I. Kontoyiannis and J. Zhang, "Arbitrary source models and Bayesian codebooks in rate-distortion theory," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, 2002, pp. 2276–2290.

[62] I. Kontoyiannis, "Pointwise redundancy in lossy data compression and universal lossy data compression," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, 2000, pp. 136–152.

[63] V. N. Koshelev, "Hierarchical coding of discrete sources," *Probl. Peredachi Inf.*, vol. 16, no. 3, 1980, pp. 31–49.

[64] V. N. Koshelev, "Estimation of mean error for a discrete successive-approximation scheme," *Probl. Peredachi Inf.*, vol. 17, no. 3, 1981, pp. 20–33.

[65] V. N. Koshelev, "On the divisibility of discrete sources with an additive single-letter distortion measure," *Probl. Peredachi Inf.*, vol. 30, no. 1, 1994, pp. 31–50.

[66] V. Kostina, "Lossy data compression: Non-asymptotic fundamental limits," Ph.D. dissertation, Department of Electrical Engineering, Princeton University, 2013.

[67] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, 2015, pp. 4316–4330.

[68] V. Kostina and E. Tuncel, "Successive refinement of abstract sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, 2019, pp. 6385–6398.

[69] V. Kostina and S. Verdú, "A new converse in rate-distortion theory," in *Conference on Information Sciences and Systems*, pp. 1–6, 2012.

[70] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, 2012, pp. 3309–3338.

[71] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, 2013, pp. 2545–2575.

[72] V. Kostina and S. Verdú, "Nonasymptotic noisy lossy source coding," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, 2016, pp. 6111–6123.

[73] O. Kosut and L. Sankar, "Asymptotics and non-asymptotics for universal fixed-to-variable source coding," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, 2017, pp. 3757–3772.

[74] S. Kuzuoka and S. Watanabe, "An information-spectrum approach to weak variable-length source coding with side-information," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, 2015, pp. 3559–3573.

[75] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, 1996, pp. 1520–1529.

[76] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, 1997, pp. 38–47.

[77] S. Q. Le, V. Y. F. Tan, and M. Motani, "Second-order coding rates for conditional rate-distortion," *arXiv:1410.2687*, 2014.

[78] S. Q. Le, V. Y. F. Tan, and M. Motani, "A case where interference does not affect the channel dispersion," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, 2015, pp. 2439–2453.

[79] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, 1974, pp. 197–199.

[80] P. Moulin, "Lower bounds on rate of fixed-length source codes under average- and 6-fidelity constraints," in *IEEE ISIT*, pp. 3220–3224, 2017.

[81] A. No, "Universality of logarithmic loss in successive refinement," *Entropy*, vol. 21, no. 2, 2019.

[82] A. No, A. Ingber, and T. Weissman, "Strong successive refinability and rate-distortion-complexity tradeoff," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, 2016, pp. 3618–3635.

[83] Y. Oohama, "Distributed source coding of correlated Gaussian remote sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, 2012, pp. 5059–5085.

[84] Y. Oohama, "Indirect and direct Gaussian distributed source coding problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, 2014, pp. 7506–7539.

[85] Y. Oohama, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, 2018.

[86] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, 1998, pp. 23–50.

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

[88] H. Palaiyanur and A. Sahai, "On the uniform continuity of the rate-distortion function," in *IEEE ISIT*, pp. 857–861, 2008.

[89] E. Perron, S. Diggavi, and E. Telatar, "The Kaspi rate-distortion problem with encoder side-information: Gaussian case," EPFL, Tech. Rep., 2005.

[90] E. Perron, S. Diggavi, and E. Telatar, "The Kaspi rate-distortion problem with encoder side-information: Binary erasure case," EPFL, Tech. Rep., 2006.

[91] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Department of Electrical Engineering, Princeton University, 2010.

[92] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, 2010, pp. 2307–2359.

[93] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," vol. 6, no. 2012-2016, 2014, p. 7.

[94] M. Raič, "A multivariate Berry–Esseen theorem with explicit constants," *Bernoulli*, vol. 25, no. 4A, 2019, pp. 2824–2853.

[95] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, 1994, pp. 253–259.

[96] S. Saito, H. Yagi, and T. Matsushima, "Variable-length lossy compression allowing positive overflow and excess distortion probabilities," in *IEEE ISIT*, pp. 1568–1572, 2017.

[97] S. Saito, H. Yagi, and T. Matsushima, "New results on variable-length lossy compression allowing positive overflow and excess distortion probabilities," in *IEEE ISIT*, pp. 359–363, 2018.

[98] Y. Sakai and V. Y. F. Tan, "Variable-length source dispersions differ under maximum and average error criteria," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, 2020, pp. 7565–7587.

[99] Y. Sakai, R. C. Yavas, and V. Y. F. Tan, "Third-order asymptotics of variable-length compression allowing errors," *IEEE Trans. Inf. Theory*, vol. 67, no. 12, 2021, pp. 7708–7722.

[100] D. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, 1968, pp. 165–167.

[101] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, 2013, pp. 838–852.

[102] J. Scarlett, A. G. i. Fàbregas, A. Somekh-Baruch, and A. Martinez, "Information-theoretic foundations of mismatched decoding," *Foundations and Trends® in Communications and Information Theory*, vol. 17, no. 2–3, 2020, pp. 149–401.

[103] J. Scarlett, V. Y. F. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, 2017, pp. 81–92.

[104] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 1, 1948, pp. 379–423.

[105] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec*, vol. 4, no. 142-163, 1959, p. 1.

[106] Y. Y. Shkel and S. Verdú, "A single-shot approach to lossy source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, 2018, pp. 129–147.

[107] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, 1973, pp. 471–480.

[108] A. J. Stam, "Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces," *J. Appl. Probab.*, vol. 19, no. 1, 1982, pp. 221–228.

[109] V. Strassen, "Asymptotische abschätzungen in Shannons informationstheorie," in *Trans. Third Prague Conf. Information Theory*, pp. 689–723, 1962.

[110] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, 1998, pp. 74–90.

[111] V. Y. F. Tan, "Moderate-deviations of lossy source coding for discrete and Gaussian sources," in *IEEE ISIT*, pp. 920–924, 2012.

[112] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1–2, 2014, pp. 1–184.

[113] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, 2014, pp. 881–903.

[114] V. Y. F. Tan and T. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, 2015, pp. 2430–2438.

[115] H. Tanizaki, *Computational methods in statistics and econometrics.* CRC Press, 2004.

[116] L. Theis and A. B. Wagner, "A coding theorem for the rate-distortion-perception function," *International Conference on Learning Representations neural compression workshop*, 2021.

[117] P. Tian and V. Kostina, "The dispersion of the Gauss–Markov source," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, 2019, pp. 6355–6384.

[118] P. Tian and V. Kostina, "Nonstationary Gauss-Markov processes: Parameter estimation and dispersion," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, 2021, pp. 2426–2449.

[119] M. Tomamichel and V. Y. F. Tan, "Second-order coding rates for channels with state," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, 2014, pp. 4427–4448.

[120] E. Tuncel and K. Rose, "Computation and analysis of the n-layer scalable rate-distortion function," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, 2003, pp. 1218–1230.

[121] E. Tuncel and K. Rose, "Error exponents in scalable source coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, 2003, pp. 289–296.

[122] R. Venkataramanan and S. S. Pradhan, "Multiple descriptions with feed-forward: A single-letter achievable rate region," in *IEEE ISIT*, pp. 692–696, 2008.

[123] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, 2003, pp. 2106–2114.

[124] S. Verdú, “Teaching lossless data compression,” *IEEE Inf. Theory Soc. Newsletter*, vol. 61, no. 1, 2011, pp. 18–19.

[125] K. B. Viswanatha, E. Akyol, and K. Rose, “The lossy common information of correlated sources,” *IEEE Trans. Inf. Theory*, vol. 60, no. 6, 2014, pp. 3238–3253.

[126] D. Wang, A. Ingber, and Y. Kochman, “The dispersion of joint source-channel coding,” in *Proc. 49th Annu. Allerton Conf.*, pp. 180–187, 2011.

[127] S. Watanabe, “Second-order region for Gray-Wyner network,” *IEEE Trans. Inf. Theory*, vol. 63, no. 2, 2017, pp. 1006–1018.

[128] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, “Nonasymptotic and second-order achievability bounds for coding with side-information,” *IEEE Trans. Inf. Theory*, vol. 61, no. 4, 2015, pp. 1574–1605.

[129] T. Weissman, “Universally attainable error exponents for rate-distortion coding of noisy sources,” *IEEE Trans. Inf. Theory*, vol. 50, no. 6, 2004, pp. 1229–1246.

[130] H. Witsenhausen, “Indirect rate distortion problems,” *IEEE Trans. Inf. Theory*, vol. 26, no. 5, 1980, pp. 518–521.

[131] J. Wolf and J. Ziv, “Transmission of noisy information to a noisy receiver with minimum distortion,” *IEEE Trans. Inf. Theory*, vol. 16, no. 4, 1970, pp. 406–411.

[132] J. K. Wolf, A. D. Wyner, and J. Ziv, “Source coding for multiple descriptions,” *Bell Syst. Tech. J.*, vol. 59, no. 8, 1980, pp. 1417–1426.

[133] Z. Wu, L. Bai, and L. Zhou, “Excess-distortion exponents for successive refinement using Gaussian codebooks,” in *IEEE ISIT*, pp. 234–239, 2022.

[134] A. D. Wyner, “Random packings and coverings of the unit n-sphere,” *Bell Syst. Tech. J.*, vol. 46, no. 9, 1967, pp. 2111–2118.

[135] A. D. Wyner, “An upper bound on the entropy series,” *Information and Control*, vol. 20, no. 2, 1972, pp. 176–181.

[136] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, 1976, pp. 1–10.

[137] R. Yaguchi and M. Hayashi, "Second order analysis for joint source-channel coding with general channel and Markovian source," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, 2019, pp. 5750–5770.

[138] E. H. Yang and Z. Zhang, "On the redundancy of lossy source coding with abstract alphabets," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, 1999, pp. 1092–1110.

[139] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *IEEE ISIT*, pp. 1287–1291, 2013.

[140] R. C. Yavas, V. Kostina, and M. Effros, "Third-order analysis of channel coding in the moderate deviations regime," in *IEEE ISIT*, pp. 2309–2314, 2022.

[141] L. Yu, "Gray–Wyner and mutual information regions for doubly symmetric binary sources and Gaussian sources," *arXiv.2212.14696*, 2022.

[142] A. A. Yushkevich, "On limit theorems connected with the concept of entropy of Markov chains," *Uspekhi Matematicheskikh Nauk*, vol. 8, no. 5, 1953, pp. 177–180.

[143] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, 1999, pp. 2629–2636.

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

[145] Z. Zhang, E.-H. Yang, and V. K. Wei, "The redundancy of source coding with a fidelity criterion-part one: Known statistics," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, 1997, pp. 71–91.

[146] Y. Zhong, F. Alajaji, and L. L. Campbell, "A type covering lemma and the excess distortion exponent for coding memoryless Laplacian sources," in *23rd Biennial Symposium on Communications*, pp. 100–103, 2006.

[147] L. Zhou and M. Motani, "Non-asymptotic converse bounds and refined asymptotics for two source coding problems," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, 2019, pp. 6414–6440.

[148] L. Zhou and M. Motani, "Kaspi problem revisited: Non-asymptotic converse bound and second-order asymptotics," in *IEEE Globecom*, pp. 1–6, 2017.

[149] L. Zhou and M. Motani, "On the multiple description coding with one semi-deterministic distortion measure," in *IEEE Globecom*, pp. 1–6, 2017.

[150] L. Zhou, V. Y. F. Tan, and M. Motani, "Discrete lossy Gray-Wyner revisited: Second-order asymptotics, large and moderate deviations," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, 2017, pp. 1766–1791.

[151] L. Zhou, V. Y. F. Tan, and M. Motani, "Second-order and moderate deviation asymptotics for successive refinement," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, 2017, pp. 2896–2921.

[152] L. Zhou, V. Y. F. Tan, and M. Motani, "Refined asymptotics for rate-distortion using Gaussian codebooks for arbitrary sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, 2019, pp. 3145–3159.

[153] L. Zhou, V. Y. F. Tan, and M. Motani, "The dispersion of mismatched joint source-channel coding for arbitrary sources and additive channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, 2019, pp. 2234–2251.