On Achievable Rate Regions for Source Coding over Networks

Thesis by

WeiHsin Gu

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2009
(Defended August 19, 2008)

Acknowledgements

This thesis contains the main research result that was done during my Ph.D study in Effros's group. It is a great pleasure to convey my gratitude to all those who helped me complete this thesis.

First, I would like to thank Professor Michelle Effros for her guidance and patience through the work. From her, I learned the spirit of doing high quality research as well as thinking rigorously and originally. She provided support and encouragement in various ways. Her passion for research motivated me to work hard constantly.

I gratefully acknowledge Professor Tracey Ho, Mayank Bakshi, and Soumya Jana for providing plenty of ideas from our discussions. Those ideas contributed greatly to this thesis. Professor Ho gave me a broad picture of the whole research topic. I learned from Mayank Bakshi much knowledge in the field of Source Coding. Soumya Jana contributed a key idea on the development of the approximation algorithms.

It is a pleasure to offer my gratitude to my labmates: Michael Fleming, Sidharth Jaggi, Asaf Cohen, Mayank Bakshi, Tao Cui, Sukwon Kim, Theodore Dikaliotis, Christopher S. Chang, Derek Leong, Svitlana Vyetrenko, and Simon Chien.

Many thanks go in particular to all my committee members: Professor Michelle Effros, Professor Robert. J McEliece, Professor Steven Low, Professor Babak Hassibi, and Professor Tracey Ho, for the help on my defense.

My special thanks go to my many best friends who shared the unforgettable and wonderful experience at Caltech for the past five years. Finally, I am very grateful to my family, Professor Chung-Chin Lu, and Professor Chi-Chao Chao for supporting me and giving me the wonderful chance to experience the great Caltech life.

Abstract

In the field of source coding over networks, a central goal is to understand the best possible performance for compressing and transmitting dependent data distributed over a network. The achievable rate region for such a network describes all link capacities that suffice to satisfy the reproduction demands. Here all the links in the networks are error-free, the data dependency is given by a joint distribution of the source random variables, and the source sequences are drawn i.i.d. according to the given source distribution. In this thesis, I study the achievable rate regions for general networks, deriving new properties for the rate regions of general network source coding problems, developing approximation algorithms to calculate these regions for particular examples, and deriving bounds on the regions for basic multi-hop and multi-path examples.

In the first part, I define a family of network source coding problems. That family contains all of the example networks in the literature as special cases. For the given family, I investigate abstract properties of the achievable rate regions for general networks. These properties include (1) continuity of the achievable rate regions with respect to both the source distribution and the distortion constraint vector and (2) a strong converse that implies the traditional strong converse. Those properties might be useful for studying a long-standing open question: whether a single-letter characterization of a given achievable rate region always exists.

In the second part, I develop a family of algorithms to approximate the achievable rate regions for some example network source coding problems based on their single-letter characterizations by using linear programming tools. Those examples contain (1) the lossless coded side information problem by Ahlswede and Körner, (2) the

Wyner-Ziv rate-distortion function, and (3) the Berger et al. bound for the lossy coded side information problem. The algorithms may apply more widely to other examples.

In the third part, I study two basic networks of different types: the two-hop and the diamond networks. The two-hop network is a basic example of line networks with single relay node on the path from the source to the destination, and the diamond network is a basic example of multi-path networks that has two paths from the source to the destination, where each of the paths contains a relay node. I derive performance bounds for the achievable rate regions for these two networks.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the point-to-point communication, when a data sequence is drawn i.i.d. according to a given probability distribution, the optimal lossless and lossy source coding performances are represented by the entropy function $H(X)$ for lossless source coding and the rate distortion function $R_X(D)$ for lossy source coding. These results were derived by Shannon in [1]. We wish to describe and calculate the achievable rate regions for generalizations of this problem where dependent data is described over a network of noiseless links. For each network source coding problem, which is defined formally in Section 2.2, we consider three types of achievable rate regions: the zero-error rate region, lossless rate region, and lossy rate region. The zero-error rate region contains all rate vectors $\mathbf{R}$ such that there exists a sequence of n-dimensional variable-length codes each of which achieves error probability precisely equal to zero with average expected description length vectors converging to $\mathbf{R}$. The lossless rate region contains all the rate vectors $\mathbf{R}$ such that there exists a sequence of length-n, rate-$\mathbf{R}$ block codes whose error probabilities can be made arbitrarily small when n grows without bound. The lossy rate region contains all the rate vectors $\mathbf{R}$ such that there exists a sequence of length-n, rate-$\mathbf{R}$ block codes that satisfy a given collection of distortion constraints asymptotically. In this thesis, we consider the scenario where the finite-alphabet sources are memoryless but not necessarily independent.

The vast majority of information theory research in the field of source coding over networks has focused on deriving single-letter bounds on the achievable rate regions. For example, the work of Slepian and Wolf treats the lossless source coding problem for the two-terminal network where the two source sequences are separately encoded and the decoder combines the two encoded messages to losslessly reproduce both of the two source sequences [2]. Gray and Wyner found an exact single-letter characterization for both lossless and lossy rate regions on a related “simple network” [3]. Ahlswede and Körner derived a single-letter characterization for the two-terminal network where the decoder needs to reconstruct only one source sequence losslessly [4]; that characterization employs an auxiliary random variable to capture the decoder's incomplete knowledge of the source that is not required to reconstruct. Wyner and Ziv derived a single-letter characterization of the optimal achievable rate for lossy source coding in the point-to-point network when side information is available only at the decoder [5]. Berger et al. derived an achievable region (inner bound) for the lossy two-terminal source coding problem in [6]. That region is known to be tight in some special cases [7]. Heegard and Berger found a single-letter characterization by using two auxiliary random variables for the network where side information may be absent [8]. Yamamoto considered a cascaded communication system with multi-hop and multi-branches [9]. For larger networks, Ahlswede et al. derived an optimal rate region for any network source coding problem where there is one source node that observes a collection of independent source random variables, all of which must be reconstructed losslessly by a family of sink nodes [10]; Ho et al. proved the cut-set bound is tight for multi-cast network with arbitrary dependency on the source random variables [11]; Bakshi and Effros generalized Ho's result to show the cut-set bound is still tight when side information random variables are available only at the end nodes [12].

In this thesis, I extend the prior results to study the rate regions for a wider family of network source coding problems. I investigate theoretical properties of rate regions, develop algorithms of approximating rate regions for particular network source coding problems, and derive inner and outer bounds for two basic examples.

In Chapters 2 and 3, I study the abstract properties of rate regions and their implications. In Chapter 2, I define a family of network source coding problems that contains the example networks and also the functional source coding problems [13, 14, 15] as special cases. I classify the problems in that family into four categories determined by separating lossless from lossy source coding and by distinguishing between what I call canonical source codes and problems that do not meet the canonical source coding definition, which are here called non-canonical source codes. Treating lossless and lossy rate regions as functions, I investigate the continuity properties of rate regions with respect to the source distribution and the distortion vector. The continuity results are critical for understanding whether rate regions for empirical distributions necessarily approximate the rate regions for the true underlying distribution. Early results of this material appear in [16, 17, 18].

I introduce a strong converse in Chapter 3. That strong converse implies the traditional strong converses for i.i.d. sources in the point-to-point [19], coded-side information [20], Slepian-Wolf [21] source coding problems. The proposed strong converse applies both to the problems mentioned above and to any multicast network with side information at the end nodes.

In Chapter 4, I develop a family of algorithms to approximate the rate regions for example distributions based on their single-letter characterizations. For rate regions characterized by auxiliary random variables, rate region calculation requires solution of an optimization problem. While derivation of rate regions has been a key area of research focus, the question of how to solve the underlying optimization problems has received far less attention. Rate region calculation turns out to be surprisingly difficult optimization problems since many regions are neither convex nor concave in the distributions of their auxiliary random variables. The well-known Arimoto-Blahut algorithm [22, 23] for calculating the channel capacity and the rate-distortion function and its extension [24] for calculating the Wyner-Ziv rate distortion function are iterative techniques for performing such optimizations. I propose an alternative approach for approximating the achievable rate region by first quantizing the space of possible distributions and then solving a finite linear programing whose solution is guaranteed to differ from the rate region by at most a factor of $(1+\epsilon)$ such that

$(1+\epsilon)$-approximation is guaranteed. This approach can be applied to a wider family of network source coding problems and may provide new tools useful for understanding some long-standing open problems, for instance, the tightness of Berger et al. bound [6]. I presented early results for the Ahlswede-Körner problem in [25, 26] and for the lossy coded side information problem in [27].

I study two basic example networks in Chapter 5: The two-hop line network and the diamond network. The two-hop line network is the simplest nontrivial example of network source coding for nodes in series. The diamond network is the simplest nontrivial example of network source coding across links in parallel. I derive inner and outer bounds for both of the problems by applying the techniques used in the literature. Parts of this work originally appeared in [28, 29, 30].

Chapter 2

A Continuity Theory

2.1 Introduction

Characterization of rate regions for source coding over networks is a primary goal in the field of source coding theory. Given a network and a collection of sources and demands, the lossless and lossy rate regions generalize Shannon's source coding and rate-distortion theorems [1] to describe, respectively, the set of achievable rate vectors for which the error probability can be made arbitrarily close to zero as block length grows without bound and the set of achievable rate vectors for which a given distortion constraint is asymptotically satisfied as block length grows without bound. The zero-error rate region is denoted by $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, the lossless rate region is denoted by $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, and the lossy rate region is denoted by $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$, where $P_{\mathbf{X},\mathbf{Y}}$ is the source and side-information distribution, here assumed to be stationary and memoryless, and $\mathbf{D}$ is the vector of distortion constraints.

In this chapter, which extends our works from [17] and [18], we investigate the continuity of $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to both $P_{\mathbf{X},\mathbf{Y}}$ and $\mathbf{D}$ for both canonical and non-canonical source codes. Here a network source coding problem is canonical if and only if every demand function can be written as a linear combination over some finite field of functions that can be calculated at some source nodes. Understanding the continuity properties of $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is important because continuity is required to guarantee that a reliable estimation of $P_{\mathbf{X},\mathbf{Y}}$ results in a reliable estimation of its rate regions. While proofs of continuity are straight forward when single-letter characterizations are available, we here study continuity for a general class of network source coding problems defined in Section 2.2. Proof of continuity is difficult in this case because single-letter characterizations are not available, optimal coding strategies are not known, and demonstrating continuity from limiting characterizations seems to be difficult.

A number of examples of single-letter rate region characterizations appear in the literature. In [2], Slepian and Wolf derive $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ for the two-terminal problem that has two separate encoders and one decoder interested in reproducing both of the source sequences. The given inequalities describe the cut-set bounds, which are tight for this network by [2], for all other multicast networks with independent and dependent sources by [10] and [11], respectively, and for multicast networks with receiver side information by [12]. The work in [4] introduces an auxiliary random variable in the single-letter characterization of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ for a network similar to the one studied by Slepian and Wolf. Other examples of single-letter characterizations of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ for non-functional source coding problems include [3, 5, 7, 8, 31, 32]. In each of these examples, the rate region is a continuous function of the source and side-information distribution when all random variables involved have finite alphabets. Rate region characterizations for the simplest lossless and lossy functional source coding problem appear in [14]. While these rate regions are also continuous in the source and side information distribution, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is not continuous in $P_{\mathbf{X},\mathbf{Y}}$ for all functional source coding problems by [33].

A function is continuous if and only if it is both inner and outer semi-continuous. Chen and Wagner demonstrated the inner semi-continuity of rate regions with respect to covariance matrix for Gaussian multi-terminal source coding problems and applied that result to investigate the tightness of some earlier derived bounds [34]. We consider only finite-alphabet source and side-information random variables.

We show that for any $P_{\mathbf{X},\mathbf{Y}}$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is continuous in $\mathbf{D}$ when (a) the network source coding problem $\mathcal{N}$ is canonical; or (b) $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$, i.e., the components of $\mathbf{D}$ are all non-zero. We prove that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ when $\mathcal{N}$ is canonical and show that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$ when (a) $\mathcal{N}$ is canonical; or (b) $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$.

A rate region, regarded as a function of distribution $P_{\mathbf{X},\mathbf{Y}}$, is s-continuous if and only if for all distributions $P_{\mathbf{X},\mathbf{Y}}$ and $Q_{\mathbf{X},\mathbf{Y}}$ with the same support, the rate regions for $P_{\mathbf{X},\mathbf{Y}}$ and $Q_{\mathbf{X},\mathbf{Y}}$ are sufficiently close when $P_{\mathbf{X},\mathbf{Y}}$ and $Q_{\mathbf{X},\mathbf{Y}}$ are sufficiently close. (See Definition 2.2.15.) We show that $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ for all $\mathbf{D}$ are all outer semi-continuous and s-continuous in $P_{\mathbf{X},\mathbf{Y}}$, which implies that the approximation of the rate regions for $P_{\mathbf{X},\mathbf{Y}}$ by the rate region for its empirical distribution is reliable. The s-continuity of $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ further implies that some graph entropies introduced in [35, 36][1] are continuous when support of the distribution is fixed.

The remainder of this chapter is structured as follows. We formulate the general network source coding problem and define continuity and s-continuity in Section 2.2. In Section 2.3, we derive some basic properties for non-functional and canonical source coding over networks. In Section 2.4, we show the continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $\mathbf{D}$ for all $\mathbf{D}$ in the canonical source coding case, and for $\mathbf{D} > \mathbf{0}$ in the non-canonical source coding case.

Section 2.5 treats continuity with respect to $P_{\mathbf{X},\mathbf{Y}}$. In Section 2.5.1, we show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous for canonical source coding and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is inner semi-continuous when (a) $\mathcal{N}$ is canonical or (b) $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$. In Section 2.5.2, we show that $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ for all $\mathbf{D}$ are all outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$. In Section 2.6, we show that $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ for all $\mathbf{D}$ are all s-continuous in $P_{\mathbf{X},\mathbf{Y}}$.

[1]The zero-error codes here are in the unrestricted-inputs scenario as defined in [35].

2.2 Formulation

Here we define a general network source coding problem and its zero-error, lossless, and lossy rate regions. Let $\mathbf{Z} = (Z_1, \ldots, Z_r)$ be a random vector with finite alphabet $\prod_{i=1}^r \mathcal{Z}_i$. We assume without loss of generality that $|\mathcal{Z}| = m$ for all $i \in \{1, \ldots, r\}$ but $\mathcal{Z}_i$ need not be the same set as $\mathcal{Z}_j$ for $i \neq j \in \{1, \ldots, r\}$. Let ϑ be a finite set that contains $\cup_{i=1}^r \mathcal{Z}_i$ as a subset and Θ denote the set of functions from $\prod_{i=1}^r \mathcal{Z}_i$ to ϑ.[2] Since $\cup_{i=1}^r \mathcal{Z}_i \subseteq \vartheta$, for each $i \in \{1, \ldots, r\}$, the function $\theta_i(z_1, \ldots, z_r) = z_i$ for all $(z_1, \ldots, z_r) \in \prod_{i=1}^r \mathcal{Z}_i$ is in Θ. For simplicity, we abbreviate the given functions to $Z_i = \theta_i$ for all $i \in \{1, \ldots, r\}$. This notation is useful later for discussing functional source coding. The sequence $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ is drawn i.i.d. according to a generic distribution $P_{\mathbf{Z}}$ of $\mathbf{Z}$, which describes all network inputs.

Fix a distortion measure $d : \vartheta \times \vartheta \to [0, \infty)$. Define $d_{\min} = \min_{a \neq b} d(a, b)$ and $d_{\max} = \max_{a \neq b} d(a, b)$. We assume that $d(a, b) = 0$ if and only if $a = b$, which implies $d_{\min} > 0$.We further assume that $d_{\max} < \infty$. The distortion between any two sequences $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ in ϑ^n is defined as $d(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n d(a_i, b_i)$.

A directed network is an ordered pair $(\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Vector $(v, v') \in \mathcal{E}$ if and only if there is a directed edge from v to v'. For each edge $e = (v, v') \in \mathcal{E}$, we call v the tail of e and v' the head of e, denoted by $v = \mathrm{tail}(e)$ and $v' = \mathrm{head}(e)$, respectively. The set of edges that end at vertex v is denoted by $\Gamma_I(v)$ and the set of edges that begin at v is denoted by $\Gamma_O(v)$, i.e.,

$$\Gamma_I(v) := \{e \in \mathcal{E} \ : \ \mathrm{head}(e) = v\}$$

$$\Gamma_O(v) := \{e \in \mathcal{E} \ : \ \mathrm{tail}(e) = v\}.$$

Let $G = (\mathcal{V}, \mathcal{E})$ be a directed acyclic network.[3] A network source coding problem

[2]Notice that ϑ can be designed to include all functions of the form $\phi_{(i_1,\ldots,i_\ell)}(\mathbf{z}) = (z_{i_1}, \ldots, z_{i_\ell})$ for all $\ell \in \{1, \ldots, r\}$ and $1 \leq i_1 < \ldots < i_\ell \leq r$.

[3]For any network with cycles, we design the codes in a chronological order and consider a corresponding acyclic network that matches the given chronological order. An example can be found in [37].

$\mathcal{N}$ is defined as $\mathcal{N} = (G, \mathcal{S}, \mathcal{D})$. Here sets $\mathcal{S}$ and $\mathcal{D}$ describe the random variable availabilities and demands, respectively. The random variable availability set $\mathcal{S}$ is a subset of $\mathcal{V} \times \{Z_1, \ldots, Z_r\}$ such that Z_i $(1 \leq i \leq r)$ is available at node $v \in \mathcal{V}$ if and only if $(v, Z_i) \in \mathcal{S}$. The demand set $\mathcal{D}$ is a subset of $\mathcal{V} \times \Theta$ such that node $v \in \mathcal{V}$ demands function $\theta \in \Theta$ if and only if $(v, \theta) \in \mathcal{D}$. Let k denote the total number of reproduction demands, i.e., $k = |\mathcal{D}|$. For each $v \in \mathcal{V}$, sets $\mathcal{S}_v \subseteq \{Z_1, \ldots, Z_r\}$ and $\mathcal{D}_v \subseteq \Theta$ summarize the random variable availabilities and demands, respectively, at node v, giving

$$\begin{aligned}\mathcal{S}_v &= \{Z_i : (v, Z_i) \in \mathcal{S}\}\\ \mathcal{D}_v &= \{\theta : (v, \theta) \in \mathcal{D}\}.\end{aligned}$$

For any set $I \subseteq \{1, \ldots, r\}$ and any $\mathbf{z} \in \prod_{i=1}^{r} \mathcal{Z}_i$, let $\mathbf{z}_I = (z_i \mid i \in I)$ and $\mathbf{z}_{I^c} = (z_i \mid i \notin I)$. We use $\mathbf{z}_I\mathbf{z}_{I^c}$ interchangeably with $\mathbf{z}$. For any $\theta \in \Theta$, we define $I(\theta)$ to be the smallest set I for which $\theta(\mathbf{z}_I, \mathbf{z}_{I^c}) = \theta(\mathbf{z}_I, \mathbf{z}'_{I^c})$ for every $\mathbf{z}_I \in \prod_{i \in I} \mathcal{Z}_i$ and every $\mathbf{z}_{I^c}$, $\mathbf{z}'_{I^c} \in \prod_{i \notin I} \mathcal{Z}_i$. Thus $\theta(\mathbf{z})$ is independent of z_i for all $i \notin I(\theta)$.

Definition 2.2.1 *Let $\mathcal{N}$ be a network source coding problem. Random variable Z_i is called a source random variable if and only if there exists a demand pair $(v, \theta) \in \mathcal{D}$ for which $i \in I(\theta)$. Otherwise, Z_i is called a side-information random variable.*

Let $1 \leq s \leq r$ be the number of source random variables, and let $t = r - s$ denote the number of side-information random variables. Henceforth, we use $\mathbf{X} = (X_1, \ldots, X_s)$ to describe the source random vector, $\mathbf{Y} = (Y_1, \ldots, Y_t)$ to specify the side-information vector, and $P_{\mathbf{X},\mathbf{Y}} = P_{\mathbf{Z}}$ to denote the probability mass function on the source and side-information vector. Adjusting this new notation, for every $\theta \in \cup_v \mathcal{D}_v$, $I(\theta)$ is now a subset of $\{1, \ldots, s\}$ and $\mathbf{x}_{I(\theta)} = (x_i \mid i \in I(\theta))$ are the symbols which θ relies. Let $\mathcal{X}_i$ and $\mathcal{Y}_j$ denote the alphabet sets of X_i and Y_j, respectively, for every $i \in \{1, \ldots, s\}$ and every $j \in \{1, \ldots, t\}$. Let

$$\mathcal{A} = \prod_{i=1}^{s} \mathcal{X}_i \times \prod_{j=1}^{t} \mathcal{Y}_j$$

denote the set of alphabet for $(\mathbf{X}, \mathbf{Y})$.

We assume that for each demand pair $(v', \theta) \in \mathcal{D}$ and each $i \in I(\theta)$, there exists a pair $(v, X_i) \in \mathcal{S}$ such that there is a path from v to v'.

The following definition defines two types of network source coding problems, called non-functional and functional network source coding problems.

Definition 2.2.2 *Let $\mathcal{N}$ be a network source coding problem. If all the demands are sources, i.e., $\cup_v \mathcal{D}_v \subseteq \{X_1, \ldots, X_s\}$, then $\mathcal{N}$ is called a* non-functional network source coding problem. *Otherwise, $\mathcal{N}$ is called a* functional network source coding problem.

Definition 2.2.3 *Let $\mathbf{R} = (R_e)_{e\in\mathcal{E}}$ be a rate vector. A rate-$\mathbf{R}$, length-n block code $\mathcal{C}$ for $\mathcal{N}$ contains a set of encoding functions $\{f_e \mid e \in \mathcal{E}\}$ and a set of decoding functions $\{g_{v,\theta} \mid (v, \theta) \in \mathcal{D}\}$.*

(i) For each $e \in \mathcal{E}$, the encoding function is a map

$$\begin{aligned} f_e \;:\; & \prod_{e'\in\Gamma_I(\mathrm{tail}(e))} \{1, 2, \ldots, 2^{nR_{e'}}\} \times \prod_{i:X_i\in\mathcal{S}_{\mathrm{tail}(e)}} \mathcal{X}_i^n \times \prod_{j:Y_j\in\mathcal{S}_{\mathrm{tail}(e)}} \mathcal{Y}_j^n \\ & \to \{1, 2, \ldots, 2^{nR_e}\}. \end{aligned}$$

(ii) For each $(v, \theta) \in \mathcal{D}$, the decoding function is a map

$$g_{v,\theta} \;:\; \prod_{e\in\Gamma_I(v)} \{1, 2, \ldots, 2^{nR_e}\} \times \prod_{i:X_i\in\mathcal{S}_v} \mathcal{X}_i^n \times \prod_{j:Y_j\in\mathcal{S}_v} \mathcal{Y}_j^n \to \vartheta^n.$$

We next define the class of variable-length codes considered in this chapter for network $\mathcal{N}$. Inorder to make arbitrary ℓ copies of every variable-length code in this class well-defined, the variable-length codes discussed in this chapter satisfy the "uniquely encodable" property described in Definition 2.2.5. Roughly speaking, the uniquely encodable property means that for any positive integer ℓ and every node v, the encoded codeword vector observed at v when applying a variable-length code ℓ times can be uniquely decomposed into ℓ codeword vectors of the same code. The definition of the uniquely encodable property relies on the uniquely decodable property for sets of codeword vectors.

For any finite set S, let $S^* = \cup_{n=1}^{\infty} S^n$ denote the set of finite-length sequences drawn from S. Without loss of generality, we treat only binary variable-length codes.

Definition 2.2.4 *Let n be a positive integer and $S_1, \ldots, S_n$ be finite sets. A set $\mathbb{C} \subset \prod_{i=1}^n S_i$ is called a uniquely decodable set if and only if for every positive integer ℓ, the map*

$$\phi_\ell : \mathbb{C}^\ell \to \prod_{i=1}^n S_i^*$$

defined by

$$\phi_\ell \left((b_1^{(1)}, \ldots, b_n^{(1)}), \ldots, (b_1^{(\ell)}, \ldots, b_n^{(\ell)}) \right) := \left((b_1^{(1)}, \ldots, b_1^{(\ell)}), \ldots, (b_n^{(1)}, \ldots, b_n^{(\ell)}) \right)$$

for all $(b_1^{(1)}, \ldots, b_n^{(1)}), \ldots, (b_1^{(\ell)}, \ldots, b_n^{(\ell)}) \in \mathbb{C}$ is a one-to-one map.

Definition 2.2.5 *A dimension-n variable-length code $\mathcal{C}$ for $\mathcal{N}$ contains a collection of codebooks $\{\mathbb{C}_e \mid \mathbb{C}_e \subset \{0,1\}^* \ \forall e \in \mathcal{E}\}$, a set of encoding functions $\{f_e \mid e \in \mathcal{E}\}$, and a set of decoding functions $\{g_{v,\theta} \mid (v,\theta) \in \mathcal{D}\}$. $\mathcal{C}$ satisfies the properties below.*

(i) For each $e \in \mathcal{E}$, the encoding function f_e is a map

$$f_e \ : \prod_{e' \in \Gamma_I(\text{tail}(e))} \mathbb{C}_{e'} \times \prod_{i: X_i \in \mathcal{S}_{\text{tail}(e)}} \mathcal{X}_i^n \times \prod_{j: Y_j \in \mathcal{S}_{\text{tail}(e)}} \mathcal{Y}_j^n \to \mathbb{C}_e.$$

(ii) For each $(v,\theta) \in \mathcal{D}$, the decoding function is a map

$$g_{v,\theta} \ : \prod_{e \in \Gamma_I(v)} \mathbb{C}_e \times \prod_{i: X_i \in \mathcal{S}_v} \mathcal{X}_i^n \times \prod_{j: Y_j \in \mathcal{S}_v} \mathcal{Y}_j^n \to \vartheta^n.$$

(iii) For every $e \in \mathcal{E}$ and every $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{A}^n$, let $c_e(\mathbf{x}^n, \mathbf{y}^n) \in \mathbb{C}_e$ denote the codeword on edge e using code $\mathcal{C}$ when the input sequence is $(\mathbf{x}^n, \mathbf{y}^n)$. For each $v \in \mathcal{V}$, the codebook consisting of all possible messages received by v, defined

by

$$\begin{aligned}\mathbb{C}[v] &:= \{((c_e(\mathbf{z}^n))_{e\in\Gamma_I(v)}, (x_i^n)_{i:X_i\in\mathcal{S}_v}, (y_j^n)_{j:Y_j\in\mathcal{S}_v}) \mid (\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{A}^n\} \\ &\subseteq \prod_{e\in\Gamma_I(v)} \mathbb{C}_e \times \prod_{i:X_i\in\mathcal{S}_v} \mathcal{X}_i^n \times \prod_{j:Y_j\in\mathcal{S}_v} \mathcal{Y}_j^n,\end{aligned}$$

is a uniquely decodable set.[4]

For each $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{A}^n$*, let* $\mathbf{L}_e(\mathbf{x}^n, \mathbf{y}^n)$ *denote the length of the codeword on* e *and let* $\mathbf{L}(\mathbf{x}^n, \mathbf{y}^n) = (\mathbf{L}_e(\mathbf{x}^n, \mathbf{y}^n))_{e\in\mathcal{E}}$ *denote the length vector using code* $\mathcal{C}$ *when input sequence is* $(\mathbf{x}^n, \mathbf{y}^n)$*.*

A function of $\mathbf{X}$ is called a canonical function if and only if it is either a source random variable or can be rewritten as a linear combination of functions of individual symbols. The next definition formalizes this idea. For any prime power q, we fix a finite field with q elements and denote it by $\mathbb{F}_q$. Notice that $\mathbb{F}_q$ is unique up to isomorphisms.

Definition 2.2.6 *A function* θ *of* $\mathbf{X}$ *in* Θ *is called a canonical function if and only if there exist a prime power* $q \geq m$*, a map* ϕ_i *from* $\mathcal{X}_i$ *to* $\mathbb{F}_q$ *for each* $i \in I(\theta)$*, and a one-to-one map* ψ *from the output alphabet of* θ *to* $\mathbb{F}_q$ *such that*

$$\psi(\theta(\mathbf{x})) = \sum_{i\in I(\theta)} \phi_i(x_i) \ \forall \ \mathbf{x} \in \prod_{i=1}^{s} \mathcal{X}_i. \tag{2.1}$$

Definition 2.2.7 *If all demands of network source coding problem* $\mathcal{N}$ *are canonical functions, then* $\mathcal{N}$ *is called a canonical network source coding problem.*

By definition, every non-functional network source coding problem is canonical. Hence the family of canonical network source coding problems contains the family of non-functional network source coding problems and some functional network source coding problems.

[4] For each $\ell \in \mathbb{N}$, the code constructed by applying $\mathcal{C}$ ℓ times is well-defined under this setting.

Definition 2.2.8 *Let $\mathcal{N}$ be a network source coding problem.*

(a) *A rate vector $\mathbf{R}$ is zero-error-achievable on pmf $P_{\mathbf{X},\mathbf{Y}}$ if and only if there exists a sequence of dimension-n , zero-error, variable-length codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ whose average expected length vectors with respect to $P_{\mathbf{X},\mathbf{Y}}$ are asymptotically no greater than $\mathbf{R}$. That is, for any $(v,\theta) \in \mathcal{D}$, let*

$$\widehat{\theta}^n(v) := g_{v,\theta}((c_e(\mathbf{X}^n))_{e\in\Gamma_I(v)}, (X_i^n)_{X_i\in S_{\text{tail}}(e)}, (Y_j^n)_{Y_j\in S_{\text{tail}}(e)})$$

denote the reproduction of $\theta^n(\mathbf{X}^n) = (\theta(\mathbf{X}_1), \ldots, \theta(\mathbf{X}_n))$ at node v using code $\mathcal{C}_n$. Then $\mathbf{R}$ is zero-error-achievable on $P_{\mathbf{X},\mathbf{Y}}$ if and only if

$$P_{\mathbf{X}^n,\mathbf{Y}^n}\left(\theta^n(\mathbf{X}^n) \neq \widehat{\theta}^n(v)\ \forall\ (v,\theta) \in \mathcal{D}\right) = 0$$
$$\limsup_{n\to\infty} \frac{1}{n} E_P d\left[\mathbf{L}_n(\mathbf{X}^n, \mathbf{Y}^n)\right] \leq \mathbf{R},$$

where E_P is the expectation with respect to $P_{\mathbf{X},\mathbf{Y}}$. The closure of the set of zero-error-achievable rate vectors (on $P_{\mathbf{X},\mathbf{Y}}$) is called the zero-error rate region, denoted by $\mathcal{R}_Z(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$.

(b) *A rate vector $\mathbf{R}$ is losslessly-achievable if and only if there exists a sequence of rate-$\mathbf{R}$, length-n block codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ whose symbol error probabilities are asymptotically zero. That is, for any $e \in \mathcal{E}$, let F_e denote the encoded message carried over edge e, and for any $(v,\theta) \in \mathcal{D}$, let*

$$\widehat{\theta}^n(v) := g_{v,\theta}((F_e)_{e\in\Gamma_I(v)}, (X_i^n)_{X_i\in S_{\text{tail}(v)}}, (Y_j^n)_{Y_j\in S_{\text{tail}(v)}})$$

denote the reproduction of $\theta^n(\mathbf{X}^n) = (\theta(\mathbf{X}_1), \ldots, \theta(\mathbf{X}_n))$ at node v using code $\mathcal{C}_n$. Then $\mathbf{R}$ is losslessly-achievable if and only if

$$\lim_{n\to\infty} P_{\mathbf{X}^n,\mathbf{Y}^n}\left(\theta^n(\mathbf{X}^n) \neq \widehat{\theta}^n(v)\right) = 0$$

for all $(v,\theta) \in \mathcal{D}$. The closure of the set of losslessly-achievable rate vectors is *called the lossless rate region, denoted by* $\mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$*.*

(c) *Let* $\mathbf{D} = (D_{v,\theta})_{(v,\theta)\in\mathcal{D}}$ *be a* k*-dimensional vector whose components are nonnegative real numbers. Rate vector* $\mathbf{R}$ *is said to be* $\mathbf{D}$*-achievable if and only if there exists a sequence of rate-*$\mathbf{R}$*, length-*n *block codes* $\{\mathcal{C}_n\}_{n=1}^{\infty}$ *such that the distortion constraint is asymptotically satisfied. That is, for every* $(v,\theta) \in \mathcal{D}$*, let* $\widehat{\theta}^n(v)$ *denote the reproduction of* $\theta^n(\mathbf{X}^n) = (\theta(\mathbf{X}_1), \ldots, \theta(\mathbf{X}_n))$ *at node* v *by* $\mathcal{C}_n$*. Then* $\mathbf{R}$ *is* $\mathbf{D}$*-achievable if and only if*

$$\limsup_{n\to\infty} \frac{1}{n} E_P d\left(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)\right) \leq D_{v,\theta}$$

for all $(v,\theta) \in \mathcal{D}$*. The closure of the set of* $\mathbf{D}$*-achievable rate vectors* $\mathbf{R}$ *is called the rate-distortion region, denoted by* $\mathcal{R}(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$*.*

In this chapter, we often abbreviate $\mathcal{R}_Z(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ to $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ when the network $\mathcal{N}$ is clear in the given context.

Let $\mathcal{M}$ denote the set of all probability distributions on $\mathcal{A}$. For any set A, we use 2^A to denote the power set of A, i.e., the set of all subsets of A. Let $\mathbb{R}_+$ denote the set of nonnegative real numbers. Define

$$\mathbb{R}_+^{\mathcal{D}} := \{\mathbf{D} = (\mathrm{D}_{(\mathrm{v},\theta)})_{(\mathrm{v},\theta)\in\mathcal{D}} \mid \mathrm{D}_{(\mathrm{v},\theta)} \geq 0 \ \forall \ (\mathrm{v},\theta) \in \mathcal{D}\}$$

to be the set of all distortion vectors $\mathbf{D}$, and

$$\mathbb{R}_+^{\mathcal{E}} := \{\mathbf{R} = (\mathrm{R}_\mathrm{e})_{\mathrm{e}\in\mathcal{E}} \mid \mathrm{R}_\mathrm{e} \geq 0 \ \forall \ \mathrm{e} \in \mathcal{E}\}$$

to be the set of all rate vectors. For all $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is a subset of $\mathbb{R}_+^{\mathcal{E}}$; therefore, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ can be considered as a function from $\mathcal{M} \times \mathbb{R}_+^{\mathcal{D}}$ to $2^{\mathbb{R}_+^{\mathcal{E}}}$. Similarly both $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ can be considered as functions from $\mathcal{M}$ to $2^{\mathbb{R}_+^{\mathcal{E}}}$.

Before defining the continuity property, we introduce the definitions of set operations in $\mathbb{R}_+^{\mathcal{E}}$ and distances[5] on $\mathcal{M}$ and $2^{\mathbb{R}_+^{\mathcal{E}}}$ used in this chapter.

Definition 2.2.9 *Let A and B be two subsets of the n-dimensional Euclidean space $\mathbb{R}^{\mathrm{n}}$.*

(a) For any n-dimensional vector $\mathbf{v} \in \mathbb{R}^{\mathrm{n}}$, define the set

$$A + \mathbf{v} := \{\mathbf{a} + \mathbf{v} \mid \mathbf{a} \in A\}.$$

(b) For any $\lambda, \mu \in \mathbb{R}_+ \cup \{0\}$, define the set

$$\lambda A + \mu B := \{\lambda\mathbf{a} + \mu\mathbf{b} \mid \mathbf{a} \in A,\ \mathbf{b} \in B\}.$$

Definition 2.2.10 *Given a positive integer n. Define*

$$\mathbf{1} := (1, \ldots, 1) \in \mathbb{R}^{\mathrm{n}} \text{ and } \mathbf{0} := (0, \ldots, 0) \in \mathbb{R}^{\mathrm{n}}.$$

Definition 2.2.11 *Let $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ be two n-dimensional real vectors.*

(a) We say that $\mathbf{a}$ is greater than or equal to $\mathbf{b}$, denoted by $\mathbf{a} \geq \mathbf{b}$, if and only if $a_i \geq b_i$ for all $i \in \{1, \ldots, n\}$.

(b) We say that $\mathbf{a}$ is greater than $\mathbf{b}$, denoted by $\mathbf{a} > \mathbf{b}$, if and only if $a_i > b_i$ for all $i \in \{1, \ldots, n\}$.

The distances on $\mathcal{M}$ and $2^{R_+^{\mathcal{E}}}$ used in this chapter are as follows.

Definition 2.2.12 *Let A and B be two subsets of the n-dimensional Euclidean space $\mathbb{R}^{\mathrm{n}}$. We use $||\mathbf{x}||$ to denote the L_2-norm of $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$. For any $\epsilon > 0$, sets A and B are said to be ϵ-close ($\epsilon > 0$) if and only if*

[5] The distance used between any two subsets in $\mathbb{R}_+^{\mathcal{E}}$ is not a metric. It is equivalent to the Hausdorff distance, which is used for compact subsets of $\mathbb{R}^{\mathrm{n}}$ and gives a way to measure the difference between two subsets in $\mathbb{R}_+^{\mathcal{E}}$.

(a) *For every* $\mathbf{a} \in A$*, there exists some* $\mathbf{b}_0 \in B$ *such that* $||\mathbf{a} - \mathbf{b}_0|| \leq \epsilon\sqrt{n}$*.*

(b) *For every* $\mathbf{b} \in B$*, there exists some* $\mathbf{a}_0 \in A$ *such that* $||\mathbf{b} - \mathbf{a}_0|| \leq \epsilon\sqrt{n}$*.*

This notion of the distance between two subsets in $\mathrm{I\!R}^n$ leads to the definitions of continuity of regions $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ (with respect to $P_{\mathbf{X},\mathbf{Y}}$), $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ (with respect to $P_{\mathbf{X},\mathbf{Y}}$), and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ (with respect to both $P_{\mathbf{X},\mathbf{Y}}$ and $\mathbf{D}$).

Definition 2.2.13 *Fix* $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$*.* $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ *is continuous in* $\mathbf{D}$ *over set* $\mathcal{D}_0 \subseteq \mathcal{D}$ *if and only if for any* $\mathbf{D} \in \mathcal{D}_0$ *and* $\epsilon > 0$*, there exists some* $\delta > 0$ *such that* $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ *and* $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}')$ *are* ϵ*-close for all* $\mathbf{D}' \in \mathcal{D}_0$ *satisfying* $||\mathbf{D} - \mathbf{D}'|| < \delta$*. If the choice of* ϵ *and* δ *can be independent of* $P_{\mathbf{X},\mathbf{Y}}$*, then* $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ *is said to be continuous in* $\mathbf{D}$ *independently of* $P_{\mathbf{X},\mathbf{Y}}$*.*

We define continuity and s-continuity with respect to $P_{\mathbf{X},\mathbf{Y}}$ here. Let $A(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathrm{I\!R}_+^{\mathcal{E}}$ be a function with a subset outcome. For example, $A(P_{\mathbf{X},\mathbf{Y}})$ may be the zero-error, lossless, or lossy rate region.

Definition 2.2.14 *Function* $A(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathrm{I\!R}_+^{\mathcal{E}}$ *is continuous in* $P_{\mathbf{X},\mathbf{Y}}$ *if and only if for any* $P_{\mathbf{X},\mathbf{Y}}$ *and* $\epsilon > 0$*, there exists some* $\delta > 0$ *such that* $A(P_{\mathbf{X},\mathbf{Y}})$ *and* $A(Q_{\mathbf{X},\mathbf{Y}})$ *are* ϵ*-close for all* $Q_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ *satisfying* $||P_{\mathbf{X},\mathbf{Y}} - Q_{\mathbf{X},\mathbf{Y}}|| < \delta$*.*

Definition 2.2.15 *Function* $A(P_{\mathbf{X},\mathbf{Y}})$ *is s-continuous in* $P_{\mathbf{X},\mathbf{Y}}$ *if and only if for any* $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ *and* $\epsilon > 0$*, there exists some* $\delta > 0$ *such that* $A(P_{\mathbf{X},\mathbf{Y}})$ *and* $A(Q_{\mathbf{X},\mathbf{Y}})$ *are* ϵ*-close for all* $Q_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ *satisfying*

$$(1-\delta)P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \leq Q_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \leq \frac{1}{1-\delta}P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \;\forall\; (\mathbf{x},\mathbf{y}) \in \mathcal{A}.$$

$\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ is known not to be continuous in $P_{\mathbf{X},\mathbf{Y}}$. An example is shown in [14]. In this chapter, we show $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous and s-continuous in $P_{\mathbf{X},\mathbf{Y}}$. (See Section 2.5.2 for the definition of outer semi-continuity.)

It is tempting to assume the continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ on general networks by analogy to the very limited collection of problems where rate regions are fully characterized and continuity is easy to check. Since this assumption is known to be incorrect by the earlier cited example of Han and Kobayashi, we reproduce that example here to demonstrate the complexity of the problem even in very small, simple networks.

Example 2.2.16 *[33, Remark 1]Consider the network source coding problem in Fig. 2.1. The function $f(X_1, X_2)$ is defined as*

$$f(0,0) = f(0,1) = 0,\ f(1,0) = 1,\ f(1,1) = 2.$$

For $\epsilon \geq 0$, consider the distribution

$$\begin{aligned} P^{\epsilon}_{X_1,X_2}(0, x_2) &= \frac{1}{2} - \epsilon\ \forall\ x_2 \in \{0,1\} \\ P^{\epsilon}_{X_1,X_2}(1, x_2) &= \epsilon\ \forall\ x_2 \in \{0,1\}. \end{aligned}$$

By [33, Theorem 1],

$$\lim_{\epsilon \to 0} \mathcal{R}_L(P^{\epsilon}_{X_1,X_2}) = \{(R_1, R_2) \mid R_1 \geq 0,\ R_2 \geq \log(2)\}.$$

On the other hand, when $\epsilon = 0$, $f = 0$ with probability 1 and hence

$$\mathcal{R}_L(P^{0}_{X_1,X_2}) = \{(R_1, R_2) \mid R_1 \geq 0,\ R_2 \geq 0\}.$$

This shows that

$$\lim_{\epsilon \to 0} \mathcal{R}_L(P^{\epsilon}_{X_1,X_2}) \subsetneq \mathcal{R}_L(P^{0}_{X_1,X_2}).$$

The demand function in the above example is not canonical, and the lossless rate region is not inner semi-continuous. (See Definition 2.5.2.) We show in Corollary 2.5.7 that the lossless rate region for any canonical network source coding problem is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$, which implies that the behaviors of lossless rate regions for canonical and non-canonical network source coding problems are different. A key

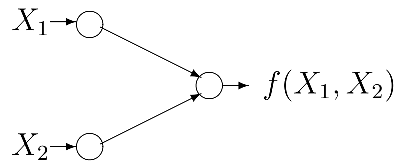


Figure 2.1: A functional two-terminal lossless problem where $\mathcal{R}_L(P_{X_1,X_2})$ is not continuous.

idea needed to prove the s-continuity result for canonical networks is captured in Lemma 2.2.18, which is proven using the technique introduced in the proof of [38, Theorem 1] together with the following theorem, which was proven in [39].

Theorem 2.2.17 *Let $\mathcal{N}$ be the lossless source coding problem shown in Fig. 2.2 and let ψ be a one-to-one map from $\mathcal{X}$ to $\mathbb{F}_q$. For any $\epsilon > 0$, there exists a sequence of rate-$(H(X|Y) + \epsilon)$, length-n block codes that encode $\psi^n(X^n)$ linearly over finite field $\mathbb{F}_q$ such that the sequence of the corresponding decoding error probabilities converges to 0 as n grows without bound.*

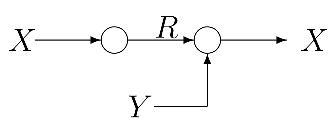


Figure 2.2: The lossless source coding problem with side information at the decoder.

Lemma 2.2.18 *Let $\mathcal{N}$ be the lossless multiterminal functional source coding problem shown in Fig. 2.3. Suppose that θ is a canonical function of $\mathbf{X}$. Then the rate vector $(H(\theta|Y) + \epsilon) \cdot \mathbf{1}$ is achievable for all $\epsilon > 0$.*

Proof. Let $q \geq m$ be a prime power, ψ be a one-to-one map from the output alphabet of θ to $\mathbb{F}_q$, and let ϕ_i be a map from $\mathcal{X}_i$ to $\mathbb{F}_q$ for each $i \in I(\theta)$ such that equation (2.1) from Definition 2.2.6 holds, viing $\psi(\theta(\mathbf{x})) = \sum_{i=1}^{s} \phi_i(x_i)$. For any

$\epsilon > 0$, let $\{\mathcal{C}_n\}$ be a sequence of rate-$(H(\theta|Y)+\epsilon)$, length-n block codes that linearly encode $\psi^n(\theta^n(\mathbf{X}^n))$ over $\mathbb{F}_q$ using encoders T_n and linearly decode the encoded message with optimal decoders $\widehat{\theta}$ to achieve error probability

$$P_e^{(n)} := \min_{\widehat{\theta}^n} P_{\mathbf{X}^n,\mathbf{Y}^n}\left(\theta^n(\mathbf{X}^n) \neq \widehat{\theta}^n\left(T_n(\psi^n(\theta^n(\mathbf{X}^n))), Y^n\right)\right)$$

satisfying

$$\lim_{n\to\infty} P_e^{(n)} = 0.$$

For each n, we construct a rate-$((H(\theta|Y)+\epsilon)\cdot\mathbf{1})$, length-$n$ block code for $\mathcal{N}$ by applying $T_n \circ \phi_i^n$ on source sequence X_i^n for each $i \in \{1,\ldots,s\}$. The end node receives the encoded messages $T_n(\phi_1^n(X_1^n))$, $\ldots$, $T_n(\phi_s^n(X_s^n))$ and calculates $\sum_{i=1}^s T_n(\phi_i(X_i^n))$. Since $T_n(\psi^n(\theta^n(\mathbf{X}^n))) = \sum_{i=1}^s T_n(\phi_i(X_i^n))$ by (2.1) and the linearity of T_n, the rate vector $(H(\theta|Y)+\epsilon)\cdot\mathbf{1}$ is losslessly achievable for $\mathcal{N}$. $\square$

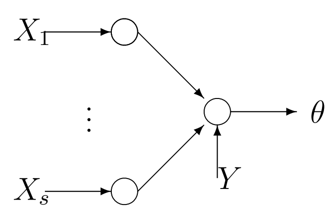


Figure 2.3: The lossless multiterminal functional source coding problem with side information at the decoder.

Remark 2.2.19 *By definition, it may seems more general to consider the rate regions achieved by the class of variable-length codes for all of our problems. In reality, for canonical lossless, canonical lossy, and functional lossy source coding on memoryless sources there is no loss of generality in restricting our attention to fixed-rate codes. We sketch the proof as follows. Since canonical lossless source coding is a special case of canonical lossy source coding (see Theorem 2.3.5) and canonical lossy source coding is a special case of functional lossy source coding, we sketch the proof for only functional lossy source coding.*

For any fixed $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$, *let* $\mathbf{R}$ *be a rate vector such that there exists a sequence of dimension-*n *variable-length codes* $\{\mathcal{C}_n\}$ *such that*

$$\limsup_{n\to\infty} \frac{1}{n} Ed(\mathbf{L}_n(\mathbf{X}^n, \mathbf{Y}^n)) \leq \mathbf{R}$$
$$\limsup_{n\to\infty} \frac{1}{n} Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} \ \forall\ (v,\theta) \in \mathcal{D},$$

where $\widehat{\theta}^n(v)$ *is the reproduction of* $\theta^n(\mathbf{X}^n)$ *at node* v *using* $\mathcal{C}_n$ *and* $\mathbf{L}_n(\mathbf{X}^n, \mathbf{Y}^n)$ *is the length vector of* $\mathcal{C}_n$ *for all* n. *For any* $\epsilon > 0$, *there exists an* $n > 0$ *such that*

$$\frac{1}{n} Ed(\mathbf{L}_n(\mathbf{X}^n, \mathbf{Y}^n)) \leq \mathbf{R} + \epsilon \cdot \mathbf{1}$$
$$\frac{1}{n} Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} + \epsilon \ \forall\ (v,\theta) \in \mathcal{D}.$$

By the weak law of large numbers, there exists an $l > 0$ *such that*

$$P_{\mathbf{X}^n,\mathbf{Y}^n}\left(\sum_{i=1}^{l} \mathbf{L}_n(\mathbf{X}_i^n, \mathbf{Y}_i^n) > ln(\mathbf{R} + 2\epsilon \cdot \mathbf{1})\right) < \epsilon \tag{2.2}$$
$$P_{\mathbf{X}^n,\mathbf{Y}^n}\left(\sum_{i=1}^{l} d(\theta^n(\mathbf{X}_i^n), \widehat{\theta}_i^n(v)) > ln(D_{v,\theta} + 2\epsilon) \ \forall\ (v,\theta) \in \mathcal{D}\right) < \epsilon, \tag{2.3}$$

where $\mathbf{X}^{nl} = (\mathbf{X}_1, \ldots, \mathbf{X}_l)$, $\mathbf{Y}^{nl} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_l)$, *and* $\widehat{\theta}_i^n(v)$ *is the reproduction of* $\theta^n(\mathbf{X}_i^n)$ *at node* v *for all* $i \in \{1, \ldots, l\}$ *and* $(v,\theta) \in \mathcal{D}$.

Let $\widetilde{\mathcal{C}}_{ln}$ *be the* (ln)*-dimensional code achieved by applying* $\mathcal{C}_n$ *on* $(\mathbf{X}_1^n, \mathbf{Y}_1^n)$, $\ldots$, $(\mathbf{X}_l^n, \mathbf{Y}_l^n)$ *sequentially. By (2.2), the variable-length code* $\widetilde{\mathcal{C}}_{ln}$ *has length vector no greater than* $ln(\mathbf{R} + 2\epsilon \cdot \mathbf{1})$ *with probability* $1 - \epsilon$. *We next construct a block code* $\widehat{\mathcal{C}}_{ln}$ *based on* $\widetilde{\mathcal{C}}_{ln}$ *as follows. For each* $e \in \mathcal{E}$, *if the code length for the encoded message on* e *is greater than* $ln(R_e + 2\epsilon)$, *then we truncate the code such that the resulting code length is* $ln(R_e + 2\epsilon)$; *if the code length for the encoded message on* e *is less than* $ln(R_e + 2\epsilon)$, *then we add a string of zeros after the encoded message to give a code with length equal to* $ln(R_e + 2\epsilon)$. *By construction, the modified block code* $\widehat{\mathcal{C}}_{ln}$

has rate $\mathbf{R} + 2\epsilon$*. Further, the probability of the set of all* $(\mathbf{x}^{nl}, \mathbf{y}^{nl})$ *satisfying*

$$\sum_{i=1}^{l} \mathbf{L}(\mathbf{x}_i^n, \mathbf{y}_i^n) \leq ln(\mathbf{R} + \epsilon \cdot \mathbf{1})$$
$$\sum_{i=1}^{l} d(\theta^n(\mathbf{x}_i^n), \widehat{\theta}_i^n(v)) \leq ln(D_{v,\theta} + \epsilon) \ \forall \ (v, \theta) \in \mathcal{D}$$

is greater than $1 - 2\epsilon$ *by (2.2) and (2.3). Hence the expected distortion vector* $\mathbf{D}(\widehat{\mathcal{C}}_{ln})$ *for* $\widehat{\mathcal{C}}_{ln}$ *satisfies*

$$\mathbf{D}(\widehat{\mathcal{C}}_{ln}) \leq ln\left((1 - 2\epsilon)(\mathbf{D} + 2\epsilon \cdot \mathbf{1}) + (2\epsilon d_{\max}) \cdot \mathbf{1}\right).$$

Since $\epsilon > 0$ *is arbitrary,* $\mathbf{R}$ *is* $\mathbf{D}$*-achievable.* □

2.3 Source Independence and the Relationship Between Lossless and Lossy Source Coding

We begin by proving that when all the source random variables are independent, the lossless rate region $\mathcal{R}_L(P_{\mathbf{X}})$ for the non-functional case depends only on the entropies of X_i for $i \in \{1, \ldots, s\}$. Furthermore, we show that in this case, the lossless rate region is a continuous function of the entropy vector $(H(X_1), \ldots, H(X_s))$. This implies that when sources are independent, separation of network coding and source coding is optimal. (This separation is not optimal in general by [40].)

We next show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0})$ when $\mathcal{N}$ is canonical. Note that by definition, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0})$ since lossless coding requires arbitrarily small block error probability, while lossy coding with $\mathbf{0}$-distortion requires only arbitrarily small average symbol error probability. This property demonstrates the relationship between lossless and lossy source coding: for canonical source codings, lossless source coding is a special case of lossy source coding. This property relies on our prior assumption that the distortion measure satisfies

$$d(a,b) = 0 \;\text{ if and only if } a = b \tag{2.4}$$

$$0 < d_{\min} \leq d(a,b) \leq d_{\max} < \infty \;\forall\; a \neq b. \tag{2.5}$$

2.3.1 $\mathcal{R}_L(P_{\mathbf{X}})$ for Independent Sources

Lemma 2.3.1 shows that introducing a side information vector $\mathbf{Y}$ which is independent of $\mathbf{X}$ cannot improve the lossless and lossy rate regions. More precisely, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}_L(P_{\mathbf{X}})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) = \mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ when $\mathbf{X}$ is independent of $\mathbf{Y}$.

Lemma 2.3.1 *Assume that $\mathbf{Y}$ is independent of $\mathbf{X}$ and $\mathcal{N}$ is a functional or non-functional network source coding problem. Then*

$$\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}_L(P_{\mathbf{X}}) \text{ and } \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) = \mathcal{R}(P_{\mathbf{X}}, \mathbf{D}).$$

Proof. Notice that by definition $\mathcal{R}_L(P_{\mathbf{X}}) \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ since the code for $P_{\mathbf{X},\mathbf{Y}}$ can ignore source $\mathbf{Y}$ and achieve performance identical to that achieved in the same network when $\mathbf{Y}$ is not known. Hence it suffices to show that $\mathcal{R}_L(P_{\mathbf{X}}) \supseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. For any rate vector $\mathbf{R}$ in the interior of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, let $\{\mathcal{C}_n\}$ be a sequence of rate-$\mathbf{R}$, length-n block codes such that $P_{\mathbf{X}^n,\mathbf{Y}^n}(E_n)$ converges to 0 as n grows without bound; here E_n is the event of decoding error using $\mathcal{C}_n$. For any n and any $\mathbf{y}^n$, let $P_{\mathbf{X}^n|\mathbf{Y}^n}(E_n|\mathbf{Y}^n = \mathbf{y}^n)$ be the conditional error probability given $\mathbf{Y}^n = \mathbf{y}^n$, and let $\mathbf{y}^n_*$ be an instance of $\mathbf{Y}^n$ that minimizes $P_{\mathbf{X}^n|\mathbf{Y}^n}(E_n|\mathbf{Y}^n = \mathbf{y}^n)$, i.e.,

$$P_{\mathbf{X}^n|\mathbf{Y}^n}(E_n|\mathbf{Y}^n = \mathbf{y}^n_*) = \min_{\mathbf{y}^n} P_{\mathbf{X}^n|\mathbf{Y}^n}(E_n|\mathbf{Y}^n = \mathbf{y}^n).$$

Now

$$\begin{aligned} P_{\mathbf{X}^n,\mathbf{Y}^n}(E_n) &= \sum_{\mathbf{y}^n} P_{\mathbf{X}^n|\mathbf{Y}^n}(E_n|\mathbf{Y}^n = \mathbf{y}^n) P_{\mathbf{Y}^n}(\mathbf{y}^n) \\ &\geq P_{\mathbf{X}^n|\mathbf{Y}^n}(E_n|\mathbf{Y}^n = \mathbf{y}^n_*). \end{aligned}$$

Define $\mathcal{C}_n(\mathbf{y}_*^n)$ as the block code $\mathcal{C}_n$ when $\mathbf{Y}^n = \mathbf{y}_*^n$. Since $\mathbf{X}$ and $\mathbf{Y}$ are independent,

$$P_{\mathbf{X}^n|\mathbf{Y}^n}(\mathbf{x}^n|\mathbf{y}_*^n) = P_{\mathbf{X}^n}(\mathbf{x}^n)\ \forall\ \mathbf{x}^n.$$

Hence the sequence of rate-$\mathbf{R}$, length-n block codes $\{\mathcal{C}_n(\mathbf{y}_*^n)\}$ for source $\mathbf{X}$ has error probabilities going to 0 as n grows without bound. Therefore, $\mathbf{R} \in \mathcal{R}_L(P_{\mathbf{X}})$. Notice that this argument relies on our assumption to guarantee that $\mathcal{C}_n(\mathbf{y}_*^n)$ has the same rate as $\mathcal{C}_n$.

A similar argument demonstrates that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) = \mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$ when $\mathbf{X}$ and $\mathbf{Y}$ are independent. Here $\mathcal{R}(P_{\mathbf{X}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is immediate, so we need only show $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$. Fix $\epsilon > 0$, and let $\mathbf{R}$ be $\mathbf{D}$-achievable for $P_{\mathbf{X},\mathbf{Y}}$. Let $\mathcal{C}_n$ be a rate-$\mathbf{R}$, length-n block code such that for any $(v, \theta) \in \mathcal{D}$,

$$\frac{1}{n} Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} + \epsilon.$$

By the weak law of large numbers[6], when l is sufficiently large, there exist $\mathbf{y}_1^n, \ldots,$ $\mathbf{y}_l^n$ such that for all $(v, \theta) \in \mathcal{D}$

$$\begin{aligned}
&\frac{1}{ln}\sum_{i=1}^{l} Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)|\mathbf{Y}^n = \mathbf{y}_i^n) \\
&\leq \frac{1}{n}\sum_{\mathbf{y}^n} P_{\mathbf{Y}^n}(\mathbf{y}^n) Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)|\mathbf{Y}^n = \mathbf{y}^n) + \epsilon \\
&= \frac{1}{n} Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) + \epsilon \\
&\leq D_{v,\theta} + 2\epsilon. \qquad (2.6)
\end{aligned}$$

Notice that since $\mathbf{X}$ and $\mathbf{Y}$ are independent, the term

$$\sum_{i=1}^{l} Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)|\mathbf{Y}^n = \mathbf{y}_i^n)$$

[6] If there is only one demand, then we can simply choose a best $\mathbf{y}^n$ that makes the distortion value minimal without applying the weak law of large numbers.

is indeed the expected distortion vector according to the distribution $P_{\mathbf{X}^n}$.

Consider the rate-$\mathbf{R}$, length-ln code composed sequentially from $\mathcal{C}_n(\mathbf{y}_1^n)$, …, $\mathcal{C}_n(\mathbf{y}_l^n)$, where for each $i \in \{1, \ldots, l\}$, $\mathcal{C}_n(\mathbf{y}_i^n)$ is the code $\mathcal{C}_n$ when $\mathbf{Y}^n = \mathbf{y}_i^n$. Since $\epsilon > 0$ is arbitrary, by inequality (2.6), $\mathbf{R}$ is $\mathbf{D}$-achievable. Thus $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X}}, \mathbf{D})$. $\square$

Theorem 2.3.2 shows that when $\mathcal{N}$ is non-functional and sources X_1, …, X_s are independent, $\mathcal{R}_L(P_{\mathbf{X}}) \subseteq \mathcal{R}_L(Q_{\mathbf{X}})$ whenever two distributions $P_{\mathbf{X}}$ and $Q_{\mathbf{X}}$ satisfy $H_P(X_i) \geq H_Q(X_i)$ for all $i \in \{1, \ldots, s\}$; here $H_P(X_i)$ and $H_Q(X_i)$ are the entropies of source X_i for distributions $P_{\mathbf{X}}$ and $Q_{\mathbf{X}}$, respectively. For each $i \in \{1, \ldots, s\}$, we denote by P_i and Q_i the marginal distributions on X_i using $P_{\mathbf{X}}$ and $Q_{\mathbf{X}}$, respectively.

Theorem 2.3.2 *Let $\mathcal{N}$ be a non-functional network source coding problem without side information. Let $P_{\mathbf{X}} = \prod_{i=1}^{s} P_i$ and $Q_{\mathbf{X}} = \prod_{i=1}^{s} Q_i$ be two distributions for independent source $\mathbf{X}$ such that*

$$H_P(X_i) \geq H_Q(X_i) \ \forall i \in \{1, \ldots, s\}.$$

Then $\mathcal{R}_L(P_{\mathbf{X}}) \subseteq \mathcal{R}_L(Q_{\mathbf{X}})$.

Proof. This proof relies on the observations that since $H_Q(X_i) \leq H_P(X_i)$, the typical set $A_{\epsilon,Q}^{(n)}(X_i)$ of X_i corresponding to Q has smaller size than the typical set $A_{\epsilon,P}^{(n)}(X_i)$ corresponding to P. Since X_1, …, X_s are independent, by constructing one-to-one maps from $A_{\epsilon,Q}^{(n)}(X_i)$ to $A_{\epsilon,P}^{(n)}(X_i)$ for each $i \in \{1, \ldots, s\}$ we can construct a one-to-one map from the typical set $A_{\epsilon,Q}^{(n)}(\mathbf{X})$ of $(X_1, \ldots, X_s)$ corresponding to Q to the typical set $A_{\epsilon,P}^{(n)}(\mathbf{X})$ corresponding to P directly. We prove only the special case when $P_i = Q_i$ for all $i \geq 2$. The general case can be proven by applying the result of this special case inductively. In this proof, we use these two observations to build a code $\mathcal{C}_n'$ for the distribution $Q_{\mathbf{X}} = Q \prod_{i=2}^{s} P_i$ from a code $\mathcal{C}_n$ for the distribution $P_{\mathbf{X}} = \prod_{i=1}^{s} P_i$.

Let $H_P = H_P(X_1)$ and $H_Q = H_Q(X_1)$. Let $A_{\epsilon,P}^{(n)}(X_1)$ and $A_{\epsilon,Q}^{(n)}(X_1)$ denote the typical sets for X_1 with respect to P_1 and Q_1, respectively. It suffices to show

$$\mathcal{R}_L(P_{\mathbf{X}}) \subseteq \mathcal{R}_L(Q_{\mathbf{X}})$$

for two distributions P_1 and Q_1 of X_1 such that $H_P \geq H_Q > 0$.

Given any losslessly achievable rate vector $\mathbf{R}$ for $P_{\mathbf{X}}$, let $\epsilon > 0$ be a nonnegative number such that $\epsilon < \min\{1/10, H_Q\}$. Choose n sufficiently large so that

$$P^n(A^{(n)}_{\epsilon,P}(X_1)) \geq 1-\epsilon \tag{2.7}$$

$$Q^n(A^{(n)}_{\epsilon,Q}(X_1)) \geq 1-\epsilon, \tag{2.8}$$

every element $x_1^n \in A^{(n)}_{\epsilon,P}(X_1)$ satisfies $P_1^n(x_1^n) \leq \frac{1}{8}$, and there exists a rate-$\mathbf{R}$, length-n block code $\mathcal{C}_n$ with $P_{\mathbf{X}^n}(E_n) < \epsilon$; here E_n is the event of a decoding error using $\mathcal{C}_n$ when the source distribution is $P_{\mathbf{X}}$. For any $x_1^n \in \mathcal{X}_1^n$, let

$$E(x_1^n) \subseteq \prod_{i=2}^{s} \mathcal{X}_i^n$$

be the collection of vectors $(x_2^n, \ldots, x_s^n)$ for which an error occurs when $(X_1^n, X_2^n, \ldots, X_s^n) = (x_1^n, x_2^n, \ldots, x_s^n)$. Then by definition,

$$\sum_{x_1^n \in A^{(n)}_{\epsilon,P}(X_1)} \sum_{(x_2^n,\ldots,x_s^n)\in E(x_1^n)} \prod_{i=1}^{s} P_i^n(x_i^n) = P_{\mathbf{X}}(E_n) < \epsilon.$$

Let $L = |A^{(n)}_{\epsilon,P}(X_1)|$, and enumerate the typical sequences as $x_1^n(1), \ldots, x_1^n(L)$. For each $j \in \{1, \ldots, L\}$, let

$$e_j = \sum_{(x_2^n,\ldots,x_s^n)\in E(x_i^n(j))} \prod_{i=2}^{s} P_i^n(x_i^n(j)),$$

and choose the order of the enumeration so that

$$e_1 \leq e_2 \leq \cdots \leq e_L.$$

For each $j \in \{1, \ldots, L\}$, set $p_j = P_1^n(x_1^n(j))$. Since $p_j \leq \frac{1}{8}$ for all $1 \leq j \leq L$ by assumption, there exists $1 \leq l \leq L$ for which

$$\frac{1}{4} \leq \sum_{j=l+1}^{L} p_j \leq \frac{1}{2}.$$

Then

$$\epsilon > \sum_{j=l+1}^{L} p_j e_j \geq \left(\sum_{j=l+1}^{L} p_j \right) e_l,$$

which implies that

$$e_l < \frac{\epsilon}{\sum_{j=l+1}^{L} p_j} \leq 4\epsilon. \tag{2.9}$$

Now since $2^{-n(H_P+\epsilon)} \leq p_j \leq 2^{-n(H_P-\epsilon)}$ for all $1 \leq j \leq L$,

$$l 2^{-n(H_P-\epsilon)} \geq \sum_{j=1}^{l} p_j \geq \frac{1}{2} \geq \frac{1}{2} \sum_{j=1}^{L} p_j \geq \frac{L}{2} 2^{-n(H_P+\epsilon)}.$$

Hence

$$l \geq L 2^{-2n\epsilon-1} \geq (1-\epsilon) 2^{nH_P - 3n\epsilon - 1}.$$

Partition the typical set $A_{\epsilon,Q}^{(n)}(X_1)$ as

$$A_{\epsilon,Q}^{(n)}(X_1) = A_1 \cup A_2 \cup \cdots \cup A_K \tag{2.10}$$

such that $|A_r| = l$ for $1 \leq r \leq K-1$ and $|A_K| \leq l$. Then

$$K = \left\lceil \frac{|A_{\epsilon,Q}^{(n)}(X_1)|}{l} \right\rceil \leq \frac{2^{n(H_Q+\epsilon)}}{(1-\epsilon)2^{n(H_P)}} 2^{3n\epsilon+1} + 1 \leq \frac{2^{4n\epsilon+1}}{1-\epsilon} + 1.$$

Let $\iota(x_1^n) = r$ if and only if $x_1^n \in A_r$. Set $B := \{x_1^n(1), \ldots, x_1^n(l)\}$. For each $r \in \{1, \ldots, K\}$, arbitrarily define a one-to-one function η_r from A_r to B. For each $x_1^n \in A_{\epsilon,Q}^{(n)}(X_1)$, let $\eta(x_1^n) = \eta_r(x_1^n)$ where $r = \iota(x_1^n)$. Finally, define the function

$$\phi : A_{\epsilon,Q}^{(n)}(X_1) \to \{1, \ldots, K\} \times B$$

as

$$\phi(x_1^n) = (\iota(x_1^n), \eta(x_1^n)).$$

By construction, ϕ is a one-to-one function.

Now construct a new code $\mathcal{C}'$ for the source distribution $Q_{\mathbf{X}}$ as follows. First, for source sequence $(x_1^n, \ldots, x_s^n)$ such that $x_1^n \in A_{\epsilon,Q}^{(n)}(X_1)$, we apply code $\mathcal{C}_n$ on $(\eta(x_1^n), x_2^n, \ldots, x_s^n)$, and then transmit the index $\iota(x_1^n)$ to every node in the network. For every $(v, X_i) \in \mathcal{D}$ with $i \neq 1$, we use $\mathcal{C}_n$ to reproduce x_i^n. If $(v, X_1) \in \mathcal{D}$, we first apply the decoding function of $\mathcal{C}_n$ to get $\widehat{x}_1^n(v)$, the reproduction of $\eta(x_1^n)$. If $\widehat{x}_1^n(v) \in A_{\epsilon,P}^{(n)}(X)$, then we recover x_1^n by applying the inverse map $\phi^{-1}(\iota(x_1^n), \widehat{x}_1^n(v))$. Otherwise, we declare an error.

The code $\mathcal{C}_n'$ has rate no greater than

$$\mathbf{R} + \frac{1}{n}\left(\log \frac{2^{n4\epsilon+1}}{1-\epsilon} + 1\right)\mathbf{1}$$

and error probability $Q\prod_{i=2}^{s} P_i(E_n)$ bounded by

$$Q\left(X_1^n \notin A_{\epsilon,Q}^{(n)}(X_1)\right) + \sum_{j=1}^{l} Q_1\left(x_1^n(j)\right) e_j \leq 5\epsilon$$

by (2.8) and (2.9). Since $\epsilon > 0$ is arbitrary, $\mathbf{R} \in \mathcal{R}_L(Q_1\prod_{i=2}^{s} P_i)$. □

The argument in the proof of Theorem 2.3.2 works only for independent sources since $(\phi(x_1^n), x_2^n, \ldots, x_n^s)$ need not be typical for $Q_{\mathbf{X}}$ when $(x_1^n, \ldots, x_s^n)$ is typical for $P_{\mathbf{X}}$ in general. Also the argument cannot be directly applied to the functional case since the functional demands cannot be calculated locally. By applying the result of Theorem 2.3.2, we have the following corollary.

Corollary 2.3.3 *Let $\mathcal{N}$ be a non-functional network source coding problem without side information. If two distributions $P_{\mathbf{X}} = \prod_{i=1}^{s} P_i$ and $Q_{\mathbf{X}} = \prod_{i=1}^{s} Q_i$ for independent source $\mathbf{X}$ satisfy $H_P(X_i) = H_Q(X_i)\ \forall\ i \in \{1, \ldots, s\}$, then $\mathcal{R}_L(P_{\mathbf{X}}) = \mathcal{R}_L(Q_{\mathbf{X}})$.*

The proof technique from Theorem 2.3.2 can also be applied to show that $\mathcal{R}_L(\prod_{i=1}^s P_i)$ is continuous in the entropy vector of the independent sources $X_1, \ldots, X_s$. Since each entropy $H_P(X_i)$ is continuous in P_i for $i \in \{1, \ldots, s\}$, this implies that $\mathcal{R}_L(\prod_{i=1}^s P_i)$ is continuous in $(P_1, \ldots, P_s)$ when $\mathcal{N}$ is non-functional, as shown in Theorem 2.3.4.

Theorem 2.3.4 *Let $\mathcal{N}$ be a non-functional network source coding problem and let $P_{\mathbf{X}} = \prod_{i=1}^s P_i$. Then $R_L(P_{\mathbf{X}})$ is continuous in the entropy vector*

$$(H_P(X_1), \ldots, H_P(X_s)).$$

In other words, for any $\epsilon > 0$ and any $P_{\mathbf{X}} = \prod_{i=1}^s P_i$ and $Q_{\mathbf{X}} = \prod_{i=1}^s Q_i$, there exists a $\delta > 0$ such that $R_L(P_{\mathbf{X}})$ and $R_L(Q_{\mathbf{X}})$ are ϵ-close whenever $|H_P(X_i) - H_Q(X_i)| < \delta$ for all $i \in \{1, \ldots, s\}$.

Proof. It suffices to consider the case

$$H_P(X_i) = H_Q(X_i) \ \forall i \in \{2, \ldots, s\}.$$

Let $H_P = H_P(X_1)$ and $H_Q = H_Q(X_1)$ and suppose that

$$H_Q \leq H_P \leq H_Q + \delta$$

for some $\delta > 0$. Then $R_L(P_{\mathbf{X}}) \subseteq R_L(Q_{\mathbf{X}})$ by Theorem 2.3.2. For any achievable rate vector $\mathbf{R} \in R_L(Q_{\mathbf{X}})$, let $\{\mathcal{C}_n\}$ be a sequence of rate-$\mathbf{R}$, length-n codes such that the error probability with respect to $Q_{\mathbf{X}}$ tends to zero as n grows without bound. For any $\tau > 0$ and $n > \frac{1}{\tau}$, since $H_P \leq H_Q + \delta$, we partition the set $A_{\tau,P}^{(n)}(X_1)$ as

$$A_{\tau,P}^{(n)}(X_1) = \cup_{i=1}^{L} A_i^{(n)},$$

where $L = 2^{n(\delta+2\tau)}$, such that each $A_i^{(n)}$ has size smaller than or equal to $A_{\tau,Q}^{(n)}(X_1)$. By building injections from $A_i^{(n)}$ to $A_{\tau,Q}^{(n)}(X_1)$ as in the proof of Theorem 2.3.2 and sending additional rate

$$\Delta\mathbf{R} = \frac{1}{n}\log\frac{2^{n(\delta+2\tau)+1}}{1-\tau}\cdot\mathbf{1}$$

throughout the network to distinguish the sets $A_1^{(n)}, \ldots, A_L^{(n)}$, we get a sequence of new codes $\{\mathcal{C}'_n\}$ of rate $\mathbf{R}+\Delta\mathbf{R}$ whose error probabilities with respect to $P_{\mathbf{X}}$ tend to zero as n grows without bound. That shows $\mathcal{R}_L(P_{\mathbf{X}})$ and $\mathcal{R}_L(Q_{\mathbf{X}})$ are $(\delta+3\tau-\log(1-\tau))$-close. This completes the proof. □

2.3.2 Comparing $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0})$

In this section, we compare $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, which requires asymptotically negligible block error probability, and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D})$, which requires asymptotically negligible per-symbol distortion. We first prove that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) = \mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0})$ for the canonical case. We then explain why this property may not hold for non-canonical coding case.

Theorem 2.3.5 *If $\mathcal{N}$ is canonical, then for all $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$*

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0}) = R_L(P_{\mathbf{X},\mathbf{Y}}).$$

Proof. We begin by proving that the desired result holds for distortion measure d if and only if it holds for the Hamming distortion measure d_H. We then show that if a reproduction with sufficiently low Hamming distortion is available at a demand node, then the additional rate required to achieve a lossless description at that node is negligible.

Let d_H denote the Hamming distance. Recall that distortion measure d satisfies $d_{\min} \le d(a,b) \le d_{\max}$ for all $a \neq b$ in ϑ by assumption, where $d_{\min} > 0$ and $d_{\max} < \infty$. Thus, for any $n > 0$ and any $\mathbf{a},\mathbf{b} \in \vartheta^n$,

$$d_{\min}\cdot d_H(\mathbf{a},\mathbf{b}) \le d(\mathbf{a},\mathbf{b}) \le d_{\max}\cdot d_H(\mathbf{a},\mathbf{b}).$$

Thus, for any two sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$,

$$\lim_{n\to\infty} \frac{1}{n} d((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = 0$$

if and only if

$$\lim_{n\to\infty} \frac{1}{n} d_H((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = 0.$$

As a result, if $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0}) = \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ when $d = d_H$, then the result applies to all distortion measures in the given class.

By definition, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0})$. We therefore need only prove that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0}) \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. Let $\mathbf{R} \in \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0})$ be arbitrary. We next show that for any $\epsilon > 0$, the rate vector $\mathbf{R} + \epsilon \cdot \mathbf{1}$ is in $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. Choose a rate-$(\mathbf{R} + \frac{\epsilon}{2} \cdot \mathbf{1})$, length-$n$ block code $\mathcal{C}$ such that for any $(v', \theta) \in \mathcal{D}$, the reproduction $\widehat{\theta}^n(v')$ of $\theta^n(\mathbf{X}^n)$ at v' satisfies

$$\frac{1}{n} E d_H(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v')) < \tau,$$

where τ is chosen to satisfy

$$ks(H(\tau) + \tau \log(m^s - 1)) < \frac{\epsilon}{2}.$$

Here we recall that k is the total number of reproduction requests, that is, k is the number of pairs $(v', \theta) \in \mathcal{D}$ and that for any $\theta \in \Theta$, $I(\theta)$ is the subset of $\{1, \ldots, s\}$ such that θ is a non-degenerate function of $X_{I(\theta)}$. For any $(v', \theta) \in \mathcal{D}$ and every $i \in I(\theta)$, choose a vertex $v(i, \theta) \in \mathcal{V}$ such that X_i can be observed by v and there is a path from $v(i, \theta)$ to v'. We use $\mathcal{P}(v', \theta, i)$ to denote this path. By assumption and Corollary A.2 in Appendix (notice that the alphabet size of θ is no greater than m^s),

$$\frac{1}{n} H(\theta^n(\mathbf{X}^n) | \widehat{\theta}^n(v')) \leq H(\tau) + \tau \log(m^s - 1).$$

Thus by Lemma 2.2.18, the additional rates along paths $\mathcal{P}(v', \theta, i)$ (for $i \in I(\theta)$) required to achieve a lossless description of $\theta^n(\mathbf{X}^n)$ at v' are asymptotically negligible. Since k and s are finite, repeating this argument for each $(v', \theta) \in \mathcal{D}$ yields a total rate bounded as

$$\begin{aligned}\mathbf{R}' &\leq \mathbf{R} + \frac{\epsilon}{2} + k\left(H(\tau) + \tau \log(m-1)\right) \cdot \mathbf{1} \\ &\leq \mathbf{R} + \epsilon \cdot \mathbf{1}.\end{aligned}$$

Hence $\mathbf{R} + \epsilon \cdot \mathbf{1} \in \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ for all $\epsilon > 0$, and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0}) \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$. □

The argument in the proof of Theorem 2.3.5 cannot be directly applied to the non-canonical source coding case since even when the average conditional entropy of the demanded function sequence given its reproduction is arbitrarily small, the additional rate used to achieve lossless reproduction of that function is not known to be arbitrarily small; the problem is that the demanded function cannot be written as a linear combination of locally calculable functions at source nodes.

2.4 Continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with Respect to $\mathbf{D}$

We next study the continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $\mathbf{D}$. Since $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is convex in $\mathbf{D}$, a naive idea is to apply convexity to show the continuity with respect to $\mathbf{D}$ as in the real-valued function case. Here we aim to prove a stronger continuity, that is, we show that the continuity is independent of $P_{\mathbf{X},\mathbf{Y}}$ and uniform over $\mathbf{D}$, which means that the choice of ϵ and δ in Definition 2.2.13 can be made independent of the values of $P_{\mathbf{X},\mathbf{Y}}$ and $\mathbf{D}$. We treat the canonical and non-canonical cases separately. In the non-canonical ones, it is not clear that whether the choice of ϵ and δ can be independent of $P_{\mathbf{X},\mathbf{Y}}$ when $\mathbf{D}$ is on the boundary of $\mathbb{R}_+^{\mathcal{D}}$, i.e., when some components of $\mathbf{D}$ are zero.

2.4.1 $\mathcal{N}$ is Canonical

Theorem 2.4.1 shows that when $\mathcal{N}$ is canonical, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is uniformly continuous in $\mathbf{D}$ and the continuity is independent of $P_{\mathbf{X},\mathbf{Y}}$. The central insight of the proof is captured in a sequence of intermediate results (Lemmas A.1 - A.4 which culminate in Lemma A.5), all of which are stated and proved in Appendix A. The key idea of the proof is that since applying only convexity and monotonicity is not sufficient to prove the continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ at $\mathbf{D}$ for those $\mathbf{D}$ on the boundary of $\mathbb{R}_+^{\mathcal{D}}$, we increase $\mathbf{D}$ slightly to $\mathbf{D} + \delta \cdot \mathbf{1}$ and show that when $\delta > 0$ is sufficiently small, the difference between $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \delta \cdot \mathbf{1})$ can be made arbitrarily small (see Lemma A.5). Since $\mathbf{D} + \delta \cdot \mathbf{1}$ is not a boundary point of $\mathbb{R}_+^{\mathcal{D}}$, we then directly apply convexity and monotonicity to show the desired result.

Theorem 2.4.1 *Let $\mathcal{N}$ be a canonical network source coding problem. Then $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is uniformly continuous in $\mathbf{D}$. This continuity is further independent of $P_{\mathbf{X},\mathbf{Y}}$.*

Proof. To prove the desired result, we show that for any $\epsilon > 0$, there exists some $\delta > 0$ such that for any $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$, and any $\mathbf{D}, \mathbf{D}' \in \mathbb{R}_+^{\mathcal{D}}$, if $||\mathbf{D} - \mathbf{D}'|| < \delta$, then $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}')$ are ϵ-close.
By Lemma A.5, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and any $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \delta \cdot \mathbf{1})$ are $\epsilon/2$-close. For any $\mathbf{D} = (D_\kappa)_{\kappa \in \mathcal{D}}$ and $\mathbf{D}' = (D'_\kappa)_{\kappa \in \mathcal{D}}$ in $\mathbb{R}_+^{\mathcal{D}}$ such that $||\mathbf{D} - \mathbf{D}'|| \le \delta$, let

$$\mathbf{D}_0 := \min\{\mathbf{D}, \mathbf{D}'\} = (\min\{D_\kappa, D'_\kappa\})_{\kappa \in \mathcal{D}}.$$

Then $||\mathbf{D} - \mathbf{D}_0|| \le \delta$ and $||\mathbf{D}' - \mathbf{D}_0|| \le \delta$. Thus

$$\begin{aligned} \mathbf{D}_0 \le \mathbf{D} \le \mathbf{D}_0 + \delta \cdot \mathbf{1} \\ \mathbf{D}_0 \le \mathbf{D}' \le \mathbf{D}_0 + \delta \cdot \mathbf{1}. \end{aligned}$$

Since

$$\begin{aligned} \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_0) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_0 + \delta \cdot \mathbf{1}) \\ \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_0) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}') \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_0 + \delta \cdot \mathbf{1}), \end{aligned}$$

Lemma A.5 implies that both $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}')$ are $\epsilon/2$-close to

$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_0)$. So $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}')$ are ϵ-close. $\square$

2.4.2 $\mathcal{N}$ is Non-Canonical

The argument in the proof of Theorem 2.4.1 cannot be directly applied to the non-canonical case since the behavior of non-canonical source coding is much different from that of canonical case when one or more distortions is precisely equal to zero. Hence the $P_{\mathbf{X},\mathbf{Y}}$-independence of the uniform continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ can be proved for only those $\mathbf{D}$ in the interior of $\mathbb{R}_+^{\mathcal{D}}$, i.e., for those $\mathbf{D}$ with all non-zero components. In Theorem 2.4.2, we prove this result by using a different metric on $\mathbb{R}_+^{\mathcal{D}}$, showing that for all $P_{\mathbf{X},\mathbf{Y}}$, whenever $\mathbf{D}$ and $\mathbf{D}'$ are sufficiently close in the L_2-norm sense and have the same supports, the corresponding rate regions are arbitrarily close even if $\mathbf{D}$ has some zero components.

Theorem 2.4.2 *Let $\mathcal{N}$ be a non-canonical network source coding problem. Let $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}} \backslash \{\mathbf{0}\}$. For any $\epsilon > 0$, there exists a $\delta > 0$ such that for any $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and any $\mathbf{D}' \in \mathbb{R}_+^{\mathcal{D}}$ satisfying $(1-\delta)\mathbf{D} \leq \mathbf{D}' \leq (1+\delta)\mathbf{D}$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}')$ are ϵ-close.*

Proof. For any $\epsilon > 0$, let $\delta > 0$ be sufficiently small so that $\delta k \log(m) < \epsilon$. Then for any $\mathbf{D}' \in \mathbb{R}_+^{\mathcal{D}}$ such that $(1-\delta)\mathbf{D} \leq \mathbf{D}' \leq (1+\delta)\mathbf{D}$,

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, (1-\delta)\mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}') \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, (1+\delta)\mathbf{D}).$$

By concavity,

$$\begin{aligned}
&\frac{1}{1+\delta}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, (1+\delta)\mathbf{D}) + \frac{\delta}{1+\delta}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \\
&(1-\delta)\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) + \delta\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, (1-\delta)\mathbf{D}).
\end{aligned}$$

Therefore,

$$\frac{1}{1+\delta}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}') + \frac{\delta}{1+\delta}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D})$$
$$(1-\delta)\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}) + \delta\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}').$$

Since $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}) \subseteq a\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D})$ for all $0 \leq a \leq 1$ and for all $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$,

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}') \subseteq \frac{1}{1+\delta}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}')$$
$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}) \subseteq (1-\delta)\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}),$$

and hence

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}') + \frac{\delta}{1+\delta}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D})$$
$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}) + \delta\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}').$$

Now, $k\log(m)\cdot\mathbf{1} \in \mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0})$ and $\delta k\log(m) < \epsilon$, together imply

$$\epsilon\cdot\mathbf{1} \in \left(\frac{\delta}{1+\delta}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0})\right) \cap \left(\delta\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{0})\right).$$

Therefore, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}')$ are ϵ-close. $\square$

Notice that by definition of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D})$,

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}) = \cap_{\mathbf{D}'\geq\mathbf{D}}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D}').$$

Additionally, the space $\mathcal{R}_+^{\mathcal{D}}$ is compact under L_2-norm. Therefore, Corollary 2.4.3 follows.

Corollary 2.4.3 *Let $\mathcal{N}$ be a network source coding problem. Fix distribution $P_{\mathbf{X},\mathbf{Y}}$. The lossy rate region $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}},\mathbf{D})$ is uniformly continuous in $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$.*

2.5 Continuity with Respect to $P_{\mathbf{X},\mathbf{Y}}$

In this section, we investigate the continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$. We begin by defining inner semi-continuity and outer semi-continuity.

Definition 2.5.1 *For any sequence of sets $\{A_\ell\}_{\ell=1}^{\infty}$, define*

$$\liminf_{\ell\to\infty} A_\ell := \cup_{\ell=1}^{\infty} \cap_{l\geq\ell} A_l$$
$$\limsup_{\ell\to\infty} A_\ell := \cap_{\ell=1}^{\infty} \cup_{l\geq\ell} A_l.$$

If both $\liminf_{\ell\to\infty} A_\ell$ and $\limsup_{\ell\to\infty} A_\ell$ exist and $\limsup_{\ell\to\infty} A_\ell = \liminf_{\ell\to\infty} A_\ell = A^*$ for some A^*, we say that $\lim_{\ell\to\infty} A_\ell$ exists and define

$$\lim_{\ell\to\infty} A_\ell := A^*.$$

Definition 2.5.2 *Set function $A(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathbb{R}^{\mathcal{E}}$ is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ if and only if for every $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and any sequence $\{P^{(\ell)}_{\mathbf{X},\mathbf{Y}}\} \subset \mathcal{M}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$*

$$A(P_{\mathbf{X},\mathbf{Y}}) \subseteq \liminf_{\ell\to\infty} A(P^{(\ell)}_{\mathbf{X},\mathbf{Y}}).$$

Definition 2.5.3 *Set function $A(P_{\mathbf{X},\mathbf{Y}}) \subseteq \mathbb{R}^{\mathcal{E}}$ is outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ if and only if for every $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and any sequence $\{P^{(\ell)}_{\mathbf{X},\mathbf{Y}}\} \subset \mathcal{M}$ that converges to $P_{\mathbf{X},\mathbf{Y}}$*

$$A(P_{\mathbf{X},\mathbf{Y}}) \supseteq \limsup_{\ell\to\infty} A(P^{(\ell)}_{\mathbf{X},\mathbf{Y}}).$$

By definition, $A(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ if and only if it is both inner and outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$.

Note that when the alphabet size is infinite, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is not necessarily continuous in $P_{\mathbf{X},\mathbf{Y}}$ even for the point-to-point network with a single source, as shown in Example 2.5.4.

Example 2.5.4 *Consider the point-to-point network with a single source, shown in Fig. 2.4. For this network, $\mathcal{N} = (\mathcal{V}, \mathcal{E}, \mathcal{S}, \mathcal{D}) = ((\{v_1, v_2\}, \{(v_1, v_2)\}), \{(v_1, X_1)\}, \{(v_2, X_1)\})$*

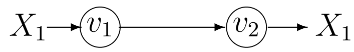


Figure 2.4: The point-to-point network

and $\mathbf{X} = X_1$. *Let the source alphabet* $\mathcal{X}_1$ *be the non-negative integers* $\mathbb{Z}_+$. *By Shannon's source coding theorem,* $\mathcal{R}_L(P_X) = \{R_{(v_1,v_2)} : \ R_{(v_1,v_2)} \geq H(X)\}$. *When* $P_{\mathbf{X}}$ *is the distribution that places probability* 1 *on the symbol* 1, $\mathcal{R}_L(P_{\mathbf{X}}) = [0, \infty)$. *Let* $M > 0$ *be fixed. We next demonstrate the existence of a sequence* $\{P^l_{\mathbf{X}}\}_{l=1}^{\infty}$ *for which* $\lim_{l\to\infty} P^l_{\mathbf{X}} = P_{\mathbf{X}}$ *while* $\mathcal{R}_L(P^l_{\mathbf{X}}) \subset [M, \infty)$ *for all* l. *To show such a sequence exists, we show that for any* $\epsilon > 0$, *there exists a finite-support distribution* $P^{\epsilon}_{\mathbf{X}}$ *on* $\mathbb{Z}_+$, *the set of nonnegative integers, such that* $\sum_{i=0}^{\infty} |P_{\mathbf{X}}(i) - P^{\epsilon}_{\mathbf{X}}(i)| \leq 2\epsilon$ *and* $H(P^{\epsilon}_{\mathbf{X}}) \geq M$. *This shows that the entropy for the distribution* $P^{\epsilon}_{\mathbf{X}}$ *can be arbitrarily large even when the distance between* $P^{\epsilon}_{\mathbf{X}}$ *and* $P_{\mathbf{X}}$ *is small.*

Given any finite-support distribution $\mathbf{q} = (q_1, q_2, \ldots)$, *consider the random variable* X *with probability distribution* $\mathbf{p} = (p_0, p_1, \ldots)$ *defined by*

$$p_0 = 1 - \epsilon, \ \ p_i = \epsilon q_i \ \ \text{for } i \geq 1.$$

Then

$$\begin{aligned} H(X) &= (1-\epsilon)\log\left(\frac{1}{1-\epsilon}\right) + \epsilon\left(\sum_{i=1}^{\infty} q_i \log\left(\frac{1}{\epsilon q_i}\right)\right) \\ &= H(\epsilon) + \epsilon H(\mathbf{q}). \end{aligned}$$

Let $M' = \max\{M, H(\epsilon) + 1\}$. *If we can find some finite-support distribution* $\mathbf{q}$ *such that*

$$H(\mathbf{q}) \geq \frac{M' - H(\epsilon)}{\epsilon},$$

then $P^{\epsilon}_{\mathbf{X}} = \mathbf{p}$ *satisfies both* $H(P^{\epsilon}_{\mathbf{X}}) \geq M$ *and* $\sum_{i=1}^{\infty} |P^{\epsilon}_{\mathbf{X}}(i) - P_{\mathbf{X}}(i)| = 2\epsilon$ *as desired. To construct such a distribution* $\mathbf{q}$, *let* $L = \lceil 2^{(M' - H(\epsilon))/\epsilon} \rceil$ *and* $q_i = \frac{1}{L}$ *for all* $1 \leq i \leq L$ *and* $q_i = 0$ *otherwise.* $\square$

2.5.1 Inner Semi-Continuity

For any length-n block code $\mathcal{C}_n$ and for every edge $e \in \mathcal{E}$ and demand $(v, \theta) \in \mathcal{D}$, let F_e denote the random variable that represents the encoded message on e and $\widehat{\theta}^n(v)$ denote the reproduction of $\theta^n(\mathbf{X}^n)$ at node v using the block code $\mathcal{C}_n$. Recall that for each $(v, \theta) \in \mathcal{D}$ and each $i \in I(\theta)$, there exist a pair $(v', X_i) \in \mathcal{S}$ and a path, denoted by $\mathcal{P}(v, \theta, i)$, that starts from v' and ends at v. If $i \notin I(\theta)$, we define $\mathcal{P}(v, \theta, i) = \emptyset$ for all $v \in \mathcal{V}$. For every $e \in \mathcal{E}$, every $i \in \{1, \ldots, s\}$, and any $\mathbf{D} = (D_\kappa)_{\kappa \in \mathcal{D}} \in \mathrm{IR}_+^{\mathcal{D}}$, let

$$\omega(e, i, \mathbf{D}) := \{(v, \theta) \in \mathcal{D} \mid e \in \mathcal{P}(v, \theta, i) \text{ and } D_{(v,\theta)} = 0\}.$$

We start by proving inner semi-continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$. For every subset A of $\mathrm{IR}_+^{\mathcal{E}}$, let $\overline{A}$ denote the Euclidean closure of A in $\mathrm{IR}_+^{\mathcal{E}}$. Define

$$\begin{aligned}\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \; := \; & \overline{\left\{\mathbf{R} = (R_e)_{e \in \mathcal{E}} \mid R_e \geq \frac{1}{n}\left(H_P(F_e) + \sum_{i=1}^{s} \sum_{(v,\theta) \in \omega(e,i,\mathbf{D})} H_P(\theta^n | \widehat{\theta}^n(v))\right)\right.} \\ & \overline{\frac{1}{n} E_P d(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} \ \forall (v, \theta) \in \mathcal{D} \text{ such that } D_{v,\theta} \neq 0} \\ & \overline{\left.\text{for some length-}n \text{ block code } \mathcal{C}\right\}}.\end{aligned}$$

In Lemma 2.5.5, we show that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) = \mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ when $\mathcal{N}$ is canonical and when $\mathcal{N}$ is non-canonical but $\mathbf{D} > \mathbf{0}$. In the proof, we apply the fact that $\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is itself continuous in $\mathbf{D}$. Since the proof technique for the continuity of $\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $\mathbf{D}$ is similar to Section 2.4, we state it without proof in Appendix B.

Lemma 2.5.5 *Let $\mathcal{N}$ be a network source coding problem and fix $\mathbf{D} \in \mathrm{IR}_+^{\mathcal{D}}$. If $\mathcal{N}$ is canonical and $\mathbf{D}$ is arbitrary or $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$, then*

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) = \mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}).$$

Proof. Let $\mathbf{D} \in \mathrm{IR}_+^{\mathcal{D}}$. If $\mathcal{N}$ is non-canonical, let $\mathbf{D} > \mathbf{0}$, otherwise let $\mathbf{D}$ be arbitrary. Let $\mathbf{R}$ be any $\mathbf{D}$-achievable rate vector. Then by definition of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$, for any

$\epsilon > 0$, there exists a rate-$\mathbf{R}$, length-n block code $\mathcal{C}$ such that

$$\frac{1}{n}Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} + \epsilon \ \forall (v,\theta) \in \mathcal{D}.$$

For every $(v,\theta) \in \mathcal{D}$ such that $D_{v,\theta} = 0$,

$$\frac{1}{n}H(\theta^n|\widehat{\theta}^n(v)) \leq H\left(\frac{\epsilon}{d_{\min}}\right) + \frac{\epsilon}{d_{\min}}\log(m^s - 1)$$

by Corollary A.2. Thus,

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) + k\left(H\left(\frac{\epsilon}{d_{\min}}\right) + \frac{\epsilon}{d_{\min}}\log(m^s - 1)\right) \cdot \mathbf{1} \subseteq \mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \epsilon \cdot \mathbf{1}) \ \forall \epsilon > 0.$$

By Lemma B.1,

$$\lim_{\epsilon \to 0} \mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \epsilon \cdot \mathbf{1}) = \mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}).$$

Hence

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq \mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}).$$

On the other hand, any $\mathbf{R}$ in the interior of $\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is $\mathbf{D}$-achievable. (This follows from Lemma 2.2.18 when $\mathcal{N}$ is canonical and $\mathbf{D}$ has zero components.) Thus $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \supseteq \mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$. □

We next use $\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ to prove that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$ when $\mathcal{N}$ is canonical or when $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$.

Theorem 2.5.6 *Fix $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$. If $\mathcal{N}$ is canonical and $\mathbf{D}$ is arbitrary or $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$, then $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$.*

Proof. The proofs of the two cases are similar. We give details only for the canonical case. Let $\mathbf{R}=(R_e)_{e\in\mathcal{E}} \in \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$. By Lemma 2.5.5, for any $\epsilon > 0$, there exist an n and a length-n block code $\mathcal{C}_n$ such that

$$R_e + \epsilon \geq \frac{1}{n}\left(H_P(F_e) + \sum_{i=1}^{s} \sum_{(v,\theta)\in\omega(e,i,\mathbf{D})} H_P(\theta^n|\widehat{\theta}^n(v))\right)$$

and for all $(v,\theta) \in \mathcal{D}$ such that $D_{v,\theta} \neq 0$

$$\frac{1}{n} E_P d(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} + \epsilon.$$

For this code $\mathcal{C}_n$, the transition probability $P_{F_e|\mathbf{X}^n,\mathbf{Y}^n}$ for all $e \in \mathcal{E}$ is fixed. For all $e \in \mathcal{E}$, since $\lim_{\ell\to\infty} P^{(\ell)}_{\mathbf{X},\mathbf{Y}} = P_{\mathbf{X},\mathbf{Y}}$, the joint probability

$$P^{(\ell)}_{\mathbf{X}^n,\mathbf{Y}^n,F_e} := P_{\mathbf{X}^n,\mathbf{Y}^n} P_{F_e|\mathbf{X}^n,\mathbf{Y}^n}$$

satisfies

$$\lim_{\ell\to\infty} P^{(\ell)}_{\mathbf{X}^n,\mathbf{Y}^n,F_e} = P_{\mathbf{X}^n,\mathbf{Y}^n,F_e},$$

where

$$P_{\mathbf{X}^n,\mathbf{Y}^n,F_e} := P_{\mathbf{X}^n,\mathbf{Y}^n} P_{F_e|\mathbf{X}^n,\mathbf{Y}^n}.$$

Thus,

$$\lim_{\ell\to\infty} \frac{1}{n} H_{P^{(\ell)}}(F_e) = \frac{1}{n} H_P(F_e). \tag{2.11}$$

Likewise,

$$\lim_{\ell\to\infty} \frac{1}{n} \left(\sum_{i=1}^{s} \sum_{(v,\theta)\in\omega(e,i,\mathbf{D})} H_{P^{(\ell)}}(\theta^n | \widehat{\theta}^n(v)) \right) = \frac{1}{n} \left(\sum_{i=1}^{s} \sum_{(v,\theta)\in\omega(e,i,\mathbf{D})} H_P(\theta^n | \widehat{\theta}^n(v)) \right)$$
$$\lim_{\ell\to\infty} \frac{1}{n} E d_{P^{(\ell)}}(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} \quad \forall (v,\theta) \in \mathcal{D}$$

by a similar argument. Hence there exists an integer ℓ' such that for all $\ell \geq \ell'$,

$$R_e + 2\epsilon \geq \frac{1}{n} \left(H_{P^{(\ell)}}(F_e) + \sum_{i=1}^{s} \sum_{(v,\theta)\in\omega(e,i,\mathbf{D})} H_{P^{(\ell)}}(\theta^n | \widehat{\theta}^n(v)) \right) \quad \forall e \in \mathcal{E}$$
$$\frac{1}{n} E_{P^{(\ell)}} d(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_{v,\theta} + 2\epsilon \quad \forall (v,\theta) \in \mathcal{D}.$$

By Lemma 2.5.5,

$$\mathbf{R} + 2\epsilon \cdot \mathbf{1} \in \bigcap_{\ell \geq \ell'} \mathcal{R}(P^{(\ell)}, \mathbf{D} + 2\epsilon \cdot \mathbf{1}).$$

By Theorem 2.4.2, when $\epsilon > 0$ is sufficiently small,

$$\mathbf{R} + 3\epsilon \cdot \mathbf{1} \in \bigcap_{\ell \geq \ell'} \mathcal{R}(P^{(\ell)}, \mathbf{D}).$$

Therefore,

$$\mathcal{R}(P, \mathbf{D}) \subseteq \liminf_{\ell \to \infty} \mathcal{R}(P^{(\ell)}, \mathbf{D}).$$

$\square$

Corollary 2.5.7 *If $\mathcal{N}$ is canonical, then $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is inner semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$.*

Proof. The result follows immediately from Theorem 2.5.6 and Theorem 2.3.5. $\square$

2.5.2 Outer Semi-Continuity

We next study outer semi-continuity with respect to source distribution $P_{\mathbf{X},\mathbf{Y}}$. We use the intermediate network source coding problem $\widehat{\mathcal{N}}$ defined below.

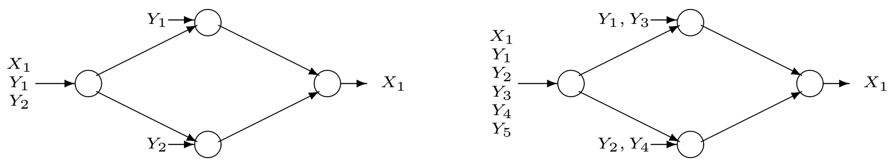
Definition 2.5.8 *Let $\mathcal{N} = (G, \mathcal{S}, \mathcal{D})$ be a fixed network source coding problem with s source and t side-information random variables. Define $\widehat{\mathcal{N}} = (G, \widehat{\mathcal{S}}, \mathcal{D})$ to be the network source coding problem with graph G, demand $\mathcal{D}$, and s source and $2t + s$ side-information random variables given by*

$$\widehat{\mathcal{S}} \;:=\; \mathcal{S} \cup \bigcup_{(v, Y_j) \in \mathcal{S}} \{(v, Y_{j+t})\} \cup \bigcup_{(v, X_i) \in \mathcal{S}} \{(v, Y_{2t+i})\}.$$

An example of $\mathcal{N}$ and $\widehat{\mathcal{N}}$ is in Fig. 2.5.

Theorems 2.5.9 - 2.5.11 show the outer semi-continuity with respect to $P_{\mathbf{X},\mathbf{Y}}$ for the lossless, lossy, and zero-error rate regions. The proof of Theorem 2.5.9 relies on Lemmas C.1 and C.2.

Theorem 2.5.9 *Rate region $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$ for all $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$ when $\mathcal{N}$ is canonical or non-canonical.*



(a) $\mathcal{N}$ (b) $\widehat{\mathcal{N}}$

Figure 2.5: The diamond network $\mathcal{N}$ and it corresponding network $\widehat{\mathcal{N}}$

Proof. Let $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and $\{P^{(\ell)}_{\mathbf{X},\mathbf{Y}}\}_{\ell=1}^{\infty} \subset \mathcal{M}$ be a sequence of distributions such that

$$\lim_{\ell\to\infty} P^{(\ell)}_{\mathbf{X},\mathbf{Y}} = P_{\mathbf{X},\mathbf{Y}}. \tag{2.12}$$

We aim to show that for any $\epsilon > 0$, there exists an integer ℓ_0 such that

$$\mathcal{R}(P^{(\ell)}_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$$

for all $\ell > \ell_0$. For $\epsilon > 0$, let ℓ be sufficiently large such that

$$P^{(\ell)}_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \geq (1-\epsilon) P_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \ \forall\ (\mathbf{x}, \mathbf{y}) \in \mathcal{A}. \tag{2.13}$$

By Lemma C.1 in Appendix C, there exists a distribution $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ such that the marginal of $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ on $(\mathbf{X}, \mathbf{Y})$ is $P^{(\ell)}_{\mathbf{X},\mathbf{Y}}$, the conditional distribution on $(\mathbf{X}, \mathbf{Y})$ given the event $\{(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}', \mathbf{Y}')\}$ is $P_{\mathbf{X},\mathbf{Y}}$, and $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}((\mathbf{X}, \mathbf{Y}) \neq (\mathbf{X}', \mathbf{Y}')) = \epsilon$. We next use this joint distribution $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ to compare the rate regions $\mathcal{R}(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$, $\mathcal{R}(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$, and $\mathcal{R}(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}, \mathbf{D})$. By definition,

$$\mathcal{R}(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq \mathcal{R}(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}, \mathbf{D}) \tag{2.14}$$

since network $\widehat{\mathcal{N}}$ has more side-information random variables than $\mathcal{N}$ and has the same demands as $\mathcal{N}$. By Lemma C.2 in Appendix C,

$$\frac{1}{1-2\epsilon}\mathcal{R}(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}, \mathbf{D}) \subseteq \mathcal{R}(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}, \frac{1}{1-2\epsilon}\mathbf{D}). \tag{2.15}$$

Together, (2.14) and (2.15) imply that

$$\mathcal{R}(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq (1-2\epsilon)\mathcal{R}(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}, \frac{1}{1-2\epsilon}\mathbf{D}).$$

This completes the proof by Theorem 2.4.2. $\square$

Theorem 2.5.10 *$\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$ when $\mathcal{N}$ is canonical or non-canonical.*

Proof. The proof is similar to that of Theorem 2.5.9. Let $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and $\{P^{(\ell)}_{\mathbf{X},\mathbf{Y}}\}_{\ell=1}^{\infty} \subset \mathcal{M}$ be a sequence of distributions such that

$$\lim_{\ell\to\infty} P^{(\ell)}_{\mathbf{X},\mathbf{Y}} = P_{\mathbf{X},\mathbf{Y}}. \tag{2.16}$$

We aim to show that for any $\epsilon > 0$, there exists an integer l_0 such that

$$\mathcal{R}_L(P^{(\ell)}_{\mathbf{X},\mathbf{Y}}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}},)$$

for all $\ell > \ell_0$. For $\epsilon > 0$, let ℓ be sufficiently large such that (2.13) holds. By Lemma C.1, there exists a distribution $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ such that the marginal of $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ on $(\mathbf{X},\mathbf{Y})$ is $P^{(\ell)}_{\mathbf{X},\mathbf{Y}}$, the conditional distribution on $(\mathbf{X},\mathbf{Y})$ given $(\mathbf{X},\mathbf{Y}) = (\mathbf{X}',\mathbf{Y}')$ is $P_{\mathbf{X},\mathbf{Y}}$, and $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}((\mathbf{X},\mathbf{Y}) \neq (\mathbf{X}',\mathbf{Y}')) = \epsilon$. We next use this joint distribution $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ to compare the rate regions $\mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}_L(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'})$. By definition,

$$\mathcal{R}_L(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}_L(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}) \tag{2.17}$$

since network $\widehat{\mathcal{N}}$ has more side-information random variables than $\mathcal{N}$ and has the same demands as $\mathcal{N}$. By Lemma C.3,

$$\frac{1}{1-2\epsilon}\mathcal{R}_L(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}) \subseteq \mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}). \tag{2.18}$$

Together, (2.17) and (2.18) imply that

$$\mathcal{R}_L(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}}) \subseteq (1-2\epsilon)\mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}).$$

$\square$

Theorem 2.5.11 *$\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ is outer semi-continuous in $P_{\mathbf{X},\mathbf{Y}}$ when $\mathcal{N}$ is canonical or non-canonical.*

Proof. The proof is similar to that of Theorem 2.5.9. Let $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and $\{P^{(\ell)}_{\mathbf{X},\mathbf{Y}}\}_{\ell=1}^{\infty} \subset \mathcal{M}$ be a sequence of distributions such that

$$\lim_{\ell\to\infty} P^{(\ell)}_{\mathbf{X},\mathbf{Y}} = P_{\mathbf{X},\mathbf{Y}}. \tag{2.19}$$

We aim to show that for any $\epsilon > 0$, there exists l_0 such that

$$\mathcal{R}_Z(P^{(\ell)}_{\mathbf{X},\mathbf{Y}}) + \epsilon\cdot\mathbf{1} \subseteq \mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}},)$$

for all $\ell > \ell_0$. For $\epsilon > 0$, let ℓ be sufficiently large such that (2.13) holds. By Lemma C.1, there exists a distribution $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ such that the marginal of $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ on $(\mathbf{X},\mathbf{Y})$ is $P^{(\ell)}_{\mathbf{X},\mathbf{Y}}$, the conditional distribution on $(\mathbf{X},\mathbf{Y})$ given $(\mathbf{X},\mathbf{Y}) = (\mathbf{X}',\mathbf{Y}')$ is $P_{\mathbf{X},\mathbf{Y}}$, and $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}((\mathbf{X},\mathbf{Y}) \neq (\mathbf{X}',\mathbf{Y}')) = \epsilon$. We next use this joint distribution $T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ to compare the rate regions $\mathcal{R}_Z(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_Z(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}})$, and $\mathcal{R}_Z(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'})$. By definition,

$$\mathcal{R}_Z(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}}) \subseteq \mathcal{R}_Z(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}) \tag{2.20}$$

since network $\widehat{\mathcal{N}}$ has more side-information random variables than $\mathcal{N}$ and has the same demands as $\mathcal{N}$. By Lemma C.4,

$$\frac{1}{1-2\epsilon}\mathcal{R}_Z(\widehat{\mathcal{N}}, T^{(\ell)}_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}) \subseteq \mathcal{R}_Z(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}). \tag{2.21}$$

Together, (2.20) and (2.21) imply that

$$\mathcal{R}_Z(\mathcal{N}, P^{(\ell)}_{\mathbf{X},\mathbf{Y}}) \subseteq (1-2\epsilon)\mathcal{R}_Z(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}).$$

□

Corollary 2.5.12 *If $\mathcal{N}$ is canonical and $\mathbf{D} \in \mathbb{R}^{\mathcal{E}}_+$ or $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$, then $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$.*

Proof. The result follows immediately from Theorems 2.5.6 and 2.5.9. □

Theorem 2.5.13 *If $\mathcal{N}$ is canonical, then $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is continuous in $P_{\mathbf{X},\mathbf{Y}}$.*

Proof. The result follows immediately from Theorem 2.3.5 and Corollary 2.5.12. □

2.6 S-Continuity with Respect to $P_{\mathbf{X},\mathbf{Y}}$

Using the results from Section 2.5.2, we next show that $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ for canonical $\mathcal{N}$ and any $\mathbf{D} \in \mathbb{R}^{\mathcal{D}}_+$ are s-continuous in $P_{\mathbf{X},\mathbf{Y}}$. We next show that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ for non-canonical $\mathcal{N}$ and any $\mathbf{D} > \mathbf{0}$ is s-continuous in $P_{\mathbf{X},\mathbf{Y}}$.

Theorem 2.6.1 *$\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ is s-continuous in $P_{\mathbf{X},\mathbf{Y}}$ for any network source coding problem $\mathcal{N}$.*

Proof. Suppose for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{A}$,

$$(1-\epsilon)P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \leq Q_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \leq \frac{1}{1-\epsilon}P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}).$$

By the argument of Theorem 2.5.11,

$$\mathcal{R}_Z(Q_{\mathbf{X},\mathbf{Y}}) \subseteq (1-2\epsilon)\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$$
$$\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}}) \subseteq (1-2\epsilon)\mathcal{R}_Z(Q_{\mathbf{X},\mathbf{Y}}).$$

This completes the proof. □

The proofs of Theorems 2.6.2 and 2.6.3 are almost identical to that of Theorem 2.6.1.

Theorem 2.6.2 *$\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ is s-continuous in $P_{\mathbf{X},\mathbf{Y}}$ for any network source coding problem $\mathcal{N}$.*

Theorem 2.6.3 *$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is s-continuous in $P_{\mathbf{X},\mathbf{Y}}$ for all $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$ and for any network source coding problem $\mathcal{N}$.*

2.7 Summary

In this chapter, we introduce a family of finite-alphabet network source coding problems that includes prior example problems as special cases. We define the zero-error rate region $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, lossless rate region $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, and lossy rate region $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ for all the members in the family and then study the continuity and s-continuity properties of those objects. We began by proving the continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $\mathbf{D}$ when (a) $\mathcal{N}$ is canonical or (b) $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$. We proved that $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ (for all $\mathbf{D}$) are all s-continuous with respect to $P_{\mathbf{X},\mathbf{Y}}$ for any network source coding problem $\mathcal{N}$. We summarize our results on the continuity with respect to $P_{\mathbf{X},\mathbf{Y}}$ in the following tables. The two entries marked "?" remain open problems.

	Canonical	Non-canonical
Inner Semi-continuous	No	No
Outer Semi-continuous	Yes	Yes
Continuous	No	No
S-continuous	Yes	Yes

Table 2.1: Continuity of $\mathcal{R}_Z(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$.

	Canonical	Non-canonical
Inner Semi-continuous	Yes	No
Outer Semi-continuous	Yes	Yes
Continuous	Yes	No
S-continuous	Yes	Yes

Table 2.2: Continuity of $\mathcal{R}_L(P_{\mathbf{X},\mathbf{Y}})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$.

	Canonical	Non-canonical (with $\mathbf{D} > \mathbf{0}$)	Non-canonical (with $\mathbf{D}$ on boundary of $\mathbb{R}_+^{\mathcal{D}}$)
Inner Semi-continuous	Yes	Yes	?
Outer Semi-continuous	Yes	Yes	Yes
Continuous	Yes	Yes	?
S-continuous	Yes	Yes	Yes

Table 2.3: Continuity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $P_{\mathbf{X},\mathbf{Y}}$.

Chapter 3

A Strong Converse

3.1 Introduction

In the traditional source coding scenario, here called the point-to-point network, comprised of one source node describing source X to one sink node across a single link, strong converses for both lossless source coding and lossy source coding have previously appeared in the literature. For example, [19] treats i.i.d. finite-alphabet source sequences, and [41], [42], and [43] treat more general source sequences. While the lossless source coding theorem describes the family of rates that can be achieved with arbitrarily small error probability, the strong converse states that for any rate outside the rate region, the probability of a correct reconstruction approaches 0 as the blocklength grows without bound. Similarly, the rate-distortion theorem describes the set of rates that can be achieved with expected distortion no greater than D while its strong converse demonstrates that the probability of observing distortion less than D at any rate outside this region approaches 0 as the blocklength grows without bound.

In this paper, we derive a strong converses for three problems: the Ahlswede-Körner (coded side information) problem, lossless source coding for multicast networks with side-information at the end nodes, and the Gray-Wyner problem. The source sequences are drawn i.i.d. according to a finite-alphabet source distribution. Generalized from the strong converse for lossy source coding of the point-to-point network in [19], the strong converses of interest state that for any distortion vector $\mathbf{D}$ when a rate vector $\mathbf{R}$ is not in the $\mathbf{D}$-achievable rate region, the probability of observing distortion at most $\mathbf{D}$ with a rate-$\mathbf{R}$ code decreases exponentially to 0 as the blocklength n grows without bound. We call such a result an exponentially strong converse to emphasize the speed of convergence of the correct probabilities for the rate vectors outside the rate region. The exponentially strong converse for a network source coding problem is useful for a variety of applications beyond basic understanding of how achievable error probability varies with rate. For example, when the exponentially strong converse holds, we can show that any demands that can be achieved across a network with rate 0 across a given link can also be achieved when that link is absent [18]. This property is actually quite subtle since it requires demonstrating that asymptotically small rates across the given link are never critical to that network's operation. Notice that this property is not trivial since single-letter characterizations of the network with that additional link may not be available even when a single-letter characterization of the network without the 0-rate link is known. As mentioned above, the exponentially strong converse holds for the point-to-point network. For the point-to-point lossless case, the intuition is that the probability of the strongly typical set $A_\epsilon^{*(n)}(X)$ for the finite-alphabet source X increases exponentially to 1 as the length n grows without bound. We denote exponent by $\tau(\epsilon) > 0$. Let $B_\epsilon^{(n)}$ denote the intersection of $A_\epsilon^{*(n)}(X)$ and the correct event for a given sequence of codes. When the code's probability of correctness equals $2^{-nc(n)}$ for some $c(n) \to 0$,

$$\frac{1}{n} \log \frac{|A_\epsilon^{*(n)}(X)|}{|B_\epsilon^{(n)}|}$$

can be made arbitrarily small when $\epsilon > 0$ is sufficiently small and n is sufficiently large. This means that the rate R that is sufficient to describe the set $B_\epsilon^{(n)}$ is asymptotically at least $\frac{1}{n} \log |A_\epsilon^{*(n)}(X)|$.

We prove that the exponentially strong converse holds for the lossless coded side information problem [4], lossless source coding for the family of multicast networks with side information at the end nodes, and the Gray-Wyner problem [3]. The cut-set bound is tight for multicast networks with side information at the end nodes by [12], and this family includes the family of multicast networks [11] as a subfamily, which includes the Slepian-Wolf problem [2] as a special case. The strong converse has been proven for the coded side information problem in [20] and for the Slepian-Wolf problem in [21]. Neither of these results shows exponential decay.

The remainder of this paper is structured as follows. We define the exponentially strong converse in Section 3.2. We briefly explain an application in Section 3.3. We show that the exponentially strong converse is true for the coded side information problem in Section 3.4. In Section 3.5, we prove the exponentially strong converse for the family of multicast networks with side information at the end nodes. Finally, we prove the exponentially strong converse for the Gray-Wyner problem in Section 3.6.

3.2 Definition and Problem Statement

Here we define the the exponentially strong converse for general non-functional lossless source coding problems defined in Section 2.2. For simplicity, when $s = 1$ (resp. $t = 1$), X_1 (resp. Y_1) will abbreviated by X (resp. Y). We first define the family of multicast networks with side information at the end nodes.

Definition 3.2.1 *A network source coding problem $\mathcal{N}$ is a multicast network with side information at the end nodes if and only if for every $v \in \mathcal{V}$*

1. $\mathcal{D}_v = \emptyset$ *or* $\mathcal{D}_v = \{X_1, \ldots, X_s\}$.
2. *If* $(v, Y_j) \in \mathcal{S}$ *for some* $j \in \{1, \ldots, t\}$, *then* $(v, v') \notin \mathcal{E}$ *for all* $v' \in \mathcal{V}$.

Any multicast network $\mathcal{N}$ is a member of this family since in multicast networks there is no side information ($t = 0$).

We next define the cut-set bound. A subset is any set $\mathcal{A} \subseteq \mathcal{V}$.

Definition 3.2.2 *Let $\mathcal{N}$ be a non-functional network source coding problem. The cut-set bound is a set of inequalities for rate vector* $\mathbf{R}$ *defined as*

$$R_{\mathcal{A}} \geq H(\mathbf{X}_{\mathcal{D}}(\mathcal{A})|\mathbf{X}_{\mathcal{S}}(\mathcal{A}^c), \mathbf{Y}_{\mathcal{S}}(\mathcal{A}^c)) \ \ \forall \ \mathcal{A} \subseteq \mathcal{V}, \tag{3.1}$$

where for any $\mathcal{A} \subseteq \mathcal{V}$,

$$\begin{aligned} R_{\mathcal{A}} &:= \sum_{(v,v') \in \mathcal{E}: v \in \mathcal{A}} R_{(v,v')} \\ \mathbf{X}_{\mathcal{S}}(\mathcal{A}^c) &:= \{X_i \mid (v, X_i) \in \mathcal{S} \text{ for some } v \in \mathcal{A}\} \\ \mathbf{Y}_{\mathcal{S}}(\mathcal{A}^c) &:= \{Y_j \mid (v, Y_j) \in \mathcal{S} \text{ for some } v \in \mathcal{A}\} \\ \mathbf{X}_{\mathcal{D}}(\mathcal{A}) &:= \{X_i \mid (v, X_i) \in \mathcal{D} \text{ for some } v \in \mathcal{A}^c\}. \end{aligned}$$

Finally we define the exponentially strong converse.

Definition 3.2.3 *We say that the exponentially strong converse holds for lossy network source coding problem* $\mathcal{N}$ *if and only if for any* $P_{\mathbf{X},\mathbf{Y}}$*, any rate vector* $\mathbf{R}$*, and any distortion vector* $\mathbf{D}$, $\mathbf{R} \notin \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ *if and only if for any sequence of rate-*$\mathbf{R}$*, length-*n *block codes* $\{\mathcal{C}_n\}$ *the probability of observing distortion less than or equal to* $\mathbf{D}$ *decreases exponentially to* 0*, i.e.,*

$$\liminf_{n\to\infty} -\frac{1}{n} \log P_{\mathbf{X}^n,\mathbf{Y}^n} \left(\frac{1}{n} d(X^n \widehat{X}^n(v')) \leq D \ \forall\ (v', X_i) \in \mathcal{D} \right) > 0,$$

where for all $(v', X_i) \in \mathcal{D}$, $\widehat{X}_i^n(v')$ *is the reproduction of* X_i^n *at node* v' *using* $\mathcal{C}_n$*.*

Our approach relies on strong typicality. We briefly mention some properties that are useful here and fix the notation as follows. Let W be a finite-alphabet random variable. For any integer n and positive number $\epsilon > 0$, let $A_\epsilon^{*(n)}(W)$ denote the strongly typical set. (For example, see [44].) Lemma 3.2.4 states that the probability of the atypical set $(A_\epsilon^{*(n)}(W))^c$ decreases exponentially to 0 as n grows without bound with exponent greater than or equal to $\tau(\epsilon)$ that depends only on the distribution of W and ϵ. We state Lemma 3.2.4 without proof.

Lemma 3.2.4 *For any* $\epsilon > 0$*, when* n *is sufficiently large*

$$\Pr(A_\epsilon^{*(n)}(W)) \geq 1 - 2^{-n\tau(\epsilon)}$$

for some $\tau(\epsilon) > 0$*.*

Let Z be another finite-alphabet random variable. For $z^n \in \mathcal{Z}^n$, let $A_\epsilon^{(n)}(W|Z = z^n)$ denote the set of sequences $w^n \in \mathcal{W}^n$ that are strongly jointly typical with z^n.

3.3 An Application

Consider the following problem: for a given network source coding problem $\mathcal{N} = ((\mathcal{V}, \mathcal{E}), \mathcal{S}, \mathcal{D})$, let $\overline{\mathcal{N}} = ((\mathcal{V}, \overline{\mathcal{E}}), \mathcal{S}, \mathcal{D})$ be the network source coding problem that is identical to $\mathcal{N}$ except that $\overline{\mathcal{E}} = \mathcal{E} \backslash \{\overline{e}\}$ for some $\overline{e} \in \mathcal{E}$. Given any $\mathbf{R} = (R_e \mid e \in \mathcal{E})$ in the achievable rate region for $\mathcal{N}$, we wish to know whether $R_{\overline{e}} = 0$ implies that the rate vector $\overline{\mathbf{R}} = (R_e)_{e \in \overline{\mathcal{E}}}$ is in the achievable rate region for $\overline{\mathcal{N}}$.

Note that by the definition of rate $R_{\overline{e}} = 0$ implies only that the number of bits transmitted across $\overline{e}$ grows sublinearly in the blocklength n; it does not imply that zero bits are sent across $\overline{e}$. So the key concern is whether a gap can arise in the limit as $R_{\overline{e}}$ approaches 0.

One way to study this question is as follows. Let $\{\mathcal{C}_n\}$ be a sequence of length-n rate-$(\mathbf{R} + c(n) \cdot \mathbf{1})$ block codes for $\mathcal{N}$ for which the probability of satisfying the distortion constraint goes to 1 as n grows without bound. Here $\mathbf{1}$ is the vector with all components equal to 1 and $\{c(n)\}$ is a sequence of non-negative numbers such that $\lim_{n\to\infty} c(n) = 0$. Choose n such that $\mathcal{C}_n$ satisfies the distortion constraint with probability $P_{\mathbf{D}}^{(n)} \geq 1/2$. The rate of code $\mathcal{C}_n$ on edge $\overline{e}$ is $c(n)$. Let $\overline{\mathcal{C}}_n$ be the code that is identical to $\mathcal{C}_n$ except that no message is transmitted across edge $\overline{e}$ and all functions that rely on that message treat the value as a constant. Then the probability $\overline{P}_{\mathbf{D}}$ that $\overline{\mathcal{C}}_n$ satisfies the distortion constraint is bounded as $\overline{P}_{\mathbf{D}} \geq 2^{-nc(n)-1} = 2^{-n(c(n)+1/n)}$ using the best guess for the missing message across the edge $\overline{e}$. If the exponentially strong converse holds for $\overline{\mathcal{N}}$, then this implies that the rate vector $\overline{\mathbf{R}}$ is achievable for $\overline{\mathcal{N}}$.

3.4 The Lossless Coded Side Information Problem

We prove the exponentially strong converse for the coded side information problem [4]. (See Figure 3.1.) Our proof follows the approach in [20], where the strong converse theorem for this particular network source coding problem is proven with a slower rate of convergence. We start with some simplifications of definitions from [20] that

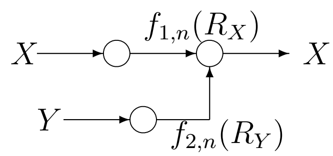


Figure 3.1: The coded side information problem.

are useful in this section.

Definition 3.4.1 *For any positive integer n, positive number $\delta > 0$, and set $\mathcal{B} \subseteq \mathcal{X}^n$, define*

$$\psi_\delta(\mathcal{B}) := \{y^n \mid \Pr(X^n \in \mathcal{B} \mid Y^n = y^n) \geq 2^{-n\delta}\}.$$

Definition 3.4.2 *For any $c > 0$, $\epsilon > 0$, and $\delta > 0$, define*

1.
$$\widehat{S}_n(c, \epsilon, \delta) := \frac{1}{n} \log \min |\mathcal{B}|,$$
where the min *is taken over all subsets $\mathcal{B} \subseteq \mathcal{X}^n$ such that*
$$-\frac{1}{n} \log \Pr\left(Y^n \in \psi_\delta(\mathcal{B}) \cap A_\epsilon^{*(n)}(Y)\right) \leq c.$$

2.
$$\widehat{T}(c) := \min H(X|U),$$
where the min *is taken over all random variables U such that $X \to Y \to U$ forms a Markov chain and*
$$I(Y; U) \leq c.$$

Theorem 3.4.3 *[20, Theorem 1]For all $\epsilon > 0$, $\delta > 0$, and $c > 0$*

$$\lim_{n\to\infty} \widehat{S}_n(c,\epsilon,\delta) = \widehat{T}(c).$$

Theorem 3.4.4 *The exponentially strong converse holds for the coded side information problem.*

Proof. Let (R_X, R_Y) be a rate pair such that there exists a sequence of length-n rate-(R_X, R_Y) block codes $\{\mathcal{C}_n\}_{i=1}^{\infty}$ such that the correct probability $\Pr(X^n = \widehat{X}^n) = 2^{-nc(n)}$ for some sequence $\{c(n)\}$ satisfying $\lim_{n\to\infty} c(n) = 0$. Here $f_{1,n}$, $f_{2,n}$, and g_n are the encoding and decoding functions of $\mathcal{C}_n$ (as shown in Figure 3.1) and

$$\widehat{X}^n = g_n(f_{1,n}(X^n), f_{2,n}(Y^n))$$

is the reproduction of X^n using code $\mathcal{C}_n$. We want to show that (R_X, R_Y) is achievable, i.e., that there exists a random variable U such that $X \to Y \to U$ forms a Markov chain and that

$$R_X \geq H(X|U),\ R_Y \geq I(Y;U).$$

For any particular value $u \in \{1, 2, \ldots, 2^{nR_Y}\}$ of the encoding function $f_{2,n}$, define

$$C^{(n)}(u) := \{x^n \mid x^n = g_n(f_{1,n}(x^n), u)\}.$$

By assumption,

$$2^{-nc(n)} \leq \sum_{y^n\in\mathcal{Y}^n} \Pr(Y^n = y^n) \Pr(C^{(n)}(f_{2,n}(y^n)) \mid Y^n = y^n).$$

Fix $\epsilon > 0$. By Lemma 3.2.4,

$$\Pr(A_\epsilon^{*(n)}(X,Y)) \geq 1 - 2^{-n\tau(\epsilon)}$$

for some $\tau(\epsilon) > 0$. Therefore when n is sufficiently large,

$$\begin{aligned}\sum_{y^n \in A_\epsilon^{*(n)}(Y)} & \Pr(Y^n = y^n)\Pr(C^{(n)}(f_{2,n}(y^n)) \mid Y^n = y^n) \\ & \geq 2^{-nc(n)} - \Pr(Y^n \notin A_\epsilon^{*(n)}(Y)) \\ & \geq 2^{-nc(n)} - 2^{-n\tau(\epsilon)} \geq 2^{-nb(n,\epsilon)} \end{aligned} \tag{3.2}$$

for some sequence of positive numbers $\{b(n,\epsilon)\}$ such that $\lim_{n\to\infty} b(n,\epsilon) = 0$ for all $\epsilon > 0$. Let $S(n,\epsilon) \subseteq A_\epsilon^{*(n)}(Y)$ be the set of all $y^n \in A_\epsilon^{*(n)}(Y)$ such that

$$\Pr(C^{(n)}(f_{2,n}(y^n)) \mid Y^n = y^n) \geq 2^{-n2b(n,\epsilon)}.$$

Then (3.2) implies

$$\Pr(S(n,\epsilon)) + (1 - \Pr(S(n,\epsilon)))2^{-n2b(n,\epsilon)} \geq 2^{-nb(n,\epsilon)}$$

which leads to

$$\Pr(S(n,\epsilon)) \geq \frac{2^{-nb(n,\epsilon)}}{1 + 2^{-n2b(n,\epsilon)}} > 2^{-n\left(b(n,\epsilon)+\frac{1}{n}\right)}$$

when n is sufficiently large. Now by definition

$$S(n,\epsilon) \subseteq \bigcup_u \left(\psi_{2b(n,\epsilon)}(C^{(n)}(u)) \cap A_\epsilon^{*(n)}(Y)\right).$$

Hence

$$\sum_u \Pr\left(\psi_{2b(n,\epsilon)}(C^{(n)}(u)) \cap A_\epsilon^{*(n)}(Y)\right) \geq 2^{-n\left(b(n,\epsilon)+\frac{1}{n}\right)},$$

where the summation is taken over all $u \in \{1,\ldots,2^{nR_Y}\}$. Thus there exists an index u^* such that

$$\Pr\left(\psi_{2b(n,\epsilon)}(C^{(n)}(u^*)) \cap A_\epsilon^{*(n)}(Y)\right) \geq 2^{-n\left(b(n,\epsilon)+\frac{1}{n}+R_Y\right)}.$$

By the definition of $\widehat{S}_n$,

$$\frac{1}{n}\log|C^{(n)}(u^*)| \geq \widehat{S}_n\left(b(n,\epsilon)+\frac{1}{n}+R_Y,\epsilon,2b(n,\epsilon)\right).$$

Choose n sufficiently large so that $\frac{1}{n}+b(n,\epsilon)<\epsilon$. By Theorem 3.4.3,

$$\frac{1}{n}\log|C^{(n)}(u^*)| \geq \widehat{T}\,(\epsilon+R_Y)+\epsilon$$

when n is sufficiently large. Since $C^{(n)}(u^*)$ is the set of x^n that can be correctly decoded when $f_{2,n}(Y^n)=u^*$, $|C^{(n)}(u^*)| \leq 2^{nR_X}$ and hence

$$R_X \geq \widehat{T}\,(\epsilon+R_Y)+\epsilon$$

when n is sufficiently large. By the definition of $\widehat{T}$, there exists an auxiliary random variable U such that $X \to Y \to U$ and

$$R_X \geq H(X|U)+\epsilon,\ \ R_Y \geq I(Y;U)+\epsilon.$$

Letting $\epsilon \to 0$ completes the proof. □

3.5 Multicast networks with Side Information at the End nodes

In this section, we consider the family of multicast networks with side information at the end nodes. The cut-set bounds for this network are tight by [12]. This result can be treated as a generalization of multicast capacity [11]. The simplest interesting example in this family is the problem of lossless source coding with side information at the decoder. (See Figure 3.2.) The infinium of the set of losslessly achievable rates is $H(X|Y)$, which corresponds to one of the two corner points in the rate region of the Slepian-Wolf problem. We prove the exponentially strong converse for this basic example and then use it to conclude that the exponentially strong converse holds for those network source coding problems where the cut-set bound is tight. This implies that the exponentially strong converse is true for the family of multicast networks with side information at the end nodes. Lemma 3.5.1 treats the simplest example, where a single encoder describes X to a decoder that knows Y.

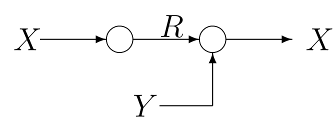


Figure 3.2: The lossless source coding problem with side information at the decoder.

Lemma 3.5.1 *The exponentially strong converse holds for the lossless source coding problem with side information at the decoder.*

Proof. Let $R > 0$. Suppose that there exists a sequence of length-n rate-R block codes $\mathcal{C}_n$ with correct probability

$$\Pr(X^n = \widehat{X}^n(X^n, Y^n)) = 2^{-nc(n)}$$

for some sequence $\{c(n)\}$ such that $\lim_{n\to\infty} c(n) = 0$, where $\widehat{X}^n(X^n, Y^n)$ is the reproduction of X^n using code $\mathcal{C}_n$. We want to show that R is in the lossless rate region by showing that $R \geq H(X|Y)$.

For any positive integer n and positive real number $\epsilon > 0$, let

$$B_\epsilon^{(n)} := A_\epsilon^{*(n)}(X, Y) \cap \{(x^n, y^n)\ : x^n = \widehat{X}^n(x^n, y^n)\}$$

be the set of strongly typical pairs (x^n, y^n) such that x^n is correctly decoded. Lemma 3.2.4 implies that

$$\Pr(A_\epsilon^{*(n)}(X, Y)) \geq 1 - 2^{-n\tau(\epsilon)}$$

for some $\tau(\epsilon) > 0$, so

$$\Pr(B_\epsilon^{(n)}) \geq 2^{-nc(n)} - 2^{-n\tau(\epsilon)} = 2^{-nb(n,\epsilon)} \tag{3.3}$$

when n is sufficiently large for some sequence of positive numbers $\{b(n,\epsilon)\}$ such that $\lim_{n\to\infty} b(n,\epsilon) = 0$ for all $\epsilon > 0$. Now

$$\begin{aligned}\Pr(B_\epsilon^{(n)}) &= \sum_{y^n \in A_\epsilon^{*(n)}(Y)} \Pr(Y^n = y^n) \Pr(B_\epsilon^{(n)} | Y^n = y^n) \\ &\geq 2^{-nb(n,\epsilon)}\end{aligned}$$

Hence there exists a $y_0^n \in A_\epsilon^{*(n)}(Y)$ such that

$$\Pr(B_\epsilon^{(n)} | Y^n = y_0^n) \geq 2^{-nb(n,\epsilon)}. \tag{3.4}$$

Since for all $x^n \in A_\epsilon^{*(n)}(X|Y^n = y_0^n)$

$$\Pr(X^n = x^n | Y^n = y_0^n) \leq 2^{-n(H(X|Y)-\epsilon)},$$

(3.4) implies that

$$|B_\epsilon^{(n)} \cap \{Y^n = y_0^n\}| \geq 2^{n(H(X|Y)-\epsilon-b(n,\epsilon))}. \tag{3.5}$$

Since $\mathcal{C}_n$ has rate R and $B_\epsilon^n \cap \{Y^n = y_0^n\}$ is by definition the set of pairs $(x^n, y_0^n) \in A_\epsilon^{*(n)}(X,Y)$ such that x^n can be correctly decoded when $Y^n = y_0^n$,

$$2^{nR} \geq |B_\epsilon^{(n)} \cap \{Y^n = y_0^n\}|.$$

Thus (3.5) implies that

$$R \geq H(X|Y) - \epsilon - b(n,\epsilon)$$

for all n and ϵ. Since $\epsilon > 0$ is arbitrary, letting $n \to \infty$ gives

$$R \geq H(X|Y).$$

□

Lemma 3.5.2 relates our condition on the probability of correct decoding to the tightness of the cut-set bound.

Lemma 3.5.2 *Let $\mathcal{N}$ be a network source coding problem and let $\mathbf{R}$ be a rate vector. For all $(v', X_i) \in \mathcal{D}$, let $\widehat{X}_i^n(v')$ be the reproduction of X_i^n at node $v' \in \mathcal{V}$. If there exists a sequence of length-n rate-$\mathbf{R}$ block codes such that the correct probability is*

$$\Pr\left(X_i^n = \widehat{X}_i^n(v') \; \forall \; (v', X_i) \in \mathcal{D}\right) = 2^{-nc(n)}$$

for some sequence $\{c(n)\}$ such that $\lim_{n\to\infty} c(n) = 0$, then $\mathbf{R}$ satisfies the cut-set bound of $\mathcal{N}$.

Proof. For any cut $\mathcal{A} \in \mathbb{A}$, since $\mathcal{A}^c$ demands the sources $\mathbf{X}_{\mathcal{A}}$ which are available in $\mathcal{A}$ and $(\mathbf{X}(\mathcal{A}^c), \mathbf{Y}(\mathcal{A}^c))$ is available in $\mathcal{A}^c$, each cut $\mathcal{A}$ corresponds a lossless source coding problem with side information on the decoder side as in Figure 3.2. Hence by Lemma 3.5.1, the overall rate $\mathbf{R}_{\mathcal{A}}$ from $\mathcal{A}$ to $\mathcal{A}^c$ must satisfy (3.1) for all $\mathcal{A} \in \mathbb{A}$. This completes the proof. □

Theorem 3.5.3 concludes this section.

Theorem 3.5.3 *The exponentially strong converse holds for the multicast network with side information at the end nodes.*

Proof. The result a direct consequence of Lemma 3.5.2 and the tightness of the cut-set bound [12]. □

3.6 Lossy Source Coding for the Gray-Wyner Network

Given $\mathbf{D} = (D_1, D_2) > \mathbf{0}$. Recall that the lossy rate region for the Gray-Wyner problem shown in Figure 3.3 is the closure of the set of all (R_0, R_1, R_2) for which

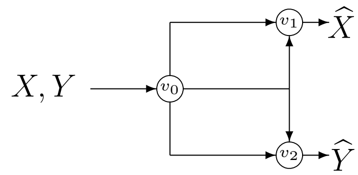


Figure 3.3: The Gray-Wyner network.

there exists a random variable U with alphabet size $|\mathcal{U}| \leq |\mathcal{X}_1||\mathcal{X}_2| + 2$, such that

$$R_0 \geq I(\mathbf{X}; U),\ R_1 \geq R_{X_1|U}(D_1),\ R_2 \geq R_{X_2|U}(D_2), \tag{3.6}$$

where $\mathbf{X} = (X_1, X_2)$ denotes the source vector and $R_{X_i|U}(D_i)$ is the conditional rate-distortion function for distortion D_i for $i \in \{1, 2\}$.

Theorem 3.6.2 shows that the exponentially strong converse holds for the lossy Gray-Wyner Problem. The idea of the proof is that since the exponent of the probability of meeting the distortion constraints is asymptotically zero, the given rate vector is $\mathbf{D}$-achievable for another distribution $\mathbf{X}$ that is close to $P_{\mathbf{X}}$. The approach follows the method of proving the converse of the region (3.6) in [3] that turns the dimension-n description of the rate vectors into a single-letter form. Hence a similar approach can be applied to prove Theorem 3.4.4. Lemma 3.6.1 is useful for proving Theorem 3.6.2.

Lemma 3.6.1 *Let W be a random variable with alphabet $\mathcal{W}$ and distribution P_W. Let $\{B^{(n)}\}$ be a sequence of sets $B^{(n)} \subseteq \mathcal{W}^n$ such that $P_{W^n}(B^{(n)}) = 2^{-nb(n)}$ for some sequence of non-negative numbers $\{b(n)\}_{n=1}^{\infty}$ satisfying $\lim_{n\to\infty} b(n) = 0$. Then there exist a sequence $\{a(n)\}_{n=1}^{\infty}$ of non-negative numbers and a sequence $\{Q_{W^n}^{(n)}\}$ of distributions on $\mathcal{W}^n$ such that*

$$\begin{aligned}
&\lim_{n\to\infty} a(n) = 0\\
&\lim_{n\to\infty} Q_{W^n}^{(n)}(B^{(n)}) = 1\\
&2^{-na(n)} P_{W^n}(w^n) \leq Q_{W^n}(w^n) \leq 2^{na(n)} P_{W^n}^{(n)}(w^n)\ \forall w^n \in \mathcal{W}^n.
\end{aligned}$$

Proof. Define

$$\begin{aligned} Q^{(n)}_{W^n}(w^n) &= \frac{2^{n(b(n)+\frac{1}{\sqrt{n}})}P_{W^n}(w^n)}{2^{n(b(n)+\frac{1}{\sqrt{n}})}P_{W^n}(B^{(n)})+(1-P_{W^n}(B^{(n)}))}, \text{ if } w^n \in B^{(n)}, \\ Q^{(n)}_{W^n}(w^n) &= \frac{P_{W^n}(w^n)}{2^{n(b(n)+\frac{1}{\sqrt{n}})}P_{W^n}(B^{(n)})+(1-P_{W^n}(B^{(n)}))}, \text{ if } w^n \notin B^{(n)}. \end{aligned}$$

Then

$$\begin{aligned} & Q^{(n)}_{W^n}(B^{(n)}) \geq \frac{2^{\sqrt{n}}}{2^{\sqrt{n}}+1} \to 1 \text{ as } n \to \infty \\ & 2^{-n(\frac{1}{\sqrt{n}}+\frac{1}{n})}P_{W^n}(w^n) \leq Q^{(n)}_{W^n}(w^n) \leq 2^{n(b(n)+\frac{1}{\sqrt{n}})}P_{W^n}(w^n) \ \forall\ w^n \in \mathcal{W}^n. \end{aligned}$$

□

Theorem 3.6.2 *The exponentially strong converse holds for the lossy Gray-Wyner problem.*

Proof. Let $\mathbf{R} = (R_0, R_1, R_2)$ be a rate vector and $P_{\mathbf{X}}$ denote the source distribution. Suppose that there exists a sequence of length-n, rate-$\mathbf{R}$ block codes such that

$$\lim_{n\to\infty} -\frac{1}{n} \Pr\left(Ed(X_i^n, \widehat{X}_i^n) \leq D_i \ \forall\ i \in \{1,2\}\right) = 0,$$

where $\widehat{X}_1^n$ and $\widehat{X}_2^n$ are reproductions of X_1^n and X_2^n at nodes v_1 and v_2, respectively. We want to show that $\mathbf{R}$ is in the region described in (3.6).
For $\epsilon > 0$, let

$$B_\epsilon^{(n)} := A_\epsilon^{*(n)}(\mathbf{X}) \cap \{\mathbf{x}^n \in \mathcal{X}_1^n \times \mathcal{X}_2^n \mid \mathbf{x}^n = (\widehat{X}_2^n(\mathbf{x}^n), \widehat{X}_2^n(\mathbf{x}^n))\}.$$

Then the same argument used to prove (3.3) leads to

$$\Pr(B_\epsilon^{(n)}) = 2^{-nb(n,\epsilon)} \tag{3.7}$$

for some sequence of non-negative numbers $\{b(n,\epsilon)\}$ such that $\lim_{n\to\infty} b(n,\epsilon) = 0$

for all $\epsilon > 0$.

By Lemma 3.6.1, there exists a sequence of non-negative numbers $\{a(n,\epsilon)\}_{n=1}^{\infty}$ and a sequence of distributions $\{Q_{\mathbf{X}^n}^{(n,\epsilon)}\}$ such that for all $\epsilon > 0$,

$$\lim_{n\to\infty} a(n,\epsilon) = 0, \ \lim_{n\to\infty} Q_{\mathbf{X}^n}^{(n,\epsilon)}(B_\epsilon^{(n)}) = 1,$$

and

$$2^{-na(n,\epsilon)} P_{\mathbf{X}^n}(\mathbf{x}^n) \leq Q_{\mathbf{X}^n}^{(n,\epsilon)}(\mathbf{x}^n) \leq 2^{na(n,\epsilon)} P_{\mathbf{X}^n}(\mathbf{x}^n) \ \forall \ \mathbf{x}^n \in \mathcal{X}_1^n \times \mathcal{X}_2^n. \tag{3.8}$$

Let $n(\epsilon)$ be a positive integer such that $a(n(\epsilon),\epsilon) < \epsilon$ and $Q_{\mathbf{X}^{n((\epsilon))}}^{(n(\epsilon),\epsilon)}(B_\epsilon^{(n(\epsilon))}) > 1-\epsilon$. Let $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}$ denote the distribution $Q_{\mathbf{X}^{n(\epsilon)}}^{(n(\epsilon),\epsilon)}$. Hence by the continuity of the lossy rate region with respect to the distortion vector, there exists a function $\tau_1(\epsilon)$ with $\lim_{\epsilon\to\infty} \tau_1(\epsilon) = 0$ such that the rate vector $n\left(\mathbf{R} + \tau_1(\epsilon)\cdot\mathbf{1}\right)$ is in the $\mathbf{D}$-achievable region for the Gray-Wyner problem with respect to distribution $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}$. Hence there exist random variables U, $\widehat{X}_1^{n(\epsilon)}$, and $\widehat{X}_2^{n(\epsilon)}$ such that

$$\begin{aligned}\mathbf{R} + \tau_1(\epsilon)\cdot\mathbf{1} \ \ \geq \ \ & \frac{1}{n(\epsilon)}\Big(I_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)};U), R_{X_1^{n(\epsilon)}|U}(n(\epsilon)D_1, Q^{(\epsilon)}), \\ & R_{X_2^{n(\epsilon)}|U}(n(\epsilon)D_2, Q^{(\epsilon)})\Big),\end{aligned}$$

where $I_{Q^{(\epsilon)}}$ and $R_{X_i^{n(\epsilon)}|U}(D_i, Q^{(\epsilon)})$ (for $i \in \{1,2\}$) are the mutual information and conditional rate-distortion functions evaluated according to distribution $Q_{\mathbf{X}^{n(\epsilon)}}^{(\epsilon)}$. Let $J(\epsilon)$ be an independent random variable uniformly distributed over $\{1,\ldots,n(\epsilon)\}$.

Define $U_i = (U, \mathbf{X}_1^{i-1})$ for all $i \in \{1, \dots, n(\epsilon)\}$. Then

$$\begin{aligned}
\frac{1}{n(\epsilon)} I_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)}; U) &= \frac{1}{n(\epsilon)} \sum_{i=1}^{n(\epsilon)} I_{Q^{(\epsilon)}}(\mathbf{X}_i; U | \mathbf{X}_1^{i-1}) \\
&= \frac{1}{n(\epsilon)} \sum_{i=1}^{n(\epsilon)} \left(I_{Q^{(\epsilon)}}(\mathbf{X}_i; U, \mathbf{X}_1^{i-1}) - I_{Q^{(\epsilon)}}(\mathbf{X}_i; \mathbf{X}_1^{i-1}) \right) \\
&= \frac{1}{n(\epsilon)} \sum_{i=1}^{n(\epsilon)} I_{Q^{(\epsilon)}}(\mathbf{X}_i; U_i) - \frac{1}{n(\epsilon)} \sum_{i=1}^{n(\epsilon)} H_{Q^{(\epsilon)}}(\mathbf{X}_i) + \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} \\
&= I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; U_{J(\epsilon)} | J(\epsilon)) - H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)} | J(\epsilon)) + \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} \\
&= -H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)} | U_{J(\epsilon)}, J(\epsilon)) + \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} \\
&= I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; U_{J(\epsilon)}, J(\epsilon)) + \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} - H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}).
\end{aligned}$$

Similarly, let $V_{1,1} = (U, X_{1,1}^{i-1})$ and $V_{2,1} = (U, X_{2,1}^{i-1})$ for $i \in \{1, \dots, n(\epsilon)\}$. Then

$$\begin{aligned}
\frac{1}{n(\epsilon)} R_{X_1^{n(\epsilon)}|U}(n(\epsilon) D_1, Q^{(\epsilon)}) &\geq R_{X_{1,J(\epsilon)}|U_{J(\epsilon)}, J(\epsilon)}(D_1, Q^{(\epsilon)}) \\
\frac{1}{n(\epsilon)} R_{X_2^{n(\epsilon)}|U}(n(\epsilon) D_2, Q^{(\epsilon)}) &\geq R_{X_{2,J(\epsilon)}|U_{J(\epsilon)}, J(\epsilon)}(D_2, Q^{(\epsilon)}).
\end{aligned}$$

Since $\mathbf{X}_{J(\epsilon)}$ has finite alphabet $\mathcal{X}_1 \times \mathcal{X}_2$, there exists a conditional distribution $Q_{V_{n(\epsilon)}|\mathbf{X}_{J(\epsilon)}}$ for random variable $V_{n(\epsilon)}$ with alphabet size $|\mathcal{X}_1||\mathcal{X}_2| + 2$ such that

$$\begin{aligned}
&I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; U_{J(\epsilon)}, J(\epsilon)) = I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; V_{n(\epsilon)}) \\
&R_{X_{1,J(\epsilon)}|U_{J(\epsilon)}, J(\epsilon)}(D_1, Q^{(\epsilon)}) \geq R_{X_{1,J(\epsilon)}|V_{n(\epsilon)}}(D_1, Q^{(\epsilon)}) \\
&R_{X_{2,J(\epsilon)}|U_{J(\epsilon)}, J(\epsilon)}(D_2, Q^{(\epsilon)}) \geq R_{X_{2,J(\epsilon)}|V_{n(\epsilon)}}(D_2, Q^{(\epsilon)}).
\end{aligned}$$

We next show that

$$\lim_{\epsilon \to 0} Q^{(\epsilon)}_{\mathbf{X}_{J(\epsilon)}}(\mathbf{x}) = P_{\mathbf{X}}(\mathbf{x}) \ \forall \ \mathbf{x} \in \mathcal{X}_1 \times \mathcal{X}_2 \tag{3.9}$$

$$\lim_{\epsilon \to 0} \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} - H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}) = 0. \tag{3.10}$$

First, for all $\mathbf{x}^{n(\epsilon)} \in \mathcal{X}_1^{n(\epsilon)} \times \mathcal{X}_2^{n(\epsilon)}$,

$$Q^{(\epsilon)}_{\mathbf{X}^{n(\epsilon)}}(\mathbf{X}_{J(\epsilon)} = \alpha | \mathbf{X}^{n(\epsilon)} = \mathbf{x}^{n(\epsilon)}) = \frac{|\{i \mid \mathbf{x}_i = \alpha\}|}{n(\epsilon)} \; \forall \alpha \in \mathcal{X}_1 \times \mathcal{X}_2.$$

Hence for all $\mathbf{x}^{n(\epsilon)} \in B_\epsilon^{(n(\epsilon))} \subseteq A_\epsilon^{*(n(\epsilon))}(\mathbf{X})$,

$$|Q^{(\epsilon)}_{\mathbf{X}^{n(\epsilon)}}(\mathbf{X}_{J(\epsilon)} = \alpha | \mathbf{X}^{n(\epsilon)} = \mathbf{x}^{n(\epsilon)}) - P_{\mathbf{X}}(\alpha)| < \frac{\epsilon}{|\mathcal{X}_1||\mathcal{X}_2|} \; \forall \alpha \in \mathcal{X}_1 \times \mathcal{X}_2.$$

The fact that $Q^{(\epsilon)}_{\mathbf{X}^{n(\epsilon)}}(B_\epsilon^{(n(\epsilon))}) > 1 - \epsilon$ leads to

$$|Q^{(\epsilon)}_{\mathbf{X}^{n(\epsilon)}}(\mathbf{X}_{J(\epsilon)} = \alpha) - P_{\mathbf{X}}(\alpha)| < \frac{\epsilon}{|\mathcal{X}_1||\mathcal{X}_2|} + \epsilon \; \forall \; \alpha \in \mathcal{X}_1 \times \mathcal{X}_2,$$

which proves (3.9). By the uniform continuity of mutual information and entropy functions on finite-alphabet random variables, (3.9) implies that

$$\begin{aligned}
&|H_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}) - H_P(\mathbf{X})| < \tau_2(\epsilon) \\
&|I_{Q^{(\epsilon)}}(\mathbf{X}_{J(\epsilon)}; V_{n(\epsilon)}) - I_P(\mathbf{X}; V_{n(\epsilon)})| < \tau_2(\epsilon) \\
&|R_{X_{1,J(\epsilon)}|V_{n(\epsilon)}}(D_1, Q^{(\epsilon)}) - R_{X_1|V_{n(\epsilon)}}(D_1, P)| < \tau_2(\epsilon) \\
&|R_{X_{2,J(\epsilon)}|V_{n(\epsilon)}}(D_2, Q^{(\epsilon)}) - R_{X_2|V_{n(\epsilon)}}(D_2, P)| < \tau_2(\epsilon)
\end{aligned}$$

for some $\tau_2(\epsilon)$ such that $\lim_{\epsilon \to \infty} \tau_2(\epsilon) = 0$, and I_P and H_P are mutual information and entropy functions evaluated according to the distribution $P_{\mathbf{X}, V_{n(\epsilon)}} = P_{\mathbf{X}} Q_{V_{n(\epsilon)}|\mathbf{X}}$. Hence for proving (3.10), it remains to show that

$$\lim_{\epsilon \to \infty} \frac{H_{Q^{(\epsilon)}}(\mathbf{X}^{n(\epsilon)})}{n(\epsilon)} = H_P(\mathbf{X}).$$

By (3.8), for all $\mathbf{x}^{n(\epsilon)} \in B_\epsilon^{(n(\epsilon))} \subseteq A_\epsilon^{*(n(\epsilon))}(\mathbf{X})$,

$$| - \frac{1}{n(\epsilon)} \log Q_{\mathbf{X}^{(\epsilon)}}(\mathbf{x}^{n(\epsilon)}) + \frac{1}{n(\epsilon)} \log P_{\mathbf{X}^{n(\epsilon)}}(\mathbf{x}^{n(\epsilon)})| \leq a(n, \epsilon) < \epsilon$$

and

$$|\frac{1}{n(\epsilon)}\log P_{\mathbf{X}^{n(\epsilon)}}(\mathbf{x}^{n(\epsilon)}) - H_P(\mathbf{X})| < \tau_3(\epsilon)$$

for some $\tau_3(\epsilon)$ such that $\lim_{n\to\infty}\tau_3(\epsilon) = 0$. Let

$$\tau_4(\epsilon) := \epsilon + \tau_3(\epsilon) + \epsilon \log|\mathcal{X}_1||\mathcal{X}_2|.$$

Since $Q^{(\epsilon)}_{\mathbf{X}^{n(\epsilon)}}(B^{(n(\epsilon))}_{\epsilon}) > 1 - \epsilon$,

$$|\frac{1}{n(\epsilon)} H_{Q^{(\epsilon)}(\mathbf{X}^{n(\epsilon)})} - H_P(\mathbf{X})| < \tau_4(\epsilon),$$

which proves (3.10). Hence the rate vector

$$\mathbf{R} + (\tau_1(\epsilon) + \tau_2(\epsilon) + \tau_4(\epsilon)) \cdot 1$$

is in the achievable rate region of the Gray-Wyner problem w.r.t. the distribution $P_{\mathbf{X}}$, which proves the desire result by letting $\epsilon \to 0$. $\square$

Chapter 4

Algorithms for Approximating Achievable Rate Regions

4.1 Introduction

The derivation of rate regions for lossless and lossy source coding problems is a central goal of network source coding theory research. While a network source coding problem is often considered to be solved once an achievable rate region and matching converse are demonstrated, these results become useful in practice only when we can evaluate them for example sources. For some problems, like Slepian and Wolf's lossless multiple access source coding problem [2], evaluating the optimal rate region for example sources is trivial since the information theoretic bound gives an explicit rate region characterization. For other problems, including lossy source coding, lossless source coding with coded side information at the receiver [4],[1] and the family of lossy source coding problems described by Jana and Blahut in [45], the information theoretic characterization describes an optimization problem whose solution is the desired bound. These optimization problems are often difficult to solve for example sources.

While single-letter characterizations and alphabet-size bounds for auxiliary random variables are often motivated by concerns about rate region evaluation, the evaluation problem itself has received surprisingly little attention in the literature. Most existing algorithms follow the strategy proposed by Blahut [23] and Arimoto [22].

[1] Source coding with coded side information at the receiver may be viewed as a type of lossy coding problem since perfect reconstruction of the side information is not required.

When applied to rate-distortion bound evaluation, this iterative descent approach progressively updates solutions for the marginal $p(z)$ on the reproduction alphabet and the conditional $p(z|x)$ on the reproduction given the source. The convexity of the objective function results in the algorithm's guaranteed convergence to the optimal solution [46]. Calculating the algorithmic complexity of this approach would require a bound on the number of iterations required to achieve convergence to the optimal solution (or a sufficiently accurate approximation).

We offer an alternative approach for rate region calculation. The proposed algorithm involves building a linear program whose solution approximates the optimal rate region to within a guaranteed factor of $(1 + \epsilon)$ times the optimal solution. The goal of achieving $(1 + \epsilon)$-accuracy using a polynomial-time algorithm is related to Csiszár and Körner's definition of computability, which they propose as a critical component of any future definition for a single-letter characterization [19, p.259–260]. Our algorithm gives a $(1+\epsilon)$-approximation of the rate region for lossless source coding with coded side information at the decoder [4]; this approach can be generalized to a lossy incast source coding problem described by Jana and Blahut in [45] and the achievable rate region for the lossy coded side information problem described by Berger et al. in [6]. Incast problems are multiple access source coding problems with one or more transmitters and a single receiver that wishes to reconstruct all sources in the network. (Reconstruction of possible receiver side information is trivial.) The lossy incast problem from [45], differs from traditional incast problems in that the sources may be statistically dependent and exactly one source is reconstructed with loss (subject to a distortion constraint) while the other source reconstructions are lossless. The lossy source coding and Wyner-Ziv problems meet this model of lossy incast problems. The rate region for this lossy incast problem relies on a single auxiliary random variable [45]. The achievable rate region for the lossy coded side information problem relies on a pair of auxiliary random variables [6].

Section 4.2 describes the algorithmic strategy. Section 4.4 describes the approximation algorithm for our lossy incast problem. Since describing the problem in its most general form increases notational complexity without adding much insight, we give details only for the Wyner-Ziv problem. Section 4.5 tackles the lossy coded side information achievability bound using tools developed for the incast problem.

4.2 Outline of Strategy

In all of the problems studied here, we begin with known information theoretic descriptions that rely on one or more auxiliary random variables. Optimization of each auxiliary random variable requires optimization of that variable's conditional distribution given one or more source random variables. Direct optimization is difficult since the desired rates are not convex or concave in the conditional distributions.

The central observation for our algorithm is that for any fixed conditional distribution on the source given a single auxiliary random variable, all rates and distortions are linear in the auxiliary random variable's marginal distribution. As a result, for any given conditional distribution, we can efficiently optimize the marginal on the auxiliary random variable using a linear program. Since the true conditional distribution of the source given the auxiliary random variable is unknown, we quantize the space of conditional distributions and find the best marginal with respect to a conditional distribution that exhibits each of these quantized distributions as the conditional given some value $z \in \mathcal{Z}$. The solution is at least as good as the solution that would be obtained if we were to first quantize the optimal conditional distribution and then run the linear program for that quantized conditional. As a result, to prove that the algorithm yields a $(1+\epsilon)$ approximation, we need only show that quantizing the optimal conditional distribution on the source given the auxiliary random variable would yield performance within a factor $(1+\epsilon)$ of the optimum.

For any finite alphabet $\mathcal{A}$, we quantize distribution $\{q(a)\}_{a\in\mathcal{A}}$ to distribution $\{\widehat{q}(a)\}_{a\in\mathcal{A}}$ as follows. First, fix parameters $\delta, \eta > 0$ and $c := 1 + \eta/|\mathcal{A}|$. These parameters are related to the approximation constant ϵ in a manner described in later sections. Fix $a_0 \in \arg\max_{a\in\mathcal{A}} q(a)$. Then

$$\widehat{q}(a) := \begin{cases} 0 & \text{if } a \neq a_0 \text{ and } q(a) < \delta \\ c^{-n} & \text{if } a \neq a_0,\ q(a) \geq \delta, \text{ and} \\ & \quad c^{-n} \leq q(a) < c^{-n+1} \\ 1 - \sum_{a\neq a_0} \hat{q}(a) & \text{if } a = a_0. \end{cases} \tag{4.1}$$

Then distribution $\hat{q}(a)$ can take on $N(\delta, \eta, |\mathcal{A}|)$ values, where

$$N(\delta, \eta, |\mathcal{A}|) \leq |\mathcal{A}| \left(\frac{-\log \delta}{\log(1 + \frac{\eta}{|\mathcal{A}|})} + 1 \right)^{|\mathcal{A}|-1}.$$

This approach quantizes smaller probability values more finely than larger probability values but maps the smallest probability values to zero. The impact of quantizing the smallest values of $q(a)$ to zero is limited since $q \log(1/q)$ approaches 0 as q approaches 0. The variation in the quantization cell size for $q(a)$ is motivated by Lemma 4.2.1.

Lemma 4.2.1 *Given distributions* $\{q(a)\}_{a\in\mathcal{A}}$ *and* $\{\widehat{q}(a)\}_{a\in\mathcal{A}}$ *on finite alphabet* $\mathcal{A}$. *If* $|q(a) - \widehat{q}(a)| \leq \epsilon q(a)$ *for all* $a \in \mathcal{A}$, *then*

$$|H(q) - H(\hat{q})| \leq \epsilon H(q) + \epsilon \log \frac{e}{1-\epsilon}.$$

Proof. Given any $x \in \mathcal{X}$. By the mean-value theorem, there is some $r_x \in [(1-\epsilon)p(x), (1+\epsilon)p(x)]$ such that

$$\begin{aligned} |p(x) \ln \frac{1}{p(x)} - q(x) \ln \frac{1}{q(x)}| &= |p(x) - q(x)|| \ln \frac{1}{r_x} - 1| \\ \leq\ & \epsilon p(x)\ \max\{|\ln \frac{1}{e(1+\epsilon)p(x)}|, |\ln \frac{1}{e(1-\epsilon)p(x)}|\} \\ \leq\ & \epsilon p(x) \ln \frac{1}{p(x)} + \epsilon p(x) + \epsilon p(x) \ln \frac{1}{1-\epsilon} \\ =\ & \epsilon p(x) \ln \frac{1}{p(x)} + \epsilon p(x) \ln \frac{e}{1-\epsilon}. \end{aligned}$$

So the result follows. $\square$

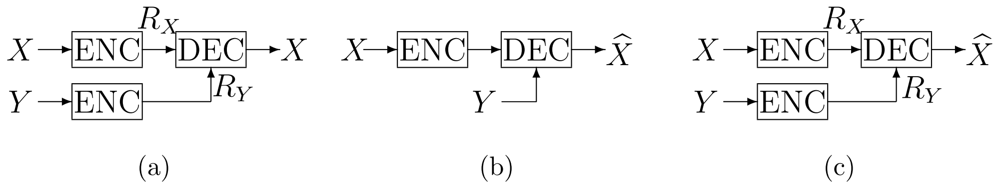


Figure 4.1: The (a) lossless coded side information, (b) Wyner-Ziv, and (c) lossy coded side information problems.

4.3 The Lossless Coded Side Information Region

Let $\mathcal{X}$ and $\mathcal{Y}$ denote the finite alphabets for the source X and side information Y, respectively. The lossless rate region for the coded side information problem [4] contains all the rate pairs (R_X, R_Y) satisfying

$$R_X \geq H(X|U),\ R_Y \geq I(Y;U)$$

for some U such that $X \to Y \to U$ forms a Markov chain. (See Figure 4.1(a).) The Lagrangian

$$J_1(\lambda) := H(X|U) + \lambda I(Y;U) \tag{4.2}$$

captures the desired constrained optimization.

Let $\mathcal{U} = \{1, \ldots, N(\delta, \eta, |\mathcal{X}|)\}$ be the alphabet for auxiliary random variable U, and for each $u \in \mathcal{U}$ let $\{Q_{Y|U}(y|u)\}_{y\in\mathcal{Y}}$ be a distinct distribution from our quantized collection (4.1). We wish to find the marginal $\{P_U(u)\}$ that minimizes $J_1(\lambda)$ for any $\lambda > 0$. Since

$$\begin{aligned} H(X|U) &= \sum_{u\in\mathcal{U}} P_U(u) H(X|U=u) \\ I(Y;U) &= H(Y) - H(Y|U) = H(Y) - \sum_{u\in\mathcal{U}} P_U(u) H(Y|U=u) \end{aligned}$$

and the constraints

$$\sum_{u\in\mathcal{U}} P_U(u) = 1,$$

$P_U(u) \geq 0$ for all $u \in \mathcal{U}$, and

$$\sum_{u \in \mathcal{U}} P_U(u) Q_{Y|U}(y|u) = p(y) \tag{4.3}$$

for all $y \in \mathcal{Y}$ are all linear functions of $P_U(u)$, we optimize $\{P_U(u)\}_{u \in \mathcal{U}}$ for the family of conditionals $\{Q_{Y|U}(y|u)\}_{(y,u) \in \mathcal{Y} \times \mathcal{U}}$ using linear programming. Notice that in the linear program, a constraint in (4.3) can be removed, and hence there are $|\mathcal{Y}| - 1$ constraints in (4.3). Since the lossless coded side information problem is a special case of the lossy coded side information problem that is discussed in Section 4.5, we state Theorem 4.3.1 without proving it.

Theorem 4.3.1 *The proposed algorithm yields a $(1+\epsilon)$-approximation algorithm for the lossless coded side information rate region in time $O(\epsilon^{-4(|\mathcal{Y}|+1)})$ as ϵ approaches 0.*

No matter what the initial size of $\mathcal{U}$, the solution to the linear program satisfies $P_U(u) = 0$ for all but $|\mathcal{Y}|$ values of $u \in \mathcal{U}$ by the following argument. The linear program has $|\mathcal{U}|$ variables and $|\mathcal{Y}| + |\mathcal{U}|$ constraints. Since there exists a solution for any linear program at a boundary point, there exists an optimal marginal $\{P_U^*(u)\}$ for which $|\mathcal{U}|$ constraints are satisfied with equality. At most $|\mathcal{Y}| - 1$ are constraints of the form $\sum_{u \in \mathcal{U}} P_U^*(u) Q_{Y|U}(y|u) = p(y)$ for some $y \in \mathcal{Y}$, and one constraint ensures that $\sum_{y \in \mathcal{Y}} P_U^*(u) = 1$. The remaining $|\mathcal{U}| - |\mathcal{Y}|$ constraints take the form $P_U^*(u) = 0$.

Figure 4.2 shows the lossless coded side information rate region (solid line) and our algorithm's $(1+\epsilon)$-approximation (circles) when $\mathcal{X} = \mathcal{Y} = \{0,1\}$ with joint distribution

$$P_{X,Y}(0,0) = 0.06,\ P_{X,Y}(0,1) = 0.24,\ P_{X,Y}(1,0) = 0.42,\ P_{X,Y}(1,1) = 0.28$$

and $\epsilon = 0.1$. The example demonstrates that the approximation is often tighter than the $(1+\epsilon)$ worst-case guarantee. While the lossless coded side information region is not difficult to calculate for these simple binary sources, the difficulty of the

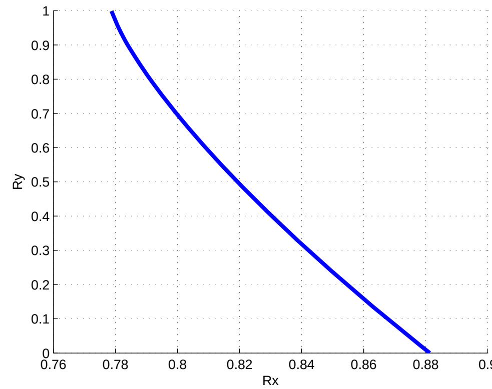


Figure 4.2: Example $(1+\epsilon)$ approximations for the lossless coded side information rate region.

calculation increases with the alphabet size.

4.4 The Wyner-Ziv Rate Region

Let $\mathcal{X}$ and $\mathcal{Y}$ denote the finite alphabets for sources X and Y. The Wyner-Ziv rate-distortion bound

$$\begin{aligned} R_{X|\{Y\}}(D) &= \min_{Z\in\Psi(X,Y)} I(X;Z|Y) \\ \Psi(X,Y) &:= \Big\{Z \Big| \, Z \to X \to Y, \\ &\qquad \exists\psi \text{ s.t. } Ed(X,\psi(Y,Z)) \le D\Big\}, \end{aligned}$$

specifies the minimal rate for describing source X to a receiver that knows side information Y and reconstructs X with expected distortion no greater than D [5]. (See Figure 4.1(b).) The Lagrangian

$$J_2(\lambda) := I(X;Z|Y) + \lambda \min_{\psi} Ed(X,\psi(Y,Z))) \tag{4.4}$$

captures the desired constrained optimization.

Let $\mathcal{Z} = \{1,\ldots,N(\delta,\eta,|\mathcal{X}|)\}$ be the alphabet for auxiliary random variable Z, and for each $z \in \mathcal{Z}$ let $\{Q_{X|Z}(x|z)\}_{x\in\mathcal{X}}$ be a distinct distribution from our quantized collection (4.1). We wish to find the marginal $\{P_Z(z)\}$ that minimizes $J_2(\lambda)$ for any $\lambda > 0$. Since $I(X;Z|Y) = H(X|Y) - H(X|Y,Z)$ and

$$\begin{aligned}
&\min_\psi Ed(X,\psi(Y,Z)) \\
&\quad = \min_\psi \sum_{z\in\mathcal{Z}} P_Z(z) E[d(X,\psi(Y,Z))|Z=z] \\
&\quad = \sum_{z\in\mathcal{Z}} P_Z(z) \min_\psi E[d(X,\psi(Y,z))|Z=z],
\end{aligned}$$

and the constraints

$$\sum_{z\in\mathcal{Z}} P_Z(z) = 1,$$

$P_Z(z) \geq 0$ for all $z \in \mathcal{Z}$, and

$$\sum_{z\in\mathcal{Z}} P_Z(z) Q_{X|Z}(x|z) = p(x)$$

for all $x \in \mathcal{X}$ are all linear functions of $P_Z(z)$, we optimize $\{P_Z(z)\}_{z\in\mathcal{Z}}$ for the family of conditionals $\{Q_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ using linear programming. The proof of Theorem 4.4.1 appears in the Appendix.

Theorem 4.4.1 *The proposed algorithm yields a $(1+\epsilon)$-approximation algorithm for the Wyner-Ziv rate region in time $O(\epsilon^{-4(|\mathcal{X}|+1)})$ as ϵ approaches 0.*

No matter what the initial size of $\mathcal{Z}$, the solution to the linear program satisfies $P_Z(z) = 0$ for all but $|\mathcal{X}|$ values of $z \in \mathcal{Z}$ by the following argument. The linear program has $|\mathcal{Z}|$ variables and $|\mathcal{X}| + |\mathcal{Z}|$ constraints. Since there exists a solution for any linear program at a boundary point, there exists an optimal marginal $\{P^*_Z(z)\}$ for which $|\mathcal{Z}|$ constraints are satisfied with equality. At most $|\mathcal{X}| - 1$ are constraints of the form $\sum_{z\in\mathcal{Z}} P^*_Z(z) Q_{X|Z}(x|z) = p(x)$ for some $x \in \mathcal{X}$, and one constraint ensures that $\sum_{x\in\mathcal{X}} P^*_Z(z) = 1$. The remaining $|\mathcal{Z}| - |\mathcal{X}|$ constraints take the form $P^*_Z(z) = 0$.

Figure 4.3 shows the Wyner-Ziv rate region (solid line) and our algorithm's $(1+\epsilon)$-

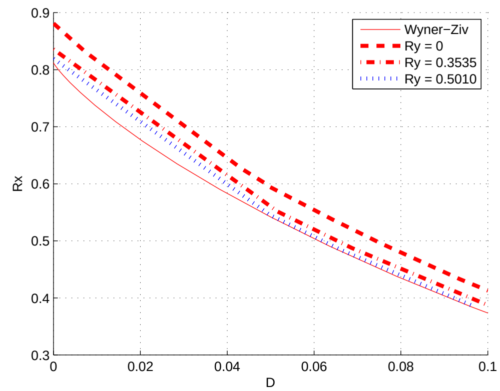


Figure 4.3: Example $(1+\epsilon)$ approximations for the Wyner-Ziv and Berger et al. achievable rate regions.

approximation (circles) when $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ with joint distribution

$$P_{X,Y}(0,0) = 0.06,\ P_{X,Y}(0,1) = 0.24,\ P_{X,Y}(1,0) = 0.42,\ P_{X,Y}(1,1) = 0.28,$$

Hamming distortion measure, and $\epsilon = 0.1$. Again, the approximation is tighter than the $(1+\epsilon)$ worst-case guarantee. While the Wyner-Ziv region is not difficult to calculate for these simple binary sources, the difficulty of the calculation increases with the alphabet size.

4.5 The Lossy Coded Side Information Region

In [6], Berger et al. derive an achievability result for the lossy coded side-information problem illustrated in Figure 4.1(c). Let $\mathcal{X}_1$ and $\mathcal{X}_2$ denote the finite alphabets for sources X_1 and X_2, respectively. The region proposed by Berger et al. is the convex hull of the rates

$$R_1 \geq I(X_1; Z_1|Z_2), \quad R_2 \geq I(X_2; Z_2),$$

for $(Z_1, Z_2) \in \Psi(X_1, X_2)$, where

$$\begin{aligned}\Psi(X_1,X_2) \quad := \quad & \Big\{(Z_1,Z_2)\Big|\, Z_1 \to X_1 \to X_2 \to Z_2, \\ & \exists \psi \text{ s.t. } Ed(X,\psi(Z_1,Z_2)) \le D\Big\}.\end{aligned}$$

We find the desired lower convex hull using Lagrangian

$$\begin{aligned}J_3(\lambda_1,\lambda_2,\lambda_3) \quad := \quad & \lambda_1 I(X_1;Z_1|Z_2) + \lambda_2 I(X_2;Z_2) \\ & +\lambda_3 \min_{\psi} Ed(X_1,\psi(Z_1,Z_2)).\end{aligned}$$

Calculating the optimal rate region for a given pair of sources (X_1, X_2) requires joint optimization of conditionals

$$\{Q_{Z_i|X_i}(z_i|x_i)\}_{(x_i,z_i)\in\mathcal{X}_i\times\mathcal{Z}_i}, \quad i \in \{1,2\}$$

where $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are the alphabets for auxiliary random variables Z_1 and $\mathcal{Z}_2$. Since joint optimization of these conditional distributions is tricky, we define a sequence of conditional distributions $\{Q_{Z_2|X_2}(z_2|x_2)\}_{(x_2,z_2)\in\mathcal{X}_2\times\mathcal{Z}_2}$ from the quantized class defined in (4.1) and then optimize $\{Q_{Z_1|X_1}(z_1|x_1)\}_{(x_1,z_1)\in\mathcal{X}_1\times\mathcal{Z}_1}$ for each. Comparing these optimal solutions yields the best pair of conditionals among all possible solutions in the class considered.

The number of possible conditionals on Z_2 given X_2 in the quantized class is $N(\delta,\eta,|\mathcal{X}_2|)^{|\mathcal{Z}_1|}$. To make this value as small as possible, we begin by bounding the alphabet size $|\mathcal{Z}_2|$. For any fixed conditional distributions $\{Q_{Z_1|X_1}(z_1|x_1)\}_{(x_1,z_1)\in\mathcal{X}_1\times\mathcal{Z}_1}$ and $\{Q_{X_2|Z_2}(x_2|z_2)\}_{(x_2,z_2)\in\mathcal{X}_2\times\mathcal{Z}_2}$, both $J_3(\lambda_1,\lambda_2,\lambda_3)$ and the distribution constraints are linear in $\{P_{Z_2}(z_2)\}_{z_2\in\mathcal{Z}_2}$. An argument analogous to the one in Section 4.4 then demonstrates that there exists an optimal solution to this linear equation in which $P_{Z_2}(z_2) = 0$ for all but at most $|\mathcal{X}_2|$ values of z_2 – giving $|\mathcal{Z}_2| \le |\mathcal{X}_2|$. [2]

Let $\mathcal{Z}_1 = \{1,\ldots,N(\delta,\eta,|\mathcal{X}_1|)\}$, and for each $z_1 \in \mathcal{Z}_1$ let $\{Q_{X_1|Z_1}(x_1|z_1)\}_{x_1\in\mathcal{X}_1}$ be

[2]We can similarly show that $|\mathcal{Z}_1| \le |\mathcal{X}_1|$, though that result is not applied in the argument that follows.

a distinct distribution from our quantized collection (4.1). Let $\mathcal{Z}_2 = \{1, \ldots, |\mathcal{X}_2|\}$. For each of the $N(\delta, \eta, |\mathcal{X}_2|)^{|\mathcal{X}_2|}$ conditionals $\{Q_{X_2|Z_2}(x_2|z_2)\}_{(x_2,z_2)\in\mathcal{X}_2\times\mathcal{Z}_2}$ in the class defined by (4.1), we run a linear program to optimize $J_3(\lambda_1, \lambda_2, \lambda_3)$ subject to the distribution constraints. The algorithm output is the best of these solutions. The proof of Theorem 4.5.1 appears in the Appendix.

Theorem 4.5.1 *The proposed algorithm runs in time $O(\epsilon^{-4(|\mathcal{X}_1|+|\mathcal{X}_2|^2+1)})$ and guarantees a $(1+\epsilon)$-approximation.*

The broken lines in Figure 4.3 show our algorithm's $(1+\epsilon)$-approximation for the achievable rate region from [6] for the example considered in the previous section. Each curve plots rate R_1 against distortion for a fixed value of R_2. In this case, the optimal region is not easily available, but the given solution is guaranteed to meet our $(1+\epsilon)$-approximation bound.

4.6 Summary

The proposed family of algorithms enables systematic calculation of the rate regions for a large class of source coding problems. The ability to calculate these regions is useful because it allows us to determine the limits of what is possible in a variety of applications – thereby enabling an objective assessment of the performance of source coding algorithms.

The given approach may also be useful for resolving theoretical questions. The coded side-information problem provides a potential example. In [47], Berger and Tung derive an inner bound (here called the Berger-Tung bound) for the lossy multiple access source coding problem. While the formulations are quite different, in [48] Jana and Blahut prove the equivalence of the inner bounds from [6] and [47]. A long-standing open question is whether the bound is tight. One possible means of proving the looseness of the bound would be to calculate it for random variables (X, Y) and compare the resulting region to the normalized region for the random variables (X^n, Y^n) where (X_i, Y_i) are drawn i.i.d. according to the same distribution as (X, Y).

If these values differ for any n, then the region is not tight. (The experiment would be inconclusive if the values are the same.) Since direct calculation of these values is difficult even for $n = 2$, the proposed algorithm may enable a solution to this problem.

Chapter 5

The Two-Hop Network and the Diamond Network

5.1 Introduction and Problem Statement

One of the central goals in network source coding theory is to bound rate-distortion regions for source coding in a given network. It is well-known that for the point-to-point network, where the source sequence is drawn i.i.d. according to a known probability mass function, the minimal rate required to describe the source with an arbitrarily small error probability of reproduction is the entropy of the source random variable. Rate distortion theory also gives a formula for $R(D)$, the minimal rate required to achieve an expected per-symbol distortion no larger than D between the source and its reproduction.

The lossless rate regions and rate-distortion regions for source coding of i.i.d. random variables in more general networks are apparently harder to describe. While complete, one-letter characterizations of the rate-distortion regions for some regions are known (for example, see [4], [3], [8], and [49]), many of these results incorporate auxiliary random variables to describe achievable rate vectors; in these cases, characterizing the lossless rate region or rate-distortion region for an example random variable requires solution of a typically non-trivial optimization problem. Lossless rate regions and rate-distortion regions for far more networks remain entirely unsolved.

To date, source coding theory has concentrated primarily on bounds for single-hop networks, assuming that every source has a direct connection to each destination. Many of today's networking applications involve multihop networks, where a data source may be separated from its destination by one or more intermediate nodes, each of which may make its own source requests. While single-hop network source coding solutions can be applied in multihop networks, such applications require explicit rate allocation for each source-destination pair, and the resulting solutions may be suboptimal. As a result, the study of source coding for multihop networks is an important, largely open area for investigation.

Multihop networks exhibit a variety of characteristics absent from prior single-hop networks: (1) a single source description may take multiple paths to its destination; (2) multiple source descriptions may share a single link en route to different destinations; and (3) intermediate nodes may process incoming descriptions and send partial descriptions on to subsequent nodes in the network. The network under investigation here concentrates on the latter two properties. To the authors' knowledge, the only prior rate-distortion theory investigations of multihop networks are Yamamoto's rate-distortion region for a single-path two-hop network without side information [9], where the network focuses on property (3), and bounds on the rate-distortion region for a two-path multihop network [28], the second network we study in this chapter.

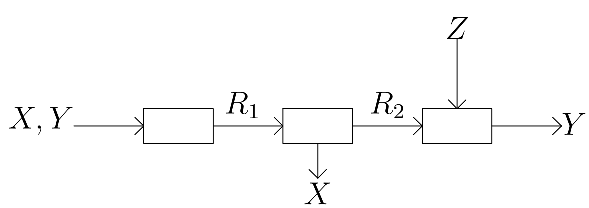


Figure 5.1: The two-hop network.

Figure 5.1 and 5.2 are two multi-hop networks of interest in this chapter, here called the two-hop network and the diamond network respectively. The first network, here called the "two-hop network", is chosen to focus on properties (2) and (3); the second example, here called the "diamond network", generalizes the two-hop

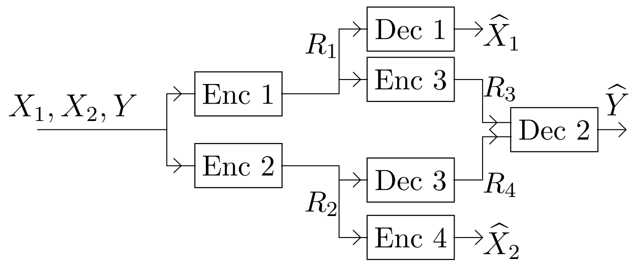


Figure 5.2: The diamond network.

network to introduce property (3). Throughout this chapter, in-arrows designate source observations, out-arrows designate source requests, and all the links are error free and directed. The goal is to study the rate-distortion regions for these two networks. Rate regions for networks with links in series and networks with links in parallel are obvious first steps in understanding rate regions for general networks.

Section 5.2 includes definitions and basic properties that are applied in bounding rate regions of the networks we investigate in this chapter. Section 5.3 derives inner and outer bounds for the rate-distortion region for the two-hop network. This network includes the one introduced by Yamamoto in [9] as a special case. The given derivation is similar to the derivation of the rate-distortion region for a network with unreliable side information at the decoder [8]. Section 5.4 applies the result of Section 5.3 to derive two inner bounds and one outer bound of the rate-distortion region for the diamond network. Section 5.5 further investigates these bounds under several assumptions on the diamond network's source random variables.

5.2 Basic Properties

In this section, we list background that is useful in the derivations that follow.

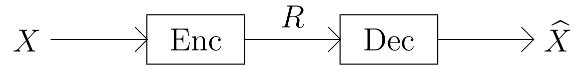


Figure 5.3: Lossy source coding for the point-to-point network.

Rate Distortion Functions

The rate-distortion function for the point-to-point network of Figure 5.3 is

$$R(D) = \min_{p(\widehat{X}|X):E(d(X,\widehat{X}))\leq D} I(X;\widehat{X}).$$

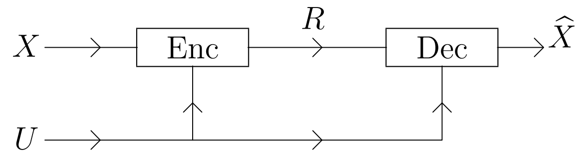


Figure 5.4: Lossy source coding with side information at the encoder and decoder.

If we modify the traditional source coding problem to allow side information U at both the encoder and the decoder, as shown in Figure 5.4, then the optimal source coding performance is the conditional rate distortion function of X given U [5]

$$R_{X|U}(D) = \min_{p(\widehat{X}|X,U):E(d(X,\widehat{X}))\leq D} I(X;\widehat{X}|U). \tag{5.1}$$

Detailed discussions of conditional rate distortion functions appear in [50], [51], [52], and [3]. We here state several properties that are particularly useful for our analysis. First, from [52], for any $D \geq 0$,

$$R_{X|U}(D) = \min_{\{D_U\}_{u\in\mathcal{U}}:E(D_U)\leq D} \sum_u \Pr\{U=u\} R_{X|U=u}(D_u), \tag{5.2}$$

where for each $u \in \mathcal{U}$, $R_{X|U=u}(D_u)$ is the rate distortion function with respect to probability distribution $P_{X|U=u}(x)$ on alphabet $\mathcal{X}$. Second, given side information sources U and V, for any $D \geq 0$,

$$R_{X|U,V}(D) \leq R_{X|U}(D). \tag{5.3}$$

The third property is stated in Lemma 5.2.1.

Lemma 5.2.1 *[3] Let $U_1, \ldots, U_n$ be n random variables with mutually disjoint alphabets $\mathcal{U}_1, \ldots, \mathcal{U}_n$. Let $\{D_1, \ldots, D_n\}$be a collection of n distortions, where $D_i \geq 0$ for all i. Let $X_1, \ldots, X_n$ be drawn i.i.d. according to distribution $P_X(\cdot)$ on alphabet $\mathcal{X}$. Let Q be a random variable uniformly distributed on $\{1, \ldots, n\}$. Define $X := X_Q$ and $U := (U_Q, Q)$. Then*

$$R_{X|U}\left(\frac{1}{n}\sum_{i=1}^{n} D_i\right) \leq \frac{1}{n}\sum_{i=1}^{n} R_{X_i|U_i}(D_i).$$

Strong Typicality

We use strong typicality (see, for example, [44]) to prove achievability results for the networks of interest in this chapter. We use strong typicality rather than weak typicality to take advantage of tighter available bounds on the size of the strongly typical set and the Markov property described in Lemma 5.2.6 below. We here set up notation and summarize a few useful results. Assume B is a finite-alphabet random variable. Let $N(\beta|b^n)$ denote the number of appearances of symbol β in string b^n. We use the notation $A_\epsilon^{*(n)}(B)$ to denote the strongly typical set for random variable B on alphabet $\mathcal{B}$, where $A_\epsilon^{*(n)}(B)$ is the set of sequences $b^n \in \mathcal{B}^n$ satisfying

(a)
$$\left|\frac{N(\beta|b^n)}{n} - p(\beta)\right| < \frac{\epsilon}{|\mathcal{B}|}$$
for every $\beta \in \mathcal{B}$ with $p(\beta) > 0$.

(b) $N(\beta|b^n) = 0$ for all $\beta \in \mathcal{B}$ with $p(\beta) = 0$.

If C is another random variable and $c^n \in A_\epsilon^{*(n)}(C)$, define

$$A_\epsilon^{*(n)}(B|c^n) := \{b^n \in B^n | (b^n, c^n) \in A_\epsilon^{*(n)}(B, C)\}.$$

Lemma 5.2.2 *[44] For any $\epsilon > 0$ and $n \in \mathbb{N}$. If $x^n \in A_\epsilon^{*(n)}(X)$, then*

$$2^{n(H(Y|X)-\epsilon')} \leq |A_\epsilon^{*(n)}(Y|x^n)| \leq 2^{n(H(Y|X)+\epsilon')},$$

where ϵ' can be made arbitrarily small by making n sufficiently large and ϵ sufficiently small.

Remark 5.2.3 *For any random variables W, $\widehat{W}$ with bounded distortion measure $d : \mathcal{W} \times \widehat{\mathcal{W}} \to [0, d_{max}]$ and every jointly strongly typical pair $(w^n, \widehat{w}^n) \in \mathcal{W}^n \times \widehat{\mathcal{W}}^n$,*

$$\begin{aligned}
&\left| \frac{1}{n} d(w^n, \widehat{w}^n) - Ed(W, \widehat{W}) \right| \\
&= \left| \sum_{\alpha \in \mathcal{W}, \beta \in \widehat{\mathcal{W}}} \left(\frac{N(\alpha, \beta | w^n, \widehat{w}^n)}{n} - p(\alpha, \beta) \right) d(\alpha, \beta) \right| \\
&\leq d_{\max} \sum_{\alpha \in \mathcal{W}, \beta \in \widehat{\mathcal{W}}} \frac{\epsilon}{|\mathcal{W}||\widehat{\mathcal{W}}|} \\
&= \epsilon \cdot d_{\max}.
\end{aligned}$$

The proofs of Corollaries 5.2.4 and 5.2.5 follow from counting arguments based on Lemma 5.2.2.

Corollary 5.2.4 *Given a probability distribution $p(x, y, w)$, fix any pair $(x^n, w^n) \in A_\epsilon^{*(n)}(X, W)$, and choose a sequence Y^n uniformly at random from the set $A_\epsilon^{*(n)}(Y|w^n)$. Then*

$$\Pr(Y^n \in A_\epsilon^{*(n)}(Y|(x^n, w^n)) \geq 2^{-n(I(X;Y|W)+\epsilon_1)},$$

where ϵ_1 can be made arbitrarily small by making n sufficiently large and ϵ sufficiently small.

Corollary 5.2.5 *Given a probability distribution $p(x, y, w)$ and a sequence $w^n \in A_\epsilon^{*(n)}(W)$, independently choose 2^{nR} sequences Y_1^n, Y_2^n,..., $Y_{2^{nR}}^n$ from the set $A_\epsilon^{*(n)}(Y|w^n)$. When $R > I(X;Y|W)$,*

$$\Pr\Big((X^n, Y_i^n, w^n) \in A_\epsilon^{*(n)}(X, Y, W) \\ \text{for some } i \in \{1, 2, \ldots, 2^{nR}\}\Big) \to 1$$

as $n \to \infty$.

Lemma 5.2.6 *[44,* **Lemma 14.8.1***] Let $X \to Y \to Z$ form a Markov chain. If, for a given $(y^n, z^n) \in A_\epsilon^{*(n)}(Y, Z)$, X^n is chosen uniformly at random from the set of x^n that are jointly typical with y^n, then*

$$Pr\{X^n \in A_\epsilon^{*(n)}(X|y^n, z^n)\} > 1 - \epsilon$$

for sufficiently large n.

5.3 Source Coding for the Two-Hop Network

Consider the network source coding problem for the two-hop network shown in Figure 5.1. The random sequence $(X_1, Y_1, Z_1), (X_2, Y_2, Z_2), \ldots$ is drawn i.i.d. according to joint probability mass function $p(x, y, z)$ on finite alphabet $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. The transmitter observes sources X and Y and describes them at rate R_1 to the middle node. The middle node uses its received description to build a reconstruction $\widehat{X}$ of source X and to create a rate-R_2 description ($R_2 \leq R_1$) for transmission to the final receiver. The final receiver combines its received description with observed side information Z to build a reconstruction $\widehat{Y}$ of source Y. We measure the accuracy of reconstructions $\widehat{X}$ and $\widehat{Y}$ using distortion measures $d_X : \mathcal{X} \times \widehat{\mathcal{X}} \to [0, \infty)$ and $d_Y : \mathcal{Y} \times \widehat{\mathcal{Y}} \to [0, \infty)$. We assume that all source random variable alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are finite. We also drop the subscripts from d for notational simplicity and use $d_{\max}$ to denote the maximal distortion value. In this section, we bound this rate-distortion region. The proof that follows combines ideas from Wyner-Ziv coding and coding with unreliable side information [8].

The proposed inner bound appears in Theorem 5.3.1 in Section 5.3.1. The proposed outer bound appears in Theorem 5.3.3 in Section 5.3.2. In addition to its direct interest, the given result can serve as a stepping stone for understanding more general multihop source coding problems.

5.3.1 Inner Bound

Theorem 5.3.1 *Rate vector (R_1, R_2) is (D_X, D_Y)-achievable for the two-hop network if there exist finite-alphabet, auxiliary random variables U and V for which*

$$\begin{aligned} R_1 &> R_{X|U}(D_X) + I(X,Y;U) + I(X,Y;V|U,Z) \\ R_2 &> I(X,Y;U|Z) + I(X,Y;V|U,Z) \end{aligned}$$

and

(i) $Z \to (X,Y) \to (U,V)$ forms a Markov chain.

(ii) There exists a function $\widehat{Y}(U,V,Z)$ such that

$$E(d(Y, \widehat{Y}(U,V,Z))) \leq D_Y.$$

Proof. Let (U,V) be a pair of random variables satisfying conditions (i) and (ii) in Theorem 5.3.1. To prove achievability, it suffices to show that for any $\delta > 0$, the vector $(R_1(\delta), R_2(\delta))$ defined by

$$\begin{aligned} R_1(\delta) &= R_{X|U}(D_X) + I(X,Y;U) + I(X,Y;V|U,Z) + 3\delta \\ R_2(\delta) &= I(X,Y;U,V|U,Z) + 2\delta \end{aligned}$$

is achievable.

Fix $n \in \mathbb{N}$. Let $\widehat{X}$ be a random variable defined by a conditional probability distribution $p(\widehat{x}|x,u)$ such that $I(X;\widehat{X}|U) = R_{X|U}(D_X) + \delta/2$ and $E(d(X,\widehat{X})) \leq D_X$. Define

$$\begin{aligned} S_1 &:= 2^{n(I(X,Y;V|U)+\delta)} & S_2 &:= 2^{n(I(X,Y;V|U,Z)+2\delta)} \\ M_1 &:= 2^{n(I(X,Y;U)+\delta)} & M_2 &:= 2^{n(I(X,Y;U|Z)+2\delta)} \\ T &:= 2^{n(I(X;\widehat{X}|U)+\delta)}. \end{aligned}$$

Generate the codebook as follows:

1. Draw M_1 sequences $U^n(1)$,...,$U^n(M_1)$ i.i.d. according to probability mass function $\prod_{i=1}^{n} p(u_i)$.

2. Color, uniformly and at random, each $m_1 \in \{1,\ldots,M_1\}$ one of M_2 distinct colors, denoting the color of m_1 by $\nu(m_1)$. Notice that $U \to (X,Y) \to Z$ implies $M_2 \leq M_1$.

3. For each $m_1 \in \{1,\ldots,M_1\}$, draw S_1 sequences $V^n(m_1,1),\ldots,V^n(m_1,S_1)$ uniformly at random from the set $A_\epsilon^{*(n)}(V|U^n(m_1))$.

4. Color, uniformly and at random, each $s_1 \in \{1,\ldots,S_1\}$ one of S_2 distinct colors, denoting the color of s_1 by $\tau(s_1)$. Notice that $U \to (X,Y) \to Z$ implies $S_2 \leq S_1$.

5. For each $m_1 \in \{1,\ldots,M\}$, draw T sequences $\widehat{X}^n(m_1,1)$,...,$\widehat{X}^n(m_1,T)$ uniformly at random from the set $A_\epsilon^{*(n)}(X|U^n(m_1))$.

Let (x^n,y^n) and z^n be the source pair to be transmitted and the observed side information, respectively. We use the following functions in defining the encoders

$$
\begin{aligned}
\psi(x^n, y^n) &:= \min\Big[\{M_1\} \cup \{m_1 \in \{1, \ldots, M_1\} : \\
&\quad (x^n, y^n, U^n(m_1)) \in A_\epsilon^{*(n)}(X, Y, U)\}\Big]. \\
\mu(x^n, y^n) &:= \min\big[\{S_1\} \cup \{j \in \{1, \ldots, S_1\} : \\
&\quad (x^n, y^n, U^n(\psi(x^n, y^n)), V^n(\psi(x^n, y^n), j)) \\
&\quad \in A_\epsilon^{*(n)}(X, U, V)\}\big]. \\
\phi(x^n, y^n) &:= \min\big[\{T\} \cup \{t \in \{1, 2, \ldots, T\} : \\
&\quad (x^n, U^n(\psi(x^n, y^n)), \widehat{X}^n(\psi(x^n, y^n), t)) \\
&\quad \in A_\epsilon^{*(n)}(X, U, \widehat{X})\}\Big]. \\
\pi(x^n, y^n) &:= \tau(\mu(x^n, y^n)).
\end{aligned}
$$

The purpose of the function τ is as follows. S_1 is the number of V-sequences needed to cover all strongly typical (X^n, Y^n) pairs when U is known. S_2 is the number of V-sequences needed to cover all strongly typical (X^n, Y^n) pairs when both U^n and Z^n are known. We randomly bin the index j that specifies V^n into one of the S_2 slots. Since the decoder has access to the side information Z^n, it is possible to recover (with high probability) j from $\tau(j)$. The reason for using the function ν is similar.

Finally, we define the encoders as

$$
\begin{aligned}
f(x^n, y^n) &:= (\psi(x^n, y^n), \pi(x^n, y^n), \phi(x^n, y^n)) \\
g(f(x^n, y^n)) &:= (\nu(\psi(x^n, y^n)), \pi(x^n, y^n)).
\end{aligned}
$$

The decoder at the middle node maps index (m_1, p, t), the received string from the encoding function f, to reconstruction

$$\widehat{X}_j^n(m_1, t).$$

When receiving (m_2, p), the decoder at the final receiver begins by finding $\widehat{m}_1$ from $\{1, \ldots, M_1\}$ such that

$$\begin{aligned}&(U^n(\widehat{m}_1), Z^n) \in A_\epsilon^{*(n)}(U, Z)\\ &\text{and } \nu(\widehat{m}_1) = m_2.\end{aligned} \tag{5.4}$$

Define the function ι_1 mapping $\{1, \ldots, M_2\}$ to $\{1, \ldots, M_1\}$ as follows

$$\iota_1(m_2) := \begin{cases} \widehat{m}_1, & \text{if } \widehat{m}_1 \text{ is the unique index satisfying (5.4)} \\ 1, & \text{otherwise.} \end{cases}$$

Then the decoder finds $\widehat{j}$ from $\{1, \ldots, S_1\}$ such that

$$\begin{aligned}&(U^n(\iota_1(m_2)), V^n(\iota_1(m_2), \widehat{j}), Z^n) \in A_\epsilon^{*(n)}(U, V, Z)\\ &\text{and } \tau(\widehat{j}) = p.\end{aligned} \tag{5.5}$$

Define the function ι_2 mapping $\{1, \ldots, M_2\} \times \{1, \ldots, S_2\}$ to $\{1, \ldots, S_1\}$ as follows

$$\iota_2(m_2, p) := \begin{cases} \widehat{j}, & \text{if } \widehat{j} \text{ is the unique index satisfying (5.5)} \\ 1, & \text{otherwise.} \end{cases}$$

The final receiver then builds its reconstruction for Y^n as

$$\widehat{Y}^n(U^n(\iota_1(m_2)), V^n(\iota_1(m_2), \iota_2(m_2, p)), Z^n).$$

Analysis of performance:

For simplicity, we denote by ψ, μ, ϕ, ι, and π the evaluated values of the corresponding functions on (x^n, y^n). Define the following error events:

1. $E_0 : (x^n, y^n, z^n) \notin A_\epsilon^{*(n)}(X, Y, Z)$.

2. $E_U :$ For all $i \in \{1, 2, \ldots, M\}$ and for all $j \in \{1, 2, \ldots, T\}$,

$$(x^n, y^n, z^n, U^n(i), V^n(i, j)) \notin A_\epsilon^{*(n)}(X, Y, Z, U, V).$$

3. E_X : For all $i \in \{1, 2, \ldots, T\}$,

$$(x^n, U^n(\psi), \widehat{X}^n(\psi, i)) \notin A_\epsilon^{*(n)}(X, U, \widehat{X}).$$

4. E_{UZ} : E_U^c happens and there exists a m_1 such that $m_1 \neq \psi$ and

$$(U^n(m_1), z^n) \in A_\epsilon^{*(n)}(U, Z)$$

   and $\nu(m_1) = \nu(\psi)$.

5. E_V : E_U^c happens and there exists a j such that $j \neq \mu$ and

$$(U^n(\psi), V^n(\psi, j), z^n) \in A_\epsilon^{*(n)}(U, V, Z)$$

   and $\tau(j) = \pi$.

6. $E_{\text{error}} = E_0 \cup E_U \cup E_X \cup E_{UZ} \cup E_V$.

Let $\mathbb{C}_n$ denote the ensemble for all such codes of length n. We prove the existence of achievable codes of the given rates by showing that $\lim_{n\to\infty} E_{\mathbb{C}_n}[\Pr(E_{\text{error}})] = 0$. First, from the basic property of typical sets, $\Pr(E_0)$ can be made arbitrarily small as n grows without bound. By Corollary 5.2.5 and Lemma 5.2.6, $\Pr(E_U)$ and $\Pr(E_X)$ can also be made arbitrarily small as n grows without bound. When $z^n \in A_\epsilon^{*(n)}(Z)$, the probability that a randomly chosen sequence u^n is jointly strongly typical with z^n is bounded by $2^{-n(I(U;Z)-\delta/2)}$ when n is sufficiently large. Hence when n is sufficiently large,

$$\begin{aligned}
&E_{\mathbb{C}_n}\left[\Pr(E_{UZ})\right] \\
&= E_{\mathbb{C}_n}\left[\Pr(\cup_{u^n\in B(\nu(\psi))}\Pr(u^n \in A_\epsilon^{*(n)}(U|z^n)))\right] \\
&\leq E_{\mathbb{C}_n}\left(|B(\nu(\psi))|\right) E_{\mathbb{C}_n}\left[\Pr(u^n \in A_\epsilon^{*(n)}(U|z^n))\right] \\
&\leq 2^{M_1-M_2}2^{-n(I(U;Z)-\frac{\delta}{2})} = 2^{-n\frac{\delta}{2}},
\end{aligned}$$

where $B(\nu(\psi))$ is the set of indices $m_1 \in \{1, \dots, M_1\}$ such that $\nu(m_1) = \nu(\psi)$. Therefore, $\lim_{n\to\infty} E[\Pr(E_{UZ})] = 0$.

Finally, by the similar argument, when $(U^n(\psi), z^n) \in A_\epsilon^{*(n)}(U, Z)$, since the probability that a randomly chosen sequence $v^n \in A_\epsilon^{*(n)}(V|U^n(\psi))$ is jointly strongly typical with $(U^n(\psi), z^n)$ is bounded by $2^{-n(I(V;Z|U)-\delta/2)}$, the expectation of the probability of the event E_V can be bounded (for n large enough) by

$$\begin{aligned} E_{\mathbb{C}_n}[\Pr(E_V)] &\leq E_{\mathbb{C}_n}[|C(\pi)|]\, 2^{-n(I(V;Z|U)-\frac{\delta}{2})} \\ &= 2^{S_1-S_2} 2^{-n(I(V;Z|U)-\frac{\delta}{2})}, \end{aligned}$$

where $C(\pi)$ is the set of indices $j \in \{1, \dots, S_1\}$ such that $\tau(j) = \pi$. Since

$$\begin{aligned} & I(X,Y;V|U) - I(X,Y;V|U,Z) \\ &= (H(V|U) - H(V|U,X,Y)) - (H(V|U,Z) \\ &\quad -H(V|X,Y,U,Z)) \\ &= H(V|U) - H(V|U,Z) \\ &\quad +(H(V|X,Y,U,Z) - H(V|U,X,Y)) \\ &= I(V;Z|U), \end{aligned}$$

$E_{\mathbb{C}_n}[\Pr(E_V)]$ can also be made arbitrarily small as n grows without bound. Thus by the union bound,

$$\lim_{n\to\infty} E_{\mathbb{C}_n}[\Pr(E_{\text{error}})] = 0.$$

Now fix a code of length n (for sufficiently large n) such that $\Pr(E_{\text{error}}) < \epsilon$. From Remark 5.2.3, the average distortion between any pair of jointly typical sequences is close to the expected distortion. Therefore,

$$
\begin{aligned}
&|Ed(X^n, \widehat{X}^n(\psi,\phi)) - D_X| \\
&= \Big|\Pr(E^c_{\text{error}}) \left[E\left(d(X^n, \widehat{X}^n(\psi,\phi)|E^c_{\text{error}}\right) - D_X\right] \\
&\qquad + \Pr(E_{\text{error}}) \left[E\left(d(X^n, \widehat{X}^n(\psi,\phi)|E_{\text{error}}\right) - D_X\right]\Big| \\
&\le \Pr(E^c_{\text{error}}) \cdot \epsilon \cdot d_{\max} + \Pr(E_{\text{error}}) \cdot d_{\max} \\
&\le \epsilon \cdot d_{\max} + \epsilon \cdot d_{\max}
\end{aligned}
$$

for n sufficiently large. Similarly,

$$
\begin{aligned}
&|Ed(Y^n, \widehat{Y}^n(U^n(\psi), V^n(\psi, \iota(\psi,\tau)), Z^n) - D_Y| \\
&< 2\epsilon d_{\max}
\end{aligned}
$$

for n sufficiently large. Since $\epsilon > 0$ can be made arbitrarily small, this coding scheme satisfies the distortion requirement (D_X, D_Y). This completes the proof. $\square$

The following corollary is a direct consequence from Theorem 5.3.1 for the case of lossless source coding.

Corollary 5.3.2 *Rate vector (R_1, R_2) is in the achievable rate region for lossless source coding in the two-hop network if there exists a finite-alphabet, auxiliary random variable U for which*

$$
\begin{aligned}
R_1 &\ge H(X|U) + I(X,Y;U) + H(Y|U,Z) \\
R_2 &\ge I(X,Y;U|Z) + H(Y|U,Z)
\end{aligned}
$$

and $Z \to (X,Y) \to U$ forms a Markov chain.

Intuitively, auxiliary random variable U represents the common information that is useful to both the middle node without use of side information Z and the final receiver with use of side information Z. Auxiliary random variable V represents the private information that is accessible only to the final receiver, which uses its knowledge of side information Z to describe the useful portion of V. In the lossy bound, the final node uses the knowledge of U and Z to reconstruct $\widehat{Y}$; the lossless bound replaces condition (ii) with an explicit inclusion of any additional rate $H(Y|U,Z)$ that may be required to reconstruct Y given the knowledge of U and Z.

5.3.2 Outer Bound

Theorem 5.3.3 *If the rate vector (R_1, R_2) is (D_X, D_Y)-achievable for the two-hop network, then for any $\epsilon > 0$, there exists a finite-alphabet, auxiliary random variable U for which*

$$
\begin{aligned}
R_1 &\geq I(X,Y;U) + I(X,Y;V|U,Z) \\
R_2 &\geq I(X,Y;V|Z)
\end{aligned}
$$

and

(i) $Z \to (X,Y) \to (U,V)$ forms a Markov chain.

(ii) There exists a function $\widehat{X}(U)$ such that

$$E(d(X,\widehat{X}(U))) \leq D_X + \epsilon.$$

(iii) There exists a function $\widehat{Y}(V,Z)$ such that

$$E(d(Y,\widehat{Y}(U,Z))) \leq D_Y + \epsilon.$$

Proof. Consider a sequence $\{\mathcal{C}_n\}_{n=1}^{\infty}$ of block length-n codes. Let

$$
\begin{aligned}
f_n &: \mathcal{X}^n \times \mathcal{Y}^n \to \{1,\ldots,2^{nR_1}\} \\
g_n &: \{1,\ldots,2^{nR_1}\} \to \{1,\ldots,2^{nR_2}\}
\end{aligned}
$$

denote the rate-R_1 encoder at the transmitter and rate-R_2 encoder at the middle node, respectively. Suppose that the distortions of the given codes approach D_X and D_Y as n grows without bound. Then for any $\epsilon > 0$, there exists an n sufficiently large such that the distortions achieved by code $\mathcal{C}_n$ are no greater than $D_X + \epsilon$ and $D_Y + \epsilon$. Let $\widehat{X}^n$ and $\widehat{Y}^n$ be the corresponding reproductions. For $i \in \{1, \ldots, n\}$, define $D_{X,i} := E[d(X_i, \widehat{X}_i)]$ and $D_{Y,i} := E[d(Y_i, \widehat{Y}_i)]$. By assumption,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} E[d(X_i, \widehat{X}_i)] &= \frac{1}{n}\sum_{i=1}^{n} D_{X,i} \leq D_X + \epsilon \\ \frac{1}{n}\sum_{i=1}^{n} E[d(Y_i, \widehat{Y}_i)] &= \frac{1}{n}\sum_{i=1}^{n} D_{Y,i} \leq D_Y + \epsilon. \end{aligned} \tag{5.6}$$

For random source sequences (X^n, Y^n), let $F = f_n(X^n, Y^n)$ and $G = g_n(F)$ represent the random variables transmitted through the first and second links respectively. Further, define $U_i := (F, Z_1^{i-1})$ and $\widehat{V}_i := (F, Z_{i+1}^n, Z_1^{i-1})$ for $1 \leq i \leq n$. Then

$$\begin{aligned} H(F) &= I(X^n, Y^n; F) \\ &= I(X^n, Y^n; F, Z^n) - I(X^n, Y^n; Z^n|F) \\ &= \sum_{i=1}^{n} \big[I(X_i, Y_i; F, Z^n | X_1^{i-1}, Y_1^{i-1}) \\ &\quad - I(X^n, Y^n; Z_i | F, Z_1^{i-1}) \big] \\ &= \sum_{i=1}^{n} \big[I(X_i, Y_i; F, Z^n, X_1^{i-1}, Y_1^{i-1}) \\ &\quad - I(X_i, Y_i; X_1^{i-1}, Y_1^{i-1}) \\ &\quad - I(X^n, Y^n; Z_i | F, Z_1^{i-1}) \big], \end{aligned}$$

where the first equality follows from the fact that F is a deterministic function of X^n and Y^n. Now since for each i, (X_i, Y_i) is independent of (X_1^{i-1}, Y_1^{i-1}),

$$I(X_i, Y_i; X_1^{i-1}, Y_1^{i-1}) = 0$$

and so

$$\begin{aligned}
&I(X_i, Y_i; F, Z^n, X_1^{i-1}, Y_1^{i-1}) - I(X_i, Y_i; X_1^{i-1}, Y_1^{i-1}) \\
&\quad = I(X_i, Y_i; F, Z^n, X_1^{i-1}, Y_1^{i-1}) \\
&\quad \geq I(X_i, Y_i; F, Z^n).
\end{aligned}$$

On the other hand, for every i, the Markov chain condition $((X_j, Y_j)_{j \neq i}, F, Z_1^{i-1}) \to (X_i, Y_i) \to Z_i$ implies that

$$\begin{aligned}
&I(X^n, Y^n; Z_i | F, Z_1^{i-1}) \\
&\quad = H(Z_i | F, Z_1^{i-1}) - H(Z_i | X_i, Y_i, F, Z_1^{i-1}, (X_j, Y_j)_{j \neq i}) \\
&\quad = H(Z_i | F, Z_1^{i-1}) - H(Z_i | X_i, Y_i, F, Z_1^{i-1}) \\
&\quad = I(X_i, Y_i; Z_i | F, Z_1^{i-1}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
H(F) \ &\geq \ \sum_{i=1}^{n} I(X_i, Y_i; F, Z_1^{i-1}, Z_i, Z_{i+1}^n) \\
&\quad - \sum_{i=1}^{n} I(X_i, Y_i; Z_i | F, Z_1^{i-1}) \\
&= \ \sum_{i=1}^{n} [I(X_i, Y_i; U_i, \widehat{V}_i, Z_i) - I(X_i, Y_i; Z_i | U_i)] \\
&= \ \sum_{i=1}^{n} \Big[I(X_i, Y_i; U_i) + I(X_i, Y_i; \widehat{V}_i, Z_i | U_i) \\
&\quad - I(X_i, Y_i; Z_i | U_i)] \\
&= \ \sum_{i=1}^{n} [I(X_i, Y_i; U_i) + I(X_i, Y_i; \widehat{V}_i | U_i, Z_i)].
\end{aligned}$$

Thus

$$nR_1 \geq H(F) \geq \sum_{i=1}^{n} [I(X_i, Y_i; U_i) + I(X_i, Y_i; \widehat{V}_i | U_i, Z_i)].$$

For every $i \in \{1, 2, \ldots, n\}$, define $V_i := (G, Z_{i+1}^n, Z_1^{i-1})$. Then we have

$$\begin{aligned}
nR_2 &\geq H(G) \geq H(G|Z^n) = I(X^n, Y^n; G|Z^n) \\
&= \sum_{i=1}^{n} I(X_i, Y_i; G|Z^n, X_1^{i-1}, Y_1^{i-1}) \\
&= \sum_{i=1}^{n} I(X_i, Y_i; G, X_1^{i-1}, Y_1^{i-1}, Z_1^{i-1}, Z_{i+1}^n|Z_i) \\
&\geq \sum_{i=1}^{n} I(X_i, Y_i; G, Z_1^{i-1}, Z_{i+1}^n|Z_i) \\
&= \sum_{i=1}^{n} I(X_i, Y_i; V_i|Z_i).
\end{aligned}$$

Now since G is a deterministic function of F, $\widehat{V}_i$ is a deterministic function of V_i and hence

$$\begin{aligned}
\sum_{i=1}^{n} I(X_i, Y_i; \widehat{V}_i|U_i, Z_i) &= \sum_{i=1}^{n} I(X_i, Y_i; \widehat{V}_i, V_i|U_i, Z_i) \\
&\geq \sum_{i=1}^{n} I(X_i, Y_i; V_i|U_i, Z_i).
\end{aligned}$$

Therefore,

$$nR_1 \geq \sum_{i=1}^{n} [I(X_i, Y_i; U_i) + I(X_i, Y_i; V_i|U_i, Z_i)].$$

Let Q denote a random variable uniformly distributed on $\{1, 2, \ldots, n\}$ that is independent of (X^n, Y^n, Z^n). Define $U := (U_Q, Q)$ and $V := (V_Q, Q)$. Since (X_i, Y_i, Z_i), $i \in \{1, 2, \ldots, n\}$, is drawn i.i.d., the joint distribution of (X_Q, Y_Q, Z_Q) is the same as that of (X, Y, Z). Furthermore, Q is independent of (X_Q, Y_Q, Z_Q), hence

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} I(X_i, Y_i; U_i) &= I(X_Q, Y_Q; U_Q|Q) \\
&= H(X_Q, Y_Q|Q) - H(X_Q, Y_Q|U_Q, Q) \\
&= H(X_Q, Y_Q) - H(X_Q, Y_Q|U) \\
&= I(X_Q, Y_Q; U).
\end{aligned}$$

Similarly, one can show that

$$\frac{1}{n}\sum_{i=1}^{n} I(X_i, Y_i; V_i|U_i, Z_i) = I(X_Q, Y_Q; V|U, Z_Q)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I(X_i, Y_i; V_i|Z_i) = I(X_Q, Y_Q; V|Z_Q).$$

Redefine $X = X_Q$, $Y = Y_Q$, and $Z = Z_Q$. Then X, Y, and Z have the same joint distribution $p(x, y, z)$. Therefore,

$$\begin{aligned}
R_1 &\geq I(X, Y; U) + I(X, Y; V|U, Z) \\
R_2 &\geq I(X, Y; V|Z).
\end{aligned}$$

Given the definition $(U_i, V_i) := (G, X_1^{i-1}, Y_1^{i-1}, Z_1^{i-1}, Z_{i+1}^n)$, $Z_i \to (X_i, Y_i) \to (U_i, V_i)$ forms a Markov chain, as does $Z \to (X, Y) \to (U, V)$.

It remains to check conditions (ii) and (iii) in the statement of the theorem. First, since $\widehat{X}_i$ is a function of F for every $i \in \{1, 2, \ldots, n\}$, by defining $\widehat{X} := \widehat{X}_Q$, (ii) is immediate.

For condition (iii), since the reproduction $\widehat{Y}^n$ of Y^n is a deterministic function of (G, Z^n), we have

$$0 = H(\widehat{Y}^n|G, Z^n) = \sum_{i=1}^{n} H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, G, Z^n),$$

which implies that

$$\begin{aligned}
& H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, U_i, V_i, Z_i) \\
& \quad \leq \quad H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, G, Z_1^{i-1}, Z_i, Z_{i+1}^n) = 0
\end{aligned}$$

for all $i \in \{1, 2, \ldots, n\}$. For $i = 1$, $H(\widehat{Y}_1|U_1, V_1, Z_1) = 0$ implies that $\widehat{Y}_1$ is a deterministic function of (U_1, V_1, Z_1). For $i > 1$, assume that $\widehat{Y}_j$ is a function of (U_j, V_j, Z_j) for all $j < i$. Then since for all $j < i$,

$$\begin{aligned}
(U_i, V_i, Z_i) &= (F, G, Z^n) = (F, G, Z^n) \\
&= (U_j, V_j, Z_j),
\end{aligned}$$

$\widehat{Y}_j$ is also a function of (U_i, V_i, Z_i). Therefore,

$$0 = H(\widehat{Y}_i|\widehat{Y}_1^{i-1}, U_i, V_i, Z_i) = H(\widehat{Y}_i|U_i, V_i, Z_i).$$

Thus by induction on i, $\widehat{Y}_i$ is a function of (U_i, V_i, Z_i). By defining $\widehat{Y} := \widehat{Y}_Q$, $\widehat{Y}$ is a function of (U, V, Z) and

$$E(d(Y, \widehat{Y})) \leq D_Y + \epsilon.$$

□

5.4 Performance Bounds for the Diamond Network

Using techniques similar to those of Section 5.3, we next derive one outer bound and two inner bounds for the rate distortion region $\mathcal{R}(D_1, D_2, D_Y)$ with distortion constraints D_1, D_2, and D_Y for the diamond network in Figure 5.2. In this diamond network, node 0 observes samples of sources X_1, X_2,and Y which are independent and identically distributed (i.i.d.) according to distribution $p(x_1, x_2, y)$. As shown in Figure 5.2, nodes 1, 2, and 3 are required to reproduce sources X_1, X_2, and Y, respectively, with corresponding distortion requirements D_1, D_2, and D_Y. Each rate value is a vector (R_1, R_2, R_3, R_4) describing the rates traversing edges $(0,1)$, $(0,2)$, $(1,3)$, and $(2,3)$, respectively. The description of Y can take two possible paths to node 3, passing either through node 1 or node 2 or both. We define $\mathcal{R}(D_1, D_2, D_Y)$ as the closure of the set of achievable rate vectors that satisfy the distortion constraint (D_1, D_2, D_Y). We aim to bound the rate distortion-region $\mathcal{R}(D_1, D_2, D_Y)$ of the diamond network.

Similar to Section 5.3, we make an assumption that all source random variable alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ are finite. Also, for notational simplicity, d denotes the distortion measure and d_{max} denotes the maximal distortion value.

5.4.1 Outer Bound

Theorem 5.4.1 *If (R_1, R_2, R_3, R_4) is a (D_1,D_2,D_Y)-achievable rate region for the diamond network, then for any $\epsilon > 0$, there exist finite-alphabet random variables U_1 and U_2 defined by conditional probability mass function $P_{U_1,U_2|X_1,X_2,Y}$ such that*

$$\begin{aligned} R_1 &\geq R_{X_1|U_1}(D_1+\epsilon) + I(X_1, X_2, Y; U_1) \\ R_2 &\geq R_{X_2|U_2}(D_2+\epsilon) + I(X_1, X_2, Y; U_2) \\ R_3 &\geq I(X_1, X_2, Y; U_1) \\ R_4 &\geq I(X_1, X_2, Y; U_2) \end{aligned} \tag{5.7}$$

and that there exists a function $\widehat{Y}$ of U_1, U_2 satisfying $Ed(\widehat{Y}, Y) \leq D_Y + \epsilon$.

Proof. Let (R_1, R_2, R_3, R_4) be a (D_1, D_2, D_Y)-achievable rate vector. For any $\epsilon > 0$, we choose sufficiently large n and dimension n encoded functions (f_1, f_2, g_1, g_2), with rate (R_1, R_2, R_3, R_4) and with distortion no greater than $(D_1+\epsilon, D_2+\epsilon, D_Y+\epsilon)$. Let $\widehat{X}_1^n = (\widehat{X}_{1,i})_{i=1}^n$, $\widehat{X}_2^n = (\widehat{X}_{2,i})_{i=1}^n$, and $\widehat{Y}^n = (\widehat{Y}_i)_{i=1}^n$ be the corresponding reproductions. For $i \in \{1, \ldots, n\}$ and for $j \in \{1, 2\}$, define $D_{j,i} := E(d(X_{j,i}, \widehat{X}_{j,i}))$ and $D_{Y,i} := E(d(Y_j, \widehat{Y}))$. Let $F_1 = f_1(X_1^n, X_2^n, Y^n)$, $F_2 = f_2(X_1^n, X_2^n, Y^n)$,

$G_1 = g_1(F_1)$, and $G_2 = g_2(F_2)$ denote the random variables representing the encoding messages. By assumption,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} E(d(X_{1,i},\widehat{X}_{1,i}) &= \frac{1}{n}\sum_{i=1}^{n} D_{1,i} \le D_1 + \epsilon \\
\frac{1}{n}\sum_{i=1}^{n} E(d(X_{2,i},\widehat{X}_{2,i}) &= \frac{1}{n}\sum_{i=1}^{n} D_{2,i} \le D_2 + \epsilon \qquad (5.8)\\
\frac{1}{n}\sum_{i=1}^{n} E(d(Y_i,\widehat{Y}_i)) &= \frac{1}{n}\sum_{i=1}^{n} D_{Y,i} \le D_Y + \epsilon.
\end{aligned}$$

For each $i \in \{1,2,\ldots,n\}$, define $U_{1,i} := ((X_{1,j},X_{2,j},Y_j)_{j=1}^{i-1}, G_1)$ and $U_{2,i} := ((X_{1,j},X_{2,j},Y_j)_{j=1}^{i-1}, G_2)$. Then

$$\begin{aligned}
H(G_1) &\ge I(G_1; X_1^n, X_2^n, Y^n) \\
&= H(X_1^n, X_2^n, Y^n) - H(X_1^n, X_2^n, Y^n|G_1) \\
&= \sum_{i=1}^{n} \big[H(X_{1,i}, X_{2,i}, Y_i|(X_{1,j}, X_{2,j}, Y_j)_{j=1}^{i-1}) \\
&\quad -H(X_{1,i}, X_{2,i}, Y_i|(X_{1,j}, X_{2,j}, Y_j)_{j=1}^{i-1}, G_1)\big] \\
&= \sum_{i=1}^{n} [H(X_{1,i}, X_{2,i}, Y_i) - H(X_{1,i}, X_{2,i}, Y_i|U_{1,i})] \\
&= \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}, Y_i; U_{1,i}) \qquad (5.9)
\end{aligned}$$

and similarly,

$$H(G_2) \ge \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}, Y_i; U_{2,i}).$$

The conditional entropy $H(\widehat{X}_1^n|G_1)$ satisfies

$$
\begin{aligned}
H(\widehat{X}_1^n|G_1) &\geq I(X_1^n;\widehat{X}_1^n|G_1) \\
&= H(X_1^n|G_1) - H(X_1^n|\widehat{X}_1^n,G_1) \\
&= \sum_{i=1}^{n}\left[H(X_{1,i}|(X_{1,j})_{j=1}^{i-1},G_1)\right. \qquad (5.10)\\
&\quad \left.-H(X_{1,i}|(X_{1,j})_{j=1}^{i-1},\widehat{X}_1^n,G_1)\right] \\
&\geq \sum_{i=1}^{n}\left[H(X_{1,i}|V_{1,i}) - H(X_{1,i}|\widehat{X}_{1,i},V_{1,i})\right] \\
&\geq \sum_{i=1}^{n} I(X_{1,i};\widehat{X}_{1,i}|V_{1,i}) \geq \sum_{i=1}^{n} R_{X_{1,i}|V_{1,i}(D_{1,i})} \\
&\geq \sum_{i=1}^{n} R_{X_{1,i}|U_{1,i}}(D_{1,i}), \qquad (5.11)
\end{aligned}
$$

where $V_{1,i} = ((X_{1,j})_{j=1}^{i-1},G_1)$ and (5.11) follows from (5.3) since $U_{1,i} = (V_{1,i},(X_{2,j},Y_j)_{j=1}^{i-1})$. Now since $\widehat{X}_1^n$ and G_1 are deterministic functions of F_1, $H(\widehat{X}_1^n,G_1|F_1) = 0$ and

$$
\begin{aligned}
nR_1 &\geq H(F_1) = H(F_1) + H(\widehat{X}_1^n,G_1|F_1) \qquad (5.12)\\
&= H(\widehat{X}_1^n,F_1,G_1) \geq H(\widehat{X}_1^n,G_1) \\
&= H(\widehat{X}_1^n|G_1) + H(G_1) \\
&\geq \sum_{i=1}^{n}(R_{X_{1,i}|U_{1,i}}(D_{1,i}) + I(X_{1,i},X_{2,i},Y_i;U_{1,i})) \\
nR_3 &\geq H(G_1) \geq \sum_{i=1}^{n} I(X_{1,i},X_{2,i},Y_i;U_{1,i}). \qquad (5.13)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
R_2 &\geq \frac{1}{n}\sum_{i=1}^{n}(R_{X_{2,i}|U_{2,i}}(D_{2,i}) + I(X_{1,i},X_{2,i},Y_i;U_{2,i}|U_{1,i})) \\
R_4 &\geq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i},X_{2,i},Y_i;U_{2,i}).
\end{aligned} \qquad (5.14)
$$

Let Q denote a random variable uniformly distributed on $\{1, 2, \ldots, n\}$ that is independent of (X_1^n, X_2^n, Y^n). From Lemma 5.2.1, for $j = 1$ or 2,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} R_{X_{j,i}|U_{j,i}}(D_{j,i}) &\geq R_{X_{j,Q},Q|U_{j,Q},Q}\left(\frac{1}{n}\sum_{i=1}^{n} D_{j,i}\right) \\ &\geq R_{X_{j,Q}|U_{j,Q},Q}(D_j + \epsilon), \end{aligned}$$

where the last inequality follows from (5.8) and the fact that $R_{X_{j,Q}|U_{j,Q},Q}(D)$ is a nonincreasing function of D. Then (5.13) and (5.14) become

$$\begin{aligned} &R_1 \geq R_{X_{1,Q}|U_{1,Q},Q}(D_1 + \epsilon) + I(X_{1,Q}, X_{1,Q}, Y_Q; U_{1,Q}|Q) \\ &R_2 \geq R_{X_{2,Q}|U_{2,Q},Q}(D_2 + \epsilon) + I(X_{1,Q}, X_{2,Q}, Y_Q; U_{2,Q}|Q) \\ &R_3 \geq I(X_{1,Q}, X_{2,Q}, Y_Q; U_{1,Q}|Q) \\ &R_4 \geq I(X_{1,Q}, X_{2,Q}, Y_Q; U_{2,Q}|Q). \end{aligned}$$

Define $U_1 := (U_{1,Q}, Q)$ and $U_2 := (U_{2,Q}, Q)$. Since $(X_{1,i}, X_{2,i}, Y_i)$, $i \in \{1, 2, \ldots, n\}$ is drawn i.i.d., the joint distribution of $(X_{1,Q}, X_{2,Q}, Y_Q)$ is the same as that of (X_1, X_2, Y). Thus for $j = 1$ or 2, if $Z_j = (X_{1,Q}, X_{2,Q}, Y_Q)$, since Z_j is independent of Q,

$$\begin{aligned} &I(Z_Q; U_{j,Q}|Q) = H(Z_Q|Q) - H(Z_Q|U_{j,Q}, Q) \\ &\quad = \quad (H(Z_Q) - H(Z_Q|U_{j,Q}, Q)) - (H(Z_Q) - H(Z_Q|Q)) \\ &\quad = \quad I(Z_Q; U_{j,Q}, Q) - I(Z_Q; Q) \\ &\quad = \quad I(Z_Q; U_{j,Q}, Q) = I(Z_Q; U_j). \end{aligned}$$

Let $X_1 = X_{1,Q}$, $X_2 = X_{2,Q}$, $Y = Y_Q$. Then X_1, X_2, and Y have the same joint distribution $p(x_1, x_2, y)$. Therefore,

$$R_1 \geq R_{X_1|U_1}(D_1 + \epsilon) + I(X_1, X_2, Y; U_1)$$
$$R_2 \geq R_{X_2|U_2}(D_2 + \epsilon) + I(X_1, X_2, Y; U_2)$$
$$R_3 \geq I(X_1, X_2, Y; U_1), R_4 \geq I(X_1, X_2, Y; U_2).$$

□

5.4.2 Inner Bounds

We propose two inner bounds in Theorems 5.4.2 and 5.4.4. To prove the first achievability result, we apply the coding scheme introduced in the proof of Theorem 5.3.1. We use four auxiliary random variables to characterize this inner bound.

Theorem 5.4.2 *(Inner bound 1) Let the sources X_1, X_2, and Y be finite-alphabet discrete random variables. Given $D_1 \geq 0$, $D_2 \geq 0$ and $D_Y \geq 0$. If the rate vector (R_1, R_2, R_3, R_4) satisfies*

$$\begin{aligned}
R_1 >& R_{X_1|U_1}(D_1) + I(X_1, X_2, Y; U_1) \\
&+ I(X_1, X_2, Y; V_1|U_1, U_2) \\
R_2 >& R_{X_2|U_2}(D_2) + I(X_1, X_2, Y; U_2) \\
&+ I(X_1, X_2, Y; V_2|U_1, U_2) \\
R_3 >& I(X_1, X_2, Y; U_1) + I(X_1, X_2, Y; V_1|U_1, U_2) \\
R_4 >& I(X_1, X_2, Y; U_2) + I(X_1, X_2, Y; V_2|U_1, U_2)
\end{aligned} \tag{5.15}$$

for some auxiliary random variables U_1, U_2, V_1, and V_2 such that

(i) $(U_1, V_1) \to (X_1, X_2, Y) \to (U_2, V_2)$ *forms a Markov chain*

(ii) there exists a function $\widehat{Y}$ of U_1, U_2, V_1, and V_2 such that

$$Ed(\widehat{Y}(U_1, U_2, V_1, V_2), Y) \leq D_Y,$$

then (R_1, R_2, R_3, R_4) *is* (D_1, D_2, D_Y)*-achievable.*

Proof. To prove the achievability of this region, it suffices to show that for any $\delta > 0$ and any (U_1, U_2, V_1, V_2) satisfying conditions (i) - (ii) above, the rate vector (R_1, R_2, R_3, R_4) defined by

$$\begin{aligned}
R_1(\delta) &= R_{X_1|U_1}(D_1) + I(X_1, X_2, Y; U_1) \\
&\quad + I(X_1, X_2, Y; V_1|U_1, U_2) + 3\delta \\
R_2(\delta) &= R_{X_2|U_2}(D_2) + I(X_1, X_2, Y; U_2) \\
&\quad + I(X_1, X_2, Y; V_2|U_1, U_2) + 3\delta \\
R_3(\delta) &= I(X_1, X_2, Y; U_1) + I(X_1, X_2, Y; V_1|U_1, U_2) \\
&\quad + 2\delta \\
R_4(\delta) &= I(X_1, X_2, Y; U_2) + I(X_1, X_2, Y; V_2|U_1, U_2) \\
&\quad + 2\delta
\end{aligned}$$

is (D_1, D_2, D_Y)-achievable.

Fix $n \in \mathbb{N}$. For $k = 1, 2$, let $\widehat{X}_k$ be a random variable defined by a conditional probability distribution $p(\widehat{x}_k|x_k, u_k)$ such that $I(X_k; \widehat{X}_k|U_k) < R_{X_k|U_k}(D_k) + \delta$ and $E(d(X_k, \widehat{X}_k)) \leq D_k$. Set

$$\begin{aligned}
S_k &:= 2^{n(I(X_1,X_2,Y;V_k|U_k)+\delta)}, \ M_k := 2^{n(I(X_1,X_2,Y;U_k)+\delta)} \\
T_k &:= 2^{n(I(X_k;\widehat{X}_k|U_k)+\delta)}, \ N_k := 2^{n(I(X_1,X_2,Y;V_k|U_1,U_2)+\delta)}.
\end{aligned}$$

Generate the codebook for $k = 1, 2$ as follows:

1. Randomly choose M_k typical sequences $U_k^n(1)$, $U_k^n(2)$, …, $U_k^n(M_k)$ i.i.d. according to the probability distribution $\prod_{i=1}^n p(u_{k,i})$.

2. For each $m \in \{1, 2, \ldots, M_k\}$, choose S_k sequences $V_k^n(m, 1)$, …, $V_k^n(m, S_k)$ uniformly at random from the set $A_\epsilon^{*(n)}(V|U_k^n(m))$.

3. For each $m \in \{1, 2, \ldots, M_k\}$, choose T_k sequences $\widehat{X}_k^n(m, 1)$, $\widehat{X}_k^n(m, 2)$, $\ldots$, $\widehat{X}_k^n(m, T_k)$ uniformly at random from the set $A_\epsilon^{*(n)}(X|U_k^n(m))$.

For $k = 1, 2$ and $j \in \{1, \ldots, S_k\}$, draw $\tau_k(j)$ uniformly at random from $\{1, \ldots, N_k\}$. Let (x_1^n, x_2^n, y^n) be the source vector to be transmitted. We use the following functions in defining the encoders :

$$
\begin{aligned}
\psi_k(x_1^n, x_2^n, y^n) &:= \min\Big[\{M_k\} \cup \{m \in \{1, \ldots, M_k\} : \\
&\quad (x_1^n, x_2^n, y^n, U_k^n(m)) \in A_\epsilon^{*(n)}(X_1, X_2, Y, U_k)\}\Big] . \\
\mu_k(x_1^n, x_2^n, y^n) &:= \min\big[\{S_k\} \cup \{s \in \{1, \ldots, S_k\} : \\
&\quad (x_1^n, x_2^n, y^n, U_k^n(\psi_k(x_1^n, x_2^n, y^n)), V_k^n(\psi_k(x_1^n, x_2^n, y^n), s)) \\
&\quad \in A_\epsilon^{*(n)}(X_1, X_2, Y, U_k, V_k)\}\big] \\
\phi_k(x_1^n, x_2^n, y^n) &:= \min\big[\{T_k\} \cup \{t \in \{1, 2, \ldots, T_k\} : \\
&\quad (x_k^n, U_k^n(\psi_k(x_1^n, x_2^n, y^n)), \widehat{X}_k^n(\psi_k(x_1^n, x_2^n, y^n), t)) \\
&\quad \in A_\epsilon^{*(n)}(X_k, U_k, \widehat{X}_k)\}\Big] \\
\pi_k(x_1^n, x_2^n, y^n) &:= \tau_k(\mu_k(x_1^n, x_2^n, y^n)).
\end{aligned}
$$

Finally, we define the encoders as

$$
\begin{aligned}
f_k(x_1^n, x_2^n, y^n) &:= (\psi_k, \pi_k, \phi_k)(x_1^n, x_2^n, y^n) \\
g_k(f_k(x_1^n, x_2^n, y^n)) &:= (\psi_k, \pi_k)(x_1^n, x_2^n, y^n).
\end{aligned}
$$

The decoding strategy is stated as follows:

1. Reproducing X_k^n for $k = 1, 2$: The decoder at node k maps the indices (m_k, p_k, t_k) to the reproduction $\widehat{X}_k^n(m_k, t_k)$.

2. Reproducing Y^n: At node 3, find $\widehat{j}_k$ $(k = 1, 2)$ from $\{1, \ldots, S_k\}$ such that

$$
\begin{aligned}
&(U_1^n(m_1), U_2^n(m_2), V_k^n(m_k, \widehat{j}_k)) \in \\
&\quad A_\epsilon^{*(n)}(U_1, U_2, V_k) \text{ and } \tau_k(\widehat{j}_k) = p_k.
\end{aligned} \tag{5.16}
$$

For $k = 1, 2$, define the function ι_k mapping from $\{1, \ldots, M_k\} \times \{1, \ldots, N_k\}$ to $\{1, \ldots, S_k\}$ as

$$\iota_k(m_k, p_k) := \begin{cases} \widehat{j}_k, & \text{if } \widehat{j}_k \text{ is the unique index satisfying (5.16)} \\ 1, & \text{otherwise.} \end{cases}$$

The receiver at node 3 builds its reconstruction for Y^n as

$$\begin{aligned} &\widehat{Y}^n(U_1^n(m_1), U_2^n(m_2), V_1^n(m_1, \iota_1(m_1, p_1)), \\ &V_2^n(m_2, \iota_2(m_2, p_2))). \end{aligned}$$

Analysis of performance:

For simplicity, we denote by μ_k, ϕ_k, ι_k, and π_k the evaluated values of the corresponding functions on (x_1^n, x_2^n, y^n) for $k = 1, 2$. Define the following atypicality events:

1. $E_0 : (x_1^n, x_2^n, y^n) \notin A_\epsilon^{*(n)}(X_1, X_2, Y)$.

2. E_{U_k} for $k = 1, 2 :$ For all $m \in \{1, \ldots, M_k\}$ and all $s \in \{1, \ldots, S_k\}$, $(x_1^n, x_2^n, y^n, U_k^n(m), V_k^n(m, s))$ is not jointly typical in $A_\epsilon^{*(n)}(X_1, X_2, Y, U_k, V_k)$.

3. E_{X_k} for $k = 1, 2 :$ For all $t \in \{1, 2, \ldots, T_k\}$,

$$(x_k^n, U_k^n(\psi_k), \widehat{X}_k^n(\psi_k, t)) \notin A_\epsilon^{*(n)}(X_k, U_k, \widehat{X}_k).$$

4. $E_{V_k} : (U_1^n(\psi_1), U_2^n(\psi_2)) \in A_\epsilon^{*(n)}(U_1, U_2)$ and there exists $V_k^n(\psi_k, j) \neq V_k^n(\psi_k, \mu_k)$ such that

$$(U_1^n(\psi_1), U_2^n(\psi_2), V_k^n(\psi_k, j)) \in A_\epsilon^{*(n)}(U_1, U_2, V_k)$$

   and $\tau_k(j) = \pi_k$.

5. $E_{12} : (x_1^n, x_2^n, y^n, U_1^n(\psi_1), V_1^n(\psi_1, \mu_1), V_2^n(\psi_2, \mu_2)) \notin A_\epsilon^{*(n)}$ or

$(x_1^n, x_2^n, y^n, U_2^n(\psi_1), V_1^n(\psi_1, \mu_1), V_2^n(\psi_2, \mu_2)) \notin A_\epsilon^{*(n)}$.

6. $E = E_0 \cup E_{U_1} \cup E_{U_2} \cup E_{X_1} \cup E_{X_2} \cup E_{V_1} \cup E_{V_2} \cup E_{12}$.

From the basic property of typical sets, $\Pr(E_0)$ can be made arbitrarily small as n grows without bound. By Corollary 5.2.5 and Lemma 5.2.6, $\Pr(E_{U_1})$, $\Pr(E_{U_2})$, $\Pr(E_{X_1})$, $\Pr(E_{X_2})$, and $\Pr(E_{12})$ can also be made arbitrarily small as n grows without bound. Finally, for $k = 1, 2$, when knowing $(U_1^n(\psi_1), U_2^n(\psi_2)) \in A_\epsilon^{*(n)}(U_1, U_2)$, since the probability that a randomly chosen sequence $v_k^n \in A_\epsilon^{*(n)}(V|U_k^n(\psi_k))$ is jointly strongly typical with $(U_1^n(\psi_1), U_2^n(\psi_2))$ is approximately $2^{nI(V_k;U_{3-k}|U_k)}$, the probability of the event E_{V_k} can be bounded (when n is large enough) by

$$\begin{aligned}\Pr(E_{V_k}) &\leq |B_k(p_k)| 2^{n(I(V_k;U_{3-k}|U_k)-\frac{\delta}{2})} \\ &= \frac{S_k}{N_k} \times 2^{n(I(V_k;U_{3-k}|U_k)-\frac{\delta}{2})},\end{aligned}$$

where $B_k(p_k)$ is the set of indices $j \in \{1, \dots, S_k\}$ such that $\tau(j) = p_k$. Since

$$\begin{aligned}&I(X_1, X_2, Y; V_k|U_k) - I(X_1, X_2, Y; V_k|U_k, U_{3-k}) \\ &= (H(V_k|U_k) - H(V_k|U_k, X_1, X_2, Y)) \\ &\quad -(H(V_k|U_1, U_2) - H(V_k|X_1, X_2, Y, U_1, U_2)) \\ &= H(V_k|U_k) - H(V_k|U_1, U_2) \\ &\quad +(H(V_k|X_1, X_2, Y, U_1, U_2) - H(V_k|U_k, X_1, X_2, Y)) \\ &= I(V_k; U_{3-k}|U_k) + 0,\end{aligned}$$

$\Pr(E_{V_k})$ can also be made arbitrarily small as n grows without bound. Hence $\lim_{n\to\infty} \Pr(E) = 0$. From Remark 5.2.3, the average distortion between any pair of jointly typical sequences is close to the expected distortion. Therefore, for $k = 1, 2$,

$$
\begin{aligned}
&|Ed(X_k^n, \widehat{X}_k^n(\psi_k, \phi_k)) - D_k| \\
&= \Big| \Pr(E^c) \Big[E \Big(d(X_k^n, \widehat{X}_k^n(\psi_k, \phi_k)) | E^c \Big) - D_k \Big] \\
&\quad + \Pr(E) \Big[E \Big(d(X_k^n, \widehat{X}_k^n(\psi_k, \phi_k)) | E \Big) - D_k \Big] \Big| \\
&\leq \Pr(E^c) \cdot \epsilon \cdot d_{\max} + \Pr(E) \cdot d_{\max} \leq \epsilon \cdot d_{\max} + \epsilon \cdot d_{\max}
\end{aligned}
$$

for n sufficiently large. Similarly,

$$
\begin{aligned}
&|Ed(Y^n, \widehat{Y}^n(U_1^n(\psi_1), U_2^n(\psi_2), V_1^n(\psi_1, \iota_1(\psi_1, \pi_1)) \\
&, V_2^n(\psi_2, \iota_2(\psi_2, \pi_2)))| < 2\epsilon \cdot d_{\max}
\end{aligned}
$$

for n sufficiently large. Since $\epsilon > 0$ can be chosen arbitrarily small, this coding scheme satisfies the distortion requirement (D_1, D_2, D_Y).

$\square$

By setting V_1 and V_2 to be constants, we have the following.

Corollary 5.4.3 *Let the sources X_1, X_2, and Y be finite-alphabet discrete random variables. Given $D_1 \geq 0$, $D_2 \geq 0$ and $D_Y \geq 0$. If the rate vector (R_1, R_2, R_3, R_4) satisfies*

$$
\begin{aligned}
&R_1 > R_{X_1|U_1}(D_1) + I(X_1, X_2, Y; U_1) \\
&R_2 > R_{X_2|U_2}(D_2) + I(X_1, X_2, Y; U_2) \\
&R_3 > I(X_1, X_2, Y; U_1), R_4 > I(X_1, X_2, Y; U_2)
\end{aligned}
$$

for some auxiliary random variables U_1 and U_2 such that

(i) $U_1 \to (X_1, X_2, Y) \to U_2$ forms a Markov chain

(ii) there exists a function $\widehat{Y}$ of U_1, U_2 such that

$$Ed(\widehat{Y}, Y) \leq D_Y,$$

then (R_1, R_2, R_3, R_4) is (D_1, D_2, D_Y)-achievable.

The second inner bound is stated in the following theorem. As in Corollary 5.4.3, we use two auxiliary random variables. Instead of assuming Markov conditions, we here use additional rate to guarantee that the sequences we pick in the encoding process associated with these two auxiliary random variables are jointly strongly typical with the source sequence.

Theorem 5.4.4 *(Inner bound 2) Let the sources X_1, X_2, Y be finite-alphabet discrete random variables. Given $D_1 \geq 0$, $D_2 \geq 0$ and $D_Y \geq 0$. If the rate vector (R_1, R_2, R_3, R_4) satisfies*

$$\begin{aligned}
&R_1 > R_{X_1|U_1}(D_1) + I(X_1, X_2, Y; U_1) \\
&R_2 > R_{X_2|U_2}(D_2) + I(X_1, X_2, Y; U_2) \\
&R_3 > I(X_1, X_2, Y; U_1) \\
&R_4 > I(X_1, X_2, Y; U_2) \\
&R_1 + R_4 > R_{X_1|U_1}(D_1) + I(X_1, X_2, Y; U_1, U_2) \\
&\qquad +I(U_1; U_2) \\
&R_2 + R_3 > R_{X_2|U_2}(D_2) + I(X_1, X_2, Y; U_1, U_2) \\
&\qquad +I(U_1; U_2) \\
&R_3 + R_4 > I(X_1, X_2, Y; U_1, U_2) + I(U_1; U_2)
\end{aligned} \tag{5.17}$$

for some auxiliary random variables U_1 and U_2 such that there exists a function $\widehat{Y}$ of U_1 and U_2 with

$$Ed(\widehat{Y}(U_1, U_2), Y) \leq D_Y,$$

then (R_1, R_2, R_3, R_4) is (D_1, D_2, D_Y)-achievable.

Proof. It is enough to show that for any $\delta > 0$ and $\lambda \in [0,1]$, the rate vector $(R_1'(\delta)$, $R_2'(\delta)$, $R_3'(\delta)$, $R_4'(\delta))$ is (D_1, D_2, D_Y)-achievable, where

$$\begin{aligned}
R_1'(\delta) &= R_{X_1|U_1}(D_1) + I(X_1, X_2, Y; U_1) + \lambda\Delta + 2\delta \\
R_2'(\delta) &= R_{X_2|U_2}(D_2) + I(X_1, X_2, Y; U_2) + (1-\lambda)\Delta + 2\delta \\
R_3'(\delta) &= I(X_1, X_2, Y; U_1) + \lambda\Delta + \delta \\
R_4'(\delta) &= I(X_1, X_2, Y; U_2) + (1-\lambda)\Delta + \delta
\end{aligned}$$

and

$$\begin{aligned}
\Delta &= \max(I(X_1, X_2, Y; U_1, U_2) + I(U_1; U_2) \\
&\quad -I(X_1, X_2, Y; U_1) - I(X_1, X_2, Y; U_2) + \delta, 0).
\end{aligned}$$

For $k = 1, 2$, let $\widehat{X}_k$ be a random variable defined by a conditional probability distribution $p(\widehat{x}_k | x_k, u_k)$ such that $I(X_k; \widehat{X}_k | U_k) < R_{X_k|U_k}(D_k) + \delta$ and $E(d(X_k, \widehat{X}_k)) \leq D_k$. Set

$$\begin{aligned}
M_1 &:= 2^{n(I(X_1,X_2,Y;U_1)+\lambda\Delta+\delta)} \\
M_2 &:= 2^{n(I(X_1,X_2,Y;U_2)+(1-\lambda)\Delta+\delta)} \\
T_1 &:= 2^{n(I(X_1;\widehat{X}_1|U_1)+\delta)}, T_2 := 2^{n(I(X_2;\widehat{X}_2|U_2)+\delta)}.
\end{aligned}$$

Generate the codebook for $k = 1, 2$ as follows:

1. Randomly choose M_k typical sequences $U_k^n(1)$, $U_k^n(2)$, $\ldots$, $U_k^n(M_k)$ according to the probability distribution $\prod_{i=1}^n p(u_{k,i})$.

2. For each $m \in \{1, 2, \ldots, M_k\}$, randomly choose T_k sequences $\widehat{X}_k^n(m, 1)$, $\widehat{X}_k^n(m, 2)$, $\ldots$, $\widehat{X}_k^n(m, T_k)$ uniformly at random from the set $A_\epsilon^{*(n)}(X | U_k^n(m))$.

Let (x_1^n, x_2^n, y^n) be the source to be transmitted. We use the following functions in defining the encoders:

$$\psi_k(x_1^n, x_2^n, y^n) := \min \Big[\{M_k\} \cup \{m \in \{1, \ldots, M_k\} : (x_1^n, x_2^n, y^n, U_k^n(m)) \in A_\epsilon^{*(n)}(X_1, X_2, Y, U_k)\} \Big].$$

$$\phi_k(x_1^n, x_2^n, y^n) := \min \Big[\{T_k\} \cup \{t \in \{1, 2, \ldots, T_k\} : (x_k^n, U_k^n(\psi_k(x_1^n, x_2^n, y^n)), \widehat{X}_k^n(\psi_k(x_1^n, x_2^n, y^n), t)) \in A_\epsilon^{*(n)}(X_k, U_k, \widehat{X}_k)\} \Big].$$

Finally, we define the encoders as

$$\begin{aligned} f_k(x_1^n, x_2^n, y^n) &:= (\phi_k, \psi_k)(x_1^n, x_2^n, y^n) \\ g_k(f_k(x_1^n, x_2^n, y^n)) &:= \psi_k(x_1^n, x_2^n, y^n). \end{aligned}$$

The decoder at node k ($k = 1, 2$) maps the indices (t_k, m_k) to the reproduction $\widehat{X}_k^n(m_k, t_k)$. At node 3, the decoder maps the indices (m_1, m_2) to the reproduction

$$\widehat{Y}^n(U_1^n(m_1), U_2^n(m_2)).$$

Analysis of performance:

For simplicity, we use μ_k, ϕ_k to denote the evaluated values of the corresponding functions on (x_1^n, x_2^n, y^n) for $k = 1, 2$. Define the following error events:

1. $E_0 : (x_1^n, x_2^n, y^n) \notin A_\epsilon^{*(n)}(X_1, X_2, Y)$.

2. E_{U_k} for $k = 1, 2 :$ For all $m \in \{1, \ldots, M_k\}$,

$$(x_1^n, x_2^n, y^n, U_k^n(m)) \notin A_\epsilon^{*(n)}(X_1, X_2, Y, U_k).$$

3. E_{X_k} for $k = 1, 2 :$ For all $t \in \{1, 2, \ldots, T_k\}$,

$$(x_k^n, U_k^n(\psi_k), \widehat{X}_k^n(\psi_k, t)) \notin A_\epsilon^{*(n)}(X_k, U_k, \widehat{X}_k).$$

4. $E_{12}: (x_1^n, x_2^n, y^n, U_1^n(\psi_1), U_2^n(\psi_2)) \notin A_\epsilon^{*(n)}(X_1, X_2, Y, U_1, U_2)$.

5. $E_{\text{error}} = E_0 \cup E_{U_1} \cup E_{U_2} \cup E_{X_1} \cup E_{X_2} \cup E_{12}$.

The same argument as in Theorem 5.4.2 shows that $\Pr(E_0)$, $\Pr(E_{U_1})$, $\Pr(E_{U_2})$, $\Pr(E_{X_1})$, $\Pr(E_{X_2})$, and $\Pr(E_{12})$ can be made arbitrarily small as n grows without bound. Since we have the following inequality

$$
\begin{aligned}
&I(X_1, X_2, Y; U_1) + I(X_1, X_2, Y; U_2) + 2\delta + \Delta \\
&> I(X_1, X_2, Y; U_1, U_2) + I(U_1; U_2),
\end{aligned}
$$

the same argument as in [53] shows that $\lim_{n\to\infty} \Pr(E_{12}) = 0$. Consequently, we have $\lim_{n\to\infty} \Pr(E_{\text{error}}) = 0$. From Remark 5.2.3, the average distortion between any pair of jointly typical sequences is close to the expected distortion. Therefore, for $k = 1, 2$,

$$
\begin{aligned}
&|Ed(X_k^n, \widehat{X}_k^n(\psi_k, \phi_k)) - D_k| \\
&= \Big|\Pr(E_{\text{error}}^c)\left[E\left(d(X_k^n, \widehat{X}_k^n(\psi_k, \phi_k)|E_{\text{error}}^c\right) - D_k\right] \\
&\quad + \Pr(E_{\text{error}})\left[E\left(d(X_k^n, \widehat{X}_k^n(\psi_k, \phi_k)|E_{\text{error}}\right) - D_k\right]\Big| \\
&\le \Pr(E_{\text{error}}^c) \cdot \epsilon \cdot d_{\max} + \Pr(E_{\text{error}}) \cdot d_{\max} \\
&\le \epsilon \cdot d_{\max} + \Pr(E_{\text{error}}) \cdot d_{\max}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&|Ed(Y^n, \widehat{Y}^n(U_1^n(\psi_1), U_2^n(\psi_2))) - D_Y| \\
&< \epsilon \cdot d_{\max} + \Pr(E_{\text{error}}) \cdot d_{\max}.
\end{aligned}
$$

Since $\epsilon > 0$ is chosen arbitrarily and $\Pr(E)$ can be made arbitrarily small for n sufficiently large, this coding scheme matches the distortion requirement (D_1, D_2, D_Y).

$\square$

Remark 5.4.5 *1. The region given in Theorem 5.4.4 may not be convex, so the inner bound derived in Theorem 5.4.4 is actually the convex closure of the set described in (5.17).*

5.5 Special Cases

Two special cases of this diamond network are discussed in this section. We have one-letter characterizations of the corresponding rate-distortion regions that are induced from the inner bounds in Section 5.4. We investigate the roles of the auxiliary random variables used in describing their rate-distortion regions.

5.5.1 Special Case I

When sources X_1 and X_2 are constant ($X_1 = C_1$ and $X_2 = C_2$ with probability one) and $R_1 = R_3$ and $R_2 = R_4$, the diamond network becomes a point-to-point network with two paths between the transmitter and the receiver, as shown in Figure 5.5. Clearly, the rate distortion region $\mathcal{R}_1(D)$ in this case is the set of vectors $(R_1,R_2) \in \mathbb{R}_+^2$ satisfying $R_1 + R_2 \geq R_X(D)$, where

$$R_X(D) = \min_{\widehat{X}: Ed(X,\widehat{X}) \leq D} I(X;\widehat{X})$$

is the rate distortion function. Although the achievable rate pairs for this trivial case are easily described, we here use the above characterization to show that our previous solutions, based on auxiliary random variables U_1 and U_2, give the optimal solution in this special case.

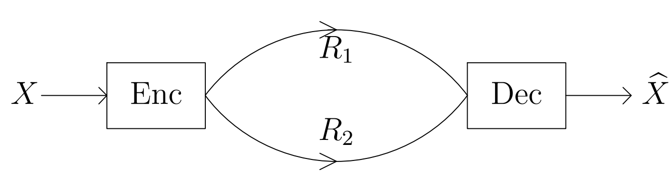


Figure 5.5: The two-path point-to-point network.

Given $D \geq 0$, define two sets of pairs of auxiliary random variables

$$\Phi_1(D) := \left\{ (U_1, U_2) \;\middle|\; \begin{array}{l} U_1 \to X \to U_2 \text{ is a Markov} \\ \text{chain and } \exists\ \widehat{X} : \mathcal{U}_1 \times \mathcal{U}_2 \\ \to \widehat{\mathcal{X}} \text{ such that} \\ Ed(X, \widehat{X}) \leq D. \end{array} \right\}$$

and

$$\Phi_2(D) := \left\{ (U_1, U_2) \;\middle|\; \begin{array}{l} \exists\ \widehat{X} : \mathcal{U}_1 \times \mathcal{U}_2 \to \widehat{\mathcal{X}} \\ \text{such that } Ed(X, \widehat{X}) \leq D. \end{array} \right\}.$$

Consider the following two sets of rate pairs

$$\mathcal{R}_1^*(D) := \overline{\bigcup_{(U_1,U_2)\in\Phi_1(D)} \left\{ (R_1, R_2) \;\middle|\; \begin{array}{l} R_1 \geq I(X; U_1) \\ R_2 \geq I(X; U_2) \end{array} \right\}}$$

$$\mathcal{R}_1^{**}(D) := \bigcup_{(U_1,U_2)\in\Phi_2(D)} \left\{ (R_1, R_2) \;\middle|\; \begin{array}{l} R_1 \geq I(X; U_1), \\ R_2 \geq I(X; U_2) \\ R_1 + R_2 \geq \\ I(X; U_1, U_2) \\ + I(U_1; U_2) \end{array} \right\},$$

where for any set $A \subseteq \mathbb{R}^4$, $\overline{A}$ denotes convex closure of A. Note that for any $(U_1, U_2) \in \Phi_1(D)$,

$$\begin{aligned} & I(X; U_1) + I(X; U_2) \\ & \quad = (H(U_1) + H(U_2)) - (H(U_1|X) + H(U_2|X)) \\ & \quad \geq H(U_1, U_2) - H(U_1, U_2|X) = I(X; U_1, U_2) \\ & \quad \geq I(X; \widehat{X}(U_1, U_2)) \geq R_X(D). \end{aligned}$$

Hence $\mathcal{R}_1^*(D) \subseteq \mathcal{R}_1(D)$. To see the converse, given any $\epsilon > 0$, choose $\widehat{X}$ satisfying

$$I(X; \widehat{X}) \leq R_X(D) + \epsilon,\ \ Ed(X, \widehat{X}) \leq D.$$

Then by letting $(U_1, U_2) = (\widehat{X}, c)$, one has $(R_X(D) + \epsilon, 0) \in \mathcal{R}_1(D)$, where c is a constant. Similarly, $(0, R_X(D) + \epsilon) \in \mathcal{R}_1(D)$. Then by convexity of $\mathcal{R}_1(D)$, $\mathcal{R}_1(D) \subseteq \mathcal{R}_1^*(D)$.

Lemma 5.5.1 $\mathcal{R}_1(D) = \mathcal{R}_1^*(D)$.

Since any rate pair in $\mathcal{R}_1(D)$ can be obtained using time-sharing, it is sufficient to pick U_1 and U_2 of alphabet sizes $|\mathcal{X}| + 1$.

Next we look at $\mathcal{R}_1^{**}(D)$. By Theorem 5.4.4, $\mathcal{R}_1^{**}(D) \subseteq \mathcal{R}_1(D)$. Following the same argument as above, we have that $(0, R_X(D))$ and $(R_X(D), 0)$ are in $\mathcal{R}_1^{**}(D)$, giving the following lemma.

Lemma 5.5.2 $\mathcal{R}_1(D) = \mathcal{R}_1^{**}(D)$.

5.5.2 Special Case II

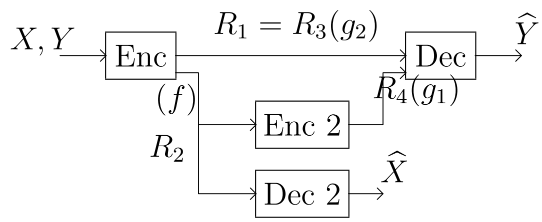


Figure 5.6: Special case II.

As a second special case, let $X_1 = C_1$ with probability 1 and set $R_1 = R_3$, but allow X_2 to be an arbitrary finite-alphabet random variable. The result is the network shown in Figure 5.6. In this network, there are one middle node at the top link and a direct link from the transmitter to the receiver. Let (D_1, D_Y) be the distortion requirements for reproducing (X, Y). Our purpose is to show that a rate vector (R_1, R_3, R_4) is in the rate distortion region if and only if

$$\begin{aligned} R_1 = R_3 &\geq R_{Y|U}(D_Y) \\ R_2 &\geq R_{X|U}(D_1) + I(X, Y; U) \\ R_4 &\geq I(X, Y; U) \end{aligned}$$

for some random variable U. To prove the converse, from Theorem 5.4.1, it suffices to check that $R_1 \geq R_{Y|U}(D_Y)$. Let f, g_1, and g_2 be the encoding functions as indicated in Figure 5.6 and let $F = f(X^n, Y^n)$, $G_1 = g_1(F)$, and $g_2 = G_2(X^n, Y^n)$ denote the encoding messages. Then, following the approach used in (5.11), we have $H(\widehat{Y}^n|G_1) \geq nR_{Y|U}(D_Y)$, where $\widehat{Y}^n$ is the reproduction of Y^n. Hence

$$\begin{aligned} nR_4 &\geq H(G_2) \geq H(G_2|G_1) = H(G_2|G_1) + H(\widehat{Y}^n|G_1, G_2) \\ &= H(\widehat{Y}^n, G_2|G_1) \geq H(\widehat{Y}^n|G_1) \\ &\geq nR_{Y|U}(D_Y). \end{aligned}$$

To show the achievability, the idea is basically the same as the proof of Theorem 5.4.2, so we briefly describe it. Let U be given and $\delta > 0$ be arbitrary. Pick random variables $\widehat{X}$ and $\widehat{Y}$ satisfying

$$\begin{aligned} &I(X; \widehat{X}|U) < R_{X|U}(D_1) + \delta,\ Ed(X, \widehat{X}) \leq D_1 \\ &I(Y; \widehat{Y}|U) < R_{Y|U}(D_Y) + \delta,\ Ed(Y, \widehat{Y}) \leq D_Y. \end{aligned}$$

Let n be given. We generate the codebook as follows:

1. Randomly generate $U^n(1)$, ..., $U^n(M)$, where $M = 2^{n(I(X,Y;U)+\delta)}$.

2. For any $i \in \{1, \ldots, M\}$, randomly generate $\widehat{X}^n(i, 1)$, ..., $\widehat{X}^n(i, T)$ from the set $A_\epsilon^{*(n)}(X|U^n(i))$, where $T = 2^{n(I(X;\widehat{X}|U)+\delta)}$.

3. For any $i \in \{1, \ldots, M\}$, randomly generate $\widehat{Y}^n(i, 1)$, ..., $\widehat{Y}^n(i, S)$ from the set $A_\epsilon^{*(n)}(Y|U^n(i))$, where $S = 2^{n(I(Y;\widehat{Y}|U)+\delta)}$.

If (x^n, y^n) is the source sequence, the transmitter finds the indices $s \in \{1, \ldots, S\}$, $t \in \{1, \ldots, t\}$, and $m \in \{1, \ldots, M\}$ such that

$$(x^n, y^n, U^n(m), \widehat{X}^n(m, t), \widehat{Y}^n(s, t)) \in A_\epsilon^{*(n)}(X, Y, U, \widehat{X}, \widehat{Y}).$$

Then he transmits s through the top link and (m, t) through the bottom link. The encoder at the middle node then transmits m to the receiver. If (m, t) is observed at the middle node, the decoder then naturally reproduces X^n as $\widehat{X}^n(m,t)$. If s is received from the top link and t is observed from the bottom at receiver, the decoder reproduces Y^n as $\widehat{Y}(s,t)$. By using this coding scheme, the same argument as in the proof of Theorem 5.4.2 implies that when n is sufficiently large, one has that with sufficiently high probability,

$$(x^n, \widehat{X}^n(m,t)) \in A_\epsilon^{*(n)}(X, \widehat{X})$$

and

$$(y^n, \widehat{Y}^n(s,t)) \in A_\epsilon^{*(n)}(Y, \widehat{Y}).$$

Then the reproductions $\widehat{X}^n(m,t)$ and $\widehat{Y}^n(s,t)$ will match the distortion requirements (D_1, D_Y).

Chapter 6

Conclusions and Future Works

We proved that the lossless rate region for canonical source coding and the lossy rate region for both canonical and non-canonical source coding are continuous in distribution in Chapter 2. Although a counterexample shows that the lossless rate region for non-canonical source coding may not be continuous in distribution, it is s-continuous, meaning that as long as two distributions are sufficiently close and have the same support, the lossless rate regions can be arbitrarily close. We also proved that the zero-error rate region is s-continuous.

The exponentially strong converse is proved for coded-side information problem, lossless source coding for multicast network with side information at the end nodes, and lossy source coding for the point-to-point communication in Chapter 3.

Chapter 4 introduces a family of algorithms to approximate the rate regions for some example network source coding problems that guarantees $(1+\epsilon)$-approximation including the lossless rate region for the coded side information problem, the Wyner-Ziv rate distortion function, and the Berger et al. bound for the lossy coded side information problem. The proposed algorithms that approximate the rate regions based on their single-letter characterizations may be improved by cleverly choosing the family of quantized conditional distributions of auxiliary random variables given the source random variables.

By applying the techniques used to solving the simple network source coding problems in the literature, we use auxiliary random variables to bound the rate regions for two basic network source coding problems that capture some important characteristics of general networks: the lined two-hop and the diamond networks. The performance bounds provide a way to bound larger networks by decomposing them into basic components.

The results in this thesis may provide a method to understand some theoretical problems in the field of source coding over networks, for instance, the existence of single-letter characterizations, tightness of some early derived bounds, and the behaviors of the achievable rate regions as functions of the source distribution and the distortion vector. Some techniques may apply to the study of the capacity regions for channel coding over networks.

Chapter 7

APPENDIX

A Lemmas for Theorem 2.4.1

The following sequence of Lemmas builds to Theorem 2.4.1, which proves that for the canonical source coding problem $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is uniformly continuous in $\mathbf{D}$. First, Lemma A.1 and Corollary A.2 bound the conditional entropy of one random vector given the other as a function of Hamming distance between them.

Lemma A.1 *Let $V^n = (V_1, V_2, \ldots, V_n)$ be a random vector in ϑ^n and let w be the per symbol expected Hamming weight of V^n*

$$w = \frac{1}{n} E d_H(V^n, \mathbf{0}) = \frac{1}{n} E|\{i \mid V_i \neq 0\}|.$$

Then

$$H(V^n) \leq n(H(w) + w \log(m^s - 1)).$$

Proof. First notice that since $V^n \in \Theta^n$, there are at most m^s possible values for V_i for every $i \in \{1, \ldots, n\}$. For every $i \in \{1, \ldots, n\}$, let $\{a_{i,0}, \ldots, a_{i,m^s-1}\}$ be the set of possible values for V_i. For each $i \in \{1, \ldots, n\}$ and each $j \in \{0, \ldots, m^s - 1\}$, let $p_{i,j}$ denote the probability $\Pr(V_i = a_{i,j})$. Then

$$w = \frac{1}{n} E d_H(V^n, \mathbf{0}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m^s-1} p_{i,j}.$$

Let $w_i := \sum_{j=1}^{m^s-1} p_{i,j}$. The maximal entropy $H(V_i)$ over all distributions with the same values $(w_1, \ldots, w_n)$ occurs when $p_{i,0} = 1 - w_i$ and $p_{i,j} = \frac{w_i}{m^s-1}$ for each $j \neq 0$. Hence

$$\begin{aligned} H(V_i) &\leq H\left(1 - w_i, \frac{w_i}{m^s - 1}, \ldots, \frac{w_i}{m^s - 1}\right) \\ &= H(w_i) + w_i \log(m^s - 1). \end{aligned}$$

Therefore, by the convexity of the entropy function,

$$\begin{aligned} H(V^n) &\leq \sum_{i=1}^{n} H(V_i) \leq \sum_{i=1}^{n} [H(w_i) + w_i \log(m^s - 1)] \\ &\leq n(H(w) + w \log(m^s - 1)). \end{aligned}$$

□

For any $V^n \in \Theta^n$, the set $\{a \in \vartheta^n \mid \Pr(V^n = a)\}$ is called the support set of V^n.

Corollary A.2 *Let V^n and W^n be two random vectors in ϑ^n with the same support set. Then*

$$\frac{1}{n} H(V^n|W^n) \leq H(\tau) + \tau \log(m^s - 1) = O(\tau \log(1/\tau)),$$

where

$$\tau := E\left[\frac{1}{n} d_H(V^n, W^n)\right].$$

Proof. Apply Lemma A.1 to the random vector $V^n - W^n$. □

For any $\mathbf{a} \in \mathbb{R}_+^{\mathcal{D}}$ and any $\delta > 0$, define $\mathbf{a}_\delta := (a_{\delta,\kappa})_{\kappa \in \mathcal{D}}$ where for each $\kappa \in \mathcal{D}$,

$$a_{\delta,\kappa} := \begin{cases} \delta, & \text{if } a_\kappa \leq \delta, \\ a_\kappa, & \text{otherwise.} \end{cases} \tag{A-1}$$

In Lemma A.3, we examine the relationship between $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta)$.

Lemma A.3 *Let $\mathcal{N}$ be a canonical network source coding problem. For any $\epsilon > 0$, there exists $\delta > 0$ such that for any $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$ and any $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta)$ are ϵ-close.*

Proof. For any $\mathbf{a} \in \mathrm{I\!R}_+^{\mathcal{D}}$ and any $\delta > 0$, let $[\mathbf{a}]_\delta = ([a]_{\delta,\kappa})_{\kappa \in \mathcal{D}}$ be defined as

$$[a]_{\delta,\kappa} = \begin{cases} 0, & \text{if } a_\kappa \leq \delta, \\ a_\kappa, & \text{otherwise.} \end{cases}$$

Let $\epsilon > 0$ be given. The majority of the proof works to show that there exists a $\delta > 0$ such that for any $\mathbf{D} = (D_\kappa)_{\kappa \in \mathcal{D}} \in \mathrm{I\!R}_+^{\mathcal{D}}$ and any $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, [\mathbf{D}]_\delta)$ are $\epsilon/2$-close. From this we can observe that both $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta)$ are $\epsilon/2$-close to $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, [\mathbf{D}]_\delta)$, and hence $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta)$ are ϵ-close as desired.

To show that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, [\mathbf{D}]_\delta)$ are $\epsilon/2$-close, note first that $[\mathbf{D}]_\delta \leq \mathbf{D}$ for any $\delta > 0$, so $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, [\mathbf{D}]_\delta) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$. Therefore, we need only to choose an appropriate $\delta > 0$ (independent of $(\mathbf{X}, \mathbf{Y})$ and $\mathbf{D}$) such that

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) + (\epsilon/2) \cdot \mathbf{1} \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, [\mathbf{D}]_\delta).$$

For any achievable rate vector $\mathbf{R} \in \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$, pick a rate-$\mathbf{R}$, length-$n$ block code $\mathcal{C}$ such that for each $\kappa = (v, \theta) \in \mathcal{D}$,

$$Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq D_\kappa + \tau d_{\min},$$

where $\widehat{\theta}^n(v)$ is the reproduction of $\theta^n(\mathbf{X}^n)$ at v and $\tau > 0$ is a constant satisfying

$$sk(H(2\tau) + 2\tau \log m) < \frac{\epsilon}{2}.$$

Our goal is to use $\mathcal{C}$ to construct another code for which the error probability of the reproduction for each $\kappa \in \mathcal{D}$ with $D_\kappa \leq \delta$ can be made arbitrarily small. By Corollary A.2, for each $\kappa = (v, \theta) \in \mathcal{D}$ such that $\frac{1}{n} Ed(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)) \leq 2\tau d_{\min}$, we know that

$$H(\theta^n(\mathbf{X}^n) | \widehat{\theta}^n(v)) \leq n\epsilon/(2sk).$$

Since reproduction $\widehat{\theta^n}(v)$ is known at node v, additional rate vector

$$\left(\frac{s}{n}H(\theta^n(\mathbf{X}^n)|\widehat{\theta^n}(v)) + \epsilon/(2k)\right) \cdot \mathbf{1}$$

suffices to describe $\theta^n(\mathbf{X}^n)$ losslessly at node v by Lemma 2.2.18. We therefore modify code $\mathcal{C}$ by adding additional $|I(\theta)|$ descriptions for reproducing $\theta^n(\mathbf{X}^n)$ losslessly given $\widehat{\theta^n}$ at node v as described in Lemma 2.2.18. For every $(v, \theta) \in \mathcal{D}$ and every $i \in I(\theta)$, the modified code sends the corresponding additional description along a path connecting v and some source node that has access to source X_i. The total additional rate vector (for all such pairs of sources and demands) on each link is less than $(\epsilon/2) \cdot \mathbf{1}$. Therefore, the rate vector $\mathbf{R} + (\epsilon/2) \cdot \mathbf{1}$ is $[\mathbf{D}]_\delta$-achievable. By letting $\delta = \tau d_{\min}$, we get the desired result. □

Lemma A.4 shows that there exists a rate vector that is $\mathbf{D}$-achievable for any distribution $P_{\mathbf{X},\mathbf{Y}}$ and any distortion vector $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$.

Lemma A.4

$$\bigcap_{P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}} \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0}) \neq \emptyset.$$

Proof. Since each component of any vector $(\mathbf{X}, \mathbf{Y})$ of source and side information symbols has alphabet size no greater than m, the rate vector $k\ \log(m)\mathbf{1} \in \mathbb{R}_+^{\mathcal{E}}$ achieves $\mathbf{0}$ distortion for any $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$. □

Lemma A.5 combines the results of Lemmas A.1 - A.4 and is applied in the proof of Theorem 2.4.1.

Lemma A.5 *Let $\mathcal{N}$ be a canonical network source coding problem. For any $\epsilon > 0$, there exists a $\delta > 0$ such that for any $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and any $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$, $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \delta \cdot \mathbf{1})$ are ϵ-close.*

Proof. Given any $\epsilon > 0$. We prove the lemma by first showing that there exists a $\tau > 0$ such that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\tau)$ are $\epsilon/2$-close for all $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$, and then showing that there exists a $0 < \delta < \tau$ such that

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1}) + \frac{\epsilon}{2} \cdot \mathbf{1} \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\tau).$$

Together, these results imply that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1})$ are ϵ-close since

$$\begin{aligned}
&\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}} + \mathbf{D}_\delta + \delta \cdot \mathbf{1}) \\
&\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1}) + \epsilon \cdot \mathbf{1} \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\tau) + \frac{\epsilon}{2} \cdot \mathbf{1} \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}).
\end{aligned}$$

This proves the desired result since

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \delta \cdot \mathbf{1}) \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1}).$$

The first result follows immediately from Lemma A.3. Precisely,

(i) there exists a $\tau > 0$ such that $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ and $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\tau)$ are $\epsilon/2$-close for all $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$.

(ii) We next show that there exists a $0 < \delta < \tau$ such that for any $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$ with $D_\kappa \geq \tau$ for all $\kappa \in \mathcal{D}$,

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \delta \cdot \mathbf{1}) + \frac{\epsilon}{2} \cdot \mathbf{1} \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}).$$

To prove this, first fix $\mathbf{R}_0 \in \bigcap_{P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}} \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{0})$; this is possible by Lemma A.4. For any $0 < \delta < \tau$, $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$ and $D_\kappa \geq \tau$ for all $\kappa \in \mathcal{D}$ together imply

$$(1 - \frac{\delta}{\tau})(\mathbf{D} + \delta \cdot \mathbf{1}) + \frac{\delta}{\tau}(\tau \cdot \mathbf{1}) \leq \mathbf{D}$$

since for any $\kappa \in \mathcal{D}$,

$$\begin{aligned} & (1 - \frac{\delta}{\tau})(D_\kappa + \delta) + \frac{\delta}{\tau}\delta - D_\kappa \\ = \quad & \delta - \frac{\delta}{\tau} D_\kappa = \delta(1 - \frac{D_\kappa}{\tau}) \leq 0. \end{aligned}$$

By the convexity and the monotonicity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ on $\mathbf{D}$, we have

$$\begin{aligned} & (1 - \frac{\delta}{\tau})\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \delta \cdot \mathbf{1}) + \frac{\delta}{\tau}\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \delta \cdot \mathbf{1}) \\ & \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, (1 - \frac{\delta}{\tau})(\mathbf{D} + \delta \cdot \mathbf{1}) + \frac{\delta}{\tau}(\delta \cdot \mathbf{1})) \\ & \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}), \end{aligned}$$

which implies for all $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$ with $D_\kappa \geq \tau$ for all $\kappa \in \mathcal{D}$,

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D} + \delta \cdot \mathbf{1}) + \frac{\delta}{\tau}\mathbf{R}_0 \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}).$$

This together with the definition of $\mathbf{D}_\tau$ (A-1) implies that for all $\mathbf{D} \in \mathbb{R}_+^{\mathcal{D}}$,

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\tau + \delta \cdot \mathbf{1}) + \frac{\delta}{\tau}\mathbf{R}_0 \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\tau).$$

By the monotonicity of $\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ in $\mathbf{D}$, the following satisfies for all $\mathbf{D}_+^{\mathcal{D}}$ and for any $0 < \delta < \tau$

$$\mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\delta + \delta \cdot \mathbf{1}) + \frac{\delta}{\tau}\mathbf{R}_0 \subseteq \mathcal{R}(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D}_\tau).$$

Thus the desired result holds for all $0 < \delta < \tau$ that satisfy

$$\frac{\delta}{\tau}\mathbf{R}_0 \leq \frac{\epsilon}{2} \cdot \mathbf{1}.$$

$\square$

B Continuity of $\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ with respect to $\mathbf{D}$

Lemma B.1 shows that the set $\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is continuous at $\mathbf{D}$ when $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$ or $\mathcal{N}$ is canonical and $\mathbf{D} \geq \mathbf{0}$. The proof is similar to Theorems 2.4.1 and 2.4.2, hence we state the lemma without a proof.

Lemma B.1 *Let $\mathcal{N}$ be a network source coding problem. If $\mathcal{N}$ is non-canonical and $\mathbf{D} > \mathbf{0}$ or $\mathcal{N}$ is canonical and $\mathbf{D} \geq \mathbf{0}$, then $\mathcal{R}^*(P_{\mathbf{X},\mathbf{Y}}, \mathbf{D})$ is continuous at $\mathbf{D}$.*

C Lemmas for Section 2.5.2

In Lemma C.1, we show that when two distributions $P_{\mathbf{X},\mathbf{Y}}$ and $Q_{\mathbf{X},\mathbf{Y}}$ are sufficiently close and the support of $P_{\mathbf{X},\mathbf{Y}}$ is a subset of that of $Q_{\mathbf{X},\mathbf{Y}}$, there exists a joint distribution $T_{\mathbf{X},\mathbf{X}',\mathbf{Y},\mathbf{Y}'}$ with marginal on $(\mathbf{X}', \mathbf{Y}')$ equal to $Q_{\mathbf{X},\mathbf{Y}}$ and conditional distribution of $(\mathbf{X}', \mathbf{Y}')$ given $(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}', \mathbf{Y}')$ equal to $P_{\mathbf{X},\mathbf{Y}}$ for which $(\mathbf{X}, \mathbf{Y})$ equals $(\mathbf{X}', \mathbf{Y}')$ with high probability.

Lemma C.1 *For any $\epsilon > 0$ and any two distributions $P_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ and $Q_{\mathbf{X},\mathbf{Y}} \in \mathcal{M}$ satisfying that*

$$(1-\epsilon)P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \leq Q_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \ \forall \ (\mathbf{x},\mathbf{y}) \in \mathcal{A},$$

there exists a joint distribution $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ of $(\mathbf{X}, \mathbf{Y})$ with alphabet $\mathcal{A}$ and $(\mathbf{X}', \mathbf{Y}')$ with alphabet $\mathcal{A} \cup \{a_0\}$, where a_0 is an extra symbol not in $\mathcal{A}$, such that

(a) $Q_{\mathbf{X},\mathbf{Y}} = T_{\mathbf{X},\mathbf{Y}}$*, the marginal of $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ on $(\mathbf{X}, \mathbf{Y})$.*

(b)

$$P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = T_{\mathbf{X},\mathbf{Y}|\{(\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')\}}(\mathbf{x},\mathbf{y}) \ \forall \ (\mathbf{x},\mathbf{y}) \in \mathcal{A},$$

where $T_{\mathbf{X},\mathbf{Y}|\{(\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')\}}$ is the conditional distribution of $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ on $(\mathbf{X}, \mathbf{Y})$ given the event

$$\{(\mathbf{X},\mathbf{Y}) = (\mathbf{X}',\mathbf{Y}')\}.$$

(c)

$$\Pr\left((\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')\right)=1-\epsilon.$$

Proof. For any $(\mathbf{x},\mathbf{y})\in\mathcal{A}$, define

$$\begin{aligned}
&T_{\mathbf{X}',\mathbf{Y}'}(\mathbf{x},\mathbf{y}) := (1-\epsilon)P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})\\
&T_{\mathbf{X},\mathbf{Y}|\mathbf{X}',\mathbf{Y}'}(\mathbf{x},\mathbf{y}|\mathbf{x},\mathbf{y}) := 1\\
&T_{\mathbf{X}',\mathbf{Y}'}(a_0) := \epsilon\\
&T_{\mathbf{X},\mathbf{Y}|\mathbf{X}',\mathbf{Y}'}(\mathbf{x},\mathbf{y}|a_0) := \frac{Q_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})-(1-\epsilon)P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{\epsilon}.
\end{aligned}$$

All other values of $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ are zero. Then this distribution $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ has marginal on $(\mathbf{X},\mathbf{Y})$ equal to $Q_{\mathbf{X},\mathbf{Y}}$. Also,

$$\Pr\left((\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')\right)=\sum_{(\mathbf{x},\mathbf{y})\in\mathcal{A}} T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}(\mathbf{x},\mathbf{y},\mathbf{x},\mathbf{y})=1-\epsilon.$$

The conditional distribution of $(\mathbf{X},\mathbf{Y})=(\mathbf{x},\mathbf{y})$ given $(\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')$ for all $(\mathbf{x},\mathbf{y})\in\mathcal{A}$ is

$$T_{\mathbf{X},\mathbf{Y}|\{(\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')\}}(\mathbf{x},\mathbf{y})=\frac{T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}(\mathbf{x},\mathbf{y})}{\Pr\left((\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')\right)}=P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}).$$

□

Lemma C.2 *Suppose $P_{\mathbf{X},\mathbf{Y}}$, $Q_{\mathbf{X},\mathbf{Y}}$, and $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ are as described in Lemma C.1. Then*

$$\frac{1}{1-2\epsilon}\mathcal{R}(\widehat{\mathcal{N}},T_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'},\mathbf{D})\subseteq\mathcal{R}(\mathcal{N},P_{\mathbf{X},\mathbf{Y}},\frac{1}{1-2\epsilon}\mathbf{D}).$$

Proof. The main idea of the proof is as follows. Since the probability of the event $(\mathbf{X},\mathbf{Y})=(\mathbf{X}',\mathbf{Y}')$ is $1-\epsilon$, when block length n is sufficiently high, the number of occurrences that $(\mathbf{X}_i,\mathbf{Y}_i)=(\mathbf{X}'_i,\mathbf{Y}'_i)$ in length-n sequence $(\mathbf{X}^n,\mathbf{Y}^n)$ is higher than $n(1-2\epsilon)$ for sufficiently high probability by the weak law of large numbers. We apply this property to show that any rate-$\mathbf{R}$, length-n block code $\mathcal{C}_n$ for $\widehat{\mathcal{N}}$ with source vector $\mathbf{X}$ and side-information vector $(\mathbf{Y}, \mathbf{Y}', \mathbf{X}')$ which achieves $\mathbf{D}$ can be applied to construct a length-$n(1-2\epsilon)$ block code for $\mathcal{N}$ with source and side information vectors $\mathbf{X}^{n(1-2\epsilon)}$ and $\mathbf{Y}^{n(1-2\epsilon)}$ that has rate no greater than $\frac{1}{1-2\epsilon}\mathbf{R}$ and satisfies the distortion constraint given by $\frac{1}{1-2\epsilon}\mathbf{D}$.

Let $\tau > 0$ and $\mathbf{R} \in \mathcal{R}(\widehat{\mathcal{N}}, T_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}, \mathbf{D})$. Let n be sufficiently large such that there exists a rate-$\mathbf{R}$, length-n block code $\mathcal{C}_n$ which satisfies

$$\Pr(\mathcal{T}_n) > 1 - \tau, \tag{A-2}$$

where $\mathcal{T}_n$ is the event

$$\mathcal{T}_n := \left\{ \frac{1}{n} d\left(\theta^n(\mathbf{X}^n), \widehat{\theta}^n(v)\right) < D_{(v,\theta)} + \tau \ \forall\ (v,\theta) \in \mathcal{D} \right\}$$

and $\widehat{\theta}^n(v)$ is the reproduction of $\theta^n(\mathbf{X}^n)$ at node v using $\mathcal{C}_n$ for all $(v,\theta) \in \mathbb{R}_+^{\mathcal{D}}$. (A-2) can be achieved by applying the weak law of large numbers on a long code constructed by repeatedly using a code which achieves $D_{(v,\theta)} + \tau/2$. For any $I \subseteq \{1, \ldots, n\}$, define

$$B_\epsilon^{(n)}(I) := \{(\mathbf{x}^n, \mathbf{x}'^n, \mathbf{y}^n, \mathbf{y}'^n) \mid (\mathbf{x}_i, \mathbf{y}_i) \neq (\mathbf{x}'_i, \mathbf{y}'_i) \text{ if and only if } i \in I\}.$$

Let

$$\mathcal{L}(I) := \{(\mathbf{a}^{|I|}, \mathbf{b}^{|I|}) \in \mathcal{A}^{|I|} \times \mathcal{A}^{|I|} \mid \mathbf{a}_i \neq \mathbf{b}_i \ \forall\ i \in I\}.$$

For any sequence $(\mathbf{a}^{|I|}, \mathbf{b}^{|I|}) \in \mathcal{L}(I)$, let

$$B_\epsilon^{(n)}(I, (\mathbf{a}^{|I|}, \mathbf{b}^{|I|})) \ := \ \left\{(\mathbf{x}^n, \mathbf{x}'^n, \mathbf{y}^n, \mathbf{y}'^n) \mid ((\mathbf{x}_I, \mathbf{y}_I), (\mathbf{x}'_I, \mathbf{y}'_I)) = (\mathbf{a}^{|I|}, \mathbf{b}^{|I|})\right\} \cap B_\epsilon^{(n)}(I)$$

denote the set of sequences $(\mathbf{x}^n, \mathbf{x}'^n, \mathbf{y}^n, \mathbf{y}'^n) \in B_\epsilon^{(n)}(I)$ such that $((\mathbf{x}_i, \mathbf{y}_i), (\mathbf{x}'_i, \mathbf{y}'_i)) = (\mathbf{a}_i, \mathbf{b}_i)$ for all $i \in I$. Let

$$B_\epsilon^{(n)} := \bigcup_{I \subseteq \{1,\ldots,n\} \ |I| < 2n\epsilon} B_\epsilon^{(n)}(I)$$

denote the set of sequences $(\mathbf{x}^n, \mathbf{x}'^n, \mathbf{y}^n, \mathbf{y}'^n)$ for which there are at most $2n\epsilon$ indices i such that $(\mathbf{x}_i, \mathbf{y}_i) \neq (\mathbf{x}'_i, \mathbf{y}'_i)$. Since $\Pr((\mathbf{X}, \mathbf{Y}) \neq (\mathbf{X}', \mathbf{Y}')) = \epsilon$,

$$\Pr\left(B_\epsilon^{(n)}\right) > 1 - \tau$$

when n is sufficiently large by the weak law of large numbers. By the union bound,

$$\Pr\left(B_\epsilon^{(n)} \cap \mathcal{T}_n\right) > 1 - 2\tau. \tag{A-3}$$

Rewrite inequality (A-3) as

$$\begin{aligned}\sum_{I \subseteq \{1,\ldots,n\}\ |I|<2n\epsilon} \sum_{(\mathbf{a}^{|I|}, \mathbf{b}^{|I|}) \in \mathcal{L}(I)} & \Pr\left(B_\epsilon^{(n)}(I, (\mathbf{a}^{|I|}, \mathbf{b}^{|I|}))\right) \Pr\left(\mathcal{T}_n | B_\epsilon^{(n)}(I, (\mathbf{a}^{|I|}, \mathbf{b}^{|I|}))\right) \\ &= \Pr\left(B_\epsilon^{(n)} \cap \mathcal{T}_n\right) > 1 - 2\tau\end{aligned}$$

Therefore, there exist $\widehat{I} \subseteq \{1, \ldots, n\}$ and
$(\mathbf{a}_0^{|\widehat{I}|}, \mathbf{b}_0^{|\widehat{I}|}) = ((\mathbf{x}_0^{|I|}, \mathbf{y}_0^{|I|}), (\mathbf{x}_0'^{|I|}, \mathbf{y}_0'^{|I|})) \in \mathcal{L}(I)$ such that $|\widehat{I}| < 2n\epsilon$ and

$$\Pr\left(\mathcal{T}_n | B_\epsilon^{(n)}(\widehat{I}, (\mathbf{a}_0^{|\widehat{I}|}, \mathbf{b}_0^{|\widehat{I}|}))\right) > 1 - 2\tau. \tag{A-4}$$

Without loss of generality, suppose that

$$I = \{n - |I| + 1, \ldots, n\}.$$

For any $(\mathbf{x}^{n-|I|}, \mathbf{y}^{n-|I|}) \in \mathcal{A}^{n-|I|}$,

$$\begin{aligned}&\Pr\left((\mathbf{X}^{n-|I|}, \mathbf{Y}^{n-|I|}) = (\mathbf{x}^{n-|I|}, \mathbf{y}^{n-|I|}) \mid (\mathbf{X}^n, \mathbf{Y}^n) \in B_\epsilon^{(n)}(\widehat{I}, (\mathbf{a}_0^{|\widehat{I}|}, \mathbf{b}_0^{|\widehat{I}|}))\right) \\ &= T_{\mathbf{X}^{n-|I|}, \mathbf{Y}^{n-|I|} | \{(\mathbf{X}^{n-|I|}, \mathbf{X}'^{n-|I|}) = (\mathbf{Y}^{n-|I|}, \mathbf{Y}'^{n-|I|})\}}(\mathbf{x}^n, \mathbf{y}^n) \\ &= P_{\mathbf{X}^{n-|I|}, vecY^{n-|I|}}(\mathbf{x}^{n-|I|}, \mathbf{y}^{n-|I|})\end{aligned} \tag{A-5}$$

Let $\widehat{\mathcal{C}}_{n-|I|}$ be the length-$(n - |I|)$ block code for input sequence $(\mathbf{X}^{n-|I|}, \mathbf{Y}^{n-|I|})$ by applying code $\mathcal{C}_n$ to

$$((\mathbf{X}^{n-|I|}, \mathbf{x}_0^{|I|}), (\mathbf{Y}^{n-|I|}, \mathbf{y}_0^{|I|}), (\mathbf{X}^{n-|I|}, \mathbf{x}_0'^{|I|}), (\mathbf{Y}^{n-|I|}, \mathbf{y}_0'^{|I|})).$$

By (A-4), the code $\widehat{\mathcal{C}}_{n-|I|}$ has expected average distortions

$$\frac{1}{n-|I|} Ed(\theta^{n-|I|}, \widehat{\theta}^{n-|I|}(v)) \leq (1-2\tau)\left(\frac{n}{n-|I|} D_{(v,\theta)} + \frac{n}{n-|I|}\tau\right) + 2\tau d_{\max}$$

for all $(v, \theta) \in \mathbb{R}_+^{\mathcal{D}}$. Therefore, since $|I| < 2n\epsilon$, $\widehat{\mathcal{C}}_{n-|I|}$ has rate

$$\frac{n}{n-|I|}\mathbf{R} \leq \frac{1}{1-2\epsilon}\mathbf{R}$$

and expected average distortion vector no greater than

$$(1-2\tau)\left(\frac{1}{1-2\epsilon}\mathbf{D} + \frac{1}{1-2\epsilon}\tau\right) + 2\tau d_{\max}$$

By (A-5), the expected distortion is evaluated according to $P_{\mathbf{X},\mathbf{Y}}$. Hence by letting $\tau \to 0$

$$\frac{1}{1-2\epsilon}\mathbf{R} \in \mathcal{R}(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}, \frac{1}{1-2\epsilon}\mathbf{D}).$$

This completes the proof. $\square$

Lemma C.3 *Suppose $P_{\mathbf{X},\mathbf{Y}}$, $Q_{\mathbf{X},\mathbf{Y}}$, and $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ are as described in Lemma C.1. Then*

$$\frac{1}{1-2\epsilon}\mathcal{R}_L(\widehat{\mathcal{N}}, T_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}) \subseteq \mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}).$$

Proof. The proof is similar to that of Lemma C.2. Let $\tau > 0$ and $\mathbf{R} \in \mathcal{R}(\widehat{\mathcal{N}}, T_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}, \mathbf{D})$. Let n be sufficiently large such that there exists a rate-$\mathbf{R}$, length-n block code $\mathcal{C}_n$ which satisfies

$$\Pr\left(\theta^n(\mathbf{X}^n) = \widehat{\theta}^n(v)\ \forall (v,\theta) \in \mathcal{D}\right) \geq 1-\tau,$$

where $\widehat{\theta}^n(v)$ is the reproduction of $\theta^n(\mathbf{X}^n)$ at node v using $\mathcal{C}_n$ for all $(v,\theta) \in \mathrm{I\!R}_+^{\mathcal{D}}$. Let

$$E_n := \left\{\theta^n(\mathbf{X}^n) \neq \widehat{\theta}^n(v) \text{ for some } (v,\theta) \in \mathcal{D}\right\}$$

denote the event of decoding error using code $\mathcal{C}_n$. Let $\mathcal{L}(I)$ (for $I \subseteq \{1,\dots,n\}$) and $B_\epsilon^{(n)}(I,(\mathbf{a}^{|I|},\mathbf{b}^{|I|}))$ (for $(\mathbf{a}^{|I|},\mathbf{b}^{|I|}) \in \mathcal{L}(I)$) be the sets defined in the proof of Theorem C.2. The same argument in Theorem C.2 leads to

$$\Pr\left(E_n^c | B_\epsilon^{(n)}(\widehat{I},(\mathbf{a}_0^{|\widehat{I}|},\mathbf{b}_0^{|\widehat{I}|}))\right) > 1 - 2\tau \tag{A-6}$$

for some $\widehat{I} \subseteq \{1,\dots,n\}$ and $(\mathbf{a}_0^{|\widehat{I}|},\mathbf{b}_0^{|\widehat{I}|}) = ((\mathbf{x}_0^{|I|},\mathbf{y}_0^{|I|}),(\mathbf{x}_0'^{|I|},\mathbf{y}_0'^{|I|})) \in \mathcal{L}(I)$ such that $|\widehat{I}| < 2n\epsilon$. Without loss of generality, suppose that

$$I = \{n - |I| + 1, \dots, n\}.$$

Let $\widehat{\mathcal{C}}_{n-|I|}$ be the length-$(n-|I|)$ block code for input sequence $(\mathbf{X}^{n-|I|},\mathbf{Y}^{n-|I|})$ by applying code $\mathcal{C}_n$ to

$$((\mathbf{X}^{n-|I|},\mathbf{x}_0^{|I|}),(\mathbf{Y}^{n-|I|},\mathbf{y}_0^{|I|}),(\mathbf{X}^{n-|I|},\mathbf{x}_0'^{|I|}),(\mathbf{Y}^{n-|I|},\mathbf{y}_0'^{|I|})).$$

By (A-6), the code $\widehat{\mathcal{C}}_{n-|I|}$ has decoding error probability no greater than 2τ and has rate no greater than

$$\frac{1}{1-2\epsilon}\mathbf{R}.$$

By (A-5), the error probability is evaluated according to $P_{\mathbf{X},\mathbf{Y}}$. Hence by letting $\tau \to 0$

$$\frac{1}{1-2\epsilon}\mathbf{R} \in \mathcal{R}_L(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}).$$

This completes the proof. $\square$

Lemma C.4 *Suppose $P_{\mathbf{X},\mathbf{Y}}$, $Q_{\mathbf{X},\mathbf{Y}}$, and $T_{\mathbf{X},\mathbf{Y},\mathbf{X}',\mathbf{Y}'}$ are as described in Lemma C.1.*

Then

$$\frac{1}{1-2\epsilon}\mathcal{R}_Z(\widehat{\mathcal{N}}, T_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}) \subseteq \mathcal{R}_Z(\mathcal{N}, P_{\mathbf{X},\mathbf{Y}}).$$

Proof. The proof is similar to that of Lemma C.2. Let $\tau > 0$ and $\mathbf{R} \in \mathcal{R}(\widehat{\mathcal{N}}, T_{\mathbf{X},\mathbf{Y},\mathbf{Y}',\mathbf{X}'}, \mathbf{D})$. Let n be sufficiently large such that there exists a dimension-n zero-error variable-length code $\mathcal{C}_n$ with length vector $\mathbf{L}^{(n)}$ which satisfies

$$\Pr(\mathcal{K}_n) > 1-\tau,$$

where

$$\mathcal{K}_n := \{\frac{1}{n}\mathbf{L}^{(n)}(\mathbf{X}^n, \mathbf{Y}^n) \leq \mathbf{R} + \tau \cdot \mathbf{1}\} \tag{A-7}$$

is the event that the average length vector is no greater than $\mathbf{R} + \tau \cdot 1$. (A-7) can be achieved by applying the weak law of large numbers on a long code constructed by repeatedly using the codewords of a zero-error variable-length code whose expected average length vector is no greater than $\mathbf{R} + (\tau/2) \cdot \mathbf{1}$. Let $\mathcal{L}(I)$ (for $I \subseteq \{1, \ldots, n\}$) and $B_\epsilon^{(n)}(I, (\mathbf{a}^{|I|}, \mathbf{b}^{|I|}))$ (for $(\mathbf{a}^{|I|}, \mathbf{b}^{|I|}) \in \mathcal{L}(I)$) be the sets defined in the proof of Theorem C.2. The same argument in Theorem C.2 leads to

$$\Pr\left(\mathcal{K}_n | B_\epsilon^{(n)}(\widehat{I}, (\mathbf{a}_0^{|\widehat{I}|}, \mathbf{b}_0^{|\widehat{I}|}))\right) > 1 - 2\tau \tag{A-8}$$

for some $\widehat{I} \subseteq \{1, \ldots, n\}$ and $(\mathbf{a}_0^{|\widehat{I}|}, \mathbf{b}_0^{|\widehat{I}|}) = ((\mathbf{x}_0^{|I|}, \mathbf{y}_0^{|I|}), (\mathbf{x}_0'^{|I|}, \mathbf{y}_0'^{|I|})) \in \mathcal{L}(I)$ such that $|\widehat{I}| < 2n\epsilon$. Without loss of generality, suppose that

$$I = \{n - |I| + 1, \ldots, n\}.$$

Let $\widehat{\mathcal{C}}_{n-|I|}$ be the length-$(n-|I|)$ block code for input sequence $(\mathbf{X}^{n-|I|}, \mathbf{Y}^{n-|I|})$ by applying code $\mathcal{C}_n$ to

$$((\mathbf{X}^{n-|I|}, \mathbf{x}_0^{|I|}), (\mathbf{Y}^{n-|I|}, \mathbf{y}_0^{|I|}), (\mathbf{X}^{n-|I|}, \mathbf{x}_0'^{|I|}), (\mathbf{Y}^{n-|I|}, \mathbf{y}_0'^{|I|})).$$

By (A-8), the zero-error variable-length code $\widehat{\mathcal{C}}_{n-|I|}$ has expected average length vector no greater than

$$\frac{1}{1-2|I|}\left((1-2\tau)(\mathbf{R}+\tau\cdot\mathbf{1})+(2\tau(s+2t)\log m)\cdot\mathbf{1}\right)$$
$$\geq \frac{1}{1-2\epsilon}\left((1-2\tau)(\mathbf{R}+\tau\cdot\mathbf{1})+(2\tau(s+2t)\log m)\cdot\mathbf{1}\right),$$

where the value $(s+2t)\log m$ uppers all possible average lengths over each edge $e\in\mathcal{E}$ since for each $e\in\mathcal{E}$, there are at most M^{s+2t} codewords using $\mathcal{C}_n$. By (A-5), the expected length is evaluated according to $P_{\mathbf{X},\mathbf{Y}}$. Hence by letting $\tau\to 0$

$$\frac{1}{1-2\epsilon}\mathbf{R}\in\mathcal{R}_Z(\mathcal{N},P_{\mathbf{X},\mathbf{Y}}).$$

This completes the proof. □

D Proof of Theorem 4.4.1

Let $J_2^*(\lambda)=\min_{\{P_Z(z)\}_{z\in\mathcal{Z}}} J_2(\lambda)$ be the optimal value of $J_2(\lambda)$ for the Wyner-Ziv rate region, and let $\widehat{J}_2(\lambda)$ be the value computed by the algorithm proposed in Section 4.4. Then $\widehat{J}_2(\lambda)\geq J_2^*(\lambda)$ since the algorithm finds an auxiliary random variable Z achieving the given Lagrangian. We next find (η,δ) to guarantee that $\widehat{J}_2(\lambda)\leq(1+\epsilon)J_2^*(\lambda)$.

Recall that $J_2(\lambda):=I(X;Z|Y)+\lambda\min_\psi Ed(X,\psi(Y,Z))$. Before bounding $\widehat{J}_2(\lambda)-$

$J_2^*(\lambda)$, rewrite $J_2(\lambda)$ as

$$\begin{aligned} J_2(\lambda) &= H(X|Y) - H(Y|X) + H(Y|Z) - H(X|Z) \\ &\quad + \lambda \min_{\psi} Ed(X, \psi(Y,Z)) \\ &= H(X|Y) - H(Y|X) + \sum_{z\in\mathcal{Z}} P_Z(z) H(Y|Z=z) \\ &\quad - \sum_{z\in\mathcal{Z}} P_Z(z) \Bigg[H(X|Z=z) \\ &\quad + \sum_{x} p(y|x) Q_{X|Z}(x|z) d(x, \psi^*(y,z)) \Bigg], \end{aligned}$$

where $\psi^*(y,z)$ is the optimizing reproduction of X given conditional $\{Q_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ and (y,z). Fix $\{Q_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ and let $\{\widehat{Q}_{X|Z}(x|z)\}_{(x,z)\in\mathcal{X}\times\mathcal{Z}}$ be the quantized conditional. Let

$$\begin{aligned} Q_{Y|Z}(y|z) &= \sum_x p(y|x) Q_{X|Z}(x|z) \\ \widehat{Q}_{Y|Z}(y|z) &= \sum_x p(y|x) \widehat{Q}_{X|Z}(x|z) \end{aligned}$$

be the corresponding conditionals on Y given Z. Then

$$\begin{aligned} (1-\eta)Q_{X|Z}(x|z) - \delta \leq \ & \widehat{Q}_{X|Z}(x|z) & \leq (1+\eta)Q_{X|Z}(x|z) \\ (1-\eta)Q_{Y|Z}(y|z) - \delta \leq \ & \widehat{Q}_{Y|Z}(y|z) & \leq (1+\eta)Q_{Y|Z}(y|z) \end{aligned}$$

for all x, y, z. Finally, let $\{P_Z^*(z)\}_{z\in\mathcal{Z}}$ be the marginal on auxiliary random variable Z that achieves $J_2^*(\lambda)$ and define $\tau := \eta \log \frac{e}{1-\eta}$. By Lemma 4.2.1, when $(\max\{|\mathcal{X}|, |\mathcal{Y}|\})\, \delta \log \frac{1}{\delta} < \tau$

$$\begin{aligned} |H(\widehat{Q}_{X|Z=z}) - H(Q_{X|Z=z})| &\leq \eta H(Q_{X|Z=z}) + 2\tau \\ |H(\widehat{Q}_{Y|Z=z}) - H(Q_{Y|Z=z})| &\leq \eta H(Q_{Y|Z=z}) \end{aligned}$$

for every $z \in \mathcal{Z}$.

Let $\mathcal{Z}' := \mathcal{Z} \cup \{z_x\}_{x\in\mathcal{X}}$ and set

$$\widehat{Q}_{X|Z}(t|z_x) = \begin{cases} 1, & \text{if } t = x \\ 0, & \text{otherwise} \end{cases}$$

$$\widehat{P}_Z(z) = \begin{cases} (1-\eta)P_Z^*(z), & \text{if } z \in \mathcal{Z} \\ P_X(x) - \sum_{z\in\mathcal{Z}} \widehat{Q}_{X|Z}(x|z)\widehat{P}_Z(z), & \text{if } z = z_x. \end{cases}$$

$$\widehat{\psi}(y,z) = \begin{cases} \psi^*(y,z) & \text{if } z \in \mathcal{Z} \\ x & \text{if } z = z_x \end{cases}$$

Since $\widehat{J}_2(\lambda)$ is optimized over all quantized distributions,

$$\widehat{J}_2(\lambda) \leq J_2(\lambda)|_{\widehat{Q}_{X|Z},\widehat{P}_Z,\widehat{\psi}}.$$

Thus

$$\begin{aligned} &\widehat{J}_2(\lambda) - J_2^*(\lambda) \\ &\quad\leq\ [(2\eta-\eta^2)H(X|Z) - \eta^2 H(Y|Z)]_{P_Z^*,Q_{X|Z}} \\ &\qquad +(2\eta-\eta^2)H(Y|X) + 4(1-\eta)\tau + |\mathcal{Y}||\mathcal{Z}|\delta(1-\eta) \\ &\quad\leq\ \eta\left((2-\eta)(|\mathcal{X}|+|\mathcal{Y}|) + 8 + |\mathcal{Y}||\mathcal{Z}|\right) \\ &\quad\leq\ \eta(2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|). \end{aligned}$$

when $\delta < \eta < 1 - \frac{e}{4}$ and $(\max\{|\mathcal{X}|, |\mathcal{Y}|\})\,\delta\log\frac{1}{\delta} < \tau$. Define

$$L^*(\lambda) := \min_{0\leq D\leq D_{\max}} \big(R_{X|Y}(D) + \lambda D\big)$$

where $R_{X|Y}(D)$ is the conditional rate-distortion function for X given Y. Then $J_2^*(\lambda) \geq L^*(\lambda)$ implies

$$\widehat{J}_2(\lambda) - J_2^*(\lambda) \leq \frac{\eta(2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|)}{L^*(\lambda)} J_2^*(\lambda).$$

We therefore wish to choose δ and η to satisfy

$$\delta < \eta = \frac{L^*(\lambda)}{2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|}\epsilon < 1 - \frac{e}{4}.$$

Define $f(x) := -x\log(x)$ for $x \in [0, 1/e)$. Function f is strictly increasing and therefore invertible. Setting

$$\eta = \min\left\{\frac{L^*(\lambda)}{2|\mathcal{X}| + 2|\mathcal{Y}| + 8 + |\mathcal{Y}||\mathcal{Z}|}\epsilon, 1 - \frac{e}{4}\right\} \quad \text{(A-9)}$$

$$\delta = \min\left\{\eta, f^{-1}\left(\min\left\{\frac{\eta}{|\mathcal{X}|}, \frac{\eta}{|\mathcal{Y}|}\right\}\right)\right\} \quad \text{(A-10)}$$

yields $(\max\{|\mathcal{X}|, |\mathcal{Y}|\})\,\delta\log\frac{1}{\delta} < \tau$ as desired and guarantees a $(1+\epsilon)$-approximation.

The interior-point solver for k variables runs in time $O(k^4)$ [54]. Since $k = |\mathcal{Z}|$ for our linear program, our algorithm runs in time $O(N(\delta, \eta, |\mathcal{X}|)^4)$. Applying the given choice of δ and η, our algorithm runs in time $O(\epsilon^{-4(|\mathcal{X}|+1)})$ as ϵ approaches 0.

E Proof of Theorem 4.5.1

Let $J_3^*(\lambda_1, \lambda_2, \lambda_3) = \min J_3(\lambda_1, \lambda_2, \lambda_3)$ be the optimal value of $J_3(\lambda_1, \lambda_2, \lambda_3)$ for the lossy coded side information region, and let $\widehat{J}_3(\lambda_1, \lambda_2, \lambda_3)$ be the value computed by the algorithm proposed in Section 4.5. Then $\widehat{J}_3(\lambda_1, \lambda_2, \lambda_3) \geq J_3^*(\lambda_1, \lambda_2, \lambda_3)$. We next find (η, δ) such that $\widehat{J}_3(\lambda_1, \lambda_2, \lambda_3) \leq (1+\epsilon)J_3^*(\lambda_1, \lambda_2, \lambda_3)$.

Let $\mathcal{Z}_1' = \mathcal{Z}_1 \cup \{z_{x_1}\}_{x_1 \in \mathcal{X}_1}$ and set

$$\widehat{Q}_{X_1|Z_1}(t|z_{x_1}) = \begin{cases} 1, & \text{if } t = x_1 \\ 0, & \text{otherwise} \end{cases}.$$

Let $P^*_{Z_1}Q^*_{Z_2|X_2}$ be a distribution on (Z_1, X_2) that achieves $J_3^*(\lambda_1, \lambda_2, \lambda_3)$. Define

$$\widehat{P}_{Z_1}(z) = \begin{cases} (1-\eta')P^*_{Z_1}(z) & \forall\ z \in \mathcal{Z}, \\ P_{X_1}(x_1) - \sum_z \widehat{Q}_{X_1|Z}(x_1|z)\widehat{P}_{Z_1}(z) & \forall\ x_1 \in \mathcal{X}_1. \end{cases}$$

Let $\tau' := \eta' \log \frac{e}{1-\eta'}$ and $\delta' := (|\mathcal{X}_1|+1)\delta\eta' = 3\eta$. Choose $\delta > 0$ such that $|\mathcal{X}_1||\mathcal{Z}_2|\delta' \log \frac{1}{\delta'} \leq \tau'$. By Lemma 4.2.1, for all $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, and $z_1 \in \mathcal{Z}_1$

$$
\begin{aligned}
&|H(\widehat{Q}_{X_1|Z_1=z_1}) - H(Q_{X_1|Z_1=z_1})| \\
&\quad \leq \quad \eta' H(Q_{X_1|Z_1=z_1}) + 2\tau' \\
&|H(\widehat{Q}_{X_1,Z_2|Z_1=z_1}) - H(Q_{X_1,Z_2|Z_1=z_1})| \\
&\quad \leq \quad \eta' H(Q_{X_1,Z_2|Z_1=z_1}) + 2\tau' \\
&|H(\widehat{Q}_{Z_2|Z_1=z_1}) - H(Q_{Z_2|Z_1=z_1})| \\
&\quad \leq \quad \eta' H(Q_{Z_2|Z_1=z_1}) + 2\tau' \\
&|H(\widehat{Q}_{Z_2|X_2=x_2}) - H(Q_{Z_2|X_2=x_2})| \\
&\quad \leq \quad \eta' H(Q_{Z_2|X_2=x_2}) + 2\tau' \\
&|H(\widehat{Q}_{Z_2}) - H(Q_{Z_2})| \leq \eta' H(Q_{Z_2}) + 2\tau'
\end{aligned}
$$

where $\widehat{Q}_{X_1,Z_2|Z_1}$, $\widehat{Q}_{Z_2|Z_1}$, and $\widehat{Q}_{Z_2}$ derive from $(\widehat{Q}_{X_1|Z_1}, \widehat{Q}_{Z_2|X_2})$. Let ψ^* be the optimal reproduction function for X_1 for $(P^*_{Z_1}, Q^*_{Z_2|X_2})$. Extend the function $\psi^*(z_1, z_2)$ to $\psi^*(z_{x_1}, z_2) = x_1$ for all $x_1 \in \mathcal{X}_1$. Now

$$
\begin{aligned}
&I(X_1; Z_1|Z_2) \\
&\quad = \quad H(X_1|Z_1) - H(X_1|Z_1, Z_2) \\
&\quad = \quad H(X_1|Z_1) - H(X_1, Z_2|Z_1) - H(Z_2|Z_1) \\
I(X_2; Z_2) \quad &= \quad H(Z_2) - H(Z_2|X_2),
\end{aligned}
$$

and hence we have

$$\begin{aligned}
\widehat{H}(X_1|Z_1) &\leq (1-\eta'^2)H^*(X_1|Z_1) + 2(1-\eta')\tau' \\
\widehat{H}(X_1, Z_2|Z_1) &\geq (1-\eta')^2 H^*(X_1, Z_2|Z_1) - 2(1-\eta')\tau' \\
\widehat{H}(Z_2|Z_1) &\geq (1-\eta')^2 H^*(Z_2|Z_1) - 2(1-\eta')\tau' \\
\widehat{H}(Z_2|X_2) &\geq (1-\eta')^2 H^*(Z_2|X_2) - 2(1-\eta')\tau' \\
& Ed(X_1, \psi^*(Z_1, Z_2))_{\widehat{P}_{Z_1}, \widehat{Q}^*_{Z_2|X_2}} \\
&= (1-\eta') Ed(X_1, \psi^*(Z_1, Z_2))_{P^*_{Z_1}, Q^*_{Z_2|X_2}}.
\end{aligned}$$

Since $\widehat{J}_3(\lambda_1, \lambda_2, \lambda_3) \leq \widehat{J}_3(\lambda_1, \lambda_2, \lambda_3)|_{\widehat{P}_{Z_1}, \{\widehat{Q}_{X_1|Z_1}\}, \widehat{Q}^*_{Z_2|X_2}}$, by taking $\eta' < 1 - \frac{e}{4}$, we have

$$\begin{aligned}
& \widehat{J}_3(\lambda_1, \lambda_2, \lambda_3) - J_3^*(\lambda_1, \lambda_2, \lambda_3) \\
\leq\ & \lambda_1 \left(2\eta' \left(|\mathcal{X}_1||\mathcal{Z}_1|\right) + |\mathcal{Z}_2|\right)) + \lambda_2(\eta'|\mathcal{Z}_2| + 2\eta'|\mathcal{Z}_2|) \\
& +(6\lambda_1 + 4\lambda_2)(1-\eta')\tau' \\
\leq\ & \eta' \left(2\lambda_1(|\mathcal{X}_1||\mathcal{Z}_1| + |\mathcal{Z}_1|) + 3\lambda_2|\mathcal{Z}_2| + (12\lambda_1 + 8\lambda_2)\right).
\end{aligned}$$

Define

$$L(\lambda_1, \lambda_2, \lambda_3) := \min_D \left[\min\{\lambda_1, \lambda_2\} R_{X_1}(D) + \lambda_3 D\right].$$

Set

$$\eta = \min\left\{\frac{4-e}{12}, \frac{\epsilon L(\lambda_1, \lambda_2, \lambda_3)}{T(\lambda_1, \lambda_2, \lambda_3)}\right\} \tag{A-11}$$

$$\delta = \frac{1}{|\mathcal{X}_1| + 1} f^{-1}\left(\frac{\eta}{3|\mathcal{X}_1||\mathcal{Z}_2|}\right), \tag{A-12}$$

where

$$\begin{aligned}
& T(\lambda_1, \lambda_2, \lambda_3) \\
:=\ & 6\lambda_1(|\mathcal{X}_1||\mathcal{Z}_1| + |\mathcal{Z}_1|) + 9\lambda_2|\mathcal{Z}_2| + (36\lambda_1 + 24\lambda_2).
\end{aligned}$$

Then the pair (η, δ) satisfies the following inequalities

$$
\begin{aligned}
\delta' &= (|\mathcal{X}_1| + 1)\delta \\
\eta' &= 3\eta < 1 - \frac{e}{4} \\
&|\mathcal{X}_1||\mathcal{Z}_2|\delta' \log \frac{1}{\delta'} \leq \tau' \\
\eta' &< \frac{\epsilon L(\lambda_1, \lambda_2, \lambda_3)}{2\lambda_1(|\mathcal{X}_1||\mathcal{Z}_1| + |\mathcal{Z}_1|) + 3\lambda_2|\mathcal{Z}_2| + (12\lambda_1 + 8\lambda_2)},
\end{aligned}
$$

which gives

$$
J_3^*(\lambda_1, \lambda_2, \lambda_3) \leq \widehat{J}_3(\lambda_1, \lambda_2, \lambda_3) \leq (1 + \epsilon) J_3^*(\lambda_1, \lambda_2, \lambda_3).
$$

In this algorithm, there are $N(\delta, \eta, |\mathcal{X}_1|)$ variables in each of the linear programs in the inner loop, and there are $N(\delta, \eta, |\mathcal{X}_2|)^{|\mathcal{X}_2|}$ quantized conditional probabilities $\widehat{Q}_{Z_2|X_2}$ in the outer loop. Applying the given choice of δ and η, our algorithm runs in time $O(\epsilon^{-4(|\mathcal{X}_1| + |\mathcal{X}_2^2| + 1)})$ as ϵ approaches 0.

Bibliography

[1] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *IRE National Convention Record*, volume 4, pages 142–163, 1959.

[2] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[3] R. M. Gray and A. D. Wyner. Source coding for a simple network. *Bell System Technical Journal*, 53(9):1681–1721, November 1974.

[4] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.

[5] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-22(1):1–10, January 1976.

[6] T. Berger, K. B. Housewright, J. K. Omura, S. Tung, and J. Wolfowitz. An upper bound on the rate distortion function for source coding with partial side information at the decoder. *IEEE Transactions on Information Theory*, IT-25(6):664–666, November 1979.

[7] T. Berger and R. Yeung. Multiterminal source encoding with one distortion criterion. *IEEE Transactions on Information Theory*, IT-35(2):228–236, March 1989.

[8] C. Heegard and T. Berger. Rate distortion when side information may be absent. *IEEE Transactions on Information Theory*, IT-31:727–734, November 1985.

[9] H. Yamamoto. Source coding theory for cascade and branching communication systems. *IEEE Transactions on Information Theory*, IT-27:299–308, May 1981.

[10] R. Ahlswede, N. Cai, S. Y. R. Li, and R. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.

[11] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, IT-52(10):4413–4430, October 2006.

[12] M. Bakshi and M. Effros. On achievable rates for multicast in the presence of side information. In *Proceedings of the IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.

[13] H. Yamamoto. Wyner-Ziv theory for a general function of the correlated sources. *IEEE Transactions on Information Theory*, IT-28(5):803–807, September 1982.

[14] A. Orlitsky and J. R. Roche. Coding for computing. *IEEE Transactions on Information Theory*, IT-47(3):903–917, March 2001.

[15] H. Feng, M. Effros, and S. Savari. Functional source coding for networks with receiver side information. *Proceedings of the Allerton Conference on Communication, Control, and Computing*, September 2004.

[16] W. Gu and M. Effros. On the concavity of lossless rate regions. In *Proceedings of the IEEE International Symposium on Information Theory*, Seattle, WA, June 2006.

[17] W. Gu and M. Effros. On the continuity of achievable rate regions for source coding over networks. In *Proceedings of the IEEE Information Theory Workshop*, Tahoe, CA, September 2007.

[18] W. Gu, M. Effros, and M. Bakshi. A continuity theory for lossless source coding over networks. *Proceedings of the Allerton Conference on Communication, Control, and Computing*, September 2008.

[19] I. Csiszár and J. Köner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* London, U.K.:Academic, 1981.

[20] R. Ahlswede, P. Gács, and J. Körner. Bounds on conditional probabilities with applications in multi-user communication. *Probability Theory and Related Fields*, 34(2):157–177, June 1976.

[21] Y. Oohama and T. Han. Universal coding for the Slepian-wolf data compression system and the strong converse theorem. *IEEE Transactions on Information Theory*, IT-40(6):1908–1919, November 1994.

[22] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, IT-18:14–20, January 1972.

[23] R.E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473, July 1972.

[24] F. Dupuis, W. Yu, and F. M. .J. Willems. Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information. In *Proceedings of the IEEE International Symposium on Information Theory*, Chicago, IL, June 2004.

[25] W. Gu, R. Koetter, M. Effros, and T. Ho. On source coding with coded side information for a binary source with binary side information. In *Proceedings of the IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[26] W. Gu and M. Effros. Approximation on computing rate regions for coded side

information problem. In *Proceedings of the IEEE Information Theory Workshop*, Tahoe, CA, September 2007.

[27] W. Gu, S. Jana, and M. Effros. On approximating the rate regions for lossy source coding with coded and uncoded side information. In *Proceedings of the IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.

[28] W. Gu and M. Effros. Source coding for a multihop network. In *Proceedings of the Data Compression Conference*, Snowbird, UT, March 2005.

[29] W. Gu and M. Effros. Source coding for a simple multi-hop network. In *Proceedings of the IEEE International Symposium on Information Theory*, Adelaide, Australia, September 2005.

[30] W. Gu and M. Effros. On multi-resolution coding and a two-hop network. In *Proceedings of the Data Compression Conference*, Snowbird, UT, March 2006.

[31] A. Sgarro. Source coding with side information at several decoders. *IEEE Transactions on Information Theory*, IT-23(2):179–182, March 1977.

[32] Y. Steinberg and N. Merhav. On successive refinement for the Wyner-Ziv problem. *IEEE Transactions on Information Theory*, IT-50(8):1636–1654, August 2004.

[33] T. Han and K. Kobayashi. A dichotomy of functions $f(x, y)$ of correlated sources (x, y). *IEEE Transactions on Information Theory*, IT-33(1):69–76, January 1987.

[34] J. Chen and A. B. Wagner. A continuity theorem and its applications to gaussian newtork source coding problems. In *Proceedings of the Information Theory and Applications*, San Diego, California, February 2008.

[35] Noga. Alon and A. Orlitsky. Source coding and graph entropies. *IEEE Transactions on Information Theory*, IT-42(5):1329–1339, September 1996.

[36] E. Tuncel and K. Rose. On variable-length coding of sources with side information at multiple decoders. In *Proceedings of the IEEE International Symposium on Information Theory*, page 140, Yokohama, Japan, June 2003.

[37] S. Robert Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, IT-49(2):371–381, February 2003.

[38] J. Körner and K. Marton. How to encode the modulo-two sum of binary sources. *IEEE Transactions on Information Theory*, 25(2):219–221, March 1979.

[39] I. Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Transactions on Information Theory*, IT-28(4):585–592, July 1982.

[40] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros. Separating distributed source coding from network coding. *IEEE Transactions on Information Theory*, IT-52(6):2785–2795, June 2006.

[41] J. C. Kieffer. Strong converses in source coding relative to a fidelity criterion. *IEEE Transactions on Information Theory*, IT-37(2):257–262, March 1991.

[42] J. C. Kieffer. Sample converses in source coding theory. *IEEE Transactions on Information Theory*, IT-37(2):263–268, March 1991.

[43] T. S. Han. *Information-Spectrum Method in Information Theory*. Springer-Verlag New Yort, LLC, 2003.

[44] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[45] S. Jana and R.E. Blahut. Achievable region for multiterminal source coding with lossless decoding in all sources except one. In *Proceedings of the IEEE Information Theory Workshop*, Tahoe, CA, September 2007.

[46] I. Csiszár. On the computation of the rate-distortion function. *IEEE Transactions on Information Theory*, IT-20:122–124, January 1974.

[47] T. Berger. Optimum quantizers and permutation codes. *IEEE Transactions on Information Theory*, 18(6):759–765, November 1972.

[48] S. Jana and R.E. Blahut. Partial side information problem: Equivalence of two inner bounds. In *Proceedings of the Conference on Information Sciences and Systems*, Princeton, NJ, March 2008.

[49] A. D. Wyner. On source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-21(3):294–300, November 1975.

[50] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[51] R. G. Gallager. *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., New York, 1968.

[52] R. M. Gray. A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions. *IEEE Transactions on Information Theory*, IT-19:480–489, July 1973.

[53] A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Transactions on Information Theory*, IT-28(6):851–857, November 1982.

[54] S. Nash and A. Sofer. *Linear and Nonlinear Programming.* McGraw-Hill, 1996.