University of California

Los Angeles

Two Problems in Multiterminal Information Theory

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

by

Thomas Alexander Courtade

2012

Abstract of the Dissertation

Two Problems in Multiterminal Information Theory

by

Thomas Alexander Courtade

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2012

Professor Richard D. Wesel, Chair

This dissertation consists of three parts. The first part considers the archetypical multiterminal source coding problem with logarithmic loss distortion constraints. A single-letter description of the achievable rate distortion region is given for finite-alphabet sources. In the course of doing so, the rate distortion region for the m-encoder CEO problem is also characterized. Several applications and examples are given, and a variety of related problems are discussed.

The second part of this dissertation considers the combinatorial problem of Coded Cooperative Data Exchange. In this problem, data which is originally distributed in a network is exchanged among nodes until universal recovery is achieved (i.e., all terminals recover all data initially present in the network). This dissertation characterizes the minimum number of exchanges which must take place in order to permit universal recovery. Explicit algorithms and tight concentration results are given for several special cases of interest.

Finally, three new lemmas are provided, each of which is interesting in its own right. Applications to multiterminal information theory are discussed.

The dissertation of Thomas Alexander Courtade is approved.

Suhas N. Diggavi

Benjamin Sudakov

Lieven Vandenberghe

Richard D. Wesel, Committee Chair

University of California, Los Angeles

2012

To Sasha

Table of Contents

List of Figures

Acknowledgments

This dissertation would not have been possible without the support of my family, colleagues, teachers, and friends. First and foremost, I want to thank my wife Sasha and my parents William and Beth, whose love, patience, and support were ever-present.

In the UCLA Department of Electrical Engineering, I want to thank my advisor Professor Richard Wesel, whose kindness and support enabled me to pursue a variety of interesting topics during my time as a graduate student. His enthusiasm for research was a great source of inspiration during the time I spent in the Communication Systems Laboratory. I also want to thank Professors Lieven Vandenberghe and Suhas Diggavi for the many positive interactions I've shared with them. While at UCLA, I have had the great fortune of knowing many other professors in the department; to all of you, I am grateful. Finally, I want to thank Deeona Columbia, Rose LaMountain, and all the other staff members whose services I have benefited from while a graduate student.

In the UCLA Department of Mathematics, I want to thank Professors James Ralston, Terence Tao, and Benny Sudakov, whose excellent courses in analysis and combinatorics played a crucial role in the development of my research and thinking.

Elsewhere, I want to thank Professor Alex Sprintson (Texas A&M) and Dariush Divsalar (JPL) for the many inspiring conversations we've shared. Dr. Thomas Halford (Trellisware Technologies) has become a colleague and friend; to him I am extremely grateful. Recently, I've had the great pleasure of working with Professor Tsachy Weissman (Stanford). Without Tsachy, the work on multiterminal source coding under logarithmic loss would have never come to fruition.

I have benefited in many ways from other students I've known. I am honored to have worked with Tsung-Yi Chen, Pavan Datta, Shijia Liu, Andrea Marinoni, Jason Pan, Zaid Towfic, Jiadong Wang, Bike Xie, and Yue Zhao, among others. To the new students in the

group – Adam Williamson, Kasra Vakilinia, and Yang Guo – I wish each of you the best of luck over the course of your graduate careers.

Finally, to my friends and colleagues at Rockwell Collins, I am eternally grateful. From 2004 to 2007, I did several internships with the Communication Systems group in the Advanced Technology Center. Those years provided me with a wealth of knowledge and experience in communication system design, which has influenced my path to the present day. I am especially thankful for my mentor Dana Jensen; if I have enjoyed any success in my career, it is due in no small part to him. After joining UCLA in 2007, Rockwell Collins proceeded to generously fund my research group, and continues to do so to the present day. Indeed, many of the projects which I have worked on as a graduate student were funded by them. Through this relationship, I have had the pleasure of working with Kent Benson, Carlos Chavez, Rob Frank, Mike Newhouse, and others. To each of them, I am grateful.

Almost surely, I have forgotten to include people in this list that deserve to be here. Please forgive the oversight.

Vita

2007	B.S. in Electrical Engineering Michigan Technological University
2007-2011	Research Assistant Department of Electrical Engineering, University of California, Los Angeles
2008	M.S. in Electrical Engineering University of California, Los Angeles
2008-2010	Teaching Assistant Department of Electrical Engineering, University of California, Los Angeles
2010-2011	Excellence in Teaching Award Department of Electrical Engineering, University of California, Los Angeles
2011-2012	Dissertation Year Fellow University of California, Los Angeles
2012	Ph.D. in Electrical Engineering University of California, Los Angeles
2012	Distinguished Ph.D. Dissertation Award in Signals & Systems Department of Electrical Engineering, University of California, Los Angeles

Selected Publications

T. A. Courtade and R. D. Wesel. “A Cross-Layer Perspective on Rateless Coding for Wireless Channels.” In Proceedings of the 2009 IEEE international conference on Communications. Dresden, Germany, June 2009.

T. A. Courtade and R. D. Wesel, "Efficient Universal Recovery in Broadcast Networks." In Proceedings of the Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2010.

T. A. Courtade and R. D. Wesel, "Optimal Allocation of Redundancy Between Packet-Level Erasure Coding and Physical-Layer Channel Coding in Fading Channels." IEEE Transactions on Communications, pages 2101 - 2109, August 2011.

T. A. Courtade and R. D.Wesel, "Multiterminal Source Coding with an Entropy-Based Distortion Measure." In Proceedings of the 2011 IEEE International Symposium on Information Theory, Saint-Petersburg, Russia, August 2011.

T. A. Courtade, J. Wang, and R. D. Wesel, "Superposition Coding to Support Multiple Streams, Priorities, and Channel Capacities in the Context of GMSK." In Proceedings of the 2011 Military Conference on Communications, Baltimore, MD, November 2011.

T. A. Courtade and T. Weissman, "Multiterminal Source Coding under Logarithmic Loss". CoRR, abs/1110.3069, 2011.

T. A. Courtade and R. D. Wesel, "Coded Cooperative Data Exchange in Multihop Networks". CoRR, abs/1203.3445, 2012.

T. A. Courtade, "Information Masking and Amplification: The Source Coding Setting". In Proceedings of the IEEE International Symposium on Information Theory, Cambridge, MA, July 1 - 6, 2012.

T. A. Courtade and T. Weissman, "Multiterminal Source Coding under Logarithmic Loss". In Proceedings of the IEEE International Symposium on Information Theory, Cambridge, MA, July 1 - 6, 2012.

CHAPTER 1

Introduction

The fundamental tradeoff between the level to which data can be compressed, and the fidelity to which it can later be reproduced, is known as rate distortion[1]. As with many of the cornerstone results in information theory, the study of the rate distortion tradeoff traces its roots back to Claude Shannon. Just over fifty years ago, Shannon proved the first rate distortion theorem in his 1959 paper [Sha59]. It is surprising then, that apparently simple extensions of this fundamental tradeoff to multiterminal settings are notoriously difficult to describe.

To wit, characterizing the rate distortion region — the set of attainable rates and distortions — for the two-encoder source coding problem depicted in Figure 1.1 is perhaps the most well-known, long-standing open problem in the field of multiterminal source coding. Indeed, it is commonly referred to as *the* multiterminal source coding problem (a tradition to which we adhere in this dissertation). Although this problem was posed nearly four decades ago, a complete description of the rate distortion region eluded researchers for any nontrivial choice of source distribution and distortion measure until very recently.

In their seminal work [WTV08], Wagner et al. characterized the rate distortion region for jointly Gaussian sources subject to mean square error distortion constraints. Notably, they showed that the extension of the single-encoder vector quantization scheme to two encoders suffices to attain any point in the achievable rate distortion region. We remark that, as Shannon showed in his 1959 paper, this type of vector quantization encoding scheme attains any point of the optimal rate distortion tradeoff in the single-encoder setting. The

[1] The term *rate* refers to the compression ratio, while *distortion* refers to the reproduction fidelity.

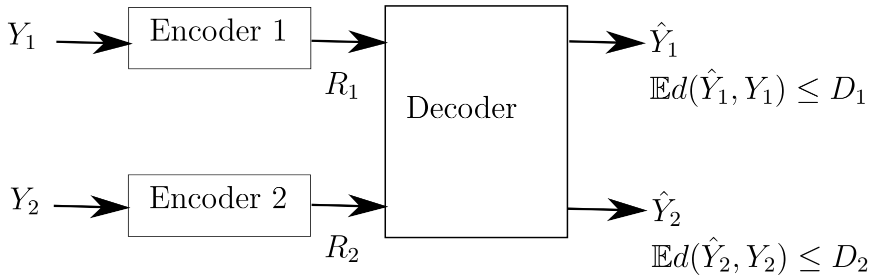
Y_1
Encoder 1
R_1
Decoder
$\hat{Y}_1$
$\mathbb{E}d(\hat{Y}_1, Y_1) \le D_1$
Y_2
Encoder 2
R_2
$\hat{Y}_2$
$\mathbb{E}d(\hat{Y}_2, Y_2) \le D_2$

Figure 1.1: The two-encoder lossy source coding network.

two-encoder extension of this scheme (where both encoders perform vector quantization) is commonly called the Berger-Tung scheme, for the researchers who first proposed and analyzed it (see [Ber77] and [Tun78]).

While the importance of [WTV08] is undeniable, the proofs contained therein relied heavily on the maximum entropy properties of the Gaussian distribution and the vast body of literature specific to Gaussian multiterminal problems (e.g., [Ooh97, PTR04, Ooh05]). Due to the inherent dependence on the peculiarities of the Gaussian distribution, the results in [WTV08] shed little light on the multiterminal source coding problem for other source distributions and/or choices of distortion measures. Indeed, it was still not clear from [WTV08] whether the Berger-Tung achievability scheme — a very natural encoding scheme — would be optimal in other settings of interest. In fact, Wagner et al. later gave a relatively simple example in [WKA11] which suggested the possibility that the Berger-Tung achievability scheme might not be optimal for finite-alphabet sources except in very special cases (e.g., where one source is to be reproduced losslessly).

In the first part of this dissertation, we address this point for the two-encoder setting. Specifically, we show that the Berger-Tung achievability scheme is optimal for all finite-alphabet sources when distortion is measured under logarithmic loss. To our knowledge, this constitutes the first time that the entire achievable rate distortion region has been described for the multiterminal source coding problem with nontrivial finite-alphabet sources and nontrivial distortion constraints. In addition to providing a single-letter characterization of the achievable rate distortion region for the multiterminal source coding problem under logarithmic loss, we also give a complete characterization of the m-encoder CEO problem (also under logarithmic loss). The CEO problem, like the multiterminal source coding problem, is another well-known – and generally open – problem in the field of network information theory.

For the reader unfamiliar with the distortion measure we consider, we briefly motivate it here. Logarithmic loss is a widely used penalty function in the theory of learning and prediction (cf. [CL06, Chapter 9]), and it has natural interpretations and applications in gambling and portfolio theory. To be more precise, logarithmic loss is a natural loss criterion in scenarios where the reconstructions are allowed to be 'soft', i.e., probability measures rather than deterministic decision values. As we demonstrate through several examples, logarithmic loss has a variety of useful applications and has implicitly appeared in the source coding literature frequently over the years (popular examples include horse racing markets and list decoding). However, it was not explicitly introduced in the context of multiterminal source coding until [CW11a].

In contrast to the lossy setting investigated in the first part of this dissertation, we restrict our attention to lossless reproduction of data in the second part. In making this sacrifice, we are able to study networks which are essentially arbitrarily connected, as opposed to the simple three-terminal network given in Figure 1.1. To be specific, we study the problem of *coded cooperative data exchange*. This problem is predominantly motivated by emerging issues in applied information theory, particularly in distributed storage and peer-to-peer networks. In our model, we assume that each node in a network starts with a subset of data. Nodes propagate data through the network by exchanging messages with their neighbors. We quantify how many such exchanges are required to achieve universal recovery, i.e., the state where each node recovers all data initially present in the network.

To see the motivation for this work, consider the task of backing up data on servers in a large data center. One common method to protect data from corruption is replication.

As the name suggests, under replication, large quantities of data are replicated in several locations so as to protect from various sources of corruption (e.g., equipment failure, power outages, natural disasters, etc.). As the quantity of information in large data centers continues to increase, the number of file transfers required to complete a periodic replication task is becoming an increasingly important consideration due to time, equipment, cost, and energy constraints. Our work in the second part of this dissertation addresses these applications.

In addition to distributed storage, there are potential applications to tactical networks. For instance, consider a scenario in which an aircraft flies over a group of nodes on the ground and tries to deliver a video stream. Each ground node might only receive a subset of the packets due to interference, obstructions, and other signal integrity issues. In order to recover the transmission, the nodes are free to communicate with their neighbors, but would like to minimize the number of transmissions in order to conserve battery power (or avoid detection, etc.). How should the nodes share information, and what is the minimum number of transmissions required so that the entire network can recover the video stream? These are the essential questions addressed.

The third and final part of this dissertation is of a technical nature. In it, we present three lemmas, each of which has applications to multiterminal information theory. The reason for separating these results from the rest of the material in this dissertation is that we feel each of these results is interesting in its own right. As such, we chose to highlight them since they may be useful in other areas.

1.1 Summary of Contributions

We briefly outline the contributions of each chapter below. We remark that Chapters 2 and 3 are largely independent from one another, but both require results which are given in Chapter 4. A background in information theory will be helpful to the reader interested in Chapter 2, while no specific background is necessary for the material in Chapter 3.

Chapter 2 Contributions

In Chapter 2, we study multiterminal source coding problems under logarithmic loss. We give a single-letter characterization of the achievable rate distortion region for the m-encoder CEO problem and the two-encoder lossy source coding problem when distortion is measured under logarithmic loss. Notably, we make no assumptions on the source distributions, other than that they have finite alphabets. For each of these problems, we obtain a strengthened converse in the sense that augmenting the reproduction alphabet does not enlarge the achievable rate distortion region.

In addition, we give a single-letter description for a fundamental entropy characterization problem. Specifically, if correlated sources are separately encoded by rate-limited encoders, we give a precise characterization of the amount of information that can be revealed about each source given the encoded representations. A consequence of this result is that relatively simple encoding functions suffice to attain any so-called achievable information pair for given rates.

Throughout Chapter 2, we give several examples and applications of logarithmic loss as a distortion measure. Specifically, we prove results related to horse racing markets, list decoding, estimating a posterior distribution, and the general multiterminal source coding problem with arbitrary distortion measures. Additionally, we characterize the achievable rate distortion regions for the multiple description problem and the two-way lossy source coding problem under logarithmic loss.

Chapter 3 Contributions

Chapter 3 presents our results on coded cooperative data exchange. To begin, we prove necessary and sufficient conditions for achieving universal recovery in an arbitrarily connected network. When the network has unit diameter (i.e., is fully connected), we provide an efficient algorithm based on submodular optimization that computes an optimal transmission schedule for the weighted cooperative data exchange problem.

When packets are randomly distributed amongst the nodes and the network topology satisfies certain regularity conditions, we prove tight concentration results on the number of transmissions required to achieve universal recovery. The interval of concentration is independent of the number of packets in the network, and we prove concentration on a single value in the single-hop setting. Moreover, we show that splitting packets does not significantly reduce the number of transmissions required to achieve universal recovery.

Finally, we relate our results to the task of distributed secrecy generation amongst a collection of nodes in the presence of an eavesdropper. Specifically, we show that the submodular optimization algorithm we present can be used to compute and generate the theoretical maximum amount of secrecy in a practical manner among nodes in a fully connected network.

Chapter 4 Contributions

In Chapter 4, we prove three lemmas, each of which has applications to multiterminal information theory. Roughly speaking, the first lemma shows that there exists an optimum ℓ_1-solution to a linear program with an underdetermined set of constraints which behaves similarly to the linear least-squares solution to a related linear system.

The second lemma is an existence result which states that, under certain assumptions, a point-wise inequality relating two different convex combinations can be "amplified" to yield an apparently stronger set of inequalities.

Finally, the third result in Chapter 4 is a concentration result. Informally, it states that if n line segments are randomly placed in the unit square, we can typically find a set of $c\sqrt{n}$ line segments which don't intersect (c is an absolute constant). While this result, unlike the other two, is not used elsewhere in this dissertation, it can be used to characterize the number of non-intersecting traffic flows in a network described by a random geometric graph.

CHAPTER 2

Multiterminal Source Coding under Logarithmic Loss

2.1 Introduction

A complete characterization of the achievable rate distortion region for the two-encoder source coding problem depicted in Figure 2.1 has remained an open problem for over three decades. Following tradition, we will refer to this two-encoder source coding network as the *multiterminal source coding problem* throughout this dissertation. Several special cases have been solved for general source alphabets and distortion measures:

- The lossless case where $D_1 = 0, D_2 = 0$. Slepian and Wolf solved this case in their seminal work [SW73].

- The case where one source is recovered losslessly: i.e., $D_1 = 0, D_2 = D_{\max}$. This case corresponds to the source coding with side information problem of Ahlswede-Körner-Wyner [AK75, Wyn75].

- The Wyner-Ziv case [WZ76] where Y_2 is available to the decoder as side information and Y_1 should be recovered with distortion at most D_1.

- The Berger-Yeung case [BY89] where D_1 is arbitrary and $D_2 = 0$. This subsumes the previous three cases

Despite the apparent progress, other seemingly fundamental cases, such as when D_1 is arbitrary and $D_2 = D_{\max}$, remain unsolved except perhaps in very special cases.

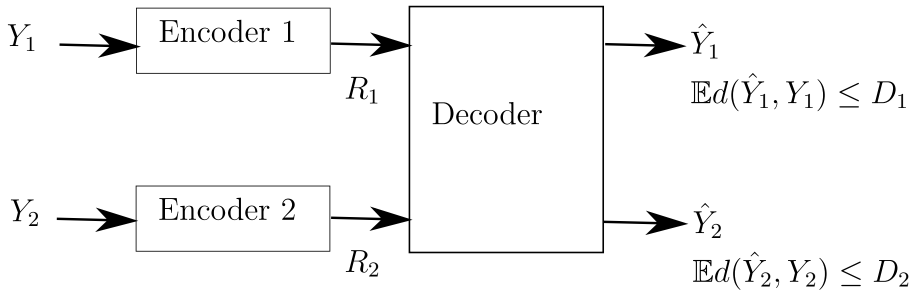
Y_1
Encoder 1
R_1
Y_2
Encoder 2
R_2
Decoder
$\hat{Y}_1$
$\mathbb{E}d(\hat{Y}_1, Y_1) \leq D_1$
$\hat{Y}_2$
$\mathbb{E}d(\hat{Y}_2, Y_2) \leq D_2$

Figure 2.1: The multiterminal source coding network.

Recently, the achievable rate distortion region for the quadratic Gaussian multiterminal source coding problem was given by Wagner, Tavildar, and Viswanath in [WTV08]. Until now, this was the only case for which the entire achievable rate distortion region was known. While this is a very important result, it is again a special case from a theoretical point of view: a specific choice of source distribution, and a specific choice of distortion measure.

In the present chapter, we determine the achievable rate distortion region of the multiterminal source coding problem for arbitrarily correlated sources with finite alphabets. However, as in [WTV08], we restrict our attention to a specific distortion measure.

At a high level, the roadmap for our argument is similar to that of [WTV08]. In particular, both arguments couple the multiterminal source coding problem to a parametrized family of CEO problems. Then, the parameter in the CEO problem is "tuned" to yield the converse result. Despite this apparent similarity, the proofs in [WTV08] rely heavily on the calculus performed on the closed-form entropy expressions which arise from the Gaussian source assumption. In our case we do not have this luxury, and our CEO tuning argument essentially relies on an existence result to yield the converse result. The success of our approach is largely due to the fact that the distortion measure we consider admits a lower bound in the form of a conditional entropy, much like the quadratic distortion measure for Gaussian sources.

2.1.1 Our Contributions

In this chapter, we characterize the achievable rate distortion region for the multiterminal source coding problem under logarithmic loss. In the process of accomplishing this, we also obtain a description of the achievable rate distortion region for the CEO problem, also under logarithmic loss. In both settings, we obtain a stronger converse than is standard for rate distortion problems in the sense that augmenting the reproduction alphabet does not enlarge the rate distortion region. Notably, we make no assumptions on the source distributions, other than that the sources have finite alphabets. In both cases, the Berger-Tung inner bound on the rate distortion region is tight. To the best of our knowledge, this constitutes the first time that the entire achievable rate distortion region has been described for general discrete alphabets under a non-trivial distortion constraint.

2.1.2 Organization

This chapter is organized as follows. In Section 2.2 we formally define the logarithmic loss function and the multiterminal source coding problem we consider. In Section 2.3 we define the CEO problem and give the rate distortion region under logarithmic loss. In Section 2.4 we return to the multiterminal source coding problem and derive the rate distortion region for the two-encoder setting. Applications to estimation, horse racing, and list decoding are given in Sections 2.3 and 2.4. In Section 2.5, we discuss connections between our results and the multiterminal source coding problem with arbitrary distortion measures. Section 2.6 considers a related source coding problem of compressing information subject to amplification and equivocation constraints. In Section 2.7, we briefly consider two other source coding problems under logarithmic loss: the multiple description problem, and the two-way lossy source coding problem. Section 2.8 delivers our concluding remarks and discusses directions for future work.

2.2 Problem Definition

Throughout this chapter, we adopt notational conventions that are standard in the literature. Specifically, random variables are denoted by capital letters (e.g., X) and their corresponding alphabets are denoted by corresponding calligraphic letters (e.g., $\mathcal{X}$). We abbreviate a sequence $(X_1, X_2, \ldots, X_n)$ of n random variables by X^n, and we denote the interval $(X_k, X_{k+1}, \ldots, X_j)$ by X_k^j. If the lower index is equal to 1, it will be omitted when there is no ambiguity (e.g., $X^j \triangleq X_1^j$). Frequently, random variables will appear with two subscripts (e.g., $Y_{i,j}$). In this case, we are referring to the j^{th} instance of random variable Y_i. We overload our notation here slightly in that $Y_{i,1}^j$ is often abbreviated as Y_i^j. However, our meaning will always be clear from context.

Let $\{(Y_{1,j}, Y_{2,j})\}_{j=1}^n = (Y_1^n, Y_2^n)$ be a sequence of n independent, identically distributed random variables with finite alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively and joint pmf $p(y_1, y_2)$. That is, $(Y_1^n, Y_2^n) \sim \prod_{i=1}^n p(y_{1,j}, y_{2,j})$.

In this chapter, we take the reproduction alphabet $\hat{\mathcal{Y}}_i$ to be equal to the set of probability distributions over the source alphabet $\mathcal{Y}_i$ for $i = 1, 2$. Thus, for a vector $\hat{Y}_i^n \in \hat{\mathcal{Y}}_i^n$, we will use the notation $\hat{Y}_{i,j}(y_i)$ to mean the j^{th} coordinate $(1 \le j \le n)$ of $\hat{Y}_i^n$ (which is a probability distribution on $\mathcal{Y}_i$) evaluated for the outcome $y_i \in \mathcal{Y}_i$. In other words, the decoder generates 'soft' estimates of the source sequences.

We will consider the logarithmic loss distortion measure defined as follows:

$$d(y_i, \hat{y}_i) = \log\left(\frac{1}{\hat{y}_i(y_i)}\right) \text{ for } i = 1, 2.$$

In particular, $d(y_i, \hat{y}_i)$ is the relative entropy (i.e., Kullback-Leibler divergence) between the empirical distribution of the event $\{Y_i = y_i\}$ and the estimate $\hat{y}_i$. Using this definition for symbol-wise distortion, it is standard to define the distortion between sequences as

$$d(y_i^n, \hat{y}_i^n) = \frac{1}{n}\sum_{j=1}^n d(y_{i,j}, \hat{y}_{i,j}) \text{ for } i = 1, 2.$$

We point out that the logarithmic loss function is a widely used penalty function in the theory of learning and prediction (cf. [CL06, Chapter 9]). Further, it is a particularly natural loss criterion in settings where the reconstructions are allowed to be ‘soft’, rather than deterministic values. Surprisingly, since distributed learning and estimation problems are some of the most oft-cited applications of lossy multiterminal source coding, it does not appear to have been studied in this context until the recent work [CW11a]. However, we note that this connection has been established previously for the single-encoder case in the study of the information bottleneck problem [HT07]. Beyond learning and prediction, a similar distortion measure has appeared before in the image processing literature [AAB06]. As we demonstrate through several examples, the logarithmic loss distortion measure has a variety of useful applications in the context of multiterminal source coding.

A rate distortion code (of blocklength n) consists of encoding functions:

$$g_i^{(n)} : \mathcal{Y}_i^n \to \left\{1, \dots, M_i^{(n)}\right\} \text{ for } i = 1, 2$$

and decoding functions

$$\psi_i^{(n)} : \left\{1, \dots, M_1^{(n)}\right\} \times \left\{1, \dots, M_2^{(n)}\right\} \to \hat{\mathcal{Y}}_i^n \text{ for } i = 1, 2.$$

A rate distortion vector (R_1, R_2, D_1, D_2) is strict-sense achievable if there exists a blocklength n, encoding functions $g_1^{(n)}, g_2^{(n)}$ and a decoder $(\psi_1^{(n)}, \psi_2^{(n)})$ such that

$$\begin{aligned} R_i &\geq \frac{1}{n} \log M_i^{(n)} \text{ for } i = 1, 2 \\ D_i &\geq \mathbb{E} d(Y_i^n, \hat{Y}_i^n) \text{ for } i = 1, 2. \end{aligned}$$

Where

$$\hat{Y}_i^n = \psi_i^{(n)}(g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n)) \text{ for } i = 1, 2.$$

Definition 1. *Let $\mathcal{RD}^\star$ denote the set of strict-sense achievable rate distortion vectors and define the set of achievable rate distortion vectors to be its closure, $\overline{\mathcal{RD}^\star}$.*

Our ultimate goal in the present chapter is to give a single-letter characterization of the region $\overline{\mathcal{RD}^\star}$. However, in order to do this, we first consider an associated CEO problem. In this sense, the roadmap for our argument is similar to that of [WTV08]. Specifically, both arguments couple the multiterminal source coding problem to a parametrized family of CEO problems. Then, the parameter in the CEO problem is “tuned” to yield the converse result. Despite this apparent similarity, the proofs are quite different since the results in [WTV08] depend heavily on the peculiarities of the Gaussian distribution.

2.3 The CEO problem

In order to attack the general multiterminal problem, we begin by studying the CEO problem (See [BZV96] for an introduction.). To this end, let $\{(X_j, Y_{1,j}, Y_{2,j})\}_{j=1}^n = (X^n, Y_1^n, Y_2^n)$ be a sequence of n independent, identically distributed random variables distributed according to the joint pmf $p(x, y_1, y_2) = p(x)p(y_1|x)p(y_2|x)$. That is, $Y_1 \leftrightarrow X \leftrightarrow Y_2$ form a Markov chain, in that order.

In this section, we consider the reproduction alphabet $\hat{\mathcal{X}}$ to be equal to the set of probability distributions over the source alphabet $\mathcal{X}$. As before, for a vector $\hat{X}^n \in \hat{\mathcal{X}}^n$, we will use the notation $\hat{X}_j(x)$ to mean the j^{th} coordinate of $\hat{X}^n$ (which is a probability distribution on $\mathcal{X}$) evaluated for the outcome $x \in \mathcal{X}$. As in the rest of this chapter, $d(\cdot, \cdot)$ is the logarithmic loss distortion measure.

A rate distortion CEO code (of blocklength n) consists of encoding functions:

$$g_i^{(n)} : \mathcal{Y}_i^n \to \left\{1, \dots, M_i^{(n)}\right\} \text{ for } i = 1, 2$$

and a decoding function

$$\psi^{(n)} : \left\{1, \dots, M_1^{(n)}\right\} \times \left\{1, \dots, M_2^{(n)}\right\} \to \hat{\mathcal{X}}^n.$$

A rate distortion vector (R_1, R_2, D) is strict-sense achievable for the CEO problem if there exists a blocklength n, encoding functions $g_1^{(n)}, g_2^{(n)}$ and a decoder $\psi^{(n)}$ such that

$$R_i \geq \frac{1}{n} \log M_i^{(n)} \text{ for } i = 1, 2$$
$$D \geq \mathbb{E} d(X^n, \hat{X}^n).$$

Where

$$\hat{X}^n = \psi^{(n)}(g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n)).$$

Definition 2. *Let $\mathcal{RD}^{\star}_{CEO}$ denote the set of strict-sense achievable rate distortion vectors and define the set of achievable rate distortion vectors to be its closure, $\overline{\mathcal{RD}}^{\star}_{CEO}$.*

2.3.1 Inner Bound

Definition 3. *Let $(R_1, R_2, D) \in \mathcal{RD}^{i}_{CEO}$ if and only if there exists a joint distribution of the form*

$$p(x, y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)p(q)$$

where $|\mathcal{U}_1| \leq |\mathcal{Y}_1|$, $|\mathcal{U}_2| \leq |\mathcal{Y}_2|$, and $|\mathcal{Q}| \leq 4$, which satisfies

$$\begin{aligned} R_1 &\geq I(U_1; Y_1|U_2, Q) \\ R_2 &\geq I(U_2; Y_2|U_1, Q) \\ R_1 + R_2 &\geq I(U_1, U_2; Y_1, Y_2|Q) \\ D &\geq H(X|U_1, U_2, Q). \end{aligned}$$

Theorem 1. *$\mathcal{RD}^{i}_{CEO} \subseteq \overline{\mathcal{RD}}^{\star}_{CEO}$. That is, all rate distortion vectors $(R_1, R_2, D) \in \mathcal{RD}^{i}_{CEO}$ are achievable.*

Before proceeding with the proof, we cite the following variant of a well-known inner bound:

Proposition 1 (Berger-Tung Inner Bound [Ber77])**.** *The rate distortion vector (R_1, R_2, D) is achievable if*

$$\begin{aligned} R_1 &\geq I(U_1; Y_1|U_2, Q) \\ R_2 &\geq I(U_2; Y_2|U_1, Q) \\ R_1 + R_2 &\geq I(U_1, U_2; Y_1, Y_2|Q) \\ D &\geq \mathbb{E}\left[d(X, f(U_1, U_2, Q)\right] \end{aligned}$$

for a joint distribution

$$p(x)p(y_1|x)p(y_2|x)p(u_1|y_1,q)p(u_2|y_2,q)p(q)$$

and reproduction function

$$f : \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \hat{\mathcal{X}}.$$

The proof of this proposition is a standard exercise in information theory, and is therefore omitted. The interested reader is directed to the text [EK12] for a modern, detailed treatment. The proposition follows from what is commonly called the Berger-Tung achievability scheme. In this encoding scheme, each encoder quantizes its observation Y_i^n to a codeword U_i^n, such that the empirical distribution of the entries in (Y_i^n, U_i^n) is very close to the true distribution $p(y_i, u_i)$. In order to communicate their respective quantizations to the decoder, the encoders essentially perform Slepian-Wolf coding. For this reason, the Berger-Tung achievability scheme is also referred to as a "quantize-and-bin" coding scheme.

Proof of Theorem 1. Given Proposition 1, the proof of Theorem 1 is immediate. Indeed, if we apply Proposition 1 with the reproduction function $f(U_1,U_2,Q) \triangleq \Pr\left[X = x|U_1,U_2,Q\right]$, we note that

$$\mathbb{E}\left[d(X, f(U_1,U_2,Q)\right] = H(X|U_1,U_2,Q),$$

which yields the desired result. □

2.3.2 A Matching Outer Bound

A particularly useful property of the logarithmic loss distortion measure is that the expected distortion is lower-bounded by a conditional entropy. A similar property is enjoyed by Gaussian random variables under quadratic distortion. In particular, if G is Gaussian, and $\hat{G}$ is such that $\mathbb{E}(\hat{G} - G)^2 \leq D$, then $\frac{1}{2}\log(2\pi e)D \geq h(G|\hat{G})$. The case for logarithmic loss is similar, and we state it formally in the following lemma which is crucial in the proof of the converse.

Lemma 1. *Let $Z = (g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n))$ be the argument of the reproduction function $\psi^{(n)}$. Then $n\mathbb{E}d(X^n, \hat{X}^n) \geq H(X^n|Z)$.*

Proof. By definition of the reproduction alphabet, we can consider the reproduction $\hat{X}^n$ to be a probability distribution on $\mathcal{X}^n$ conditioned on the argument Z. In particular, if $\hat{x}^n = \psi^{(n)}(z)$, define $s(x^n|z) \triangleq \prod_{j=1}^n \hat{x}_j(x_j)$. It is readily verified that s is a probability measure on $\mathcal{X}^n$. Then, we obtain the following lower bound on the expected distortion conditioned on $Z = z$:

$$\begin{aligned}
\mathbb{E}\left[d(X^n, \hat{X}^n)|Z = z\right] &= \frac{1}{n}\sum_{j=1}^{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\log\left(\frac{1}{\hat{x}_j(x_j)}\right) \\
&= \frac{1}{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\sum_{j=1}^{n}\log\left(\frac{1}{\hat{x}_j(x_j)}\right) \\
&= \frac{1}{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\log\left(\frac{1}{s(x^n|z)}\right) \\
&= \frac{1}{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\log\left(\frac{p(x^n|z)}{s(x^n|z)}\right) + \frac{1}{n}H(X^n|Z = z) \\
&= \frac{1}{n}D\left(p(x^n|z)\|s(x^n|z)\right) + \frac{1}{n}H(X^n|Z = z) \\
&\geq \frac{1}{n}H(X^n|Z = z),
\end{aligned}$$

where $p(x^n|z) = \Pr\left(X^n = x^n|Z = z\right)$ is the true conditional distribution. Averaging both sides over all values of Z, we obtain the desired result. □

Definition 4. *Let $(R_1, R_2, D) \in \mathcal{RD}_{CEO}^o$ if and only if there exists a joint distribution of the form*

$$p(x)p(y_1|x)p(y_2|x)p(u_1|y_1, q)p(u_2|y_2, q)p(q),$$

which satisfies

$$\left.\begin{aligned} R_1 &\geq I(U_1;Y_1|X,Q) + H(X|U_2,Q) - D \\ R_2 &\geq I(U_2;Y_2|X,Q) + H(X|U_1,Q) - D \\ R_1 + R_2 &\geq I(U_1;Y_1|X,Q) + I(U_2;Y_2|X,Q) + H(X) - D \\ D &\geq H(X|U_1,U_2,Q). \end{aligned}\right\} \tag{2.1}$$

Theorem 2. *If (R_1, R_2, D) is strict-sense achievable for the CEO problem, then $(R_1, R_2, D) \in \mathcal{RD}^o_{CEO}$.*

Proof. Suppose the point (R_1, R_2, D) is strict-sense achievable. Let A be a nonempty subset of $\{1, 2\}$, and let $F_i = g_i^{(n)}(Y_i^n)$ be the message sent by encoder $i \in \{1, 2\}$. Define $U_{i,j} \triangleq (F_i, Y_i^{j-1})$ and $Q_j \triangleq (X^{j-1}, X_{j+1}^n) = X^n \backslash X_j$. To simplify notation, let $Y_A = \cup_{i \in A} Y_i$ (similarly for U_A and F_A).

With these notations established, we have the following string of inequalities:

$$
\begin{aligned}
n\sum_{i\in A} R_i &\geq \sum_{i\in A} H(F_i) \\
&\geq H(F_A) \\
&\geq I(Y_A^n; F_A|F_{A^c}) \\
&= I(X^n, Y_A^n; F_A|F_{A^c}) && (2.2)\\
&= I(X^n; F_A|F_{A^c}) + \sum_{i\in A} I(F_i; Y_i^n|X^n) && (2.3)\\
&= H(X^n|F_{A^c}) - H(X^n|F_1, F_2) + \sum_{i\in A}\sum_{j=1}^{n} I(Y_{i,j}; F_i|X^n, Y_i^{j-1}) \\
&\geq H(X^n|F_{A^c}) + \sum_{i\in A}\sum_{j=1}^{n} I(Y_{i,j}; F_i|X^n, Y_i^{j-1}) - nD && (2.4)\\
&= \sum_{j=1}^{n} H(X_j|F_{A^c}, X^{j-1}) + \sum_{i\in A}\sum_{j=1}^{n} I(Y_{i,j}; F_i|X^n, Y_i^{j-1}) - nD \\
&= \sum_{j=1}^{n} H(X_j|F_{A^c}, X^{j-1}) + \sum_{i\in A}\sum_{j=1}^{n} I(Y_{i,j}; U_{i,j}|X_j, Q_j) - nD && (2.5)\\
&\geq \sum_{j=1}^{n} H(X_j|U_{A^c,j}, Q_j) + \sum_{i\in A}\sum_{j=1}^{n} I(Y_{i,j}; U_{i,j}|X_j, Q_j) - nD. && (2.6)
\end{aligned}
$$

The nontrivial steps above can be justified as follows:

- (2.2) follows since F_A is a function of Y_A^n.
- (2.3) follows since F_i is a function of Y_i^n and $F_1 \leftrightarrow X^n \leftrightarrow F_2$ form a Markov chain (since $Y_1^n \leftrightarrow X^n \leftrightarrow Y_2^n$ form a Markov chain).
- (2.4) follows since $nD \geq H(X^n|F_1, F_2)$ by Lemma 1.
- (2.5) follows from the Markov chain $Y_{i,j} \leftrightarrow X^n \leftrightarrow Y_i^{j-1}$, which follows from the i.i.d. nature of the source sequences.
- (2.6) simply follows from the fact that conditioning reduces entropy.

Therefore, dividing both sides by n, we have:

$$\sum_{i\in A} R_i \geq \frac{1}{n}\sum_{j=1}^{n} H(X_j|U_{A^c,j},Q_j) + \sum_{i\in A}\frac{1}{n}\sum_{j=1}^{n} I(Y_{i,j};U_{i,j}|X_j,Q_j) - D.$$

Also, using Lemma 1 and the fact that conditioning reduces entropy:

$$D \geq \frac{1}{n}H(X^n|F_1,F_2) \geq \frac{1}{n}\sum_{j=1}^{n} H(X_j|U_{1,j},U_{2,j},Q_j).$$

Observe that Q_j is independent of $(X_j, Y_{1,j}, Y_{2,j})$ and, conditioned on Q_j, we have the long Markov chain $U_{1,j} \leftrightarrow Y_{1,j} \leftrightarrow X_j \leftrightarrow Y_{2,j} \leftrightarrow U_{2,j}$. Finally, by a standard time-sharing argument, we conclude by saying that if (R_1, R_2, D) is strict-sense achievable for the CEO problem, then

$$\begin{aligned} R_1 &\geq I(U_1;Y_1|X,Q) + H(X|U_2,Q) - D \\ R_2 &\geq I(U_2;Y_2|X,Q) + H(X|U_1,Q) - D \\ R_1 + R_2 &\geq I(U_1;Y_1|X,Q) + I(U_2;Y_2|X,Q) + H(X) - D \\ D &\geq H(X|U_1,U_2,Q). \end{aligned}$$

for some joint distribution $p(q)p(x,y_1,y_2)p(u_1|y_1,q)p(u_2|y_2,q)$. □

Theorem 3. $\mathcal{RD}^o_{CEO} = \mathcal{RD}^i_{CEO} = \overline{\mathcal{RD}}^\star_{CEO}$.

Proof. We first remark that the cardinality bounds on the alphabets in the definition of $\mathcal{RD}^i_{CEO}$ can be imposed without any loss of generality. This is a consequence of [Jan09, Lemma 2.2] and is discussed in detail in Appendix 2.A.

Therefore, we can ignore the cardinality constraints while showing $\mathcal{RD}^o_{CEO} \subseteq \mathcal{RD}^i_{CEO}$ to prove the theorem. To this end, fix $p(q)$, $p(u_1|y_1,q)$, and $p(u_2|y_2,q)$ and consider the extreme points[1] of polytope defined by the inequalities (2.1):

$$\begin{aligned}
P_1 &= \Big(0, 0, I(Y_1; U_1|X, Q) + I(Y_2; U_2|X, Q) + H(X)\Big)\\
P_2 &= \Big(I(Y_1; U_1|Q), 0, I(U_2; Y_2|X, Q) + H(X|U_1, Q)\Big)\\
P_3 &= \Big(0, I(Y_2; U_2|Q), I(U_1; Y_1|X, Q) + H(X|U_2, Q)\Big)\\
P_4 &= \Big(I(Y_1; U_1|Q), I(Y_2; U_2|U_1, Q), H(X|U_1, U_2, Q)\Big)\\
P_5 &= \Big(I(Y_1; U_1|U_2, Q), I(Y_2; U_2|Q), H(X|U_1, U_2, Q)\Big),
\end{aligned}$$

where the point P_j is a triple $(R_1^{(j)}, R_2^{(j)}, D^{(j)})$. We say a point $(R_1^{(j)}, R_2^{(j)}, D^{(j)})$ is *dominated by* a point in $\mathcal{RD}^i_{CEO}$ if there exists some $(R_1, R_2, D) \in \mathcal{RD}^i_{CEO}$ for which $R_1 \leq R_1^{(j)}$, $R_2 \leq R_2^{(j)}$, and $D \leq D^{(j)}$. Observe that each of the extreme points $P_1, \ldots, P_5$ is dominated by a point in $\mathcal{RD}^i_{CEO}$:

- First, observe that P_4 and P_5 are both in $\mathcal{RD}^i_{CEO}$, so these points are not problematic.
- Next, observe that the point $(0, 0, H(X))$ is in $\mathcal{RD}^i_{CEO}$, which can be seen by setting all auxiliary random variables to be constant. This point dominates P_1.
- By using auxiliary random variables $(\hat{U}_1, \hat{U}_2, Q) = (U_1, \emptyset, Q)$, the point $(I(Y_1; U_1|Q), 0, H(X|U_1, Q))$ is in $\mathcal{RD}^i_{CEO}$, and dominates the point P_2. By a symmetric argument, the point P_3 is also dominated by a point in $\mathcal{RD}^i_{CEO}$.

Since $\mathcal{RD}^o_{CEO}$ is the convex hull of all such extreme points (i.e., the convex hull of the union of extreme points over all appropriate joint distributions), the theorem is proved. □

Remark 1. *Theorem 3 can be extended to the general case of m-encoders. Details are provided in Appendix 2.B.*

[1] For two encoders, it is easy enough to enumerate the extreme points by inspection. However, this can be formalized by a submodularity argument, which is given in Appendix 2.B.

2.3.3 A stronger converse result for the CEO problem

As defined, our reproduction sequence $\hat{X}^n$ is restricted to be a product distribution on $\mathcal{X}^n$. However, for a blocklength n code, we can allow $\hat{X}^n$ to be *any* probability distribution on $\mathcal{X}^n$ and the converse result still holds. In this case, we define the sequence distortion as follows:

$$d(x^n, \hat{x}^n) = \frac{1}{n} \log \left(\frac{1}{\hat{x}^n(x^n)} \right),$$

which is compatible with the original definition when $\hat{X}^n$ is a product distribution. The reader can verify that the result of Lemma 1 is still true for this more general distortion alphabet by setting $s(x^n|z) = \hat{x}^n(x^n)$ in the corresponding proof. Since Lemma 1 is the key tool in the CEO converse result, this implies that the converse holds even if $\hat{X}^n$ is allowed to be any probability distribution on $\mathcal{X}^n$ (rather than being restricted to the set of product distributions).

When this stronger converse result is taken together with the achievability result, we observe that restricting $\hat{X}^n$ to be a product distribution is in fact optimal and can achieve all points in $\overline{\mathcal{RD}}^{\star}_{CEO}$.

2.3.4 An Example: Distributed compression of a posterior distribution

Suppose two sensors observe sequences Y_1^n and Y_2^n respectively, which are conditionally independent given a hidden sequence X^n. The sensors communicate with a fusion center through rate-limited links of capacity R_1 and R_2 respectively. Given sequences (Y_1^n, Y_2^n) are observed, the sequence X^n cannot be determined in general, so the fusion center would like to estimate the posterior distribution $p(x^n|Y_1^n, Y_2^n)$. However, since the communication links are rate-limited, the fusion center generally cannot compute $p(x^n|Y_1^n, Y_2^n)$ exactly. In this case, the fusion center would like to generate an estimate $\hat{p}(x^n|g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n))$ that should approximate $p(x^n|Y_1^n, Y_2^n)$ in the sense that

$$\mathbb{E}D\Big(p(x^n|Y_1^n, Y_2^n)\Big\|\hat{p}(x^n|g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n))\Big) \leq n\varepsilon,$$

where, consistent with standard notation (e.g. [CT06]), we define

$$\begin{aligned}\mathbb{E}D\Big(p(x^n|Y_1^n,Y_2^n)\Big\|\hat{p}(x^n|g_1^{(n)}(Y_1^n),g_2^{(n)}(Y_2^n))\Big)&\\ &= \mathbb{E}_{p(x^n,y_1^n,y_2^n)}\log\frac{p(x^n|Y_1^n,Y_2^n)}{\hat{p}(x^n|g_1^{(n)}(Y_1^n),g_2^{(n)}(Y_2^n))}\\ &= \sum_{x^n,y_1^n,y_2^n} p(x^n,y_1^n,y_2^n)\log\frac{p(x^n|y_1^n,y_2^n)}{\hat{p}(x^n|g_1^{(n)}(y_1^n),g_2^{(n)}(y_2^n))}.\end{aligned}$$

The relevant question here is the following. What is the minimum distortion ε that is attainable given R_1 and R_2?

Considering the CEO problem for this setup, we have:

$$\begin{aligned}\mathbb{E}d(X^n,\hat{X}^n) &= \frac{1}{n}\sum_{(x^n,y_1^n,y_2^n)} p(x^n,y_1^n,y_2^n)\log\left(\frac{1}{\hat{x}^n(x^n)}\right)\\ &= \frac{1}{n}\mathbb{E}D\Big(p(x^n|Y_1^n,Y_2^n)\Big\|\hat{X}^n(x^n)\Big)+\frac{1}{n}H(X^n|Y_1^n,Y_2^n).\end{aligned}$$

Identifying $\hat{p}(x^n|g_1^{(n)}(Y_1^n),g_2^{(n)}(Y_2^n)) \leftarrow \hat{X}^n(x^n)$, we have:

$$\mathbb{E}D\Big(p(x^n|Y_1^n,Y_2^n)\Big\|\hat{p}(x^n|g_1^{(n)}(Y_1^n),g_2^{(n)}(Y_2^n))\Big) = n\mathbb{E}d(X^n,\hat{X}^n) - nH(X|Y_1,Y_2).$$

Thus, finding the minimum possible distortion reduces to an optimization problem over $\overline{\mathcal{RD}}^{\star}_{CEO}$. In particular, the minimum attainable distortion ε^* is given by

$$\varepsilon^* = \inf\Big\{D : (R_1,R_2,D)\in\overline{\mathcal{RD}}^{\star}_{CEO}\Big\} - H(X|Y_1,Y_2). \tag{2.7}$$

Moreover, the minimum distortion is obtained by estimating each x_j separately. In other words, there exists an optimal estimate $\hat{p}^*(x^n|\cdot,\cdot)$ (which is itself a function of optimal encoding functions $g_1^{*(n)}(\cdot)$ and $g_2^{*(n)}(\cdot)$) that can be expressed as a product distribution

$$\hat{p}^*(x^n|\cdot,\cdot) = \prod_{j=1}^n \hat{p}_j^*\left(x_j|g_1^{*(n)}(\cdot),g_2^{*(n)}(\cdot)\right).$$

For this choice of $\hat{p}^*(x^n|\cdot,\cdot)$, we have the following relationship:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n \mathbb{E}D\Big(p(x_j|Y_{1,j},Y_{2,j})\Big\|\hat{p}_j^*(x(j)|g_1^{*(n)}(Y_1^n),g_2^{*(n)}(Y_2^n))\Big) = \varepsilon^*.$$

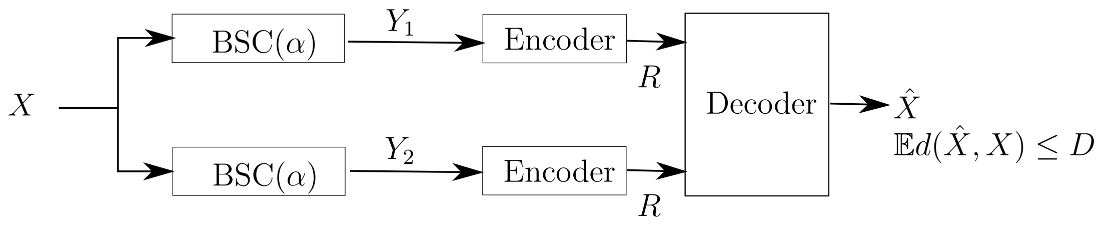

Figure 2.2: An example CEO problem where $X \sim \text{Bernoulli}(\frac{1}{2})$, $\Pr(Y_i = X) = (1 - \alpha)$, and both encoders are subject to the same rate constraint.

To make this example more concrete, consider the scenario depicted in Figure 2.2, where $X \sim \text{Bernoulli}(\frac{1}{2})$ and Y_i is the result of passing X through a binary symmetric channel with crossover probability α for $i = 1, 2$. To simplify things, we constrain the rates of each encoder to be at most R bits per channel use.

By performing a brute-force search over a fine mesh of conditional distributions $\{p(u_i|y_i)\}_{i=1}^2$, we numerically approximate the set of (R, D) pairs such that (R, R, D) is in the achievable region $\overline{\mathcal{RD}}_{CEO}^{\star}$ corresponding to the network in Figure 2.2. The lower convex envelope of these (R, D) pairs is plotted in Figure 2.3 for $\alpha \in \{0.01, 0.1, 0.25\}$. Continuing our example above for this concrete choice of source parameters, we compute the minimum achievable Kullback-Leibler distance ε^* according to (2.7). The result is given in Figure 2.4.

These numerical results are intuitively satisfying in the sense that, if Y_1, Y_2 are high-quality estimates of X (e.g., $\alpha = 0.01$), then a small increase in the allowable rate R results in a large relative improvement of $\hat{p}(x|\cdot, \cdot)$, the decoder's estimate of $p(x|Y_1, Y_2)$. On the other hand, if Y_1, Y_2 are poor-quality estimates of X (e.g., $\alpha = 0.25$), then we require a large increase in the allowable rate R in order to obtain an appreciable improvement of $\hat{p}(x|\cdot, \cdot)$.

One field where this example is directly applicable is machine learning. In this case, X_j could represent the class of object j, and $Y_{1,j}, Y_{2,j}$ are observable attributes. In machine learning, one typically estimates the probability that an object belongs to a particular class

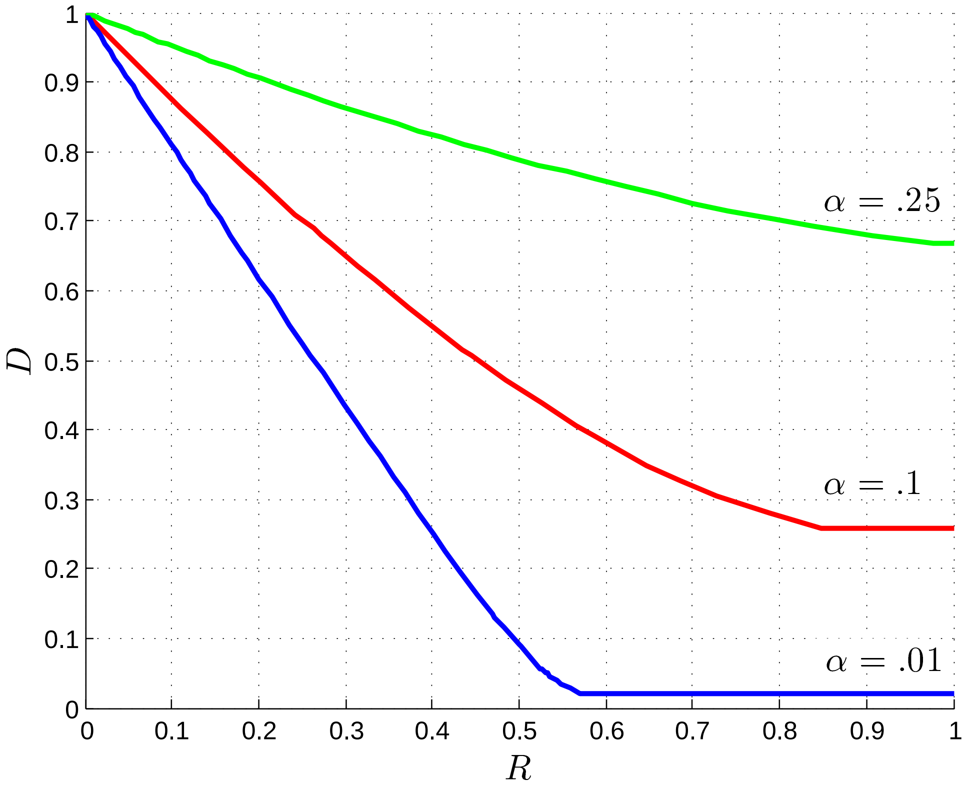

Figure 2.3: The distortion-rate function of the network in Figure 2.2 computed for $\alpha \in \{0.01, 0.1, 0.25\}$.

given a set of observable attributes. For this type of estimation problem, relative entropy is a natural penalty criterion.

Another application is to horse racing with conditionally independent, rate-limited side informations. In this case, the doubling rate of the gambler's wealth can be expressed in terms of the logarithmic loss distortion measure. This example is consistent with the original interpretation of the CEO problem, where the CEO makes consecutive business decisions (investments) having outcomes X^n, with the objective of maximizing the wealth of the company. We omit the details.

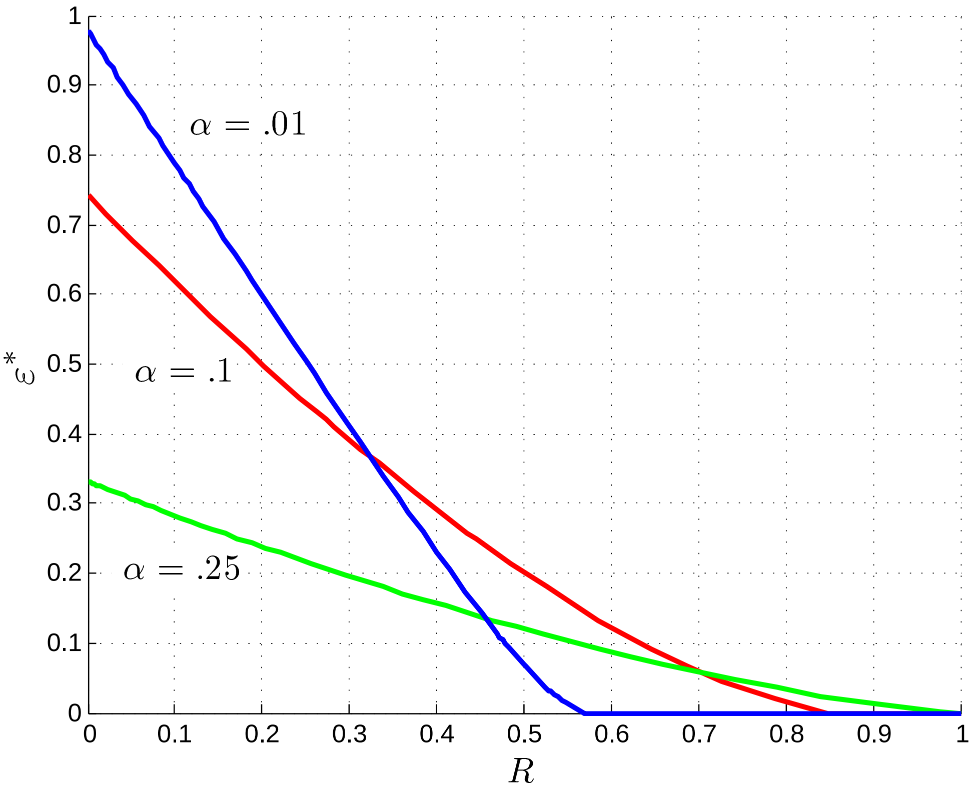

Figure 2.4: The minimum achievable Kullback-Leibler distance computed according to (2.7), i.e., the curves here are those of Figure 2.3, lowered by the constant $H(X|Y_1, Y_2)$.

2.3.5 An Example: Joint estimation of the encoder observations

Suppose one wishes to estimate the encoder observations (Y_1, Y_2). In this case, the rate region simplifies considerably. In particular, if we tolerate a distortion D in our estimate of the pair (Y_1, Y_2), then the achievable rate region is the same as the Slepian-Wolf rate region with each rate constraint relaxed by D bits. Formally:

Theorem 4. *If $X = (Y_1, Y_2)$, then $\overline{\mathcal{RD}}^{\star}_{CEO}$ consists of all vectors (R_1, R_2, D) satisfying*

$$\begin{aligned} R_1 &\geq H(Y_1|Y_2) - D \\ R_2 &\geq H(Y_2|Y_1) - D \\ R_1 + R_2 &\geq H(Y_1, Y_2) - D \\ D &\geq 0. \end{aligned}$$

Proof. First, note that Theorem 3 implies that $\overline{\mathcal{RD}}^{\star}_{CEO}$ is equivalent to the the union of (R_1, R_2, D) triples satisfying (2.1) taken over all joint distributions $p(q)p(x, y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)$. Now, since $X = (Y_1, Y_2)$, each of the inequalities (2.1) can be lower bounded as follows:

$$\begin{aligned} R_1 &\geq I(Y_1; U_1|Y_1, Y_2, Q) + H(Y_1, Y_2|U_2, Q) - D \\ &= H(Y_2|U_2, Q) + H(Y_1|Y_2) - D \\ &\geq H(Y_1|Y_2) - D \\ R_2 &\geq I(Y_2; U_2|Y_1, Y_2, Q) + H(Y_1, Y_2|U_1, Q) - D \\ &= H(Y_1|U_1, Q) + H(Y_2|Y_1) - D \\ &\geq H(Y_2|Y_1) - D \\ R_1 + R_2 &\geq I(U_1; Y_1|Y_1, Y_2, Q) + I(U_2; Y_2|Y_1, Y_2, Q) + H(Y_1, Y_2) - D \\ &= H(Y_1, Y_2) - D \\ D &\geq H(Y_1, Y_2|U_1, U_2, Q) \\ &\geq 0. \end{aligned}$$

Finally, observe that by setting $U_i = Y_i$ for $i = 1, 2$, we can achieve any point in this relaxed region (again, a consequence of Theorem 3). □

We remark that this result was first proved in [CW11a] by Courtade and Wesel using a different method.

2.3.6 An Example: The Information Bottleneck

If we consider the CEO problem with a single observed source (i.e., $Y_2 = \emptyset$), then the achievable rate distortion region given by Theorem 3 is characterized by all (R_1, D) pairs satisfying

$$\begin{aligned} R_1 &\geq I(Y_1; U_1) \\ D &\geq H(X|U_1) \end{aligned}$$

for some U_1 satisfying the Markov chain $X \leftrightarrow Y_1 \leftrightarrow U_1$. Alternatively, by making the substitution $\tau = H(X) - D$, this tradeoff can be characterized as follows:

$$R_1(\tau) = \min_{p(u_1|y_1): I(U_1;X) \geq \tau} I(Y_1; U_1). \tag{2.8}$$

Expression 2.8 is known as the *Information Bottleneck Function* (cf. [GNT03]). Intuitively, U_1 is a description of X which is generated (stochastically) from the observation Y_1. The function $R_1(\tau)$ describes the tradeoff between the complexity and the accuracy of the description U_1. Ideally, U_1 should capture the relevant information about X present in the observation Y_1.

The concept of the Information Bottleneck was first introduced by Tishby et al. in [TPB99], and the first formal rate distortion theorem on the topic was later proved by Gilad-Bachrach et al. in [GNT03]. We remark that algorithms motivated by the Information Bottleneck Method have been successfully applied to a wide variety problems. Examples include word clustering for text classification [ST01], galaxy spectra classification [SST01], neural code analysis [GCT01], and speech recognition [HT05]. Since Theorem 3 (and the m-encoder extension given in Appendix 2.B) generalize the tradeoff (2.8) to a distributed setting, our results could be applied to similar problems. Particularly those for which processing and computation occurs in a distributed or parallel manner.

2.4 Multiterminal Source Coding

With Theorem 3 in hand, we are now in a position to characterize the achievable rate distortion region $\overline{\mathcal{RD}}^\star$ for the multiterminal source coding problem under logarithmic loss. As before, we prove an inner bound first.

2.4.1 Inner Bound

Definition 5. *Let $(R_1, R_2, D_1, D_2) \in \mathcal{RD}^i$ if and only if there exists a joint distribution of the form*

$$p(y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)p(q)$$

where $|\mathcal{U}_1| \leq |\mathcal{Y}_1|$, $|\mathcal{U}_2| \leq |\mathcal{Y}_2|$, and $|\mathcal{Q}| \leq 5$, which satisfies

$$\begin{aligned} R_1 &\geq I(U_1; Y_1|U_2, Q) \\ R_2 &\geq I(U_2; Y_2|U_1, Q) \\ R_1 + R_2 &\geq I(U_1, U_2; Y_1, Y_2|Q) \\ D_1 &\geq H(Y_1|U_1, U_2, Q) \\ D_2 &\geq H(Y_2|U_1, U_2, Q). \end{aligned}$$

Theorem 5. *$\mathcal{RD}^i \subseteq \overline{\mathcal{RD}}^\star$. That is, all rate distortion vectors in $\mathcal{RD}^i$ are achievable.*

Again, we require an appropriate version of the Berger-Tung inner bound:

Proposition 2 (Berger-Tung Inner Bound [Ber77])**.** *The rate distortion vector (R_1, R_2, D_1, D_2) is achievable if*

$$\begin{aligned} R_1 &\geq I(U_1; Y_1|U_2, Q) \\ R_2 &\geq I(U_2; Y_2|U_1, Q) \\ R_1 + R_2 &\geq I(U_1, U_2; Y_1, Y_2|Q) \\ D_1 &\geq \mathbb{E}\left[d(Y_1, f_1(U_1, U_2, Q)\right] \\ D_2 &\geq \mathbb{E}\left[d(Y_2, f_2(U_1, U_2, Q)\right]. \end{aligned}$$

for a joint distribution

$$p(y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)p(q)$$

and reproduction functions

$$f_i : \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Q} \to \hat{\mathcal{Y}}_i, \text{ for } i = 1, 2.$$

Proof of Theorem 5. To prove the theorem, we simply apply Proposition 2 with the reproduction functions $f_i(U_1, U_2, Q) := \Pr[Y_i = y_i|U_1, U_2, Q]$. □

2.4.2 A Matching Outer Bound

The main result of this chapter is the following theorem.

Theorem 6. $\mathcal{RD}^i = \overline{\mathcal{RD}}^\star$.

Proof. As in the proof of Theorem 3, we note that the cardinality bounds on the alphabets in the definition of $\mathcal{RD}^i$ can be imposed without any loss of generality. This is discussed in detail in Appendix 2.A. Thus, we do not need to consider these constraints in the proof of the converse.

Assume (R_1, R_2, D_1, D_2) is strict-sense achievable. Observe that proving that $(R_1, R_2, D_1, D_2) \in \mathcal{RD}^i$ will prove the theorem, since $\mathcal{RD}^i \subseteq \overline{\mathcal{RD}}^\star$ and $\overline{\mathcal{RD}}^\star$ is closed by definition.

For convenience, define $\mathcal{P}(R_1, R_2)$ to be the set of joint distributions of the form

$$p(y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)p(q)$$

with $|\mathcal{U}_1| \leq |\mathcal{Y}_1|$, $|\mathcal{U}_2| \leq |\mathcal{Y}_2|$, and $|\mathcal{Q}| \leq 4$ satisfying

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

$$R_2 \geq I(U_2; Y_2|U_1, Q)$$

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

We remark that $\mathcal{P}(R_1, R_2)$ is compact. We also note that it will suffice to show the existence of a joint distribution in $\mathcal{P}(R_1, R_2)$ satisfying $H(Y_1|U_1, U_2, Q) \leq D_1$ and $H(Y_2|U_1, U_2, Q) \leq D_2$ to prove that $(R_1, R_2, D_1, D_2) \in \mathcal{RD}^i$.

With foresight, consider random variable X defined as follows

$$X = \begin{cases} (Y_1, 1) & \text{with probability } t \\ (Y_2, 2) & \text{with probability } 1-t. \end{cases} \tag{2.9}$$

In other words, $X = (Y_B, B)$, where B is a Bernoulli random variable independent of Y_1, Y_2. Observe that $Y_1 \leftrightarrow X \leftrightarrow Y_2$ form a Markov chain, and thus, we are able to apply Theorem 3.

Since (R_1, R_2, D_1, D_2) is strict-sense achievable, the decoder can construct reproductions $\hat{Y}_1^n, \hat{Y}_2^n$ satisfying

$$\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}d(Y_{i,j}, \hat{Y}_{i,j}) \leq D_i \text{ for } i = 1, 2.$$

Fix the encoding operations and set $\hat{X}_j\left((y_1, 1)\right) = t\hat{Y}_{1,j}(y_1)$ and $\hat{X}_j\left((y_2, 2)\right) = (1-t)\hat{Y}_{2,j}(y_2)$. Then for the CEO problem defined by (X, Y_1, Y_2):

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}d(X_j, \hat{X}_j) \\
&= \frac{t}{n}\sum_{j=1}^{n} \mathbb{E}\log\left(\frac{1}{t\hat{Y}_{1,j}(Y_{1,j})}\right) + \frac{1-t}{n}\sum_{j=1}^{n} \mathbb{E}\log\left(\frac{1}{(1-t)\hat{Y}_{2,j}(Y_{2,j})}\right) \\
&= h_2(t) + \frac{t}{n}\sum_{j=1}^{n} \mathbb{E}d(Y_{1,j}, \hat{Y}_{1,j}) + \frac{1-t}{n}\sum_{j=1}^{n} \mathbb{E}d(Y_{2,j}, \hat{Y}_{2,j}) \\
&\leq h_2(t) + tD_1 + (1-t)D_2
\end{aligned}$$

where $h_2(t)$ is the binary entropy function. Hence, for this CEO problem, distortion $h_2(t) + tD_1 + (1-t)D_2$ is achievable and Theorem 3 yields a joint distribution[2] $P_t \in$

[2] Henceforth, we use the superscript (t) to explicitly denote the dependence of the auxiliary random variables on the distribution parametrized by t.

$\mathcal{P}(R_1, R_2)$ satisfying

$$\begin{aligned}h_2(t) + tD_1 + (1-t)D_2 &\geq H(X|U_1^{(t)}, U_2^{(t)}, Q^{(t)}) \\ &= h_2(t) + tH(Y_1|U_1^{(t)}, U_2^{(t)}, Q^{(t)}) \\ &\quad + (1-t)H(Y_2|U_1^{(t)}, U_2^{(t)}, Q^{(t)}),\end{aligned}$$

where the second equality follows by by definition of X in (2.9). For convenience, define $H_1(P_t) \triangleq H(Y_1|U_1^{(t)}, U_2^{(t)}, Q^{(t)})$ and $H_2(P_t) \triangleq H(Y_2|U_1^{(t)}, U_2^{(t)}, Q^{(t)})$. Note the following two facts:

1. By continuity of entropy, the functions $H_1(\cdot)$ and $H_2(\cdot)$ are continuous on the compact domain $\mathcal{P}(R_1, R_2)$.

2. The above argument proves the existence of a function $\varphi : [0,1] \to \mathcal{P}(R_1, R_2)$ which satisfies

$$tH_1(\varphi(t)) + (1-t)H_2(\varphi(t)) \leq tD_1 + (1-t)D_2 \text{ for all } t \in [0,1].$$

These two facts satisfy the requirements of Lemma 17 (see Chapter 4), and hence there exists $P_{t_1} \in \mathcal{P}(R_1, R_2)$, $P_{t_2} \in \mathcal{P}(R_1, R_2)$, and $\theta \in [0,1]$ for which

$$\begin{aligned}\theta H_1(P_{t_1}) + (1-\theta)H_1(P_{t_2}) &\leq D_1 \\ \theta H_2(P_{t_1}) + (1-\theta)H_2(P_{t_2}) &\leq D_2.\end{aligned}$$

Timesharing[3] between distributions P_{t_1} and P_{t_2} with probabilities θ and $(1-\theta)$, respectively, yields a distribution $P^* \in \mathcal{P}(R_1, R_2)$ which satisfies $H_1(P^*) \leq D_1$ and $H_2(P^*) \leq D_2$. This proves the theorem. □

2.4.3 A stronger converse

For the CEO problem, we are able to obtain a stronger converse result as discussed in Section 2.3.3. We can obtain a similar result for the multiterminal source coding problem.

[3] The timesharing scheme can be embedded in the timesharing variable Q, increasing the cardinality of $\mathcal{Q}$ by a factor of two.

To be precise, let $\hat{\mathcal{Y}}_1^{*n}$ and $\hat{\mathcal{Y}}_2^{*n}$ denote the set of probability measures on $\mathcal{Y}_1^n$ and $\mathcal{Y}_2^n$, respectively. Let d_1^*, d_2^* be the (extended)-log loss distortion measures defined as follows:

$$\begin{aligned} d_1^*(y_1^n, \hat{y}_1^n) &= \frac{1}{n} \log \left(\frac{1}{\hat{y}_1^n(y_1^n)} \right) \\ d_2^*(y_2^n, \hat{y}_2^n) &= \frac{1}{n} \log \left(\frac{1}{\hat{y}_2^n(y_2^n)} \right), \end{aligned}$$

where $\hat{y}_1^n(y_1^n)$ is the probability assigned to outcome $y_1^n \in \mathcal{Y}_1^n$ by the probability measure $\hat{y}_1^n \in \hat{\mathcal{Y}}_1^{*n}$. Similarly for $\hat{y}_2^n(y_2^n)$. Note that this extends the standard definition of logarithmic loss to sequence reproductions.

Definition 6. *We say that a tuple (R_1, R_2, D_1, D_2) is sequence-achievable if, for any $\epsilon > 0$, there exist encoding functions*

$$\begin{aligned} f_1 &: \mathcal{Y}_1^n \to \{1, \dots, 2^{nR_1}\} \\ f_2 &: \mathcal{Y}_2^n \to \{1, \dots, 2^{nR_2}\}, \end{aligned}$$

and decoding functions

$$\begin{aligned} \phi_1 &: \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\} \to \hat{\mathcal{Y}}_1^{*n} \\ \phi_2 &: \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\} \to \hat{\mathcal{Y}}_2^{*n}, \end{aligned}$$

which satisfy

$$\begin{aligned} \mathbb{E}\ d_1^*(Y_1^n, \hat{Y}_1^n) &\leq D_1 + \epsilon \\ \mathbb{E}\ d_2^*(Y_2^n, \hat{Y}_2^n) &\leq D_2 + \epsilon, \end{aligned}$$

where

$$\begin{aligned} \hat{Y}_1^n &= \phi_1(f_1(Y_1^n), f_2(Y_2^n)) \\ \hat{Y}_2^n &= \phi_2(f_1(Y_1^n), f_2(Y_2^n)). \end{aligned}$$

Theorem 7. *If (R_1, R_2, D_1, D_2) is sequence-achievable, then $(R_1, R_2, D_1, D_2) \in \mathcal{RD}^i = \overline{\mathcal{RD}}^\star$.*

Proof. The theorem is an immediate consequence of Theorem 6 and Lemmas 2 and 3, which are given below. □

We refer to Theorem 7 as the "strengthened converse" of Theorem 6. Indeed, it states that enlarging the set of possible reproduction sequences to include non-product distributions cannot attain better performance than when the decoder is restricted to choosing a reproduction sequence from the set of product distributions. This result has several interesting implications, which we describe in detail in the following sections.

Lemma 2. *If $(R_1, R_2, \tilde{D}_1, D_2)$ is sequence-achievable, then there exists a joint distribution*

$$p(y_1, y_2, u_1, u_2, q) = p(q)p(y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)$$

and a $D_1 \leq \tilde{D}_1$ which satisfies

$$\begin{aligned} D_1 &\geq H(Y_1|U_1, U_2, Q) \\ D_2 &\geq D_1 + H(Y_2|U_1, U_2, Q) - H(Y_1|U_1, U_2, Q), \end{aligned}$$

and

$$\begin{aligned} R_1 &\geq H(Y_1|U_2, Q) - D_1 \\ R_2 &\geq I(Y_2; U_2|Y_1, Q) + H(Y_1|U_1, Q) - D_1 \\ R_1 + R_2 &\geq I(Y_2; U_2|Y_1, Q) + H(Y_1) - D_1. \end{aligned}$$

Proof. For convenience, let $F_1 = f_1(Y_1^n)$ and $F_2 = f_2(Y_2^n)$, where f_1, f_2 are the encoding functions corresponding to a scheme which achieves $(R_1, R_2, \tilde{D}_1, D_2)$ (in the sequence-reproduction sense). Define $D_1 = \frac{1}{n}H(Y_1^n|F_1, F_2)$, so that:

$$nD_1 = H(Y_1^n|F_1, F_2). \tag{2.10}$$

Since $n\tilde{D}_1 \geq H(Y_1^n|F_1, F_2)$ by the strengthened version[4] of Lemma 1, we have $D_1 \leq \tilde{D}_1$ as desired. By definition of D_1, we immediately obtain the following inequality:

$$nD_1 = \sum_{i=1}^{n} H(Y_{1,i}|F_1, F_2, Y_{1,i+1}^n) \geq \sum_{i=1}^{n} H(Y_{1,i}|F_1, F_2, Y_2^{i-1}, Y_{1,i+1}^n). \tag{2.11}$$

[4] See the comment in Section 2.3.3.

Next, recall the Csiszár sum identity:

$$\sum_{i=1}^{n} I(Y_{1,i+1}^{n};Y_{2,i}|Y_2^{i-1},F_1,F_2) = \sum_{i=1}^{n} I(Y_2^{i-1};Y_{1,i}|Y_{1,i+1}^{n},F_1,F_2).$$

This, together with (2.10), implies the following inequality:

$$nD_2 \geq nD_1 + \sum_{i=1}^{n} H(Y_{2,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}) - H(Y_{1,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}), \tag{2.12}$$

which we can verifiy as follows:

$$\begin{aligned}
nD_2 &\geq H(Y_2^n|F_1,F_2) = \sum_{i=1}^{n} H(Y_{2,i}|F_1,F_2,Y_2^{i-1}) \\
&= \sum_{i=1}^{n} H(Y_{2,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}) + I(Y_{1,i+1}^{n};Y_{2,i}|F_1,F_2,Y_2^{i-1}) \\
&= \sum_{i=1}^{n} H(Y_{2,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}) + I(Y_2^{i-1};Y_{1,i}|Y_{1,i+1}^{n},F_1,F_2) \\
&= H(Y_1^n|F_1,F_2) + \sum_{i=1}^{n} H(Y_{2,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}) - H(Y_{1,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}) \\
&= nD_1 + \sum_{i=1}^{n} H(Y_{2,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}) - H(Y_{1,i}|F_1,F_2,Y_2^{i-1},Y_{1,i+1}^{n}).
\end{aligned}$$

Next, observe that we can lower bound R_1 as follows:

$$\begin{aligned}
nR_1 &\geq H(F_1) \geq I(Y_1^n;F_1|F_2) \\
&= \sum_{i=1}^{n} H(Y_{1,i}|F_2,Y_1^{i-1}) - H(Y_1^n|F_1,F_2) \\
&\geq \sum_{i=1}^{n} H(Y_{1,i}|F_2,Y_1^{i-1},Y_2^{i-1}) - nD_1 && (2.13) \\
&= \sum_{i=1}^{n} H(Y_{1,i}|F_2,Y_2^{i-1}) - nD_1 && (2.14) \\
&\geq \sum_{i=1}^{n} H(Y_{1,i}|F_2,Y_2^{i-1},Y_{1,i+1}^{n}) - nD_1. && (2.15)
\end{aligned}$$

In the above string of inequalities, (2.13) follows from (2.10) and the fact that conditioning reduces entropy. Equality (2.14) follows since $Y_{1,i} \leftrightarrow F_2, Y_2^{i-1} \leftrightarrow Y_1^{i-1}$ form a Markov chain (in that order).

Next, we can obtain a lower bound on R_2:

$$\begin{aligned}
nR_2 &\geq H(F_2) \geq H(F_2|F_1) = H(F_2|F_1, Y_1^n) + I(Y_1^n; F_2|F_1) \\
&\geq I(Y_2^n; F_2|F_1, Y_1^n) + I(Y_1^n; F_2|F_1) \\
&= I(Y_2^n; F_2|Y_1^n) + I(Y_1^n; F_2|F_1) \qquad (2.16) \\
&= \sum_{i=1}^{n} I(Y_{2,i}; F_2|Y_1^n, Y_2^{i-1}) + H(Y_{1,i}|F_1, Y_{1,i+1}^n) - nD_1 \qquad (2.17) \\
&\geq \sum_{i=1}^{n} I(Y_{2,i}; F_2|Y_1^n, Y_2^{i-1}) + H(Y_{1,i}|F_1, Y_2^{i-1}, Y_{1,i+1}^n) - nD_1 \\
&= \sum_{i=1}^{n} I(Y_{2,i}; F_2, Y_1^{i-1}, Y_2^{i-1}|Y_{1,i}, Y_2^{i-1}, Y_{1,i+1}^n) + H(Y_{1,i}|F_1, Y_2^{i-1}, Y_{1,i+1}^n) - nD_1 \qquad (2.18) \\
&\geq \sum_{i=1}^{n} I(Y_{2,i}; F_2, Y_2^{i-1}|Y_{1,i}, Y_2^{i-1}, Y_{1,i+1}^n) + H(Y_{1,i}|F_1, Y_2^{i-1}, Y_{1,i+1}^n) - nD_1. \qquad (2.19)
\end{aligned}$$

In the above string of inequalities, (2.17) follows from (2.10) and the chain rule. (2.18) follows from the i.i.d. property of the sources, and (2.19) follows by monotonicity of mutual information.

A lower bound on the sum-rate $R_1 + R_2$ can be obtained as follows:

$$\begin{aligned}
n(R_1 + R_2) &\geq H(F_1) + H(F_2) \geq H(F_2) + H(F_1|F_2) \\
&\geq I(F_2; Y_1^n, Y_2^n) + I(F_1; Y_1^n|F_2) \\
&= I(F_2; Y_1^n) + I(F_2; Y_2^n|Y_1^n) + I(F_1; Y_1^n|F_2) \\
&= I(F_2; Y_2^n|Y_1^n) + I(F_1, F_2; Y_1^n) \\
&\geq \sum_{i=1}^{n} I(Y_{2,i}; F_2, Y_2^{i-1}|Y_{1,i}, Y_2^{i-1}, Y_{1,i+1}^n) + H(Y_{1,i}) - nD_1. \qquad (2.20)
\end{aligned}$$

Where (2.20) follows in a manner similar to (2.16)-(2.19) in the lower bound on R_2.

Now, define $U_{1,i} \triangleq F_1$, $U_{2,i} \triangleq (F_2, Y_2^{i-1})$, and $Q_i \triangleq (Y_2^{i-1}, Y_{1,i+1}^n)$. Then we can summarize our results so far as follows. Inequalities (2.11) and (2.12) become

$$\begin{aligned} D_1 &\geq \frac{1}{n}\sum_{i=1}^{n} H(Y_{1,i}|U_{1,i},U_{2,i},Q_i) \\ D_2 &\geq D_1 + \frac{1}{n}\sum_{i=1}^{n} H(Y_{2,i}|U_{1,i},U_{2,i},Q_i) - H(Y_{1,i}|U_{1,i},U_{2,i},Q_i), \end{aligned}$$

and inequalities (2.15), (2.19), and (2.20) can be written as:

$$\begin{aligned} R_1 &\geq \frac{1}{n}\sum_{i=1}^{n} H(Y_{1,i}|U_{2,i},Q_i) - D_1 \\ R_2 &\geq \frac{1}{n}\sum_{i=1}^{n} I(Y_{2,i};U_{2,i}|Y_{1,i},Q_i) + H(Y_{1,i}|U_{1,i},Q_i) - D_1 \\ R_1 + R_2 &\geq \frac{1}{n}\sum_{i=1}^{n} I(Y_{2,i};U_{2,i}|Y_{1,i},Q_i) + H(Y_{1,i}) - D_1. \end{aligned}$$

Next, we note that $U_{1,i} \leftrightarrow Y_{1,i} \leftrightarrow Y_{2,i} \leftrightarrow U_{2,i}$ form a Markov chain (in that order) conditioned on Q_i. Moreover, Q_i is independent of $Y_{1,i}, Y_{2,i}$. Hence, a standard timesharing argument proves the lemma. □

Lemma 3. *Fix* (R_1, R_2, D_1, D_2)*. If there exists a joint distribution of the form*

$$p(y_1, y_2, u_1, u_2, q) = p(q)p(y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)$$

which satisfies

$$D_1 \geq H(Y_1|U_1, U_2, Q) \tag{2.21}$$

$$D_2 \geq D_1 + H(Y_2|U_1, U_2, Q) - H(Y_1|U_1, U_2, Q), \tag{2.22}$$

and

$$R_1 \geq H(Y_1|U_2, Q) - D_1 \tag{2.23}$$

$$R_2 \geq I(Y_2; U_2|Y_1, Q) + H(Y_1|U_1, Q) - D_1 \tag{2.24}$$

$$R_1 + R_2 \geq I(Y_2; U_2|Y_1, Q) + H(Y_1) - D_1, \tag{2.25}$$

then $(R_1, R_2, D_1, D_2) \in \mathcal{RD}^i$*.*

Proof. Let $\mathcal{P}$ denote the polytope of rate pairs which satisfy the inequalities (2.23)-(2.25). It suffices to show that if (r_1, r_2) is a vertex of $\mathcal{P}$, then $(r_1, r_2, D_1, D_2) \in \mathcal{RD}^i$. For convenience, let $[x]^+ = \max\{x, 0\}$. There are only two extreme points of $\mathcal{P}$:

$$\begin{aligned}
r_1^{(1)} &= \Big[H(Y_1|U_2, Q) - D_1\Big]^+ \\
r_2^{(1)} &= I(Y_2; U_2|Y_1, Q) + H(Y_1) - D_1 - r_1^{(1)},
\end{aligned}$$

and

$$\begin{aligned}
r_1^{(2)} &= I(Y_2; U_2|Y_1, Q) + H(Y_1) - D_1 - r_2^{(2)}, \\
r_2^{(2)} &= \Big[I(Y_2; U_2|Y_1, Q) + H(Y_1|U_1, Q) - D_1\Big]^+.
\end{aligned}$$

We first analyze the extreme point $(r_1^{(1)}, r_2^{(1)})$:

- **Case 1.1:** $r_1^{(1)} = 0$. In this case, we have $r_2^{(1)} = I(Y_2; U_2|Y_1, Q) + H(Y_1) - D_1$. This can be expressed as:

  $$r_2^{(1)} = (1 - \theta) I(Y_2; U_2|Q),$$

  where

  $$\theta = \frac{D_1 - I(Y_2; U_2|Y_1, Q) - H(Y_1) + I(Y_2; U_2|Q)}{I(Y_2; U_2|Q)}.$$

  Since $r_1^{(1)} = 0$, we must have $D_1 \geq H(Y_1|U_2, Q)$. This implies that

  $$\theta \geq \frac{H(Y_1|U_2, Q) - I(Y_2; U_2|Y_1, Q) - H(Y_1) + I(Y_2; U_2|Q)}{I(Y_2; U_2|Q)} = 0.$$

  Also, we can assume without loss of generality that $D_1 \leq H(Y_1)$, hence $\theta \in [0, 1]$. Applying the Berger-Tung achievability scheme, we can achieve the following distortions:

$$\begin{aligned}
D_1^\theta &= \theta H(Y_1) + (1-\theta)H(Y_1|U_2,Q)\\
&= H(Y_1|U_2,Q) + \theta I(Y_1;U_2|Q)\\
&\le H(Y_1|U_2,Q) + D_1 - I(Y_2;U_2|Y_1,Q) - H(Y_1) + I(Y_2;U_2|Q) \qquad (2.26)\\
&= D_1 - I(Y_2;U_2|Y_1,Q) - I(Y_1;U_2|Q) + I(Y_2;U_2|Q)\\
&= D_1,
\end{aligned}$$

where (2.26) follows since $I(Y_1;U_2|Q) \le I(Y_2;U_2|Q)$ by the data processing inequality.

$$\begin{aligned}
D_2^\theta &= \theta H(Y_2) + (1-\theta)H(Y_2|U_2,Q)\\
&= H(Y_2|U_2,Q) + \theta I(Y_2;U_2|Q)\\
&= H(Y_2|U_2,Q) + D_1 - I(Y_2;U_2|Y_1,Q) - H(Y_1) + I(Y_2;U_2|Q)\\
&= H(Y_2) + D_1 - I(Y_2;U_2|Y_1,Q) - H(Y_1)\\
&= H(Y_2|Y_1,U_2,Q) + D_1 - H(Y_1|Y_2)\\
&= H(Y_2|Y_1,U_1,U_2,Q) + D_1 - H(Y_1|Y_2) \qquad (2.27)\\
&\le H(Y_2|Y_1,U_1,U_2,Q) + D_1 - H(Y_1|Y_2,U_1,U_2,Q)\\
&= H(Y_2|U_1,U_2,Q) + D_1 - H(Y_1|U_1,U_2,Q)\\
&\le D_2, \qquad (2.28)
\end{aligned}$$

where (2.27) follows since $U_1 \leftrightarrow (Y_1,U_2,Q) \leftrightarrow Y_2$, and (2.28) follows from (2.22).

- **Case 1.2:** $r_1^{(1)} \ge 0$. In this case, we have $r_2^{(1)} = I(Y_2;U_2|Y_1,Q) + I(Y_1;U_2|Q) = I(Y_2;U_2|Q)$. Also, we can write $r_1^{(1)}$ as:

$$r_1^{(1)} = (1-\theta)I(Y_1;U_1|U_2,Q),$$

where

$$\theta = \frac{D_1 - H(Y_1|U_2,Q) + I(Y_1;U_1|U_2,Q)}{I(Y_1;U_1|U_2,Q)}.$$

Since $r_1^{(1)} \geq 0$, we must have $D_1 \leq H(Y_1|U_2, Q)$. This implies that

$$\theta \leq \frac{H(Y_1|U_2, Q) - H(Y_1|U_2, Q) + I(Y_1; U_1|U_2, Q)}{I(Y_1; U_1|U_2, Q)} = 1.$$

Also, (2.21) implies that $D_1 \geq H(Y_1|U_1, U_2, Q)$, hence $\theta \in [0, 1]$. Applying the Berger-Tung achievability scheme, we can achieve the following distortions:

$$\begin{aligned} D_1^\theta &= \theta H(Y_1|U_2, Q) + (1 - \theta) H(Y_1|U_1, U_2, Q) \\ &= H(Y_1|U_1, U_2, Q) + \theta I(Y_1; U_1|U_2, Q) \\ &= H(Y_1|U_1, U_2, Q) + D_1 - H(Y_1|U_2, Q) + I(Y_1; U_1|U_2, Q) \\ &= D_1, \end{aligned}$$

and

$$\begin{aligned} D_2^\theta &= \theta H(Y_2|U_2, Q) + (1 - \theta) H(Y_2|U_1, U_2, Q) \\ &= H(Y_2|U_1, U_2, Q) + \theta I(Y_2; U_1|U_2, Q) \\ &\leq H(Y_2|U_1, U_2, Q) + D_1 - H(Y_1|U_2, Q) + I(Y_1; U_1|U_2, Q) \qquad (2.29) \\ &= H(Y_2|U_1, U_2, Q) + D_1 - H(Y_1|U_1, U_2, Q) \\ &\leq D_2, \qquad (2.30) \end{aligned}$$

where (2.29) follows since $I(Y_2; U_1|U_2, Q) \leq I(Y_1; U_1|U_2, Q)$ by the data processing inequality, and (2.30) follows from (2.22).

In a similar manner, we now analyze the second extreme point $(r_1^{(2)}, r_2^{(2)})$:

- **Case 2.1:** $r_2^{(2)} = 0$. In this case, we have $r_1^{(2)} = I(Y_2; U_2|Y_1, Q) + H(Y_1) - D_1$. This can be expressed as:

  $$r_1^{(2)} = (1 - \theta) I(Y_1; U_1|Q),$$

  where

  $$\theta = \frac{D_1 - I(Y_2; U_2|Y_1, Q) - H(Y_1) + I(Y_1; U_1|Q)}{I(Y_1; U_1|Q)}.$$

Since $r_2^{(2)} = 0$, we must have $D_1 \geq H(Y_1|U_1, Q) + I(Y_2; U_2|Y_1, Q)$. This implies that

$$\theta \geq \frac{H(Y_1|U_1, Q) + I(Y_2; U_2|Y_1, Q) - I(Y_2; U_2|Y_1, Q) - H(Y_1) + I(Y_1; U_1|Q)}{I(Y_1; U_1|Q)} = 0.$$

Also, we can assume without loss of generality that $D_1 \leq H(Y_1)$, hence

$$\theta \leq \frac{H(Y_1) - I(Y_2; U_2|Y_1, Q) - H(Y_1) + I(Y_1; U_1|Q)}{I(Y_1; U_1|Q)} \leq 1,$$

and therefore $\theta \in [0, 1]$. Applying the Berger-Tung achievability scheme, we can achieve the following distortions:

$$\begin{aligned} D_1^\theta &= \theta H(Y_1) + (1 - \theta) H(Y_1|U_1, Q) \\ &= H(Y_1|U_1, Q) + \theta I(Y_1; U_1|Q) \\ &= H(Y_1|U_1, Q) + D_1 - I(Y_2; U_2|Y_1, Q) - H(Y_1) + I(Y_1; U_1|Q) \\ &= D_1 - I(Y_2; U_2|Y_1, Q) \\ &\leq D_1, \end{aligned}$$

and

$$\begin{aligned} D_2^\theta &= \theta H(Y_2) + (1 - \theta) H(Y_2|U_1, Q) \\ &= H(Y_2|U_1, Q) + \theta I(Y_2; U_1|Q) \\ &\leq H(Y_2|U_1, Q) + D_1 - I(Y_2; U_2|Y_1, Q) - H(Y_1) + I(Y_1; U_1|Q) && (2.31) \\ &= H(Y_2|Y_1, U_2, Q) + D_1 - H(Y_1|Y_2, U_1, Q) \\ &= H(Y_2|Y_1, U_1, U_2, Q) + D_1 - H(Y_1|Y_2, U_1, U_2, Q) && (2.32) \\ &= H(Y_2|U_1, U_2, Q) + D_1 - H(Y_1|U_1, U_2, Q) \\ &\leq D_2, && (2.33) \end{aligned}$$

where (2.31) follows since $I(Y_2; U_1|Q) \leq I(Y_1; U_1|Q)$ by the data processing inequality, (2.32) follows since $U_1 \leftrightarrow (Y_1, U_2, Q) \leftrightarrow Y_2$ and $U_2 \leftrightarrow (Y_2, U_1, Q) \leftrightarrow Y_1$, and (2.33) follows from (2.22).

- **Case 2.2:** $r_2^{(2)} \geq 0$. In this case, we have $r_1^{(2)} = I(Y_1; U_1|Q)$. Also, we can write $r_2^{(2)}$ as:

$$r_2^{(2)} = (1-\theta) I(Y_2; U_2|U_1, Q),$$

where

$$\theta = \frac{D_1 - H(Y_1|U_1, Q) - I(Y_2; U_2|Y_1, Q) + I(Y_2; U_2|U_1, Q)}{I(Y_2; U_2|U_1, Q)}.$$

Since $r_2^{(2)} \geq 0$, we must have $D_1 \leq H(Y_1|U_1, Q) + I(Y_2; U_2|Y_1, Q)$. This implies that $\theta \leq 1$. Also, (2.21) implies that $D_1 \geq H(Y_1|U_1, U_2, Q)$, yielding

$$\theta \geq \frac{H(Y_1|U_1, U_2, Q) - H(Y_1|U_1, Q) - I(Y_2; U_2|Y_1, Q) + I(Y_2; U_2|U_1, Q)}{I(Y_2; U_2|U_1, Q)} = 0.$$

Therefore, $\theta \in [0, 1]$. Applying the Berger-Tung achievability scheme, we can achieve the following distortions:

$$\begin{aligned} D_1^\theta &= \theta H(Y_1|U_1, Q) + (1-\theta) H(Y_1|U_1, U_2, Q) \\ &= H(Y_1|U_1, U_2, Q) + \theta I(Y_1; U_2|U_1, Q) \\ &\leq H(Y_1|U_1, U_2, Q) + D_1 - H(Y_1|U_1, Q) - I(Y_2; U_2|Y_1, Q) + I(Y_2; U_2|U_1, Q) \qquad (2.34) \\ &= D_1, \end{aligned}$$

where (2.34) follows since $I(Y_1; U_2|U_1, Q) \leq I(Y_2; U_2|U_1, Q)$ by the data processing inequality.

$$\begin{aligned} D_2^\theta &= \theta H(Y_2|U_1, Q) + (1-\theta) H(Y_2|U_1, U_2, Q) \\ &= H(Y_2|U_1, U_2, Q) + \theta I(Y_2; U_2|U_1, Q) \\ &= H(Y_2|U_1, U_2, Q) + D_1 - H(Y_1|U_1, Q) - I(Y_2; U_2|Y_1, Q) + I(Y_2; U_2|U_1, Q) \\ &= H(Y_2|U_1, U_2, Q) + D_1 - H(Y_1|U_1, U_2, Q) \\ &\leq D_2, \qquad (2.35) \end{aligned}$$

where (2.35) follows from (2.22).

Thus, this proves that the Berger-Tung compression scheme can achieve any rate distortion tuple (r_1, r_2, D_1, D_2) for $(r_1, r_2) \in \mathcal{P}$. Since $\mathcal{RD}^i$ is, by definition, the set of rate distortion tuples attainable by the Berger-Tung achievability scheme, we must have that $(R_1, R_2, D_1, D_2) \in \mathcal{RD}^i$. This proves the lemma. □

2.4.3.1 A Brief Remark on the Proof of the Strengthened Converse

We note that the proof of Theorem 7 offers a direct proof of the converse of Theorem 6, and as such we do not require a CEO result (Theorem 3) or a "black box" tuning argument (Lemma 17). At the heart of this alternative proof lies the Csiszár sum identity (and a careful choice of auxiliary random variables) which provides a coupling between the attainable distortions for each source. In the original proof of Theorem 6, this coupling is accomplished by the tuning argument through Lemma 17.

Interestingly, the two proofs are quite similar in spirit, with the key differences being the use of the Csiszár sum identity versus the tuning argument. Intuitively, the original tuning argument allows a "clumsier" choice of auxiliary random variables which leads to a more elegant and intuitive proof, but appears incapable of establishing the strengthened converse. On the other hand, applying the Csiszár sum identity requires a very careful choice of auxiliary random variables which, in turn, affords a finer degree of control over various quantities. Ultimately, this allows us to prove the strengthened converse given by Theorem 7.

2.4.4 An Example: The Daily Double

The *Daily Double* is a single bet that links together wagers on the winners of two consecutive horse races. Winning the Daily Double is dependent on both wagers winning together. In general, the outcomes of two consecutive races can be correlated (e.g. due to track conditions), so a gambler can potentially use this information to maximize his expected winnings. Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be the set of horses running in the first and second races respectively. If horses y_1 and y_2 win their respective races, then the payoff is $o(y_1, y_2)$ dollars for each dollar invested in outcome $(Y_1, Y_2) = (y_1, y_2)$.

There are two betting strategies one can follow:

1. The gambler can wager a fraction $b_1(y_1)$ of his wealth on horse y_1 winning the first race and parlay his winnings by betting a fraction $b_2(y_2)$ of his wealth on horse y_2 winning the second race. In this case, the gambler's wealth relative is $b_1(Y_1)b_2(Y_2)o(Y_1, Y_2)$ upon learning the outcome of the Daily Double. We refer to this betting strategy as the *product-wager*.

2. The gambler can wager a fraction $b(y_1, y_2)$ of his wealth on horses (y_1, y_2) winning the first and second races, respectively. In this case, the gambler's wealth relative is $b(Y_1, Y_2)o(Y_1, Y_2)$ upon learning the outcome of the Daily Double. We refer to this betting strategy as the *joint-wager*.

Clearly the joint-wager includes the product-wager as a special case. However, the product-wager requires less effort to place, so the question is: how do the two betting strategies compare?

To make things interesting, suppose the gamblers have access to rate-limited information about the first and second race outcomes at rates R_1, R_2 respectively. Further, assume that $R_1 \leq H(Y_1)$, $R_2 \leq H(Y_2)$, and $R_1 + R_2 \leq H(Y_1, Y_2)$. For (R_1, R_2) and $p(y_1, y_2)$ given, let $\mathcal{P}(R_1, R_2)$ denote the set of joint pmf's of the form

$$p(q, y_1, y_2, u_1, u_2) = p(q)p(y_1, y_2)p(u_1|y_1, q)p(u_1|y_1, q)$$

which satisfy

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

$$R_2 \geq I(Y_2; U_2|U_1, Q)$$

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

for alphabets $\mathcal{U}_1$,$\mathcal{U}_2$, $\mathcal{Q}$ satisfying $|\mathcal{U}_i| \leq |\mathcal{Y}_i|$ and $|Q| \leq 5$.

Typically, the quality of a bet in a horse race market is measured by the associated doubling rate (cf. [CT06]). Theorem 6 implies that the optimal doubling rate for the product-wager is given by:

$$\begin{aligned} W^*_{\text{p-w}}(p(y_1,y_2)) &= \sum_{y_1,y_2} p(y_1,y_2) \log b_1^*(y_1) b_2^*(y_2) o(y_1,y_2) \\ &= \mathbb{E} \log o(Y_1,Y_2) - \inf_{p \in \mathcal{P}(R_1,R_2)} \left\{ H(Y_1|U_1,U_2,Q) + H(Y_2|U_1,U_2,Q) \right\}. \end{aligned}$$

Likewise, Theorem 4 implies that the optimal doubling rate for the joint-wager is given by:

$$\begin{aligned} W^*_{\text{j-w}}(p(y_1,y_2)) &= \sum_{y_1,y_2} p(y_1,y_2) \log b^*(y_1,y_2) o(y_1,y_2) \\ &= \mathbb{E} \log o(Y_1,Y_2) + \min\{R_1 - H(Y_1|Y_2), R_2 - H(Y_2|Y_1), \\ &\qquad\qquad R_1 + R_2 - H(Y_1,Y_2)\}. \end{aligned}$$

It is important to note that we do not require the side informations to be the same for each type of wager, rather, the side informations are only provided at the same rates. Thus, the gambler placing the joint-wager receives side information at rates (R_1, R_2) that maximizes his doubling rate, while the gambler placing the product-wager receives (potentially different) side information at rates (R_1, R_2) that maximizes his doubling rate. However, as we will see shortly, for any rates (R_1, R_2), there always exists rate-limited side information which simultaneously allows each type of gambler to attain their maximum doubling rate.

By combining the expressions for $W^*_{\text{p-w}}(p(y_1,y_2))$ and $W^*_{\text{j-w}}(p(y_1,y_2))$, we find that the difference in doubling rates is given by:

$$\begin{aligned}
\Delta(R_1, R_2) &= W^*_{\text{j-w}}(p(y_1,y_2)) - W^*_{\text{p-w}}(p(y_1,y_2))\\
&= \min\Big\{R_1 - H(Y_1|Y_2), R_2 - H(Y_2|Y_1), R_1 + R_2 - H(Y_1,Y_2)\Big\}\\
&\quad + \inf_{p\in\mathcal{P}(R_1,R_2)} \{H(Y_1|U_1,U_2,Q) + H(Y_2|U_1,U_2,Q)\} \qquad (2.36)\\
&= \inf_{p\in\mathcal{P}(R_1,R_2)} \min\Big\{R_1 - I(Y_1;U_1|U_2,Q) + I(Y_1;Y_2) - I(Y_1;U_2,Q) + H(Y_2|U_1,U_2,Q),\\
&\quad R_2 - I(Y_2;U_2|U_1,Q) + I(Y_2;Y_1) - I(Y_2;U_1,Q) + H(Y_1|U_1,U_2,Q),\\
&\quad R_1 + R_2 - I(Y_1,Y_2;U_1,U_2|Q) + I(Y_1;Y_2|U_1,U_2,Q)\Big\}\\
&= \inf_{p\in\mathcal{P}(R_1,R_2)} I(Y_1;Y_2|U_1,U_2,Q). \qquad (2.37)
\end{aligned}$$

The final equality (2.37) can be deduced by looking at each of the three terms in the minimum individually. To this end, note that:

- We have $R_1 \geq I(Y_1;U_1|U_2,Q)$ for any $p \in \mathcal{P}(R_1,R_2)$, and $I(Y_1;Y_2) \geq I(Y_1;U_2,Q)$ by the data processing inequality. Hence

$$\begin{aligned}
&R_1 - I(Y_1;U_1|U_2,Q) + I(Y_1;Y_2) - I(Y_1;U_2,Q) + H(Y_2|U_1,U_2,Q)\\
&\quad \geq H(Y_2|U_1,U_2,Q)\\
&\quad \geq I(Y_1;Y_2|U_1,U_2,Q).
\end{aligned}$$

- By a symmetric argument, we can lower bound the second term in a similar fashion:

$$\begin{aligned}
&R_2 - I(Y_2;U_2|U_1,Q) + I(Y_2;Y_1) - I(Y_2;U_1,Q) + H(Y_1|U_1,U_2,Q)\\
&\quad \geq I(Y_1;Y_2|U_1,U_2,Q).
\end{aligned}$$

- By Lemma 9 (see Appendix 2.D), the infimum in (2.36) is attained by a $p \in \mathcal{P}(R_1,R_2)$ satisfying $R_1 + R_2 = I(Y_1,Y_2;U_1,U_2|Q)$. Therefore, the third term in the minimum is given by

$$R_1 + R_2 - I(Y_1,Y_2;U_1,U_2|Q) + I(Y_1;Y_2|U_1,U_2,Q) = I(Y_1;Y_2|U_1,U_2,Q),$$

  and (2.37) follows immediately.

Let $p^* \in \mathcal{P}(R_1, R_2)$ be the distribution that attains the infimum in (2.36) (such a p^* always exists), then (2.37) yields

$$\begin{aligned}
&W^*_{\text{j-w}}(p(y_1, y_2)) - W^*_{\text{p-w}}(p(y_1, y_2)) \\
&= \sum_{u_1,u_2,q} p^*(u_1, u_2, q) \sum_{y_1,y_2} p^*(y_1, y_2|u_1, u_2, q) \log \frac{p^*(y_1, y_2|u_1, u_2, q)}{p^*(y_1|u_1, u_2, q)p^*(y_2|u_1, u_2, q)} \\
&= \mathbb{E}_{p^*} \log o(Y_1, Y_2)p^*(Y_1, Y_2|U_1, U_2, Q) \\
&\quad - \mathbb{E}_{p^*} \log o(Y_1, Y_2)p^*(Y_1|U_1, U_2, Q)p^*(Y_2|U_1, U_2, Q).
\end{aligned}$$

Hence, we can interpret the auxiliary random variables corresponding to p^* as optimal rate-limited side informations for *both* betting strategies. Moreover, optimal bets for each strategy are given by

1. $b^*(y_1, y_2) = p^*(y_1, y_2|u_1, u_2, q)$ for the joint-wager, and
2. $b^*_1(y_1) = p^*(y_1|u_1, u_2, q),\ b^*_2(y_2) = p^*(y_2|u_1, u_2, q)$ for the product-wager.

Since $\mathcal{P}(R_1, R_2) \subseteq \mathcal{P}(R'_1, R'_2)$ for $R_1 \leq R'_1$ and $R_2 \leq R'_2$, the function $\Delta(R_1, R_2)$ is nonincreasing in R_1 and R_2. Thus, the benefits of using the joint-wager over the product-wager diminish in the amount of side-information available. It is also not difficult to show that $\Delta(R_1, R_2)$ is jointly convex in (R_1, R_2).

Furthermore, for rate-pairs (R_1, R_2) and (R'_1, R'_2) satisfying $R_1 < R'_1$ and $R_2 < R'_2$, there exist corresponding optimal joint- and product-wagers $b^*(y_1, y_2)$ and $b^*_1(y_1)b^*_2(y_2)$, and $b^{*'}(y_1, y_2)$ and $b^{*'}_1(y_1)b^{*'}_2(y_2)$, respectively, satisfying

$$D\Big(b^{*'}(y_1, y_2)\Big|\Big|b^{*'}_1(y_1)b^{*'}_2(y_2)\Big) < D\Big(b^*(y_1, y_2)\Big|\Big|b^*_1(y_1)b^*_2(y_2)\Big). \tag{2.38}$$

Hence, roughly speaking, the joint-wager and product-wager look “more alike” as the amount of side information is increased. The proof of the strict inequality in (2.38) can be inferred from the proof of Lemma 9 in Appendix 2.D.

To conclude this example, we note that $\Delta(R_1, R_2)$ enjoys a great deal of symmetry near the origin in the sense that side information from either encoder contributes approximately the same amount to the improvement of the product-wager. We state this formally as a theorem:

Theorem 8. *Define $\rho_m(Y_1, Y_2)$ to be the Hirschfeld-Gebelein-Rényi maximal correlation between random variables Y_1 and Y_2. Then, $\Delta(R_1, R_2) \geq I(Y_1; Y_2) - \rho_m^2(Y_1, Y_2) \cdot (R_1 + R_2)$. Moreover, this bound is tight as $(R_1, R_2) \to (0, 0)$.*

Proof. If $R_2 = 0$, then it is readily verified that $\Delta(R_1, 0)$ can be expressed as follows:

$$\Delta(R_1, 0) = I(Y_1; Y_2) - \max_{\substack{p(u_1|y_1): I(U_1;Y_1)=R_1, \\ U_1 \to Y_1 \to Y_2,\ |\mathcal{U}_1| \leq |\mathcal{Y}_1|+1}} I(U_1; Y_2).$$

By symmetry:

$$\Delta(0, R_2) = I(Y_1; Y_2) - \max_{\substack{p(u_2|y_2): I(U_2;Y_2)=R_2, \\ U_2 \to Y_2 \to Y_1,\ |\mathcal{U}_2| \leq |\mathcal{Y}_2|+1}} I(U_2; Y_1).$$

Here, we can apply a result of Erkip [Erk96, Theorem 10] to evaluate the gradient of $\Delta(R_1, R_2)$ at $(R_1, R_2) = (0, 0)$:

$$\frac{\partial}{\partial R_1} \Delta(R_1, R_2) \bigg|_{(R_1,R_2)=(0,0)} = \frac{\partial}{\partial R_2} \Delta(R_1, R_2) \bigg|_{(R_1,R_2)=(0,0)} = -\rho_m^2(Y_1, Y_2). \tag{2.39}$$

Note, since $\Delta(R_1, 0)$ and $\Delta(0, R_2)$ are each convex in their respective variable and $\Delta(0, 0) = I(Y_1; Y_2)$, we have

$$\begin{aligned} \Delta(R_1, 0) &\geq I(Y_1; Y_2) - \rho_m^2(Y_1, Y_2) R_1 \\ \Delta(0, R_2) &\geq I(Y_1; Y_2) - \rho_m^2(Y_1, Y_2) R_2. \end{aligned} \tag{2.40}$$

Taking this one step further, for $\nu_1, \nu_2 > 0$, we can evaluate the one-sided derivative:

$$\lim_{\lambda \downarrow 0} \frac{\Delta(\lambda \nu_1, \lambda \nu_2) - \Delta(0, 0)}{\lambda} = -\rho_m^2(Y_1, Y_2) \cdot (\nu_1 + \nu_2). \tag{2.41}$$

We remark that (2.41) does not follow immediately from (2.39) since the point at which we are taking the derivatives (i.e., the origin) does not lie in an open neighborhood of the domain. Nonetheless, the expected result holds.

Since $\Delta(R_1, R_2)$ is convex, we obtain an upper bound on the one-sided derivative as follows:

$$\begin{aligned}
\lim_{\lambda \downarrow 0} \frac{\Delta(\lambda\nu_1, \lambda\nu_2) - \Delta(0,0)}{\lambda} &\leq \lim_{\lambda \downarrow 0} \frac{\frac{1}{2}\Delta(2\lambda\nu_1, 0) + \frac{1}{2}\Delta(0, 2\lambda\nu_2) - \Delta(0,0;p)}{\lambda} \\
&= \frac{1}{2}\lim_{\lambda \downarrow 0} \frac{\Delta(\lambda 2\nu_1, 0) - \Delta(0,0)}{\lambda} + \frac{1}{2}\lim_{\lambda \downarrow 0} \frac{\Delta(0, \lambda 2\nu_2) - \Delta(0,0)}{\lambda} \\
&= -\rho_m^2(Y_1, Y_2) \cdot (\nu_1 + \nu_2),
\end{aligned}$$

where the final equality follows by (2.39) and the positive homogeneity of the directional derivative.

Therefore, to complete the proof of (2.41), it suffices to prove the lower bound

$$\lim_{\lambda \downarrow 0} \frac{\Delta(\lambda\nu_1, \lambda\nu_2) - \Delta(0,0)}{\lambda} \geq -\rho_m^2(Y_1, Y_2) \cdot (\nu_1 + \nu_2).$$

To this end, fix $\lambda, \nu_1, \nu_2 > 0$ and observe that

$$\begin{aligned}
&\frac{\Delta(\lambda\nu_1, \lambda\nu_2) - \Delta(0,0)}{\lambda} \\
&\quad= \frac{1}{\lambda} \inf_{p \in \mathcal{P}(\lambda\nu_1, \lambda\nu_2)} \Big\{ I(Y_1; Y_2 | U_1, U_2 | Q) - I(Y_1; Y_2) \Big\} && (2.42) \\
&\quad= \frac{1}{\lambda} \inf_{p \in \mathcal{P}(\lambda\nu_1, \lambda\nu_2)} \Big\{ I(Y_1, Y_2; U_1, U_2 | Q) - I(Y_1; U_1, U_2 | Q) - I(Y_2; U_1, U_2 | Q) \Big\} \\
&\quad= (\nu_1 + \nu_2) - \frac{1}{\lambda}\Big(I_{p^*}(Y_1; U_1, U_2 | Q) + I_{p^*}(Y_2; U_1, U_2 | Q) \Big) && (2.43) \\
&\quad= (\nu_1 + \nu_2) - \frac{1}{\lambda}\Big(I_{p^*}(Y_1; U_1 | U_2, Q) + I_{p^*}(Y_1; U_2 | Q) \\
&\qquad\qquad + I_{p^*}(Y_2; U_2 | U_1, Q) + I_{p^*}(Y_2; U_1 | Q) \Big) \\
&\quad\geq (\nu_1 + \nu_2) - \rho_m^2(Y_1, Y_2)\,(2\nu_1 + 2\nu_2) \\
&\qquad - \frac{(1 - \rho_m^2(Y_1, Y_2))}{\lambda}\,(I_{p^*}(Y_1; U_1 | U_2, Q) + I_{p^*}(Y_2; U_2 | U_1, Q)) && (2.44) \\
&\quad= -\rho_m^2(Y_1, Y_2)\,(\nu_1 + \nu_2) + (1 - \rho_m^2(Y_1, Y_2))\,(\nu_1 + \nu_2) \\
&\qquad - \frac{(1 - \rho_m^2(Y_1, Y_2))}{\lambda}\,(I_{p^*}(Y_1; U_1 | U_2, Q) + I_{p^*}(Y_2; U_2 | U_1, Q)) \\
&\quad\geq -\rho_m^2(Y_1, Y_2)\,(\nu_1 + \nu_2)\,. && (2.45)
\end{aligned}$$

In the above string of inequalities

- (2.42) follows by definition of $\Delta(R_1, R_2)$.

- Equality (2.43) follows since Lemma 9 guarantees that the infimum is attained in (2.42) for some $p^* \in \mathcal{P}(\lambda\nu_1, \lambda\nu_2)$ satisfying $I_{p^*}(Y_1, Y_2; U_1, U_2|Q) = \lambda(\nu_1 + \nu_2)$. Here, we write $I_{p^*}(Y_1, Y_2; U_1, U_2|Q)$ to denote the mutual information $I(Y_1, Y_2; U_1, U_2|Q)$ evaluated for the distribution p^*.

- To see that (2.44) holds, note that

  $$I_{p^*}(Y_2; U_2|Q) = \lambda\nu_1 + \lambda\nu_2 - I_{p^*}(Y_1; U_1|U_2, Q),$$

  and thus

  $$\begin{aligned} &I(Y_1; Y_2) - \rho_m^2(Y_1, Y_2)\left(\lambda\nu_1 + \lambda\nu_2 - I_{p^*}(Y_1; U_1|U_2, Q)\right) \\ &\quad \le \Delta(0, \lambda\nu_1 + \lambda\nu_2 - I_{p^*}(Y_1; U_1|U_2, Q)) && (2.46) \\ &\quad = I(Y_1; Y_2) - \max_{\substack{p(\tilde{u}_2|y_2): I(Y_2;\tilde{U}_2) \le \lambda\nu_1 + \lambda\nu_2 - I_{p^*}(Y_1;U_1|U_2,Q), \\ \tilde{U}_2 \leftrightarrow Y_2 \leftrightarrow Y_1}} I(\tilde{U}_2; Y_1) && (2.47) \\ &\quad \le I(Y_1; Y_2) - I_{p^*}(Y_1; U_2|Q), && (2.48) \end{aligned}$$

  which implies

  $$-\rho_m^2(Y_1, Y_2)\left(\lambda\nu_1 + \lambda\nu_2 - I_{p^*}(Y_1; U_1|U_2, Q)\right) \le -I_{p^*}(Y_1; U_2|Q).$$

  The above steps are justified as follows:

  - (2.46) follows from (2.40).
  - (2.47) follows by definition of the function $\Delta(0, x)$.
  - (2.48) follows since Q is independent of Y_1, Y_2 (by definition of p^*), and thus $\tilde{U}_2 = (U_2, Q)$ lies in the set over which we take the maximum in (2.47).

  By symmetry, we conclude that

  $$\begin{aligned} &-\left(I_{p^*}(Y_1; U_2|Q) + I_{p^*}(Y_2; U_1|Q)\right) \\ &\quad \ge -\rho_m^2(Y_1, Y_2)\left(2\lambda\nu_1 + 2\lambda\nu_2 - I_{p^*}(Y_1; U_1|U_2, Q) - I_{p^*}(Y_2; U_2|U_1, Q)\right), \end{aligned}$$

  and (2.44) follows.

- (2.45) follows since $\lambda\nu_1 \geq I_{p^*}(Y_1; U_1|U_2, Q)$ and $\lambda\nu_2 \geq I_{p^*}(Y_2; U_2|U_1, Q)$ for $p^* \in \mathcal{P}(\lambda\nu_1, \lambda\nu_2)$.

□

2.4.5 An Application: List Decoding

In the previous example, we did not take advantage of the strengthened converse result (i.e., Theorem 7) which we proved in Section 2.4.3. In this section, we give an application that requires this strengthening.

Formally, a 2-list code (of blocklength n consists) of encoding functions:

$$g_i^{(n)} : \mathcal{Y}_i^n \to \left\{1, \ldots, M_i^{(n)}\right\} \text{ for } i = 1, 2$$

and list decoding functions

$$\begin{aligned} L_1^{(n)} &: \left\{1, \ldots, M_1^{(n)}\right\} \times \left\{1, \ldots, M_2^{(n)}\right\} \to 2^{\mathcal{Y}_1^n} \\ L_2^{(n)} &: \left\{1, \ldots, M_1^{(n)}\right\} \times \left\{1, \ldots, M_2^{(n)}\right\} \to 2^{\mathcal{Y}_2^n}. \end{aligned}$$

A list decoding tuple $(R_1, R_2, \Delta_1, \Delta_2)$ is achievable if, for any $\epsilon > 0$, there exists a 2-list code of blocklength n satisfying the rate constraints

$$\begin{aligned} \frac{1}{n} \log M_1^{(n)} &\leq R_1 + \epsilon \\ \frac{1}{n} \log M_2^{(n)} &\leq R_2 + \epsilon, \end{aligned}$$

and the probability of list-decoding error constraints

$$\begin{aligned} \Pr\left[Y_1^n \notin L_1^{(n)}\left(g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n)\right)\right] &\leq \epsilon, \\ \Pr\left[Y_2^n \notin L_2^{(n)}\left(g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n)\right)\right] &\leq \epsilon. \end{aligned}$$

with list sizes

$$\frac{1}{n}\log|L_1^{(n)}| \leq \Delta_1 + \epsilon$$
$$\frac{1}{n}\log|L_2^{(n)}| \leq \Delta_2 + \epsilon.$$

With a 2-list code so defined, the following theorem shows that the 2-list decoding problem and multiterminal source coding problem under logarithmic loss are equivalent (inasmuch as the achievable regions are identical):

Theorem 9. *The list decoding tuple $(R_1, R_2, \Delta_1, \Delta_2)$ is achievable if and only if*

$$\begin{aligned} R_1 &\geq I(U_1; Y_1|U_2, Q) \\ R_2 &\geq I(U_2; Y_2|U_1, Q) \\ R_1 + R_2 &\geq I(U_1, U_2; Y_1, Y_2|Q) \\ \Delta_1 &\geq H(Y_1|U_1, U_2, Q) \\ \Delta_2 &\geq H(Y_2|U_1, U_2, Q). \end{aligned}$$

for some joint distribution

$$p(y_1, y_2, u_1, u_2, q) = p(y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)p(q),$$

where $|\mathcal{U}_1| \leq |\mathcal{Y}_1|$, $|\mathcal{U}_2| \leq |\mathcal{Y}_2|$, and $|\mathcal{Q}| \leq 5$.

Remark 2. *We note that a similar connection to list decoding can be made in the context of the CEO problem.*

To prove the theorem, we require a slightly modified version of [KSC08, Lemma 1]:

Lemma 4. *If the list decoding tuple $(R_1, R_2, \Delta_1, \Delta_2)$ is achieved by a sequence of 2-list codes $\{g_1^{(n)}, g_2^{(n)}, L_1^{(n)}, L_2^{(n)}\}_{n\to\infty}$, then*

$$H(Y_1^n|g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n)) \leq |L_1^{(n)}| + n\epsilon_n$$
$$H(Y_2^n|g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n)) \leq |L_2^{(n)}| + n\epsilon_n,$$

where $\epsilon_n \to 0$ as $n \to \infty$.

Proof. The proof is virtually identical to that of [KSC08, Lemma 1], and is therefore omitted. □

Proof of Theorem 9. First observe that the direct part is trivial. Indeed, for a joint distribution $p(y_1, y_2, u_1, u_2, q) = p(y_1, y_2)p(u_1|y_1, q)p(u_2|y_2, q)p(q)$, apply the Berger-Tung achievability scheme and take $L_i^{(n)}$ to be the set of y_i^n sequences which are jointly typical with the decoded quantizations (U_1^n, U_2^n). This set has cardinality no larger than $2^{n(H(Y_i|U_1,U_2,Q)+\epsilon)}$, which proves achievability.

To see the converse, note that setting

$$\hat{Y}_i^n = \Pr\left[Y_i^n|g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n)\right]$$

achieves a logarithmic loss of $\frac{1}{n}H(Y_i^n|g_1^{(n)}(Y_1^n), g_2^{(n)}(Y_2^n))$ for source i in the setting where reproductions are not restricted to product distributions. Applying Theorem 7 together with Lemma 4 yields the desired result. □

2.5 Relationship to the General Multiterminal Source Coding Problem

In this section, we relate our results on logarithmic loss to multiterminal source coding problems with arbitrary distortion measures and reproduction alphabets.

As before, we let $\{Y_{1,j}, Y_{2,j}\}_{j=1}^n$ be a sequence of n independent, identically distributed random variables with finite alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, respectively, and joint pmf $p(y_1, y_2)$.

In this section, the reproduction alphabets $\breve{\mathcal{Y}}_i$, $i = 1, 2$, are arbitrary. We also consider generic distortion measures:

$$\breve{d}_i : \mathcal{Y}_i \times \breve{\mathcal{Y}}_i \to \mathbb{R}^+ \text{ for } i = 1, 2,$$

where $\mathbb{R}^+$ denotes the set of nonnegative real numbers. The sequence distortion is then defined as follows:

$$\breve{d}_i(y_i^n, \breve{y}_i^n) = \frac{1}{n}\sum_{j=1}^{n} \breve{d}_i(y_{i,j}, \breve{y}_{i,j}).$$

We will continue to let $d(\cdot,\cdot)$ and $\hat{\mathcal{Y}}_1, \hat{\mathcal{Y}}_2$ denote the logarithmic loss distortion measure and the associated reproduction alphabets, respectively.

A rate distortion code (of blocklength n) consists of encoding functions:

$$\breve{g}_i^{(n)} : \mathcal{Y}_i^n \to \left\{1, \ldots, M_i^{(n)}\right\} \text{ for } i = 1, 2$$

and decoding functions

$$\breve{\psi}_i^{(n)} : \left\{1, \ldots, M_1^{(n)}\right\} \times \left\{1, \ldots, M_2^{(n)}\right\} \to \breve{\mathcal{Y}}_i^n \text{ for } i = 1, 2.$$

A rate distortion vector (R_1, R_2, D_1, D_2) is strict-sense achievable if there exists a blocklength n, encoding functions $\breve{g}_1^{(n)}, \breve{g}_2^{(n)}$ and a decoder $(\breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)})$ such that

$$R_i \geq \frac{1}{n}\log M_i^{(n)} \text{ for } i = 1, 2 \tag{2.49}$$

$$D_i \geq \mathbb{E}\breve{d}_i(Y_i^n, \breve{Y}_i^n) \text{ for } i = 1, 2. \tag{2.50}$$

Where

$$\breve{Y}_i^n = \breve{\psi}_i^{(n)}(\breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)) \text{ for } i = 1, 2.$$

For these functions, we define the quantity

$$\beta_i\left(\breve{g}_1^{(n)}, \breve{g}_2^{(n)}, \breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)}\right) := \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\log\left(\sum_{y_i \in \mathcal{Y}_i} 2^{-\breve{d}_i(y_i, \breve{Y}_{i,j})}\right) \text{ for } i = 1, 2. \tag{2.51}$$

Now, let $\beta_i(R_1, R_2, D_1, D_2)$ be the infimum of the $\beta_i\left(\breve{g}_1^{(n)}, \breve{g}_2^{(n)}, \breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)}\right)$'s, where the infimum is taken over all codes that achieve the rate distortion vector (R_1, R_2, D_1, D_2).

At this point it is instructive to pause and consider some examples.

Example 1 (Binary Sources and Hamming Distortion). *For $i = 1, 2$, let $\breve{\mathcal{Y}}_i = \mathcal{Y}_i = \{0, 1\}$ and let $\breve{d}_i$ be the α-scaled Hamming distortion measure:*

$$\breve{d}_i(y_i, \breve{y}_i) = \begin{cases} 0 & \text{if } \breve{y}_i = y_i, \\ \alpha & \text{if } \breve{y}_i \neq y_i. \end{cases}$$

In this case,

$$\sum_{y_i \in \mathcal{Y}_i} 2^{-\breve{d}_i(y_i, \breve{Y}_{i,j})} = 2^0 + 2^{-\alpha}, \tag{2.52}$$

so $\beta_i(R_1, R_2, D_1, D_2) = \log(1 + 2^{-\alpha})$ for any (R_1, R_2, D_1, D_2). This notion that $\beta_i(R_1, R_2, D_1, D_2)$ is a constant extends to all distortion measures for which the columns of the $|\mathcal{Y}_i| \times |\breve{\mathcal{Y}}_i|$ distortion matrix are permutations of one another.

Example 2 (Binary Sources and Erasure Distortion). *For $i = 1, 2$, let $\mathcal{Y}_i = \{0, 1\}$, $\breve{\mathcal{Y}}_i = \{0, 1, e\}$ and let $\breve{d}_i$ be the standard erasure distortion measure:*

$$\breve{d}_i(y_i, \breve{y}_i) = \begin{cases} 0 & \text{if } \breve{y}_i = y_i \\ 1 & \text{if } \breve{y}_i = e \\ \infty & \text{if } \breve{y}_i \in \{0, 1\} \text{ and } \breve{y}_i \neq y_i. \end{cases}$$

In this case,

$$\sum_{y_i \in \mathcal{Y}_i} 2^{-\breve{d}_i(y_i, \breve{Y}_{i,j})} = \begin{cases} 2^{-\infty} + 2^0 = 1 & \text{if } \breve{Y}_{i,j} \in \{0, 1\} \\ 2^{-1} + 2^{-1} = 1 & \text{if } \breve{Y}_{i,j} = e. \end{cases} \tag{2.53}$$

so $\beta_i(R_1, R_2, D_1, D_2) = 0$ for any (R_1, R_2, D_1, D_2). This result can easily be extended to erasure distortion on larger alphabets by setting the penalty to $\log |\mathcal{Y}_i|$ when $\breve{Y}_i = e$.

Theorem 10. *Suppose (R_1, R_2, D_1, D_2) is strict-sense achievable for the general multiterminal source coding problem. Then*

$$\left.\begin{aligned} R_1 &\geq I(U_1; Y_1 | U_2, Q) \\ R_2 &\geq I(U_2; Y_2 | U_1, Q) \\ R_1 + R_2 &\geq I(U_1, U_2; Y_1, Y_2 | Q) \\ D_1 &\geq H(Y_1 | U_1, U_2, Q) - \beta_1(R_1, R_2, D_1, D_2) \\ D_2 &\geq H(Y_2 | U_1, U_2, Q) - \beta_2(R_1, R_2, D_1, D_2) \end{aligned}\right\} \tag{2.54}$$

for some joint distribution $p(y_1, y_2)p(q)p(u_1|y_1, q)p(u_2|y_2, q)$ with $|\mathcal{U}_i| \leq |\mathcal{Y}_i|$ and $|\mathcal{Q}| \leq 5$.

Proof. Since (R_1, R_2, D_1, D_2) is strict-sense achievable, there exists a blocklength n, encoding functions $\breve{g}_1^{(n)}, \breve{g}_2^{(n)}$ and a decoder $(\breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)})$ satisfying (2.49)-(2.50). Given these functions, the decoder can generate reproductions $\breve{Y}_1^n, \breve{Y}_2^n$ satisfying the average distortion constraints (2.50). From the reproduction $\breve{Y}_i^n$, we construct the reproduction $\hat{Y}_i^n$ as follows:

$$\hat{Y}_j(y_i) = \frac{2^{-\breve{d}_i(y_i, \breve{Y}_{i,j})}}{\sum_{y_i' \in \mathcal{Y}_i} 2^{-\breve{d}_i(y_i', \breve{Y}_{i,j})}}.$$

Now, using the logarithmic loss distortion measure, observe that $\hat{Y}_i^n$ satisfies

$$\begin{aligned}
\mathbb{E}d(Y_i^n, \hat{Y}_i^n) &= \frac{1}{n}\sum_{j=1}^n \mathbb{E}\log\left(2^{\breve{d}_i(Y_{i,j}, \breve{Y}_{i,j})}\right) + \frac{1}{n}\sum_{j=1}^n \mathbb{E}\log\left(\sum_{y_i' \in \mathcal{Y}_i} 2^{-\breve{d}_i(y_i', \breve{Y}_{i,j})}\right)\\
&= \frac{1}{n}\sum_{j=1}^n \mathbb{E}\breve{d}_i(Y_{i,j}, \breve{Y}_{i,j}) + \beta_i\left(\breve{g}_1^{(n)}, \breve{g}_2^{(n)}, \breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)}\right)\\
&\le D_i + \beta_i\left(\breve{g}_1^{(n)}, \breve{g}_2^{(n)}, \breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)}\right)\\
&:= \tilde{D}_i.
\end{aligned}$$

Thus, $(R_1, R_2, \tilde{D}_1, \tilde{D}_2)$ is achievable for the multiterminal source coding problem with the logarithmic loss distortion measure. Applying Theorem 6 and taking the infimum over all coding schemes that achieve (R_1, R_2, D_1, D_2) proves the theorem. □

This outer bound is interesting because the region is defined over the same set of probability distributions that define the Berger-Tung inner bound. While the β_i's can be difficult to compute in general, we have shown that they can be readily determined for many popular distortion measures. As an application, we now give a quantitative approximation of the rate distortion region for binary sources subject to Hamming distortion constraints. Before proceeding, we prove the following lemma.

Lemma 5. *Suppose $(R_1, R_2, \tilde{D}_1, \tilde{D}_2)$ is strict-sense achievable for the multiterminal source coding problem with binary sources and $\breve{d}_i$ equal to the α_i-scaled Hamming distortion measure, for $i = 1, 2$. Then the Berger-Tung achievability scheme can achieve a point*

(R_1, R_2, D_1, D_2) satisfying

$$D_i - \tilde{D}_i \leq \left(\frac{\alpha_i}{2} - 1\right) H_i + \log(1 + 2^{-\alpha_i})$$

for some $H_i \in [0,1]$, $i = 1, 2$.

Proof. By Theorem 10, $(R_1, R_2, \tilde{D}_1, \tilde{D}_2)$ satisfy (2.54) for some joint distribution $p(y_1, y_2)p(q)p(u_1|y_1, q)p(u_2|y_2, q)$. For this distribution, define the reproduction functions

$$\breve{Y}_i(U_1, U_2, Q) = \arg\max_{y_i} p(y_i|U_1, U_2, Q) \text{ for } i = 1, 2. \tag{2.55}$$

Then, observe that for $i = 1, 2$:

$$\begin{aligned}
\mathbb{E}\breve{d}_i(Y_i, \breve{Y}_i) &= \sum_{u_1,u_2,q} p(u_1, u_2, q) \left[\alpha_i \cdot \min_{y_i} p(y_i|u_1, u_2, q) + 0 \cdot \max_{y_i} p(y_i|u_1, u_2, q)\right] \\
&= \alpha_i \sum_{u_1,u_2,q} p(u_1, u_2, q) \cdot \min_{y_i} p(y_i|u_1, u_2, q) \\
&\leq \frac{\alpha_i}{2} \sum_{u_1,u_2,q} p(u_1, u_2, q) \cdot H(Y_i|U_1, U_2, Q = u_1, u_2, q) \qquad (2.56) \\
&= \frac{\alpha_i}{2} H(Y_i|U_1, U_2, Q).
\end{aligned}$$

Where (2.56) follows from the fact that $2p \leq h_2(p)$ for $0 \leq p \leq 0.5$. Thus, $D_i = \frac{\alpha_i}{2} H(Y_i|U_1, U_2, Q)$ is achievable for rates (R_1, R_2) using the Berger-Tung achievability scheme. Combining this with the fact that $\tilde{D}_i \geq H(Y_i|U_1, U_2, Q) - \log(1 + 2^{-\alpha_i})$, we see that

$$D_i - \tilde{D}_i \leq \frac{\alpha_i}{2} H(Y_i|U_1, U_2, Q) - H(Y_i|U_1, U_2, Q) + \log(1 + 2^{-\alpha_i}).$$

□

Lemma 5 allows us to give a quantitative outer bound on the achievable rate distortion region in terms of the Berger-Tung inner bound.

Corollary 1. *Suppose $(R_1, R_2, \tilde{D}_1^{(1)}, \tilde{D}_2^{(1)})$ is strict-sense achievable for the multiterminal source coding problem with binary sources and $\breve{d}_i$ equal to the standard 1-scaled Hamming* *distortion measure, for $i = 1, 2$. Then the Berger-Tung achievability scheme can achieve a point $(R_1, R_2, D_1^{(1)}, D_2^{(1)})$, where*

$$D_i^{(1)} - \tilde{D}_i^{(1)} \leq \frac{1}{2} \log \left(\frac{5}{4} \right) < 0.161 \textit{ for } i = 1, 2.$$

Proof. For rates (R_1, R_2), note that distortions $(\tilde{D}_1, \tilde{D}_2)$ are strict-sense achievable for the α_i-scaled Hamming distortion measures if and only if distortions $(\tilde{D}_1^{(1)}, \tilde{D}_2^{(1)}) = (\frac{1}{\alpha_1}\tilde{D}_1, \frac{1}{\alpha_2}\tilde{D}_2)$ are strict-sense achievable for the 1-scaled Hamming distortion measure. Likewise, the point (R_1, R_2, D_1, D_2) is achieved by the Berger-Tung coding scheme for the α_i-scaled Hamming distortion measures if and only if $(R_1, R_2, \frac{1}{\alpha_1}D_1, \frac{1}{\alpha_2}D_2)$ is achieved by the Berger-Tung coding scheme for the 1-scaled Hamming distortion measure.

Thus, applying Lemma 5, we can use the Berger-Tung achievability scheme to achieve a point $(R_1, R_2, D_1^{(1)}, D_2^{(1)})$ satisfying

$$\begin{aligned} D_i^{(1)} - \tilde{D}_i^{(1)} &= \frac{1}{\alpha_i} \left(D_i - \tilde{D}_i \right) \\ &\leq \frac{1}{\alpha_i} \left(\frac{\alpha_i}{2} - 1 \right) H_i + \frac{1}{\alpha_i} \log(1 + 2^{-\alpha_i}) \\ &= \left(\frac{1}{2} - \frac{1}{\alpha_i} \right) H_i + \frac{1}{\alpha_i} \log(1 + 2^{-\alpha_i}) \end{aligned} \tag{2.57}$$

for some $H_i \in [0, 1]$. We can optimize (2.57) over α_i to find the minimum gap for a given H_i. Maximizing over $H_i \in [0, 1]$ then gives the worst-case gap. Straightforward calculus yields the saddle-point:

$$\begin{aligned} &\max_{H_i \in [0,1]} \inf_{\alpha_i > 0} \left\{ \left(\frac{1}{2} - \frac{1}{\alpha_i} \right) H_i + \frac{1}{\alpha_i} \log(1 + 2^{-\alpha_i}) \right\} \\ &\quad = \inf_{\alpha_i > 0} \max_{H_i \in [0,1]} \left\{ \left(\frac{1}{2} - \frac{1}{\alpha_i} \right) H_i + \frac{1}{\alpha_i} \log(1 + 2^{-\alpha_i}) \right\} \\ &\quad = \frac{1}{2} \log \left(\frac{5}{4} \right) < 0.161, \end{aligned}$$

which is achieved for $\alpha_i = 2$ and any $H \in [0, 1]$. □

In the above results, we did not require the strengthening that Theorem 7 provides over Theorem 6. However, our next result takes this strengthened converse into account.

To motivate the final results of this section, we begin with a discussion of the two-encoder quadratic Gaussian source coding problem. To this end, suppose Y_1, Y_2 are jointly Gaussian – each with unit variance and correlation ρ – and distortion is measured under mean square error. In this setting, the set of achievable rate distortion tuples (R_1, R_2, D_1, D_2) is given by

$$R_1 \geq \frac{1}{2}\log\left(\frac{1}{D_1}\left(1-\rho^2+\rho^2 2^{-2R_2}\right)\right) \tag{2.58}$$

$$R_2 \geq \frac{1}{2}\log\left(\frac{1}{D_2}\left(1-\rho^2+\rho^2 2^{-2R_1}\right)\right) \tag{2.59}$$

$$R_1 + R_2 \geq \frac{1}{2}\log\left(\frac{(1-\rho^2)\beta(D_1,D_2)}{2D_1D_2}\right), \tag{2.60}$$

where

$$\beta(D_1,D_2) = 1 + \sqrt{1 + \frac{4\rho^2 D_1 D_2}{(1-\rho^2)^2}}.$$

Long before the converse result was completed in [WTV08], it was known that any (R_1, R_2, D_1, D_2) satisfying (2.58)-(2.60) was achievable. Indeed, (2.58)-(2.60) correspond to a set of points in the Berger-Tung inner bound [Ber77, Tun78] achieved by Gausian test channels. Moreover, roughly a decade before the sum-rate lower bound (2.60) was established in [WTV08], it was proven by Oohama [Ooh97] that (2.58)-(2.59) were necessary conditions for (R_1, R_2, D_1, D_2) to be achievable. Thus, in the period between [Ooh97] and [WTV08], ad-hoc lower bounds on the sum-rate could be established as follows.

Noting that the right hand sides of (2.58) and (2.59) are convex in R_2 and R_2, respectively, it is straightforward to establish the necessity of

$$R_1 + \rho^2 R_2 \geq \frac{1}{2}\log\left(\frac{1}{D_1}\right) \tag{2.61}$$

$$R_2 + \rho^2 R_1 \geq \frac{1}{2}\log\left(\frac{1}{D_2}\right) \tag{2.62}$$

in order for (R_1, R_2, D_1, D_2) to be achievable. Thus, a simple sum-rate lower bound in the

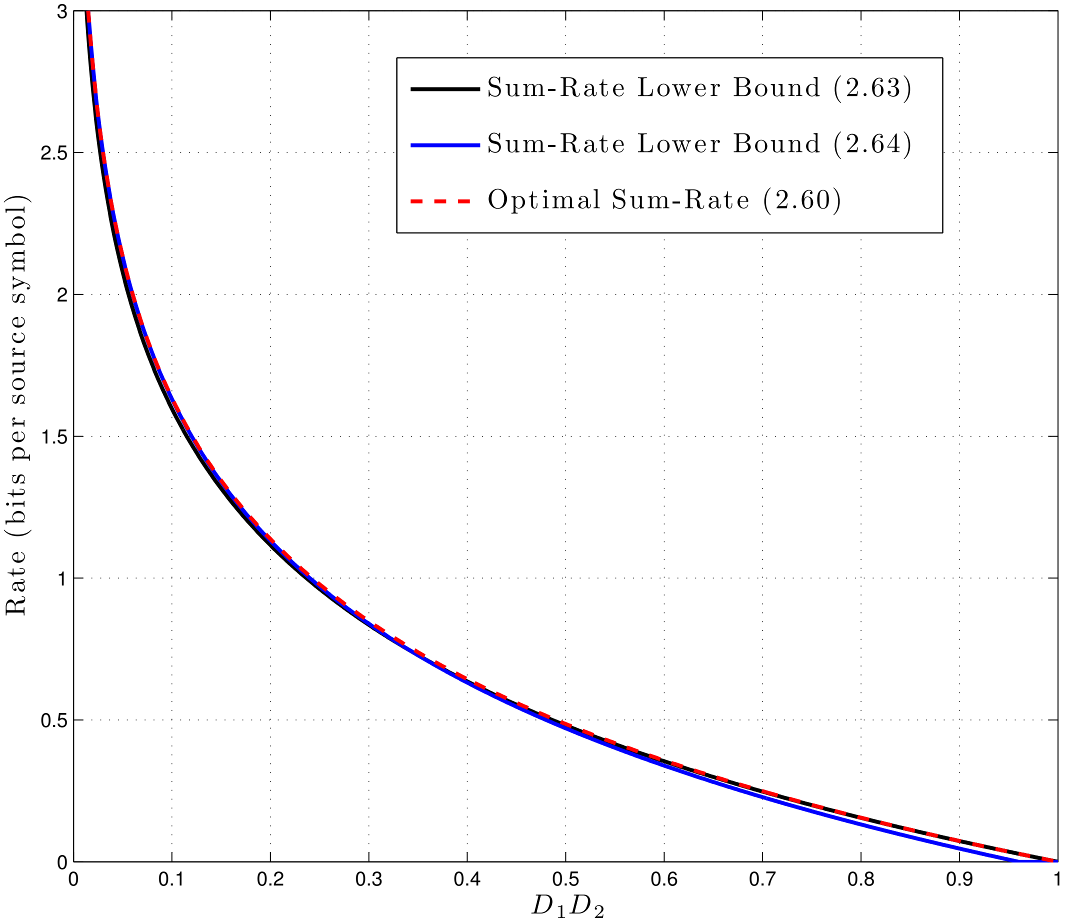

Figure 2.5: Comparison of Eqns. (2.60), (2.63), and (2.64) for $\rho = 1/5$.

quadratic Gaussian setting is given by

$$R_1 + R_2 \geq \frac{1}{(1+\rho^2)} \left(\frac{1}{2} \log \left(\frac{1}{D_1} \right) + \frac{1}{2} \log \left(\frac{1}{D_2} \right) \right) \tag{2.63}$$
$$= \frac{1}{2(1+\rho^2)} \log \left(\frac{1}{D_1 D_2} \right).$$

Furthermore, since the sum-rate is trivially lower bounded by the rate distortion function of one source plus the Wyner-Ziv rate distortion function of the other source (given the former source as side information), we have the necessary condition

$$\begin{aligned} R_1 + R_2 &\geq \frac{1}{2}\log\left(\frac{1}{D_1}\right) + \frac{1}{2}\log\left(\frac{1-\rho^2}{D_2}\right) \\ &= \frac{1}{2}\log\left(\frac{1}{D_2}\right) + \frac{1}{2}\log\left(\frac{1-\rho^2}{D_1}\right) \\ &= \frac{1}{2}\log\left(\frac{1-\rho^2}{D_1 D_2}\right). \end{aligned} \tag{2.64}$$

In Figures 2.5 and 2.6 we have plotted the lower bounds (2.60), (2.63), and optimal sum-rate (2.64) for $\rho = 1/5$ and $\rho = 4/5$, respectively. As evidenced by the plots, taking the maximum of the necessary conditions (2.63) and (2.64) yields a strikingly tight bound on (2.60) (i.e., the optimal sum-rate constraint).

Unfortunately, our derivation of (2.63) was somewhat ad-hoc and required necessity of (2.58) and (2.59), which was established by Oohama in [Ooh97] many years after the multiterminal source coding problem was posed. Thus, it is desirable to establish a generalization of (2.61) and (2.62) to arbitrary sources and distortion measures which does not require known converse results for the specific problem instance under consideration. This leads us to state our final result of the section.

Definition 7. *Define $\rho_m(Y_1, Y_2)$ to be the Hirschfeld-Gebelein-Rényi maximal correlation (cf. [Ren59]) between random variables Y_1 and Y_2.*

Theorem 11. *Suppose (R_1, R_2, D_1, D_2) is achievable for the general multiterminal source coding problem. Then*

$$R_1 + \rho_m^2(Y_1, Y_2) R_2 \geq I(Y_1; \breve{Y}_1, \breve{Y}_2)$$
$$R_2 + \rho_m^2(Y_1, Y_2) R_1 \geq I(Y_2; \breve{Y}_1, \breve{Y}_2)$$

for some conditional pmf $p(\breve{y}_1, \breve{y}_2 | y_1, y_2)$ satisfying

$$\mathbb{E} d_1(Y_1, \breve{Y}_1) \leq D_1$$
$$\mathbb{E} d_2(Y_2, \breve{Y}_2) \leq D_2.$$

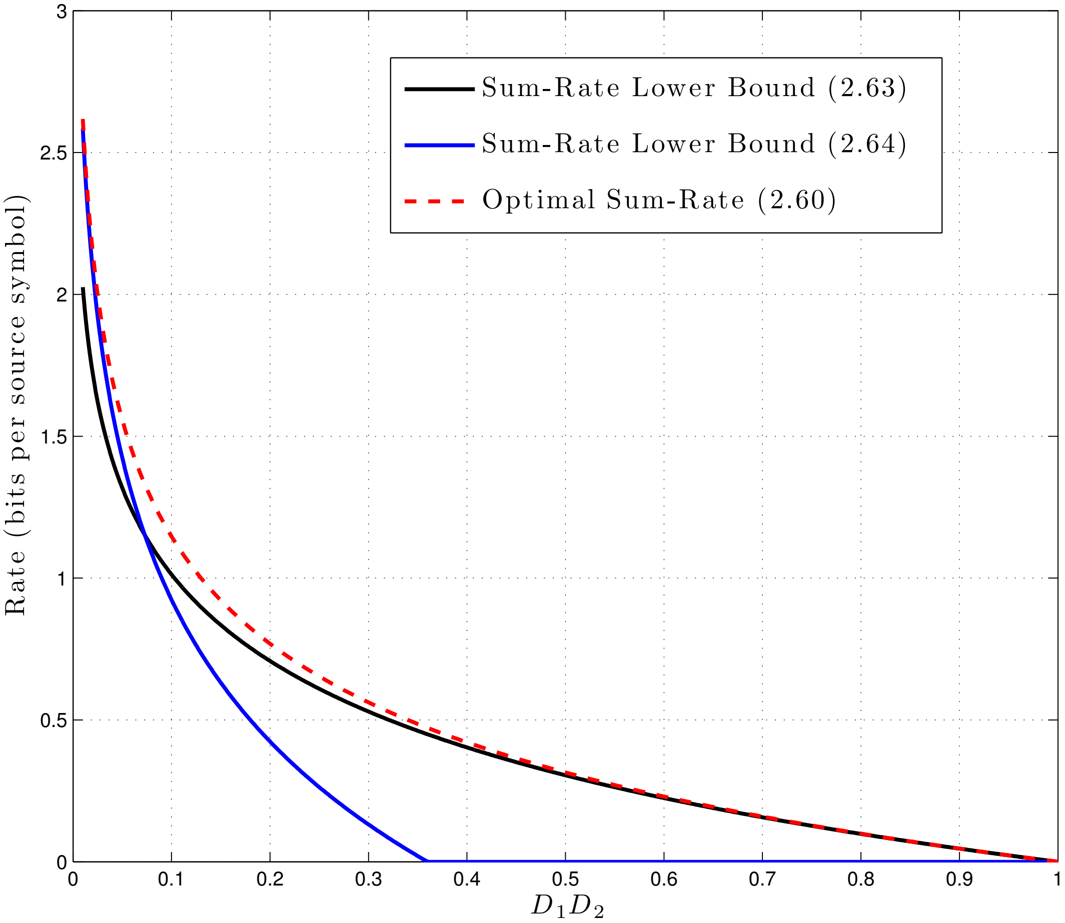

Figure 2.6: Comparison of Eqns. (2.60), (2.63), and (2.64) for $\rho = 4/5$.

To see how Theorem 11 relates to the above discussion, we make the following definition.

Definition 8. *Let $R(D_i; \breve{d}_i)$ denote the rate distortion function for source Y_i under distortion measure $\breve{d}_i$. That is,*

$$\breve{d}_i : \mathcal{Y}_i \times \breve{\mathcal{Y}}_i \to [0, \infty) \quad \textit{for } i = 1, 2,$$

and

$$R(D_i; \breve{d}_i) \triangleq \min_{p(\breve{y}_i|y_i):\mathbb{E}d_i(Y_i,\breve{Y}_i)\leq D_i} I(Y_i; \breve{Y}_i) \quad \textit{for } i = 1, 2.$$

With this definition in hand, we observe that an immediate corollary of Theorem 11 is that any achievable (R_1, R_2, D_1, D_2) necessarily satisfies

$$R_1 + \rho_m^2(Y_1, Y_2)R_2 \geq R(D_1; \breve{d}_1)$$
$$R_2 + \rho_m^2(Y_1, Y_2)R_1 \geq R(D_2; \breve{d}_2).$$

Noting that $\rho_m^2(Y_1, Y_2) = \rho^2$ for the Gaussian setting described above, we see that Theorem 11 generalizes (2.61) and (2.62) without requiring previously known converse results.

Remarkably, Theorem 11 makes no assumptions on the distortion measures or source distribution (other than the standard assumption that the source is i.i.d.). Moreover, the necessary conditions implied by Theorem 11 are easily computable for finite alphabet sources. Indeed, the squared maximal correlation $\rho_m^2(Y_1, Y_2)$ corresponds to an eigenvalue of a linear system defined by the joint pmf $p(y_1, y_2)$ [Ren59], and minimizing $\lambda_1 I(Y_1; \breve{Y}_1, \breve{Y}_2) + \lambda_2 I(Y_2; \breve{Y}_1, \breve{Y}_2)$ subject to the desired distortion constraints is a convex optimization problem for $\lambda_1, \lambda_2 \geq 0$.

Another immediate Corollary to Theorem 11 is the following intuitive lower bound on the sum-rate which extends (2.63) to the general setting.

Corollary 2. *If (R_1, R_2, D_1, D_2) is achievable for the general multiterminal source coding problem, then*

$$R_1 + R_2 \geq \frac{1}{1 + \rho_m^2(Y_1, Y_2)} \left(R(D_1; \breve{d}_1) + R(D_2; \breve{d}_2) \right).$$

Corollary 2 gives a rough estimate of the price paid (with respect to sum rate) by compressing each source separately. Indeed, at most $\frac{\rho_m^2(Y_1,Y_2)}{1+\rho_m^2(Y_1,Y_2)} \left(R(D_1; \breve{d}_1) + R(D_2; \breve{d}_2) \right)$ total bits of rate can be saved by exploiting correlation (if any) between sources. Since $\rho_m(Y_1, Y_2) \in [0, 1]$ by definition, the coefficient $\frac{\rho_m^2(Y_1,Y_2)}{1+\rho_m^2(Y_1,Y_2)}$ can be quite small if the sources under consideration are not strongly correlated. The practical implication is that compressing each source individually can attain near-optimal performance if the sources are not strongly correlated. For instance, if $\rho_m(Y_1, Y_2) = 0.2$, then compressing each source individually requires a sum-rate at most 3.85% in excess of that required by an optimal scheme, regardless of which distortion measures are used. One simple example of a source with $\rho_m(Y_1, Y_2) = 0.2$ is the following: $\mathcal{Y}_1 = \mathcal{Y}_2 = \{1, 2, 3, 4\}$ with joint pmf

$$p(y_1, y_2) = \begin{cases} 1/10 & \text{if } y_1 = y_2 \\ 1/20 & \text{if } y_1 \neq y_2. \end{cases}$$

We note that the outer bound given by Theorem 11 and Corollary 2 should be tight in the high-disortion (i.e., low-rate) regime. Therefore, our results complement others which characterize the rate distortion region in the high-resolution regime (e.g., [ZB99]). This is concretely demonstrated in Figure 2.6.

Proof of Theorem 11. Fix $\epsilon > 0$ and let $(\breve{g}_1^{(n)}, \breve{g}_2^{(n)}, \breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)})$ be a code which attains distortions $D_1 + \epsilon$ and $D_2 + \epsilon$ for sources Y_1 and Y_2, respectively. By considering the log-loss sequence reproductions $\hat{Y}_i^n = \Pr[Y_i^n = y_i^n | \breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)]$ for each source $i = 1, 2$, Theorem 7 immediately implies that there exists a joint distribution $p \triangleq p(y_1, y_2)p(q)p(u_1|y_1, q)p(u_2|y_2, q)$ with $|\mathcal{U}_i| \leq |\mathcal{Y}_i|$ and $|\mathcal{Q}| \leq 5$ which satisfies[5]:

$$\begin{aligned} R_1 &\geq I(Y_1; U_1|U_2, Q) \\ R_2 &\geq I(Y_2; U_2|U_1, Q) \\ R_1 + R_2 &= I(Y_1, Y_2; U_1, U_2|Q) \\ \frac{1}{n} H(Y_1^n | \breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)) &\geq H(Y_1|U_1, U_2, Q) \\ \frac{1}{n} H(Y_2^n | \breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)) &\geq H(Y_2|U_1, U_2, Q). \end{aligned}$$

Since the sources are memoryless, the last two inequalities above are equivalent to

$$\begin{aligned} I(Y_1; U_1, U_2|Q) &\geq \frac{1}{n} I(Y_1^n; \breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)) \qquad (2.65) \\ I(Y_2; U_1, U_2|Q) &\geq \frac{1}{n} I(Y_2^n; \breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)). \end{aligned}$$

[5] Establishing the equality in the sum-rate constraint is straightforward. It can be accomplished along the same lines as Lemma 9.

By definition of p, we have $R_1 + R_2 = I(Y_1, Y_2; U_1, U_2|Q)$, and hence

$$\begin{aligned} R_2 &= I(Y_1, Y_2; U_1, U_2|Q) - R_1 \\ &= I(Y_2; U_2|Q) - (R_1 - I(Y_1; U_1|U_2, Q)) \\ &\leq I(Y_2; U_2|Q). \end{aligned} \tag{2.66}$$

Recall the following result of Erkip [Erk96, Theorem 10]. For all real $x \geq 0$, the following holds:

$$\max_{\substack{p(\tilde{u}_2|y_2): I(Y_2;\tilde{U}_2) \leq x, \\ \tilde{U}_2 \leftrightarrow Y_2 \leftrightarrow Y_1}} I(\tilde{U}_2; Y_1) \leq \rho_m^2(Y_1, Y_2)x. \tag{2.67}$$

Applying this inequality, we obtain

$$\begin{aligned} \rho_m^2(Y_1, Y_2) I(Y_2; U_2|Q) &= \rho_m^2(Y_1, Y_2) I(Y_2; U_2, Q) && (2.68) \\ &\geq \max_{\substack{p(\tilde{u}_2|y_2): I(Y_2;\tilde{U}_2) \leq I(Y_2;U_2,Q), \\ \tilde{U}_2 \leftrightarrow Y_2 \leftrightarrow Y_1}} I(\tilde{U}_2; Y_1) && (2.69) \\ &\geq I(U_2, Q; Y_1) && (2.70) \\ &= I(U_2; Y_1|Q) && (2.71) \end{aligned}$$

where

- (2.68) and (2.71) follow since Q and (Y_1, Y_2) are independent by definition of p.
- (2.69) follows from (2.67).
- (2.70) follows since $\tilde{U}_2 = (U_2, Q)$ lies in the set over which we take the maximum in (2.69).

Summarizing the above string of inequalities, we have

$$I(Y_1; U_2|Q) \leq \rho_m^2(Y_1, Y_2) I(Y_2; U_2|Q). \tag{2.72}$$

Next, observe that

$$\begin{aligned}
R_1 + R_2 &= I(Y_1, Y_2; U_1, U_2|Q) \\
&= I(Y_1; U_1, U_2|Q) + I(Y_2; U_1, U_2|Y_1, Q) \\
&= I(Y_1; U_1, U_2|Q) + I(Y_2; U_2|Q) - I(Y_1; U_2|Q) \\
&\geq I(Y_1; U_1, U_2|Q) + I(Y_2; U_2|Q) - \rho_m^2(Y_1, Y_2) I(Y_2; U_2|Q) \qquad (2.73) \\
&= I(Y_1; U_1, U_2|Q) + (1 - \rho_m^2(Y_1, Y_2)) I(Y_2; U_2|Q) \\
&\geq I(Y_1; U_1, U_2|Q) + (1 - \rho_m^2(Y_1, Y_2)) R_2, \qquad (2.74)
\end{aligned}$$

where (2.73) follows from (2.72), and (2.74) follows from (2.66) and the fact that $\rho_m(Y_1, Y_2) \in [0, 1]$. Rearranging and applying (2.65), we obtain the inequality

$$\begin{aligned}
R_1 + \rho_m^2(Y_1, Y_2) R_2 &\geq I(Y_1; U_1, U_2|Q) \\
&\geq \frac{1}{n} I(Y_1^n; \breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)). \qquad (2.75)
\end{aligned}$$

By a similar argument, we obtain the symmetric inequality:

$$R_2 + \rho_m^2(Y_1, Y_2) R_1 \geq \frac{1}{n} I(Y_2^n; \breve{g}_1^{(n)}(Y_1^n), \breve{g}_2^{(n)}(Y_2^n)). \qquad (2.76)$$

Next, since $\breve{Y}_j^n = \breve{\psi}_j^{(n)}(\breve{g}_1(Y_1^n), \breve{g}_2(Y_2^n))$ for $j = 1, 2$, the data processing inequality implies that

$$\begin{aligned}
\frac{1}{n} I(Y_1^n; \breve{g}_1(Y_1^n), \breve{g}_2(Y_2^n)) &\geq \frac{1}{n} I(Y_1^n; \breve{Y}_1^n, \breve{Y}_2^n) \\
&= \frac{1}{n} \sum_{i=1}^{n} I(Y_{1,i}; \breve{Y}_1^n, \breve{Y}_2^n | Y_1^{i-1}) \\
&= \frac{1}{n} \sum_{i=1}^{n} I(Y_{1,i}; \breve{Y}_1^n, \breve{Y}_2^n, Y_1^{i-1}) \\
&\geq \frac{1}{n} \sum_{i=1}^{n} I(Y_{1,i}; \breve{Y}_{1,i}, \breve{Y}_{2,i}).
\end{aligned}$$

By a similar argument, we have

$$\frac{1}{n} I(Y_2^n; \breve{g}_1(Y_1^n), \breve{g}_2(Y_2^n)) \geq \frac{1}{n} \sum_{i=1}^{n} I(Y_{2,i}; \breve{Y}_{1,i}, \breve{Y}_{2,i}).$$

Also, by definition of $(\breve{g}_1^{(n)}, \breve{g}_2^{(n)}, \breve{\psi}_1^{(n)}, \breve{\psi}_2^{(n)})$, we note that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\breve{d}_1(Y_{1,i}, \breve{Y}_{1,i}) &\leq D_1 + \epsilon\\ \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\breve{d}_2(Y_{2,i}, \breve{Y}_{2,i}) &\leq D_2 + \epsilon.\end{aligned}$$

For each $j = 1, \ldots, n$, define

$$p(\breve{Y}_1 = \breve{y}_1, \breve{Y}_2 = \breve{y}_2 | Y_{1,j} = y_1, Y_{2,j} = y_2) = \frac{1}{n}\sum_{i=1}^{n}\Pr\left[\breve{Y}_{1,i} = \breve{y}_1, \breve{Y}_{2,i} = \breve{y}_2 | Y_{1,i} = y_1, Y_{2,i} = y_2\right].$$

Recalling that $I(X;Y)$ is convex in $p(x|y)$ for fixed $p(y)$ and Y_1^n is i.i.d., we can apply Jensen's inequality to obtain

$$\begin{aligned}\frac{1}{n}I(Y_1^n; \breve{g}_1(Y_1^n), \breve{g}_2(Y_2^n)) &\geq \frac{1}{n}\sum_{i=1}^{n} I(Y_{1,i}; \breve{Y}_{1,i}, \breve{Y}_{2,i})\\ &\geq \frac{1}{n}\sum_{i=1}^{n} I(Y_{1,i}; \breve{Y}_1, \breve{Y}_2)\\ &= I(Y_1; \breve{Y}_1, \breve{Y}_2), \qquad (2.77)\end{aligned}$$

where

$$p(\breve{Y}_1 = \breve{y}_1, \breve{Y}_2 = \breve{y}_2 | Y_1 = y_1, Y_2 = y_2) \triangleq p(\breve{Y}_1 = \breve{y}_1, \breve{Y}_2 = \breve{y}_2 | Y_{1,j} = y_1, Y_{2,j} = y_2).$$

Similarly,

$$\frac{1}{n}I(Y_2^n; \breve{g}_1(Y_1^n), \breve{g}_2(Y_2^n)) \geq I(Y_2; \breve{Y}_1, \breve{Y}_2). \qquad (2.78)$$

Next, we apply linearity of expectation to obtain

$$\mathbb{E}\breve{d}_j(Y_j, \breve{Y}_j) \leq D_j + \epsilon \text{ for } j = 1, 2. \qquad (2.79)$$

Finally, we combine (2.75), (2.76), (2.77), (2.78), and (2.79) to obtain the desired result:

$$\begin{aligned}R_1 + \rho_m^2(Y_1, Y_2)R_2 &\geq I(Y_1; \breve{Y}_1, \breve{Y}_2)\\ R_2 + \rho_m^2(Y_1, Y_2)R_1 &\geq I(Y_2; \breve{Y}_1, \breve{Y}_2)\end{aligned}$$

for some conditional pmf $p(\breve{y}_1, \breve{y}_2|y_1, y_2)$ satisfying

$$\mathbb{E}d_1(Y_1, \breve{Y}_1) \leq D_1 + \epsilon$$
$$\mathbb{E}d_2(Y_2, \breve{Y}_2) \leq D_2 + \epsilon.$$

Since ϵ can be taken arbitrarily small, the theorem is proved. □

2.6 Information Amplification and Masking

In Section 2.4, we considered the goal of minimizing the attainable distortion for each source subject to rate constraints at the encoders. However, we can also consider a related setting where the objective is to minimize one distortion, say D_1, while simultaneously maximizing the other distortion D_2. In general, there is a tension between these two objectives which is characterized by the *amplification-masking tradeoff.* In this section, we give a single-letter description of this tradeoff. As an application, we apply this result, together with Theorem 7, to solve a fundamental entropy characterization problem.

The well known source coding with side information problem has an achievable rate region given by

$$R_x \geq H(X|U), \quad R_y \geq I(Y;U)$$

as originally shown by Ahlswede and Körner [AK75], and independently by Wyner [Wyn75]. In this setting, the side information encoder merely serves as a helper with the sole purpose of aiding in the recovery of X^n at the decoder. However, for given rates (R_x, R_y), there may be many different coding schemes which permit recovery of X^n at the decoder. In some cases, it may be desirable to select a coding scheme that reveals very little information about the side information Y^n to the decoder. We refer to this objective as *masking* the side information.

To motivate this goal, consider the following example. Suppose X is an attribute of an online customer that an advertiser would like to specifically target (e.g., gender), and Y is other detailed information about the same customer (e.g., credit history). Companies A and B separately have databases X^n and Y^n corresponding to n different customers (the databases could be indexed by IP address, for example). The advertiser pays Companies A and B to learn as much about the database X^n as possible. Now, suppose governing laws prohibit the database Y^n from being revealed too extensively. In this case, the material given to the advertiser must be chosen so that at most a prescribed amount of information is revealed about Y^n.

In general, a masking constraint on Y^n may render near-lossless reconstruction of X^n impossible. This motivates the study the *amplification-masking tradeoff*. That is, the tradeoff between amplifying (or revealing) information about X^n while simultaneously masking the side information Y^n.

Similar problems have been previously considered in the information theory literature on secrecy and privacy. For example, Sankar et al. determine the utility-privacy tradeoff for the case of a single encoder in [SRP11]. In their setting, the random variable X is a vector with a given set of coordinates that should be masked and another set that should be revealed (up to a prescribed distortion). In this context, our study of the amplification-masking tradeoff is a distributed version of [SRP11], in which utility is measured by the information revealed about the database X^n. The problem we consider here is distinct from those typically studied in the information-theoretic secrecy literature, in that the masking (i.e., equivocation) constraint corresponds to the intended decoder, rather than an eavesdropper.

We remark that the results presented in this section are inspired in part by the recent, complementary works [KSC08] and [MS07] which respectively study amplification and masking of channel state information in the Gel'fand-Pinsker channel. We borrow our terminology from those works. On a related note, the tension between amplification and masking of channel state information in the presence of an eavesdropper was studied recently in [KSV11].

This section of the dissertation is largely self-contained and is organized as follows. Section 2.6.1 formally defines the problems considered and delivers the main results of the section. The corresponding proofs are given in Section 2.6.2.

2.6.1 Problem Statement and Results

Throughout this section we adopt notational conventions that are standard in the literature. Specifically, random variables are denoted by capital letters (e.g., X) and their corresponding alphabets are denoted by corresponding calligraphic letters (e.g., $\mathcal{X}$). We abbreviate a sequence $(X_1, \ldots, X_n)$ of n random variables by X^n, and we let $\delta(\epsilon)$ represent a quantity satisfying $\lim_{\epsilon \to 0} \delta(\epsilon) = 0$. Other notation will be introduced where necessary.

For a joint distribution $p(x, y)$ on finite alphabets $\mathcal{X} \times \mathcal{Y}$, consider the source coding setting where separate Encoders 1 and 2 have access to the sequences X^n and Y^n, respectively. We make the standard assumption that the sequences (X^n, Y^n) are drawn i.i.d. according to $p(x, y)$ (i.e., $X^n, Y^n \sim \prod_{i=1}^n p(x_i, y_i)$), and n can be taken arbitrarily large.

The first of the following three subsections characterizes the amplification-masking tradeoff. This result is applied to solve a fundamental entropy characterization in the second subsection. The final subsection comments on the connection between information amplification and list decoding. Proofs of the main results are postponed until Section 2.6.2.

2.6.1.1 The Amplification-Masking Tradeoff

Formally, a $(2^{nR_x}, 2^{nR_y}, n)$ code is defined by its encoding functions

$$f_x : \mathcal{X}^n \to \{1, \ldots, 2^{nR_x}\} \text{ and } f_y : \mathcal{Y}^n \to \{1, \ldots, 2^{nR_y}\}.$$

A rate-amplification-masking tuple $(R_x, R_y, \Delta_A, \Delta_M)$ is achievable if, for any $\epsilon > 0$, there exists a $(2^{nR_x}, 2^{nR_y}, n)$ code satisfying the amplification criterion:

$$\Delta_A \leq \frac{1}{n} I\left(X^n; f_x(X^n), f_y(Y^n)\right) + \epsilon, \tag{2.80}$$

and the masking criterion:

$$\Delta_M \geq \frac{1}{n} I\left(Y^n; f_x(X^n), f_y(Y^n)\right) - \epsilon. \tag{2.81}$$

Thus, we see that the amplification-masking problem is an entropy characterization problem similar to that considered in [CK81, Chapter 15].

Definition 9. *The achievable amplification-masking region $\mathcal{R}_{AM}$ is the closure of the set of all achievable rate-amplification-masking tuples $(R_x, R_y, \Delta_A, \Delta_M)$.*

Theorem 12. *$\mathcal{R}_{AM}$ consists of the rate-amplification-masking tuples $(R_x, R_y, \Delta_A, \Delta_M)$ satisfying*

$$\left.\begin{aligned} R_x &\geq \Delta_A - I(X;U) \\ R_y &\geq I(Y;U) \\ \Delta_M &\geq \max\left\{I(Y;U,X) + \Delta_A - H(X), I(Y;U)\right\} \\ \Delta_A &\leq H(X). \end{aligned}\right\} \tag{2.82}$$

for some joint distribution $p(x,y,u) = p(x,y)p(u|y)$, where $|\mathcal{U}| \leq |\mathcal{Y}| + 1$.

Observe that $\mathcal{R}_{AM}$ characterizes the entire tradeoff between amplifying X^n and masking Y^n. We remark that maximum amplification $\Delta_A = H(X)$ does not necessarily imply that X^n can be recovered near-losslessly at the encoder. However, if an application demands near lossless reproduction of the sequence X^n, Theorem 12 can be strengthened to include this case. To this end, define a rate-masking triple (R_x, R_y, Δ_M) to be achievable if, for any $\epsilon > 0$, there exists a $(2^{nR_x}, 2^{nR_y}, n)$ code satisfying the masking criterion (2.81), and a decoding function

$$\hat{X}^n : \{1, 2, \ldots, 2^{nR_x}\} \times \{1, 2, \ldots, 2^{nR_y}\} \to \mathcal{X}^n$$

which satisfies the decoding-error criterion

$$\Pr\left[X^n \neq \hat{X}^n(f_x(X^n), f_y(Y^n))\right] \leq \epsilon.$$

Definition 10. *The achievable rate-masking region $\mathcal{R}_M$ is the closure of the set of all achievable rate-masking triples (R_x, R_y, Δ_M).*

Corollary 3. *$\mathcal{R}_M$ consists of the rate-masking triples (R_x, R_y, Δ_M) satisfying*

$$R_x \geq H(X|U)$$
$$R_y \geq I(Y;U)$$
$$\Delta_M \geq I(Y;X,U)$$

for some joint distribution $p(x,y,u) = p(x,y)p(u|y)$, where $|\mathcal{U}| \leq |\mathcal{Y}| + 1$.

2.6.1.2 An Entropy Characterization Result

As we previously noted, the amplification-masking tradeoff solves a multi-letter entropy characterization problem by reducing it to single-letter form. The reader is directed to [CK81] for an introduction to entropy characterization problems. Here, we apply our results to yield a fundamental characterization of the information revealed about X^n and Y^n, respectively, by arbitrary encoding functions f_x and f_y (of rates R_x, R_y).

Definition 11. *Define the region $\mathcal{R}^\star(R_x, R_y)$ as follows. The pair $(\Delta_X, \Delta_Y) \in \mathcal{R}^\star(R_x, R_y)$ if and only if, for any $\epsilon > 0$, there exists a $(2^{nR_x}, 2^{nRy}, n)$ code satisfying*

$$\left|\Delta_X - \frac{1}{n} I(X^n; f_x(X^n), f_y(Y^n))\right| \leq \epsilon, \text{ and}$$
$$\left|\Delta_Y - \frac{1}{n} I(Y^n; f_x(X^n), f_y(Y^n))\right| \leq \epsilon.$$

Let $\overline{\mathcal{R}^\star}(R_x, R_y)$ be the closure of $\mathcal{R}^\star(R_x, R_y)$.

Ultimately we obtain a single-letter description of $\overline{\mathcal{R}^\star}(R_x, R_y)$. However, in order to do so, we require some notation. To this end, let:

$$\mathcal{R}_{AM}(R_x, R_y) = \left\{(\Delta_X, \Delta_Y) : (R_x, R_y, \Delta_X, \Delta_Y) \in \mathcal{R}_{AM}\right\}.$$

Symmetrically, let $\mathcal{R}_{MA}$ be the region where X^n is subject to masking Δ_X and Y^n is subject to amplification Δ_Y. Let

$$\mathcal{R}_{MA}(R_x, R_y) = \left\{(\Delta_X, \Delta_Y) : (R_x, R_y, \Delta_X, \Delta_Y) \in \mathcal{R}_{MA}\right\}.$$

Finally, let $\mathcal{R}_{AA}(R_x, R_y)$ consist of all pairs (Δ_X, Δ_Y) satisfying

$$\begin{aligned}
R_x &\geq I(U_x; X|U_y, Q) \\
R_y &\geq I(U_y; Y|U_x, Q) \\
R_x + R_y &\geq I(U_x, U_y; X, Y|Q) \\
\Delta_X &\leq I(X; U_x, U_y|Q) \\
\Delta_Y &\leq I(Y; U_x, U_y|Q)
\end{aligned}$$

for some joint distribution of the form

$$p(x, y, u_x, u_y, q) = p(x, y)p(u_x|x, q)p(u_y|y, q)p(q),$$

where $|\mathcal{U}_x| \leq |\mathcal{X}|$, $|\mathcal{U}_y| \leq |\mathcal{Y}|$, and $|\mathcal{Q}| \leq 5$.

Theorem 13. *The region $\overline{\mathcal{R}^\star}(R_x, R_y)$ has a single-letter characterization given by*

$$\begin{aligned}
&\overline{\mathcal{R}^\star}(R_x, R_y) = \\
&\qquad \mathcal{R}_{AM}(R_x, R_y) \cap \mathcal{R}_{MA}(R_x, R_y) \cap \mathcal{R}_{AA}(R_x, R_y).
\end{aligned}$$

Moreover, restriction of the encoding functions to vector-quantization and/or random binning is sufficient to achieve any point in $\overline{\mathcal{R}^\star}(R_x, R_y)$.

The second statement of Theorem 13 is notable since it states that relatively simple encoding functions (i.e., vector quantization and/or binning) can asymptotically reveal the same amount of information about X^n and Y^n, respectively, as encoding functions that are only restricted in rate. In contrast, this is not true for the setting of three or more sources, as the modulo-sum problem studied by Körner and Marton [KM79] provides a counterexample where the Berger-Tung achievability scheme [Ber77] is not optimal. Thus, obtaining a characterization like Theorem 13 for three or more sources represents a formidable challenge.

We remark that the points in $\overline{\mathcal{R}^\star}(R_x, R_y)$ with $\Delta_X = H(X)$ and/or $\Delta_Y = H(Y)$ also capture the more stringent constraint(s) of near-lossless reproduction of X^n and/or Y^n, respectively. This is a consequence of Corollary 3.

To give a concrete example of $\overline{\mathcal{R}^\star}(R_x, R_y)$, consider the following joint distribution:

$P_{X,Y}(x,y)$	$x = 0$	$x = 1$
$y = 0$	$1/3$	0
$y = 1$	$1/6$	$1/2.$

(2.83)

By performing a brute-force search over the auxiliary random variables defining $\overline{\mathcal{R}^\star}(R_x, R_y)$ for the distribution $P_{X,Y}$, we have obtained numerical approximations of $\overline{\mathcal{R}^\star}(\cdot, \cdot)$ for several different pairs of (R_x, R_y). The results are given in Figure 2.7.

2.6.1.3 Connection to List Decoding

We briefly comment on the connection between an amplification constraint and list decoding. As discussed in detail in [KSC08], the amplification criterion (2.80) is essentially equivalent to the requirement for a list decoder

$$L_n : \{1, \dots, 2^{nR_x}\} \times \{1, \dots, 2^{nR_y}\} \to 2^{\mathcal{X}^n}$$

with list size and probability of error respectively satisfying

$$\begin{aligned} \log |L_n| &\le n(H(X) - \Delta_A + \epsilon), \text{and} \\ \Pr\left[X^n \notin L_n(f_x(X^n), f_y(Y^n))\right] &\le \epsilon. \end{aligned}$$

Thus maximizing the amplification of X^n subject to given rate and masking constraints can be thought of as characterizing the best list decoder in that setting.

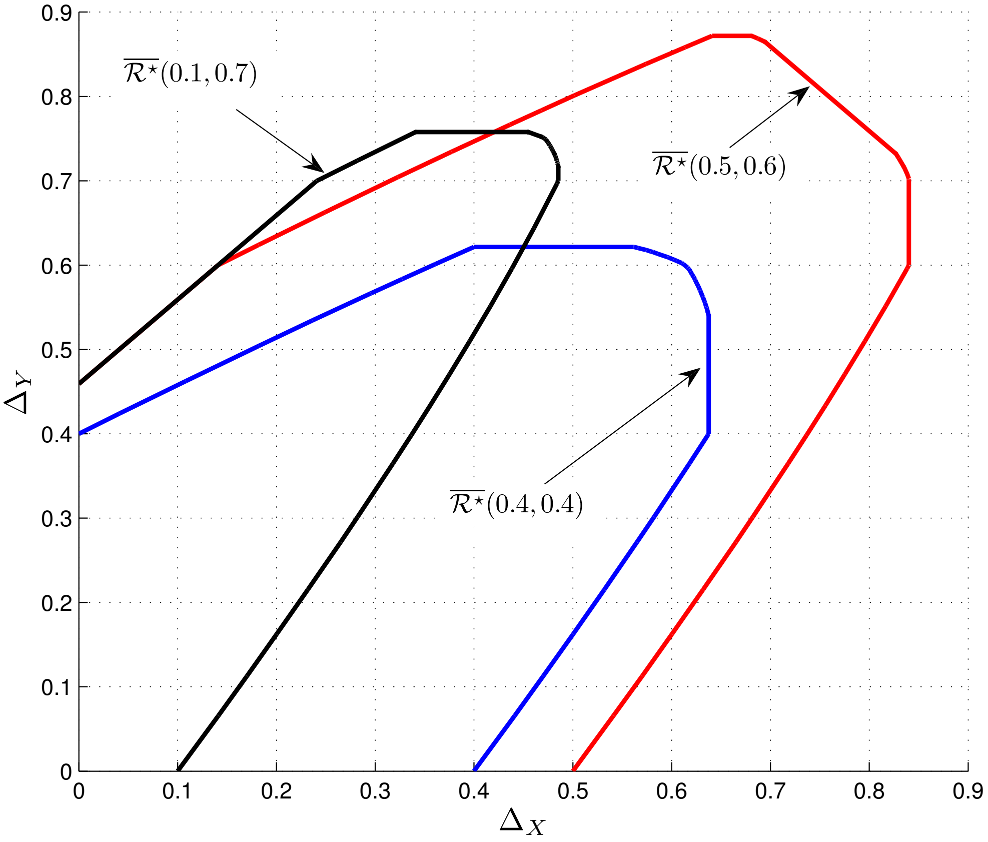

Figure 2.7: The region $\overline{\mathcal{R}^\star}(R_x, R_y)$ for joint distribution $P_{X,Y}$ given by (2.83) and three different pairs of rates. Rate pairs (R_x, R_y) equal to $(0.1, 0.7)$, $(0.4, 0.4)$, and $(0.5, 0.6)$ define the convex regions bounded by the black, blue, and red curves, respectively.

2.6.2 Proofs of the Amplification-Masking Results

Proof of Theorem 12. Converse Part: Suppose $(R_x, R_y, \Delta_A, \Delta_M)$ is achievable. For convenience, define $F_x = f_x(X^n)$, $F_y = f_y(Y^n)$, and $U_i = (F_y, Y^{i-1})$.

First, note that $\Delta_A \leq H(X)$ is trivially satisfied. Next, the constraint on R_x is given by:

$$\begin{aligned}
nR_x &\geq H(F_x) \geq H(F_x|F_y) \\
&= \sum_{i=1}^{n} H(X_i|F_y, X^{i-1}) - H(X^n|F_x, F_y) \\
&\geq \sum_{i=1}^{n} H(X_i|F_y, Y^{i-1}, X^{i-1}) - H(X^n|F_x, F_y) \\
&= I(X^n; F_x, F_y) - \sum_{i=1}^{n} I(X_i; U_i) && (2.84) \\
&\geq n(\Delta_A - \epsilon) - \sum_{i=1}^{n} I(X_i; U_i). && (2.85)
\end{aligned}$$

Equality (2.84) follows since $X_i \leftrightarrow F_y, Y^{i-1} \leftrightarrow X^{i-1}$ form a Markov chain, and inequality (2.85) follows since amplification Δ_A is achievable.

The constraint on R_y is trivial:

$$\begin{aligned}
nR_y \geq H(F_y) \geq I(F_y; Y^n) &= \sum_{i=1}^{n} I(Y_i; F_y|Y^{i-1}) \\
= \sum_{i=1}^{n} I(Y_i; F_y, Y^{i-1}) &= \sum_{i=1}^{n} I(Y_i; U_i).
\end{aligned}$$

Similarly, we obtain the first lower bound on Δ_M:

$$n(\Delta_M + \epsilon) \geq I(Y^n; F_x, F_y) \geq I(Y^n; F_y) = \sum_{i=1}^{n} I(Y_i; U_i).$$

The second lower bound on Δ_M requires slightly more work, and can be derived as follows:

$$\begin{aligned}
n(\Delta_M+\epsilon) &\geq I(Y^n;F_x,F_y) \\
&= I(Y^n;X^n,F_y) + I(X^n;F_x,F_y) - I(X^n;F_x,Y^n) \\
&\geq I(Y^n;X^n,F_y) + n\Delta_A - I(X^n;F_x,Y^n) - n\epsilon \qquad (2.86)\\
&\geq \sum_{i=1}^{n} I(Y_i;X^n,F_y|Y^{i-1}) + n\Delta_A - H(X^n) - n\epsilon \\
&\geq \sum_{i=1}^{n} I(Y_i;X_i,U_i) + \Delta_A - H(X_i) - \epsilon,
\end{aligned}$$

where (2.86) follows since amplification Δ_A is achievable.

Observing that the Markov condition $U_i \leftrightarrow Y_i \leftrightarrow X_i$ is satisfied for each i, a standard timesharing argument proves the existence of a random variable U such that $U \leftrightarrow Y \leftrightarrow X$ forms a Markov chain and (2.82) is satisfied.

Direct Part: Fix $p(u|y)$ and suppose $(R_x, R_y, \Delta_A, \Delta_M)$ satisfy (2.82) with strict inequality. Next, fix $\epsilon > 0$ sufficiently small so that it is less than the minimum slack in said inequalities, and set $\tilde{R} = I(Y;U) + \epsilon$. Our achievability scheme uses a standard random coding argument which we sketch below.

Codebook generation. Randomly and independently, bin the typical x^n's uniformly into $2^{n(\Delta_A - I(X;U)+\epsilon)}$ bins. Let $b(x^n)$ be the index of the bin which contains x^n. For $l \in \{1,\dots,2^{n\tilde{R}}\}$, randomly and independently generate $u^n(l)$, each according to $\prod_{i=1}^n p_U(u_i)$.

Encoding. Encoder 1, upon observing the sequence X^n, sends the corresponding bin index $b(X^n)$ to the decoder. If X^n is not typical, an error is declared. Encoder 2, upon observing the sequence Y^n, finds an $L \in \{1,\dots,2^{n\tilde{R}}\}$ such that $(Y^n, U^n(L))$ are jointly ϵ-typical, and sends the unique index L to the decoder. If more than one such L exists, ties are broken arbitrarily. If no such L exists, then an error is declared.

This coding scheme clearly satisfies the given rates. Further, each encoder errs with arbitrarily small probability as $n \to \infty$. Hence, we only need to check that the amplification and masking constraints are satisfied. To this end, let $\mathcal{C}$ be the random codebook. We first check that the amplification and masking constraints are separately satisfied when averaged over random codebooks $\mathcal{C}$.

To see that the (averaged) amplification constraint is satisfied, consider the following:

$$\begin{aligned} I(X^n; F_x, F_y|\mathcal{C}) &= H(X^n|\mathcal{C}) - H(X^n|b(X^n), L, \mathcal{C}) \\ &\geq nH(X) - n(H(X) - \Delta_A + \delta(\epsilon)) \qquad (2.87) \\ &= n(\Delta_A - \delta(\epsilon)), \end{aligned}$$

where (2.87) follows since X^n is independent of $\mathcal{C}$ and, averaged over codebooks, there are at most $2^{n(H(X)-\Delta_A+\delta(\epsilon))}$ sequences x^n in bin $b(X^n)$ which are typical with $U^n(L)$, where $L \in \{1, \dots, 2^{n\tilde{R}}\}$.

To see this is true, we appeal to the proof strategy of [EK12, Lemma 22.3] and make adjustments where necessary. For convenience, define $\tilde{R}_x = \Delta_A - I(X;U) + \epsilon$ and recall that ϵ was chosen sufficiently small so that $\tilde{R}_x < H(X|U)$. Note that we can express the random codebook $\mathcal{C}$ as a pair of random codebooks $\mathcal{C} = (\mathcal{C}_B, \mathcal{C}_{VQ})$, where $\mathcal{C}_B$ is the "binning codebook" at Encoder 1, and $\mathcal{C}_{VQ}$ is the "vector-quantization codebook" at Encoder 2.

Let $E_1 = 1$ if $(X^n, U^n(L)) \notin \mathcal{T}_\epsilon^{(n)}$ and $E_1 = 0$ otherwise, where $\mathcal{T}_\epsilon^{(n)}$ denotes the set of ϵ-jointly typical (x^n, u^n) sequence pairs. Note that $\Pr(\{E_1 = 1\})$ tends to 0 as $n \to \infty$. Consider

$$\begin{aligned} &H(X^n|L, b(X^n), \mathcal{C}) \\ &\quad \leq H(X^n, E_1|L, b(X^n), \mathcal{C}) \\ &\quad \leq 1 + n\Pr(\{E_1 = 1\})H(X) \\ &\qquad + \sum_{(l,b,c_{VQ})} p(l, b, c_{VQ}|E_1 = 0) \times H(X^n|L = l, b(X^n) = b, E_1 = 0, \mathcal{C}_{VQ} = c_{VQ}, \mathcal{C}_B). \end{aligned}$$

Now, let $N(l, b, c_{VQ}, \mathcal{C}_B)$ be the number of sequences $x^n \in \mathcal{B}(b) \cap \mathcal{T}_\epsilon^{(n)}(X|u^n(l))$, where $\mathcal{B}(b)$ denotes the bin of x-sequences which is labeled by index b and $u^n(l)$ is the codeword in the (fixed) codebook c_{VQ} with index l. Note that $N(l, b, c_{VQ}, \mathcal{C}_B)$ is a binomial random variable, where the source of randomness comes from the random codebook $\mathcal{C}_B$. Define

$$E_2(l,b,c_{VQ},\mathcal{C}_B)=\begin{cases}1 & \text{if } N(l,b,c_{VQ},\mathcal{C}_B)\geq 2\mathbb{E}\left[N(l,b,c_{VQ},\mathcal{C}_B)\right],\\ 0 & \text{otherwise.}\end{cases}$$

Due to the binomial distribution of $N(l,b,c_{VQ},\mathcal{C}_B)$, it is readily verified that

$$\begin{aligned}\mathbb{E}\left[N(l,b,c_{VQ},\mathcal{C}_B)\right]&=2^{-n\tilde{R}_x}\left|\mathcal{T}_\epsilon^{(n)}(X|u^n(l))\right|,\\ \mathrm{Var}(N(l,b,c_{VQ},\mathcal{C}_B))&\leq 2^{-n\tilde{R}_x}\left|\mathcal{T}_\epsilon^{(n)}(X|u^n(l))\right|.\end{aligned}$$

Then, by the Chebyshev lemma [EK12, Appendix B],

$$\Pr(\{E_2(l,b,c_{VQ},\mathcal{C}_B)=1\})\leq\frac{\mathrm{Var}(N(l,b,c_{VQ},\mathcal{C}_B))}{\left(\mathbb{E}\left[N(l,b,c_{VQ},\mathcal{C}_B)\right]\right)^2}\leq 2^{-n(H(X|U)-\tilde{R}_x-\delta(\epsilon))},$$

which tends to zero as $n\to\infty$ if $\tilde{R}_x<H(X|U)-\delta(\epsilon)$, which is satisfied for ϵ sufficiently small. Now consider

$$\begin{aligned}&H(X^n|L=l,b(X^n)=b,E_1=0,\mathcal{C}_{VQ}=c_{VQ},\mathcal{C}_B)\\ &\quad\leq H(X^n,E_2|L=l,b(X^n)=b,E_1=0,\mathcal{C}_{VQ}=c_{VQ},\mathcal{C}_B)\\ &\quad\leq 1+n\Pr(\{E_2=1\})H(X)+H(X^n|L=l,b(X^n)=b,E_1=0,E_2=0,\mathcal{C}_{VQ}=c_{VQ},\mathcal{C}_B)\\ &\quad\leq 1+n\Pr(\{E_2=1\})H(X)+n(H(X|U)-\tilde{R}_x+\delta(\epsilon)),\end{aligned}$$

which implies that

$$\begin{aligned}&H(X^n|L,b(X^n),\mathcal{C})\\ &\quad\leq 2+n(\Pr(\{E_1=1\})+\Pr(\{E_2=1\}))H(X)+n(H(X|U)-\tilde{R}_x+\delta(\epsilon))\\ &\quad\leq 2+n(\Pr(\{E_1=1\})+\Pr(\{E_2=1\}))H(X)+n(H(X)-\Delta_A+\delta(\epsilon)).\end{aligned}$$

Taking $n\to\infty$ completes the proof of (2.87).

We now turn our attention to the masking criterion. First note the following inequality:

$$\begin{aligned}
I(Y^n; F_x, F_y|\mathcal{C}) &= I(Y^n; L|\mathcal{C}) + I(Y^n; b(X^n)|L, \mathcal{C}) \\
&\le I(Y^n; L|\mathcal{C}) + H(b(X^n)|\mathcal{C}) - H(b(X^n)|Y^n, \mathcal{C}) \\
&= I(Y^n; L|\mathcal{C}) + I(X^n; Y^n) - H(X^n) + H(b(X^n)|\mathcal{C}) - H(b(X^n)|Y^n, \mathcal{C}) + H(X^n|Y^n) \\
&\le I(Y^n; L|\mathcal{C}) + I(X^n; Y^n) - H(X^n) + H(b(X^n)|\mathcal{C}) - I(b(X^n); X^n|Y^n, \mathcal{C}) + H(X^n|Y^n) \\
&= I(Y^n; L|\mathcal{C}) + I(X^n; Y^n) - H(X^n) + H(b(X^n)|\mathcal{C}) + H(X^n|Y^n, b(X^n), \mathcal{C}) \qquad (2.88)
\end{aligned}$$

Two of the terms in (2.88) can be bounded as follows: First, since $L \in \{1, \dots, 2^{n\tilde{R}}\}$, we have

$$I(Y^n; L|\mathcal{C}) \le n\tilde{R} = n(I(Y; U) + \epsilon).$$

Second, there are $2^{n(\Delta_A - I(X;U)+\epsilon)}$ bins at Encoder 1 by construction, and hence $H(b(X^n)|\mathcal{C}) \le n(\Delta_A - I(X; U) + \epsilon)$. Therefore, substituting into (2.88) and simplifying, we have:

$$I(Y^n; F_x, F_y|\mathcal{C}) \le n(I(Y; U, X) + \Delta_A - H(X)) + H(X^n|Y^n, b(X^n), \mathcal{C}) + n2\epsilon. \qquad (2.89)$$

We now consider three separate cases. First, assume $\Delta_A \le I(U; X)$. Then,

$$I(Y; X, U) + \Delta_A - H(X) \le I(Y; X, U) - H(X|U) = I(Y; U) - H(X|Y),$$

and (2.89) becomes

$$I(Y^n; F_x, F_y|\mathcal{C}) \le nI(Y; U) - I(X^n; b(X^n)|Y^n, \mathcal{C}) + n2\epsilon \le nI(Y; U) + n2\epsilon.$$

Next, suppose that $\Delta_A \ge I(X; U) + H(X|Y)$. In this case, there are greater than $2^{n(H(X|Y)+\epsilon)}$ bins in which the X^n sequences are distributed. Hence, knowing Y^n and $b(X^n)$ is sufficient to determine X^n with high probability (i.e., we have a Slepian-Wolf binning at Encoder 1). Therefore, $H(X^n|Y^n, b(X^n), \mathcal{C}) \le n\epsilon$, and (2.89) becomes

$$I(Y^n; F_x, F_y|\mathcal{C}) \le n(I(Y; X, U) + \Delta_A - H(X)) + n3\epsilon.$$

Finally, suppose $\Delta_A = I(X;U) + \theta H(X|Y)$ for some $\theta \in [0,1]$. In this case, we can timeshare between a code $\mathcal{C}_1$ designed for amplification $\Delta'_A = I(X;U)$ with probability θ, and a code $\mathcal{C}_2$ designed for amplification $\Delta''_A = I(X;U) + H(X|Y)$ with probability $1-\theta$ to obtain a code $\mathcal{C}$ with the same average rates and averaged amplification

$$\begin{aligned} I(X^n; F_x, F_y|\mathcal{C}) &= \theta I(X^n; F_x, F_y|\mathcal{C}_1) + (1-\theta) I(X^n; F_x, F_y|\mathcal{C}_2) \\ &\geq n(I(X;U) + \theta H(X|Y) - \delta(\epsilon)) = n(\Delta_A - \delta(\epsilon)). \end{aligned}$$

Then, applying the inequalities obtained in the previous two cases, we obtain:

$$\begin{aligned} I(Y^n; F_x, F_y|\mathcal{C}) &= \theta I(Y^n; F_x, F_y|\mathcal{C}_1) + (1-\theta) I(Y^n; F_x, F_y|\mathcal{C}_2) \\ &\leq \theta n I(Y;U) + (1-\theta) n (I(Y;X,U) + \Delta''_A - H(X)) + 3n\epsilon \\ &= nI(Y;U) + 3n\epsilon. \end{aligned}$$

Combining these three cases proves that

$$\begin{aligned} \frac{1}{n} I(Y^n; F_x, F_y|\mathcal{C}) &\leq \max\{I(Y;U,X) + \Delta_A - H(X), I(Y;U)\} + 3\epsilon \\ &\leq \Delta_M + 3\epsilon. \end{aligned}$$

To show that there exists a code which satisfies the amplification and masking constraints simultaneously, we construct a super-code $\bar{\mathcal{C}}$ of blocklength Nn by concatenating N randomly, independently chosen codes of length n (each constructed as described above). By the weak law of large numbers and independence of the concatenated coded blocks,

$$\begin{aligned} \Pr\left(\left\{\bar{c} : \frac{1}{Nn} I(X^{Nn}; \bar{F}_x, \bar{F}_y|\bar{\mathcal{C}} = \bar{c}) > \Delta_A - \delta(\epsilon)\right\}\right) \geq 3/4 \\ \Pr\left(\left\{\bar{c} : \frac{1}{Nn} I(Y^{Nn}; \bar{F}_x, \bar{F}_y|\bar{\mathcal{C}} = \bar{c}) < \Delta_M + \delta(\epsilon)\right\}\right) \geq 3/4 \end{aligned}$$

for N and n sufficiently large. Thus, there must exist one super-code which simultaneously satisfies both desired constraints. This completes the proof that $(R_x, R_y, \Delta_A, \Delta_M)$ is achievable. Finally, we invoke the Support Lemma [CK81] to see that $|\mathcal{Y}| - 1$ letters are sufficient to preserve $p(y)$. Plus, we require two more letters to preserve the values of $H(X|U)$ and $I(Y;U|X)$. □

Proof of Corollary 3. By setting $\Delta_A = H(X)$, [AK75, Theorem 2] implies that X^n can be reproduced near losslessly. A simplified version of the argument in the direct part of the proof of Theorem 12 shows that the masking criterion will be satisfied for the standard coding scheme. The converse of Theorem 12 continues to apply □

Proof of Theorem 13. First, we remark that Theorem 7 states that $\mathcal{R}_{AA}(R_x, R_y)$ is the closure of pairs (Δ_X, Δ_Y) such that there exists a $(2^{nR_x}, 2^{nR_y}, n)$ code satisfying

$$\begin{aligned}\Delta_X &\leq \frac{1}{n} I(X^n; f_x(X^n), f_y(Y^n)) + \epsilon, \\ \Delta_Y &\leq \frac{1}{n} I(Y^n; f_x(X^n), f_y(Y^n)) + \epsilon\end{aligned}$$

for any $\epsilon > 0$.

Suppose $(\Delta_X, \Delta_Y) \in \mathcal{R}^\star(R_x, R_y)$. By definition of $\mathcal{R}^\star(R_x, R_y)$, Theorem 12, and the above statement, (Δ_X, Δ_Y) also lies in each of the sets $\mathcal{R}_{AM}(R_x, R_y)$, $\mathcal{R}_{MA}(R_x, R_y)$, and $\mathcal{R}_{AA}(R_x, R_y)$. Since each of these sets are closed by definition, we must have

$$\begin{aligned}&\overline{\mathcal{R}^\star}(R_x, R_y) \subseteq \\ &\qquad \mathcal{R}_{AM}(R_x, R_y) \cap \mathcal{R}_{MA}(R_x, R_y) \cap \mathcal{R}_{AA}(R_x, R_y).\end{aligned}$$

Since each point in the sets $\mathcal{R}_{AM}(R_x, R_y)$, $\mathcal{R}_{MA}(R_x, R_y)$, and $\mathcal{R}_{AA}(R_x, R_y)$ is achievable by vector quantization and/or random binning, the second statement of the Theorem is proved.

To show the reverse inclusion, fix $\epsilon > 0$ and suppose $(\Delta_X, \Delta_Y) \in \mathcal{R}_{AM}(R_x, R_y) \cap \mathcal{R}_{MA}(R_x, R_y) \cap \mathcal{R}_{AA}(R_x, R_y)$. This implies the existence of $(2^{n_{AM}R_x}, 2^{n_{AM}R_y}, n_{AM})$,

$(2^{n_{MA}R_x}, 2^{n_{MA}R_y}, n_{MA})$, and $(2^{n_{AA}R_x}, 2^{n_{AA}R_y}, n_{AA})$ codes satisfying:

$$\begin{aligned}
\Delta_X &\leq \frac{1}{n_{AM}} I(X^{n_{AM}}; f_x^{AM}(X^{n_{AM}}), f_y^{AM}(Y^{n_{AM}})) + \epsilon, \\
\Delta_Y &\geq \frac{1}{n_{AM}} I(Y^{n_{AM}}; f_x^{AM}(X^{n_{AM}}), f_y^{AM}(Y^{n_{AM}})) - \epsilon. \\
\Delta_X &\geq \frac{1}{n_{MA}} I(X^{n_{MA}}; f_x^{MA}(X^{n_{MA}}), f_y^{MA}(Y^{n_{MA}})) - \epsilon, \\
\Delta_Y &\leq \frac{1}{n_{MA}} I(Y^{n_{MA}}; f_x^{MA}(X^{n_{MA}}), f_y^{MA}(Y^{n_{MA}})) + \epsilon, \\
\Delta_X &\leq \frac{1}{n_{AA}} I(X^{n_{AA}}; f_x^{AA}(X^{n_{AA}}), f_y^{AA}(Y^{n_{AA}})) + \epsilon, \\
\Delta_Y &\leq \frac{1}{n_{AA}} I(Y^{n_{AA}}; f_x^{AA}(X^{n_{AA}}), f_y^{AA}(Y^{n_{AA}})) + \epsilon.
\end{aligned}$$

Also, by taking f_x^{MM}, f_y^{MM} to be constants, we trivially have a $(2^{n_{MM}R_x}, 2^{n_{MM}R_y}, n_{MM})$ code satisfying

$$\begin{aligned}
\Delta_X &\geq \frac{1}{n_{MM}} I(X^{n_{MM}}; f_x^{MM}(X^{n_{MM}}), f_y^{MM}(Y^{n_{MM}})), \\
\Delta_Y &\geq \frac{1}{n_{MM}} I(Y^{n_{MM}}; f_x^{MM}(X^{n_{MM}}), f_y^{MM}(Y^{n_{MM}})).
\end{aligned}$$

It is readily verified that, by an appropriate timesharing between these four codes, there exists a $(2^{nR_x}, 2^{nR_y}, n)$ code satisfying

$$\begin{aligned}
\left| \Delta_X - \frac{1}{n} I(X^n; f_x(X^n), f_y(Y^n)) \right| &\leq \delta(\epsilon), \text{ and} \\
\left| \Delta_Y - \frac{1}{n} I(Y^n; f_x(X^n), f_y(Y^n)) \right| &\leq \delta(\epsilon).
\end{aligned}$$

This completes the proof of the theorem. □

Remark 3. *To conclude this section, we remark that in the state amplification and masking problems considered in [KSC08] and [MS07], the authors obtain explicit characterizations of the achievable regions when the channel state and noise are independent Gaussian random variables. Presumably, this could also be accomplished in our setting using known results on Gaussian multiterminal source coding, however, a compete investigation into this matter is left to future work.*

2.7 Other Multiterminal Problems under Logarithmic Loss

In light of the fact that the CEO and multiterminal source coding problems admit solutions under logarithmic loss, it is natural to ask whether other open source coding problems can be solved when distortion is measured under logarithmic loss. It turns out that the answer is affirmative, and we present two here which have particularly concise solutions. This is by no means an exhaustive list of problems which can easily be solved under logarithmic loss.

2.7.1 Multiple Description Coding

The multiple description problem for a discrete memoryless source X is depicted in Figure 2.8. Briefly, a $(2^{nR_1}, 2^{nR_2}, n)$ multiple description code consists of two encoders:

$$\begin{aligned} g_1 &: X^n \mapsto M_1 \in \{1, \ldots, 2^{nR_1}\} \\ g_2 &: X^n \mapsto M_2 \in \{1, \ldots, 2^{nR_2}\}, \end{aligned}$$

and three decoders:

$$\begin{aligned} \psi_0 &: (M_1, M_2) \mapsto \hat{X}_0^n \\ \psi_1 &: M_1 \mapsto \hat{X}_1^n \\ \psi_2 &: M_2 \mapsto \hat{X}_2^n. \end{aligned}$$

A rate distortion tuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes satisfying

$$\limsup_{n \to \infty} \mathbb{E}\left[d(X^n, \hat{X}_j^n)\right] \leq D_j \text{ for } j = 0, 1, 2.$$

Although a characterization of the achievable rate distortion region is unknown in general[6], it is readily described when distortion is measured under logarithmic loss.

[6] The reader is directed to [EK12, Chapter 13] for an introduction to known results for the multiple description problem.

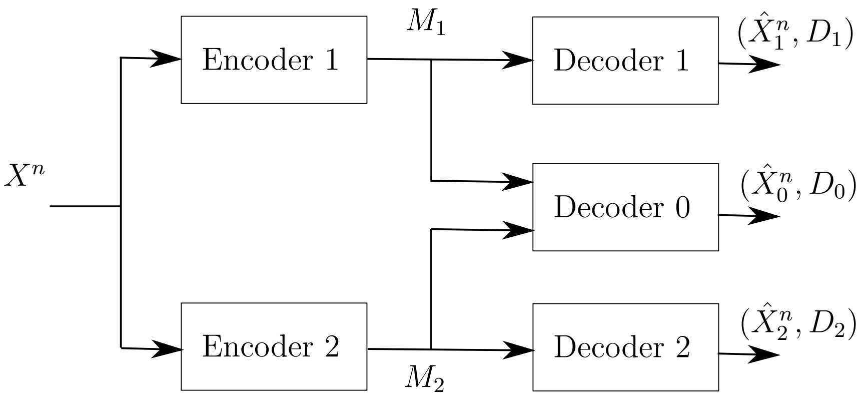

Figure 2.8: The multiple description coding network.

Theorem 14. *For the multiple description problem, the achievable rate distortion region under logarithmic loss consists of all tuples* $(R_1, R_2, D_0, D_1, D_2)$ *satisfying:*

$$\begin{aligned} D_1 &\geq [H(X) - R_1]^+ \\ D_2 &\geq [H(X) - R_2]^+ \\ D_0 &\geq [H(X) - (R_1 + R_2)]^+ , \end{aligned}$$

where $[x]^+ = \max x, 0$*.*

We remark that the achievable region coincides with the El Gamal-Cover Inner Bound for multiple description coding (cf. [GC82, EK12]).

Proof. Converse Part. By Lemma 1, we have

$$nD_1 \geq H(X^n|M_1) = H(X^n) - I(M_1; X^n) \geq nH(X) - nR_1.$$

In a similar fashion, we can obtain the desired lower bounds on D_2 and D_0.

Direct Part. Let M_1 be a lossless encoding of the first $n_1 = nR_1/H(X)$ symbols of X^n. By the lossless source coding theorem, this can be achieved with rate R_1. Next, let M_2 be a lossless encoding of the last $n_2 = nR_2/H(X)$ symbols of X^n. This can be achieved with rate R_2.

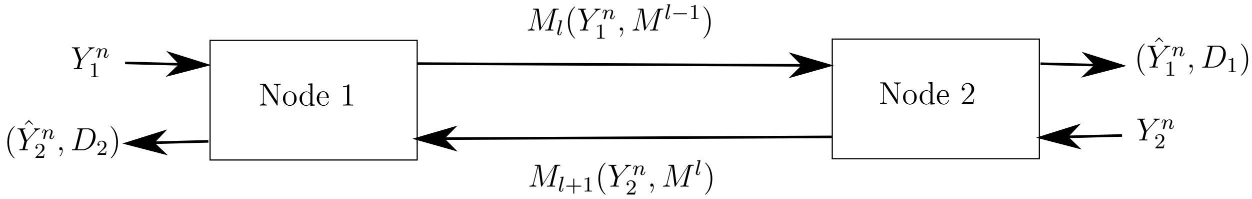

Figure 2.9: Two-way lossy source coding.

This scheme allows decoder 1 to achieve distortion $H(X) - R_1$, since zero distortion is incurred for the first n_1 symbols of X^n, and distortion $H(X)$ is incurred (on average) for the remaining $(n - n_1)$ symbols. By a similar argument, distortions $H(X) - R_2$ and $[H(X) - (R_1 + R_2)]^+$ are achieved by decoders 2 and 0, respectively. □

2.7.2 Two-Way Lossy Source Coding

The interactive source coding problem illustrated in Figure 2.9 is known as the Two-Way Lossy Source Coding problem (cf. [EK12, Chapter 20]). As with the multiple description problem, a characterization of the achievable rate distortion region for the two-way source coding problem is not known in general.

Without loss of generality, we can assume that the communication takes place in rounds $l = 1, 2, \dots, m$, where Node 1 transmits during odd rounds, and Node 2 transmits during even rounds. To be more specific, a $(2^{nr_1}, \dots, 2^{nr_m}, n)$ interactive coding scheme can be described as follows:

- During round l, where l is odd, Node 1 sends a message $M_l = M_l(Y_1^n, M^{l-1}) \in \{1, \dots, 2^{nr_l}\}$. The notation $M_l(Y_1^n, M^{l-1})$ signifies that M_l can be a function of Y_1^n and all previous messages.
- During round l, where l is even, Node 2 sends a message $M_l = M_l(Y_2^n, M^{l-1}) \in \{1, \dots, 2^{nr_l}\}$. Again, the notation $M_l(Y_2^n, M^{l-1})$ signifies that M_l can be a function of Y_2^n and all previous messages.

- After round m concludes, Node 1 generates a reproduction sequence $\hat{Y}_2^n$ as a function of $(Y_1^n, M_1, \dots, M_m)$. Similarly, Node 2 generates a reproduction sequence $\hat{Y}_1^n$ as a function of $(Y_2^n, M_1, \dots, M_m)$.

A rate distortion tuple (R_1, R_2, D_1, D_2) is achievable if there exists a sequence of $(2^{nr_1}, \dots, 2^{nr_m}, n)$ interactive coding schemes satisfying

$$\sum_{\substack{l=1 \\ l \text{ odd}}}^{m} r_l \leq R_1$$

$$\sum_{\substack{l=2 \\ l \text{ even}}}^{m} r_l \leq R_2$$

$$\limsup_{n\to\infty} \mathbb{E}\left[d(Y_j^n, \hat{Y}_j^n)\right] \leq D_j \text{ for } j = 1, 2.$$

Theorem 15. *For the two-way lossy source coding problem, the achievable rate distortion region under logarithmic loss consists of all tuples (R_1, R_2, D_1, D_2) satisfying:*

$$R_1 \geq H(Y_1|Y_2) - D_1$$

$$R_2 \geq H(Y_2|Y_1) - D_2.$$

Proof. Converse Part. To see the converse, enhance Node 1 by providing it with additional side information Y_2^n. Clearly, Node 1 does not require multiple transmission rounds in this enhanced setting, since it can simulate any messages it would receive from Node 2 in advance. Thus, let $M_1 = M_1(Y_1^n, Y_2^n)$ be the message sent by Node 1. Now, Lemma 1 implies that

$$nD_1 \geq H(Y_1^n|M_1, Y_2^n) = H(Y_1^n|Y_2^n) - I(Y_1^n; M_1|Y_2^n) \geq n(H(Y_1|Y_2) - R_1)$$

for this enhanced setting. Therefore, it immediately follows that $R_1 \geq H(Y_1|Y_2) - D_1$ for the two-way source coding problem. The inequality $R_2 \geq H(Y_2|Y_1) - D_2$ follows by a symmetric argument.

Direct Part. Achievability follows by letting each node perform Wyner-Ziv coding (no interaction is required). □

Remark 4. *Theorem 15 is due to Yeow Khiang Chia [Chi12].*

2.8 Concluding Remarks

One immediate direction for further work would be to extend our results on the multiterminal source coding problem to more than two encoders. For the CEO problem, our results can be extended to an arbitrary number of encoders[7]. On the other hand, generalizing the results for the two-encoder source coding problem with distortion constraints on Y_1 and Y_2 poses a significant challenge. The obvious point of difficulty in the proof is extending the tuning argument to higher dimensions so that it yields a distribution with the desired properties. In fact, a "quick-fix" to the tuning argument alone would not be sufficient since this would imply that the Berger-Tung inner bound is tight for more than two encoders. This is known to be false (even for the logarithmic loss distortion measure) since the Berger-Tung achievability scheme is not optimal for the lossless modulo-sum problem studied by Körner and Marton in [KM79].

Another potential direction would be to extend the multiple description result to $m > 2$ descriptions and 2^m-1 decoders. Although the case for $m = 2$ descriptions is readily solved, the problem appears nontrivial for $m > 2$ descriptions.

Acknowledgement

A large part of the material in this chapter is based on the works [CW11c, CW12a] by Courtade and Weissman. Section 2.6 is based on [Cou12]. The author thanks Yeow Khiang Chia for providing the result on two-way lossy source coding in Section 2.7.2.

[7] This extension is proved in Appendix 2.B.

2.A Cardinality Bounds on Auxiliary Random Variables

In order to obtain tight cardinality bounds on the auxiliary random variables used throughout this chapter, we refer to a recent result by Jana. In [Jan09], the author carefully applies the Caratheodory-Fenchel-Eggleston theorem in order to obtain tight cardinality bounds on the auxiliary random variables in the Berger-Tung inner bound. This result extends the results and techniques employed by Gu and Effros for the Wyner-Ahlswede-Körner problem [GE07], and by Gu, Jana, and Effros for the Wyner-Ziv problem [GJE08]. We now state Jana's result, appropriately modified for our purposes:

Consider an arbitrary joint distribution $p(v, y_1, \ldots, y_m)$ with random variables $V, Y_1, \ldots, Y_m$ coming from alphabets $\mathcal{V}, \mathcal{Y}_1, \ldots, \mathcal{Y}_m$ respectively.

Let $d_l : \mathcal{V} \times \hat{\mathcal{V}}_l \to \mathbb{R}$, $1 \le l \le L$ be arbitrary distortion measures defined for possibly different reproduction alphabets $\hat{\mathcal{V}}_l$.

Definition 12. *Define $\mathcal{A}^\star$ to be the set of $(m+L)$-vectors $(R_1, \ldots, R_m, D_1, \ldots, D_L)$ satisfying the following conditions:*

1. *auxiliary random variables $U_1, \ldots, U_m$ exist such that*

$$\sum_{i \in \mathcal{I}} R_i \ge I(Y_{\mathcal{I}}; U_{\mathcal{I}} | U_{\mathcal{I}^c}), \text{ for all } \mathcal{I} \subseteq \{1, \ldots, m\}, \text{ and}$$

2. *mappings $\psi_l : \mathcal{U}_1 \times \cdots \times \mathcal{U}_m \to \hat{\mathcal{V}}_l$, $1 \le l \le L$ exist such that*

$$\mathbb{E} d_l(V, \psi_l(U_1, \ldots, U_m)) \le D_l$$

for some joint distribution

$$p(v, y_1, \ldots, y_m) \prod_{j=1}^{m} p(u_j | y_j).$$

Lemma 6 (Lemma 2.2 from [Jan09])**.** *Every extreme point of $\mathcal{A}^\star$ corresponds to some choice of auxiliary variables $U_1, \ldots, U_m$ with alphabet sizes $|\mathcal{U}_j| \le |\mathcal{Y}_j|$, $1 \le j \le m$.*

In order to obtain the cardinality bounds for the CEO problem, we simply let $L = 1$, $V = X$, and $\hat{\mathcal{V}}_1 = \hat{\mathcal{X}}$. Defining

$$d_1(x, \hat{x}) = \log\left(\frac{1}{\hat{x}(x)}\right),$$

we see that $\overline{\mathcal{RD}}^{\star}_{CEO} = \mathrm{conv}\,(\mathcal{A}^{\star})$, where $\mathrm{conv}\,(\mathcal{A}^{\star})$ denotes the convex hull of $\mathcal{A}^{\star}$. Therefore, Lemma 6 implies that all extreme points of $\overline{\mathcal{RD}}^{\star}_{CEO}$ are achieved with a choice of auxiliary random variables $U_1, \ldots, U_m$ with alphabet sizes $|\mathcal{U}_j| \leq |\mathcal{Y}_j|$, $1 \leq j \leq m$. By timesharing between extreme points, any point in $\overline{\mathcal{RD}}^{\star}_{CEO}$ can be achieved for these alphabet sizes.

Obtaining the cardinality bounds for the multiterminal source coding problem proceeds in a similar fashion. In particular, let $L = m = 2$, $V = (Y_1, Y_2)$, and $\hat{\mathcal{V}}_j = \hat{\mathcal{Y}}_j$, $j = 1, 2$. Defining

$$d_j((y_1, y_2), \hat{y}_j) = \log\left(\frac{1}{\hat{y}_j(y_j)}\right) \text{ for } j = 1, 2,$$

we see that $\overline{\mathcal{RD}}^{\star} = \mathrm{conv}\,(\mathcal{A}^{\star})$. In this case, Lemma 6 implies that all extreme points of $\overline{\mathcal{RD}}^{\star}$ are achieved with a choice of auxiliary random variables U_1, U_2 with alphabet sizes $|\mathcal{U}_j| \leq |\mathcal{Y}_j|$, $1 \leq j \leq 2$. By timesharing between extreme points, any point in $\overline{\mathcal{RD}}^{\star}$ can be achieved for these alphabet sizes.

In order to obtain cardinality bounds on the timesharing variable Q, we can apply Caratheodory's theorem (cf. [Wit80]). In particular, if $C \subset \mathbb{R}^n$ is compact, then any point in $\mathrm{conv}(C)$ is a convex combination of at most $n + 1$ points of C. Taking C to be the closure of the set of extreme points of $\mathcal{A}^{\star}$ is sufficient for our purposes (boundedness of C can be dealt with by a standard truncation argument).

2.B Extension of CEO Results to m Encoders

In this appendix, we prove the generalization of Theorem 3 to m encoders, which essentially amounts to extending Lemma 2 to the general case. We begin by stating the m-encoder generalizations of Theorems 1 and 2, the proofs of which are trivial extensions of the proofs given for the two-encoder case and are therefore omitted.

Definition 13. *Let $\mathcal{R}^i_{CEO,m}$ be the set of all $(R_1, \ldots, R_m, D)$ satisfying*

$$\sum_{i\in\mathcal{I}} R_i \geq I(Y_{\mathcal{I}}; U_{\mathcal{I}}|U_{\mathcal{I}^c}, Q) \text{ for all } \mathcal{I} \subseteq \{1, \ldots, m\}$$
$$D \geq H(X|U_1, \ldots, U_m, Q).$$

for some joint distribution $p(q)p(x)\prod_{i=1}^m p(y_i|x)p(u_i|y_i, q)$.

Theorem 16. *All rate distortion vectors $(R_1, \ldots, R_m, D) \in \mathcal{R}^i_{CEO,m}$ are achievable.*

Definition 14. *Let $\mathcal{R}^o_{CEO,m}$ be the set of $(R_1, \ldots, R_m, D)$ satisfying*

$$\sum_{i\in\mathcal{I}} R_i \geq \sum_{i\in\mathcal{I}} I(U_i; Y_i|X, Q) + H(X|U_{\mathcal{I}^c}, Q) - D \text{ for all } \mathcal{I} \subseteq \{1, \ldots, m\} \tag{2.90}$$
$$D \geq H(X|U_1, \ldots, U_m, Q). \tag{2.91}$$

for some joint distribution $p(q)p(x)\prod_{i=1}^m p(y_i|x)p(u_i|y_i, q)$.

Theorem 17. *If $(R_1, \ldots, R_m, D)$ is strict-sense achievable, then $(R_1, \ldots, R_m, D) \in \mathcal{R}^o_{CEO,m}$.*

Given the definitions of $\mathcal{R}^i_{CEO,m}$ and $\mathcal{R}^o_{CEO,m}$, the generalization of Theorem 3 to m encoders is an immediate consequence of the following lemma:

Lemma 7. $\mathcal{R}^o_{CEO,m} \subseteq \mathcal{R}^i_{CEO,m}$.

Proof. Suppose $(R_1, \ldots, R_m, D) \in \mathcal{R}^o_{CEO,m}$, then by definition there exists $p(q)$ and conditional distributions $\{p(u_i|y_i, q)\}_{i=1}^m$ so that (2.90) and (2.91) are satisfied. For the joint distribution corresponding to $p(q)$ and conditional distributions $p\{(u_i|y_i, q)\}_{i=1}^m$, define $\mathcal{P}_D \subset \mathbb{R}^m$ to be the polytope defined by the inequalities (2.90). Now, to show $(R_1, \ldots, R_m, D) \in \mathcal{R}^i_{CEO,m}$, it suffices to show that each extreme point of $\mathcal{P}_D$ is dominated by a point in $\mathcal{R}^i_{CEO,m}$ that achieves distortion at most D.

To this end, define the set function $f : 2^{[m]} \to \mathbb{R}$ as follows:

$$\begin{aligned} f(\mathcal{I}) &:= I(Y_{\mathcal{I}}; U_{\mathcal{I}}|U_{\mathcal{I}^c}, Q) - (D - H(X|U_1, \ldots, U_m, Q)) \\ &= \sum_{i\in\mathcal{I}} I(U_i; Y_i|X, Q) + H(X|U_{\mathcal{I}^c}, Q) - D. \end{aligned}$$

It can be verified that the function f and the function $f^+(\mathcal{I}) = \max\{f(\mathcal{I}), 0\}$ are supermodular functions (see Appendix 2.C). By construction, $\mathcal{P}_D$ is equal to the set of $(R_1, \ldots, R_m)$ which satisfy:

$$\sum_{i \in \mathcal{I}} R_i \geq f^+(\mathcal{I}).$$

It follows by basic results in submodular optimization (see Appendix 2.C) that, for a linear ordering $i_1 \prec i_2 \prec \cdots \prec i_m$ of $\{1, \ldots, m\}$, an extreme point of $\mathcal{P}_D$ can be greedily computed as follows:

$$\tilde{R}_{i_j} = f^+(\{i_1, \ldots, i_j\}) - f^+(\{i_1, \ldots, i_{j-1}\}) \text{ for } j = 1, \ldots, m.$$

Furthermore, all extreme points of $\mathcal{P}_D$ can be enumerated by looking over all linear orderings $i_1 \prec i_2 \prec \cdots \prec i_m$ of $\{1, \ldots, m\}$. Each ordering of $\{1, \ldots, m\}$ is analyzed in the same manner, hence we assume (for notational simplicity) that the ordering we consider is the natural ordering $i_j = j$.

Let j be the first index for which $\tilde{R}_j > 0$. Then, by construction,

$$\tilde{R}_k = I(U_k; Y_k | U_{k+1}, \ldots, U_m, Q) \text{ for all } k > j.$$

Furthermore, we must have $f(\{1, \ldots, j'\}) \leq 0$ for all $j' < j$. Thus, $\tilde{R}_j$ can be expressed as

$$\begin{aligned}
\tilde{R}_j &= \sum_{i=1}^{j} I(Y_i; U_i | X, Q) + H(X | U_{j+1}, \ldots, U_m, Q) - D \\
&= I(Y_j; U_j | U_{j+1}, \ldots, U_m, Q) + f(\{1, \ldots, j-1\}) \\
&= (1-\theta) I(Y_j; U_j | U_{j+1}, \ldots, U_m, Q),
\end{aligned}$$

where $\theta \in [0, 1)$ is defined as:

$$\begin{aligned}
\theta &= \frac{-f(\{1, \ldots, j-1\})}{I(Y_j; U_j | U_{j+1}, \ldots, U_m, Q)} \\
&= \frac{D - H(X | U_1, \ldots, U_m, Q) - I(U_1, \ldots, U_{j-1}; Y_1, \ldots, Y_{j-1} | U_j, \ldots, U_m, Q)}{I(Y_j; U_j | U_{j+1}, \ldots, U_m, Q)}.
\end{aligned}$$

By the results of Theorem 16, the rates $(\tilde{R}_1, \ldots, \tilde{R}_m)$ permit the following coding scheme: For a fraction $(1-\theta)$ of the time, a codebook can be used that allows the decoder to recover $U_j^n, \ldots, U_m^n$ with high probability. The other fraction θ of the time, a codebook can be used that allows the decoder to recover $U_{j+1}^n, \ldots, U_m^n$ with high probability. As $n \to \infty$, this coding scheme can achieve distortion

$$\begin{aligned}
\tilde{D} &= (1-\theta)H(X|U_j, \ldots, U_m, Q) + \theta H(X|U_{j+1}, \ldots, U_m, Q) \\
&= H(X|U_j, \ldots, U_m, Q) + \theta I(X; U_j|U_{j+1}, \ldots, U_m, Q) \\
&= H(X|U_j, \ldots, U_m, Q) + \frac{I(X; U_j|U_{j+1}, \ldots, U_m, Q)}{I(Y_j; U_j|U_{j+1}, \ldots, U_m, Q)} \times \\
&\quad [D - H(X|U_1, \ldots, U_m, Q) - I(U_1, \ldots, U_{j-1}; Y_1, \ldots, Y_{j-1}|U_j, \ldots, U_m, Q)] \\
&\leq H(X|U_j, \ldots, U_m, Q) + D - H(X|U_1, \ldots, U_m, Q) \\
&\quad - I(U_1, \ldots, U_{j-1}; Y_1, \ldots, Y_{j-1}|U_j, \ldots, U_m, Q) \qquad (2.92) \\
&= D + I(X; U_1, \ldots U_{j-1}|U_j, \ldots, U_m, Q) \\
&\quad - I(U_1, \ldots, U_{j-1}; Y_1, \ldots, Y_{j-1}|U_j, \ldots, U_m, Q) \\
&= D - I(U_1, \ldots, U_{j-1}; Y_1, \ldots, Y_{j-1}|X, U_j, \ldots, U_m, Q) \\
&\leq D. \qquad (2.93)
\end{aligned}$$

In the preceding string of inequalities (2.92) follows since U_j is conditionally independent of everything else given (Y_j, Q), and (2.93) follows from the non-negativity of mutual information.

Therefore, for every extreme point $(\tilde{R}_1, \ldots, \tilde{R}_m)$ of $\mathcal{P}_D$, the point $(\tilde{R}_1, \ldots, \tilde{R}_m, D)$ lies in $\mathcal{R}_{CEO,m}^i$. This proves the lemma. □

Finally, we remark that the results of Appendix 2.A imply that it suffices to consider auxiliary random variables $U_1, \ldots, U_m$ with alphabet sizes $|\mathcal{U}_j| \leq |\mathcal{Y}_j|$, $1 \leq j \leq m$. The timesharing variable Q requires an alphabet size bounded by $|\mathcal{Q}| \leq m + 2$.

2.C Supermodular Functions

In this appendix, we review some basic results in submodular optimization that were used in Appendix 2.B to prove Lemma 7. We tailor our statements toward supermodularity, since this is the property we require in Appendix 2.B.

We begin by defining a supermodular function.

Definition 15. *Let $E = \{1, \dots, n\}$ be a finite set. A function $s : 2^E \to \mathbb{R}$ is* supermodular *if for all $S, T \subseteq E$*

$$s(S) + s(T) \leq s(S \cap T) + s(S \cup T). \tag{2.94}$$

One of the fundamental results in submodular optimization is that a greedy algorithm minimizes a linear function over a supermodular polyhedron. By varying the linear function to be minimized, all extreme points of the supermodular polyhedron can be enumerated. In particular, define the supermodular polyhedron $\mathcal{P}(s) \subset \mathbb{R}^n$ be the set of $x \in \mathbb{R}^n$ satisfying

$$\sum_{i \in T} x_i \geq s(T) \text{ for all } T \subseteq E.$$

The following theorem provides an algorithm that enumerates the extreme points of $\mathcal{P}(s)$.

Theorem 18 (See [Sch03, Fuj10, McC05])**.** *For a linear ordering $e_1 \prec e_2 \prec \cdots \prec e_n$ of the elements in E, Algorithm 2.C.1 returns an extreme point v of $\mathcal{P}(s)$. Moreover, all extreme points of $\mathcal{P}(s)$ can be enumerated by considering all linear orderings of the elements of E.*

Algorithm 2.C.1: GREEDY*$(s, E, \prec)$*

```
comment: Returns extreme point $v$ of $\mathcal{P}(s)$ corresponding to the ordering $\prec$.
for $i = 1, \dots n$
    Set $v_i = s(\{e_1, e_2, \dots, e_i\}) - s(\{e_1, e_2, \dots, e_{i-1}\})$
return $(v)$
```

Proof. See [Sch03, Fuj10, McC05]. □

Theorem 18 is the key tool we employ to establish Lemma 7. In order to apply it, we require the following lemma.

Lemma 8. *For any joint distribution of the form* $p(q)p(x)\prod_{i=1}^{m} p(y_i|x)p(u_i|y_i,q)$ *and fixed* $D \in \mathbb{R}$*, define the set function* $f : 2^{[m]} \to \mathbb{R}$ *as:*

$$\begin{aligned} f(\mathcal{I}) &:= I(Y_{\mathcal{I}}; U_{\mathcal{I}} | U_{\mathcal{I}^c}, Q) - (D - H(X|U_1, \ldots, U_m, Q)) \qquad (2.95) \\ &= \sum_{i \in \mathcal{I}} I(U_i; Y_i | X, Q) + H(X|U_{\mathcal{I}^c}, Q) - D, \end{aligned}$$

and the corresponding non-negative set function $f^+ : 2^{[m]} \to \mathbb{R}$ *as* $f^+ = \max\{f, 0\}$*. The functions* f *and* f^+ *are supermodular.*

Proof. In order to verify that f is supermodular, it suffices to check that the function $f'(\mathcal{I}) = I(Y_{\mathcal{I}}; U_{\mathcal{I}} | U_{\mathcal{I}^c}, Q)$ is supermodular since the latter two terms in (2.95) are constant. To this end, consider sets $T, S \subseteq \{1, \ldots, m\}$ and observe that:

$$\begin{aligned} f'(S) + f'(T) &= I(Y_S; U_S | U_{S^c}, Q) + I(Y_T; U_T | U_{T^c}, Q) \\ &= H(U_S|U_{S^c}, Q) - H(U_S|Y_S, Q) + H(U_T|U_{T^c}, Q) - H(U_T|Y_T, Q) \\ &= H(U_S|U_{S^c}, Q) + H(U_T|U_{T^c}, Q) \\ &\quad - H(U_{S\cup T}|Y_{S\cup T}, Q) - H(U_{S\cap T}|Y_{S\cap T}, Q) \qquad (2.96) \\ &= H(U_{S\setminus T}|U_{S^c}, Q) + H(U_{S\cap T}|U_{(S\cap T)^c}, Q) + H(U_T|U_{T^c}, Q) \\ &\quad - H(U_{S\cup T}|Y_{S\cup T}, Q) - H(U_{S\cap T}|Y_{S\cap T}, Q) \qquad (2.97) \\ &= H(U_{S\setminus T}|U_{S^c}, Q) + H(U_T|U_{T^c}, Q) - H(U_{S\cup T}|Y_{S\cup T}, Q) \\ &\quad + I(U_{S\cap T}; Y_{S\cap T}|U_{(S\cap T)^c}, Q) \\ &\le H(U_{S\setminus T}|U_{(S\cup T)^c}, Q) + H(U_T|U_{T^c}, Q) - H(U_{S\cup T}|Y_{S\cup T}, Q) \\ &\quad + I(U_{S\cap T}; Y_{S\cap T}|U_{(S\cap T)^c}, Q) \qquad (2.98) \\ &= I(U_{S\cup T}; Y_{S\cup T}|U_{(S\cup T)^c}, Q) + I(U_{S\cap T}; Y_{S\cap T}|U_{(S\cap T)^c}, Q) \\ &= f'(S\cap T) + f'(S\cup T). \end{aligned}$$

The labeled steps above can be justified as follows:

- (2.96) follows since U_i is conditionally independent of everything else given (Y_i, Q).
- (2.97) is simply the chain rule.
- (2.98) follows since conditioning reduces entropy.

Next, we show that $f^+ = \max\{f, 0\}$ is supermodular. Observe first that f is monotone increasing, i.e., if $S \subset T$, then $f(S) \leq f(T)$. Thus, fixing $S, T \subseteq \{1, \ldots, m\}$, we can assume without loss of generality that

$$f(S \cap T) \leq f(S) \leq f(T) \leq f(S \cup T).$$

If $f(S \cap T) \geq 0$, then (2.94) is satisfied for $s = f^+$ by the supermodularity of f. On the other hand, if $f(S \cup T) \leq 0$, then (2.94) is a tautology for $s = f^+$. Therefore, it suffices to check the following three cases:

- Case 1: $f(S \cap T) \leq 0 \leq f(S) \leq f(T) \leq f(S \cup T)$. In this case, the supermodularity of f and the fact that $f^+ \geq f$ imply:

$$\begin{aligned} f^+(S \cup T) + f^+(S \cap T) &\geq f(S \cup T) + f(S \cap T) \\ &\geq f(S) + f(T) = f^+(S) + f^+(T). \end{aligned}$$

- Case 2: $f(S \cap T) \leq f(S) \leq 0 \leq f(T) \leq f(S \cup T)$. Since f is monotone increasing, we have:

$$f^+(S \cup T) + f^+(S \cap T) = f(S \cup T) + 0 \geq f(T) + 0 = f^+(S) + f^+(T).$$

- Case 3: $f(S \cap T) \leq f(S) \leq f(T) \leq 0 \leq f(S \cup T)$. By definition of f^+:

$$f^+(S \cup T) + f^+(S \cap T) = f(S \cup T) + 0 \geq 0 + 0 = f^+(S) + f^+(T).$$

Hence, $f^+ = \max\{f, 0\}$ is supermodular.

□

2.D A Lemma for the Daily Double

For a given joint distribution $p(y_1, y_2)$ on the finite alphabet $\mathcal{Y}_1 \times \mathcal{Y}_2$, let $\mathcal{P}(R_1, R_2)$ denote the set of joint pmf's of the form

$$p(q, y_1, y_2, u_1, u_2) = p(q)p(y_1, y_2)p(u_1|y_1, q)p(u_1|y_1, q)$$

which satisfy

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

for given finite alphabets $\mathcal{U}_1$,$\mathcal{U}_2$, $\mathcal{Q}$.

Lemma 9. *For R_1, R_2 satisfying $R_1 \leq H(Y_1)$, $R_2 \leq H(Y_2)$, and $R_1 + R_2 \leq H(Y_1, Y_2)$, the infimum*

$$\inf_{p \in \mathcal{P}(R_1, R_2)} \{H(Y_1|U_1, U_2, Q) + H(Y_2|U_1, U_2, Q)\}$$

is attained by some $p^* \in \mathcal{P}(R_1, R_2)$ which satisfies $R_1 + R_2 = I(Y_1, Y_2; U_1^*, U_2^*|Q^*)$, where U_1^*, U_2^*, Q^* correspond to the auxiliary random variables defined by p^*.

Proof. First, note that the infimum is always attained since $\mathcal{P}(R_1, R_2)$ is compact and the objective function is continuous on $\mathcal{P}(R_1, R_2)$. Therefore, let U_1^*, U_2^*, Q^* correspond to the auxiliary random variables which attain the infimum.

If $H(Y_1|U_1^*, U_2^*, Q^*) + H(Y_2|U_1^*, U_2^*, Q^*) = 0$, then we must have $I(Y_1, Y_2; U_1^*, U_2^*|Q^*) = H(Y_1, Y_2)$. Thus, $R_1 + R_2 = I(Y_1, Y_2; U_1^*, U_2^*|Q^*)$.

Next, consider the case where $H(Y_1|U_1^*, U_2^*, Q^*) + H(Y_2|U_1^*, U_2^*, Q^*) > 0$. Assume for sake of contradiction that $R_1 + R_2 > I(Y_1, Y_2; U_1^*, U_2^*|Q^*)$. For any $p \in \mathcal{P}(R_1, R_2)$:

$$I(Y_1; U_1|U_2, Q) + I(Y_2; U_2|U_1, Q) \leq I(Y_1, Y_2; U_1, U_2|Q).$$

Hence, at most one of the remaining rate constraints can be satisfied with equality. If none of the rate constraints are satisfied with equality, then define

$$(\tilde{U}_1, \tilde{U}_2) = \begin{cases} (U_1^*, U_2^*) & \text{with probability } 1-\epsilon \\ (Y_1, Y_2) & \text{with probability } \epsilon. \end{cases}$$

For $\epsilon > 0$ sufficiently small, the distribution $\tilde{p}$ corresponding to the auxiliary random variables $\tilde{U}_1, \tilde{U}_2, Q^*$ is still in $\mathcal{P}(R_1, R_2)$. However, $\tilde{p}$ satisfies

$$H(Y_1|\tilde{U}_1, \tilde{U}_2, Q^*) + H(Y_2|\tilde{U}_1, \tilde{U}_2, Q^*) < H(Y_1|U_1^*, U_2^*, Q^*) + H(Y_2|U_1^*, U_2^*, Q^*),$$

which contradicts the optimality of p^*.

Therefore, assume without loss of generality that

$$\begin{aligned} R_1 &= I(Y_1; U_1^*|U_2^*, Q^*) \\ R_1 + R_2 &> I(Y_1, Y_2; U_1^*, U_2^*|Q^*). \end{aligned}$$

This implies that $R_2 > I(Y_2; U_2^*|Q^*)$. Now, define

$$\tilde{U}_2 = \begin{cases} U_2^* & \text{with probability } 1-\epsilon \\ Y_2 & \text{with probability } \epsilon. \end{cases}$$

Note that for $\epsilon > 0$ sufficiently small:

$$\begin{aligned} I(Y_2; U_2^*|Q^*) &< I(Y_2; \tilde{U}_2|Q^*) < R_2 \\ I(Y_1, Y_2; U_1^*, U_2^*|Q^*) &< I(Y_1, Y_2; U_1^*, \tilde{U}_2|Q^*) < R_1 + R_2, \end{aligned}$$

and for any $\epsilon \in [0, 1]$:

$$\begin{aligned} R_1 = I(Y_1; U_1^*|U_2^*, Q^*) &\geq I(Y_1; U_1^*|\tilde{U}_2, Q^*) \\ H(Y_1|U_1^*, U_2^*, Q^*) + H(Y_2|U_1^*, U_2^*, Q^*) &\geq H(Y_1|U_1^*, \tilde{U}_2, Q^*) + H(Y_2|U_1^*, \tilde{U}_2, Q^*). \end{aligned} \tag{2.99}$$

Since $R_2 \leq H(Y_2)$, as ϵ is increased from 0 to 1, at least one of the following must occur:

1. $I(Y_2; \tilde{U}_2|Q^*) = R_2$.

2. $I(Y_1, Y_2; U_1^*, \tilde{U}_2 | Q^*) = R_1 + R_2$.

3. $I(Y_1; U_1 | \tilde{U}_2, Q^*) < R_1$.

If either of events 1 or 2 occur first then the sum-rate constraint is met with equality (since they are equivalent in this case). If event 3 occurs first, then all rate constraints are satisfied with strict inequality and we can apply the above argument to contradict optimality of p^*. Since (2.99) shows that the objective is nonincreasing in ϵ, there must exist a $\tilde{p} \in \mathcal{P}(R_1, R_2)$ which attains the infimum and satisfies the sum-rate constraint with equality. $\square$

CHAPTER 3

Coded Cooperative Data Exchange

3.1 Introduction

Consider a connected network of n nodes that all wish to recover k desired packets. Each node begins with a subset of the desired packets and broadcasts messages to its neighbors over discrete, memoryless, and interference-free channels. Furthermore, every node knows which packets are already known by each node and knows the topology of the network. How many transmissions are required to disseminate the k packets to every node in the network? How should this be accomplished? These are the essential questions addressed. We refer to this as the *Coded Cooperative Data Exchange* problem, or just the Cooperative Data Exchange problem.

This work is motivated in part by emerging issues in distributed data storage. Consider the problem of backing up data on servers in a large data center. One commonly employed method to protect data from corruption is replication. Using this method, large quantities of data are replicated in several locations so as to protect from various sources of corruption (e.g., equipment failure, power outages, natural disasters, etc.). As the quantity of information in large data centers continues to increase, the number of file transfers required to complete a periodic replication task is becoming an increasingly important consideration due to time, equipment, cost, and energy constraints. The results presented in this chapter address these issues.

This model also has natural applications in the context of tactical networks, and we give one of them here. Consider a scenario in which an aircraft flies over a group of nodes on the ground and tries to deliver a video stream. Each ground node might only receive a subset of the transmitted packets due to interference, obstructions, and other signal integrity issues. In order to recover the transmission, the nodes are free to communicate with their neighbors, but would like to minimize the number of transmissions in order to conserve battery power (or avoid detection, etc.). How should the nodes share information, and what is the minimum number of transmissions required so that the entire network can recover the video stream?

Beyond the examples mentioned above, the results presented herein can also be applied to practical secrecy generation amongst a collection of nodes. We consider this application in detail in Section 3.4.

3.1.1 Related Work

Distributed data exchange problems have received a great deal of attention over the past several years. The powerful techniques afforded by network coding [ACR00, RMY03] have paved the way for cooperative communications at the packet-level.

The coded cooperative data exchange problem (also called the universal recovery problem in [CXW10, CW10, CW11b]) was originally introduced by El Rouayheb *et al.* in [ECS07, ESS10] for a fully connected network (i.e., a single-hop network). For this special case, a randomized algorithm for finding an optimal transmission scheme was given in [SSB10b], and the first deterministic algorithm was recently given in [SSB10a]. In the concluding remarks of [SSB10a], the weighted universal recovery problem (in which the objective is to minimize the weighted sum of transmissions by nodes) was posed as an open problem. However, this was solved using a variant of the same algorithm in [OS11], and independently by the present authors using a submodular algorithm in [CW11b].

The coded cooperative data exchange problem is related to the index coding problem originally introduced by Birk and Kol in [BK06]. Specifically, generalizing the index coding problem to permit each node to be a transmitter (instead of having a single server) and further generalizing so that the network need not be a single hop network leads to a class of problems that includes our problem as a special case in which each node desires to receive all packets.

One significant result in index coding is that nonlinear index coding outperforms the best linear index code in certain cases [LS09, ALS08]. As discussed above, our problem is a special case of the generalized index coding problem, and it turns out that linear encoding does achieve the minimum number of transmissions required for universal recovery and this solution is computable in polynomial time for some important cases.

In this chapter, we apply principles of cooperative data exchange to generate secrecy in the presence of an eavesdropper. In this context, the secrecy generation problem was originally studied in [CN04]. In [CN04], Csiszar and Narayan gave single-letter characterizations of the secret-key and private-key capacities for a network of nodes connected by an error-free broadcast channel. While general and powerful, these results left two practical issues as open questions. First, (as with many information-theoretic investigations) the results require the nodes to observe arbitrarily long sequences of i.i.d. source symbols, which is generally not practical. Second, no efficient algorithm is provided in [CN04] which achieves the respective secrecy capacities. More recent work in [YN05, YR10] addressed the latter point.

3.1.2 Our Contributions

We provide necessary and sufficient conditions for achieving universal recovery[1] in arbitrarily connected multihop networks. We specialize these necessary and sufficient conditions to obtain precise results in the case where the underlying network topology satisfies some modest regularity conditions.

[1] In this chapter, we use the term *universal recovery* to refer to the ultimate condition where every node has successfully recovered all packets.

For the case of a fully connected network, we provide an algorithm based on submodular optimization which solves the cooperative data exchange problem. This algorithm is unique from the others previously appearing in the literature (cf. [SSB10b, SSB10a, OS11]) in that it exploits submodularity. As a corollary, we provide exact concentration results when packets are randomly distributed in a network.

In this same vein, we also obtain tight concentration results and approximate solutions when the underlying network is d-regular and packets are distributed randomly.

Furthermore, if packets are divisible (allowing transmissions to consist of partial packets), we prove that the traditional cut-set bounds can be achieved for any network topology. In the case of d-regular and fully connected networks, we show that splitting packets does not typically provide any significant benefits.

Finally, for the application to secrecy generation, we leverage the results of [CN04] in the context of the cooperative data exchange problem for a fully connected network. In doing so, we provide an efficient algorithm that achieves the secrecy capacity without requiring any quantities to grow asymptotically large.

3.1.3 Organization

This chapter is organized as follows. Section 3.2 formally introduces the problem and provides basic definitions and notation. Section 3.3 presents our main results. Section 3.4 discusses the application of our results to secrecy generation by a collection of nodes in the presence of an eavesdropper. Section 3.5 contains the relevant proofs. Section 3.6 delivers the conclusions and discusses directions for future work.

3.2 System Model and Definitions

Before we formally introduce the problem, we establish some notation. Let $\mathbb{N} = 0, 1, 2, \ldots$ denote the set of natural numbers. For two sets A and B, the relation $A \subset B$ implies that

A is a proper subset of B (i.e., $A \subseteq B$ and $A \neq B$). For a set A, the corresponding power set is denoted $2^A := \{B : B \subseteq A\}$. We use the notation $[m]$ to denote the set $\{1, \ldots, m\}$.

This chapter considers a network of n nodes. The network must be connected, but it need not be fully connected (i.e., it need not be a complete graph). A graph $\mathcal{G} = (V, E)$ describes the specific connections in the network, where V is the set of vertices $\{v_i : i \in \{1, \ldots, n\}\}$ (each corresponding to a node) and E is the set of edges connecting nodes. We assume that the edges in E are undirected, but our results can be extended to directed graphs.

Each node wishes to recover the same k desired packets, and each node begins with a (possibly empty) subset of the desired packets. Formally, let $P_i \subseteq \{p_1, \ldots, p_k\}$ be the (indexed) set of packets originally available at node i, and $\{P_i\}_{i=1}^n$ satisfies $\bigcup_{i=1}^n P_i = \{p_1, \ldots, p_k\}$. Each $p_j \in \mathbb{F}$, where $\mathbb{F}$ is some finite field (e.g. $\mathbb{F} = \mathrm{GF}(2^m)$). For our purposes, it suffices to assume $|\mathbb{F}| \geq 2n$. The set of packets initially missing at node i is denoted $P_i^c := \{p_1, \ldots, p_k\} \backslash P_i$.

Throughout this chapter, we assume that each packet $p_i \in \{p_1, \ldots, p_k\}$ is equally likely to be any element of $\mathbb{F}$. Moreover, we assume that packets are independent of one another. Thus, no correlation between different packets or prior knowledge about unknown packets can be exploited.

To simplify notation, we will refer to a given problem instance (i.e., a graph and corresponding sets of packets available at each node) as a network $\mathcal{T} = \{\mathcal{G}, P_1, \ldots, P_n\}$. When no ambiguity is present, we will refer to a network by $\mathcal{T}$ and omit the implicit dependence on the parameters $\{\mathcal{G}, P_1, \ldots, P_n\}$.

Let the set $\Gamma(i)$ be the neighborhood of node i. There exists an edge $e \in E$ connecting two vertices $v_i, v_j \in V$ iff $i \in \Gamma(j)$. For convenience, we put $i \in \Gamma(i)$. Node i sends (possibly coded) packets to its neighbors $\Gamma(i)$ over discrete, memoryless, and interference-free channels. In other words, if node i transmits a message, then every node in $\Gamma(i)$ receives that message. If S is a set of nodes, then we define $\Gamma(S) = \cup_{i \in S} \Gamma(i)$. In a similar

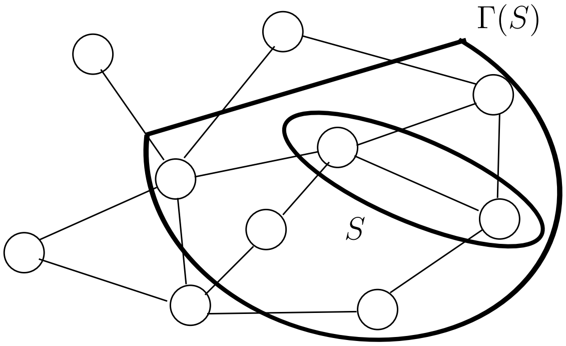

Figure 3.1: For the given graph, a set of vertices S and its neighborhood $\Gamma(S)$ are depicted. The set $\partial(S)$ (i.e., the boundary of S) consists of the four vertices in $\Gamma(S)$ which are not in S.

manner, we define $\partial(S) = \Gamma(S) \backslash S$ to be the boundary of the vertices in S. An example of sets S, $\Gamma(S)$, and $\partial(S)$ is given in Figure 3.1.

We seek to determine the minimum number of transmissions required to achieve universal recovery (when every node has learned all k packets). We primarily consider the case where packets are deemed indivisible. In this case, a single transmission by user i consists of sending a packet (some $z \in \mathbb{F}$) to all nodes $j \in \Gamma(i)$. This motivates the following definition.

Definition 16. *Given a network $\mathcal{T}$, the minimum number of transmissions required to achieve universal recovery is denoted $M^*(\mathcal{T})$.*

To clarify this concept, we briefly consider two examples:

Example 3 (Line Network)**.** *Suppose $\mathcal{T}$ is a network of nodes connected along a line as follows: $V = \{v_1, v_2, v_3\}$, $E = \{(v_1, v_2), (v_2, v_3)\}$, $P_1 = \{p_1\}$, $P_2 = \emptyset$, and $P_3 = \{p_2\}$. Note that each node must transmit at least once in order for all nodes to recover $\{p_1, p_2\}$, hence $M^*(\mathcal{T}) \geq 3$. Suppose node 1 transmits p_1 and node 3 transmits p_2. Then (upon receipt of p_1 and p_2 from nodes 1 and 3, respectively) node 2 transmits $p_1 \oplus p_2$, where $\oplus$ indicates addition in the finite field $\mathbb{F}$. This strategy requires 3 transmissions and allows each user to recover $\{p_1, p_2\}$. Hence $M^*(\mathcal{T}) = 3$.*

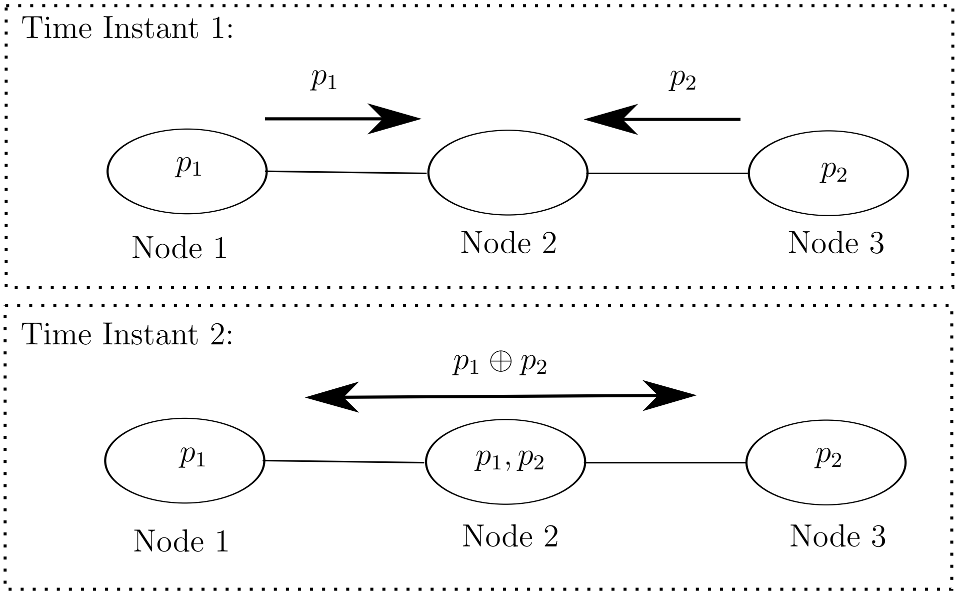

Figure 3.2: An illustration of the transmission scheme employed in Example 3. During the first time instant, Nodes 1 and 3 broadcast packets p_1 and p_2, respectively. During the second time instant, Node 2 broadcasts the XOR of packets p_1 and p_2. This scheme requires three transmissions and achieves universal recovery.

Example 3 demonstrates a transmission schedule that uses two *rounds* of communication. The transmissions by node i in a particular round of communication can depend only on the information available to node i prior to that round (i.e. P_i and previously received transmissions from neighboring nodes). In other words, the transmissions are causal. The transmission scheme employed in Example 3 is illustrated in Figure 3.2.

Example 4 (Fully Connected Network)**.** *Suppose $\mathcal{T}$ is a 3-node fully connected network in which $\mathcal{G}$ is a complete graph on 3 vertices, and $P_i = \{p_1, p_2, p_3\} \backslash p_i$. Clearly one transmission is not sufficient, thus $M^*(\mathcal{T}) \geq 2$. It can be seen that two transmissions suffice: let node 1 transmit p_2 which lets node 2 have $P_2 \cup p_2 = \{p_1, p_2, p_3\}$. Now, node 2 transmits $p_1 \oplus p_3$, allowing nodes 1 and 3 to each recover all three packets. Thus $M^*(\mathcal{T}) = 2$. Since each transmission was only a function of the packets originally available at the corresponding node, this transmission strategy can be accomplished in a single round of communica-*

tion.

In the above examples, we notice that the transmission schemes are partially characterized by a schedule of which nodes transmit during which round of communication. We formalize this notion with the following definition:

Definition 17 (Transmission Schedule)**.** *A set of integers $\{b_i^j : i \in [n], j \in [r], b_i^j \in \mathbb{N}\}$ is called a transmission schedule for r rounds of communication if node i makes exactly b_i^j transmissions during communication round j.*

When the parameters n and r are clear from context, a transmission schedule will be denoted by the shorthand notation $\{b_i^j\}$. Although finding a transmission schedule that achieves universal recovery is relatively easy (e.g., each node transmits all packets in their possession at each time instant), finding one that achieves universal recovery with $M^*(\mathcal{T})$ transmissions can be extremely difficult. This is demonstrated by the following example:

Example 5 (Optimal Cooperative Data Exchange is NP-Hard.)**.** *Suppose $\mathcal{T}$ is a network with $k = 1$ corresponding to a bipartite graph with left and right vertex sets V_L and V_R respectively. Let $P_i = p_1$ for each $i \in V_L$, and let $P_i = \emptyset$ for each $i \in V_R$. In this case, $M^*(\mathcal{T})$ is given by the minimum number of sets in $\{\Gamma(i)\}_{i \in V_L}$ which cover all vertices in V_R. Thus, finding $M^*(\mathcal{T})$ is at least as hard as the Minimum Set Cover problem, which is NP-complete [Kar72].*

Several of our results are stated in the context of *randomly distributed packets.* Assume $0 < q < 1$ is given. Our model is essentially that each packet is available independently at each node with probability q. However, we must condition on the event that each packet is available to at least one node. Thus, when packets are randomly distributed, the underlying probability measure is given by

$$\Pr\left[p_i \in \bigcup_{j \in S} P_j\right] = \frac{1 - (1-q)^{|S|}}{1 - (1-q)^n} \tag{3.1}$$

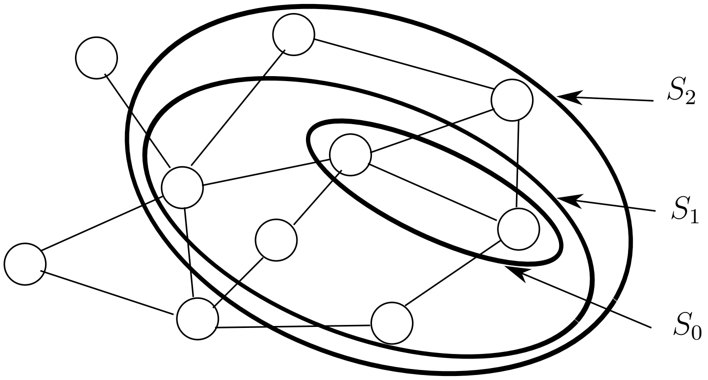

Figure 3.3: An example of a sequence $(S_0, S_1, S_2) \in \mathcal{S}^{(2)}(\mathcal{G})$ for a particular choice of graph $\mathcal{G}$.

for all $i \in [k]$ and all nonempty $S \subseteq V = [n]$.

Finally, we introduce one more definition which links the network topology with the number of communication rounds, r.

Definition 18. *For a graph $\mathcal{G} = (V, E)$ on n vertices, define $\mathcal{S}^{(r)}(\mathcal{G}) \subset (2^V)^{r+1}$ as follows: $(S_0, S_1, \dots, S_r) \in \mathcal{S}^{(r)}(\mathcal{G})$ if and only if the sets $\{S_i\}_{i=0}^r$ satisfy the following two conditions:*

$$\emptyset \subset S_i \subset V \quad \textit{for each } 0 \leq i \leq r, \textit{ and}$$
$$S_{i-1} \subseteq S_i \subseteq \Gamma(S_{i-1}) \quad \textit{for each } 1 \leq i \leq r.$$

In words, any element in $\mathcal{S}^{(r)}(\mathcal{G})$ is a nested sequence of subsets of vertices of $\mathcal{G}$. Moreover, the constraint that each set in the sequence is contained in its predecessor's neighborhood implies that the sets cannot expand too quickly relative to the topology of $\mathcal{G}$.

To make the definition of $\mathcal{S}^{(r)}(\mathcal{G})$ more concrete, we have illustrated a sequence $(S_0, S_1, S_2) \in \mathcal{S}^{(2)}(\mathcal{G})$ for a particular choice of graph $\mathcal{G}$ in Figure 3.3.

3.3 Main Results

In this section, we present our main results. Proofs are delayed until Section 3.5.

3.3.1 Necessary and Sufficient Conditions for Universal Recovery

First, we provide necessary and sufficient conditions for achieving universal recovery in a network $\mathcal{T}$. It turns out that these conditions are characterized by a particular set of transmission schedules $\mathcal{R}_r(\mathcal{T})$ which we define as follows:

Definition 19. *For a network $\mathcal{T} = \{\mathcal{G}, P_1, \ldots, P_n\}$, define the region $\mathcal{R}_r(\mathcal{T}) \subseteq \mathbb{N}^{n \times r}$ to be the set of all transmission schedules $\{b_i^j\}$ satisfying:*

$$\sum_{j=1}^{r} \sum_{i \in S_j^c \cap \Gamma(S_{j-1})} b_i^{(r+1-j)} \geq \left| \bigcap_{i \in S_r} P_i^c \right| \quad \textit{for each } (S_0, \ldots, S_r) \in \mathcal{S}^{(r)}(\mathcal{G}).$$

Theorem 19. *For a network $\mathcal{T}$, a transmission schedule $\{b_i^j\}$ permits universal recovery in r rounds of communication if and only if $\{b_i^j\} \in \mathcal{R}_r(\mathcal{T})$.*

Theorem 19 reveals that the set of transmission schedules permitting universal recovery is characterized precisely by the region $\mathcal{R}_r(\mathcal{T})$. In fact, given a transmission schedule in $\mathcal{R}_r(\mathcal{T})$, a corresponding coding scheme that achieves universal recovery can be computed in polynomial time using the algorithm in [JSC05] applied to the network coding graph discussed in the proof of Theorem 19. Alternatively, one could employ random linear network coding over a sufficiently large field size [HMK06]. If transmissions are made in a manner consistent with a schedule in $\mathcal{R}_r(\mathcal{T})$, universal recovery will be achieved with high probability.

Thus, the problem of achieving universal recovery with the minimum number of transmissions reduces to solving a combinatorial optimization problem over $\mathcal{R}_r(\mathcal{T})$. As this problem was shown to be NP-hard in Example 5, we do not attempt to solve it in its most general form. Instead, we apply Theorem 19 to obtain surprisingly simple characterizations for several cases of interest.

Before proceeding, we provide a quick example showing how the traditional cut-set bounds can be recovered from Theorem 19.

Example 6 (Cut-Set Bounds)**.** *Considering the constraint defining $\mathcal{R}_r(\mathcal{T})$ in which the nested subsets that form $\mathcal{S}^{(r)}(\mathcal{G})$ are all identical. That is, $(S, S, \ldots, S) \in \mathcal{S}^{(r)}(\mathcal{G})$ for some nonempty $S \subset V$. We see that any transmission schedule $\{b_i^j\} \in \mathcal{R}_r(\mathcal{T})$ must satisfy the familiar cut-set bounds:*

$$\sum_{j=1}^{r} \sum_{i \in \partial(S)} b_i^j \geq \left| \bigcap_{i \in S} P_i^c \right|. \tag{3.2}$$

In words, the total number of packets that flow into the set of nodes S must be greater than or equal to the number of packets that the nodes in S are collectively missing.

3.3.2 Fully Connected Networks

When $\mathcal{T}$ is a fully connected network, the graph $\mathcal{G}$ is a complete graph on n vertices. This is perhaps one of the most practically important cases to consider. For example, in a wired computer network, clients can multicast their messages to all other terminals which are cooperatively exchanging data. In wireless networks, broadcast is a natural transmission mode. Indeed, there are protocols tailored specifically to wireless networks which support reliable network-wide broadcast capabilities (cf. [HCP10, HC10, HH10, BBC08]). It is fortunate then, that the cooperative data exchange problem can be solved in polynomial time for fully connected networks:

Theorem 20. *For a fully connected network $\mathcal{T}$, a transmission schedule requiring only $M^*(\mathcal{T})$ transmissions can be computed in polynomial time. Necessary and sufficient conditions for universal recovery in this case are given by the cut-set constraints* (3.2)*. Moreover, a single round of communication is sufficient to achieve universal recovery with $M^*(\mathcal{T})$ transmissions.*

For the fully connected network in Example 4, we remarked that only one round of transmission was required. Theorem 20 states that this trend extends to any fully connected network.

An algorithm for solving the cooperative data exchange problem for fully connected networks is presented in Appendix 3.A. We remark that the algorithm is sufficiently general that it can also solve the cooperative data exchange problem where the objective is to minimize the weighted sum of nodes' transmissions.

Although Theorem 20 applies to arbitrary sets of packets $P_1, \ldots, P_n$, it is insightful to consider the case where packets are randomly distributed in the network. In this case, the minimum number of transmissions required for universal recovery converges in probability to a simple function of the (random) sets $P_1, \ldots, P_n$.

Theorem 21. *If $\mathcal{T}$ is a fully connected network and packets are randomly distributed, then*

$$M^*(\mathcal{T}) = \left\lceil \frac{1}{n-1} \sum_{i=1}^{n} |P_i^c| \right\rceil .$$

with probability approaching 1 *as the number of packets $k \to \infty$.*

3.3.3 d-Regular Networks

Given that precise results can be obtained for fully connected networks, it is natural to ask whether these results can be extended to a larger class of networks which includes fully connected networks as a special case. In this section, we partially answer this question in the affirmative. To this end, we define d-regular networks.

Definition 20 (d-Regular Networks). *A network $\mathcal{T}$ is said to be d-regular if $\partial(i) = d$ for each $i \in V$ and $\partial(S) \geq d$ for each nonempty $S \subset V$ with $|S| \leq n - d$. In other words, a network $\mathcal{T}$ is d-regular if the associated graph $\mathcal{G}$ is d-regular and d-vertex-connected.*

Immediately, we see that the class of d-regular networks includes fully connected networks as a special case with $d = n - 1$. Further, the class of d-regular networks includes many frequently studied network topologies (e.g., cycles, grids on tori, etc.).

Unfortunately, the deterministic algorithm of Theorem 20 does not appear to extend to d-regular networks. However, a slightly weaker concentration result similar to Theorem 21 can be obtained when packets are randomly distributed. Before stating this result, consider the following Linear Program (LP) with variable vector $x \in \mathbb{R}^n$ defined for a network $\mathcal{T}$:

$$\text{minimize} \quad \sum_{i=1}^{n} x_i \tag{3.3}$$

$$\text{subject to:} \quad \sum_{i \in \partial(j)} x_i \geq \left|P_j^c\right| \quad \text{for each } j \in V. \tag{3.4}$$

Let $M_{LP}(\mathcal{T})$ denote the optimal value of this LP. Interpreting x_i as $\sum_j b_i^j$, the constraints in the LP are a subset of the cut-set constraints of (3.2) which are a subset of the necessary constraints for universal recovery given in Theorem 19. Furthermore, the integer constraints on the x_i's are relaxed. Thus $M_{LP}(\mathcal{T})$ certainly bounds $M^*(\mathcal{T})$ from below. Surprisingly, if $\mathcal{T}$ is a d-regular network and the packets are randomly distributed, $M^*(\mathcal{T})$ is very close to this lower bound with high probability:

Theorem 22. *If $\mathcal{T}$ is a d-regular network and the packets are randomly distributed, then*

$$M^*(\mathcal{T}) < M_{LP}(\mathcal{T}) + n$$

with probability approaching 1 *as the number of packets $k \to \infty$.*

We make two important observations. First, the length of the interval in which $M^*(\mathcal{T})$ is concentrated is independent of k. Hence, even though the number of packets k may be extremely large, $M^*(\mathcal{T})$ can be estimated accurately. Second, as k grows large, $M^*(\mathcal{T})$ is dominated by the *local* topology of $\mathcal{T}$. This is readily seen since the constraints defining $M_{LP}(\mathcal{T})$ correspond only to nodes' immediate neighborhoods. The importance of the local neighborhood was also seen in [RSW05] where network coding capacity for certain random networks is shown to concentrate around the expected number of nearest neighbors of the source and the terminals.

3.3.4 Large (Divisible) Packets

We now return to general networks with arbitrarily distributed packets. However, we now consider the case where packets are "large" and can be divided into several smaller pieces (e.g., packets actually correspond to large files). To formalize this, assume that each packet can be partitioned into t chunks of equal size, and transmissions can consist of a single chunk (as opposed to an entire packet). In this case, we say the packets are t-divisible. To illustrate this point more clearly, we return to Example 4, this time considering 2-divisible packets.

Example 7 (2-Divisible Packets)**.** *Let $\mathcal{T}$ be the network of Example 4 and split each packet into two halves: $p_i \to (p_i^{(1)}, p_i^{(2)})$. Denote this new network $\mathcal{T}'$ with corresponding sets of packets:*

$$P_i' = \{p_1^{(1)}, p_1^{(2)} p_2^{(1)}, p_2^{(2)}, p_3^{(1)}, p_3^{(2)}\} \backslash \{p_i^{(1)}, p_i^{(2)}\}.$$

Three chunk transmissions allow universal recovery as follows: Node 1 transmits $p_2^{(2)} \oplus p_3^{(2)}$. Node 2 transmits $p_1^{(1)} \oplus p_3^{(1)}$. Node 3 transmits $p_1^{(2)} \oplus p_2^{(1)}$. It is readily verified from (3.2) that 3 *chunk-transmissions are required to permit universal recovery. Thus, $M^*(\mathcal{T}') = 3$. Hence, if we were allowed to split the packets of Example 4 into two halves, it would suffice to transmit 3 chunks. Normalizing the number of transmissions by the number of chunks per packet, we say that universal recovery can be achieved with* 1.5 *packet transmissions.*

Motivated by this example, define $M_t^*(\mathcal{T})$ to be the minimum number of (normalized) packet-transmissions required to achieve universal recovery in the network $\mathcal{T}$ when packets are t-divisible. For the network $\mathcal{T}$ in Example 4, we saw above that $M_2^*(\mathcal{T}) = 1.5$.

It turns out, if packets are t-divisible and t is large, the cut-set bounds (3.2) are "nearly sufficient" for achieving universal recovery. To see this, let $M_{\text{cut-set}}(\mathcal{T})$ be the optimal value of the LP:

$$\text{minimize} \quad \sum_{i=1}^{n} x_i \tag{3.5}$$

$$\text{subject to:} \quad \sum_{i \in \partial(S)} x_i \geq \left| \bigcap_{i \in S} P_i^c \right| \quad \text{for each nonempty } S \subset V. \tag{3.6}$$

Clearly $M_{\text{cut-set}}(\mathcal{T}) \leq M_t^*(\mathcal{T})$ for any network $\mathcal{T}$ with t-divisible packets because the LP producing $M_{\text{cut-set}}(\mathcal{T})$ relaxes the integer constraints and is constrained only by (3.2) rather than the full set of constraints given in Theorem 19. However, there exist transmission schedules which can approach this lower bound. Stated more precisely:

Theorem 23. *For any network $\mathcal{T}$, the minimum number of (normalized) packet-transmissions required to achieve universal recovery with t-divisible packets satisfies*

$$\lim_{t\to\infty} M_t^*(\mathcal{T}) = M_{cut\text{-}set}(\mathcal{T}).$$

Precisely how large t is required to be in order to approach $M_{\text{cut-set}}(\mathcal{T})$ within a specified tolerance is not clear for general networks. However, an immediate consequence of Theorem 21 is that $t = n-1$ is sufficient to achieve this lower bound with high probability when packets are randomly distributed in a fully connected network.

Finally, we remark that it is a simple exercise to construct examples where the cut-set bounds alone are not sufficient to characterize transmission schedules permitting universal recovery when packets are not divisible (e.g., a 4-node line network with packets p_1 and p_2 at the left-most and right-most nodes, respectively). Thus, t-divisibility of packets provides the additional degrees of freedom necessary to approach the cut-set bounds more closely.

3.3.5 Remarks

One interesting consequence of our results is that splitting packets does not significantly reduce the required number of packet-transmissions for many scenarios. Indeed, at most one transmission can be saved if the network is fully connected (under any distribution of packets). If the network is d-regular, we can expect to save fewer than n transmissions if packets are randomly distributed (in fact, at most one transmission per node). It seems possible that this result could be strengthened to include arbitrary distributions of packets in d-regular networks (as opposed to randomly distributed packets), but a proof has not been found.

The limited value of dividing packets has practical ramifications since there is usually some additional communication overhead associated with dividing packets (e.g. additional headers, etc. for each transmitted chunk are required). Thus, if the packets are very large, say each packet is a video file, our results imply that entire coded packets can be transmitted without significant loss, avoiding any additional overhead incurred by dividing packets.

3.4 An Application: Secrecy Generation

In this section, we consider the setup of the cooperative data exchange problem for a fully connected network $\mathcal{T}$, but we consider a different goal. In particular, we wish to generate a secret-key among the nodes that cannot be derived by an eavesdropper privy to all of the transmissions among nodes. Also, like the nodes themselves, the eavesdropper is assumed to know the indices of the packets initially available to each node. The goal is to generate the maximum amount of "secrecy" that cannot be determined by the eavesdropper.

The theory behind secrecy generation among multiple terminals was originally established in [CN04] for a very general class of problems. Our results should be interpreted as a practical application of the theory originally developed in [CN04]. Indeed, our results and proofs are special cases of those in [CN04] which have been streamlined to deal with the scenario under consideration. The aim of the present section is to show how secrecy can be generated in a *practical* scenario. In particular, we show that it is possible to efficiently generate the maximum amount of secrecy (as established in [CN04]) among nodes in a fully connected network $\mathcal{T} = \{\mathcal{G}, P_1, \dots, P_n\}$. Moreover, we show that this is possible in the non-asymptotic regime (i.e., there are no ϵ's and we don't require the number of packets or nodes to grow arbitrarily large). Finally, we note that it is possible to generate perfect secrecy instead of ϵ-secrecy without any sacrifice.

3.4.1 Practical Secrecy Results

In this subsection, we state two results on secrecy generation. Proofs are again postponed until Section 3.5. We begin with some definitions[2]. Let $\mathbf{F}$ denote the set of all transmissions (all of which are available to the eavesdropper by definition). A function K of the packets $\{p_1, \ldots, p_k\}$ in the network is called a secret key (SK) if K is recoverable by all nodes after observing $\mathbf{F}$, and it satisfies the (perfect) secrecy condition

$$I(K; \mathbf{F}) = 0, \tag{3.7}$$

and the uniformity condition

$$\Pr\left(K = key\right) = \frac{1}{|\mathcal{K}|} \text{ for all } key \in \mathcal{K}, \tag{3.8}$$

where $\mathcal{K}$ is the alphabet of possible keys.

We define $C_{SK}(P_1, \ldots, P_n)$ to be the secret-key capacity for a particular distribution of packets. We will drop the notational dependence on $P_1, \ldots, P_n$ where it doesn't cause confusion. By this we mean that a secret-key K can be generated if and only if $\mathcal{K} = \mathbb{F}^{C_{SK}}$. In other words, the nodes can generate at most C_{SK} packets worth of secret-key. Our first result of this section is the following:

Theorem 24. *The secret-key capacity is given by:* $C_{SK}(P_1, \ldots, P_n) = k - M^*(\mathcal{T})$.

Next, consider the related problem where a subset $D \subset V$ of nodes is compromised. In this problem, the eavesdropper has access to $\mathbf{F}$ and P_i for $i \in D$. In this case, the secret-key should also be kept hidden from the nodes in D (or else the eavesdropper could also recover it). Thus, for a subset of nodes D, let $P_D = \bigcup_{i \in D} P_i$, and call K a private-key (PK) if it is a secret-key which is only recoverable by the nodes in $V \backslash D$, and also satisfies the stronger secrecy condition:

$$I(K; \mathbf{F}, P_D) = 0. \tag{3.9}$$

[2] We attempt to follow the notation of [CN04] where appropriate.

Similar to above, define $C_{PK}(P_1, \ldots, P_n, D)$ to be the private-key capacity for a particular distribution of packets and subset of nodes D. Again, we mean that a private-key K can be generated if and only if $\mathcal{K} = \mathbb{F}^{C_{PK}}$. In other words, the nodes in $V \backslash D$ can generate at most C_{PK} packets worth of private-key. Note that, since P_D is known to the eavesdropper, each node $i \in D$ can transmit its respective set of packets P_i without any loss of secrecy capacity.

Define a new network $\mathcal{T}_D = \{\mathcal{G}_D, \{P_i^{(D)}\}_{i \in V \backslash D}\}$ as follows. Let $\mathcal{G}_D$ be the complete graph on $V \backslash D$, and let $P_i^{(D)} = P_i \backslash P_D$ for each $i \in V \backslash D$. Thus, $\mathcal{T}_D$ is a fully connected network with $n - |D|$ nodes and $k - |P_D|$ packets. Our second result of this section is the following:

Theorem 25. *The private-key capacity is given by:*

$$C_{PK}(P_1, \ldots, P_n, D) = (k - |P_D|) - M^*(\mathcal{T}_D).$$

The basic idea for private-key generation is that the users in $V \backslash D$ should generate a secret-key from $\{p_1, \ldots, p_k\} \backslash P_D$.

By the definitions of the SK and PK capacities, Theorem 20 implies that it is possible to compute these capacities efficiently. Moreover, as we will see in the achievability proofs, these capacities can be achieved by performing coded cooperative data exchange amongst the nodes. Thus, the algorithm developed in Appendix 3.A combined with the algorithm in [JSC05] can be employed to efficiently solve the secrecy generation problem we consider.

We conclude this subsection with an example to illustrate the results.

Example 8. *Consider again the network of Example 4 and assume* $\mathbb{F} = \{0, 1\}$ *(i.e., each packet is a single bit). The secret-key capacity for this network is* 1 *bit. After performing universal recovery, the eavesdropper knows* p_2 *and the parity* $p_1 \oplus p_3$*. A perfect secret-key is* $K = p_1$ *(we could alternatively use* $K = p_3$*). If any of the nodes are compromised by the eavesdropper, the private-key capacity is 0.*

We remark that the secret-key in the above example can in fact be attained by all nodes using only one transmission (i.e., universal recovery is not a prerequisite for secret-key generation). However, it remains true that only one bit of secrecy can be generated.

3.5 Proofs of Main Results

3.5.1 Necessary and Sufficient Conditions for Universal Recovery

Proof of Theorem 19. This proof is accomplished by reducing the problem at hand to an instance of a single-source network coding problem and invoking the Max-Flow Min-Cut Theorem for network information flow [ACR00].

First, fix the number of communication rounds r to be large enough to permit universal recovery. For a network $\mathcal{T}$, construct the network-coding graph $\mathcal{G}^{NC} = (V_{NC}, E_{NC})$ as follows. The vertex set, V_{NC} is defined as:

$$V_{NC} = \{s, u_1, \dots, u_k\} \cup \bigcup_{j=0}^{r} \{v_1^j, \dots, v_n^j\} \cup \bigcup_{j=1}^{r} \{w_1^j, \dots, w_n^j\}.$$

The edge set, E_{NC}, consists of directed edges and is constructed as follows:

- For each $i \in [k]$, there is an edge of unit capacity[3] from s to u_i.
- If $p_i \in P_j$, then there is an edge of infinite capacity from u_i to v_j^0.
- For each $j \in [r]$ and each $i \in [n]$, there is an edge of infinite capacity from v_i^{j-1} to v_i^j.
- For each $j \in [r]$ and each $i \in [n]$, there is an edge of capacity b_i^j from v_i^{j-1} to w_i^j.
- For each $j \in [r]$ and each $i \in [n]$, there is an edge of infinite capacity from w_i^j to $v_{i'}^j$ iff $i' \in \Gamma(i)$.

The interpretation of this graph is as follows: the vertex u_i is introduced to represent packet p_i, the vertex v_i^j represents node i after the j^{th} round of communication, and the

[3] An edge of unit capacity can carry one field element $z \in \mathbb{F}$ per unit time.

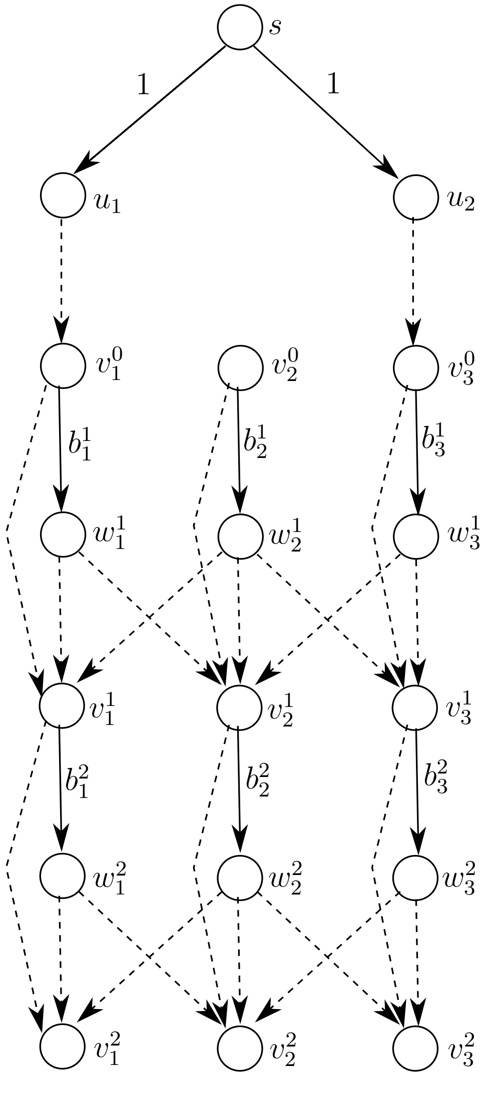

Figure 3.4: The graph $\mathcal{G}^{NC}$ corresponding to the line network of Example 3. Edges represented by broken lines have infinite capacity. Edges with finite capacities are labeled with the corresponding capacity value.

vertex w_i^j represents the broadcast of node i during the j^{th} round of communication. If the b_i^j's are chosen such that the graph $\mathcal{G}^{NC}$ admits a network coding solution which supports a multicast of k units from s to $\{v_1^r, \ldots, v_n^r\}$, then this network coding solution also solves the universal recovery problem for the network $\mathcal{T}$ when node i is allowed to make at most b_i^j transmissions during the j^{th} round of communication. The graph $\mathcal{G}^{NC}$ corresponding to the line network of Example 3 is given in Figure 3.4.

We now formally prove the equivalence of the network coding problem on $\mathcal{G}^{NC}$ and the universal recovery problem defined by $\mathcal{T}$.

Suppose a set of encoding functions $\{f_i^j\}$ and a set decoding functions $\{\phi_i\}$ describe a transmission strategy which solves the universal recovery problem for a network $\mathcal{T}$ in r rounds of communication. Let b_i^j be the number of transmissions made by node i during the j^{th} round of communication, and let $\mathcal{I}_i^j$ be all the information known to node i prior to the j^{th} round of communication (e.g. $\mathcal{I}_i^1 = P_i$). The function f_i^j is the encoding function for user i during the j^{th} round of communication (i.e. $f_i^j(\mathcal{I}_i^j) \in \mathbb{F}^{b_i^j}$), and the decoding functions satisfy:

$$\phi_i\left(\mathcal{I}_i^r, \cup_{i' \in \Gamma(i)} \{f_{i'}^r(\mathcal{I}_{i'}^r)\}\right) = \{p_1, \ldots, p_k\}.$$

Note that, given the encoding functions and the P_i's, the $\mathcal{I}_i^j$'s can be defined recursively as:

$$\mathcal{I}_i^{j+1} = \mathcal{I}_i^j \cup \bigcup_{i' \in \Gamma(i)} \{f_{i'}^j(\mathcal{I}_{i'}^j)\}.$$

The functions $\{f_i^j\}$ and $\{\phi_i\}$ can be used to generate a network coding solution which supports k units of flow from s to $\{v_1^r, \ldots, v_n^r\}$ on $\mathcal{G}^{NC}$ as follows:

For each vertex $v \in V_{NC}$, let $\mathrm{IN}(v)$ be whatever v receives on its incoming edges. Let g_v be the encoding function at vertex v, and $g_v(e, \mathrm{IN}(v))$ be the encoded message which vertex v sends along e (e is an outgoing edge from v).

If e is an edge of infinite capacity emanating from v, let $g_v(e, \mathrm{IN}(v)) = \mathrm{IN}(v)$.

Let s send p_i along edge (s,u_i). At this point, we have $\mathrm{IN}(v_i^0) = P_i = \mathcal{I}_i^1$. For each $i \in [n]$, let $g_{v_i^0}((v_i^0, w_i^1), \mathrm{IN}(v_i^0)) = f_i^1(\mathcal{I}_i^1)$. By a simple inductive argument, defining the encoding functions $g_{v_i^j}((v_i^j, w_i^{j+1}), \mathrm{IN}(v_i^j))$ to be equal to f_i^{j+1} yields the result that $\mathrm{IN}(v_i^r) = \left(\mathcal{I}_i^r, \cup_{i' \in \Gamma(i)}\{f_{i'}^r(\mathcal{I}_{i'}^r)\}\right)$. Hence, the decoding function ϕ_i can be used at v_i^r to allow error-free reconstruction of the k-unit flow.

The equivalence argument is completed by showing that a network coding solution which supports a k-unit multicast flow from s to $\{v_1^r, \dots, v_n^r\}$ on $\mathcal{G}^{NC}$ also solves the universal recovery problem on $\mathcal{T}$. This is argued in a similar manner as above, and is therefore omitted.

Since we have shown that the universal recovery problem on $\mathcal{T}$ is equivalent to a network coding problem on $\mathcal{G}^{NC}$, the celebrated max-flow min-cut result of Ahlswede et. al [ACR00] is applicable. In particular, a fixed vector $\{b_i^j\}$ admits a solution to the universal recovery problem where node i makes at most b_i^j transmissions during the j^{th} round of communication if and only if any cut separating s from some v_i^r in $\mathcal{G}^{NC}$ has capacity at least k.

What remains to be shown is that the inequalities defining $\mathcal{R}_r(\mathcal{T})$ are satisfied if and only if any cut separating s from some v_i^r in $\mathcal{G}^{NC}$ has capacity at least k.

To this end, suppose we have a cut (S, S^c) satisfying $s \in S^c$ and $v_i^r \in S$ for some $i \in [n]$. We will modify the cut (S, S^c) to produce a new cut (S', S'^c) with capacity less than or equal to the capacity of the original cut (S, S^c).

Define the set $S_0 \subseteq [n]$ as follows: $i \in S_0$ iff $v_i^r \in S$ (by definition of S, we have that $S_0 \neq \emptyset$).

Initially, let $S' = S$. Modify the cut (S', S'^c) as follows:

M1) If $i \in \Gamma(S_0)$, then place w_i^r into S'.

M2) If $i \notin \Gamma(S_0)$, then place w_i^r into S'^c.

Modifications M1 and M2 are justified (respectively) by J1 and J2:

J1) If $i \in \Gamma(S_0)$, then there exists an edge of infinite capacity from w_i^r to some $v_{i'}^r \in S$. Thus, moving w_i^r to $\mathcal{S}'$ (if necessary) does not increase the capacity of the cut.

J2) If $i \notin \Gamma(S_0)$, then there are no edges from w_i^r to S, hence we can move w_i^r into S'^c (if necessary) without increasing the capacity of the cut.

Modifications M1 and M2 guarantee that $w_i^r \in S'$ iff $i \in \Gamma(S_0)$. Thus, assume that (S', S'^c) satisfies this condition and further modify the cut as follows:

M3) If $i \in S_0$, then place v_i^{r-1} into S'.

M4) If $i \notin \Gamma(S_0)$, then place v_i^{r-1} into S'^c.

Modifications M3 and M4 are justified (respectively) by J3 and J4:

J3) If $i \in S_0$, then there exists an edge of infinite capacity from v_i^{r-1} to $v_i^r \in S$. Thus, moving v_i^{r-1} to $\mathcal{S}'$ (if necessary) does not increase the capacity of the cut.

J4) If $i \notin \Gamma(S_0)$, then there are no edges from v_i^{r-1} to S' (since $w_i^r \notin S'$ by assumption), hence we can move v_i^{r-1} into S'^c (if necessary) without increasing the capacity of the cut.

At this point, define the set $S_1 \subseteq [n]$ as follows: $i \in S_1$ iff $v_i^{r-1} \in S'$. Note that the modifications of S' guarantee that S_1 satisfies $S_0 \subseteq S_1 \subseteq \Gamma(S_0)$.

This procedure can be repeated for each layer of the graph resulting in a sequence of sets $\emptyset \subsetneq S_0 \subseteq \cdots \subseteq S_r \subseteq [n]$ satisfying $S_j \subseteq \Gamma(S_{j-1})$ for each $j \in [r]$.

We now perform a final modification of the cut (S', S'^c):

M5) If $p_j \in \cup_{i \in S_r} P_i$, then place u_j into S'.

M6) If $p_j \notin \cup_{i \in S_r} P_i$, then place u_j into S'^c.

Modifications M5 and M6 are justified (respectively) by J5 and J6:

J5) If $p_j \in \cup_{i \in S_r} P_i$, then there is an edge of infinite capacity from u_j to S' and moving u_j into S' (if necessary) does not increase the capacity of the cut.

J6) If $p_j \notin \cup_{i \in S_r} P_i$, then there are no edges from u_j to S', hence moving u_j (if necessary) into S'^c cannot increase the capacity of the cut.

A quick calculation shows that the modified cut (S', S'^c) has capacity greater than or equal to k iff:

$$\sum_{j=1}^{r} \sum_{i \in S_j^c \cap \Gamma(S_{j-1})} b_i^{r+1-j} \geq \left| \bigcap_{i \in S_r} P_i^c \right|. \tag{3.10}$$

Since every modification of the cut either preserved or reduced the capacity of the cut, the original cut (S, S^c) also has capacity greater than or equal to k if the above inequality is satisfied. In Figure 3.5, we illustrate a cut (S, S^c) and its modified minimal cut (S', S'^c) for the graph $\mathcal{G}^{NC}$ corresponding to the line network of Example 3.

By the equivalence of the universal recovery problem on a network $\mathcal{T}$ to the network coding problem on $\mathcal{G}^{NC}$ and the max-flow min-cut theorem for network information flow, if a transmission scheme solves the universal recovery problem on $\mathcal{T}$, then the associated b_i^j's must satisfy the constraints of the form given by (3.10). Conversely, for any set of b_i^j's which satisfy the constraints of the form given by (3.10), there exists a transmission scheme using exactly those numbers of transmissions which solves the universal recovery problem for $\mathcal{T}$. Thus the constraints of (3.10), and hence the inequalities defining $\mathcal{R}_r(\mathcal{T})$, are satisfied if and only if any cut separating s from some v_i^r in $\mathcal{G}^{NC}$ has capacity at least k.

Remark 5. *Since $\left| \bigcap_{i \in [n]} P_i^c \right| = 0$, constraints where $S_r = [n]$ are trivially satisfied. Therefore, we can restrict our attention to sequences of sets where $S_r \subsetneq [n]$.*

□

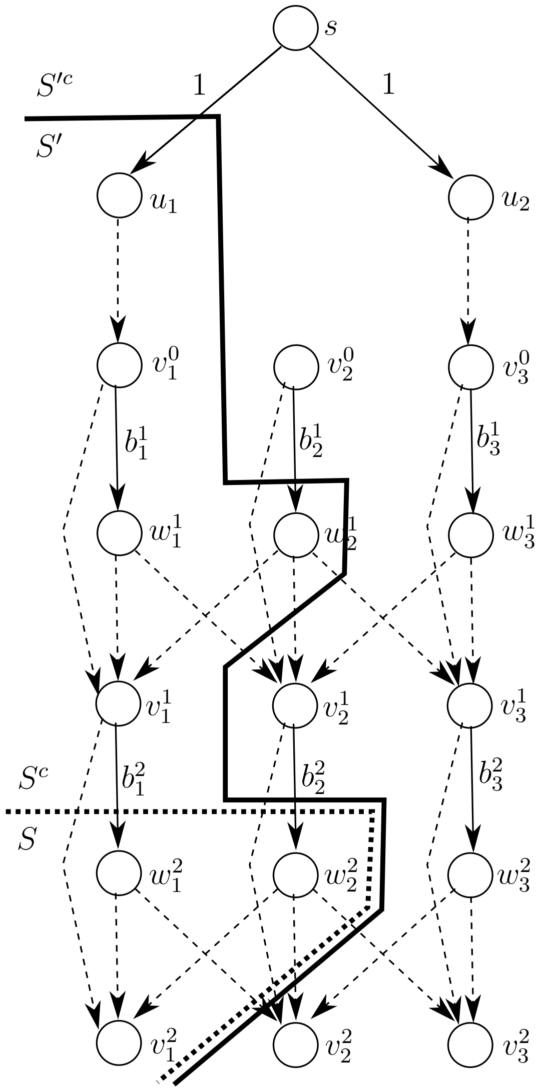

Figure 3.5: The graph $\mathcal{G}^{NC}$ corresponding to the line network of Example 3 with original cut (S, S^c) and the corresponding modified minimal cut (S', S'^c). In this case, $S_0 = S_1 = S_2 = \{1\}$. Upon substitution into (3.10), this choice of S_0, S_1, S_2 yields the inequality $b_2^1 + b_2^2 \geq 1$.

3.5.2 Fully Connected Networks

Proof of Theorem 20. In the case where $\mathcal{T}$ is a fully connected network, we have that $S_j^c \cap \Gamma(S_{j-1}) = S_j^c$ for any nonempty $S \subset V$. Therefore, the constraints defining $\mathcal{R}_r(\mathcal{T})$ become:

$$\sum_{j=1}^{r} \sum_{i \in S_j^c} b_i^{r+1-j} \geq \left| \bigcap_{i \in S_r} P_i^c \right|. \tag{3.11}$$

Now, suppose a transmission schedule $\{b_i^j\} \in \mathcal{R}_r(\mathcal{T})$ and consider the modified transmission schedule $\{\tilde{b}_i^j\}$ defined by: $\tilde{b}_i^r = \sum_{j=1}^r b_i^j$ and $\tilde{b}_i^j = 0$ for $j < r$. By construction, $S_{j+1}^c \subseteq S_j^c$ in the constraints defining $\mathcal{R}_r(\mathcal{T})$. Therefore, using the definition of $\{\tilde{b}_i^j\}$, we have:

$$\sum_{i \in S_1^c} \tilde{b}_i^r \geq \sum_{j=1}^{r} \sum_{i \in S_j^c} b_i^{r+1-j} \geq \left| \bigcap_{i \in S_r} P_i^c \right|.$$

Thus the modified transmission schedule is also in $\mathcal{R}_r(\mathcal{T})$. Since $\left|\bigcap_{i \in S_1} P_i^c\right| \geq \left|\bigcap_{i \in S_r} P_i^c\right|$, when $\mathcal{T}$ is a fully connected network, it is sufficient to consider constraints of the form:

$$\sum_{i \in S^c} b_i^1 \geq \left| \bigcap_{i \in S} P_i^c \right| \quad \text{for all nonempty } S \subset V. \tag{3.12}$$

This proves the latter two statements of the theorem: that the cut-set constraints are necessary and sufficient for universal recovery when $\mathcal{T}$ is a fully connected network, and that a single round of communication is sufficient to achieve universal recovery with $M^*(\mathcal{T})$ transmissions.

With these results established, an optimal transmission schedule can be obtained by solving the following integer linear program:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} b_i \\ \text{subject to:} \quad & \sum_{i \in S^c} b_i \geq \left| \bigcap_{i \in S} P_i^c \right| \quad \text{for each nonempty } S \subset V. \end{aligned} \tag{3.13}$$

In order to accomplish this, we identify $B_i \leftarrow P_i^c$ and set $w_i = 1$ for $i \in [n]$ and apply the submodular algorithm presented in Appendix 3.A. □

Now we consider fully connected networks in which packets are randomly distributed according to (3.1), which is parametrized by q. The proof of Theorem 21 requires the following lemma:

Lemma 10. *If $0 < q < 1$ is fixed, then there exists some $\delta > 0$ such that the following inequality holds for all $\ell \in \{2, \ldots, n-1\}$:*

$$\frac{n-\ell}{n-1} \geq \frac{(1-q)^{\ell} - (1-q)^n}{1-q-(1-q)^n} + \delta.$$

Proof. Applying Jensen's inequality to the strictly convex function $f(x) = (1-q)^x$ using the convex combination $\ell = \theta \cdot 1 + (1-\theta) \cdot n$ yields:

$$\frac{(1-q)^{\ell} - (1-q)^n}{1-q-(1-q)^n} < \frac{n-\ell}{n-1}.$$

Taking δ to be the minimum gap in the above inequality for the values $\ell \in \{2, \ldots, n-1\}$ completes the proof. □

Proof of Theorem 21. We begin by showing that the LP

$$\text{minimize} \quad \sum_{i=1}^{n} b_i \tag{3.14}$$

$$\text{subject to:} \quad \sum_{i \in S^c} b_i \geq \left| \bigcap_{i \in S} P_i^c \right| \quad \text{for each nonempty } S \subset V. \tag{3.15}$$

has an optimal value of $\frac{1}{n-1} \sum_{i=1}^{n} |P_i^c|$ with high probability. To this end, note that the inequalities

$$\sum_{\substack{i=1 \\ i \neq j}}^{n} b_i \geq |P_j^c| \quad \text{for } 1 \leq j \leq n. \tag{3.16}$$

are a subset of the inequality constraints (3.15). Summing both sides of (3.16) over $1 \leq j \leq n$ reveals that any feasible vector $b \in \mathbb{R}^n$ for LP (3.14)-(3.15) must satisfy:

$$\sum_{i=1}^{n} b_i \geq \frac{1}{n-1} \sum_{i=1}^{n} |P_i^c|. \tag{3.17}$$

This establishes a lower bound on the optimal value of the LP. We now identify a solution that is feasible with probability approaching 1 as $k \to \infty$ while achieving the lower bound of (3.17) with equality. To begin note that

$$\tilde{b}_j = \frac{1}{n-1} \sum_{i=1}^{n} |P_i^c| - |P_j^c| \tag{3.18}$$

is a solution to the system of linear equations given by (3.16) and achieves (3.17) with equality. Now, we prove that $(\tilde{b}_1, \ldots, \tilde{b}_n)$ is a feasible solution to LP (3.14) with high probability. To be specific, we must verify that

$$\sum_{i \in S^c} \tilde{b}_i \geq \left| \bigcap_{i \in S} P_i^c \right| \tag{3.19}$$

holds with high probability for all subsets $S \subset V$ satisfying $2 \leq |S| \leq n-1$ (the case $|S| = 1$ is satisfied by the definition of $\{\tilde{b}_i\}_{i=1}^n$). Substitution of (3.18) into (3.19) along with some algebra yields that the following equivalent conditions must hold:

$$\left(\frac{n-|S|}{n-1}\right) \sum_{i=1}^{n} \frac{1}{k} |P_i^c| - \sum_{i \in S^c} \frac{1}{k} |P_i^c| \geq \frac{1}{k} \left| \bigcap_{i \in S} P_i^c \right|. \tag{3.20}$$

To this end, note that for any S, $\left|\bigcap_{i \in S} P_i^c\right|$ is a random variable which can be expressed as $\left|\bigcap_{i \in S} P_i^c\right| = \sum_{j=1}^{k} X_j^S$, where X_j^S is an indicator random variable taking the value 1 if $p_j \in \bigcap_{i \in S} P_i^c$ and 0 otherwise. From (3.1) we have:

$$\Pr\left(X_j^S = 1\right) = \frac{(1-q)^{|S|} - (1-q)^n}{1-(1-q)^n}.$$

By the weak law of large numbers, for any $\eta > 0$:

$$\Pr\left(\left| \frac{1}{k} \left| \bigcap_{i \in S} P_i^c \right| - \frac{(1-q)^{|S|} - (1-q)^n}{1-(1-q)^n} \right| > \eta \right) < \epsilon_k, \tag{3.21}$$

where $\epsilon_k \to 0$ as $k \to \infty$. Thus, by the union bound, Lemma 10, and taking η sufficiently small, the following string of inequalities holds with arbitrarily high probability as $k \to \infty$:

$$\begin{aligned}
&\left(\frac{n-|S|}{n-1}\right)\sum_{i=1}^{n}\frac{1}{k}\left|P_i^c\right| - \sum_{i\in S^c}\frac{1}{k}\left|P_i^c\right| \\
&\geq \left(\frac{n-|S|}{n-1}\right)\left(\frac{(1-q)-(1-q)^n}{1-(1-q)^n} - (2n-1)\eta\right) \\
&\geq \frac{(1-q)^{|S|}-(1-q)^n}{1-(1-q)^n} + \eta \\
&\geq \frac{1}{k}\left|\bigcap_{i\in S} P_i^c\right|.
\end{aligned}$$

These steps are justified as follows: for η sufficiently small the first and last inequalities hold with high probability by (3.21), and the second inequality follows from Lemma 10 with $\ell = |S|$. This proves that (3.20) holds, and therefore $(\tilde{b}_1, \ldots, \tilde{b}_n)$ is a feasible solution to LP (3.14) with high probability. Now, taking Corollary 4 in Appendix 3.A together with Theorem 20 completes the proof. □

3.5.3 d-Regular Networks

Lemma 11. *Assume packets are randomly distributed in a d-regular network $\mathcal{T}$. For any $\epsilon > 0$, there exists an optimal solution x^* to LP (3.3-3.4) which satisfies*

$$\left\| x^* - \frac{1}{d}\mathbb{E}[|P_1^c|]\mathbb{1} \right\|_\infty < \epsilon k$$

with probability approaching 1 as $k \to \infty$, where $\mathbb{E}$ indicates expectation.

Proof. Let $\vec{P} = (|P_1^c|, \ldots, |P_n^c|)^T$ and let A be the adjacency matrix of $\mathcal{G}$ (i.e., $a_{i,j} = 1$ if $(i,j) \in E$ and 0 otherwise). Observe that A is symmetric and $A\mathbb{1} = d\mathbb{1}$, where $\mathbb{1}$ denotes a column vector of 1's. With this notation, LP (3.3) can be rewritten as:

$$\begin{aligned}
\text{minimize } \ & \mathbb{1}^T x \\
\text{subject to: } \ & Ax \succeq \vec{P},
\end{aligned} \tag{3.22}$$

where "$a \succeq b$" for vectors $a, b \in \mathbb{R}^n$ means that $a_i \geq b_i$ for $i = 1, \ldots, n$.

Let A^+ denote the Moore-Penrose pseudoinverse of A. Observe that the linear least squares solution to $Ax \approx \vec{P}$ is given by:

$$\begin{aligned} \bar{x}_{LS} &= A^+\vec{P} \\ &= A^+\mathbb{E}\vec{P} \ + A^+\left(\vec{P} - \mathbb{E}\vec{P}\right) \\ &= \frac{1}{d}\mathbb{E}\vec{P} + A^+\left(\vec{P} - \mathbb{E}\vec{P}\right). \end{aligned}$$

For the last step above, note that $\mathbb{E}\vec{P}$ is an eigenvector of A with eigenvalue d so $\mathbb{E}\vec{P}$ will also be an eigenvector of A^+ with eigenvalue $\frac{1}{d}$. Hence,

$$\begin{aligned} \|x_{LS} - \frac{1}{d}\mathbb{E}\vec{P}\|_2 &= \|A^+\left(\vec{P} - \mathbb{E}\vec{P}\right)\|_2 \\ &\le \|A^+\|_2\|\vec{P} - \mathbb{E}\vec{P}\|_2. \end{aligned}$$

Combining this with the triangle inequality implies that, for any vector y,

$$\begin{aligned} \|y - \frac{1}{d}\mathbb{E}\vec{P}\|_\infty &\le \|y - \bar{x}_{LS}\|_\infty + \|\bar{x}_{LS} - \frac{1}{d}\mathbb{E}\vec{P}\|_\infty \\ &\le \|y - \bar{x}_{LS}\|_\infty + \|\bar{x}_{LS} - \frac{1}{d}\mathbb{E}\vec{P}\|_2 \\ &\le \|y - \bar{x}_{LS}\|_\infty + \|A^+\|_2\|\vec{P} - \mathbb{E}\vec{P}\|_2. \end{aligned}$$

Therefore, Lemma 16 (see Chapter 4) guarantees the existence of an optimal solution x^* to LP (3.22) (and consequently LP (3.3)) which satisfies:

$$\begin{aligned} \|x^* - \frac{1}{d}\mathbb{E}\vec{P}\|_\infty &\le \|x^* - \bar{x}_{LS}\|_\infty + \|A^+\|_2\|\vec{P} - \mathbb{E}\vec{P}\|_2 \\ &\le c_A\|A\bar{x}_{LS} - \vec{P}\|_2 + \|A^+\|_2\|\vec{P} - \mathbb{E}\vec{P}\|_2 \\ &\le c_A\|\frac{1}{d}A\mathbb{E}\vec{P} - \vec{P}\|_2 + \|A^+\|_2\|\vec{P} - \mathbb{E}\vec{P}\|_2 \\ &= c_A\|\mathbb{E}\vec{P} - \vec{P}\|_2 + \|A^+\|_2\|\vec{P} - \mathbb{E}\vec{P}\|_2, \end{aligned}$$

where c_A is a constant depending only on A. By the weak law of large numbers, $\|\vec{P} - \mathbb{E}\vec{P}\|_2 \le \epsilon k$ with probability tending to 1 as $k \to \infty$ for any $\epsilon > 0$. Noting that $\mathbb{E}\vec{P} = \mathbb{E}[|P_1^c|]\mathbb{1}$ completes the proof. □

Proof of Theorem 22. We begin with some observations and definitions:

- First, recall that our model for randomly distributed packets (3.1) implies that

$$\mathbb{E}\left[\left|\bigcap_{i\in S} P_i^c\right|\right] = k\frac{(1-q)^{|S|}-(1-q)^n}{1-(1-q)^n} \quad \text{for all nonempty } S \subset V. \tag{3.23}$$

- With this in mind, there exists a constant $c_q > 0$ such that

$$\mathbb{E}\left[|P_1^c|\right] \geq (1+c_q)\mathbb{E}\left[\left|\bigcap_{i\in S} P_i^c\right|\right] \quad \text{for all } S \subset V, |S| \geq 2. \tag{3.24}$$

- Next, Lemma 10 implies the existence of a constant $\delta_q > 0$ such that for any $S \subset V$ with $2 \leq |S| \leq n-1$:

$$\frac{n-|S|}{n-1} \geq \frac{(1-q)^{|S|}-(1-q)^n}{(1-q)-(1-q)^n} + \delta_q = \frac{\mathbb{E}\left[\left|\bigcap_{i\in S} P_i^c\right|\right]}{\mathbb{E}\left[|P_1^c|\right]} + \delta_q. \tag{3.25}$$

- The weak law of large numbers implies that

$$\left(1+\frac{\min\{\delta_q, c_q\}}{4}\right)\mathbb{E}\left[\left|\bigcap_{i\in S} P_i^c\right|\right] \geq \left|\bigcap_{i\in S} P_i^c\right| \tag{3.26}$$

  with probability approaching 1 as $k \to \infty$.

- Finally, for the proof below, we will take the number of communication rounds sufficiently large to satisfy

$$r \geq \max\left\{\frac{2d}{n\delta_q}, \frac{2n(1+c_q)}{dc_q}\right\}. \tag{3.27}$$

Fix $\epsilon > 0$. Lemma 11 guarantees that there exists an optimal solution x^* to LP (3.3) satisfying

$$\left\| x^* - \frac{1}{d}\mathbb{E}[|P_1^c|]\mathbb{1}\right\|_\infty < \epsilon k \tag{3.28}$$

with probability tending to 1 in k. Now, it is always possible to construct a transmission schedule $\{b_i^j\}$ which satisfies $\sum_j b_i^j = \lceil x_i^* \rceil$ and $\lfloor \frac{1}{r}x_i^* \rfloor \leq b_i^j \leq \lceil \frac{1}{r}x_i^* \rceil$ for each i, j. Observe that $\sum_{i,j} b_i^j < n + \sum_i x_i^*$. Thus, proving that $\{b_i^j\} \in \mathcal{R}_r(\mathcal{T})$ with high probability will prove the theorem.

Since the network is d-regular, $|\partial(S)| \geq d$ whenever $|S| \leq n-d$ and $|\partial(S_1)| \geq n - |S_2|$ whenever $|S_2| \geq n-d$ and $S_1 \subseteq S_2$. We consider the cases where $2 \leq |S_r| \leq n-d$ and $n-d < |S_r| \leq n-1$ separately. The case where $|S_r| = 1$ coincides precisely with the constraints (3.4), and hence is satisfied by definition of $\{b_i^j\}$.

Considering the case where $2 \leq |S_r| \leq n-d$, we have the following string of inequalities:

$$\begin{aligned}
\sum_{j=1}^{r} \sum_{i\in S_j^c\cap\Gamma(S_{j-1})} b_i^{(r+1-j)} &\geq \sum_{j=1}^{r} \sum_{i\in S_j^c\cap\Gamma(S_{j-1})} \left\lfloor \frac{1}{r} x_i^* \right\rfloor && (3.29)\\
&\geq \frac{1}{r}\sum_{j=1}^{r} \sum_{i\in S_j^c\cap\Gamma(S_{j-1})} x_i^* - nr && (3.30)\\
&= \frac{1}{r}\sum_{j=1}^{r} \sum_{i\in \partial(S_{j-1})} x_i^* - \frac{1}{r}\sum_{i\in S_r\cap S_0^c} x_i^* - nr && (3.31)\\
&\geq \frac{1}{r}\sum_{j=1}^{r} \sum_{i\in \partial(S_{j-1})} \frac{1}{d}\mathbb{E}[|P_1^c|] - \frac{1}{r}\sum_{i\in S_r\cap S_0^c} \frac{1}{d}\mathbb{E}[|P_1^c|] - nk\epsilon - nr && (3.32)\\
&\geq \frac{1}{rd}\mathbb{E}[|P_1^c|]\left(\sum_{j=1}^{r} |\partial(S_{j-1})| - n\right) - nr(k\epsilon+1) && (3.33)\\
&\geq \frac{1+c_q}{rd}\mathbb{E}\left[\left|\bigcap_{i\in S_r} P_i^c\right|\right]\left(\sum_{j=1}^{r} |\partial(S_{j-1})| - n\right) - nr(k\epsilon+1) && (3.34)\\
&\geq \frac{1+c_q}{rd}\mathbb{E}\left[\left|\bigcap_{i\in S_r} P_i^c\right|\right](rd-n) - nr(k\epsilon+1) && (3.35)\\
&\geq \left(1+\frac{c_q}{2}\right)\mathbb{E}\left[\left|\bigcap_{i\in S_r} P_i^c\right|\right] - nr(k\epsilon+1) && (3.36)\\
&\geq \left(1+\frac{c_q}{4}\right)\mathbb{E}\left[\left|\bigcap_{i\in S_r} P_i^c\right|\right] && (3.37)\\
&\geq \left|\bigcap_{i\in S_r} P_i^c\right|. && (3.38)
\end{aligned}$$

The above string of inequalities holds with probability tending to 1 as $k \to \infty$. They can be justified as follows:

- (3.29) follows by definition of $\{b_i^j\}$.
- (3.30) follows since $\left\lfloor \frac{1}{r} x_i^* \right\rfloor \geq \frac{1}{r} x_i^* - 1$ and $|S_j^c \cap \Gamma(S_{j-1})| \leq n$.
- (3.31) follows from writing $\cup_{j=1}^r S_j^c \cap \Gamma(S_{j-1})$ as $\left(\cup_{j=1}^r \partial(S_{j-1})\right) \setminus (S_r \cap S_0^c)$ and expanding the sum.
- (3.32) follows from (3.28).
- (3.33) is true since $|S_0^c \cap S_r| \leq n$.
- (3.34) follows from (3.24).
- (3.35) follows from $|\partial(S_{j-1})| \geq d$ by d regularity and the assumption that $2 \leq |S_r| \leq n - d$.
- (3.36) follows from our choice of r given in (3.27).
- (3.37) follows since $\frac{c_q}{4} \mathbb{E}\left[\left|\bigcap_{i \in S_r} P_i^c\right|\right] \geq nr(k\epsilon + 1)$ with high probability for ϵ sufficiently small.
- (3.38) follows from (3.26).

Next, consider the case where $n-d \le |S_r| \le n-1$. Starting from (3.33), we obtain:

$$\begin{aligned}
\sum_{j=1}^{r} \sum_{i \in S_j^c \cap \Gamma(S_{j-1})} b_i^{(r+1-j)} &\ge \frac{1}{rd}\mathbb{E}[|P_1^c|]\left(\sum_{j=1}^{r} |\partial(S_{j-1})| - n\right) - nr(k\epsilon + 1) && (3.39)\\
&\ge \frac{1}{rd}\mathbb{E}[|P_1^c|]\Big(r(n-|S_r|) - n\Big) - nr(k\epsilon + 1) && (3.40)\\
&= \mathbb{E}[|P_1^c|]\left(\frac{n-|S_r|}{d} - \frac{n}{rd}\right) - nr(k\epsilon + 1) && (3.41)\\
&\ge \mathbb{E}[|P_1^c|]\left(\frac{n-|S_r|}{n-1} - \frac{n}{rd}\right) - nr(k\epsilon + 1) && (3.42)\\
&\ge \mathbb{E}[|P_1^c|]\left(\frac{\mathbb{E}\left[\left|\bigcap_{i \in S_r} P_i^c\right|\right]}{\mathbb{E}[|P_1^c|]} + \delta_q - \frac{n}{rd}\right) - nr(k\epsilon + 1) && (3.43)\\
&\ge \mathbb{E}[|P_1^c|]\left(\frac{\mathbb{E}\left[\left|\bigcap_{i \in S_r} P_i^c\right|\right]}{\mathbb{E}[|P_1^c|]} + \frac{\delta_q}{2}\right) - nr(k\epsilon + 1) && (3.44)\\
&\ge \left(1 + \frac{\delta_q}{4}\right)\mathbb{E}\left[\left|\bigcap_{i \in S_r} P_i^c\right|\right] && (3.45)\\
&\ge \left|\bigcap_{i \in S_r} P_i^c\right|. && (3.46)
\end{aligned}$$

The above string of inequalities holds with probability tending to 1 as $k \to \infty$. They can be justified as follows:

- (3.39) is simply (3.33) repeated for convenience.
- (3.40) follows since $n-d \le |S_r| \le n-1$ and hence d-regularity implies that $|\partial(S_{j-1})| \ge (n-|S_r|)$.
- (3.42) follows since $d \le n-1$.
- (3.43) follows from (3.25).
- (3.44) follows from from our definition of r given in (3.27).
- (3.45) follows since $\frac{\delta_q}{4}\mathbb{E}[P_1^c] \ge nr(k\epsilon+1)$ with high probability for ϵ sufficiently small.
- (3.46) follows from (3.26).

Thus, we conclude that, for ϵ sufficiently small, the transmission schedule $\{b_i^j\}$ satisfies each of the inequalities defining $\mathcal{R}_r(\mathcal{T})$ with probability tending to 1. Since the number of such inequalities is finite, an application of the union bound completes the proof that $\{b_i^j\} \in \mathcal{R}_r(\mathcal{T})$ with probability tending to 1 as $k \to \infty$. □

3.5.4 Divisible Packets

Proof of Theorem 23. Fix any $\epsilon > 0$ and let x^* be an optimal solution to LP (3.5). Put $b_i = x_i^* + \epsilon$. Note that b_i is nonnegative. This follows by considering the set $S \backslash \{i\}$ in the inequality constraint (3.6), which implies $x_i^* \geq 0$.

Now, take an integer $r \geq \epsilon^{-1} n \max_{1\leq i \leq n} b_i$. If packets are t-divisible, we can find a transmission schedule $\{b_i^j\}$ such that $\frac{1}{r} b_i \leq b_i^j \leq \frac{1}{r} b_i + \frac{1}{t}$ for all $i \in [n], j \in [r]$.

Thus, for any $(S_0, \cdots, S_r) \in \mathcal{S}^{(r)}(\mathcal{G})$ we have the following string of inequalities:

$$\begin{aligned}
\sum_{j=1}^{r} \sum_{i \in S_j^c \cap \Gamma(S_{j-1})} b_i^{(r+1-j)} &\geq \frac{1}{r} \sum_{j=1}^{r} \sum_{i \in S_j^c \cap \Gamma(S_{j-1})} b_i \\
&= \frac{1}{r} \sum_{j=1}^{r} \sum_{i \in \partial(S_{j-1})} b_i - \frac{1}{r} \sum_{i \in S_0^c \cap S_r} b_i \\
&= \frac{1}{r} \sum_{j=1}^{r} \sum_{i \in \partial(S_{j-1})} x_i^* + \frac{\epsilon}{r} \sum_{j=1}^{r} |\partial(S_{j-1})| - \frac{1}{r} \sum_{i \in S_0^c \cap S_r} b_i \\
&\geq \frac{1}{r} \sum_{j=1}^{r} \left| \bigcap_{i \in S_{j-1}} P_j^c \right| + \epsilon - \frac{n}{r} \max_{1 \leq i \leq n} b_i \\
&\geq \left| \bigcap_{i \in S_r} P_j^c \right|.
\end{aligned}$$

Hence, Theorem 19 implies that the transmission schedule $\{b_i^j\}$ is sufficient to achieve universal recovery. Noting that

$$\sum_{i,j} b_i^j \leq \sum_{i=1}^{n} b_i + \frac{nr}{t} \leq \sum_{i=1}^{n} x_i^* + n \left(\frac{r}{t} + \epsilon \right)$$

completes the proof of the theorem. □

3.5.5 Secrecy Generation

In this subsection, we prove Theorems 24 and 25. We again remark that our proofs can be seen as special cases of those in [CN04] which have been adapted for the problem at hand. For notational convenience, define $P = \{p_1, \ldots, p_k\}$. We will require the following lemma.

Lemma 12. *Given a packet distribution $P_1, \ldots, P_n$, let K be a secret-key achievable with communication $\mathbf{F}$. Then the following holds:*

$$H(K|\mathbf{F}) = H(P) - \sum_{i=1}^{n} x_i. \tag{3.47}$$

for some vector $x = (x_1, \ldots, x_n)$ which is feasible for the following ILP:

$$\textit{minimize} \quad \sum_{i=1}^{n} x_i \tag{3.48}$$

$$\textit{subject to:} \quad \sum_{i \in S} x_i \geq \left| \bigcap_{i \in S^c} P_i^c \right| \quad \textit{for each nonempty } S \subset V. \tag{3.49}$$

Moreover, if K is a PK (with respect to a set D) and each node $i \in D$ transmits its respective set of packets P_i, then

$$H(K|\mathbf{F}) = H(P|P_D) - \sum_{i \in V \backslash D} x_i. \tag{3.50}$$

for some vector $x = (x_1, \ldots, x_n)$ which is feasible for the ILP:

$$\textit{minimize} \quad \sum_{i \in V \backslash D} x_i \tag{3.51}$$

$$\textit{subject to:} \quad \sum_{i \in S} x_i \geq \left| \bigcap_{i \in S^c} P_i^c \right| \quad \textit{for each nonempty } S \subset V \backslash D. \tag{3.52}$$

Remark 6. *We remark that* (3.49) *and* (3.52) *are necessary and sufficient conditions for achieving universal recovery in the networks $\mathcal{T}$ and $\mathcal{T}_D$ considered in Theorems 24 and 25, respectively. Thus, the optimal values of ILPs* (3.48) *and* (3.51) *are equal to $M^*(\mathcal{T})$ and $M^*(\mathcal{T}_D)$, respectively.*

Proof. We assume throughout that all entropies are with respect to the base-$|\mathbb{F}|$ logarithm (i.e., information is measured in packets). For this and the following proofs, let $\mathbf{F} = (F_1, \ldots, F_n)$ and $F_{[1,i]} = (F_1, \ldots, F_i)$, where F_i denotes the transmissions made by node i. For simplicity, our proof does not take into account interactive communication, but can be modified to do so. Allowing interactive communication does not change the results. See [CN04] for details.

Since K and $\mathbf{F}$ are functions of P:

$$\begin{aligned} H(P) &= H(\mathbf{F}, K, P_1, \ldots, P_n) && (3.53) \\ &= \sum_{i=1}^{n} H(F_i|F_{[1,i-1]}) + H(K|\mathbf{F}) + \sum_{i=1}^{n} H(P_i|\mathbf{F}, K, P_{[1,i-1]}). && (3.54) \end{aligned}$$

Set $x_i = H(F_i|F_{[1,i-1]}) + H(P_i|\mathbf{F}, K, P_{[1,i-1]})$. Then, the substituting x_i into the above equation yields:

$$H(K|\mathbf{F}) = H(P) - \sum_{i=1}^{n} x_i. \qquad (3.55)$$

To show that $x = (x_1, \ldots, x_n)$ is a feasible vector for ILP (3.48), we write:

$$\begin{aligned} \left| \bigcap_{i \in S^c} P_i^c \right| &= H(P_S|P_{S^c}) && (3.56) \\ &= H(\mathbf{F}, K, P_S|P_{S^c}) && (3.57) \\ &= \sum_{i=1}^{n} H(F_i|F_{[1,i-1]}, P_{S^c}) + H(K|\mathbf{F}, P_{S^c}) + \sum_{i \in S} H(P_i|\mathbf{F}, K, P_{[1,i-1]}, P_{S^c \cap [i+1,n]}) && (3.58) \\ &\leq \sum_{i \in S} H(F_i|F_{[1,i-1]}) + \sum_{i \in S} H(P_i|\mathbf{F}, K, P_{[1,i-1]}) && (3.59) \\ &= \sum_{i \in S} x_i. && (3.60) \end{aligned}$$

In the above inequality, we used the fact that conditioning reduces entropy, the fact that K is a function of $(\mathbf{F}, P_{S^c})$ for any $S \neq V$, and the fact that F_i is a function of P_i (by the assumption that communication is not interactive).

To prove the second part of the lemma, we can assume $D = \{1, \ldots, \ell\}$. The assumption that each node i in D transmits all of the packets in P_i implies $F_i = P_i$. Thus, for $i \in D$ we have $x_i = H(P_i | P_{[1,i-1]})$. Repeating the above argument, we obtain

$$H(K|\mathbf{F}) = H(P) - H(P_D) - \sum_{i \in V \setminus D} x_i \tag{3.61}$$

$$= H(P|P_D) - \sum_{i \in V \setminus D} x_i, \tag{3.62}$$

completing the proof of the lemma. □

Proof of Theorem 24. Converse Part. Suppose K is a secret-key achievable with communication $\mathbf{F}$. Then, by definition of a SK and Lemma 12 we have

$$C_{SK} = H(K) = H(K|\mathbf{F}) = H(P) - \sum_{i=1}^{n} x_i \leq H(P) - M^*(\mathcal{T}) = k - M^*(\mathcal{T}). \tag{3.63}$$

Achievability Part. By definition, universal recovery can be achieved with $M^*(\mathcal{T})$ transmissions. Moreover, the communication $\mathbf{F}$ can be generated as a linear function of P (see the proof of Theorem 19 and [JSC05]). Denote this linear transformation by $\mathbf{F} = \mathcal{L}P$. Note that $\mathcal{L}$ only depends on the indices of the packets available to each node, not the values of the packets themselves (see [JSC05]). Let $\mathcal{P}_{\mathbf{F}} = \{P' : \mathcal{L}P' = \mathbf{F}\}$ be the set of all packet distributions which generate $\mathbf{F}$.

By our assumption that the packets are i.i.d. uniform from $\mathbb{F}$, each $P' \in \mathcal{P}_{\mathbf{F}}$ is equally likely given $\mathbf{F}$ was observed. Since $\mathbf{F}$ has dimension $M^*(\mathcal{T})$, $|\mathcal{P}_{\mathbf{F}}| = \mathbb{F}^{k-M^*(\mathcal{T})}$. Thus, we can set $\mathcal{K} = \mathbb{F}^{k-M^*(\mathcal{T})}$ and label each $P' \in \mathcal{P}_{\mathbf{F}}$ with a unique element in $\mathcal{K}$. The label for the actual P (which is reconstructed by all nodes after observing $\mathbf{F}$) is the secret-key. Thus, $C_{SK} \geq k - M^*(\mathcal{T})$.

We remark that this labeling can be done efficiently by an appropriate linear transformation mapping P to K. □

Proof of Theorem 25. Converse Part. Suppose K is a private-key. Then, by definition of a PK and Lemma 12,

$$\begin{aligned} C_{PK} = H(K) = H(K|\mathbf{F}) = H(P|P_D) - \sum_{i \in V \backslash D} x_i \\ \leq H(P|P_D) - M^*(\mathcal{T}_D) = (k - |P_D|) - M^*(\mathcal{T}_D). \end{aligned}$$

Achievability Part. Let each node $i \in D$ transmit P_i so that we can update $P_j \leftarrow P_j \cup P_D$ for each $j \in V \backslash D$. Now, consider the universal recovery problem for only the nodes in $V \backslash D$. $M^*(\mathcal{T}_D)$ is the minimum number of transmissions required among the nodes in $V \backslash D$ so that each node in $V \backslash D$ recovers P. At this point, the achievability proof proceeds identically to the SK case. □

3.6 Concluding Remarks

In this chapter, we derive necessary and sufficient conditions for achieving universal recovery in an arbitrarily connected network. For the case when the network is fully connected, we provide an efficient algorithm based on submodular optimization which efficiently solves the cooperative problem. This algorithm and its derivation yield tight concentration results for the case when packets are randomly distributed. Moreover, concentration results are provided when the network is d-regular and packets are distributed randomly. If packets are divisible, we prove that the traditional cut-set bounds are achievable. As a consequence of this and the concentration results, we show that splitting packets does not typically provide a significant benefit when the network is d-regular. Finally, we discuss an application to secrecy generation in the presence of an eavesdropper. We demonstrate that our submodular algorithm can be used to generate the maximum amount of secrecy in an efficient manner.

It is conceivable that the coded cooperative data exchange problem can be solved (or approximated) in polynomial time if the network is d-regular, but packets aren't necessarily randomly distributed. This is one possible direction for future work.

Acknowledgement

The authors would like to thank Kent Benson, Alex Sprintson, Pavan Datta, and Chi-Wen Su for the helpful conversations and suggestions which led to this work. The material in this chapter is based on the works [CXW10, CW10, CW11b, CW12b].

3.A An Efficiently Solvable Integer Linear Program

In this appendix, we introduce a special ILP and provide an efficient algorithm for solving it. This algorithm can be used to efficiently solve the cooperative data exchange problem when the underlying graph is fully-connected. We begin by introducing some notation[4].

Let $E = \{1, \ldots, n\}$ be a finite set with n elements. We denote the family of all subsets of E by 2^E. We frequently use the compact notation $E \backslash U$ and $U + i$ to denote the sets $E \cap U^c$ and $U \cup \{i\}$ respectively. For a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, define the corresponding functional $x : 2^E \to \mathbb{R}$ as:

$$x(U) := \sum_{i \in U} x_i, \text{ for } U \subseteq E. \tag{3.64}$$

Throughout this section, we let $\mathcal{F} = 2^E - \{\emptyset, E\}$ denote the family of nonempty proper subsets of E. Let $\mathcal{B} = \{B_1, \ldots, B_n\}$. No special structure is assumed for the B_i's except that they are finite.

With the above notation established, we consider the following Integer Linear Program (ILP) in this section:

$$\text{minimize} \left\{ \sum_{i \in E} w_i x_i : x(U) \geq \left| \bigcap_{i \in E \backslash U} B_i \right|, \forall\, U \in \mathcal{F}, x_i \in \mathbb{Z} \right\}. \tag{3.65}$$

It is clear that any algorithm that efficiently solves this ILP also solves ILP (3.13) by putting $B_i \leftarrow P_i^c$ and $w = \mathbb{1}$.

[4] We attempt to keep the notation generic in order to emphasize that the results in this appendix are not restricted to the context of the cooperative data exchange problem.

3.A.1 Submodular Optimization

Our algorithm for solving ILP (3.65) relies heavily on submodular function optimization. To this end, we give a very brief introduction to submodular functions here.

A function $g : 2^E \to \mathbb{R}$ is said to be submodular if, for all $X, Y \in 2^E$,

$$g(X) + g(Y) \geq g(X \cap Y) + g(X \cup Y). \tag{3.66}$$

Over the past three decades, submodular function optimization has received a significant amount of attention. Notably, several polynomial time algorithms have been developed for solving the Submodular Function Minimization (SFM) problem

$$\min \{g(U) : U \subseteq E\}. \tag{3.67}$$

We refer the reader to [McC05, Fuj10, Sch03] for a comprehensive overview of SFM and known algorithms. As we will demonstrate, we can solve ILP (3.65) via an algorithm that iteratively calls a SFM routine. The most notable feature of SFM algorithms is their ability to solve problems with exponentially many constraints in polynomial time. One of the key drawbacks of SFM is that the problem formulation is very specific. Namely, SFM routines typically require the function g to be submodular on *all* subsets of the set E.

3.A.2 The Algorithm

We begin by developing an algorithm to solve an equality constrained version of ILP (3.65). We will remark on the general case at the conclusion of this section. To this end, let M be a positive integer and consider the following ILP:

$$\text{minimize} \quad w^T x \tag{3.68}$$

$$\text{subject to: } x(U) \geq \left| \bigcap_{i \in E \setminus U} B_i \right| \text{ for all } U \in \mathcal{F}, \text{ and} \tag{3.69}$$

$$x(E) = M. \tag{3.70}$$

Remark 7. *We assume $w_i \geq 0$, else in the case without the equality constraint we could allow the corresponding $x_i \to +\infty$ and the problem is unbounded from below.*

```
Algorithm 3.A.1: SolveILP($\mathcal{B}, E, M, w$)

 comment: Define $f : 2^E \to \mathbb{R}$ as in equation (3.71).

 $x \leftarrow$ ComputePotentialX($f, M, w$)
 if CheckFeasible($f, x$)
   then return ($x$)
   else return (Problem Infeasible)
```

Theorem 26. *Algorithm 3.A.1 solves the equality constrained ILP* (3.68) *in polynomial time. If feasible, Algorithm 3.A.1 returns an optimal x. If infeasible, Algorithm 3.A.1 returns "Problem Infeasible".*

Proof. The proof is accomplished in three steps:

1. First, we show that if our algorithm returns an x, it is feasible.
2. Second, we prove that if a returned x is feasible, it is also optimal.
3. Finally, we show that if our algorithm does not return an x, then the problem is infeasible.

Each step is given its own subsection. □

Algorithm 3.A.1 relies on three basic subroutines given below:

```
Algorithm 3.A.2: ComputePotentialX($f, M, w$)

comment: If feasible, returns $x$ satisfying (3.69) and (3.70) that minimizes $w^T x$.

comment: Order elements of $E$ so that $w_1 \geq w_2 \geq \cdots \geq w_n$.

for $i \leftarrow n$ to 2
  do comment: Define $f_i(U) := f(U + i)$ for $U \subseteq \{i, \ldots, n\}$.
     $x_i \leftarrow \mathrm{SFM}(f_i, \{i, \ldots, n\})$
$x_1 \leftarrow M - \sum_{i=2}^{n} x_i$
return ($x$)
```

```
Algorithm 3.A.3: CheckFeasible($f, x$)

comment: Check if $x(U) \leq f(U)$ for all $U \in \mathcal{F}$ with $1 \in U$.

comment: Define $f_1(U) := f(U + 1)$ for $U \subseteq E$.

if $\mathrm{SFM}(f_1, E) < 0$
  then return ( false )
  else return ( true )
```

```
Algorithm 3.A.4: SFM($f, V$)

comment: Minimize submodular function $f$ over groundset $V$ (cf. [McC05]).

$v \leftarrow \min \{f(U) : U \subseteq V\}$
return ($v$)
```

3.A.3 Feasibility of a Returned x

In this section, we prove that if Algorithm 3.A.1 returns a vector x, it must be feasible. We begin with some definitions.

Definition 21. *A pair of sets $X, Y \subset E$ is called* ***crossing*** *if $X \cap Y \neq \emptyset$ and $X \cup Y \neq E$.*

Definition 22. *A function $g : 2^E \to \mathbb{R}$ is* ***crossing submodular*** *if*

$$g(X) + g(Y) \geq g(X \cap Y) + g(X \cup Y)$$

for X, Y crossing.

We remark that minimization of crossing submodular functions is well established, however it involves a lengthy reduction to a standard submodular optimization problem. However, the crossing family $\mathcal{F}$ admits a straightforward algorithm, which is what we provide in Algorithm 3.A.1. We refer the reader to [Sch03] for complete details on the general case.

For M a positive integer, define

$$f(U) := M - \left|\bigcap_{i \in U} B_i\right| - x(U), \text{ for } U \in \mathcal{F}. \tag{3.71}$$

Lemma 13. *The function f is crossing submodular on $\mathcal{F}$.*

Proof. For $X, Y \in \mathcal{F}$ crossing:

$$\begin{aligned}
f(X) + f(Y) &= M - \left|\bigcap_{i \in X} B_i\right| - x(X) + M - \left|\bigcap_{i \in Y} B_i\right| - x(Y) \\
&= M - \left|\bigcap_{i \in X} B_i\right| - x(X \cap Y) + M - \left|\bigcap_{i \in Y} B_i\right| - x(X \cup Y) \\
&\geq M - \left|\bigcap_{i \in X \cap Y} B_i\right| - x(X \cap Y) + M - \left|\bigcap_{i \in X \cup Y} B_i\right| - x(X \cup Y) \\
&= f(X \cap Y) + f(X \cup Y).
\end{aligned}$$

□

Observe that, with f defined as above, the constraints of ILP (3.68) can be equivalently written as:

$$f(U) = M - \left|\bigcap_{i\in U} B_i\right| - x(U) \geq 0 \text{ for all } U \in \mathcal{F}, \text{ and} \tag{3.72}$$

$$x(E) = M. \tag{3.73}$$

Without loss of generality, assume the elements of E are ordered lexicographically so that $w_1 \geq w_2 \geq \cdots \geq w_n$. At iteration i in Algorithm 3.A.2, $x_j = 0$ for all $j \leq i$. Thus, setting

$$\begin{aligned} x_i \leftarrow & \min_{U\subseteq\{i,\ldots,n\}} \{f_i(U)\} && (3.74)\\ = & \min_{U\subseteq\{i,\ldots,n\}:i\in U} \{f(U)\} && (3.75)\\ = & \min_{U\subseteq\{i,\ldots,n\}:i\in U} \left\{M - \left|\bigcap_{i\in U} B_i\right| - x(U)\right\} && (3.76)\end{aligned}$$

and noting that the returned x satisfies $x(E) = M$, rearranging (3.76) guarantees that

$$x(E\backslash U) \geq \left|\bigcap_{i\in U} B_i\right|, \text{ for all } U \subseteq \{i, \ldots, n\}, i \in U \tag{3.77}$$

as desired. Iterating through $i \in \{2, \ldots, n\}$ guarantees (3.77) holds for $2 \leq i \leq n$.

Remark 8. *In the feasibility check routine (Algorithm 3.A.3), we must be able to evaluate $f_1(E)$. The reader can verify that putting $f(E) = 0$ preserves submodularity.*

Now, in order for the feasibility check to return **true**, we must have

$$\begin{aligned} \min_{U\subseteq E} \{f_1(U)\} = & \min_{U\subseteq E:1\in U} \{f(U)\} && (3.78)\\ = & \min_{U\subseteq E:1\in U} \left\{M - \left|\bigcap_{i\in U} B_i\right| - x(U)\right\} && (3.79)\\ \geq & \ 0, && (3.80)\end{aligned}$$

implying that

$$x(E\backslash U) \geq \left|\bigcap_{i\in U} B_i\right|, \text{ for all } U \subseteq E, 1 \in U. \tag{3.81}$$

Combining (3.77) and (3.81) and noting that $x(E) = M$ proves that x is indeed feasible. Moreover, x is integral as desired.

3.A.4 Optimality of a Returned x

In this section, we prove that if Algorithm 3.A.1 returns a feasible x, then it is also optimal. First, we require two more definitions and a lemma.

Definition 23. *A constraint of the form* (3.72) *corresponding to U is said to be* ***tight*** *for U if*

$$f(U) = M - \left| \bigcap_{i \in U} B_i \right| - x(U) = 0. \tag{3.82}$$

Lemma 14. *If x is feasible, X, Y are crossing, and their corresponding constraints are tight, then the constraints corresponding to $X \cap Y$ and $X \cup Y$ are also tight.*

Proof. Since the constraints corresponding to X and Y are tight, we have

$$0 = f(X) + f(Y) \geq f(X \cap Y) + f(X \cup Y) \geq 0. \tag{3.83}$$

The first inequality is due to submodularity and the last inequality holds since x is feasible. This implies the result. □

Definition 24. *A family of sets $\mathcal{L}$ is* ***laminar*** *if $X, Y \in \mathcal{L}$ implies either $X \cap Y = \emptyset$, $X \subset Y$, or $Y \subset X$.*

At iteration k ($1 < k \leq n$) of Algorithm 3.A.2, let U_k be the set where (3.76) achieves its minimum. Note that $k \in U_k \subseteq \{k, \ldots, n\}$. By construction, the constraint corresponding to U_k is tight. Also, the constraint $x(E) = M$ is tight. From the U_k's and E we can construct a laminar family as follows: if $U_j \cap U_k \neq \emptyset$ for $j < k$, then replace U_j with $\tilde{U}_j \leftarrow U_k \cup U_j$. By Lemma 14, the constraints corresponding to the sets in the newly constructed laminar family are tight. Call this family $\mathcal{L}$. For each $i \in E$, there is a unique smallest set in $\mathcal{L}$ containing i. Denote this set L_i. Since $k \in U_k \subseteq \{k, \ldots, n\}$, $L_i \neq L_j$ for $i \neq j$. Note that $L_1 = E$ and $L_i \subset L_j$ only if $j < i$.

For each $L_i \in \mathcal{L}$ there is a unique smallest set L_j such that $L_i \subset L_j$. We call L_j the least upper bound on L_i.

Now, consider the dual linear program to (3.68):

$$\text{maximize} \quad -\sum_{U\in\mathcal{F}} \pi_U \left(M - \left| \bigcap_{i\in U} B_i \right| \right) - \pi_E M \tag{3.84}$$

$$\text{subject to:} \quad \sum_{U\in\mathcal{F}:i\in U} \pi_U + \pi_E + w_i = 0, \text{ for } 1 \leq i \leq n \tag{3.85}$$

$$\pi_U \geq 0 \text{ for } U \in \mathcal{F}, \text{ and } \pi_E \text{ free.} \tag{3.86}$$

For each $L_i \in \mathcal{L}$, let the corresponding dual variable $\pi_{L_i} = w_j - w_i$, where L_j is the least upper bound on L_i. By construction, $\pi_{L_i} \geq 0$ since it was assumed that $w_1 \geq \cdots \geq w_n$. Finally, let $\pi_E = -w_1$ and $\pi_U = 0$ for $U \notin \mathcal{L}$.

Now, observe that:

$$\sum_{U\in\mathcal{F}:i\in U} \pi_U + \pi_E + w_i = 0 \tag{3.87}$$

as desired for each i. Thus, π is dual feasible. Finally, note that $\pi_U > 0$ only if $U \in \mathcal{L}$. However, the primal constraints corresponding to the sets in $\mathcal{L}$ are tight. Thus, (x, π) form a primal-dual feasible pair satisfying complementary slackness conditions, and are therefore optimal.

3.A.5 No Returned x = Infeasibility

Finally, we prove that if the feasibility check returns **false**, then ILP (3.68) is infeasible. Note by construction that the vector x passed to the feasibility check satisfies

$$M - \left| \bigcap_{i\in U} B_i \right| - x(U) \geq 0 \text{ for all nonempty } U \subseteq \{2, \ldots, n\}, \tag{3.88}$$

and $x(E) = M$. Again, let U_k be the set where (3.76) achieves its minimum and let $\mathcal{L}$ be the laminar family generated by these U_k's and E exactly as before. Again, the constraints corresponding to the sets in $\mathcal{L}$ are tight (this can be verified in a manner identical to the proof of Lemma 14). Now, since x failed the feasibilty check, there exists some exceptional set T with $1 \in T$ for which

$$M - \left|\bigcap_{i\in T} B_i\right| - x(T) < 0. \tag{3.89}$$

Generate a set L_T as follows: Initialize $L_T \leftarrow T$. For each $L_i \in \mathcal{L}, L_i \neq E$, if $L_T \cap L_i \neq \emptyset$, update $L_T \leftarrow L_T \cup L_i$. Now, we can add L_T to family $\mathcal{L}$ while preserving the laminar property. We pause to make two observations:

1. By an argument similar to the proof of Lemma 14, we have that
$$M - \left|\bigcap_{i\in L_T} B_i\right| - x(L_T) < 0.$$

2. The sets in $\mathcal{L}$ whose least upper bound is E form a partition of E. We note that L_T is a nonempty class of this partition. Call this partition $\mathcal{P}_{\mathcal{L}}$.

Again consider the dual constraints, however, let $w_i = 0$ (this does not affect feasibility). For each $L \in \mathcal{P}_{\mathcal{L}}$ define the associated dual variable $\pi_L = \alpha$, and let $\pi_E = -\alpha$. All other dual variables are set to zero. It is easy to check that this π is dual feasible. Now, the dual objective function becomes:

$$-\sum_{U\in\mathcal{F}} \pi_U \left((M - \left|\bigcap_{i\in U} B_i\right|\right) - \pi_E M = -\alpha \sum_{L\in\mathcal{P}_{\mathcal{L}}} \left(M - \left|\bigcap_{i\in L} B_i\right| - x(L) + x(L)\right) + \alpha M \tag{3.90}$$

$$= -\alpha\left(M - \left|\bigcap_{i\in L_T} B_i\right| - x(L_T)\right) - \alpha x(E) + \alpha M \tag{3.91}$$

$$= -\alpha\left(M - \left|\bigcap_{i\in L_T} B_i\right| - x(L_T)\right) \tag{3.92}$$

$$\to +\infty \text{ as } \alpha \to \infty. \tag{3.93}$$

Thus, the dual is unbounded and therefore the primal problem must be infeasible.

As an immediate corollary we obtain the following:

Corollary 4. *The optimal values of the ILP:*

$$\min\left\{x(E) : x(U) \geq \left|\cap_{i\in E\setminus U} B_i\right|, U \in \mathcal{F}, x_i \in \mathbb{Z}\right\}$$

and the corresponding LP relaxation:

$$\min\left\{x(E) : x(U) \geq \left|\cap_{i\in E\setminus U} B_i\right|, U \in \mathcal{F}, x_i \in \mathbb{R}\right\}$$

differ by less than 1.

Proof. Algorithm 3.A.1 is guaranteed to return an optimal x if the intersection of the polytope and the hyperplane $x(E) = M$ is nonempty. Thus, if M^* is the minimum such M, then the optimal value of the LP must be greater than $M^* - 1$. □

3.A.6 Solving the General ILP

Finally, we remark on how to solve the general case of the ILP without the equality constraint given in (3.65). First, we state a simple convexity result.

Lemma 15. *Let $p_w^*(M)$ denote the optimal value of ILP* (3.68) *when the equality constraint is $x(E) = M$. We claim that $p_w^*(M)$ is a convex function of M.*

Proof. Let M_1 and M_2 be integers and let $\theta \in [0, 1]$ be such that $M_\theta = \theta M_1 + (1-\theta)M_2$ is an integer. Let $x^{(1)}$ and be $x^{(2)}$ optimal vectors that attain $p_w^*(M_1)$ and $p_w^*(M_2)$ respectively. Let $x^{(\theta)} = \theta x^{(1)} + (1-\theta)x^{(2)}$. By convexity, $x^{(\theta)}$ is feasible, though not necessarily integer. However, by the results from above, optimality is always attained by an integral vector. Thus, it follows that:

$$\theta p_w^*(M_1) + (1-\theta)p_w^*(M_2) = \theta w^T x^{(1)} + (1-\theta)w^T x^{(2)} = w^T x^{(\theta)} \geq p_w^*(M_\theta). \tag{3.94}$$

□

Noting that $p_w^*(M)$ is convex in M, we can perform bisection on M to solve the ILP in the general case. For our purposes, it suffices to have relatively loose upper and lower bounds on M since the complexity only grows logarithmically in the difference. A simple lower bound on M is given by $M \geq \max_i |B_i|$.

3.A.7 Complexity

Our aim here is not to give a detailed complexity analysis of our algorithm. This is due to the fact that the complexity is dominated by the the SFM over the set E in Algorithm 3.A.3. Therefore, the complexity of Algorithm 3.A.1 is essentially the same as the complexity of the SFM solver employed.

However, we have performed a series of numerical experiments to demonstrate that Algorithm 3.A.1 performs quite well in practice. In our implementation, we ran the Fujishige-Wolfe (FW) algorithm for SFM [FHI06] based largely on a Matlab routine by A. Krause [KS10]. While the FW algorithm has not been proven to run in polynomial time, it has been shown to work quite well in practice [FHI06] (similar to the Simplex algorithm for solving Linear Programs). Whether or not FW has worst-case polynomial complexity is an open problem to date. We remark that there are several SFM algorithms that run in strongly polynomial time which could be used if a particular application requires polynomially bounded worst-case complexity [McC05].

In our series of experiments, we chose $B_i \subset F$ randomly, where $|F| = 50$. We let $n = |E|$ range from 10 to 190 in increments of 10. For each value of n, we ran 10 experiments. The average computation time is shown in Figure 3.6, with error bars indicating one standard deviation. We consistently observed that the computations run in approximately $O(n^{1.85})$ time. Due to the iterative nature of the SFM algorithm, we anticipate that the computation time could be significantly reduced by implementing the algorithm in C/C++ instead of Matlab. However, the $O(n^{1.85})$ trend should remain the same. Regardless, we are able to solve the ILP problems under consideration with an astonishing 2^{190} constraints in approximately one minute.

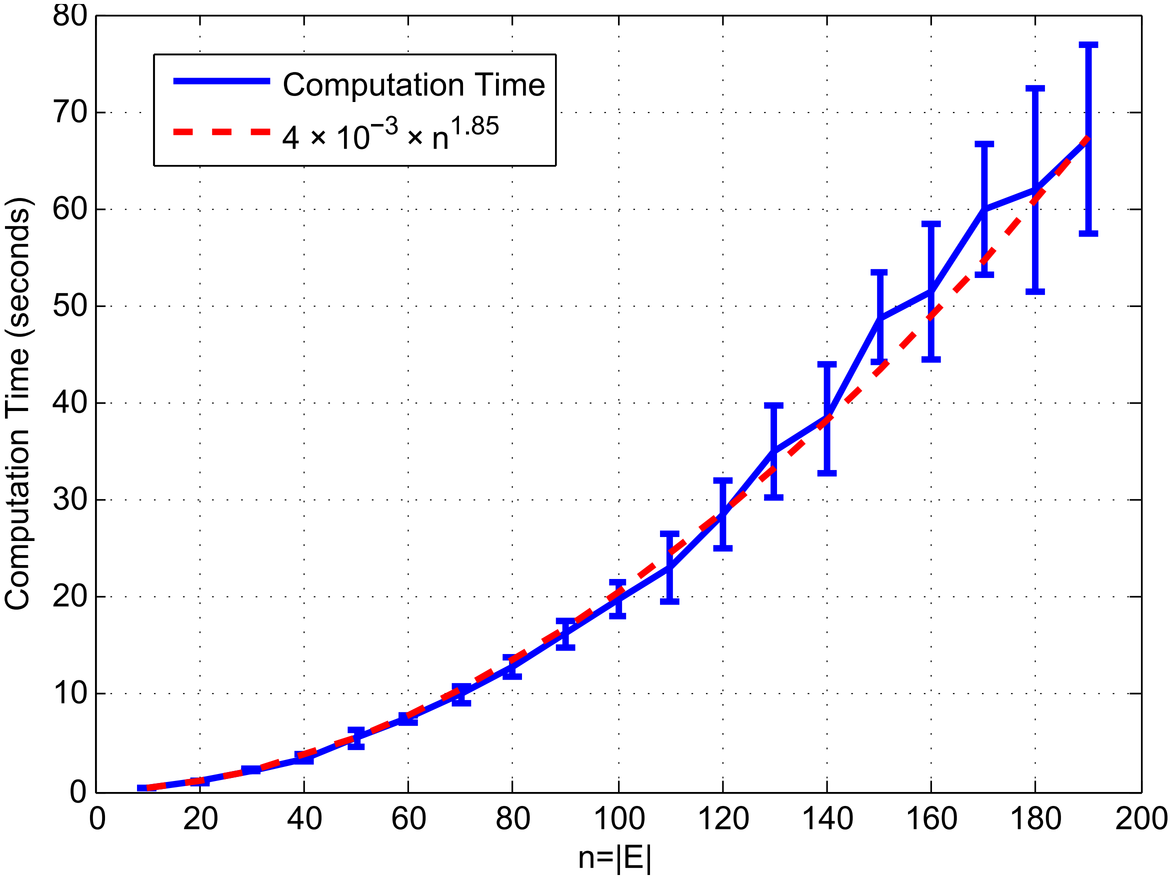

Figure 3.6: Experimental results. For the red dotted line, the multiplicative constant α and exponent β were chosen to minimize the MSE $\sum_{i=1}^{n} |\log(\alpha n^{\beta}) - \log(\hat{m}_n)|^2$, where $\hat{m}_n$ is the sample mean of the computation times for $|E| = n$.

CHAPTER 4

Three Interesting Lemmas

4.1 Introduction

I have always admired results that can be simply stated and appreciated, regardless of their potential applications. In this chapter, I have compiled three such results which I've developed over the course of my research. For one reason or another, I find each of the lemmas presented herein interesting. Therefore, I feel they merit their own chapter instead of being relegated to an appendix, hidden amongst the gritty details of this dissertation.

Fitting with the common theme, each of the results presented in this chapter has an application to multiterminal information theory. Indeed, the first two results are crucial tools employed in Chapters 2 and 3. The final result of this chapter does not appear elsewhere in this dissertation. However, it has become a favorite problem of mine and I couldn't resist the temptation to include it.

Each section of this chapter is self-contained and follows a common format. Specifically, each section begins with a statement of the lemma, a brief discussion ensues, and a proof concludes.

4.2 A Linear Programming Approximation

Lemma 16. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with nonnegative entries and all column sums equal to d. Let $\bar{x}_y$ be the vector of minimum Euclidean norm which minimizes $\|Ax_y-$*

$y\|_2$. There exists an optimal solution x^* to the linear program

$$\begin{aligned} &\textit{minimize} \;\; \mathbb{1}^T x \\ &\textit{subject to:} \;\; Ax \succeq y \end{aligned} \tag{4.1}$$

which satisfies

$$\|x^* - \bar{x}_y\|_\infty \leq c_A \|A\bar{x}_y - y\|_2,$$

where c_A is a constant depending only on A.

4.2.1 Remarks

The lemma is trivial if the matrix A is nonsingular. However, the case where A is singular is interesting to consider. In this case, any nonzero vector v in the nullspace of A must satisfy $\mathbb{1}^T v = 0$ since $\mathbb{1}$ is an eigenvector of A corresponding to eigenvalue d. Thus, if x^* is an optimal solution to LP (4.1), then so is $x^* + v$. Hence, the set of optimal solutions is unbounded in the sense that for any $B > 0$, there exists an optimal solution to LP (4.1) with entries greater than B and other entries less than $-B$. Thus, Lemma 16 is useful because it always guarantees the existence of an optimal solution that is well-behaved in the sense that it is quantifiably close to the linear least-squares solution to $Ax \approx y$.

In my case, the motivating application was where $y = \mathbb{1} + \varepsilon$, $d = 1$, and the vector ε is a small perturbation (see Chapter 3 for details). In this case, it is straightforward to show that there exists an optimal solution x^* to LP (4.1) satisfying

$$\|x^* - \mathbb{1}\|_\infty \leq \delta(\|\varepsilon\|),$$

where $\delta(\|\varepsilon\|) \to 0$ as $\|\varepsilon\| \to 0$. Although this seems intuitive, if ε is not in the range of A, then the previous remark demonstrates that there always exists optimal solutions satisfying $\|x^* - \mathbb{1}\|_\infty > B$ for any $B > 0$. Thus, out of infinitely many optimal solutions to LP (4.1), Lemma 16 allows us to choose one which is suitably well behaved.

4.2.2 Proof

Proof of Lemma 16. To begin the proof, we make a few definitions. Let λ be the absolute value of the nonzero eigenvalue of A with smallest modulus (at least one exists since d is an eigenvalue). Define $\mathcal{N}(A)$ to be the nullspace of A, and let $\mathcal{N}^{\perp}(A)$ denote its orthogonal complement. Finally, let A^{+} denote the Moore-Penrose pseudoinverse of A (see [Lau05]).

Fix $\bar{x}_y \in \mathbb{R}^n$, and note that x^* is an optimal solution to LP (4.1) if and only if $x^* - \bar{x}_y$ is an optimal solution to the linear program

$$\begin{aligned} \text{minimize} \;\; & \mathbb{1}^T(x + \bar{x}_y) \\ \text{subject to:} \;\; & A(x + \bar{x}_y) \succeq y \end{aligned}$$

with variable $x \in \mathbb{R}^n$. With this in mind, put $\bar{x}_y = A^{+}y$ and define $b = y - A\bar{x}_y$. By definition of the pseudoinverse, $\bar{x}_y$ is the vector of minimum Euclidean norm which minimizes $\|Ax_y - y\|_2$. Moreover, $b \in \mathcal{N}(A)$.

Thus, in order to prove the lemma, it suffices to show the existence of an optimal solution x^* to the linear program

$$\begin{aligned} \text{minimize} \;\; & \mathbb{1}^T x \\ \text{subject to:} \;\; & Ax \succeq b \end{aligned} \tag{4.2}$$

which also satisfies the additional constraints

$$|x_i| \leq c_A \|b\|_2 \;\; \text{for } i = 1, \ldots, n,$$

where c_A is a constant depending only on A.

Claim 1. *There exists an optimal solution x^* to Linear Program* (4.2) *which satisfies*

$$x_i^* \leq (d\lambda)^{-1} n \|b\|_\infty \;\; \textit{for } i = 1, \ldots, n. \tag{4.3}$$

The proof relies heavily on duality. The reader is directed to [BV04] or any other standard text for details.

To prove the claim, consider LP (4.2). By premultiplying the inequality constraint by $d^{-1}\mathbb{1}^T$ on both sides, we see that $\mathbb{1}^T x \geq d^{-1}\mathbb{1}^T b > -\infty$. Thus, the objective is bounded from below, which implies that strong duality holds. Thus, let $\tilde{z}$ be an optimal solution to the dual LP of (4.2):

$$\begin{aligned} \text{maximize} \quad & b^T z \\ \text{subject to:} \quad & Az = \mathbb{1} \\ & z \succeq 0 \end{aligned} \tag{4.4}$$

with dual variable $z \in \mathbb{R}^n$.

Next, consider the dual LP of (4.2) with the additional inequality constraints corresponding to (4.3):

$$\begin{aligned} \text{maximize} \quad & b^T z - (d\lambda)^{-1} n \|b\|_\infty \mathbb{1}^T y \\ \text{subject to:} \quad & Az = \mathbb{1} + y \\ & z \succeq 0 \\ & y \succeq 0 \end{aligned} \tag{4.5}$$

with dual variables $z \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. Equivalently, by setting $z = \tilde{z} + \Delta z$ and observing that $y = A\Delta z$, we can write the dual LP (4.5) as

$$\begin{aligned} \text{maximize} \quad & b^T \tilde{z} + b^T \Delta z - (d\lambda)^{-1} n \|b\|_\infty \mathbb{1}^T A \Delta z \\ \text{subject to:} \quad & A\Delta z \succeq 0 \\ & \tilde{z} + \Delta z \succeq 0 \end{aligned} \tag{4.6}$$

with dual variables $\Delta z \in \mathbb{R}^n$. We prove the claim by showing that the dual LPs (4.4) and (4.6) have the same optimal value. Since strong duality holds, the corresponding primal problems must also have the same optimal value.

Without loss of generality, we can uniquely decompose $\Delta z = \Delta z_1 + \Delta z_2$ where $\Delta z_1 \in \mathcal{N}(A)$ and $\Delta z_2 \in \mathcal{N}^\perp(A)$. Since $b \in \mathcal{N}(A)$, we have $b^T \Delta z_2 = 0$ and we can rewrite (4.6) yet again as

$$\begin{aligned} \text{maximize} \quad & b^T\tilde{z} + b^T\Delta z_1 - (d\lambda)^{-1}n\|b\|_\infty \mathbb{1}^T A\Delta z_2 && (4.7)\\ \text{subject to:} \quad & A\Delta z_2 \succeq 0 \\ & \tilde{z} + \Delta z_1 + \Delta z_2 \succeq 0 && (4.8)\\ & \Delta z_1 \in \mathcal{N}(A), \Delta z_2 \in \mathcal{N}^\perp(A). \end{aligned}$$

By definition of λ, for any unit vector $u \in \mathcal{N}^\perp(A)$ with $\|u\|_2 = 1$ we have $\|Au\|_2 \geq \lambda$. Using this and the fact that $A\Delta z_2 \succeq 0$ for all feasible Δz_2, we have the following inequality:

$$\mathbb{1}^T A\Delta z_2 = \|A\Delta z_2\|_1 \geq \|A\Delta z_2\|_2 \geq \lambda\|\Delta z_2\|_2.$$

Thus, the objective (4.7) can be upper bounded as follows:

$$b^T\tilde{z} + b^T\Delta z_1 - (d\lambda)^{-1}n\|b\|_\infty \mathbb{1}^T A\Delta z_2 \leq b^T\tilde{z} + b^T\Delta z_1 - d^{-1}n\|b\|_\infty\|\Delta z_2\|_2. \qquad (4.9)$$

Next, we obtain an upper bound on $b^T\Delta z_1$. To this end, observe that constraint (4.8) implies that $\tilde{z} + \Delta z_1 \succeq -\mathbb{1}\|\Delta z_2\|_\infty$. Motivated by this, consider the following ϵ-perturbed LP:

$$\begin{aligned} \text{minimize} \quad & -b^T v && (4.10)\\ \text{subject to:} \quad & \tilde{z} + v \succeq -\epsilon\mathbb{1} \\ & v \in \mathcal{N}(A). \end{aligned}$$

with variable v. Let $p^*(\epsilon)$ denote the optimal value of the ϵ-perturbed problem. First observe that $p^*(0) = 0$. To see this, note that if $\tilde{z} + v \succeq 0$, then $b^T v \leq 0$, else we would contradict the optimality of $\tilde{z}$ since $z = \tilde{z} + v$ is a feasible solution to the dual LP (4.4) in this case. Now, weak duality implies

$$-b^T v \geq p^*(\epsilon) \geq p^*(0) - \epsilon\mathbb{1}^T w^*, \qquad (4.11)$$

where w^* corresponds to an optimal solution to the dual LP of the unperturbed primal LP (4.10), given by:

$$\begin{aligned} \text{maximize} \quad & -\tilde{z}^T(Aw - b) \\ \text{subject to:} \quad & Aw \succeq b. \end{aligned} \tag{4.12}$$

Hence, (4.11) implies that

$$b^T \Delta z_1 \leq \|\Delta z_2\|_\infty \mathbb{1}^T w^* \tag{4.13}$$

if $\Delta z_1, \Delta z_2$ are feasible for LP (4.7).

By definition of $\tilde{z}$, $\tilde{z}^T A = \mathbb{1}^T$, and hence a vector w^* is optimal for (4.12) if and only if it also optimizes:

$$\begin{aligned} \text{minimize} \quad & \mathbb{1}^T w \\ \text{subject to:} \quad & Aw \succeq b. \end{aligned}$$

Combining this with (4.13), we have

$$b^T \Delta z_1 \leq \|\Delta z_2\|_\infty \mathbb{1}^T w^* \leq \|\Delta z_2\|_\infty \mathbb{1}^T w$$

for any vector w satisfying $Aw \succeq b$. Trivially, $w = d^{-1}\|b\|_\infty \mathbb{1}$ satisfies this, and hence we obtain:

$$b^T \Delta z_1 \leq d^{-1} n \|b\|_\infty \|\Delta z_2\|_\infty.$$

Finally, we substitute this into (4.9) and see that

$$\begin{aligned} b^T z &\leq b^T \tilde{z} + d^{-1} n \|b\|_\infty \|\Delta z_2\|_\infty - d^{-1} n \|b\|_\infty \|\Delta z_2\|_2 \\ &\leq b^T \tilde{z} + d^{-1} n \|b\|_\infty \|\Delta z_2\|_2 - d^{-1} n \|b\|_\infty \|\Delta z_2\|_2 \\ &\leq b^T \tilde{z} \end{aligned}$$

for all vectors z which are feasible for the dual LP (4.5). This completes the proof of Claim 1.

Claim 2. *There exists an optimal solution x^* to Linear Program* (4.2) *which satisfies*

$$|x_i| \leq c_A \|b\|_2 \quad \textit{for } i = 1, \ldots, n \tag{4.14}$$

for some constant c_A depending only on A.

First note that $\|b\|_\infty \leq \|b\|_2$ for any $b \in \mathbb{R}^n$, hence it suffices to prove the claim for the infinity norm. Claim 1 shows that each of the x_i's can be upper bounded by $(d\lambda)^{-1} n \|b\|_\infty$ without affecting the optimal value of LP (4.2). To see the lower bound, let a_j^T be a row of A with entry $a_{ji} \geq d/n$ in the i^{th} coordinate (at least one exists for each i since the columns of A sum to d). Now, the inequality constraint $Ax \succeq b$ combined with the upper bound on each x_i implies:

$$a_{ji} x_i + (d - a_{ji}) \lambda^{-1} n \|b\|_\infty \geq a_j^T x \geq b_j \geq -\|b\|_\infty. \tag{4.15}$$

Since $a_{ji} \geq d/n$, (4.15) implies:

$$x_i \geq -\lambda^{-1} n(n-1) \|b\|_\infty.$$

Hence, we can take $c_A = \lambda^{-1} n \times \max\{n-1, d^{-1}\}$. This proves Claim 2, and, by our earlier remarks, proves the lemma. □

4.3 Amplifying a Pointwise Convexity Constraint

Lemma 17. *Let $r_1, r_2 \in \mathbb{R}$ be given, and suppose $f_1 : K \to \mathbb{R}$ and $f_2 : K \to \mathbb{R}$ are continuous functions defined on a compact domain $K \subset \mathbb{R}^n$. If there exists a function $h : [0,1] \to K$ satisfying*

$$t \left(f_1 \circ h\right)(t) + (1-t) \left(f_2 \circ h\right)(t) \leq t r_1 + (1-t) r_2 \quad \textit{for all } t \in [0,1], \tag{4.16}$$

then there exists $x_1^*, x_2^* \in K$ and $t^* \in [0,1]$ for which

$$t^* f_1(x_1^*) + (1-t^*) f_1(x_2^*) \leq r_1$$
$$t^* f_2(x_1^*) + (1-t^*) f_2(x_2^*) \leq r_2.$$

4.3.1 Remarks

At first glance, this lemma appears somewhat bizarre. Indeed, the set K need only be compact (e.g., connectedness is not required) and h can be an arbitrarily complicated function, as long as it satisfies (4.16). The strange nature of the lemma is echoed by the proof in that we merely prove the existence of the desired x_1^*, x_2^* and t^*; no further information is obtained. Stripped to its core, the existence of the desired x_1^*, x_2^* and t^* essentially follows from the pigeon-hole principle, which manifests itself in the sequential compactness of K.

Despite its strange nature, Lemma 17 was crucial in establishing the converse result for the multiterminal source coding problem under logarithmic loss. Specifically, Lemma 17 demonstrates that an appropriately parametrized CEO problem can be tuned to yield the converse result for the two-encoder source coding problem with separate distortion constraints (see Chapter 2 for details). In this application, K was a closed subset of a finite-dimensional probability simplex and f_1, f_2 were conditional entropies evaluated for probability distributions in K.

Finally, we remark that the Lemma 17 can be generalized to a certain extent. For example, the function h need only be defined on a dense subset of $[0,1]$ and the set K can be a more general sequentially compact space.

4.3.2 Proof

Proof of Lemma 17. Since f_1, f_2 are continuous[1] and K is compact, there exists $M < \infty$ such that f_1 and f_2 are bounded from above and below by M and $-M$, respectively. Fix $\epsilon > 0$, and partition the interval $[0,1]$ as $0 = t_1 < t_2 < \cdots < t_m = 1$, such that $|t_{j+1} - t_j| < \frac{\epsilon}{M}$. For convenience define $x_{t_j} := h(t_j)$ when t_j is in the partition.

[1] Although not required for our purposes, we can assume f_1 and f_2 are defined and continuous over all of $\mathbb{R}^n$. This is a consequence of the Tietze extension theorem.

Now, for $i = 1, 2$ define piecewise-linear functions $g_1(t), g_2(t)$ on [0,1] by:

$$g_i(t) = \begin{cases} f_i(x_{t_j}) & \text{if } t_j \text{ is in the partition} \\ \theta f_i(x_{t_j}) + (1-\theta) f_i(x_{t_{j+1}}) & \text{if } t \text{ is in the interval } (t_j, t_{j+1}), \end{cases} \tag{4.17}$$

where $\theta \in (0,1)$ is chosen so that $t = \theta t_j + (1-\theta) t_{j+1}$ when t is in the interval (t_j, t_{j+1}).

With $g_1(t)$ and $g_2(t)$ defined in this manner, suppose $t = \theta t_j + (1-\theta)t_{j+1}$ for some j and $\theta \in [0,1]$. Then some straightforward algebra yields:

$$\begin{aligned} tg_1(t) + (1-t)g_2(t) &= (\theta t_j + (1-\theta)t_{j+1}) \left(\theta f_1(x_{t_j}) + (1-\theta) f_1(x_{t_{j+1}})\right) \\ &\quad + (1 - \theta t_j - (1-\theta)t_{j+1}) \left(\theta f_2(x_{t_j}) + (1-\theta) f_2(x_{t_{j+1}})\right) \\ &= \theta^2 \left[t_j f_1(x_{t_j}) + (1-t_j) f_2(x_{t_j})\right] \\ &\quad + (1-\theta)^2 \left[t_{j+1} f_1(x_{t_{j+1}}) + (1-t_{j+1}) f_2(x_{t_{j+1}})\right] \\ &\quad + \theta(1-\theta) \left[(1-t_j) f_2(x_{t_{j+1}}) + (1-t_{j+1}) f_2(x_{t_j})\right. \\ &\qquad\qquad \left. + t_{j+1} f_1(x_{t_j}) + t_j f_1(x_{t_{j+1}})\right] \\ &\le \theta^2 \left[t_j f_1(x_{t_j}) + (1-t_j) f_2(x_{t_j})\right] \\ &\quad + (1-\theta)^2 \left[t_{j+1} f_1(x_{t_{j+1}}) + (1-t_{j+1}) f_2(x_{t_{j+1}})\right] \\ &\quad + \theta(1-\theta) \left[(1-t_{j+1}) f_2(x_{t_{j+1}}) + (1-t_j) f_2(x_{t_j})\right. \\ &\qquad\qquad \left. + t_j f_1(x_{t_j}) + t_{j+1} f_1(x_{t_{j+1}})\right] + \epsilon \\ &\le \theta^2 \left[t_j r_1 + (1-t_j) r_2\right] \\ &\quad + (1-\theta)^2 \left[t_{j+1} r_1 + (1-t_{j+1}) r_2\right] \\ &\quad + \theta(1-\theta) \left[(1-t_{j+1}) r_2 + (1-t_j) r_2\right. \\ &\qquad\qquad \left. + t_j r_1 + t_{j+1} r_1\right] + \epsilon \\ &= (\theta t_j + (1-\theta)t_{j+1}) r_1 + (1 - \theta t_j - (1-\theta)t_{j+1}) r_2 + \epsilon \\ &= t r_1 + (1-t) r_2 + \epsilon, \end{aligned}$$

where the first inequality follows since $|t_{j+1} - t_j|$ is small, and the second inequality follows from the the fact that (4.16) holds for each t_j in the partition. Notably, this implies that

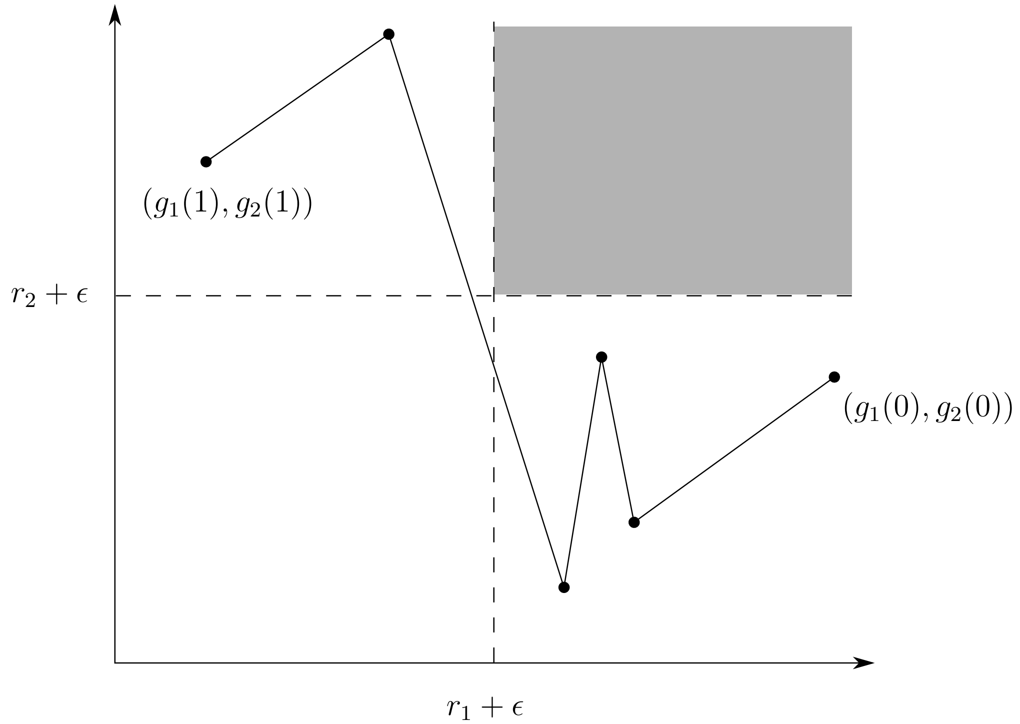

Figure 4.1: A parametric plot of the function $\varphi : t \mapsto (g_1(t), g_2(t))$. Since $\varphi(t)$ is continuous, starts with $g_2(0) \leq D_2 + \epsilon$, ends with $g_1(1) \leq D_1 + \epsilon$, and doesn't intersect the shaded area, $\varphi(t)$ must pass through the lower-left region.

it is impossible to have

$$g_1(t) > r_1 + \epsilon \text{ and } g_2(t) > r_2 + \epsilon$$

hold simultaneously for any $t \in [0, 1]$, else we would obtain a contradiction. Also, since we included the endpoints $t_1 = 0$ and $t_m = 1$ in the partition, we have the following two inequalities:

$$g_1(1) \leq r_1, \text{ and } g_2(0) \leq r_2.$$

Combining these observations with the fact that $g_1(t)$ and $g_2(t)$ are continuous, there must exist some $t^* \in [0, 1]$ for which

$$g_1(t^*) \leq r_1 + \epsilon, \quad \text{and} \quad g_2(t^*) \leq r_2 + \epsilon$$

simultaneously. An illustration of this is given in Figure 4.1.

Applying this result, we can find a sequence $\{x_1^{(n)}, x_2^{(n)}, t^{(n)}\}_{n=1}^{\infty}$ satisfying

$$\begin{aligned} t^{(n)} f_1(x_1^{(n)}) + (1 - t^{(n)}) f_1(x_2^{(n)}) &\leq r_1 + \frac{1}{n} \\ t^{(n)} f_2(x_1^{(n)}) + (1 - t^{(n)}) f_2(x_2^{(n)}) &\leq r_2 + \frac{1}{n} \end{aligned}$$

for each $n \geq 1$. Since $K \times K \times [0,1]$ is sequentially compact, there exists a convergent subsequence $\{n_j\}_{j=1}^{\infty}$ such that $(x_1^{(n_j)}, x_2^{(n_j)}, t^{(n_j)}) \to (x_1^*, x_2^*, t^*) \in K \times K \times [0,1]$. The continuity of f_1 and f_2 then apply to yield the desired result. □

4.4 Random Line Segments in the Unit Square

Let $Q = [0,1] \times [0,1]$ denote the unit square and let $\mathcal{L}_n$ be a set of n line segments in Q. Two line segments are said to be *crossing* if they intersect at any point. A subset of line segments is called *non-crossing* if no two segments in the subset are crossing.

Here, we consider the scenario where the endpoints of the n line segments are randomly distributed, independently and uniformly, in Q. To this end, define $N(\mathcal{L}_n)$ to be the size of the largest non-crossing subset of segments:

$$N(\mathcal{L}_n) = \max_{\mathcal{U} \subseteq \mathcal{L}_n} \left\{ |\mathcal{U}| : \ell_1, \ell_2 \text{ do not cross for all } \ell_1, \ell_2 \in \mathcal{U} \right\},$$

where $|\mathcal{U}|$ denotes the number of line segments in the subset $\mathcal{U}$.

Lemma 18. *There exists some constant $c \in [1/2, 15]$ such that*

$$\frac{N(\mathcal{L}_n)}{\sqrt{n}} \to c \text{ a.s.}$$

Moreover, $N(\mathcal{L}_n)$ is tightly concentrated about $c\sqrt{n}$ in the sense that $|N(\mathcal{L}_n) - c\sqrt{n}| \gg n^{1/4}$ with probability approaching zero as $n \to \infty$.

4.4.1 Remarks

Unlike the other two results in this chapter, Lemma 18 is not required elsewhere in this dissertation. However, the toy problem of counting the number of non-crossing line segments has become a favorite of mine, and as such, I could not resist the temptation to include it here.

My motivation for this result was the following. Consider a random geometric graph on $m > 2n$ vertices, and assume n randomly selected pairs of vertices are selected. How many pairs can be found for which the shortest paths connecting each vertex pair do not intersect the shortest paths connecting any other vertex pair? It can be shown that, for m sufficiently large, shortest paths connecting two vertices can be approximated by a straight line. Thus, Lemma 18 can be applied to estimate the maximum number of vertex pairs with disjoint shortest paths. The reader is directed to [HCC11] for complete details on this application.

Regarding the proof, the lower bound $N(\mathcal{L}_n) \geq 1/2\sqrt{n}$ w.h.p. is a relatively standard application of the probabilistic method (cf. [AS08]). The upper bound $N(\mathcal{L}_n) \leq 15\sqrt{n}$ w.h.p. is more delicate and requires a recent result by Sharir et al. [SW06].

Once we have these two bounds, the rest follows with relatively little effort. An application of Talagrand's Inequality [AS08, Section 7.7] reveals that $N(\mathcal{L}_n)$ is tightly concentrated about its median $m_n \in [\sqrt{n}/2, 15\sqrt{n}]$ in an interval of order no larger than $n^{1/4}$. Further, by applying the subadditive ergodic theorem (in a manner similar in spirit to [Dur10, Example 7.5.2] albeit to a four-dimensional Poisson process), one can show that there exists some constant $c \in [1/2, 15]$ such that

$$\frac{N(\mathcal{L}_n)}{\sqrt{n}} \to c \text{ a.s.}$$

Hence, almost everything is known about asymptotic the behavior of $N(\mathcal{L}_n)$, except the exact value of c. This is left as an open problem. This problem is similar in spirit to finding the length of the longest increasing subsequence in a random permutation. The reader is referred to [Ham72, VK77, LS77, BJ95] for more details.

4.4.2 Proof

Proof of Lemma 18. Given the sketch of the complete proof in the remarks above, we will only prove here that $1/2 \leq N(\mathcal{L}_n)/\sqrt{n} \leq 15$ with probability tending to 1 as $n \to \infty$.

Throughout the proof, we make the distinction between *left* and *right* endpoints of line segments. This is somewhat arbitrary, but simplifies the argument significantly. Thus, a line segment is generated according to the following process: (Step 1) the left endpoint is chosen uniformly from Q, and (Step 2) the right endpoint is chosen independently and uniformly from Q.

Claim 3. *With probability tending to 1 as $n \to \infty$, $N(\mathcal{L}_n) \geq \sqrt{n}/2$.*

Partition Q into $\sqrt{n}$ disjoint horizontal strips[2], each having height $1/\sqrt{n}$ and width 1. Note that if a line segment ℓ is contained in a single strip, then it will not intersect line segments contained in any other strip. Then $N(\mathcal{L}_n) \geq Y$, where Y is the number of strips that contain line segments. Observe that

$$\begin{aligned}
&\Pr[\text{Line } \ell \text{ in strip } j] \\
&= \Pr[\{\text{left endpoint of } \ell \text{ in strip } j\} \wedge \{\text{right endpoint of } \ell \text{ in strip } j\}] \\
&= \Pr[\text{left endpoint of } \ell \text{ in strip } j] \times \Pr[\text{right endpoint of } \ell \text{ in strip } j] \\
&= \frac{1}{n}.
\end{aligned}$$

Where we used the fact that the probability a given point falls in a particular strip is $1/\sqrt{n}$ and points are chosen independently. Then, the probability that a given strip does not contain any line segments is:

$$(1 - 1/n)^n \approx 1/e.$$

[2] The idea of partitioning Q into strips to find roughly $\sqrt{n}$ segments is originally due to Professor Michael Neely [Nee11].

Further, note that:

$$\begin{aligned}
&\Pr\left[\{\text{Line } \ell \text{ not in strip } i\} \wedge \{\text{Line } \ell \text{ not in strip } j\}\right] \\
&= 1 - \Pr\left[\{\text{Line } \ell \text{ in strip } i\} \vee \{\text{Line } \ell \text{ in strip } j\}\right] \\
&= 1 - \left(\Pr\left[\text{Line } \ell \text{ in strip } i\right] + \Pr\left[\text{Line } \ell \text{ in strip } j\right]\right) \\
&= 1 - 2/n.
\end{aligned}$$

Where we used the fact that the events $\{\text{Line } \ell \text{ in strip } i\}$ and $\{\text{Line } \ell \text{ in strip } j\}$ are disjoint. Then for any pair of strips (i, j), the probability that neither strip i nor strip j contains any line segments is

$$(1 - 2/n)^n \approx 1/e^2.$$

Let X_i be the indicator random variable taking value 1 if strip i contains no line segments and taking the value 0 otherwise. Note that $\mathbb{E}X_i \approx e^{-1}$, $\mathrm{Var}(X_i) \approx e^{-1}(1 - e^{-1})$, and

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right] \\
&= (1 - 2/n)^n - (1 - 1/n)^{2n} \\
&\leq e^{-2} - e^{-2} + o(1) \\
&= o(1).
\end{aligned}$$

Then, letting $X = \sum_{i=1}^{\sqrt{n}} X_i$ be the number of strips that don't contain any line segments, and noting that

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) \\
&\leq \sqrt{n}\left(\frac{1}{e}\left(1 - \frac{1}{e}\right) + o(1)\right) + \sqrt{n}(\sqrt{n} - 1)o(1) \\
&\leq \sqrt{n} + n \cdot o(1),
\end{aligned}$$

Chebyshev's inequality yields:

$$\begin{aligned}\Pr\left[|X - \mathbb{E}X| \geq \frac{1}{10}\sqrt{n}\right] &\leq 100 \times \frac{\mathrm{Var}(X)}{n} \\ &\leq 100 \times \frac{\sqrt{n} + n \cdot o(1)}{n} \\ &\to 0.\end{aligned}$$

Therefore, with probability tending to 1,

$$X \leq \left(1 + \frac{1}{10} + o(1)\right)\frac{\sqrt{n}}{e} \leq \frac{\sqrt{n}}{2}.$$

Hence, $Y = \sqrt{n} - X \geq \sqrt{n}/2$ with probability tending to 1. This proves the claim.

Claim 4. *With probability tending to 1 as $n \to \infty$, $N(\mathcal{L}_n) \leq 15\sqrt{n}$.*

From [SW06], there exists an absolute constant c such that for any $2k$ points in the plane, the number of non-crossing left-right[3] perfect matchings is upper-bounded by $c \cdot 29^k$. Consider any realization of n line segments in the plane and further consider the $2k$ (k left and k right) endpoints corresponding to any subset S consisting of k line segments. Conditioned on the locations of the left and right endpoints, every left-right perfect matching of these $2k$ points is equally likely, and thus the probability that these k segments are non-crossing is upper bounded by:

$$\Pr\left[S \text{ is non-crossing}\right] \leq \frac{c \cdot 29^k}{k!}$$

since there are $k!$ left-right perfect matchings on the $2k$ endpoints.

Stirling's formula states

$$\lim_{k\to\infty} \frac{k!}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k} = 1,$$

and thus

$$\frac{c \cdot 29^k}{k!} \leq o(1) \cdot \left(\frac{29 \cdot e}{k}\right)^k.$$

[3] A left-right perfect matching distinguishes between left endpoints and right endpoints in edges. In other words, an edge is only allowed to match a left endpoint to a right endpoint.

Recalling the crude upper bound $\binom{n}{k} \le \left(\frac{n \cdot e}{k}\right)^k$, a union bound gives:

$$\begin{aligned}
\Pr[\exists\ k \text{ non-crossing segments}] &\le \binom{n}{k} \frac{c \cdot 29^k}{k!} \\
&\le o(1) \cdot \left(\frac{29 \cdot n \cdot e^2}{k^2}\right)^k
\end{aligned}$$

Letting $k = 15\sqrt{n}$, we have

$$\begin{aligned}
\Pr[\exists\ 15\sqrt{n} \text{ non-crossing segments}] &\le o(1) \cdot \left(\frac{29 \cdot e^2}{15^2}\right)^{15\sqrt{n}} \\
&\le o(1) \cdot (.96)^{15\sqrt{n}} \to 0.
\end{aligned}$$

This proves the claim. □

CHAPTER 5

Conclusion

5.1 Summary of the Results

In this dissertation, we investigated both lossy and lossless source coding problems. In the former case, we studied source coding problems where distortion is measured under logarithmic loss. We characterized the achievable rate distortion region for the two-encoder multiterminal source coding problem and, along the way, we also characterized the rate distortion region for the m-encoder CEO problem (both under logarithmic loss). We gave several applications of these results, including betting on correlated horse races, estimation of a posterior distribution, and list decoding. In addition, we demonstrated that our results can be applied to obtain quantitative bounds on the general multiterminal source coding problem with arbitrary distortion measures.

We also derived a single-letter description for a fundamental entropy characterization problem. Namely, for correlated sources encoded separately by rate-limited encoders, we precisely characterized the amount of information that can be revealed about each source by the encoded representations. A consequence of this result is that relatively simple encoding functions suffice to attain any achievable information pair for given rates.

In the context of lossless coding, we studied the problem of cooperative data exchange. Here, we gave necessary and sufficient conditions for achieving universal recovery in an arbitrarily connected network. When the network has unit diameter (i.e., is fully connected), we derived an algorithm based on submodular optimization that can efficiently compute an optimal transmission schedule.

When packets are randomly distributed and the network topology satisfies certain regularity conditions, we proved tight concentration results on the number of transmissions required to achieve universal recovery. Notably, the concentration interval is independent of the number of packets in the network. In the special case of a single-hop network, this concentration is shown to occur at a single, computable value. We also showed in these settings that splitting packets does not significantly reduce the number of transmissions required to achieve universal recovery.

As an application, we related our results to the task of distributed secrecy generation amongst a collection of nodes in the presence of an eavesdropper. Specifically, we showed that the submodular optimization algorithm we derived can be used to compute and generate the theoretical maximum amount of secrecy among nodes in a practical manner.

Finally, we gave three technical lemmas, each of which has applications to multiterminal information theory. We intentionally highlighted these lemmas in their own chapter, as each appears to be interesting in its own right. Perhaps these results will enjoy additional applications in information theory and other fields.

5.2 Future Directions

Multiterminal information theory continues to be a rich source of problems, and will continue to be so for the foreseeable future. We take this opportunity to briefly mention some potential directions for research specifically related to this dissertation.

In the context of the multiterminal source coding problem, one immediate direction for further work would be to extend our results to more than two encoders. Indeed, we have shown for the CEO problem that our results can be extended to an arbitrary number of encoders. However, generalizing the results for the two-encoder source coding problem with separate distortion constraints on each source poses a formidable challenge. In fact, any extension of our converse alone would not be sufficient since this would imply that the Berger-Tung inner bound is tight for more than two encoders. This is known to be false (even under logarithmic loss) since the Berger-Tung achievability scheme is not optimal for the lossless modulo-sum problem studied by Körner and Marton in [KM79].

For the cooperative data exchange problem, it seems plausible that an efficient algorithm can be found for computing optimal transmission schedules in other settings of interest. For example, in the case of regular networks or trees with deterministic packet distributions. On the other hand, we have seen that minimizing the number of transmissions required to permit universal recovery problem is NP-hard in general. Therefore, an interesting first step would be to try to find an example demonstrating that this problem is NP-hard for regular networks.

References

[AAB06] T. Andre, M. Antonini, M. Barlaud, and R.M. Gray. "Entropy-Based Distortion Measure for Image Coding." In *Image Processing, 2006 IEEE International Conference on*, pp. 1157 –1160, Oct. 2006.

[ACR00] Rudolf Ahlswede, Ning Cai, Shuo yen Robert Li, and Raymond W. Yeung. "Network Information Flow." *IEEE TRANSACTIONS ON INFORMATION THEORY*, **46**(4):1204–1216, 2000.

[AK75] R. Ahlswede and J. Korner. "Source coding with side information and a converse for degraded broadcast channels." *Information Theory, IEEE Transactions on*, **21**(6):629 – 637, Nov 1975.

[ALS08] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim. "Broadcasting with Side Information." In *Foundations of Computer Science, 2008. FOCS '08. IEEE 49th Annual IEEE Symposium on*, pp. 823 –832, oct. 2008.

[AS08] Noga Alon and Joel H. Spencer. *The Probabilistic Method.* John Wiley & Sons, 3rd edition, 2008.

[BBC08] A. Blair, T. Brown, K.M. Chugg, T.R. Halford, and M. Johnson. "Barrage relay networks for cooperative transport in tactical MANETs." In *Military Communications Conference, 2008. MILCOM 2008. IEEE*, pp. 1 –7, nov. 2008.

[Ber77] T. Berger. *Multiterminal Source Coding. In G. Longo (Ed.), The Information Theory Approach to Communications.* Springer-Verlag, New York, NY, USA, 1977.

[BJ95] Bela Bollobas and Svante Janson. "On the length of the longest increasing subsequence in a random permutation." In *Combinatorics, Probability and Computing*, p. 121128. Univ. Press, 1995.

[BK06] Y. Birk and T. Kol. "Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients." *Information Theory, IEEE Transactions on*, **52**(6):2825 – 2830, june 2006.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[BY89] T. Berger and R.W. Yeung. "Multiterminal source encoding with one distortion criterion." *Information Theory, IEEE Transactions on*, **35**(2):228 –236, Mar 1989.

[BZV96] T. Berger, Zhen Zhang, and H. Viswanathan. "The CEO problem [multiterminal source coding]." *Information Theory, IEEE Transactions on*, **42**(3):887 –902, may 1996.

[Chi12] Yeow Khiang Chia. *Personal communication.* April 2012.

[CK81] I. Csiszar and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, NY, USA, 1981.

[CL06] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games.* Cambridge University Press, New York, NY, USA, 2006.

[CN04] I. Csiszar and P. Narayan. "Secrecy capacities for multiple terminals." *Information Theory, IEEE Transactions on*, **50**(12):3047 – 3061, dec. 2004.

[Cou12] T.A. Courtade. "Information Masking and Amplification: The Source Coding Setting." In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012.

[CT06] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, 2nd edition, 2006.

[CW10] T.A. Courtade and R.D. Wesel. "Efficient universal recovery in broadcast networks." In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pp. 1542 –1549, 29 2010-oct. 1 2010.

[CW11a] T.A. Courtade and R.D. Wesel. "Multiterminal source coding with an entropy-based distortion measure." In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 2040 –2044, Aug. 2011.

[CW11b] T.A. Courtade and R.D. Wesel. "Weighted Universal Recovery, Practical Secrecy, and an Efficient Algorithm for Solving Both." In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, Oct. 2011.

[CW11c] Thomas A. Courtade and Tsachy Weissman. "Multiterminal Source Coding under Logarithmic Loss." *CoRR*, **abs/1110.3069**, 2011.

[CW12a] T.A. Courtade and Tsachy Weissman. "Multiterminal Source Coding under Logarithmic Loss." In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012.

[CW12b] Thomas A. Courtade and Richard D. Wesel. "Coded Cooperative Data Exchange in Multihop Networks." *CoRR*, **abs/1203.3445**, 2012.

[CXW10] T.A. Courtade, Bike Xie, and R.D. Wesel. "Optimal exchange of packets for universal recovery in broadcast networks." In *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, pp. 2250 –2255, 31 2010-nov. 3 2010.

[Dur10] Richard Durrett. *Probability: Theory and Examples.* Cambridge University Press, 4th edition, 2010.

[ECS07] S.Y. El Rouayheb, M.A.R. Chaudhry, and A. Sprintson. "On the Minimum Number of Transmissions in Single-Hop Wireless Coding Networks." In *Information Theory Workshop, 2007. ITW '07. IEEE*, pp. 120 –125, sept. 2007.

[EK12] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2012.

[Erk96] Elza Erkip. *The Efficiency of Information in Investment*. PhD thesis, Stanford University, 1996.

[ESS10] S. El Rouayheb, A. Sprintson, and P. Sadeghi. "On coding for cooperative data exchange." In *Information Theory Workshop (ITW), 2010 IEEE*, pp. 1 –5, Jan. 2010.

[FHI06] Satoru Fujishige, Takumi Hayashi, and Shigueo Isotani. "The Minimum-Norm-Point Algorithm Applied to Submodular Function Minimization and." In *Kyoto University, Kyoto Japan*, 2006.

[Fuj10] S. Fujishige. *Submodular Functions and Optimization*. Elsevier Science, Berlin, 2nd edition, 2010.

[GC82] A.E. Gamal and T. Cover. "Achievable rates for multiple descriptions." *Information Theory, IEEE Transactions on*, **28**(6):851 – 857, nov 1982.

[GCT01] Amir Globerson, Gal Chechik, Naftali Tishby, Orna Steinberg, and Eilon Vaadia. "Distributional Clustering of Movements Based on Neural Responses.", 2001.

[GE07] W. Gu and M. Effros. "On Approximating the Rate Region for Source Coding with Coded Side Information." In *Information Theory Workshop, 2007. ITW '07. IEEE*, pp. 432 –435, Sept. 2007.

[GJE08] W. Gu, S. Jana, and M. Effros. "On approximating the rate regions for lossy source coding with coded and uncoded side information." In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pp. 2162 –2166, July 2008.

[GNT03] Ran Gilad-bachrach, Amir Navot, and Naftali Tishby. "An information theoretic tradeoff between complexity and accuracy." In *In Proceedings of the COLT*, pp. 595–609. Springer, 2003.

[Ham72] J M Hammersley. "A few seedlings of research." In *Univ. of California*, pp. 345–394. Press, 1972.

[HC10] T.R. Halford and K.M. Chugg. "Barrage Relay Networks." In *Information Theory and Applications Workshop (ITA), 2010*, pp. 1 –8, 31 2010-feb. 5 2010.

[HCC11] T. R. Halford, T. A. Courtade, and K. M. Chugg. "Barrage relay networks." *Under Preparation*, 2011.

[HCP10] T.R. Halford, K.M. Chugg, and A. Polydoros. “Barrage relay networks: System & protocol design.” In *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, pp. 1133 –1138, sept. 2010.

[HH10] T.R. Halford and G. Hwang. “Barrage relay networks for unmanned ground systems.” In *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, pp. 1274 –1280, 31 2010-nov. 3 2010.

[HMK06] T. Ho, M. Medard, R. Koetter, D.R. Karger, M. Effros, Jun Shi, and B. Leong. “A Random Linear Network Coding Approach to Multicast.” *Information Theory, IEEE Transactions on*, **52**(10):4413 –4430, oct. 2006.

[HT05] Ron M. Hecht and Naftali Tishby. “Extraction of Relevant Speech Features using the Information Bottleneck Method.” In *In Proceedings of ”InterSpeech, (Lisbon)*, pp. 353–356, 2005.

[HT07] Peter Harremoes and Naftali Tishby. “The Information Bottleneck Revisited or How to Choose a Good Distortion Measure.” In *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pp. 566 –570, June 2007.

[Jan09] S. Jana. “Alphabet sizes of auxiliary random variables in canonical inner bounds.” In *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pp. 67 –71, March 2009.

[JSC05] S. Jaggi, P. Sanders, P.A. Chou, M. Effros, S. Egner, K. Jain, and L.M.G.M. Tolhuizen. “Polynomial time algorithms for multicast network code construction.” *Information Theory, IEEE Transactions on*, **51**(6):1973 – 1982, june 2005.

[Kar72] R. M. Karp. “Reducibility Among Combinatorial Problems.” In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pp. 85–103. Plenum Press, 1972.

[KM79] J. Korner and K. Marton. “How to encode the modulo-two sum of binary sources (Corresp.).” *Information Theory, IEEE Transactions on*, **25**(2):219 – 221, Mar 1979.

[KS10] Andreas Krause and Soeren Sonnenburg. “SFO: A Toolbox for Submodular Function Optimization, The.” *Journal of Machine Learning Research*, pp. 1141–1144, 2010.

[KSC08] Young-Han Kim, A. Sutivong, and T.M. Cover. “State Amplification.” *Information Theory, IEEE Transactions on*, **54**(5):1850 –1859, may 2008.

[KSV11] Onur Ozan Koyluoglu, Rajiv Soundararajan, and Sriram Vishwanath. “State Amplification Subject To Masking Constraints.” *CoRR*, **abs/1112.4090**, 2011.

[Lau05] Alan J. Laub. *Matrix analysis - for scientists and engineers.* SIAM, 2005.

[LS77] B F Logan and L A Shepp. “A variational problem for random Young tableaux. Adv.” In *In Proc. International Congress of Mathematicians, Zurich*, pp. 1384–1394, 1977.

[LS09] E. Lubetzky and U. Stav. “Nonlinear Index Coding Outperforming the Linear Optimum.” *Information Theory, IEEE Transactions on*, **55**(8):3544 –3551, aug. 2009.

[McC05] S. McCormick. *Submodular Function Minimization. In Discrete Optimization, K. Aardal, G. Nemhauser, and R. Weismantel, eds. Handbooks in Operations Research and Management Science*, volume 12. Elsevier, 2005.

[MS07] N. Merhav and S. Shamai. “Information Rates Subject to State Masking.” *Information Theory, IEEE Transactions on*, **53**(6):2254 –2261, june 2007.

[Nee11] Michael Neely. *Personal communication.* 2011.

[Ooh97] Y. Oohama. “Gaussian multiterminal source coding.” *Information Theory, IEEE Transactions on*, **43**(6):1912 –1923, nov 1997.

[Ooh05] Y. Oohama. “Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder.” *Information Theory, IEEE Transactions on*, **51**(7):2577 –2593, july 2005.

[OS11] D. Ozgul and A. Sprintson. “An algorithm for cooperative data exchange with cost criterion.” In *Information Theory and Applications Workshop (ITA), 2011*, pp. 1 –4, feb. 2011.

[PTR04] V. Prabhakaran, D. Tse, and K. Ramachandran. “Rate region of the quadratic Gaussian CEO problem.” In *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, p. 119, june-2 july 2004.

[Ren59] A. Rényi. “On measures of dependence.” *Acta Mathematica Hungarica*, **10**:441–451, 1959. 10.1007/BF02024507.

[RMY03] Shuo yen Robert Li, Senior Member, Raymond W. Yeung, and Ning Cai. “Linear network coding.” *IEEE Transactions on Information Theory*, **49**:371–381, 2003.

[RSW05] Aditya Ramamoorthy, Jun Shi, and Richard D. Wesel. “On the Capacity of Network Coding for Random Networks.” *IEEE TRANSACTIONS ON INFORMATION THEORY*, **51**(8):2878–2885, 2005.

[Sch03] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency.* Springer-Verlag, Berlin, 2003.

[Sha59] Claude E. Shannon. "Coding Theorems for a Discrete Source with a Fidelity Criterion." *Institute of Radio Engineers, International Convention Record*, **7, part 4**:142–163, March 1959.

[SRP11] Lalitha Sankar, S. Raj Rajagopalan, and H. Vincent Poor. "A Theory of Privacy and Utility in Databases." *CoRR*, **abs/1102.3751v1**, 2011.

[SSB10a] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb. "Deterministic Algorithm for Coded Cooperative Data Exchange." In *ICST QShine*, Nov. 2010.

[SSB10b] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb. "A randomized algorithm and performance bounds for coded cooperative data exchange." In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 1888 –1892, june 2010.

[SST01] Noam Slonim, Rachel Somerville, Naftali Tishby, and Ofer Lahav. "Objective classification of galaxy spectra using the information bottleneck method." *Monthly Notes of the Royal Astronomical Society*, **323**:2001, 2001.

[ST01] Noam Slonim and Naftali Tishby. "The power of word clusters for text classification." In *In 23rd European Colloquium on Information Retrieval Research*, 2001.

[SW73] D. Slepian and J. Wolf. "Noiseless coding of correlated information sources." *Information Theory, IEEE Transactions on*, **19**(4):471 – 480, Jul 1973.

[SW06] Micha Sharir and Emo Welzl. "On the Number of Crossing-Free Matchings, Cycles, and Partitions." *SIAM J. Comput.*, **36**:695–720, September 2006.

[TPB99] N. Tishby, F. Pereira, and W. Bialek. "The information bottleneck method." In *In Proceedings of the 37th annual Allerton Conference on Communication, Control, and Computing*, pp. 368–377, 1999.

[Tun78] S.-Y. Tung. *Multiterminal Source Coding.* PhD thesis, Cornell University, Ithaca, NY, 1978.

[VK77] A M Versik and S V Kerov. "Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux, Dokl. Akad. Nauk." Technical report, (Russian) Bela Bollobas, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, 16 Mill Lane, Cambridge, 1977.

[Wit80] H. Witsenhausen. "Some aspects of convexity useful in information theory." *Information Theory, IEEE Transactions on*, **26**(3):265 – 271, May 1980.

[WKA11] A.B. Wagner, B.G. Kelly, and Y. Altug. "Distributed Rate-Distortion With Common Components." *Information Theory, IEEE Transactions on*, **57**(7):4035 –4057, july 2011.

[WTV08] A.B. Wagner, S. Tavildar, and P. Viswanath. "Rate Region of the Quadratic Gaussian Two-Encoder Source-Coding Problem." *Information Theory, IEEE Transactions on*, **54**(5):1938 –1961, May 2008.

[Wyn75] A. Wyner. "On source coding with side information at the decoder." *Information Theory, IEEE Transactions on*, **21**(3):294 – 300, May 1975.

[WZ76] A. Wyner and J. Ziv. "The rate-distortion function for source coding with side information at the decoder." *Information Theory, IEEE Transactions on*, **22**(1):1 – 10, Jan. 1976.

[YN05] Chunxuan Ye and P. Narayan. "Secret key and private key constructions for simple multiterminal source models." In *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pp. 2133 –2137, sept. 2005.

[YR10] Chunxuan Ye and A. Reznik. "A simple secret key construction system for broadcasting model." In *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pp. 1 –6, march 2010.

[ZB99] R. Zamir and T. Berger. "Multiterminal source coding with high resolution." *Information Theory, IEEE Transactions on*, **45**(1):106 –117, jan 1999.