On the Tightness of Inner and Outer Bounds for Broadcast Channels with Three and More Receivers

WANG, Zizhou

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Information Engineering

The Chinese University of Hong Kong

June 2010

Abstract of thesis entitled:

On the Tightness of Inner and Outer Bounds for Broadcast Channels with Three and More Receivers

Submitted by WANG, Zizhou

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in June 2010

This thesis focused on a fundamental problem of network information theory called broadcast channel, which models the communication from a single sender to multiple receivers (say, from a cellular tower to cell phone users in its coverage area). The goal is to determine the set of achievable communication data rates, so each receiver can decode the messages it requires with high fidelity. From a purely theoretical standpoint, however, this problem of characterizing the feasible rate region (capacity region) had stumped researchers for over three decades.

The main contributions in this thesis consist of the following three parts:

The first part studied the existing inner and outer bounds

to the capacity region for 3-receiver broadcast channels with 2-degraded message sets, in an attempt to find the deficiencies with the current techniques of establishing the bounds. We produced a simple example where we were able to explicitly evaluate these bounds to show that they are indeed different. For a class of channels where the bounds differ, we used a new argument to show that the inner bound is tight and outer bound is weak.

The second part considered a broadcast channel consisting of k receivers that lie in a less noisy sequence. The capacity region for this scenario had been unknown since the mid 1970s, when $k \geq 3$. We solved this open problem for the case $k = 3$. Indeed we proved that superposition coding is optimal for a class of broadcast channels with a sequence of less noisy receivers. This class contains the $k = 3$ case, thus resolving its capacity region.

The last part considered a k-receiver broadcast channel with two unmatched degraded components, and degraded message sets where receiver Y_s, $s \in \{1, \cdots, k\}$ requires messages $(M_s, \cdots, M_k)$. We established the capacity region for this class of broadcast channels by showing that superposition coding is optimal. In the process of proving the achievability, we showed a general superposition coding region for any broadcast channels with degraded message requirement.

Acknowledgement

My foremost gratitude is to my advisor, Professor Shuo-Yen Robert Li. He has been a constant source of encouragement and motivation throughout my graduate life, with his enthusiasm for teaching me the right ways to study the problems and kindly sharing of his living experiences. It is a pity that even with him I did not become an expert in the fields of network coding and algebraic switching, but I did learn a very useful way to look at some math and engineering problems.

Moreover, I would like to acknowledge my co-advisor, Professor Chandra Nair, who led me into the very exciting field of network information theory, and worked with me closely throughout the last two and a half years. This thesis is the outcome of my collaboration with him. I benefited greatly from his patient guidance on every aspect of the thesis. However, what I have learned from him is not just the solution to some problems, but his insights, inspiration, his way of conducting research and living. In fact I could never forget those days that we have been talking for five hours to study one problem, the way he worked harder on my talk than myself and many times that I gave the presentation with only one audience. All I can say is that I could hardly ask for any more from such kind advisor.

Besides, I worked in a harmonious laboratory, the Switching Laboratory. The excellent atmosphere here for studying, working, and living is created and maintained by my advisor Professor Shuo-Yen Robert Li together with my colleagues Dr.Jian Zhu, Dr.Xuesong Jonathan Tan, Dr.Qifu Sun, Mr.Zhengfeng Qian, Dr.Siuting Ho, Mr.Ziyu Shao, Ms.Shuqin Li, Ms.Xiaoming Wu, Ms.Lokman Janice Law, Mr.Tong Liang, Mr.Qiwei Li and Mr.Yanlin Geng.

Finally, I would like to thank my mother for her constant support and understanding. I can hardly appreciate more for her selfless care and love.

This work is dedicated to my parents.

摘要

本論文研究課題是網絡信息論中的廣播信道，即模擬一個基站到多個手機用戶之間的通信，其目標是確定在該網絡中可以實現的最大信息傳輸速度（即信道容量），明確目前應用的通信技術可能的提升空間。雖然很多研究者經過了三十余年的努力，但這個看似簡單的理論問題迄今為止仍未得到完全解決。

本論文的主要貢獻由以下三個部分組成：

第一部分的研究模型是由三個接收機組成的廣播信道並帶有降級信息集合的要求。具體來說該類信息傳送機制是：接收機 1 和 2 需要解調出所有的信息而接收器 3 只需要解調公共的信息。在該模型下我們提出了一個簡單的信道並在該信道上成功的確定了現有容量區域上下界的具體差值。對於一類滿足容量區域上下界不同的信道，我們證明出其信道容量正是容量區域的下界。

第二部分的研究模型是由 k個接收 機組成的廣播信道並且接收機按照抗噪能力排序。該類信道的容量問題即使對於簡單的由 3 個接 收機組成的廣播信道自 70 年代中期以來就一直未能得到解決。我們在該模型下定義了一類新的廣播信道並得出了其信道容量。由於上面提及的 3 個接收器組成的廣播信道是此類新的信道的特例，由此信道容量問題得以解決。

第三部分的研究模型是由 k個接收 機組成的廣播積信道並帶有降級信息集合的要求。我們解決了該類信道的容量問題並在證明該類信道信息傳輸速度的可實現性過程中，推導得出廣義信道容量區域的下界。

Contents

List of Figures

Chapter 1

Introduction

The demand for higher data rate over wireless networks has increased dramatically primarily due to two facts: first, the density of wireless users has grown several-fold; and second, the data demand per user has gone up considerably. Since wireless bandwidth is a limited, scarce, and expensive resource, one of the fundamental design challenges in communication is to devise strategies to maximize the throughput (data rate) in a given scenario. To this end, it is also critical for one to know the fundamental limits of communication in the scenario.

There used to be a dichotomy between the rates for practical communication scenarios and the fundamental limits of communication as the latter was in most cases infeasible to implement. However, remarkable advances in technology (both hardware and algorithms) over the past decade has significantly closed the gap between practical algorithms and fundamental limits. This has rekindled the interest in multiuser information theory.

Unlike point-to-point communication, most fundamental communication problems in multiuser setting remain unsolved and there is no general theory yet. A very important setting is the so called broadcast channel, which models communication from a single sender to multiple receivers (say, from a cellular tower to cell phone users in its coverage area). The characterization of the optimal rates, even for the case of two receivers, has been a long standing open problem. Moreover, almost all available tools and techniques such as superposition coding and random binning date back to the 1970s and early 1980s. Thus to compute the fundamental communication limits one has to develop new techniques and understand optimal strategies, which then can be incorporated into practical implementations.

As an attempt to understand the fundamental limits of reliable communication, the present thesis studies the best known inner and outer bounds for broadcast channels. Here the quantity of interest is to enlarge the available set of communication strategies in large networks as well as to establish tight characterizations of achievable rates, i.e. capacity region, for some new classes of broadcast channels with 3 or more receivers.

1.1 Background

1.1.1 Point-to-point communication

In his celebrated work, Shannon[24] proved two basic theorems regarding "point-to-point" communications, which forms the basis of single-user information theory.

- Source coding theorem: The maximum compression rate of a stationary ergodic source sequence that can be successfully reconstructed at a receiver is equal to the entropy rate of the source.

- Channel coding theorem: The maximum rate at which one can transmit information reliably through a noisy discrete memoryless channel is equal to the capacity of the channel.

Given the source statistics in source coding, it is easy to compute the entropy rate of the source. Similarly, the channel transition matrix yields the capacity of the channel. Thus the fundamental limits of a single user communication system serve as benchmarks to compare the performance of any feasible communication strategy. The recent advances in coding techniques using better algorithms, higher processing power, etc, have almost achieved the fundamental limits in implementation.

1.1.2 Multiuser information theory

The field of multiuser information theory was again started by Shannon[23], though his original problem of a two-way communication channel remains unsolved. Any multiuser network should consist of at least one of the following two components: a sender transmitting (intentionally or unintentionally) to multiple receivers, or a receiver hearing (wittingly or unwittingly) from multiple senders.

- Many transmitters, one receiver: Commonly referred to as the multiple access channel (MAC), which models communication from many senders to a single receivers (say, from cell phone users to a common cellular tower). The capacity region of MAC is one of the few multi-user scenarios where one has a complete characterization of the capacity[1][2][14]. The capacity region in the MAC has a close connection to the max-flow min-cut bounds in graph theory and the maximum information flow is indeed given by the minimum cut inequalities.

- One transmitters, many receivers: Commonly referred to as the broadcast channel (BC)[4], which models communication from a single sender to the multiple receivers (say, from a cellular tower to cell phone users in its coverage area). The characterization of the capacity, even for the case of two receivers, is still open.

1.1.3 Broadcast channel

Cover[4] introduced the notion of a broadcast channel where one sender transmits information to multiple receivers. Formally, a k-receiver broadcast channel consists of an input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}_1, \cdots, \mathcal{Y}_k$ and a probability transition function $p(y_1, \cdots, y_k|x)$. A $\big((2^{nR_0}, 2^{nR_1}, \cdots, 2^{nR_k}), n\big)$ code for a broadcast channel with common information consists of the following:

1. A common message set $\mathcal{W}_0 = \{1, \cdots, 2^{nR_0}\}$ and k private message sets $\mathcal{W}_k = \{1, \cdots, 2^{nR_k}\}$;
2. Messages $(W_0, W_1, \cdots, W_k)$ are independent of each other and are uniformly distributed over $(\mathcal{W}_0, \mathcal{W}_1, \cdots, \mathcal{W}_k)$ respectively;
3. An encoder maps each message tuple $(W_0, W_1, \cdots, W_k)$ into a codeword $X^n \in \mathcal{X}^n$;
4. k decoders: decoder at receiver $Y_i, i \in \{1, \cdots, k\}$, i.e. $g_i : \mathcal{Y}_i^n \to \mathcal{W}_0 \times \mathcal{W}_i$ maps the received sequence $y_i^n \in \mathcal{Y}_i^n$ into

an estimate message pair $(\hat{W}_0, \hat{W}_i) \in \mathcal{W}_0 \times \mathcal{W}_i$.

The probability of error $P_e^{(n)}$ is defined as the probability that the decoded message is not equal to the transmitted message, i.e.,

$$P_e^{(n)} = \mathbf{P}\left(\{g_1(Y_1^n) \neq (\mathcal{W}_0, \mathcal{W}_1)\} \cup \cdots \cup \{g_k(Y_k^n) \neq (\mathcal{W}_0, \mathcal{W}_k)\}\right)$$

A rate tuple $(R_0, R_1, \cdots, R_k)$ is said to be *achievable* for the broadcast channel if there exists a sequence of $\big((2^{nR_0}, 2^{nR_1}, \cdots, 2^{nR_k}), n\big)$ codes with $P_e^{(n)} \to 0$. The *capacity region* of the broadcast channel is defined as the closure of the set of achievable rates. However the characterization of the capacity region, even for the $k = 2$ case, is still open.

The capacity region was known for some special cases such as degraded broadcast channel[3][10], less noisy[11], more capable[8], essentially less noisy[17], and essentially more capable[17] etc. It was shown that superposition coding strategy proposed by Cover[4] is indeed optimal here. The best known achievable strategy for general 2-receiver broadcast channels by Marton[15] combines the ideas of superposition coding and random binning. An outer bound by Korner and Marton[15] was the best known outer bound for over two decades. Recently Nair and El Gamal[19] introduced a new outer bound and showed that it is strictly better than Korner-Marton outer bound for the binary skew-symmetric channel (BSSC)[19]. Following this work, a series of outer bounds[16][13] were reported to the capacity region of the broadcast channels. However it is still unknown whether any of these outer bounds is strictly better than Nair-El Gamal outer bound.

1.2 Previous work on broadcast channel with degraded message sets

Consider a 2-receiver broadcast channel with messages (M_1, M_2) where the message M_2 is required by both receivers and message M_1 is only required by receiver Y_1. This means that the messages required by the receivers are *degraded*. This scenario was studied by Korner and Marton[12] and they showed the optimality of superposition coding region[4] that is the union of rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_2 &\leq I(U; Y_2) \\ R_1 + R_2 &\leq I(X; Y_1) \\ R_1 + R_2 &\leq I(X; Y_1|U) + I(U; Y_2) \end{aligned} \tag{1.1}$$

for some $p(u)p(x|u)p(y_1, y_2|x)$. More recently, some investigations were made to the degraded message set problem with 3 and more receivers[6][22][18]. The capacity regions for some classes of broadcast channels with 3 and more receivers were established in[6][22][20] by showing that the straightforward extension of the superposition coding region is optimal. In [18], Nair and El Gamal introduced an idea called indirect decoding and showed that the straightforward extension of the superposition coding region is suboptimal. However the new achievable regions obtained by indirect decoding become quite complex even for the simple case of 3 receivers (Y_1, Y_2, Y_3) with 2 degraded message sets where the message M_2 is required by all receivers and message M_1 is only required by receivers Y_1. So a natural question is the following,

Could one make progress on some new classes of broadcast channels where the achievable region given by indirect decoding coincides with that given by superposition coding?

1.3 Previous work on broadcast channel with less noisy sequence

Korner and Marton[11] introduced a class of 2-receiver broadcast channels called less noisy broadcast channel where a receiver Y_1 is said to be *less noisy* than receiver Y_2 if $I(U; Y_1) \geq I(U; Y_2)$

holds for all U such that $U \to X \to (Y_1, Y_2)$ forms a Markov chain (we denote this partial order less noisy relationship by $Y_1 \succeq Y_2$). The capacity region for this scenario was established (Proposition 3 in [11]) to be the union of rate pairs (R_1, R_2) satisfying

$$R_1 \le I(X; Y_1|U)$$
$$R_2 \le I(U; Y_2)$$

for some $p(u)p(x|u)p(y_1, y_2|x)$. The capacity region of k-receiver less noisy broadcast channel, i.e. $Y_1 \succeq Y_2 \succeq \cdots \succeq Y_k$, however, is unknown when $k \ge 3$. Thus an open issue is:

Is it possible to solve the capacity region for k-receiver less noisy broadcast channel, when $k \ge 3$?

1.4 Thesis organization

In this thesis, we aim at providing answers to the two fundamental issues mentioned in the previous sections. Here is a summary of results established in this thesis.

- For a class of broadcast channels shown in Figure 2.1 we provide the explicit expression of the bounds.(Section 2.3 and 2.4)

- For a class of broadcast channels shown in Figure 2.1 we establish that the inner bound is tight and outer bound is weak.(Section 2.5)

- We prove that superposition coding is optimal for a class of broadcast channels with a sequence of less noisy receivers. This class contains the 3-receiver less noisy broadcast channel, thus resolving its capacity region.(Section 3.1)

- We show a general superposition coding achievable region for k-receiver broadcast channels with degraded message requirement.(Section 4.1)

- For a k-receiver broadcast channel with two unmatched degraded components, and degraded message sets where receiver $Y_s, s \in \{1, \cdots, k\}$ requires messages $(M_s, \cdots, M_k)$, we establish that superposition coding is indeed optimal. (Section 4.1)

□ **End of chapter.**

Chapter 2

3-receiver BC with degraded message sets

In this chapter we consider a broadcast channel with 3 receivers and 2 messages (M_0, M_1) where two of the three receivers need to decode messages (M_0, M_1) while the remaining one just needs to decode the message M_0. The capacity region for this class of broadcast channels is an open problem and the best known inner and outer bounds are presented below.

2.1 Existing bounds

We obtain the following inner bound using superposition coding.

Bound 1. *The union of the following set of rate pairs* (R_0, R_1)

satisfying

$$\begin{aligned}
R_0 &\le I(U;Y_3)\\
R_1 &\le \min\{I(X;Y_1|U), I(X;Y_2|U)\}\\
R_0 + R_1 &\le \min\{I(X;Y_1), I(X;Y_2)\}
\end{aligned}$$

over all pairs of random variables (U, X) *such that* $U \to X \to (Y_1, Y_2, Y_3)$ *forms a Markov chain constitutes an inner bound to the capacity region.*

We obtain the following outer bound in a similar fashion as the traditional outer bounds obtained for the 2-receiver broadcast channels [8, 15, 19].

Bound 2. *The union over the set of rate pairs* (R_0, R_1) *satisfying*

$$\begin{aligned}
R_0 &\le \min\{I(U_1;Y_3), I(U_2;Y_3)\}\\
R_0 + R_1 &\le \min\{I(U_1;Y_3) + I(X;Y_1|U_1),\\
&\qquad I(U_2;Y_3) + I(X;Y_2|U_2)\}\\
R_0 + R_1 &\le \min\{I(X;Y_1), I(X;Y_2)\}
\end{aligned}$$

over all possible choices of random variables (U_1, U_2, X) *such that* $(U_1, U_2) \to X \to (Y_1, Y_2, Y_3)$ *forms a Markov chain constitutes an outer bound for this scenario.*

The identification $U_{1i} = (M_0, Y_1^{i-1}, Y_{3,i+1}^n)$ and $U_{2i} = (M_0, Y_2^{i-1}, Y_{3,i+1}^n)$ suffices to obtain this outer bound.

Remark 1. *It is also possible to include the constraint*

$$R_0 \leq \min\{I(U_1; Y_1), I(U_2; Y_2)\}$$

into the outer bound. However, it is quite straightforward to show that the region obtained by adding this inequality is identical to the bound we presented.

Bounds 1 and 2 are tight in all of the following special cases,

- Receiver Y_1 is a *less noisy* receiver than Y_3 and Y_2 is a *less noisy* receiver than Y_3 [6, 18],
- Y_3 is a deterministic function of X,
- Y_1 is a *more capable* receiver than Y_2 (or vice-versa),
- Y_3 is a *more capable* receiver than Y_2 (or Y_1).

The last two cases are very straightforward and the proof is omitted. When Y_3 is a deterministic function of X, note that it is not difficult to show that the following region is obtained by taking the convex closure of the regions given by setting (i) $U = Y_3$ and (ii) $U = \emptyset$ in Bound 1.

$$R_0 \leq H(Y_3)$$
$$R_0 + R_1 \leq \min\{I(X; Y_1), I(X; Y_2)\}$$

This clearly forms an outer bound and thus gives the capacity region.

2.2 A class of channels

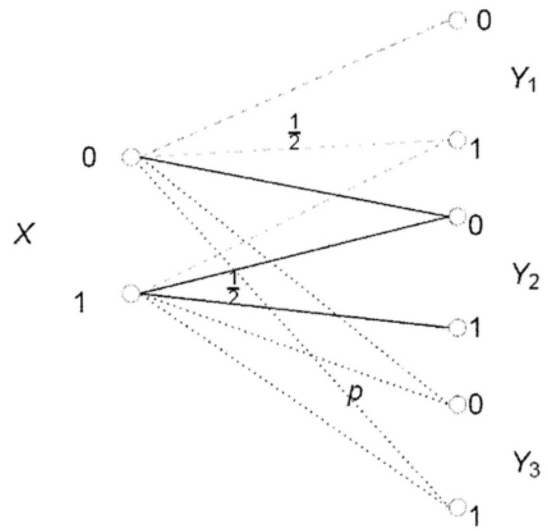

Figure 2.1: A class of 3-receiver broadcast channels

One class of channels that does not fall into any of the above cases is the following channel shown in Figure 2.1 below. The channel $X \to (Y_1, Y_2)$ represents a BSSC channel and the channel $X \to Y_3$ represents a *binary symmetric* (BSC) with crossover probability p, with $0 \leq p \leq \frac{1}{2}$.

Remark 2. *An interesting observation is the use of BSSC to obtain a class of broadcast channels where, as we shall see, the* *inner and outer bounds do not match and the inner bound is in fact tight.*

2.3 Evaluation of the inner bound

In the evaluation of the inner bound, we divide the range $0 \leq p \leq \frac{1}{2}$ into two regions, $0 \leq p \leq p_{max}$ and $p_{max} \leq p \leq \frac{1}{2}$, where $p_{max} \in [0, \frac{1}{2}]$ is the unique solution of

$$1 - h(p) = h\left(\frac{1}{4}\right) - \frac{1}{2};$$

i.e. the value of p at which capacity of the BSC matches the term $\max_{p(x)} \min\{I(X; Y_1), I(X; Y_2)\}$. The numerical value of $p_{max} \approx 0.184$.

2.3.1 Evaluation of the inner bound, $0 \leq p \leq p_{max}$

In the region $0 \leq p \leq p_{max}$ it is straightforward to see that the inner bound reduces to the following region (obtained via a time-division between the two auxiliary channels: (i) $U = \emptyset$ and (ii) $U = X$, and in each case, setting $\mathrm{P}(X = 0) = 0.5$),

$$R_0 + R_1 \leq h\left(\frac{1}{4}\right) - \frac{1}{2},$$

which clearly matches the outer bound (Bound 2). *Thus for $0 \leq p \leq p_{max} \approx 0.184$, the inner and outer bounds match and thus give the capacity region.*

2.3.2 Evaluation of the inner bound, $p_{max} \leq p \leq \frac{1}{2}$

Let $\mathcal{U} = \{1, 2, ..., m\}$ and let $\mathrm{P}(U = i) = u_i$ and $\mathrm{P}(X = 0|U = i) = s_i$. Further, let

$$h(x) = -x \log_2 x - (1-x) \log_2 (1-x)$$

denote the binary entropy function.

Using these notations we have,

$$\begin{aligned}
I(U; Y_3) &= h(\sum_i u_i(s_i(1-p) + (1-s_i)p)) \\
&\quad - \sum_i u_i h(s_i(1-p) + (1-s_i)p) \\
I(X; Y_1|U) &= \sum_i u_i h(\frac{s_i}{2}) - \sum_i u_i s_i \\
I(X; Y_2|U) &= \sum_i u_i h(\frac{1-s_i}{2}) - \sum_i u_i(1-s_i) \\
I(X; Y_1) &= h(\sum_i \frac{u_i s_i}{2}) - \sum_i u_i s_i \\
I(X; Y_2) &= h(\sum_i \frac{u_i(1-s_i)}{2}) - \sum_i u_i(1-s_i).
\end{aligned}$$

Define $\tilde{\mathcal{U}} = \{1, 2, ..., m\} \times \{1, 2\}$, $\mathrm{P}(\tilde{U} = (i, 1)) = \frac{u_i}{2}$, $\mathrm{P}(X = 0|\tilde{U} = (i, 1)) = s_i$, $\mathrm{P}(\tilde{U} = (i, 2)) = \frac{u_i}{2}$, and $\mathrm{P}(X = 0|\tilde{U} = (i, 2)) = 1 - s_i$. This induces an $\tilde{X}$ with $P(\tilde{X} = 0) = \frac{1}{2}$. It is straightforward to see the following:

$$\begin{aligned}
I(\tilde{U};\tilde{Y}_3) &\geq I(U;Y_3) \\
I(\tilde{X};\tilde{Y}_1|\tilde{U}) &= I(\tilde{X};\tilde{Y}_2|\tilde{U}) = \frac{1}{2}(I(X;Y_1|U) + I(X;Y_2|U)) \\
&\geq \min\{I(X;Y_1|U), I(X;Y_2|U)\} \\
I(\tilde{X};\tilde{Y}_1) &= I(\tilde{X};\tilde{Y}_2) \geq \frac{1}{2}(I(X;Y_1) + I(X;Y_2))
\end{aligned}$$

From this it follows that for every U replacing U by $\tilde{U}$ leads to a larger achievable region. Hence to evaluate Bound 1, it suffices to maximize over all auxiliary random variables of the form U defined by: $\mathcal{U} = \{1, 2, ..., m\} \times \{1, 2\}$, $\mathrm{P}(U = (i, 1)) = \frac{u_i}{2}$, $\mathrm{P}(X = 0|U = (i, 1)) = s_i$, $\mathrm{P}(U = (i, 2)) = \frac{u_i}{2}$, and $\mathrm{P}(X = 0|U = (i, 2)) = 1 - s_i$.

Under this notation we have the following expression for the rate region given in Bound 1,

$$\begin{aligned}
R_0 &\leq I(U;Y_3) \\
&= h(\frac{1}{2}) - \sum_i u_i h(s_i(1-p) + (1-s_i)p), \\
R_1 &\leq \min\{I(X;Y_1|U), I(X;Y_2|U)\} \\
&= \sum_i \frac{u_i}{2}\left(h(\frac{s_i}{2}) + h(\frac{1-s_i}{2})\right) - \frac{1}{2}, \\
R_0 + R_1 &\leq \min\{I(X;Y_1), I(X;Y_2)\} \\
&= h(\frac{1}{4}) - \frac{1}{2}.
\end{aligned}$$

Using the symmetry of the function $h(x) = h(1-x)$ we note that

$$h(s_i(1-p)+(1-s_i)p) = h((1-s_i)(1-p)+s_ip)$$

and thus the above region is constant under the transformation $s_i \to 1 - s_i$, implying we can restrict s_i to take values only in $0 \leq s_i \leq \frac{1}{2}$.

Before we proceed to determine the boundary of this region, we prove the following lemma.

2.3.3 An inequality for a class of functions

Lemma 1. *Let $f(x)$ and $g(x)$ be two non-negative and strictly increasing functions that are differentiable in the region $x \in [x_1, x_2]$. Further assume that $\frac{f^{(1)}(x)}{g^{(1)}(x)}$ is a decreasing function, where $f^{(1)}(x)$ and $g^{(1)}(x)$ denote the derivatives of the function. Given any $u, 0 \leq u \leq 1$, let x_{int} be uniquely defined according to $f(x_{int}) = uf(x_1) + (1-u)f(x_2)$. Then the following holds,*

$$g(x_{int}) \leq ug(x_1) + (1-u)g(x_2).$$

Proof. We have $u(f(x_{int})-f(x_1)) = (1-u)(f(x_2)-f(x_{int}))$, and we wish to show that $u(g(x_{int})-g(x_1)) \leq (1-u)(g(x_2)-g(x_{int}))$. Since all the terms are positive, this reduces to showing

$$\frac{f(x_{int}) - f(x_1)}{g(x_{int}) - g(x_1)} \geq \frac{f(x_2) - f(x_{int})}{g(x_2) - g(x_{int})}.$$

However, this is immediate as shown below.

From the fact that $\frac{f^{(1)}(x)}{g^{(1)}(x)}$ is a decreasing function, we have

$$\frac{\int_{x_1}^{x_{int}} f^{(1)}(x)dx}{\int_{x_1}^{x_{int}} g^{(1)}(x)dx} \geq \frac{f^{(1)}(x_{int})}{g^{(1)}(x_{int})} \geq \frac{\int_{x_{int}}^{x_2} f^{(1)}(x)dx}{\int_{x_{int}}^{x_2} g^{(1)}(x)dx}$$

$\square$

Repeated applications of Lemma 1 leads to the following corollary - potentially of independent interest.

Corollary 1. *Let $f(x)$ and $g(x)$ be two non-negative and strictly increasing functions that are differentiable in the region $x \in [x_1, x_2]$. Further assume that $\frac{f^{(1)}(x)}{g^{(1)}(x)}$ is a decreasing function, where as before $f^{(1)}(x)$ and $g^{(1)}(x)$ denote the derivatives of the function. Given any $u_i \geq 0, \sum_i u_i = 1$, and $y_i \in [x_1, x_2]$, let x_{int} be uniquely defined according to $f(x_{int}) = \sum_i u_i f(y_i)$. Then the following holds*

$$g(x_{int}) \leq \sum_i u_i g(y_i).$$

2.3.4 Determining the boundary rate pairs

We use Corollary 1 to determine the boundary of the region. We make the following identifications, let $f(x) = h(\frac{x}{2}) + h(\frac{1-x}{2}) - 1$, and $g(x) = h(x(1-p) + (1-x)p)$. Observe that $f(x)$ and $g(x)$ are increasing differentiable functions in the region $[0, \frac{1}{2}]$.

Claim 1. *For $\frac{1}{6} \le p \le \frac{1}{2}$, the ratio of the derivatives $\frac{f^{(1)}(x)}{g^{(1)}(x)}$ is a decreasing function.*

The proof of this fact is found in Appendix A.

(Numerical simulations indicate that this is true for $p_{min} \le p \le \frac{1}{2}$ for $p_{min} \approx 0.05$, but for the purposes of establishing the inner bound clearly this region of p suffices, as $\frac{1}{6} \le p_{max} \approx 0.184$).

Proposition 1. *For $\frac{1}{6} \le p \le \frac{1}{2}$, the function $\Phi(y) = g(f^{-1}(y)), 0 \le y \le 2h(\frac{1}{4}) - 1$, is convex in y.*

The proof of this proposition is given in Appendix B. This is very similar to Mrs. Gerber's lemma [25].

Now let s_{int} be defined according to

$$h(\frac{s_{int}}{2}) + h(\frac{1-s_{int}}{2}) = \sum_i u_i \left(h(\frac{s_i}{2}) + h(\frac{1-s_i}{2}) \right).$$

Then from Corollary 1, for $p_{min} \le p \le \frac{1}{2}$ we have

$$\begin{aligned} &h\big(\frac{1}{2}\big) - \sum_i u_i h(s_i(1-p) + (1-s_i)p) \\ &\quad \le h\big(\frac{1}{2}\big) - h(s_{int}(1-p) + (1-s_{int})p). \end{aligned}$$

This implies that the optimal auxiliary channel $U \to X$ is a BSC with a cross-over probability s and $\mathrm{P}(U=0) = \frac{1}{2}$. Thus for $p_{max} \le p \le \frac{1}{2}$, the boundary is characterized by the pair of points of the form,

$$\begin{aligned}R_0 &= 1 - h(s(1-p) + (1-s)p),\\ R_1 &= \min\left\{\frac{1}{2}\left(h(\frac{s}{2}) + h(\frac{1-s}{2}) - 1\right),\right.\\ &\left. h(\frac{1}{4}) - \frac{3}{2} + h(s(1-p) + (1-s)p)\right\},\end{aligned} \tag{2.1}$$

for $0 \leq s \leq \frac{1}{2}$. The second term in R_1 comes from taking into account the sum rate constraint,

$$R_0 + R_1 \leq h(\frac{1}{4}) - \frac{1}{2}.$$

A simple calculation shows that for $p_o \leq p \leq \frac{1}{2}$ one can ignore the sum rate constraint, where $p_o = \frac{\sqrt{3}-1}{2\sqrt{3}} \approx 0.211$. This p_o corresponds to the smallest value of p where the convex region characterized by the pairs

$$\begin{aligned}R_0 &= 1 - h(s(1-p) + (1-s)p),\\ R_1 &= \frac{1}{2}\left(h(\frac{s}{2}) + h(\frac{1-s}{2}) - 1\right).\end{aligned}$$

has a slope of -1 at the point $(R_0, R_1) = \left(0, h(\frac{1}{4}) - \frac{1}{2}\right)$.

Therefore the inner bound has three different expressions:

- $0 \leq p \leq p_{max}$: the inner bound reduces to $R_0 + R_1 \leq h(\frac{1}{4}) - \frac{1}{2}$,

- $p_{max} \leq p \leq p_o$: the inner bound is given by equation (2.1) where all inequalities are necessary,

- $p_o \leq p \leq \frac{1}{2}$: the inner bound is characterized by pair of points of the form

$$R_0 = 1 - h(s(1-p) + (1-s)p),$$
$$R_1 = \frac{1}{2}\left(h(\frac{s}{2}) + h(\frac{1-s}{2}) - 1\right).$$

2.4 Evaluation of the outer bound

The evaluation of the outer bound (i.e. Bound 2) follows roughly similar lines to that of Bound 1. Firstly, we show that we can restrict ourselves to $p(x)$ such that $\mathrm{P}(X=0) = \frac{1}{2}$ and Bound 2 reduces to the union of rate pairs satisfying

Bound 3.

$$R_0 \leq I(U_1; Y_3)$$
$$R_0 + R_1 \leq I(U_1; Y_3) + I(X; Y_1|U_1)$$
$$R_0 + R_1 \leq I(X; Y_1) = h(\frac{1}{4}) - \frac{1}{2}.$$

over all (U_1, X) *such that* $\mathrm{P}(X = 0) = \frac{1}{2}$ *and* $U_1 \rightarrow X \rightarrow (Y_1, Y_2, Y_3)$ *forms a Markov chain.*

Then we will show that the boundary of Bound 3 can be realized by binary U_1 and $0 < s_1 < s_2 = 1$, where $\mathrm{P}(X = 0|U_1 = i) = s_i, i \in (1, 2)$ (i.e. the optimal auxiliary channel $U_1 \rightarrow X$ is a Z channel), thus providing an explicit expression of Bound

2. However since we are interested in the fact that the bounds differ, so finally we will show that the boundary of Bound 2 yields a strictly larger region than Bound 1.

Remark 3. *Restricting to* $\mathrm{P}(X = 0) = \frac{1}{2}$ *it is clear that the region described by Bound 3 is at least as large as the one described by Bound 2. For the reverse direction, given a* U_1 *one can construct a* U_2 *by setting* $\mathrm{P}(U_2 = i) = \mathrm{P}(U_1 = i)$ *and* $\mathrm{P}(X = 0|U_2 = i) = 1 - \mathrm{P}(X = 0|U_1 = i)$ *and the region described in Bound 2 by this triple* (U_1, U_2, X) *will match the region in Bound 3. Note that the existence of the triple* (U_1, U_2, X) *is guaranteed by the consistency of distribution of* X *in* (U_1, X) *and* (U_2, X). *For instance, one can first generate* X *and then generate* U_1 *and* U_2 *conditionally independent of* X.

2.4.1 Restricting the marginal distribution of X

Given any triple of random variables (U_1, U_2, X), we construct a related triple of random variables $(\tilde{U}_1, \tilde{U}_2, \tilde{X})$ with $\mathrm{P}(X = 0) = \frac{1}{2}$ such that the rate pairs described by $(\tilde{U}_1, \tilde{U}_2, \tilde{X})$ dominate the rate pairs described by (U_1, U_2, X). Further for this triple $(\tilde{U}_1, \tilde{U}_2, \tilde{X})$ the region described by Bounds 2 and 3 are identical.

Consider an independent binary random variable W such that $\mathrm{P}(W = 0) = \frac{1}{2}$. Now set $\mathrm{P}(\bar{U}_1 = i, \bar{U}_2 = j|W = 0) = \mathrm{P}(U_1 =$

$i, U_2 = j)$ and $\mathrm{P}(\bar{U}_1 = j, \bar{U}_2 = i|W = 1) = \mathrm{P}(U_1 = i, U_2 = j)$. Further set $\mathrm{P}(X = 0|\bar{U}_1 = i, \bar{U}_2 = j, W = 0) = \mathrm{P}(X = 0|U_1 = i, U_2 = j)$ and $\mathrm{P}(X = 0|\bar{U}_1 = j, \bar{U}_2 = i, W = 1) = 1 - \mathrm{P}(X = 0|U_1 = i, U_2 = j)$. Now, define $\tilde{U}_1 = (W, \bar{U}_1), \tilde{U}_2 = (W, \bar{U}_2)$ and $\tilde{X}$ to be the induced distribution on X. It is straightforward to check the following,

$$\begin{aligned}
\mathrm{P}(\tilde{X} = 0) &= \frac{1}{2}, \\
I(\tilde{U}_1; \tilde{Y}_3) &= I(\tilde{U}_2; \tilde{Y}_3) \\
&= 1 - \frac{1}{2}\left(H(Y_3|U_1) + H(Y_3|U_2)\right) \\
&\geq \min\{I(U_1; Y_3), I(U_2; Y_3)\}, \\
I(\tilde{X}; \tilde{Y}_1|\tilde{U}_1) &= I(\tilde{X}; \tilde{Y}_2|\tilde{U}_2) \\
&= \frac{1}{2}\left(I(X; Y_1|U_1) + I(X; Y_2|U_2)\right) \\
&\geq \min\{I(X; Y_1|U_1), I(X; Y_2|U_2)\}, \\
I(\tilde{X}; \tilde{Y}_1) &= I(\tilde{X}; \tilde{Y}_2) \\
&\geq \frac{1}{2}\left(I(X; Y_1) + I(X; Y_2)\right) \\
&\geq \min\{I(X; Y_1), I(X; Y_2)\}.
\end{aligned}$$

This construction establishes that we can restrict ourselves to $\mathrm{P}(X = 0) = \frac{1}{2}$.

2.4.2 Optimal auxiliary channel

For the evaluation of Bound 2 one needs to take the union of the following rate pairs satisfying

$$\begin{aligned} R_0 &\leq I(U_1;Y_3) \\ R_1 &\leq I(X;Y_1|U_1) \\ R_0+R_1 &\leq I(X;Y_1) = h(\frac{1}{4}) - \frac{1}{2} \end{aligned}$$

over all (U_1, X) such that $\mathrm{P}(X=0)=\frac{1}{2}$ and $U_1 \to X \to (Y_1, Y_2, Y_3)$ forms a Markov chain.

Let $\mathcal{U}_1 = \{1, 2, ..., m\}$, $\mathrm{P}(U_1 = i) = u_i$ and $\mathrm{P}(X = 0|U_1 = i) = s_i$ such that $\sum_i u_i s_i = \frac{1}{2}$. Under this notation we have,

$$\begin{aligned} R_0 &\leq f_1 = 1 - \sum_i u_i h(s_i * p), \\ R_1 &\leq f_2 = \sum_i u_i h(\frac{s_i}{2}) - \frac{1}{2}, \\ R_0 + R_1 &\leq f_3 = h(\frac{1}{4}) - \frac{1}{2}. \end{aligned} \tag{2.2}$$

The following theorem provides an explicit expression of the boundary of Bound 2.

Theorem 1. *For a class of broadcast channels, as shown in Figure 2.1, the boundary of Bound 2 is realized by binary U_1, and further $0 < s_1 < s_2 = 1$ (i.e. the optimal auxiliary channel $U_1 \to X$ is a Z channel).*

Before we proceed to prove this theorem, we present and prove some preliminary results.

Let $J(x) = \log\frac{1-x}{x}$, $\alpha(x) = -J(x*p)$, $\beta(x) = -\frac{h(s_2*p)-h(x*p)}{(1-2p)(s_2-x)}$, $\gamma(x) = -J(\frac{x}{2})$ and $\delta(x) = -\frac{h(\frac{s_2}{2})-h(\frac{x}{2})}{\frac{1}{2}(s_2-x)}$. Under this notation we prove the following Claims 2 and 3.

Claim 2. *For $0 < s \leq s_2$, the ratio $\frac{\gamma(t)-\gamma(s)}{\alpha(t)-\alpha(s)}$ is a strictly decreasing function of t, $t \in [s, s_2]$.*

The proof is given in Appendix C.

Claim 3. *For $0 < s \leq s_2$, the ratio $\frac{\beta^{(1)}(s)}{\delta^{(1)}(s)}$ is a strictly increasing function and $\frac{\gamma^{(1)}(s)}{\delta^{(1)}(s)}$ is a strictly decreasing function.*

The proof is given in Appendix D.

We are now ready to prove Theorem 1. Our technique of proof is similar to the Karush-Kuhn-Tucker (KKT) conditions for the occurrence of a local maxima/minima, and could possibly be transformed directly into KKT conditions. Essentially, for any $U_1 \to X$ not satisfying our theorem, we will produce a valid perturbation direction in which all the quantities of interest to us remain non-decreasing (in fact at least one of them will be strictly increasing). Hence the only valid points on the boundary have to satisfy the theorem.

Proof. Without loss of generality, we can assume that any (u_i, s_i) satisfy $0 \leq s_1 < s_2 < \cdots < s_k \leq 1$.

Consider any (u_i, s_i) that contains a particular class of (u_1, u_2, s_1, s_2) such that $0 < u_1 + u_2 \leq 1$ and $0 < s_1 < s_2 < 1$, in fact, we will perturb u_1, u_2, s_1, s_2 such that we keep $u_1 + u_2$ constant. We rewrite region (2.2) with constraints $\sum_i u_i s_i = \frac{1}{2}$ and $\sum_i u_i = 1$ as follows

$$
\begin{aligned}
f_1 &= 1 - u_1 h(s_1 * p) - u_2 h(s_2 * p) - \sum_{i \geq 3} u_i h(s_i * p) \\
f_2 &= u_1 h(\frac{s_1}{2}) + u_2 h(\frac{s_2}{2}) + \sum_{i \geq 3} u_i h(\frac{s_i}{2}) - \frac{1}{2} \\
f_3 &= u_1 s_1 + u_2 s_2 + \sum_{i \geq 3} u_i s_i - \frac{1}{2} \\
f_4 &= u_1 + u_2 + \sum_{i \geq 3} u_i - 1
\end{aligned}
$$

The partial derivatives of the four functions with respect to u_1, u_2, s_1, s_2 will lead to the following matrix

$$
\begin{bmatrix}
-h(s_1 * p) & -h(s_2 * p) & -(1-2p)J(s_1 * p) & -(1-2p)J(s_2 * p) \\
h(\frac{s_1}{2}) & h(\frac{s_2}{2}) & \frac{1}{2}J(\frac{s_1}{2}) & \frac{1}{2}J(\frac{s_2}{2}) \\
s_1 & s_2 & 1 & 1 \\
1 & 1 & 0 & 0
\end{bmatrix}
$$

So if we can show for some $\varepsilon_1, \varepsilon_2, \varepsilon_3$ and ε_4 such that

$$
\begin{bmatrix}
-h(s_1 * p) & -h(s_2 * p) & -(1-2p)J(s_1 * p) & -(1-2p)J(s_2 * p) \\
h(\frac{s_1}{2}) & h(\frac{s_2}{2}) & \frac{1}{2}J(\frac{s_1}{2}) & \frac{1}{2}J(\frac{s_2}{2}) \\
s_1 & s_2 & 1 & 1 \\
1 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4
\end{bmatrix}
=
\begin{bmatrix}
> 0 \\ > 0 \\ 0 \\ 0
\end{bmatrix}
\tag{2.3}
$$

then there exists a perturbation direction such that the first two terms f_1 and f_2 are strictly increasing while the remaining terms f_3 and f_4 keep unchanged. Elementary manipulations of (2.3): setting $Col1 = Col1 - Col2, Col3 = Col3 - Col4$, and then setting $Col1 = Col1 - (s_1 - s_2)Col4$ one can see that it suffices to show that the determinant of

$$\begin{bmatrix} h(s_2 * p) - h(s_1 * p) - (1-2p)(s_2 - s_1)J(s_2 * p) & (1-2p)(J(s_2 * p) - J(s_1 * p)) \\ h(\frac{s_1}{2}) - h(\frac{s_2}{2}) + \frac{1}{2}(s_2 - s_1)J(\frac{s_2}{2}) & \frac{1}{2}(J(\frac{s_1}{2}) - J(\frac{s_2}{2})) \end{bmatrix}$$

is non-zero. Indeed we will show

$$\frac{J(s_1 * p) - J(s_2 * p)}{\frac{h(s_2*p)-h(s_1*p)}{(1-2p)(s_2-s_1)} - J(s_2 * p)} < \frac{J(\frac{s_1}{2}) - J(\frac{s_2}{2})}{\frac{h(\frac{s_2}{2})-h(\frac{s_1}{2})}{\frac{1}{2}(s_2-s_1)} - J(\frac{s_2}{2})} \tag{2.4}$$

Recalling the definitions, $\alpha(x) = -J(x*p)$, $\beta(x) = -\frac{h(s_2*p)-h(x*p)}{(1-2p)(s_2-x)}$, $\gamma(x) = -J(\frac{x}{2})$ and $\delta(x) = -\frac{h(\frac{s_2}{2})-h(\frac{x}{2})}{\frac{1}{2}(s_2-x)}$, (2.4) reduces to showing

$$\frac{\alpha(s_2) - \alpha(s_1)}{\beta(s_2) - \beta(s_1)} < \frac{\gamma(s_2) - \gamma(s_1)}{\delta(s_2) - \delta(s_1)} \tag{2.5}$$

From Claim 2, one can see that (2.5) holds as $s_1 \to s_2$ and hence we have

$$\frac{\alpha^{(1)}(s)}{\beta^{(1)}(s)} < \frac{\gamma^{(1)}(s)}{\delta^{(1)}(s)}$$

Since all the terms involved in the above inequality are positive, (2.5) reduces to showing

$$\frac{\int_{s_1}^{s_2} \frac{\beta^{(1)}(s)\gamma^{(1)}(s)}{\delta^{(1)}(s)} ds}{\int_{s_1}^{s_2} \beta^{(1)}(s) ds} < \frac{\int_{s_1}^{s_2} \gamma^{(1)}(s) ds}{\int_{s_1}^{s_2} \delta^{(1)}(s) ds}$$

From FKG inequality [9], we see that it suffices to show that for $0 < s \leq s_2$, the ratio $\frac{\beta^{(1)}(s)}{\delta^{(1)}(s)}$ is a strictly increasing function and $\frac{\gamma^{(1)}(s)}{\delta^{(1)}(s)}$ is a strictly decreasing function. However, this is immediate as shown in Claim 3.

Conclusion 1: There is no U such that $s_1 < s_2 < \cdots < s_k < 1$ lie on the boundary for $k \geq 2$.

Remark 4. *It is also possible to include the case $s_1 = 0$, however, it is straightforward to show that the result obtained by substituting $s_1 = 0$ with corresponding $\varepsilon_3 > 0$ (Note that this is the only valid direction) into (2.3) is identical to the conclusion we presented.*

Thus either:

1. $u_1 = u_2 = \frac{1}{2}$, and $s_1 = 0, s_2 = 1$ which corresponds to the point on the boundary $R_0 = \min\left\{1 - h(p), h(\frac{1}{4}) - \frac{1}{2}\right\}, R_1 = 0$. Or

2. $u_1 + u_2 = 1$, and $0 < s_1 < s_2 = 1$.

Combining the above analysis with conclusion 1, we see that it suffices to consider binary U_1 and $0 < s_1 < s_2 = 1$ (i.e. the optimal auxiliary channel $U_1 \to X$ is a Z channel) to characterize the boundary of Bound 2. $\square$

2.4.3 Comparison with the inner bound

To show that Bound 2 gives a larger region, let

$$\mathrm{P}(U_1 = 1) = u, \mathrm{P}(U_1 = 2) = 1 - u,$$
$$\mathrm{P}(X = 0|U_1 = 1) = 1, \mathrm{P}(X = 0|U_1 = 2) = s.$$

where $s = \frac{0.5-u}{1-u}$ for $0 \leq u \leq 0.5$. Substituting this choice into Bound 3, we obtain the boundary *Region A* given by,

$$\begin{aligned} R_0 &\leq 1 - (1-u)h(s * p) - uh(p), \\ R_1 &\leq (1-u)h\big(\frac{s}{2}\big) - \frac{1}{2} + u, \\ R_0 + R_1 &\leq h\big(\frac{1}{4}\big) - \frac{1}{2}. \end{aligned}$$

As in Bound 1, the sum rate constraint $R_0 + R_1 \leq h\big(\frac{1}{4}\big) - \frac{1}{2}$ can be ignored in Region A for $p_m \leq p \leq \frac{1}{2}$ where $p_m \approx 0.2384$ solves

$$h(p) = 1 - \frac{1}{2}\log\big(\frac{4}{3}\big).$$

Figure 2.2 plots Region A and Bound 1 for $p = \frac{1}{4}$. Observe that Region A is strictly larger than Bound 1, and hence Bounds 1 and 2 do not match for the scenario shown in Figure 2.1. This implies the following corollary.

Corollary 2. *There exists a class of channels, given in Figure 2.1, for which Bounds 1 and 2 do not match.*

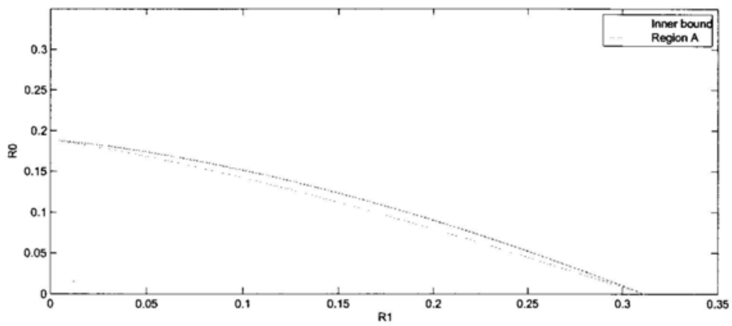

Figure 2.2: Comparing Bound 1 and Region A for $p = \frac{1}{4}$

2.5 Revisiting the outer bound

Theorem 2. *The capacity region of the broadcast channel in Figure 2.1 is the set of rate pairs* (R_0, R_1) *satisfying*

$$\begin{aligned} R_0 &\leq I(U;Y_3) \\ R_1 &\leq \min\{I(X;Y_1|U), I(X;Y_2|U)\} \\ R_0 + R_1 &\leq \min\{I(X;Y_1), I(X;Y_2)\} \end{aligned}$$

over all pairs of random variables (U, X) *such that* $U \to X \to (Y_1, Y_2, Y_3)$.

Proof. Let $\pi : \{0,1\} \mapsto \{0,1\}; \pi(0) = 1, \pi(1) = 0$.

Consider an ϵ-codebook $\{x^n_{m_0,m_1}, 1 \leq m_0 \leq 2^{nR_0}, 1 \leq m_1 \leq 2^{nR_1}, \mathcal{A}_{m_0,m_1} \subseteq \mathcal{Y}_1^n, \mathcal{B}_{m_0,m_1} \subseteq \mathcal{Y}_2^n, \mathcal{C}_{m_0} \subseteq \mathcal{Y}_3^n\}$, where the disjoint sets $\mathcal{A}_{m_0,m_1}, \mathcal{B}_{m_0,m_1}, \mathcal{C}_{m_0}$ represent the decoding maps. From the skew symmetry of the channels $X \to (Y_1, Y_2)$ and the symmetry in channel $X \to Y_3$, it is clear that $\{\pi(x^n_{m_0,m_1}), 1 \le m_0 \le 2^{nR_0}, 1 \le m_1 \le 2^{nR_1}, \pi(\mathcal{B}_{m_0,m_1}) \subseteq \mathcal{Y}_1^n, \pi(\mathcal{A}_{m_0,m_1}) \subseteq \mathcal{Y}_2^n, \pi(\mathcal{C}_{m_0}) \subseteq \mathcal{Y}_3^n\}$ represents a valid ϵ-codebook as well.

From these two codes, construct a pseudo-codebook (with error bounded by $\frac{1}{2} + \epsilon$) and size $2^{nR_0} \times 2^{nR_1+1}$ as follows: The codewords are indexed by $x^n_{m_0,(m_1,b)}$ where $b = 0, 1$. When $b = 0$ the codeword $x^n_{m_0,(m_1,b=0)} = x^n_{m_0,m_1}$ and when $b = 1$, we have $x^n_{m_0,(m_1,b=1)} = \pi(x^n_{m_0,m_1})$. The decoding maps for this codebook are created as follows: If $y_1^n \in \mathcal{A}_{m_0^1,m_1^1} \cap \pi(\mathcal{B}_{m_0^2,m_1^2})$ then the receiver chooses one of the two message pairs $(m_0^1, m_1^1), (m_0^2, m_1^2)$ with equal probability. Otherwise it picks the message pair corresponding to the unique set $\mathcal{A}_{m_0^1,m_1^1}$ or $\pi(\mathcal{B}_{m_0^2,m_1^2})$ that it belongs to. A similar decoding strategy applies for receivers Y_2 and Y_3 as well.

The key feature is the symmetry of the codebook. If $x^n \in \mathbb{C}$ then $\pi(x^n) \in \mathbb{C}$ and corresponds to the same message M_0.

Now observe that $H(M_0, M_1|Y_1^n) \le H(M_0, M_1, b|Y_1^n) \le 1 + H(M_0, M_1|Y_1^n, b) = 1 + n(R_0 + R_1)\epsilon_n$. Therefore we obtain the same outer bound (Bound 2) using Fano's inequality and identification of the auxiliary random variables as before.

In particular, the identifications of the auxiliary random variables remain the following,

$$U_{1i} = (M_0, Y_{31}^{\,i-1}, Y_{1i+1}^{\,n}), U_{2i} = (M_0, Y_{31}^{\,i-1}, Y_{2i+1}^{\,n})$$

Now for the skew-symmetric channels and a symmetric codebook, observe that

$$\begin{aligned}
&\mathrm{P}\Big(M_0 = m_0, Y_{31}^{\,i-1} = y_{31}^{i-1}, Y_{1i+1}^{\,n} = y_{1i+1}^{n}, X_i = x_i\Big) \\
&= \sum_{x_1^n \backslash x_i} \mathrm{P}\Big(M_0 = m_0, X_1^n = x_1^n, Y_{31}^{\,i-1} = y_{31}^{i-1}, Y_{1i+1}^{\,n} = y_{1i+1}^{n}\Big) \\
&\overset{(a)}{=} \sum_{x_1^n \backslash x_i} \mathrm{P}\Big(M_0 = m_0, X_1^n = x_1^n\Big) \prod_{j=1}^{i-1} \mathrm{P}(Y_{3j} = y_{3j} | X_j = x_j) \\
&\quad \times \prod_{k=i+1}^{n} \mathrm{P}(Y_{1k} = y_{1k} | X_k = x_k) \\
&\overset{(b)}{=} \sum_{x_1^n \backslash x_i} \mathrm{P}\Big(M_0 = m_0, X_1^n = \pi(x_1^n)\Big) \prod_{j=1}^{i-1} \mathrm{P}(Y_{3j} = \pi(y_{3j}) | X_j = \pi(x_j)) \\
&\quad \times \prod_{k=i+1}^{n} \mathrm{P}(Y_{2k} = \pi(y_{1k}) | X_k = \pi(x_k)) \\
&= \sum_{x_1^n \backslash x_i} \mathrm{P}\Big(M_0 = m_0, X_1^n = \pi(x_1^n), Y_{31}^{\,i-1} = \pi(y_{31}^{\,i-1}), Y_{2i+1}^{\,n} = \pi(y_{1i+1}^{\,n})\Big) \\
&\overset{(c)}{=} \mathrm{P}\Big(M_0 = m_0, Y_{31}^{\,i-1} = \pi(y_{31}^{i-1}), Y_{2i+1}^{\,n} = \pi(y_{1i+1}^{\,n}), X_i = \pi(x_i)\Big).
\end{aligned}$$

Here (a) follows from the discrete memoryless property of the channel; and (b) follows from (i) symmetry of the code, (ii) symmetry of the channel $X \to Y_3$ with respect to $\pi(\cdot)$, and (iii)

the skew symmetry between receivers Y_1, Y_2 i.e.

$$\mathrm{P}(Y_2 = \pi(y)|X = \pi(x)) = \mathrm{P}(Y_1 = y|X = x);$$

and (c) is a consequence of $\pi(\cdot)$ being a bijection.

Therefore the random variables (U_1, X) and (U_2, X) are identical up to re-labeling. Since the mutual information and entropy do not depend on the labeling, it follows that

$$\begin{aligned} I(U_1; Y_3) &= I(U_2; Y_3) \\ I(X; Y_2|U_1) &= I(X; Y_2|U_2). \end{aligned}$$

Therefore we obtain the following revised outer bound.

Bound 4. *The union over the set of rate pairs* (R_0, R_1) *satisfying*

$$\begin{aligned} R_0 &\leq I(U_1; Y_3) \\ R_0 + R_1 &\leq \min\{I(U_1; Y_3) + I(X; Y_1|U_1), \\ &\qquad I(U_1; Y_3) + I(X; Y_2|U_1)\} \\ R_0 + R_1 &\leq \min\{I(X; Y_1), I(X; Y_2)\} \end{aligned}$$

over all possible choices of random variables (U_1, X) *such that* $U_1 \to X \to (Y_1, Y_2, Y_3)$ *forms a Markov chain constitutes an outer bound for this channel.*

It is straightforward to see (using the boundary points) that Bound 4 matches Bound 1 and forms the capacity region. □

Remark 5. *This technique of proof can be extended to other skew-symmetric channels as well, i.e one for which such a* $\pi(\cdot)$ *exists.*

□ **End of chapter.**

Chapter 3

k-receiver less noisy BC

In this chapter, we consider a discrete memoryless broadcast channel with k receivers $Y_1, \cdots, Y_k$.

Definition 1. *A receiver Y_s is said to be less noisy[11] than receiver Y_t if $I(U;Y_s) \geq I(U;Y_t)$ for all $U \to X \to (Y_s, Y_t)$.*

We denote this relationship (partial-order) by $Y_s \succeq Y_t$.

Remark 6. *Observe that this partial order only depends on marginal distributions $p(y_s|x)$ and $p(y_t|x)$.*

Definition 2. *A k-receiver broadcast channel is said to belong to class $\mathcal{C}$ if there exists $k-1$ virtual receivers $V_1, ..., V_{k-1}$ satisfying:*

- *$X \to V_1 \to ... \to V_{k-1}$ forms a Markov chain and*
- *The following "interleaved" less noisy condition holds:*

$$Y_1 \succeq V_1 \succeq Y_2 \succeq \cdots Y_{k-1} \succeq V_{k-1} \succeq Y_k. \tag{3.1}$$

This class contains some interesting sequences of less noisy receivers as mentioned below.

1. A sequence of degraded receivers, i.e. $X \to Y_1 \to \ldots \to Y_k$; set $V_i = Y_{i+1}$,
2. A sequence of "nested" less noisy receivers, i.e. $Y_s \succeq (Y_{s+1}, \ldots, Y_k)$; set $V_i = (Y_{i+1}, \ldots, Y_k)$,
3. A 3-receiver less noisy sequence, i.e. $Y_1 \succeq Y_2 \succeq Y_3$; set $V_1 = V_2 = Y_2$.

We present a couple of results before we prove the capacity region for the class $\mathcal{C}$ with private message requirements.

Fact 1. *From the definition of less noisy scenario, by conditioning on U_2, it follows that whenever $(U_1, U_2) \to X \to (Y_s, Y_t)$ forms a Markov chain*

$$I(U_1; Y_t|U_2) \leq I(U_1; Y_s|U_2). \tag{3.2}$$

Lemma 2. *If a receiver $Y_s \succeq Y_t$ then*[1]

$$\begin{aligned} I(Y_{t,1}^{i-1}; Y_{s,i}|U) &\leq I(Y_{t,1}^{i-2}, Y_{s,i-1}; Y_{s,i}|U) \leq \cdots \\ &\leq I(Y_{t,1}, Y_{s,2}^{i-1}; Y_{s,i}|U) \leq I(Y_{s,1}^{i-1}; Y_{s,i}|U) \end{aligned}$$

whenever $(U, Y_{t,1}^{p-1}, Y_{s,p+1}^{i-1}, Y_{s,i}) \to X_p \to (Y_{t,p}, Y_{s,p})$ forms a Markov chain for $1 \leq p \leq i-1$.

[1]The notation $Y_{t,p}^{i}$ denotes $(Y_{t,p}, Y_{t,p+1}, \ldots, Y_{t,i})$.

Proof. For all p such that $1 \leq p \leq i-1$, observe that

$$\begin{aligned}
& I(Y_{t,1}^{p}, Y_{s,p+1}^{i-1}; Y_{s,i}|U) \\
& = I(Y_{t,1}^{p-1}, Y_{s,p+1}^{i-1}; Y_{s,i}|U) + I(Y_{t,p}; Y_{s,i}|U, Y_{t,1}^{p-1}, Y_{s,p+1}^{i-1}) \\
& \leq I(Y_{t,1}^{p-1}, Y_{s,p+1}^{i-1}; Y_{s,i}|U) + I(Y_{s,p}; Y_{s,i}|U, Y_{t,1}^{p-1}, Y_{s,p+1}^{i-1}) \\
& = I(Y_{t,1}^{p-1}, Y_{s,p}^{i-1}; Y_{s,i}|U)
\end{aligned}$$

where the inequality follows from (3.2) as $(U, Y_{t,1}^{p-1}, Y_{s,p+1}^{i-1}, Y_{s,i}) \to X_p \to (Y_{t,p}, Y_{s,p})$ forms a Markov chain for $1 \leq p \leq i-1$. $\square$

Lemma 3. *For any broadcast channel belonging to class $\mathcal{C}$, the following Markov chain*

$$V_{s,1}^{i-1} \to V_{s-1,1}^{i-1} \to X^n, Y_1^n, \cdots, Y_k^n, M_1, \cdots, M_k$$

holds for $1 \leq s \leq k-1$; (set $V_0 = X$).

Proof. From Remark 6 we have that the less noisy ordering only depends on the marginals. Hence the probability distribution $p(x^n, y_1^n, \cdots, y_k^n, v_1^n, \cdots, v_{k-1}^n)$ can be factorized as

$$\begin{aligned}
& p(x^n, y_1^n, \cdots, y_k^n, v_1^n, \cdots, v_{k-1}^n) \\
& = \prod_{i=1}^{n} p(x_i|x_{i-1}) p(y_{1,i}, \cdots, y_{k,i}, v_{1,i}, \cdots, v_{k-1,i}|x_i) \\
& = \prod_{i=1}^{n} p(x_i|x_{i-1}) p(y_{1,i}, \cdots, y_{k,i}|x_i) p(v_{1,i}, \cdots, v_{k-1,i}|x_i) \\
& = \prod_{i=1}^{n} p(x_i|x_{i-1}) p(y_{1,i}, \cdots, y_{k,i}|x_i) p(v_{1,i}|x_i) \prod_{j=2}^{k-1} p(v_{j,i}|v_{j-1,i})
\end{aligned}$$

where the first inequality is due to the fact that the channel is DMC without feedback, the second follows from the assumptions that the less noisy scenario depends on the marginal distributions, and the last follows from the Markov chain $X \to V_1 \to \cdots \to V_{k-1}$. Then given this structure, we have

$$V_{s,1}^{i-1} \to V_{s-1,1}^{i-1} \to X^n, Y_1^n, \cdots, Y_k^n, M_1, \cdots, M_k$$

is Markov. □

3.1 Main result

Theorem 3. *For any broadcast channel belonging to class $\mathcal{C}$ with independent message requirements, the capacity region is the set of rate tuples $R_1, ..., R_k$ satisfying*

$$R_s \leq I(U_s; Y_s | U_{s+1}), \quad 1 \leq s \leq k$$

where $U_1 = X, U_{k+1} = \emptyset$ and the sequence $U_k \to U_{k-1} \cdots \to U_2 \to X \to (Y_1, ..., Y_k)$ forms a Markov chain.

3.2 Proof of Theorem 3

The achievability is straightforward using superposition coding and jointly typical decoding. We shall refer the reader to [5] for details. Since $Y_s \succeq Y_j, s \leq j \leq k$, the receiver Y_s successively decodes messages M_j (equivalently the sequences U_j^n) from $j = k$ to $j = s$. Each step is correct with high probability since

$$\begin{aligned} R_j &= I(U_j; Y_j | U_{j+1}) - \epsilon \\ &\leq I(U_j; Y_s | U_{j+1}) - \epsilon, \end{aligned}$$

when $s \leq j \leq k$. Therefore the rate tuples given in Theorem 3 are indeed achievable.

Now we show the converse. Let $M_{s+1}^k = (M_{s+1}, ..., M_k)$. Using Fano's inequality, observe that for $1 \leq s \leq k$

$$\begin{aligned} nR_s &\leq I(M_s; Y_{s,1}^n | M_{s+1}^k) + n\epsilon_n \\ &= \sum_{i=1}^n I(M_s; Y_{s,i} | M_{s+1}^k, Y_{s,1}^{i-1}) + n\epsilon_n \\ &= \sum_{i=1}^n \big(I(M_s, Y_{s,1}^{i-1}; Y_{s,i} | M_{s+1}^k) - I(Y_{s,1}^{i-1}; Y_{s,i} | M_{s+1}^k) \big) + n\epsilon_n \\ &\overset{(a)}{\leq} \sum_{i=1}^n \big(I(M_s, Y_{s,1}^{i-1}; Y_{s,i} | M_{s+1}^k) - I(V_{s,1}^{i-1}; Y_{s,i} | M_{s+1}^k) \big) + n\epsilon_n \\ &\overset{(b)}{\leq} \sum_{i=1}^n \big(I(M_s, V_{s-1,1}^{i-1}; Y_{s,i} | M_{s+1}^k) - I(V_{s,1}^{i-1}; Y_{s,i} | M_{s+1}^k) \big) + n\epsilon_n \\ &\overset{(c)}{=} \sum_{i=1}^n \big(I(M_s, V_{s-1,1}^{i-1}, V_{s,1}^{i-1}; Y_{s,i} | M_{s+1}^k) - I(V_{s,1}^{i-1}; Y_{s,i} | M_{s+1}^k) \big) + n\epsilon_n \\ &= \sum_{i=1}^n I(M_s, V_{s-1,1}^{i-1}; Y_{s,i} | M_{s+1}^k, V_{s,1}^{i-1}) + n\epsilon_n \\ &= \sum_{i=1}^n I(U_{s,i}; Y_{s,i} | U_{s+1,i}) + n\epsilon_n, \end{aligned}$$

where $U_{s,i} = (M_s^k, V_{s-1,1}^{i-1})$. Here the inequalities (a) and (b) follows from Lemma 2 and the less noisy condition $V_{s-1} \succeq Y_s \succeq V_s$. The equality (c) follows from the Lemma 3.

Define Q to be a uniform random variable taking values in $\{1, .., n\}$ and independent of all other random variables. As usual, we set $U_s = (U_{s,Q}, Q)$ and $X = X_Q$. Since $X \rightarrow V_1 \rightarrow \cdots \rightarrow V_{k-1}$ is a Markov chain it follows that $U_k \rightarrow U_{k-1} \rightarrow \cdots \rightarrow U_2 \rightarrow X$ forms a Markov chain as well. This completes the proof of the converse.

□ **End of chapter.**

Chapter 4

Product of two unmatched BC with k receivers and degraded message sets

The product of two unmatched k-receiver degraded broadcast channels is defined as a broadcast channel $(\mathscr{X}, p(y_1, \cdots, y_k|x), \mathscr{Y}_1 \times \cdots \times \mathscr{Y}_k)$ where

$$\mathscr{X} = \bar{\mathscr{X}} \times \widetilde{\mathscr{X}}, \mathscr{Y}_i = \bar{\mathscr{Y}_i} \times \widetilde{\mathscr{Y}_i}, 1 \leq i \leq k \tag{4.1}$$

$$p(y_1, \cdots, y_k|x) = p(\bar{y}_1|\bar{x}) \cdots p(\bar{y}_k|\bar{y}_{k-1}) p(\widetilde{y}_k|\widetilde{x}) \cdots p(\widetilde{y}_1|\widetilde{y}_2)$$

This channel was initially studied by Poltyrev[21] for the $k = 2$ case where the subchannels $\bar{X} \to \bar{Y}_1 \to \bar{Y}_2$ and $\widetilde{X} \to \widetilde{Y}_2 \to \widetilde{Y}_1$ form Markov chains. Unlike the classical degraded broadcast channel[4], the overall channel is non-degraded since the order of degradation is different across the subchannels. Poltyrev[21] established the capacity region for the $k = 2$ case with private message requirements. The result was subsequently generalized by El Gamal[7] for the same case with general message requirements (i.e. a common message intended for both receivers and a private message intended for each receiver). Extending the El Gamal's result to more than two receivers has been a long standing open problem. Given this dismal situation, one may make progress by studying the problem with degraded message sets where the idea of superposition coding turns out to be transparent. Motivated by this, we study the channel defined in (4.1) with degraded message sets where receiver Y_s, $s \in \{1, \cdots, k\}$ requires messages $(M_s, \cdots, M_k)$. The main questions we consider are the following:

- What is the achievable rate region for this channel?
- Is this region optimal?

In this chapter we establish tight characterization of the superposition coding region for the channel defined in (4.1) with degraded message sets (Theorem 5). The achievability part is a consequence of the general superposition coding region established in Theorem 4 for any broadcast channels with degraded message sets, while in the converse we follow a rather standard information-theoretical argument by making use of degraded conditions.

4.1 Statement of results

Theorem 4. *For a k-receiver broadcast channel with degraded message sets where the receiver Y_s, $s \in \{1, \cdots, k\}$ requires messages $(M_s, \cdots, M_k)$, the following rate tuples $R_1, \cdots, R_k$ are achievable using superposition coding.*

$$\sum_{j \geq s} R_j \leq I(U_s; Y_s|U_{i_1}) + I(U_{i_1}; Y_{i_1}|U_{i_2}) + \cdots + I(U_{i_{l-1}}; Y_{i_{l-1}}|U_{i_l}) + I(U_{i_l}; Y_{i_l})$$

where $\{i_1, \cdots, i_l\} \in \mathcal{I}_s = \big\{\{i_1, \cdots, i_l\} : s = i_1 < i_2 < \cdots < i_l \leq k, 1 \leq l \leq k\big\}$ and the sequence $U_k \to \cdots \to U_2 \to U_1 = X \to (Y_1, \cdots, Y_k)$ forms a Markov chain.

Theorem 5. *For the product of two unmatched k-receiver degraded broadcast channels (defined in (4.1)) with degraded message sets where the receiver Y_s, $s \in \{1, \cdots, k\}$ requires messages $(M_s, \cdots, M_k)$, the capacity region is actually the superposition coding region and the region can be further simplified to the* *union of rate tuples* $R_1, \cdots, R_k$ *satisfying*

$$\sum_{j \geq s} R_j \leq I(U_s; \bar{Y}_s | U_{s+1}) + \cdots + I(U_{s+l'-2}; \bar{Y}_{s+l'-2} | U_{s+l'-1})$$
$$+ I(U_{s+l'-1}; \bar{Y}_{s+l'-1}) + I(\widetilde{X}; \widetilde{Y}_{s+l'-1}), \quad 1 \leq l' \leq k - s + 1 \tag{4.2}$$

where the sequences

$$U_k \to \cdots \to U_2 \to U_1 = \bar{X} \to \bar{Y}_1 \to \bar{Y}_2 \to \cdots \to \bar{Y}_k \tag{4.3}$$
$$\widetilde{X} \to \widetilde{Y}_k \to \cdots \to \widetilde{Y}_2 \to \widetilde{Y}_1$$

form Markov chains.

Remark 7. *Observe that for the* $k = 2$ *case, the capacity region is the union of rate pairs* (R_1, R_2) *satisfying*

$$R_2 \leq I(U; \bar{Y}_2) + I(\widetilde{X}; \widetilde{Y}_2)$$
$$R_1 + R_2 \leq I(\bar{X}; \bar{Y}_1) + I(\widetilde{X}; \widetilde{Y}_2)$$
$$R_1 + R_2 \leq I(\bar{X}; \bar{Y}_1 | U) + I(U; \bar{Y}_2) + I(\widetilde{X}; \widetilde{Y}_2)$$

for some $p(u)p(\bar{x}|u)p(\widetilde{x})p(y_1, y_2|x)$*. This can be derived directly from [7] and also coincides with the region given by* (1.1) *with the setting* $U' = (U, \widetilde{X}), X' = (\bar{X}, \widetilde{X}), Y_1' = (\bar{Y}_1, \widetilde{Y}_1)$ *and* $Y_2' = (\bar{Y}_2, \widetilde{Y}_2)$ *where* $U \to \bar{X} \to \bar{Y}_1 \to \bar{Y}_2$ *and* $\widetilde{X} \to \widetilde{Y}_2 \to \widetilde{Y}_1$ *are Markov and further* $\widetilde{X}$ *is independent of* $(U, \bar{X})$*.*

4.2 Proof of Theorem 4

The proof of achievability follows from superposition coding and jointly typical decoding. Let $\mathcal{I}_s = \Big\{\{i_1, \cdots, i_l\} : s = i_1 < i_2 < \cdots < i_l \leq k, 1 \leq l \leq k\Big\}$. Further let

$$A_s = \min_{\{i_1,\cdots,i_l\}\in\mathcal{I}_s} \left(I(U_s;Y_s|U_{i_1}) + \sum_{j=1}^{l-1} I(U_{i_j};Y_{i_j}|U_{i_{j+1}}) + I(U_{i_l};Y_{i_l})\right)$$

Define a sequence $\{B_k, B_{k-1}, \cdots, B_1\}$ such that $B_p = \min_{q\leq s}\{A_s\}$ for all $s \in \{1, \cdots, k\}$ to guarantee positive rate tuples $R_1, \cdots, R_k$. We will show that the following corner points

$$R_p = B_p - \sum_{j\geq p+1} R_j - \frac{\epsilon}{2^p} \tag{4.4}$$

are achievable for some $\epsilon > 0$.

Code generation: Fix $p(u_k)p(u_{k-1}|u_k)\cdots p(x|u_2)$. Generate $2^{nR_k} = 2^{n(B_k-\frac{\epsilon}{2^k})}$ independent codewords $U_k^n(m_k)$, $m_k \in \{1, 2, \cdots, 2^{nR_k}\}$ according to $\prod_{i=1}^{n} p(u_{k,i})$. For each $U_k^n(m_k)$, generate $2^{nR_{k-1}} = 2^{n(B_{k-1}-R_k-\frac{\epsilon}{2^{k-1}})}$ independent codewords $U_{k-1}^n(m_k, m_{k-1})$, $m_{k-1} \in \{1, \cdots, 2^{nR_{k-1}}\}$ according to $\prod_{i=1}^{n} p(u_{k-1,i}|u_{k,i}(m_k))$. Subsequently for each codeword $\big(U_{p+1}^n(m_k, \cdots, m_{p+1})\big), 2 \leq p \leq k-2$, generate $2^{nR_p} = 2^{n(B_p-\sum_{j\geq p+1} R_j-\frac{\epsilon}{2^p})}$ independent codewords $U_p^n(m_k, \cdots, m_p)$, $m_p \in \{1, 2, \cdots, 2^{nR_p}\}$ according to $\prod_{i=1}^{n} p(u_{p,i}|u_{p+1,i}(m_k, \cdots, m_{p+1}))$. Finally for each codeword $\big(U_2^n(m_k, \cdots, m_2)\big)$, generate $2^{nR_1} =$

$2^{n(B_1-\sum\limits_{j\geq 2}R_j-\frac{\epsilon}{2})}$ independent codewords $X^n(m_k,\cdots,m_1)$, $m_1\in\{1,2,\cdots,2^{nR_1}\}$ according to $\prod\limits_{i=1}^{n}p(x_i|u_{2,i}(m_k,\cdots,m_2))$.

Encoding: To send the messages $(m_1,m_2,\cdots,m_k)\in[1,2^{nR_1}]\times[1,2^{nR_2}]\times\cdots\times[1,2^{nR_k}]$, the sender sends $X^n(m_1,m_2,\cdots,m_k)$.

Decoding and analysis of error probability: For $1\leq p\leq k$, the receiver Y_p declares that $(m_k,\cdots,m_p)$ is sent if it is the unique rate tuple such that Y_p^n is jointly typical with the sequences $\big(U_k^n(m_k),\cdots,U_p^n(m_k,\cdots,m_p)\big)$, i.e. $(U_k^n(m_k),\cdots,U_p^n(m_k,\cdots,m_p),Y_p^n)\in A_\epsilon^n$. If there are none such tuple or more than one tuple, an error is declared. Without loss of generality, we assume that a message tuple $(m_k,m_{k-1},\cdots,m_p)=(1,1,\cdots,1)$ is sent. We first define the event

$$E_{m_k,\cdots,m_p}=\{(U_k^n(m_k),\cdots,U_p^n(m_k,\cdots,m_p),Y_p^n)\in A_\epsilon^n\}$$

Then the possible error events can be classified into the following two groups:

- The received codewords are not jointly typical with the transmitted sequences, i.e. $E_{1,1,\cdots,1}^c$.
- There exist some $(m_k,\cdots,m_p)\neq(1,\cdots,1)$ such that the codeword $X^n(m_k,\cdots,m_p)$ is jointly typical with Y_p^n.

Thus we can bound the probability of decoding error at receiver p as

$$
\begin{aligned}
P_e^{(n)} &= \mathrm{P}\left(E^c_{1,1,\cdots,1} \cup \cup_{(m_k,\cdots,m_p)\neq(1,\cdots,1)} E_{m_k,\cdots,m_p}\right) \\
&\le \mathrm{P}(E^c_{1,1,\cdots,1}) + \sum_{m_k\neq 1} \mathrm{P}(E_{m_k,\cdots}) + \sum_{m_{k-1}\neq 1} \mathrm{P}(E_{1,m_{k-1},\cdots}) \\
&\quad + \cdots + \sum_{m_p\neq 1} \mathrm{P}(E_{1,\cdots,1,m_p})
\end{aligned}
$$

Now by the LLN, the first term $\mathrm{P}(E^c_{1,1,\cdots,1}) \to 0$ as $n \to \infty$. Further by the packing lemma we have that for all a such that $p \le a \le k$, $\sum_{m_a\neq 1} \mathrm{P}(E_{m_{\{b:b>a\}}=1,m_a\neq 1,\cdots}) \to 0$ if

$$
\begin{aligned}
\sum_{j=p}^{a} R_j &\le I(U_p; Y_p|U_{a+1}, \cdots, U_k) \\
&= I(U_p; Y_p|U_{a+1}) \qquad (4.5)
\end{aligned}
$$

where $U_{k+1} = \emptyset$. Observe that

$$
\begin{aligned}
\sum_{j=p}^{a} R_j &= \sum_{j=p}^{k} R_j - \sum_{j=a+1}^{k} R_j \\
&= B_p - B_{a+1} + \frac{\epsilon}{2^{a+1}} - \frac{\epsilon}{2^p} \\
&\le \min\{A_1, \cdots, A_p\} - \min\{A_1, \cdots, A_p, A_{p+1}, \cdots, A_{a+1}\}
\end{aligned}
$$

Obviously (4.5) holds if $\min\{A_1, \cdots, A_p\} \le \min\{A_{p+1}, \cdots, A_{a+1}\}$. Thus it reduces to showing

$$
\min\{A_1, A_2, \cdots, A_p\} - \min\{A_{p+1}, \cdots, A_{a+1}\} \le I(U_p; Y_p|U_{a+1})
$$

Let $\min\{A_{p+1}, \cdots, A_{a+1}\} = A_{q+1}$. Obviously $p \le q \le a$. Observe that

$$\begin{aligned}
&\min\{A_1, A_2, \cdots, A_p\} - \min\{A_{p+1}, \cdots, A_{a+1}\} \\
&\le A_p - A_{q+1} \\
&= \min_{\{i_1,i_2,\cdots,i_l\}\in\mathcal{I}_p} \Big(I(U_p; Y_p|U_{i_1}) + \sum_{j=1}^{l-1} I(U_{i_j}; Y_{i_j}|U_{i_{j+1}}) + I(U_{i_l}; Y_{i_l})\Big) \\
&\quad - \min_{\{i_1,i_2,\cdots,i_l\}\in\mathcal{I}_{q+1}} \Big(I(U_{q+1}; Y_{q+1}|U_{i_1}) + \sum_{j=1}^{l-1} I(U_{i_j}; Y_{i_j}|U_{i_{j+1}}) + I(U_{i_l}; Y_{i_l})\Big) \\
&\le \min_{\{i_1,i_2,\cdots,i_l\}\in\mathcal{I}_p, i_2=q+1} \Big(I(U_p; Y_p|U_{i_1}) + \sum_{j=1}^{l-1} I(U_{i_j}; Y_{i_j}|U_{i_{j+1}}) + I(U_{i_l}; Y_{i_l})\Big) \\
&\quad - \min_{\{i_1,i_2,\cdots,i_l\}\in\mathcal{I}_{q+1}} \Big(I(U_{q+1}; Y_{q+1}|U_{i_1}) + \sum_{j=1}^{l-1} I(U_{i_j}; Y_{i_j}|U_{i_{j+1}}) + I(U_{i_l}; Y_{i_l})\Big) \\
&= I(U_p; Y_p|U_{q+1}) \\
&\le I(U_p; Y_p|U_{a+1})
\end{aligned}$$

where the last inequality follows from the following fact

$$\begin{aligned}
I(U_p; Y_p|U_{q+1}) &= I(U_p; Y_p|U_{q+1}, U_{a+1}) \\
&= I(U_p; Y_p|U_{a+1}) - I(U_{s+1}; Y_p|U_{a+1}) \\
&\le I(U_p; Y_p|U_{a+1})
\end{aligned}$$

Thus the receiver Y_p decodes the intended messages with arbitrarily small probability of error and hence (4.4) is achievable.

4.3 Proof of Theorem 5

4.3.1 Achievability

From Theorem 4 it is straightforward to see that given any distribution $p(u_1, \cdots, u_k)$ satisfying $U_k \to \cdots \to U_2 \to U_1 = \bar{X} \to (\bar{Y}_1, \cdots, \bar{Y}_k)$ and $\widetilde{X} \to (\widetilde{Y}_1, \cdots, \widetilde{Y}_k)$, the following rate tuples $R_1, \cdots, R_k$ are achievable.

$$\sum_{j \geq s} R_j \leq I(U_s; \bar{Y}_s | U_{i_1}) + I(U_{i_1}; \bar{Y}_{i_1} | U_{i_2}) + \cdots + I(U_{i_{l-1}}; \bar{Y}_{i_{l-1}} | U_{i_l}) + I(U_{i_l}; \bar{Y}_{i_l}) + I(\widetilde{X}; \widetilde{Y}_{i_l}) \tag{4.6}$$

where $\{i_1, \cdots, i_l\} \in \mathcal{I}_s = \big\{\{i_1, \cdots, i_l\} : s = i_1 < i_2 < \cdots < i_l \leq k, 1 \leq l \leq k\big\}$.

Therefore to show that the region described by (4.6) with the Markov condition (4.3) is equivalent to the region defined by (4.2) it suffices to show that all inequalities are redundant except for the case when $i_1 = s, i_{p+1} = i_p + 1$ for all $p \in \{1, 2, \cdots, l-1\}$. Indeed we will show

$$\begin{aligned} & I(U_{i_1}; \bar{Y}_{i_1} | U_{i_1+1}) + I(U_{i_1+1}; \bar{Y}_{i_1+1} | U_{i_1+2}) \\ & \qquad + \cdots + I(U_{i_1+l'-2}; \bar{Y}_{i_1+l'-2} | U_{i_1+l'-1}) \\ & \leq I(U_{i_1}; \bar{Y}_{i_1} | U_{i_2}) + I(U_{i_2}; \bar{Y}_{i_2} | U_{i_3}) + \cdots + I(U_{i_{l-1}}; \bar{Y}_{i_{l-1}} | U_{i_1+l'-1}) \end{aligned}$$

where $i_1 = s, i_l = i_1 + l' - 1$. This would be established by proving the following inequality

$$\sum_{j=i_p}^{i_{p+1}-1} I(U_j;\bar{Y}_j|U_{j+1}) \leq I(U_{i_p};\bar{Y}_{i_p}|U_{i_{p+1}})$$

for all $p \in \{1, 2, \cdots, l-1\}$.

However this is immediate as shown below.

It is trivial true if $i_{p+1} = i_p + 1$. Now let $i_{p+1} = i_p + q$, where $q \geq 2$. Observe that

$$\begin{aligned}
&\sum_{j=0}^{q-1} I(U_{i_p+j};\bar{Y}_{i_p+j}|U_{i_p+j+1}) - I(U_{i_p};\bar{Y}_{i_p}|U_{i_p+q}) \\
&= \sum_{j=1}^{q-1} I(U_{i_p+j};\bar{Y}_{i_p+j}|U_{i_p+j+1}) - I(U_{i_p+1};\bar{Y}_{i_p}|U_{i_p+q}) \\
&\leq \sum_{j=2}^{q-1} I(U_{i_p+j};\bar{Y}_{i_p+j}|U_{i_p+j+1}) - I(U_{i_p+2};\bar{Y}_{i_p+1}|U_{i_p+q}) \\
&\leq \cdots \\
&\leq I(U_{i_{p+q-1}};\bar{Y}_{i_{p+q-1}}|U_{i_p+q}) - I(U_{i_{p+q-1}};\bar{Y}_{i_{p+q-2}}|U_{i_p+q}) \\
&\leq 0.
\end{aligned}$$

where the first equality follows from the fact that $I(U_{i_p};\bar{Y}_{i_p}|U_{i_p+1}) = I(U_{i_p};\bar{Y}_{i_p}|U_{i_p+1}, U_{i_p+q}) = I(U_{i_p};\bar{Y}_{i_p}|U_{i_p+q}) - I(U_{i_p+1};\bar{Y}_{i_p}|U_{i_p+q})$ and the last inequality follows from the fact that $U_{i_p+q} \rightarrow \bar{Y}_{i_{p+q-2}} \rightarrow \bar{Y}_{i_{p+q-1}}$ forms a Markov chain. Thus completes the proof of achievability.

4.3.2 Converse

We now show the converse. Indeed we wish to show that given any sequence of $\left((2^{nR_1}, \cdots, 2^{nR_k}), n, \mathrm{P}_e\right)$ code with $\mathrm{P}_e \to 0$, the rate tuples $R_1, \cdots, R_k$ must satisfy (4.2) with the Markov condition (4.3).

Let $M_s^k = (M_s, ..., M_k)$. Further let $U_{j,i} = (M_j^k, \bar{Y}_{j,1}^{i-1})$. Using Fano's inequality, observe that for $l' \geq 2$,

$$
\begin{aligned}
n\sum_{j\geq s} R_j \leq & \sum_{j=s}^{s+l'-2} I(M_j; \bar{Y}_{j,1}^n, \widetilde{Y}_{j,1}^n | M_{j+1}^k) \\
& + I(M_{s+l'-1}^k; \bar{Y}_{s+l'-1,1}^n, \widetilde{Y}_{s+l'-1,1}^n) + n\epsilon_n \\
= & \sum_{j=s}^{s+l'-2} I(M_j; \widetilde{Y}_{j,1}^n | \bar{Y}_{j,1}^n, M_{j+1}^k) + I(M_{s+l'-1}^k; \widetilde{Y}_{s+l'-1,1}^n | \bar{Y}_{s+l'-1,1}^n) \\
& + \sum_{j=s}^{s+l'-2} I(M_j; \bar{Y}_{j,1}^n | M_{j+1}^k) + I(M_{s+l'-1}^k; \bar{Y}_{s+l'-1,1}^n) + n\epsilon_n
\end{aligned}
$$

Note that,

$$
\begin{aligned}
& \sum_{j=s}^{s+l'-2} I(M_j; \bar{Y}_{j,1}^n | M_{j+1}^k) + I(M_{s+l'-1}^k; \bar{Y}_{s+l'-1,1}^n) \\
= & \sum_{j=s}^{s+l'-2} \sum_{i=1}^{n} I(M_j; \bar{Y}_{j,i} | M_{j+1}^k, \bar{Y}_{j,1}^{i-1}) + \sum_{i=1}^{n} I(M_{s+l'-1}^k; \bar{Y}_{s+l'-1,i} | \bar{Y}_{s+l'-1,1}^{i-1}) \\
= & \sum_{j=s}^{s+l'-2} \sum_{i=1}^{n} I(M_j; \bar{Y}_{j,i} | M_{j+1}^k, \bar{Y}_{j,1}^{i-1}, \bar{Y}_{j+1,1}^{i-1}) + \sum_{i=1}^{n} I(M_{s+l'-1}^k; \bar{Y}_{s+l'-1,i} | \bar{Y}_{s+l'-1,1}^{i-1}) \qquad (4.7) \\
\leq & \sum_{j=s}^{s+l'-2} \sum_{i=1}^{n} I(M_j, \bar{Y}_{j,1}^{i-1}; \bar{Y}_{j,i} | M_{j+1}^k, \bar{Y}_{j+1,1}^{i-1}) + \sum_{i=1}^{n} I(M_{s+l'-1}^k, \bar{Y}_{s+l'-1,1}^{i-1}; \bar{Y}_{s+l'-1,i}) \\
= & \sum_{j=s}^{s+l'-2} \sum_{i=1}^{n} I(U_{j,i}; \bar{Y}_{j,i} | U_{j+1,i}) + \sum_{i=1}^{n} I(U_{s+l'-1,i}; \bar{Y}_{s+l'-1,i})
\end{aligned}
$$

And further,

$$
\begin{aligned}
&\sum_{j=s}^{s+l'-2} I(M_j;\widetilde{Y}_{j,1}^n|\bar{Y}_{j,1}^n, M_{j+1}^k) + I(M_{s+l'-1}^k;\widetilde{Y}_{s+l'-1,1}^n|\bar{Y}_{s+l'-1,1}^n) \\
&\le I(M_s^{s+l'-2},\bar{Y}_{s+1,1}^n,\cdots,\bar{Y}_{s+l'-3,1}^n;\widetilde{Y}_{s+l'-2,1}^n|M_{s+l'-1}^k,\bar{Y}_{s+l'-2,1}^n) \\
&\quad + I(M_{s+l'-1}^k;\widetilde{Y}_{s+l'-1,1}^n|\bar{Y}_{s+l'-1,1}^n) \\
&= I(M_s^{s+l'-2},\bar{Y}_{s+1,1}^n,\cdots,\bar{Y}_{s+l'-3,1}^n;\widetilde{Y}_{s+l'-2,1}^n|M_{s+l'-1}^k,\bar{Y}_{s+l'-2,1}^n,\bar{Y}_{s+l'-1,1}^n) \\
&\quad + I(M_{s+l'-1}^k;\widetilde{Y}_{s+l'-1,1}^n|\bar{Y}_{s+l'-1,1}^n) \qquad (4.8) \\
&\le I(M_s^{s+l'-2},\bar{Y}_{s,1}^n,\bar{Y}_{s+1,1}^n,\cdots,\bar{Y}_{s+l'-2,1}^n;\widetilde{Y}_{s+l'-1,1}^n|M_{s+l'-1}^k,\bar{Y}_{s+l'-1,1}^n) \\
&\quad + I(M_{s+l'-1}^k;\widetilde{Y}_{s+l'-1,1}^n|\bar{Y}_{s+l'-1,1}^n) \\
&= I(M_s^k,\bar{Y}_{s,1}^n,\bar{Y}_{s+1,1}^n,\cdots,\bar{Y}_{s+l'-2,1}^n;\widetilde{Y}_{s+l'-1,1}^n|\bar{Y}_{s+l'-1,1}^n) \\
&\le \sum_{i=1}^{n} I(\widetilde{X}_i;\widetilde{Y}_{s+l'-1,i})
\end{aligned}
$$

where the first inequality follows from the fact that for all $p \in \{0,1,\cdots,l'-3\}$ we have

$$
\begin{aligned}
&I(M_s^{s+p},\bar{Y}_{s,1}^n,\cdots,\bar{Y}_{s+p-1,1}^n;\widetilde{Y}_{s+p,1}^n|M_{s+p+1}^k,\bar{Y}_{s+p,1}^n) \\
&\quad + I(M_{s+p+1};\widetilde{Y}_{s+p+1,1}^n|M_{s+p+2}^k,\bar{Y}_{s+p+1,1}^n) \\
&= I(M_s^{s+p},\bar{Y}_{s,1}^n,\cdots,\bar{Y}_{s+p-1,1}^n;\widetilde{Y}_{s+p,1}^n|M_{s+p+1}^k,\bar{Y}_{s+p,1}^n,\bar{Y}_{s+p+1,1}^n) \\
&\quad + I(M_{s+p+1};\widetilde{Y}_{s+p+1,1}^n|M_{s+p+2}^k,\bar{Y}_{s+p+1,1}^n) \\
&\le I(M_s^{s+p},\bar{Y}_{s,1}^n,\cdots,\bar{Y}_{s+p,1}^n;\widetilde{Y}_{s+p,1}^n|M_{s+p+1}^k,\bar{Y}_{s+p+1,1}^n) \\
&\quad + I(M_{s+p+1};\widetilde{Y}_{s+p+1,1}^n|M_{s+p+2}^k,\bar{Y}_{s+p+1,1}^n) \\
&\le I(M_s^{s+p},\bar{Y}_{s,1}^n,\cdots,\bar{Y}_{s+p,1}^n;\widetilde{Y}_{s+p+1,1}^n|M_{s+p+1}^k,\bar{Y}_{s+p+1,1}^n) \\
&\quad + I(M_{s+p+1};\widetilde{Y}_{s+p+1,1}^n|M_{s+p+2}^k,\bar{Y}_{s+p+1,1}^n) \\
&= I(M_s^{s+p+1},\bar{Y}_{s,1}^n,\cdots,\bar{Y}_{s+p,1}^n;\widetilde{Y}_{s+p+1,1}^n|M_{s+p+2}^k,\bar{Y}_{s+p+1,1}^n)
\end{aligned}
$$

Combining equations (4.7) and (4.8) we obtain for $l' \geq 2$

$$n\sum_{j\geq s} R_j \leq \sum_{j=s}^{s+l'-2}\sum_{i=1}^{n} I(U_{j,i};\bar{Y}_{j,i}|U_{j+1,i}) + \sum_{i=1}^{n} I(U_{s+l'-1,i};\bar{Y}_{s+l'-1,i})$$
$$+\sum_{i=1}^{n} I(\widetilde{X}_i;\widetilde{Y}_{s+l'-1,i}) + n\epsilon_n$$

For $l' = 1$, it reduces to show the converse of

$$\sum_{j\geq s} R_j \leq I(U_s;\bar{Y}_s) + I(\widetilde{X};\widetilde{Y}_s)$$

However this is immediate by similarly using Fano's inequality,

$$\begin{aligned}
n\sum_{j\geq s} R_j &\leq I(M_s^k;\bar{Y}_{s,1}^n,\widetilde{Y}_{s,1}^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(M_s^k;\bar{Y}_{s,i}|\bar{Y}_{s,1}^{i-1}) + I(M_s^k;\widetilde{Y}_{s,1}^n|\bar{Y}_{s,1}^n) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(M_s^k,\bar{Y}_{s,1}^{i-1};\bar{Y}_{s,i}) + I(M_s^k,\bar{Y}_{s,1}^n;\widetilde{Y}_{s,1}^n) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(M_s^k,\bar{Y}_{s,1}^{i-1};\bar{Y}_{s,i}) + \sum_{i=1}^{n} I(\widetilde{X}_i;\widetilde{Y}_{s,i}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(U_{s,i};\bar{Y}_{s,i}) + \sum_{i=1}^{n} I(\widetilde{X}_i;\widetilde{Y}_{s,i}) + n\epsilon_n
\end{aligned}$$

Define Q to be a uniform random variable taking values in $\{1,..,n\}$ and independent of all other random variables. As usual, we set $U_s = (U_{s,Q},Q)$, $\widetilde{X} = \widetilde{X}_Q$, $\bar{Y}_s = \bar{Y}_{s,Q}$ and $\widetilde{Y}_s = \widetilde{Y}_{s,Q}$. This completes the proof of the converse.

□ **End of chapter.**

Chapter 5

Summary

In this thesis we have considered some very interesting open problems in the area of broadcast channel, in an attempt to understand the fundamental limits of this network. The main objective is to identify a new class of multi-user network based solutions to improve the existing bounds as well as to establish the capacity region for some new classes of broadcast channels. We have identified a simple example where we are able to explicitly evaluate the best known inner and outer bounds to show that the bounds do not match for a class of 3-receiver broadcast channels with 2 degraded message sets. Due to the difficulty in explicitly evaluating the bounds no such example was known previously. We have further shown that for a class of broadcast channels where the bounds differ, the inner bound is tight and outer bound is weak. This makes us better understand the deficiencies with the current techniques of establishing the outer bound.

We have also established the capacity region for a class of broadcast channels with a sequence of less noisy receivers. This generalizes the result about the optimality of superposition coding for k-receiver degraded broadcast channels. Indeed the optimality of the superposition coding has been unknown for $k(\geq 3)$-receiver less noisy broadcast channel since the mid 1970s. Our result have established the optimality of superposition coding for the case $k = 3$, as it can be shown to be in the new class we defined.

In addition we have solved the capacity region of the product of k-receiver degraded broadcast channel with degraded message sets where receiver $Y_s, s \in \{1, \cdots, k\}$ requires messages $(M_s, \cdots, M_k)$. In the process of proving the achievability, we showed a general superposition coding region for k-receiver broadcast channels with degraded message requirement.

In summary, we have provided promising answers to the two open questions we set out to answer in the introduction part of this thesis:

Question: *Could one make progress on some new classes of* *broadcast channels where the achievable region given by indirect decoding coincides with that given by superposition coding?*

Some classes of broadcast channels studied in this thesis showed that one can make significant progress on the scenario where the idea of superposition coding yields the best known achievable region scheme. The idea of this coding strategy turns out to be transparent and seems to be optimal for the channels with general degraded messages requirement.

Question: *Is it possible to solve the capacity region of* $k(\geq 3)$*-receiver less noisy broadcast channel is the straightforward extension of Korner and Marton result?*

It is possible but one has to find new ideas to solve the capacity region for this class of broadcast channels. For instance, the capacity of the case $k = 3$ was solved by using the idea of virtual receivers in the identification of the converse. However the characterization of capacity region is still open for $k(\geq 4)$-receiver less noisy broadcast channel.

□ **End of chapter.**

Appendix A

Proof of Claim 1

In this section we show that when $\frac{1}{6} \leq p \leq \frac{1}{2}$, the ratio $\frac{f^{(1)}(x)}{g^{(1)}(x)}$ is a decreasing function of $x, x \in [0, \frac{1}{2}]$. Recalling the definitions, $f(x) = h(\frac{x}{2}) + h(\frac{1-x}{2}) - 1$, and $g(x) = h(x * p)$. As $f(x)$ and $g(x)$ are strictly increasing in $x \in [0, \frac{1}{2}]$, it suffices to show that

$$\frac{f^{(2)}(x)}{f^{(1)}(x)} \leq \frac{g^{(2)}(x)}{g^{(1)}(x)}, \tag{A.1}$$

where $f^{(2)}(x), g^{(2)}(x)$ denote the second derivatives of the function.

Let $J(x) = \log \frac{1-x}{x}$ and $U(x) = x(1-x)$. Using this notation and substituting for the derivatives, (A.1) reduces to showing

$$\frac{J(x * p)U(x * p)}{1 - 2p} \geq \frac{2\Big(J\big(\frac{x}{2}\big) - J\big(\frac{1-x}{2}\big)\Big)}{\frac{1}{U(\frac{x}{2})} + \frac{1}{U(\frac{1-x}{2})}}. \tag{A.2}$$

Now observe that as $x \to \frac{1}{2}$ both $J(x * p)$ and $J\big(\frac{x}{2}\big) - J\big(\frac{1-x}{2}\big)$ tend to zero and all other terms remain positive. Thus we have an equality at $x = \frac{1}{2}$. To show the inequality for $x \in [0, \frac{1}{2}]$ it suffices to prove that the *derivative of the left hand side (L.H.S.)* of (A.2) is smaller than *derivative of the right hand side (R.H.S.)* of (A.2).

The derivative of the L.H.S. is given by

$$\frac{d}{dx}\frac{J(x*p)U(x*p)}{1-2p} = -1 + J(x*p)(1-2(x*p)).$$

Let us define $R(x)$ to be the derivative of the R.H.S., i.e.

$$\frac{d}{dx}\frac{2\Big(J(\frac{x}{2}) - J(\frac{1-x}{2})\Big)}{\frac{1}{U(\frac{x}{2})} + \frac{1}{U(\frac{1-x}{2})}} = R(x).$$

We wish to show that

$$-1 + J(x*p)(1-2(x*p)) \leq R(x), \tag{A.3}$$

for all $\frac{1}{6} \leq p \leq \frac{1}{2}$ and $x \in [0, \frac{1}{2}]$. Given any $x \in [0, \frac{1}{2}]$, observe that $J(x*p)(1-2(x*p))$ is a decreasing function of p for $0 \leq p \leq \frac{1}{2}$. Thus establishing (A.3) for $p = \frac{1}{6}$ suffices.

Let $S(x) = -1 + J(x*\frac{1}{6})(1-2(x*\frac{1}{6}))$. Figure A.1 plots $S(x)$ and $R(x)$.

Thus we have $S(x) \leq R(x)$ for $0 \leq x \leq \frac{1}{2}$. This completes the proof of Claim 1.

□ **End of chapter.**

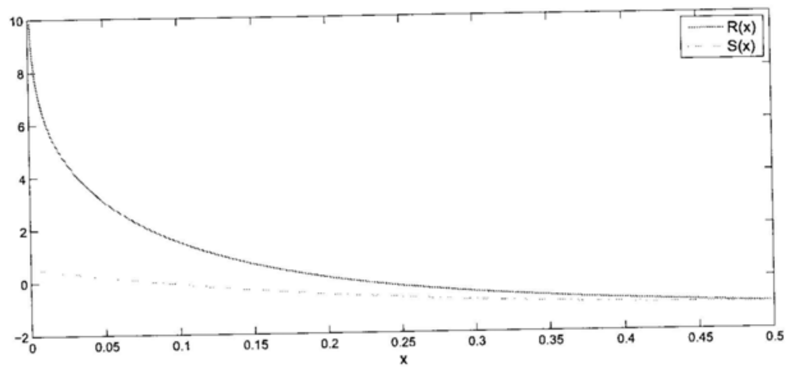

Figure A.1: Comparing $R(x)$ and $S(x)$

Appendix B

Proof of Proposition 1

In this section we show that $\frac{d^2\Phi(y)}{dy^2} \geq 0$, for $0 < y < 2h(\frac{1}{4}) - 1$ and $\frac{1}{6} \leq p \leq \frac{1}{2}$. As $f(x)$ is a strictly increasing function in $x \in [0, \frac{1}{2}]$, thus we can define $x = f^{-1}(y) \in [0, \frac{1}{2}]$ unambiguously. Observe that

$$\frac{d}{dy}\Phi(y) = \frac{g^{(1)}(f^{-1}(y))}{f^{(1)}(f^{-1}(y))}$$

Thus

$$\begin{aligned}
\frac{d^2\Phi(y)}{dy^2} &= \frac{g^{(2)}(f^{-1}(y))f^{(1)}(f^{-1}(y)) - g^{(1)}(f^{-1}(y))f^{(2)}(f^{-1}(y))}{(f^{(1)}(f^{-1}(y)))^3} \\
&= \frac{g^{(1)}(f^{-1}(y))f^{(1)}(f^{-1}(y))}{(f^{(1)}(f^{-1}(y)))^3}\left(\frac{g^{(2)}(f^{-1}(y))}{g^{(1)}(f^{-1}(y))} - \frac{f^{(2)}(f^{-1}(y))}{f^{(1)}(f^{-1}(y))}\right) \\
&= \frac{g^{(1)}(x)f^{(1)}(x)}{(f^{(1)}(x))^3}\left(\frac{g^{(2)}(x)}{g^{(1)}(x)} - \frac{f^{(2)}(x)}{f^{(1)}(x)}\right).
\end{aligned}$$

Since both $f^{(1)}(x) > 0$ and $g^{(1)}(x) > 0$ for $x \in (0, \frac{1}{2})$, it will suffice to establish that the term in brackets is ≥ 0 for $\frac{1}{6} \leq p \leq \frac{1}{2}$.

However, this is immediate as shown in Claim 1.

□ **End of chapter.**

Appendix C

Proof of Claim 2

In this section we show that when $0 \leq p \leq \frac{1}{2}$, the ratio $\frac{\gamma(t)-\gamma(s)}{\alpha(t)-\alpha(s)}$ is a strictly decreasing in $t \in [s, s_2]$, where $\gamma(x) = -J(\frac{x}{2})$ and $\alpha(x) = -J(x * p)$.

Let $F(s,t,p) = \ln(J(\frac{s}{2}) - J(\frac{t}{2})) - \ln(J(s*p) - J(t*p))$, we wish to show

$$\frac{\partial}{\partial t}F(s,t,p) = \frac{\frac{1}{U(\frac{t}{2})}}{\int_s^t \frac{1}{U(\frac{v}{2})}dv} - \frac{\frac{1}{U(t*p)}}{\int_s^t \frac{1}{U(v*p)}dv} < 0$$

where $U(x) = x(1-x)$. Since all the terms are positive, this reduces to showing that the ratio $\frac{U(\frac{v}{2})}{U(v*p)}$ is a strictly increasing in $v \in (0,1)$.

Let $K(v,p) = \ln(\frac{v}{2}) + \ln(1-\frac{v}{2}) - \ln(v*p) - \ln(1-v*p)$, we wish to show that

$$\frac{\partial}{\partial v}K(v,p) = \frac{1}{v} - \frac{1}{2-v} - (\frac{1-2p}{v*p} - \frac{1-2p}{1-v*p}) > 0$$

Further let $G(v,p) = \frac{1-2p}{v*p} - \frac{1-2p}{1-v*p}$, observe that

$$\begin{aligned}\frac{\partial}{\partial p}G(v,p) &= \frac{-2}{v*p} + \frac{2}{1-v*p} - \frac{(1-2p)(1-2v)}{(v*p)^2} \\ &\quad - \frac{(1-2p)(1-2v)}{(1-v*p)^2}\end{aligned}$$

If $0 < v \leq \frac{1}{2}$, then $\frac{\partial}{\partial p}G(v,p) \leq 0$ and we have

$$G(v,0) = \frac{1}{v} - \frac{1}{1-v} < \frac{1}{v} - \frac{1}{2-v}$$

If $\frac{1}{2} \leq v < 1$, then $\frac{\partial}{\partial p}G(v,p) \geq 0$ and we have

$$G(v,\frac{1}{2}) = 0 < \frac{1}{v} - \frac{1}{2-v}$$

Thus $\frac{\partial}{\partial v}K(v,p) > 0$ for $v \in (0,1)$. This completes the proof of Claim 2.

□ **End of chapter.**

Appendix D

Proof of Claim 3

In this section we show that when $0 < s \leq s_2$, the ratio $\frac{\beta^{(1)}(s)}{\delta^{(1)}(s)}$ is a strictly increasing function and $\frac{\gamma^{(1)}(s)}{\delta^{(1)}(s)}$ is a strictly decreasing function, where $\beta(s) = -\frac{h(s_2*p)-h(s*p)}{(1-2p)(s_2-s)}$, $\gamma(s) = -J(\frac{s}{2})$ and $\delta(s) = -\frac{h(\frac{s_2}{2})-h(\frac{s}{2})}{\frac{1}{2}(s_2-s)}$. As all the terms $\beta(s)$, $\gamma(s)$ and $\delta(s)$ are strictly increasing in $s \in (0, s_2]$, it suffices to show that

$$\frac{\beta^{(2)}(s)}{\beta^{(1)}(s)} > \frac{\delta^{(2)}(s)}{\delta^{(1)}(s)}, \tag{D.1}$$

and

$$\frac{\gamma^{(2)}(s)}{\gamma^{(1)}(s)} < \frac{\delta^{(2)}(s)}{\delta^{(1)}(s)} \tag{D.2}$$

Substituting the derivatives, (D.1) and (D.2) reduce to showing,

$$\begin{aligned}(s*p)(1-s*p)\int_s^{s_2}(J(s*p)-J(t*p))dt \\ > (1-2p)s(1-\frac{s}{2})\int_s^{s_2}(J(\frac{s}{2})-J(\frac{t}{2}))dt\end{aligned} \tag{D.3}$$

$$\frac{\int_s^{s_2}(J(\frac{s}{2})-J(\frac{t}{2}))dt}{\int_s^{s_2}2(t-s)dt} > \frac{1}{s+s_2-ss_2} \tag{D.4}$$

respectively.

However, both are immediate as shown below.

From Claim 2 that the ratio $\frac{J(\frac{s}{2})-J(\frac{t}{2})}{J(s*p)-J(t*p)}$ is a strictly decreasing in $t \in [s, s_2]$, we have

$$\begin{aligned}\frac{J(\frac{s}{2})-J(\frac{t}{2})}{J(s*p)-J(t*p)} &< \lim_{t\to s}\frac{J(\frac{s}{2})-J(\frac{t}{2})}{J(s*p)-J(t*p)}\\ &= \frac{(s*p)(1-s*p)}{(1-2p)s(1-\frac{s}{2})}\end{aligned}$$

and hence (D.3) is established.

For (D.4), it reduces to showing

$$J(\frac{s}{2})-J(\frac{t}{2}) \geq \frac{2(t-s)}{s+s_2(1-s)} \tag{D.5}$$

for $s \leq t \leq s_2$. Given any $t \in [s, s_2]$, observe that $s+s_2(1-s) \geq s+t(1-s)$. Thus establishing $\log_2\left(1+\frac{2(t-s)}{(2-t)s}\right) \geq \frac{2(t-s)}{s+t(1-s)}$ suffices.

Let $R = \frac{t-s}{(2-t)s}$. Clearly this holds for $R \geq \frac{1}{2}$. For $0 \leq R \leq \frac{1}{2}$ we have $\log_2(1+2R) \geq 2R \geq \frac{2R}{1+R}$ and hence establishes (D.5).

□ **End of chapter.**

Bibliography

[1] R. F. Ahlswede. Multi-way communication channels. *2nd Internation Symposium on Information Theory, Tsahkadsor*, 1971.

[2] R. F. Ahlswede. Multi-way communication channels. *Hungarian Acad. Sci.*, pages 23–52, 1973.

[3] P. F. Bergmans. Coding theorem for broadcast channels with degraded components. *IEEE Trans. Info. Theory*, IT-15:197–207, March, 1973.

[4] T. Cover. Broadcast channels. *IEEE Trans. Info. Theory*, IT-18:2–14, January, 1972.

[5] T. Cover and J. Thomas. *Elements of Information Theory.* Wiley Interscience, 1991.

[6] S. Diggavi and D. Tse. On opportunistic codes and broadcast codes with degraded message sets. *Information theory workshop (ITW)*, 2006.

[7] A. El Gamal. Capacity of the product and sum of two unmatched broadcast channels. *Problemy Perdachi Informatsi*, 6:3–23, Jan-March, 1980.

[8] A. El Gamal. The capacity of a class of broadcast channels. *IEEE Trans. Info. Theory*, IT-25:166–169, March, 1979.

[9] C. M. Fortuin, J. Ginibre, and P. Kasteleyn. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22, 1971.

[10] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Probl. Peredac. Inform.*, 10(3):3–14, 1974.

[11] J. Körner and K. Marton. Comparison of two noisy channels. *Topics in Inform. Theory(ed. by I. Csiszar and P.Elias), Keszthely, Hungary*, pages 411–423, August, 1975.

[12] J. Körner and K. Marton. General broadcast channels with degraded message sets. *IEEE Trans. Info. Theory*, IT-23:60–64, Jan, 1977.

[13] Y. Liang, G. Kramer, and S. Shamai. Capacity outer bounds for broadcast channels. *Proceedings of IEEE Inf. Theory Workshop, Porto, Portugal*, pages 2–4, 2008.

[14] H. Liao. A coding theorem for multiple access communications. *IEEE Symposium on Information Theory, Asilomar*, 1972.

[15] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Info. Theory*, IT-25:306–311, May, 1979.

[16] C. Nair. An outer bound for 2-receiver discrete memoryless broadcast channels. *CoRR*, abs/0807.3593, 2008.

[17] C. Nair. Capacity regions of two new classes of 2-receiver broadcast channels. *International Symposium on Information Theory*, pages 1839–1843, 2009.

[18] C. Nair and A. El Gamal. The capacity of a class of 3-receiver broadcast channels with degraded message sets. *International Symposium on Information Theory*, pages 1706–1710, 2008.

[19] C. Nair and A. El Gamal. An outer bound to the capacity region of the broadcast channel. *IEEE Trans. Info. Theory*, IT-53:350–355, January, 2007.

[20] C. Nair and Z. Wang. On the inner and outer bounds of 3-receiver broadcast channels with 2-degraded message

sets. *International Symposium on Information Theory*, pages 1844–1848, 2009.

[21] G. S. Poltyrev. Capacity for parallel broadcast channels with degraded components. *Probl. Peredachi Inf., 13*, 2:23–35, 1977.

[22] V. Prabhakaran, S. Diggavi, and D. Tse. Broadcasting with degraded message sets: A deterministic approach. *Proceedings of the 45th Annual Allerton Conference on Communication, Control and Computing*, 2007.

[23] C. E. Shannon. Two-way communication channels. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:611–644, 1961.

[24] C. E. Shannon. A mathermatical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October, 1948.

[25] A. Wyner and J. Ziv. A theorem on the entropy of certain binary sequences and applications: Part I. *IEEE Trans. Info. Theory*, IT-19:772–777, Nov 1973.