On the Evaluation of Marton's Inner Bound for Binary Input Broadcast Channels

GENG, Yanlin

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
August 2012

Abstract of thesis entitled:

On the Evaluation of Marton's Inner Bound for Binary Input Broadcast Channels

Submitted by GENG, Yanlin

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in August 2012

This thesis concerns the evaluation of Marton's inner bound for binary input broadcast channel without common message. This inner bound is the best one for two-receiver broadcast channel, while the best outer bound is UV outer bound. Recently we have shown that UV outer bound is not optimal, however the optimality of Marton's inner bound is still unknown.

In the first part, we introduce a binary inequality obtained by Jog and Nair for binary-skew symmetric broadcast channel, which helps to show for the first time that Marton's inner bound is strictly included in UV outer bound. We generalize this inequality to be true for arbitrary binary input broadcast channel. The method applied here is perturbation analysis, which helps to characterize the properties of non-trivial cases in the proof.

In the second part, we study a class of broadcast channel consisting of binary input symmetric-output channels. We show that whether Marton's inner bound is strictly included in UV outer bound is closely related to the more capable partial order, and we find a second example that demonstrates the strict inclusion.

To evaluate the inner bound beyond the sum-rate, we consider the supporting hyperplanes of the boundary points and conjecture the binary inequality to a stronger one, where we utilize the notion of concave envelope. We prove the extended inequality for certain cases.

The main contribution of the thesis is in the development of new tools and techniques for evaluating certain achievable regions as well as for proving certain information inequalities that are not based on convexity.

中文摘要

本論文考慮對於二值輸入廣播信道在沒有公共信息要求的情況下，如何評估 Marton 内界。對於雙用户廣播信道而言，該内界是最好的，而最好的外界是 UV 外界。最近我們證明了 UV 外界不是容量區域，但是 Marton 内界是否是容量區域尚未可知。

在論文的第一部份，我們介紹了一個由 Jog 和 Nair 獲得的基於二值輸入斜對稱廣播信道的不等式，該不等式被用於首次證明 Marton 内界嚴格包含在 UV 外界里。我們將該不等式推廣到任意二值輸入廣播信道。在證明中，我們採用擾動分析的方法，幫助刻劃了不等式在非平凡情況下的性質。

在第二部份，我們專注于研究輸出對稱的二值輸入廣播信道。我們證明了 Marton 内界是否嚴格包含于 UV 外界里是與特定偏序密切相關的，同時找到了另一個嚴格包含的例子。

對於評估内界而不僅僅是其中的總傳輸率，我們考慮邊界的支撐超平面，然後提出一個猜想，利用凸包的概念推廣了之前提及的不等式。對於大部份情況，我們證明了該猜想。

本論文的主要貢獻在於，我們拓展了評估特定可達傳輸率的新工具和方法，同時證明了某些非基於凸性質的不等式。

Acknowledgement

When I came to CUHK in 2009, I knew little about information theory. During the last three years, my supervisor, Prof. Chandra Nair, introduced me to this field, more specific, to broadcast channel. I learned a lot of things from him. Among them, the important one is, his way of thinking about and trying to solve the problems. I appreciate it and am still trying to “borrow” it.

As a senior group member, Zizhou Wang helped me a lot on living, studying and doing research in CUHK. I am very happy to work with him.

I would like to thank Prof. Raymond Yeung. I took his course on information theory, which is interesting and different from others.

As a visitor to CUHK, Prof. Amin Gohari often provides brilliant solutions to the problems I am working on. I am glad that I know him.

I would also like to thank these guys in ITCSC, including officer Venus Hong, colleagues Yufei Cai, Chengwei Guo, Sida Liu, Yang Liu, Chenglong Ma, Lingxiao Xia, Yuanming Yu. Also I would like to thank Fan Cheng in IE department. With these guys I have a nice everyday life.

Last but not least, I would like to thank my parents and sister for their long lasting support.

This work is dedicated to my parents and sister.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cover [1] introduced the broadcast channel as a communication model where there are one transmitter and multiple receivers. The single-letter characterization for capacity region of broadcast channel is still unknown, except in some scenarios where there are some partial orders between component channels. However, we do have some inner bounds and outer bounds which sandwich the capacity region, and among them Marton's inner bound [12] and UV outer bound [14] are the best ones. It is known through a particular broadcast channel that these two bounds differ, and only until recently we know that UV outer bound is not the capacity region [4]. Thus it is interesting to investigate Marton's inner bound. While in general this is difficult, we restrict ourselves to the binary input case.

In the following we will introduce some bounds and partial

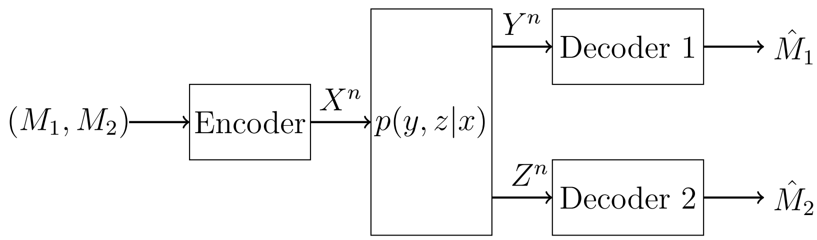

Figure 1.1: Broadcast channel with private messages

orders on broadcast channel. For more details, one may also refer to Chapter 5,8,9 in [3].

1.1 Broadcast channel and capacity region

For the purpose of this thesis, we focus on the following two-receiver discrete memoryless broadcast channel with only private message requirements (Figure 1.1).

Definition 1 ([1]). A *broadcast channel* consists of an input alphabet $\mathcal{X}$ and output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, all of finite sizes, and a probability transition function $p(y, z|x)$.

A $(2^{nR_1}, 2^{nR_2}, n)$ code consists of an encoder

$$x^n : [2^{nR_1}] \times [2^{nR_2}] \to \mathcal{X}^n,$$

and two decoders

$$\hat{\mathcal{W}}_1 : \mathcal{Y}^n \to [2^{nR_1}],$$

$$\hat{\mathcal{W}}_2 : \mathcal{Z}^n \to [2^{nR_2}].$$

The averaged probability of error $P_e^{(n)}$ is defined as

$$P_e^{(n)} = \mathrm{P}(\hat{M}_1 \neq M_1 \text{ or } \hat{M}_2 \neq M_2)$$

where (M_1, M_2) is assumed to be uniform over $[2^{nR_1}] \times [2^{nR_2}]$.

A rate pair (R_1, R_2) is said to be achievable for the broadcast channel if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $P_e^{(n)} \to 0$. The *capacity region* (CR) of the broadcast channel is the closure of the set of achievable rate pairs.

The multi-letter characterization of capacity region is the limit (as $n \to \infty$) of the union of rate pairs satisfying

$$\begin{aligned} R_1 &\leq \frac{1}{n} I(U; Y^n) \\ R_2 &\leq \frac{1}{n} I(V; Z^n) \end{aligned}$$

over $p(u)p(v)p(x^n|u, v)$. However, this does not help evaluation, and the single-letter one is still unknown. Instead we have some inner bounds and outer bounds.

1.2 Inner bounds to capacity region

An inner bound refers to a region in which rate pairs are achievable by some particular coding scheme. Hence an inner bound is contained in the capacity region. In this section we introduce some known inner bounds to the capacity region.

Time-division region (TD) is characterized by the following set of points

Bound 1 (TD)**.** *The following rate pairs are achievable*

$$R_1 \leq \alpha \cdot \max_{p(x)} I(X;Y)$$
$$R_2 \leq (1-\alpha) \cdot \max_{p(x)} I(X;Z)$$

for all $\alpha \in [0,1]$.

The rates are achieved by transmitting at channel capacity to one receiver for fraction α of the time, and at channel capacity to the other one for the remaining fraction.

Randomized time-division region (RTD) corresponds to a time-division strategy except that the slots for which communication occurs to one receiver is also drawn from a codebook which conveys additional information. The rates are characterized by

Bound 2 (RTD)**.** *The following rate pairs are achievable*

$$\begin{aligned} R_1 &\leq I(W;Y) + \mathrm{P}(W{=}0) I(X;Y|W{=}0) \\ R_2 &\leq I(W;Z) + \mathrm{P}(W{=}1) I(X;Z|W{=}1) \\ R_1 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} + \mathrm{P}(W{=}0) I(X;Y|W{=}0) \\ &\quad + \mathrm{P}(W{=}1) I(X;Z|W{=}1) \end{aligned}$$

for all $p(w,x)$ *over binary* W.

Here W characterizes the slots which distinguish communication to one receiver over the other. By taking $W \sim Bern(1-\alpha)$, and $p(x|w)$ as capacities achieving distributions, RTD reduces to TD region.

The following Marton's inner bound (MIB) [12] is the best known achievable rate region.

Bound 3 (MIB)**.** *The following rate pairs are achievable*

$$\begin{aligned} R_1 &\leq I(U, W; Y) \\ R_2 &\leq I(V, W; Z) \\ R_1 + R_2 &\leq \min\{I(W; Y), I(W; Z)\} \\ &\quad + I(U; Y|W) + I(V; Z|W) - I(U; V|W), \end{aligned}$$

for all $p(w, u, v, x)$*.*

Observe that setting $U = X$, $V = \emptyset$ when $W = 0$ and $V = X$, $U = \emptyset$ when $W = 1$ reduces MIB to the RTD region.

The description above already suggests that

$$TD \subseteq RTD \subseteq MIB \subseteq CR$$

While whether Marton's inner bound is strictly included in capacity region is still unknown.

1.3 Outer bounds to capacity region

An outer bound refers to a region that, any achievable rate pair must lie in the region. Hence an outer bound contains the capacity region. In this section we introduce some known outer bounds to the capacity region.

Körner–Marton outer bound (KMOB) [12] is the following bound:

Bound 4 (KMOB)**.** *The region $\mathcal{O}_Y \cap \mathcal{O}_Z$ forms an outer bound to the capacity region, where $\mathcal{O}_Y$ is the union of rate pairs satisfying*

$$R_1 \leq I(U;Y)$$
$$R_2 \leq I(X;Z)$$
$$R_1 + R_2 \leq I(U;Y) + I(X;Z|U)$$

over $p(u,x)$, and similarly $\mathcal{O}_Z$ is

$$R_1 \leq I(X;Y)$$
$$R_2 \leq I(V;Z)$$
$$R_1 + R_2 \leq I(V;Z) + I(X;Y|V)$$

over $p(v,x)$.

The following UV outer bound (UVOB) [14] is the best known outer bound

Bound 5 (UVOB)**.** *The union of rate pairs satisfying*

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

over $p(u,v,x)$ *forms an outer bound to capacity region.*

It is clear that $UVOB \subseteq \mathcal{O}_Y$ and $UVOB \subseteq \mathcal{O}_Z$, hence

$$CR \subseteq UVOB \subseteq KMOB$$

We already know that UV outer bound is not optimal [4]. There are also some other outer bounds for general broadcast channels, however we don't know if they are better than UV outer bound.

1.4 Partial orders

In this section we introduce some partial orders on component channels of a broadcast channel, where we know the capacity region. In these partial orders, there is a "stronger" receiver, say Y, and the capacity region is achieved by superposition coding [12], which is

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

over $p(v, x)$.

In the following, by receiver Y we also mean the component channel $X \to Y$. We will not clarify unless necessary. And by saying $X \to Y \to Z$ we mean Markov chain.

The following definitions on degraded can be found in Section 5.4 of [3].

Definition 2. Receiver Z is a *physically degraded* version of Y if $X \to Y \to Z$.

Definition 3. Receiver Z is a *statistically degraded* version of Y if there exists a virtual receiver $\tilde{Z}$ such that $p(\tilde{z}|x) = p(z|x)$ and $X \to Y \to \tilde{Z}$.

Definition 4. Receiver Z is a *degraded* version of Y if Z is physically or statistically degraded version of Y.

Since for discrete memoryless broadcast channel, capacity region only depends on marginals $p(y|x)$ and $p(z|x)$, thus when refer to degraded case, we use $X \to Y \to Z$ for simplicity.

Definition 5 ([11]). Receiver Y is *less noisy* than Z if for all $p(u,x)$ we have $I(U;Y) \geq I(U;Z)$.

Definition 6 ([11]). Receiver Y is *more capable* than Z if for all $p(x)$ we have $I(X;Y) \geq I(X;Z)$.

Definition 7 ([13]). A class of distributions $\mathcal{P} = \{p(x)\}$ is said to be a *sufficient class* for broadcast channel $X \to (Y,Z)$, if for any distribution $q(u,v,x)$ there exists a $p(u,v,x)$ such that $p(x) \in \mathcal{P}$ and the following values

$$I(U;Y), I(X;Y|U), I(V;Y), I(X;Y|V)$$
$$I(U;Z), I(X;Z|U), I(V;Z), I(X;Z|V)$$

are non-decreasing when distributions change from q to p.

Definition 8 ([13]). Receiver Y is *essentially less noisy* than Z if there exists a sufficient class $\mathcal{P}$ such that whenever $p(x) \in \mathcal{P}$, for all $p(u,x)$ we have $I(U;Y) \geq I(U;Z)$.

For the notation, we use $\overset{ln}{\geq}, \overset{mc}{\geq}, \overset{eln}{\geq}$ to denote the corresponding relationship, say $Y \overset{ln}{\geq} Z$ means Y is less noisy than Z.

Remark 1. There are some notes here. (a) A sufficient class is a set of distributions that are sufficient for evaluating certain region $\mathcal{R}$. (b) For degraded, less noisy and more capable, the underlying sufficient class is the whole space of distributions on

X, that is, a common sufficient class. (c) Thus for essentially less noisy, it makes sense to consider broadcast channels having a common sufficient class. (d) Since we are considering region $\mathcal{R}$, it is natural to say that receivers Z and R are equivalent under class $\mathcal{P}$, if for any receiver Y such that $\mathcal{P}$ is a common sufficient class for $X \to (Y, Z)$ and $X \to (Y, R)$, they have the same region $\mathcal{R}$. (e) Thus one may say that essentially less noisy is a partial order (under a common sufficient class).

It is clear that

$$\begin{aligned} X \to Y \to Z \implies Y \overset{ln}{\geq} Z &\implies Y \overset{mc}{\geq} Z \\ Y \overset{ln}{\geq} Z &\implies Y \overset{eln}{\geq} Z \end{aligned}$$

However, more capable and essentially less noisy are not comparable in general. In Chapter 3 we will show that these two relationships go in *reverse* directions for some special receivers, which is counter intuitive since $Y \overset{mc}{\geq} Z$ and $Y \overset{eln}{\geq} Z$ are both saying that Y is "stronger" than Z. Hence these special receivers can be served as examples where we revise the definitions of "stronger" receivers.

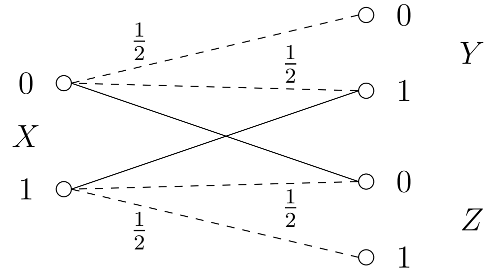

Figure 1.2: Binary skew-symmetric broadcast channel (BSSC)

1.5 Examples where inner and outer bounds differ

It is from the results in [15, 7, 10] that people know Marton's inner bound is not equal to UV outer bound. The counter example there is the binary skew-symmetric broadcast channel (BSSC) in Figure 1.2. They showed that the maximum sum-rates $R_1 + R_2$ evaluate to 0.3616, 0.3725, 0.3744 for Marton's inner bound, UV outer bound, Körner–Marton outer bound, respectively [5].

The evaluation of sum-rate for Marton's inner bound utilizes the following inequality for BSSC

$$I(U;Y) + I(V;Z) - I(U;V) \leq \max\{I(X;Y), I(X;Z)\} \quad (1.1)$$

In Chapter 2, we prove this inequality to be true for arbitrary broadcast channel with binary input X. In Chapter 4, we generalize this inequality further, and make a conjecture which helps to evaluate Marton's inner bound for binary input broadcast channels. In particular the conjecture is true for BSSC.

Another class of examples is provided in Chapter 3, where we show that, for a special class of broadcast channels, Marton's inner bound coincides with UV outer bound if and only if the receivers are more capable comparable. By providing a broadcast channel where the receivers are not more capable comparable, we find a second example where the two bounds differ (and actually one can get more examples).

Recently it is shown [4] that UV outer bound is not optimal, hence it becomes more important to investigate Marton's inner bound. By studying examples other than BSSC, we hope that better understanding can be made on Marton's inner bound.

To the end of this chapter, some notations and abbreviations are provided in Table 1.1 for reference.

Table 1.1: Notation

Notation	Meaning
X^n	$(X_1, X_2, \dots, X_n)$
$[n]$	$\{1, 2, \dots, n\}$
$Bern(\alpha)$	Bernoulli distribution
x	$\mathrm{P}(X{=}0)$ for binary X
$\bar{t}$	$1 - t$
$a * b$	$a(1-b) + (1-a)b$
$h(t)$	binary entropy function
$p_{y\|x}$	$p(y\|x)$
p_{uv}	$\mathrm{P}(U{=}u, V{=}v)$
$Y \overset{ln}{\geq} Z$	Y less noisy than Z
$Y \overset{mc}{\geq} Z$	Y more capable than Z
$Y \overset{eln}{\geq} Z$	Y essentially less noisy than Z
$\mathfrak{C}[f]$	concave envelope
$\subset$	strict inclusion
$\subseteq$	inclusion
iff	if and only if
TD	time-division
RTD	randomized time-division
MIB	Marton's inner bound
CR	capacity region
UVOB	UV outer bound
KMOB	Körner–Marton outer bound
BSSC	binary skew-symmetric broadcast channel
BISO	binary input symmetric output

Chapter 2

A binary inequality

The main result of this chapter is the following Theorem 1, which generalizes (1.1) to be true for *every* binary input broadcast channel.

Theorem 1. *For Markov chain* $(U,V) \to X \to (Y,Z)$ *with binary* X*, the following inequality holds:*

$$I(U;Y) + I(V;Z) - I(U;V) \le \max\{I(X;Y), I(X;Z)\}. \quad (2.1)$$

To evaluate the sum-rate for Marton's inner bound, we need the cardinality bounds from Theorem 2.

Theorem 2 ([14])**.** *It suffices to consider* $|\mathcal{U}| \le |\mathcal{X}|$, $|\mathcal{V}| \le |\mathcal{X}|$, *and* $|\mathcal{W}| \le |\mathcal{X}|$ *to achieve the supremum of the sum-rate for Marton's inner bound.*

Combining Theorem 1 and Theorem 2, Corollary 1 establishes that the maximum sum-rate given by Marton's coding strategy matches that given via the randomized time-division strategy [14], a much simpler achievable strategy for any binary input broadcast channel.

Corollary 1. *The maximum value of the sum-rate for Marton's inner bound for any binary input broadcast channel is given by*

$$\max_{p(w,x)}\{\min\{I(W;Y),I(W;Z)\} \\ +\mathrm{P}(W{=}0)I(X;Y|W{=}0)+\mathrm{P}(W{=}1)I(X;Z|W{=}1)\}$$

where $|\mathcal{W}|=2$.

Proof. Let $\bar{R}$ be the maximum sum-rate obtained by the randomized time-division strategy (Bound 2) and R be that by Marton's inner bound (Bound 3). We need to show $R=\bar{R}$. Clearly $R\geq\bar{R}$ as $\bar{R}$ is a restriction of the choice of (U,V,W).

From Theorem 2, to evaluate the Marton's sum-rate for binary input broadcast channel it suffices to look at $|\mathcal{W}|\leq 2$. Consider a (U,V,W) that achieves the maximum sum-rate R. Without loss of generality we consider two cases:

Case 1: $I(X;Y|W{=}w)\geq I(X;Z|W{=}w)$ for $w=0,1$. Clearly

$$R=\min\{I(W;Y),I(W;Z)\} \\ +I(U;Y|W)+I(V;Z|W)-I(U;V|W)$$

$$
\begin{aligned}
&= \min\{I(W;Y), I(W;Z)\} \\
&\quad + \mathrm{P}(W{=}0)\big(I(U;Y|W{=}0) + I(V;Z|W{=}0) - I(U;V|W{=}0)\big) \\
&\quad + \mathrm{P}(W{=}1)\big(I(U;Y|W{=}1) + I(V;Z|W{=}1) - I(U;V|W{=}1)\big) \\
&\overset{(a)}{\leq} \min\{I(W;Y), I(W;Z)\} \\
&\quad + \mathrm{P}(W{=}0)I(X;Y|W{=}0) + \mathrm{P}(W{=}1)I(X;Y|W{=}1) \\
&\leq \min\{I(W;Y), I(W;Z)\} + I(X;Y|W) \leq I(X;Y) \leq \bar{R}
\end{aligned}
$$

where (a) follows from Theorem 1 and case specification.

Case 2: $I(X;Y|W{=}0) \geq I(X;Z|W{=}0)$ and $I(X;Y|W{=}1) \leq I(X;Z|W{=}1)$. Observe that

$$
\begin{aligned}
R &= \min\{I(W;Y), I(W;Z)\} \\
&\quad + I(U;Y|W) + I(V;Z|W) - I(U;V|W) \\
&= \min\{I(W;Y), I(W;Z)\} \\
&\quad + \mathrm{P}(W{=}0)\big(I(U;Y|W{=}0) + I(V;Z|W{=}0) - I(U;V|W{=}0)\big) \\
&\quad + \mathrm{P}(W{=}1)\big(I(U;Y|W{=}1) + I(V;Z|W{=}1) - I(U;V|W{=}1)\big) \\
&\overset{(b)}{\leq} \min\{I(W;Y), I(W;Z)\} \\
&\quad + \mathrm{P}(W{=}0)I(X;Y|W{=}0) + \mathrm{P}(W{=}1)I(X;Z|W{=}1) \\
&\leq \bar{R},
\end{aligned}
$$

where (b) follows from Theorem 1 and case specification.

The other two cases follow similarly. Thus $R \leq \bar{R}$ and the proof is completed. □

In the rest part we are going to prove Theorem 1. The idea is to fix the broadcast channel $p(y, z|x)$ and show that for all fixed $p(x)$ we have that

$$\max_{p(u,v|x)} I(U;Y) + I(V;Z) - I(U;V) \leq \max\{I(X;Y), I(X;Z)\}$$

Denote LHS and RHS as the left-hand side and right-hand side of the inequality (2.1), respectively, that is

$$\begin{aligned} LHS =& I(U;Y) + I(V;Z) - I(U;V) \\ RHS =& \max\{I(X;Y), I(X;Z)\} \end{aligned}$$

For brevity let

$$p_{uv} = \mathrm{P}(U{=}u, V{=}v), \quad p_{y|x} = p(y|x), \quad p_{z|x} = p(z|x).$$

Also since the couplings of variables are (U, Y) and (V, Z), we use notation

$$p_{uy} = \mathrm{P}(U{=}u, Y{=}y), \quad p_{vz} = \mathrm{P}(V{=}v, Z{=}z)$$

unless they conflict notation p_{uv}.

Remark 2. As LHS and RHS are continuous in $p_{y|x}$ and $p_{z|x}$ for any fixed $p(u, v, x)$, it suffices to prove the inequality when $p_{y|x}$ and $p_{z|x}$ are positive.

The following theorem is one of the main results in [7]. For a self-contained shorter proof, one may see Fact 1 and Claim 1 in [10].

Theorem 3. *To maximize $I(U;Y)+I(V;Z)-I(U;V)$ over all $p(u,v|x)$ such that $(U,V) \to X \to (Y,Z)$, it suffices to consider $|\mathcal{U}| \leq |\mathcal{X}|, |\mathcal{V}| \leq |\mathcal{X}|$ and $X = f(U,V)$, a deterministic function of (U,V).*

Since X is binary, for expressing f, we use the notation: $U \wedge V$ (and), $U \vee V$ (or), $U \oplus V$ (xor), $\bar{U}$ (not).

The following claim will be used in the proof. We include it here to avoid messing the main part of the proof.

Claim 1. *For broadcast channel $X \to (Y,Z)$ with positive transition probabilities, let function $X = f(U,V)$ and p.m.f. $p(u,v)$ maximize LHS. If $p(u) > 0$ and $p(v) > 0$ for a pair (u,v), then $p(u,v) > 0$.*

Proof. The proof uses perturbation to show that we can increase LHS otherwise. Suppose $p(u_1,v_1) = 0$ and $p(u_1) > 0$, $p(v_1) > 0$. Then we must have $v_2 \neq v_1$ such that $p(u_1,v_2) > 0$. Let $f(u_1,v_2) = x_1$. Perturbate p at two points

$$q(u,v,x) = \begin{cases} p(u,v,x) - \epsilon & (u,v,x) = (u_1,v_2,x_1) \\ \epsilon & (u,v,x) = (u_1,v_1,x_1) \\ p(u,v,x) & \text{otherwise} \end{cases}$$

Notice that $p(x)$ is maintained. Now for LHS, we have (for simplicity we use natural logarithm)

$$
\begin{aligned}
& LHS(q) - LHS(p) \\
& = H_q(U,V) - H_q(U,Y) - H_q(V,Z) \\
& \quad - H_p(U,V) + H_p(U,Y) + H_p(V,Z) \\
& = \epsilon\{-\ln\epsilon + \ln p_{u_1 v_2} + \sum_z p_{z|x_1} \ln \frac{p_{v_1 z}}{p_{v_2 z}}\} + o(\epsilon)
\end{aligned}
$$

Observe that the first derivative is positive infinity, hence we can increase the sum-rate. □

The outline of the proof is:

1. We first prove the inequality for some special settings, or "trivial" cases. (Section 2.1)

2. We show that it suffices to prove for the nontrivial cases $X = U \wedge V$ and $X = U \oplus V$. (Section 2.2)

3. For $X = U \oplus V$, nontrivial maximum of LHS is not achievable when $p(u,v) > 0$. (Section 2.3)

4. For $X = U \wedge V$, nontrivial maximum of LHS is not achievable when $p(u,v) > 0$. (Section 2.4)

2.1 Proof of special settings

Since $U \to X \to Y$ and $V \to X \to Z$ are Markov chains, from data processing inequality, we know

$$
\begin{aligned}
I(U;Y) \le I(X;Y), &\quad I(U;Y) \le I(U;X), \\
I(V;Z) \le I(X;Z), &\quad I(V;Z) \le I(V;X).
\end{aligned} \tag{2.2}
$$

With these inequalities, we first prove Theorem 1 for some special settings. Denote $X \perp Y$ as independence.

SS1: $p_{y|0} \equiv p_{y|1}$. Then $X \perp Y$, thus $I(U;Y) = I(X;Y) = 0$. From (2.2) and the non-negativity of $I(U;V)$ we have $I(V;Z) - I(U;V) \le I(X;Z)$, i.e. Theorem 1 holds. Similarly Theorem 1 holds when $p_{z|0} \equiv p_{z|1}$.

SS2: $U \perp X$. Then $I(U;Y) = I(U;X) = 0$. Again from (2.2) and the non-negativity of $I(U;V)$ Theorem 1 holds. Similarly when $V \perp X$, Theorem 1 also holds.

2.2 Two nontrivial cases

According to Theorem 3, to prove the inequality (2.1), it suffices to consider $X = f(U, V)$ with binary U and V. Notice there are 16 possible functions f, and they can be classified into the following equivalent (due to relabeling) groups

G_1: $X = 0, X = 1$

G_2: $X = U, X = \bar{U}, X = V, X = \bar{V}$

G_3: $X = U \wedge V, X = \bar{U} \wedge V, X = U \wedge \bar{V}, X = \bar{U} \wedge \bar{V}$

G_4: $X = U \vee V, X = \bar{U} \vee V, X = U \vee \bar{V}, X = \bar{U} \vee \bar{V}$

G_5: $X = U \oplus V, X = \bar{U} \oplus V$

The reason that these are equivalent groups is that, in each group, all the cases can be reduced to the first case by using some bijections. For example, in G_3, let the distributions of (U, V) be $p(u, v)$ and $r(u, v)$ for $X = U \wedge V$ and $X = \bar{U} \wedge V$, respectively. The bijection is $p_{00} \leftrightarrow r_{10}$, $p_{01} \leftrightarrow r_{11}$, $p_{10} \leftrightarrow r_{00}$, $p_{11} \leftrightarrow r_{01}$. Thus, we just need to prove Theorem 1 for the first function in each group.

Further, notice for the case $X = U \vee V$ with $q(u, v)$, by bijection $p_{00} \leftrightarrow q_{11}$, $p_{01} \leftrightarrow q_{01}$, $p_{10} \leftrightarrow q_{10}$, $p_{11} \leftrightarrow q_{00}$, we can also use the same proof as for the case $X = U \wedge V$. That is, we use the fact that $X = U \vee V \Leftrightarrow \bar{X} = \bar{U} \wedge \bar{V}$ to reduce the proof of the OR case of one channel to the AND case of another broadcast channel obtained by flipping U, V and X.

So it remains to consider the first cases of groups except G_4.

The first two cases are trivial. For $X = 0$, the theorem is reduced to $-I(U; V) \leq 0$. For $X = U$, i.e. $I(U; Y) = I(X; Y)$, the theorem follows from the data processing inequality, $I(V;Z) \leq I(V;U) = I(V;X)$ (see (2.2)). Now for cases in G_3 and G_5, if $p(x) = 0$ for some x, then they reduce to G_1; if $p_u = 0$ (or $p_v = 0$) for some u (or v), then they reduce to cases in G_1 or G_2. By Claim 1, finally we just need to consider the following two nontrivial cases:

C_3: $X = U \wedge V$ with $p(x) > 0$ and $p(u,v) > 0$

C_5: $X = U \oplus V$ with $p(x) > 0$ and $p(u,v) > 0$

We are going to prove that there is no nontrivial local maximum for these two cases.

2.3 Proof of XOR case

Just as in [10] we will consider an additive perturbation, first for any fixed $X = f(U,V)$ subject to $p(x) > 0$ and $p(u,v) > 0$, then restricted to $X = U \oplus V$.

Consider an additive perturbation $q(u,v,x) = p(u,v,x) + \varepsilon\lambda(u,v,x)$ for some $\varepsilon \geq 0$. For the notation, $\lambda_{uvx} = \lambda(u,v,x)$, p_u means the marginal p.m.f. of U given $p(u,v,x,y,z)$, and any other marginal p.m.f. is similar.

For a valid perturbation, we require that $\lambda_{uvx} \geq 0$ if the corresponding $p(u,v,x)$ is zero, which is

$$\lambda_{uvx} \geq 0, \quad \text{if } f(u,v) \neq x$$

Further let us require the perturbation maintains $p(x)$ (hence $H(Y)$ and $H(Z)$), that is

$$\sum_{uv} \lambda_{uvx} = 0, \quad \forall x \in \mathcal{X} \tag{2.3}$$

For any perturbation that satisfies the above conditions at any local maximum $p(u,v,x)$, it must be true that the first derivative cannot be positive. This implies that $\sum_{xuv} \lambda_{uvx} C_{uvx} \leq 0$, where

$$C_{uvx} = -\log p_{uv} + \sum_{y} p_{y|x} \log p_{uy} + \sum_{z} p_{z|x} \log p_{vz}$$

For $x \in \mathcal{X}$, choose one pair (u_x, v_x) such that $f(u_x, v_x) = x$. This is possible since $p(x) > 0$. From (2.3), we express $\lambda_{u_x v_x x}$ using other λ_{uvx} variables

$$\lambda_{u_x v_x x} = -\sum_{uv \neq u_x v_x} \lambda_{uvx}$$

Substituting it into $\sum_{xuv} \lambda_{uvx} C_{uvx} \leq 0$, we have

$$\sum_{xuv: uv \neq u_x v_x} \lambda_{uvx}(C_{uvx} - C_{u_x v_x x}) \leq 0$$

Above holds for any signed $\{\lambda_{uvx} : f(u,v) = x, (u,v) \neq (u_x, v_x)\}$ and any nonnegative $\{\lambda_{uvx} : f(u,v) \neq x\}$, it implies

$$C_{uvx} = C_{u_x v_x x}, \quad \text{if } f(u,v) = x$$

$$C_{uvx} \leq C_{u_x v_x x}, \quad \text{if } f(u,v) \neq x$$

So now we have the following claims:

Claim 2. *Let $f(u,v) = x$, for any (u_1, v_1) we have $C_{u_1v_1x} \leq C_{uvx}$, that is*

$$\log \frac{p_{u_1v_1}}{p_{uv}} \geq \sum_y p_{y|x} \log \frac{p_{u_1y}}{p_{uy}} + \sum_z p_{z|x} \log \frac{p_{v_1z}}{p_{vz}}$$

Claim 3. *If $f(u_1, v_1) = f(u_2, v_2) = x$, then*

$$p_{u_1v_1}p_{u_2v_2} \leq p_{u_1v_2}p_{u_2v_1}$$

where the equality holds iff $C_{u_1v_2x} = C_{u_2v_1x} = C_{u_1v_1x} (= C_{u_2v_2x})$.

Proof. The proof is finished by noticing that $C_{u_1v_1x} + C_{u_2v_2x} \geq C_{u_1v_2x} + C_{u_2v_1x}$. $\square$

Now return to $X = U \oplus V$, notice that $f(0,0) = f(1,1) = 0$, hence by Claim 3 we have for p_{uv} that $p_{00}p_{11} \leq p_{01}p_{10}$; also $f(0,1) = f(1,0) = 1$, hence $p_{00}p_{11} \geq p_{01}p_{10}$. Thus we have

$$p_{00}p_{11} = p_{01}p_{10} \tag{2.4}$$

By Claim 3, this holds iff $C_{010} = C_{100} = C_{000} = C_{110}$ and $C_{001} = C_{111} = C_{011} = C_{101}$. In particular, $C_{000} = C_{010}$ and $C_{001} = C_{011}$ imply that

$$\log \frac{p_{00}}{p_{01}} = \sum p_{z|0} \log \frac{p_{0z}}{p_{1z}} = \sum p_{z|1} \log \frac{p_{0z}}{p_{1z}}.$$

Notice $p_{0z} = p_{00}p_{z|0} + p_{10}p_{z|1}$, take a weighted sum, we get

$$(p_{00} + p_{10}) \log \frac{p_{00}}{p_{01}} = \sum p_{0z} \log \frac{p_{0z}}{p_{1z}}$$

From above and using K-L divergence, we have

$$\log \frac{p_{00}}{p_{01}} \geq \log \frac{p_{00} + p_{10}}{p_{11} + p_{01}} = \log \frac{p_{00}}{p_{01}}$$

where the last step holds since $p_{00}p_{11} = p_{01}p_{10}$. Now that the K-L divergence inequality is indeed an equality, we require

$$\frac{p_{00}}{p_{01}} \equiv \frac{p_{0z}}{p_{1z}} = \frac{p_{00}p_{z|0} + p_{10}p_{z|1}}{p_{11}p_{z|0} + p_{01}p_{z|1}}$$

From the above we obtain

$$(p_{01} - p_{11})(p_{z|0} - p_{z|1}) \equiv 0. \tag{2.5}$$

Similarly from $C_{100} = C_{110}$ and $C_{101} = C_{111}$, we can obtain

$$(p_{10} - p_{11})(p_{y|0} - p_{y|1}) \equiv 0. \tag{2.6}$$

Now we have two cases

1: $p_{y|0} \equiv p_{y|1}$, or $p_{z|0} \equiv p_{z|1}$. In this case the Theorem holds (special setting SS1).

2: $p_{01} = p_{11}$, $p_{10} = p_{11}$. Combining this with $p_{00}p_{11} = p_{01}p_{10}$ one obtains that $p_{uv} = 1/4$, and as a result U, V and X are mutually independent. The Theorem holds (special setting SS2).

If neither of these two cases is satisfied, there would be no local maxima.

2.4 Proof of AND case

Similarly we will show that nontrivial local maxima can't be achieved when $p(x) > 0$ and $p(u, v) > 0$. In this case, $\mathrm{P}(X{=}1) = p_{11}$. Now we fix $p_{11} \in (0, 1)$. Take (p_{10}, p_{01}) as the free variables, with $p_{00} = 1 - p_{11} - p_{01} - p_{10}$. Notice that $H(Y)$ and $H(Z)$ are fixed, the local maxima of LHS is the same as that of

$$
\begin{aligned}
&J(p_{10}, p_{01}) := H(U, V) - H(U, Y) - H(V, Z) \\
&= -p_{00} \log p_{00} - p_{01} \log p_{01} - p_{10} \log p_{10} - p_{11} \log p_{11} \\
&\quad + \sum (p_{00} + p_{01}) p_{y|0} \log\{(p_{00} + p_{01}) p_{y|0}\} \\
&\quad + \sum (p_{10} p_{y|0} + p_{11} p_{y|1}) \log\{p_{10} p_{y|0} + p_{11} p_{y|1}\} \\
&\quad + \sum (p_{00} + p_{10}) p_{z|0} \log\{(p_{00} + p_{10}) p_{z|0}\} \\
&\quad + \sum (p_{01} p_{z|0} + p_{11} p_{z|1}) \log\{p_{01} p_{z|0} + p_{11} p_{z|1}\}.
\end{aligned}
$$

At any local maximum, the gradient ∇J and Hessian matrix $\nabla^2 J$ must satisfy

$$\nabla J = 0, \quad \nabla^2 J \preceq 0,$$

where $\nabla^2 J \preceq 0$ denotes that $\nabla^2 J$ is negative semi-definite. We now compute the gradient and the Hessian to investigate locations of the local maxima.

1. *First Derivative*:

Differentiating with respect to free variables:

$$\frac{\partial J}{\partial p_{10}} = \log \frac{p_{00}}{p_{10}} - \sum p_{y|0} \log \frac{(p_{00}+p_{01})p_{y|0}}{p_{10}p_{y|0}+p_{11}p_{y|1}}$$
$$\frac{\partial J}{\partial p_{01}} = \log \frac{p_{00}}{p_{01}} - \sum p_{z|0} \log \frac{(p_{00}+p_{10})p_{z|0}}{p_{01}p_{z|0}+p_{11}p_{z|1}}$$

The condition $\nabla J = 0$ implies that

$$\log \frac{p_{00}}{p_{10}} = \sum p_{y|0} \log \frac{(p_{00}+p_{01})p_{y|0}}{p_{10}p_{y|0}+p_{11}p_{y|1}} \tag{2.7}$$
$$\log \frac{p_{00}}{p_{01}} = \sum p_{z|0} \log \frac{(p_{00}+p_{10})p_{z|0}}{p_{01}p_{z|0}+p_{11}p_{z|1}}. \tag{2.8}$$

Remark 3. Equalities above are obvious from Claim 2 by noticing that $0 \wedge 0 = 0 \wedge 1 = 1 \wedge 0 = 0$. This is expected as Claim 2 is a result from first derivative.

Using the concavity of logarithm, we have

$$\begin{aligned} \frac{p_{00}}{p_{10}} &\leq \sum \frac{(p_{00}+p_{01})p_{y|0}^2}{p_{10}p_{y|0}+p_{11}p_{y|1}} \\ \frac{p_{00}}{p_{01}} &\leq \sum \frac{(p_{00}+p_{10})p_{z|0}^2}{p_{01}p_{z|0}+p_{11}p_{z|1}} \end{aligned} \tag{2.9}$$

where the equalities hold iff (using Remark 2)

$$p_{y|0} \equiv p_{y|1} \cdot C_Y, \quad p_{z|0} \equiv p_{z|1} \cdot C_Z,$$

for some constants C_Y, C_Z respectively. However since $\sum p_{y|0} = \sum p_{y|1} = 1$ we obtain that $C_Y = 1$ (similarly $C_Z = 1$). Thus equalities hold iff

$$p_{y|0} \equiv p_{y|1}, \quad p_{z|0} \equiv p_{z|1}. \tag{2.10}$$

2. Second Derivative:

We now compute the Hessian $G = \nabla^2 J$, The second derivatives are

$$G_{11} = \frac{\partial^2 J}{\partial p_{10}^2} = -\frac{1}{p_{00}} - \frac{1}{p_{10}} + \frac{1}{p_{00}+p_{01}} + \sum \frac{p_{y|0}^2}{p_{10}p_{y|0} + p_{11}p_{y|1}}$$
$$G_{12} = G_{21} = -\frac{1}{p_{00}}$$
$$G_{22} = \frac{\partial^2 J}{\partial p_{01}^2} = -\frac{1}{p_{00}} - \frac{1}{p_{01}} + \frac{1}{p_{00}+p_{10}} + \sum \frac{p_{z|0}^2}{p_{01}p_{z|0} + p_{11}p_{z|1}}$$

As $p_{01} > 0$, we have $G_{11} \le -\frac{1}{p_{00}} - \frac{1}{p_{10}} + \frac{1}{p_{00}+p_{01}} + \frac{1}{p_{10}} < 0$. Similarly we have $G_{22} < 0$. Since $G_{11} < 0$ and $G_{22} < 0$, G is negative semi-definite iff $\det(G) \ge 0$.

From (2.9) we have

$$\begin{aligned} G_{11} &\ge -\frac{1}{p_{00}} - \frac{1}{p_{10}} + \frac{1}{p_{00}+p_{01}} + \frac{p_{00}}{p_{10}(p_{00}+p_{01})} \\ &= -\frac{p_{01}(p_{00}+p_{10})}{p_{00}p_{10}(p_{00}+p_{01})} \end{aligned}$$

And similarly

$$G_{22} \ge -\frac{p_{10}(p_{00}+p_{01})}{p_{00}p_{01}(p_{00}+p_{10})}$$

It is clear that equalities in the above two inequalities hold iff (2.10) holds.

Since $G_{11}, G_{22} < 0$ we have

$$G_{11}G_{22} \le \frac{p_{01}(p_{00}+p_{10})}{p_{00}p_{10}(p_{00}+p_{01})} \cdot \frac{p_{10}(p_{00}+p_{01})}{p_{00}p_{01}(p_{00}+p_{10})} = \frac{1}{p_{00}^2} = G_{12}^2,$$

with equality holding only if (2.10) holds. Thus $\det(G) \le 0$.

When $\det(G) < 0$, there is no local maximum for $p(u, v) > 0$. When $\det(G) = 0$, channel parameters satisfy (2.10), and the inequality is true from the special setting SS1.

This completes the proof of Theorem 1.

Chapter 3

BISO broadcast channel

In this chapter, we focus on a sub-class of binary input broadcast channels: binary input symmetric output (BISO) [6] broadcast channels. We study in detail two partial orders: more capable and essentially less noisy. We establish a slew of results and some of the interesting ones are summarized below. Notice that by demonstrating that one channel is more capable than the other, we indirectly establish its capacity region as the capacity region for the more capable class is known [2].

- Any BISO channel with capacity C is more capable than the binary symmetric channel with capacity C. (Corollary 2).

- The binary erasure channel with capacity C is more capable than any BISO channel with capacity C. (Corollary 3)

- Any two BISO channels with the same capacity and whose outputs have cardinality at most 3, are more capable comparable. (Corollary 4)

- For any two BISO channels with same capacity, a receiver Y is more capable than receiver Z *if and only if* receiver Z is essentially less noisy than Y. (They go in *reverse* directions.) (Lemma 4)

- Superposition coding region is the capacity region for a BISO broadcast channel if any one of the channels is either a BSC or a BEC. (Corollary 5)

- For two BISO channels with the same capacity, superposition coding is optimal if and only if the channels are more capable comparable. (Corollary 6)

- For two BISO channels of same capacity, Marton's inner bound differs from UV outer bound [14] unless the channels are more capable comparable. (Theorem 6)

- We also show that it suffices to consider $U \to X$ to be BSC when we wish to compute the boundary of the superposition coding region for BISO broadcast channels. (Lemma 7.) This vastly generalizes a result of Wyner and Ziv [16] for degraded BSC broadcast channel.

3.1 BISO channel

Definition 9. A discrete memoryless channel with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} = \{k : -l \leq k \leq l\}$ is said to be *binary input symmetric output* if

$$p_k := \mathrm{P}(Y{=}k|X{=}0) = \mathrm{P}(Y{=}-k|X{=}1), -l \leq k \leq l.$$

By BISO broadcast channel, we mean the component channels are BISO channels. Binary symmetric channel (BSC) and Binary erasure channel (BEC) are examples of BISO broadcast channels.

Remark 4. As $k = 0$ can be split into 0^+ and 0^- with equal probability $\frac{1}{2}p_0$, we just consider $k = \pm 1, ..., \pm l$ and use $\{p_k, p_{-k} : k = 1, \ldots, l\}$ to denote the transition probabilities. Sometimes shortened to $\{p_k, p_{-k}\}_k$.

Let $x = \mathrm{P}(X{=}0)$. Consider $(Q, \tilde{X})$ such that $Q \sim Bern(\frac{1}{2})$, and $(\tilde{X}|Q{=}0) \sim Bern(\bar{x})$, and $(\tilde{X}|Q{=}1) \sim Bern(x)$. Now $\tilde{X}$ is uniformly distributed. Due to channel symmetry we have

$$I(X; Y) = I(\tilde{X}; \tilde{Y}|Q) \leq I(\tilde{X}; \tilde{Y}).$$

Hence uniform input distribution is the capacity achieving distribution for any BISO channel.

Partition P of an interval $[a, b]$ is a finite sequence of points $\{t_k\}_k$ such that $a = t_0 < t_1 < t_2 < \ldots < t_N = b$. A partition

P is finer than Q if points of partition P contain those of Q. A common refinement of two partitions P and Q is a new partition consisting of all the points of P and Q.

Definition 10. For a BISO channel with transition probabilities $\{p_k, p_{-k}\}_k$, rearrange $h(\frac{p_k}{p_k+p_{-k}})$ in the ascending order and denote the permutation as π. *BISO partition* is defined as the partition of $[0,1]$ with points $t_k = \sum_{i=1}^{k}(p_{\pi_i} + p_{-\pi_i})$, and $t_0 = 0$. *BISO curve* is defined as the stepwise function $f(t)$ such that $f(t) = h(\frac{p_{\pi_k}}{p_{\pi_k}+p_{-\pi_k}})$ on $(t_{k-1}, t_k]$, and $f(0) = 0$.

For the channel $BSC(p)$, we have the partition as $t_0 = 0, t_1 = 1$ and the curve as $f(t) = h(p)$ on $(0,1]$. For the channel $BEC(e)$, we have the partition as $t_0 = 0, t_1 = 1-e, t_2 = 1$, and the curve as $f(t) = 0$ on $(0, 1-e]$ and $f(t) = 1$ on $(1-e, 1]$.

Definition 11. For a BISO channel with BISO curve $f(t)$, the *Lorenz curve* $F(t)$ is defined as $F(t) = \int_0^t f(\tau)\mathrm{d}\tau$. (cumulative curve)

Property 1. *Since $f(t) \in [0,1]$ is non-decreasing we have*

(1) $F(t)$ is non-negative, piecewise linear, and convex.

(2) The slope of the line segments of $F(t)$ is at most 1.

By definition of BISO curve, the length of interval $(t_{k-1}, t_k]$ is $(p_{\pi_k} + p_{-\pi_k})$. Therefore

$$\begin{aligned}
I(X;Y) &= \sum_{k>0} (p_k + p_{-k}) h(x * h^{-1}(h(\frac{p_k}{p_k + p_{-k}}))) \\
&\qquad - \sum_{k>0} (p_k + p_{-k}) h(\frac{p_k}{p_k + p_{-k}}) \qquad (3.1) \\
&= \int_0^1 h(x * h^{-1}(f(\tau)))\mathrm{d}\tau - \int_0^1 f(\tau)\mathrm{d}\tau \\
&= \int_0^1 h(x * h^{-1}(f(\tau)))\mathrm{d}\tau - F(1)
\end{aligned}$$

Thus a finer partition does not change $I(X;Y)$ and in particular channel capacity. Indeed capacity, achieved by $x = \frac{1}{2}$, is $1-F(1)$.

3.2 Partial orders on BISO broadcast channel

3.2.1 More capable comparability

We will establish a sufficient condition in Theorem 4 for determining whether two BISO channels are comparable using the more capable partial order. Towards this, the following three lemmas are needed.

Lemma 1. *Given BISO channels $X \to Y$ and $X \to Z$ with BISO curves $f(t)$ and $g(t)$, respectively. Let the common refinement of these two BISO partitions be $\{t_k : k = 0, \ldots, \hat{N}\}$, and*

$\xi_k = t_k - t_{k-1}$. *Then*

$$F(t_i) = \sum_{k=1}^{i} \xi_k f(t_k) \le \sum_{k=1}^{i} \xi_k g(t_k) = G(t_i), \quad i = 1, \dots, \hat{N}$$

if and only if the Lorenz curve $F(t) \le G(t)$ *for all* $t \in [0,1]$.

Proof. The *if* direction is obvious. For the other direction, we prove by contradiction. Let t^* be a point where $F(t^*) > G(t^*)$. Clearly $t^* \in (t_{j-1}, t_j)$ for some j. Since $F(t_{j-1}) \le G(t_{j-1})$, it is necessary that $f(t) > g(t)$ for $t \in (t_{j-1}, t_j)$. However integrating from t^* to t_j, we have that $F(t_j) > G(t_j)$, which contradicts. □

The following lemma is well-known.

Lemma 2 (Lemma 2 in [16]). *The function* $h(x * h^{-1}(y))$ *is strictly convex in* y. *(Key ingredient of Mrs. Gerber's lemma)*

Lemma 3 (Lemma 1 in [9]). *Let* $x_1, ..., x_l$ *and* $y_1, ..., y_l$ *be non-decreasing sequences of real numbers. Let* $\xi_1, ..., \xi_l$ *be a sequence of real numbers such that*

$$\sum_{j=k}^{l} \xi_j x_j \ge \sum_{j=k}^{l} \xi_j y_j, \quad 1 \le k \le l$$

with equality for $k = 1$. *Then for any convex function* Λ,

$$\sum_{j=1}^{l} \xi_j \Lambda(x_j) \ge \sum_{j=1}^{l} \xi_j \Lambda(y_j).$$

Theorem 4. *Given BISO channels $X \to Y$ and $X \to Z$ with Lorenz curves $F(t)$ and $G(t)$, respectively. Further let $F(1) = G(1)$, i.e. channels have same capacity. If $F(t) \leq G(t)$ then Y is more capable than Z.*

Proof. Using Lemma 1 we know that

$$F(t_i) = \sum_{k=1}^{i} \xi_k f(t_k) \leq \sum_{k=1}^{i} \xi_k g(t_k) = G(t_i), \quad i = 1, \ldots, \hat{N}$$

and since $F(1) = G(1)$ we have equality at $i = \hat{N}$. Using Lemma 3 and by noticing that $f(t_k)$ and $g(t_k)$ are both nondecreasing we have

$$\sum_{j=1}^{\hat{N}} \xi_j \Lambda(f(t_j)) \geq \sum_{j=1}^{\hat{N}} \xi_j \Lambda(g(t_j))$$

for any convex function Λ. Taking $\Lambda(y) = h(x * h^{-1}(y)) - y$ we obtain that

$$\sum_{j=1}^{\hat{N}} \xi_j h\big(x * h^{-1}(f(t_j))\big) - \sum_{j=1}^{\hat{N}} \xi_j f(t_j)$$
$$\geq \sum_{j=1}^{\hat{N}} \xi_j h\big(x * h^{-1}(g(t_j))\big) - \sum_{j=1}^{\hat{N}} \xi_j g(t_j).$$

From (3.1) this is equivalent to

$$I(X;Y) \geq I(X;Z), \forall p(x).$$

Thus the theorem is established. $\square$

For reasons that will be apparent later (Lemma 5) it is useful to zoom in the subclass of BISO channels that have the same channel capacity C. For instance $BSC(p)$, with $1 - h(p) = C$, belongs to this class. Similarly for $BEC(e)$ with $1 - e = C$.

Notation: Let $BISO(C)$ denote an arbitrary BISO channel that has capacity C. To abuse notation, we denote $BSC(C)$ and $BEC(C)$ as the binary symmetric channel and the binary erasure channel with capacity C, respectively.

Corollary 2. *$BISO(C) \overset{mc}{\geq} BSC(C)$.*

Proof. From Theorem 4 it suffices that the Lorenz curves satisfy $G(t) \leq F_{BSC}(t), t \in [0, 1]$. Observe that $G(0) = F_{BSC}(0) = 0$, $G(1) = F_{BSC}(1)$ and that $F_{BSC}(t)$ is the straight-line connecting 0 and $F_{BSC}(1)$. The convexity of Lorenz curve $G(t)$ implies that $G(t) \leq F_{BSC}(t), t \in [0, 1]$. □

Corollary 3. *$BEC(C) \overset{mc}{\geq} BISO(C)$.*

Proof. Similar to above it suffices that the Lorenz curves satisfy $F_{BEC}(t) \leq G(t), t \in [0, 1]$. $F_{BEC}(t) = 0, t \in [0, 1 - e]$ and hence $F_{BEC}(t) \leq G(t), t \in [0, 1 - e]$. Combining $F_{BEC}(1) = G(1)$ and (comparing slopes) $F'_{BEC}(t) = f_{BEC}(t) = 1 \geq g(t) = G'(t), t \in (1 - e, 1]$, we also have $F_{BEC}(t) \leq G(t), t \in [1 - e, 1]$. □

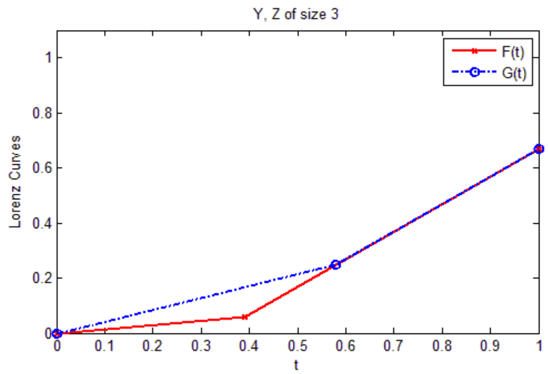

Figure 3.1: Lorenz curves for BISO channels with the same capacity and output of size 3.

Corollary 4. *Two $BISO(C)$ channels whose output alphabet sizes are at most 3 are always more capable comparable.*

Proof. For BISO channel $X \to Y$ with transition probabilities $\{p_{-1}, p_0, p_1\}$, $k = 0$ is split equally into 0^+ and 0^-. Thus the Lorenz curve $F(t)$ contains two sloping lines: one with slope $h(\frac{p_{0+}}{p_{0+}+p_{0-}}) = 1$, and the other not bigger than 1. Given two Lorenz curves of this kind, $F(t)$ and $G(t)$, with $F(1) = G(1)$, then either $F(t) \leq G(t)$ for all $t \in [0, 1]$ or $F(t) \geq G(t)$ for all $t \in [0, 1]$ (Figure 3.1). According to Theorem 4, these two channels are more capable comparable. □

Remark 5. Not all BISO channels with the same capacity are more capable comparable. A counter example is the following: Consider BISO channels $X \to (Y, Z)$ with transition probabilities according to:

$$\mathrm{P}(Y{=}i|X{=}0) = a_i, -2 \leq i \leq 2$$

$$\mathrm{P}(Z{=}j|X{=}0) = b_j, -2 \leq j \leq 2$$

where $a_{-2} = 0.061$, $a_{-1} = a_1 = \frac{1}{2}(1 - 10a_{-2})$, $a_2 = 9a_{-2}$ and $b_{-2} = 0.0634977$, $b_{-1} = \frac{1}{5}(1 - b_{-2})$, $b_1 = \frac{4}{5}(1 - b_{-2})$, $b_2 = 0$. One can verify that the channels have same capacity, but are not more capable comparable.

3.2.2 More capable and essentially less noisy

In this section we will establish that these two partial orders, restricted to BISO channels with capacity C, are *inverse* of each other. This is counter-intuitive as more capable and essentially less noisy are two notions of saying that one receiver is superior to another receiver.

Below (for a complete argument see Lemma 1 in [13]) we note that the uniform input distribution forms a sufficient class for any BISO broadcast channel. Thus in the following when talking about essentially less noisy, we automatically assume it is under this sufficient class.

Claim 4. *For any BISO broadcast channel, the uniform input distribution* $\mathrm{P}(X{=}0) = \frac{1}{2}$ *forms a sufficient class.*

Proof. The following construction suffices. Let $j, k \in \{0, 1\}$, then define p.m.f.

$$Q(\tilde{U}{=}(u,j), \tilde{V}{=}(v,k), \tilde{X}{=}x) = \begin{cases} \frac{1}{2}\mathrm{P}(U{=}u, V{=}v, X{=}x \oplus j) & j = k \\ 0 & j \neq k \end{cases}$$

$\square$

Lemma 4.

$$BISO_1(C) \overset{mc}{\geq} BISO_2(C) \iff BISO_2(C) \overset{eln}{\geq} BISO_1(C)$$

Proof. Assume component channels Y and Z have same capacity C and $Y \overset{mc}{\geq} Z$. When $\mathrm{P}(X{=}0) = \frac{1}{2}$ we have for all U such that $U \to X \to (Y, Z)$

$$\begin{aligned} I(U;Y) &= I(X;Y) - I(X;Y|U) \\ &= C - I(X;Y|U) \\ &= I(X;Z) - I(X;Y|U) \\ &= I(U;Z) + I(X;Z|U) - I(X;Y|U) \\ &\leq I(U;Z) \end{aligned}$$

where the last inequality follows from $Y \overset{mc}{\geq} Z$. From Claim 4, $\mathrm{P}(X{=}0) = \frac{1}{2}$ is a sufficient class of input distribution, by definition, $Z \overset{eln}{\geq} Y$.

Assume $Z \overset{eln}{\geq} Y$. The proof follows by contradiction. Suppose there is a value x such that when $\mathrm{P}(X{=}0) = x$, we have

$I(X;Z) - I(X;Y) = \delta > 0$, then consider a $U \sim Bern(\frac{1}{2})$ such that $\mathrm{P}(X{=}0|U{=}0) = x = \mathrm{P}(X{=}1|U{=}1)$. Observe that, from the symmetry $I(X;Z|U) - I(X;Y|U) = \delta > 0$. However since $\mathrm{P}(X{=}0) = \frac{1}{2}$, using a similar decomposition we see that

$$\begin{aligned} I(U;Y) &= I(U;Z) + I(X;Z|U) - I(X;Y|U) \\ &= I(U;Z) + \delta > I(U;Z) \end{aligned}$$

contradicting the assumption $Z \overset{eln}{\geq} Y$. □

The following lemma is an immediate consequence of Corollaries 2, 3, and Lemma 4.

Lemma 5.

(1) $BEC(C) \overset{mc}{\geq} BISO(C) \overset{mc}{\geq} BSC(C)$*,*

(2) $BSC(C) \overset{eln}{\geq} BISO(C) \overset{eln}{\geq} BEC(C)$*.*

This leads us to one of the main results in this section.

Theorem 5. *For any three numbers* $0 \leq C_1 \leq C_2 \leq C_3$ *we have*

(1) $BEC(C_3) \overset{mc}{\geq} BISO(C_2) \overset{mc}{\geq} BSC(C_1)$*,*

(2) $BSC(C_3) \overset{eln}{\geq} BISO(C_2) \overset{eln}{\geq} BEC(C_1)$*.*

Proof. Suppose $C_a < C_b$, then $BSC(C_a)$ and $BEC(C_a)$ are degraded versions of $BSC(C_b)$ and $BEC(C_b)$ respectively. Hence from Lemma 5 we have

$$BEC(C_3) \overset{mc}{\geq} BEC(C_2) \overset{mc}{\geq} BISO(C_2) \overset{mc}{\geq} BSC(C_2) \overset{mc}{\geq} BSC(C_1)$$
$$BSC(C_3) \overset{eln}{\geq} BSC(C_2) \overset{eln}{\geq} BISO(C_2) \overset{eln}{\geq} BEC(C_2) \overset{eln}{\geq} BEC(C_1) \qquad \square$$

The following corollary is immediate.

Corollary 5. *Superposition coding region is the capacity region for a BISO broadcast channel if any one of the channels is either a BSC or a BEC.*

Proof. Superposition coding is optimal both for more capable comparable channels [2] and for essentially less noisy comparable channels [13]. From Theorem 5, if any one of the channels is either a BSC or a BEC, then the channels are either more capable comparable or essentially less noisy comparable. □

Remark 6. In [13] the capacity region of a BSC/BEC broadcast channel was established. Corollary 5 generalizes this result to only requiring that one of the BISO channels is a BEC or a BSC.

3.3 Comparison of bounds for BISO broadcast channel

The following lemma and Lemma 7 in Appendix generalize the result by Wyner and Ziv [16] for BSC broadcast channels. In [2] it was shown that superposition coding is indeed optimal when the two channels are more capable comparable.

Lemma 6. *Consider a two-receiver BISO broadcast channel. Consider the following region formed by taking the union of rate pairs (R_1, R_2) satisfying*

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

over all $p(u,x)$. Then the same region can be realized by restricting to a binary U such that $U \to X \sim BSC$ and $\mathrm{P}(X{=}0) = \frac{1}{2}$.

Proof. The proof is presented in the Appendix. □

Remark 7. For BISO broadcast channels since $\mathrm{P}(X{=}0) = \frac{1}{2}$ is a common sufficient distribution, it can be shown that the UV outer bound matches the Körner–Marton outer bound.

Consider $\mathrm{P}(X{=}0) = \frac{1}{2}$ and $U \to X \sim BSC(s_1)$ where $s_1 = \mathrm{P}(X{=}1|U{=}0)$, similarly let $V \to X \sim BSC(s_2)$. Let $I(U;Y) = f_1(s_1)$, and $I(V;Z) = f_2(s_2)$. It is clear from symmetry that $f_1(s) = f_1(1-s), f_2(s) = f_2(1-s)$.

From Lemma 6 it follows that UVOB can be written as the union of rate pairs R_1, R_2 satisfying

$$\begin{aligned} R_1 &\le f_1(s_1) \\ R_2 &\le f_2(s_2) \\ R_1 + R_2 &\le f_1(s_1) + C - f_2(s_1) \\ R_1 + R_2 &\le f_2(s_2) + C - f_1(s_2). \end{aligned} \tag{3.2}$$

for some $0 \le s_1, s_2 \le \frac{1}{2}$.

Let

$$I = \{s \in [0, 0.5] : f_1(s) > f_2(s)\}$$
$$J = \{s \in [0, 0.5] : f_1(s) < f_2(s)\}.$$

The following result relates the equivalence of the various bounds and their relation to whether the channels are more capable comparable. (Randomized time division region is the same as the Marton's inner bound due to Corollary 1.)

Theorem 6. *For a BISO broadcast channel with component channels $BISO_1(C)$ and $BISO_2(C)$, the followings are equivalent:*

(a) $BISO_1(C)$ and $BISO_2(C)$ are not more capable comparable

(b) $TD \subset UVOB$

(c) There exists $s_1 \in I, s_2 \in J$ such that $f_1(s_1) + f_2(s_2) > C$

(d) $TD \subset MIB$

(e) $MIB \subset UVOB$.

Proof. The proof is presented in the Appendix. □

Corollary 6. *For a BISO broadcast channel with component channels having the same capacity, superposition coding is optimal if and only if the channels are more capable comparable.*

Proof. If superposition coding region is indeed the capacity region, then we have $R_1 + R_2 \leq I(X;Y) \leq C$. Further since the two channels have the same capacity, we have the TD region is optimal. From Theorem 6 we have that the channels are more capable comparable. □

Remark 8. A characterization of when superposition coding is optimal for two-receiver broadcast channels is open in general. It is known that superposition coding is optimal when the channels are either essentially more capable comparable or essentially less noisy comparable [13] - two incompatible notions. However a converse statement is still unknown.

Observation 1. From Remark 5 we know that there exists a pair of channels $BISO_1(C)$ and $BISO_2(C)$ which are not more capable comparable. Hence from Theorem 6 we know that the capacity region is strictly larger than TD. However, if we replace $BISO_2(C)$ by $BEC(C)$, a more capable channel, then the capacity of the broadcast channel formed by $BISO_1(C)$ and $BEC(C)$ is the TD region (Corollary 3). Thus replacing by a more capable channel can *strictly* reduce the capacity region.

3.4 A new partial order

We now introduce a natural operational partial order among broadcast channels.

Definition 12. Receiver $\tilde{Z}$ is a *better receiver* than Z if the capacity region of $X \to (Y, \tilde{Z})$ contains that of $X \to (Y, Z)$ for every channel $X \to Y$. In other words, if we replace receiver Z by receiver $\tilde{Z}$ then the capacity region will not decrease.

Remark 9. Note that the capacity region of a broadcast channel just depends on the marginal distributions $X \to Y$, $X \to Z$, and hence the definition makes sense.

From Observation 1 we know that a more capable receiver is not necessarily a better receiver. However we will show that if $\tilde{Z}$ is a less noisy receiver than Z, then $\tilde{Z}$ is indeed a better receiver than Z.

Claim 5. *If $\tilde{Z}$ is a less noisy receiver than Z, then $\tilde{Z}$ is a better receiver than Z.*

Proof. The capacity region of a discrete memoryless broadcast channel has the following n-letter characterization. Consider the region $\mathcal{R}_n$ defined as the union of rate pairs (R_1, R_2) that satisfy

$$R_1 \leq \frac{1}{n} I(U; Y^n)$$
$$R_2 \leq \frac{1}{n} I(V; Z^n)$$

for some $p(u)p(v)p(x^n|u, v)$. It is known that the capacity region is $\lim_n \mathcal{R}_n$. (It is clear that this is achievable, and a converse follows by setting $U = M_1$ and $V = M_2$ and applying Fano's inequality.) Observe for $j = n, \ldots, 1$

$$\begin{aligned} I(V; Z^j, \tilde{Z}_{j+1}^n) &= I(V; Z^{j-1}, \tilde{Z}_{j+1}^n) + I(V; Z_j | Z^{j-1}, \tilde{Z}_{j+1}^n) \\ &\leq I(V; Z^{j-1}, \tilde{Z}_{j+1}^n) + I(V; \tilde{Z}_j | Z^{j-1}, \tilde{Z}_{j+1}^n) \\ &= I(V; Z^{j-1}, \tilde{Z}_j^n). \end{aligned}$$

By taking extreme points of this chain we obtain $I(V; Z^n) \leq I(V; \tilde{Z}^n)$. Claim follows from the expression of the capacity region stated above. $\square$

Appendix

Proof to Lemma 6

Proof. Let $\tilde{U} = (U, Q)$, where $Q \sim Bern(\frac{1}{2})$ is independent of U such that

$$\mathrm{P}(X{=}0|(U,Q){=}(u,0)) = \mathrm{P}(X{=}0|U{=}u)$$
$$\mathrm{P}(X{=}0|(U,Q){=}(u,1)) = 1 - \mathrm{P}(X{=}0|U{=}u)$$

This induces an $\tilde{X} \sim Bern(\frac{1}{2})$. It is straightforward that

$$I(\tilde{U};\tilde{Z}) \geq I(U;Z),$$
$$I(\tilde{X};\tilde{Y}|\tilde{U}) = I(X;Y|U),$$
$$I(\tilde{X};\tilde{Y}) \geq I(X;Y).$$

Thus for every pair of (U, X), replacing it to $(\tilde{U}, \tilde{X})$ leads to a larger achievable region. Denote this class of $(\tilde{U}, \tilde{X})$ as $\mathcal{Q}$.

Hence it suffices to maximize over (U, X) in $\mathcal{Q}$. Since X is uniform, the third inequality remains constant. Therefore, to compute the extreme points, we proceed to compute the distribution (U, X) in $\mathcal{Q}$ that maximizes $\lambda I(U;Z) + \big(I(U;Z) + I(X;Y|U)\big)$. Reformulate it as

$$(\lambda+1)I(X;Z) + I(X;Y|U) - (\lambda+1)I(X;Z|U).$$

Let $f(p) = I(X;Y) - (\lambda+1)I(X;Z)$, where $p = \mathrm{P}(X{=}0)$. Notice that $f(p) = f(1-p)$, suppose $p_\lambda \in [0, \frac{1}{2}]$ and $1 - p_\lambda$ maximize $f(p)$. Construct $U \to X \sim BSC(p_\lambda)$, then $I(X;Y|U) - (\lambda+1)I(X;Z|U)$ is maximized; construct $U \sim Bern(\frac{1}{2})$, then $I(X;Z)$ is maximized since $\mathrm{P}(X{=}0) = \frac{1}{2}$. Notice this construction falls into class $\mathcal{Q}$, hence finishes the proof. □

The same proof can also be used to establish the following lemma.

Lemma 7. *Consider a two-receiver BISO broadcast channel. Consider the following superposition coding region formed by taking the union of rate pairs* (R_1, R_2) *satisfying*

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

over all $p(u,x)$*. Then the same region can be realized by restricting to a binary* U *such that* $U \to X \sim BSC$ *and* $\mathrm{P}(X{=}0) = \frac{1}{2}$.

Proof to Theorem 6

Proof. (a) $\Rightarrow$ (b): Recalling: Let

$$\begin{aligned} I &= \{s \in [0, 0.5] : f_1(s) > f_2(s)\} \\ J &= \{s \in [0, 0.5] : f_1(s) < f_2(s)\}. \end{aligned}$$

Since the channels are not more-capable comparable, we know that there esists $s_1 \in I$ and $s_2 \in J$. Construct $\tilde{U} \to X$, where $\tilde{U} = U' \times Q$ with binary U' and Q, and probabilities

$$
\begin{aligned}
\mathrm{P}(\tilde{U}{=}(0,0)) &= \frac{1-\varepsilon}{2} & \qquad \mathrm{P}(X{=}0|\tilde{U}{=}(0,0)) &= 1\\
\mathrm{P}(\tilde{U}{=}(0,1)) &= \frac{\varepsilon}{2} & \qquad \mathrm{P}(X{=}0|\tilde{U}{=}(0,1)) &= s_1\\
\mathrm{P}(\tilde{U}{=}(1,0)) &= \frac{1-\varepsilon}{2} & \qquad \mathrm{P}(X{=}1|\tilde{U}{=}(1,0)) &= 1\\
\mathrm{P}(\tilde{U}{=}(1,1)) &= \frac{\varepsilon}{2} & \qquad \mathrm{P}(X{=}1|\tilde{U}{=}(1,1)) &= s_1.
\end{aligned}
$$

Thus, $U' \mapsto X \sim BSC(0)$ conditioned on the event $Q = 0$, $U' \mapsto X \sim BSC(1 - s_1)$ conditioned on $Q = 1$, and further U' is independent of Q with $\mathrm{P}(U'{=}0) = \frac{1}{2}$. We can see that Q is independent of X and hence of Y, Z; thus $I(Q;Y) = I(Q;Z) = 0$. Now

$$
\begin{aligned}
I(\tilde{U};Y) = I(U', Q; Y) &= I(U';Y|Q) + I(Q;Y)\\
&= I(U';Y|Q)\\
&= (1-\varepsilon)I(X;Y) + \varepsilon I(U';Y|Q{=}1)\\
&= (1-\varepsilon)C + \varepsilon f_1(s_1).
\end{aligned}
$$

Similarly, we obtain

$$
I(\tilde{U};Z) = (1-\varepsilon)C + \varepsilon f_2(s_1).
$$

Thus we have

$$
\begin{aligned}
R_1 &\le (1-\varepsilon)C + \varepsilon f_1(s_1) \\
R_2 &\le f_2(s_2) \\
R_1 + R_2 &\le I(\tilde{U};Y) + I(X;Z|\tilde{U}) \\
&= I(\tilde{U};Y) + I(X;Z) - I(\tilde{U};Z) \\
&= (1-\varepsilon)C + \varepsilon f_1(s_1) + C - [(1-\varepsilon)C + \varepsilon f_2(s_1)] \\
&= C + \varepsilon[f_1(s_1) - f_2(s_1)] \qquad (> C) \\
R_1 + R_2 &\le I(V;Z) + I(X;Y|V) \\
&= f_2(s_2) + C - f_1(s_2) \qquad (> C).
\end{aligned}
$$

To show that we can have $(1-\varepsilon)C + \varepsilon f_1(s_1) + f_2(s_2) > C$, we just need to choose small ε to ensure $f_2(s_2) > \varepsilon[C - f_1(s_1)]$. Since this is clearly possible, we have $UVOB \supset TD$.

(b) $\Rightarrow$ (c): From Equation (3.2), we have the following expression of the boundary of the outer bound,

$$
\begin{aligned}
R_1 &\le I(U;Y) = f_1(s_1) \\
R_2 &\le I(V;Z) = f_2(s_2) \\
R_1 + R_2 &\le I(U;Y) + I(X;Z|U) = f_1(s_1) + C - f_2(s_1) \\
R_1 + R_2 &\le I(V;Z) + I(X;Y|V) = f_2(s_2) + C - f_1(s_2)
\end{aligned}
$$

Clearly for every $s_1 \in I, s_2 \in J$ if $f_1(s_1) + f_2(s_2) \le C$ then from above $UVOB = TD$. However since $UVOB \supset TD$, there exists $s_1 \in I, s_2 \in J$ such that $f_1(s_1) + f_2(s_2) > C$.

(c) $\Rightarrow$ (d): In general, $TD \subseteq RTD \subseteq MIB$. So now it suffices to show there exists an example where the sum rate of RTD region is strictly larger than TD region.

We now compute the maximum sum rate of the RTD region. From Corollary 1 we know that this matches the maximum sum rate of the MIB region.

Consider an auxiliary channel $W \to X$ such that

$$\mathrm{P}(W{=}0) = a, \quad \mathrm{P}(W{=}1) = 1 - a$$
$$\mathrm{P}(X{=}0|W{=}0) = s_2, \quad \mathrm{P}(X{=}0|W{=}1) = s_1$$

where $as_2 + (1-a)s_1 = \frac{1}{2}$.

It is straightforward to check the following

$$I(X;Y|W{=}0) = C - f_1(s_2), \; I(X;Y|W{=}1) = C - f_1(s_1)$$
$$I(X;Z|W{=}0) = C - f_2(s_2), \; I(X;Z|W{=}1) = C - f_2(s_1),$$
$$I(X;Y) = I(X;Z) = C.$$

Then observe that

$$\begin{aligned} &I(W;Y) + \mathrm{P}(W{=}0)I(X;Y|W{=}0) + \mathrm{P}(W{=}1)I(X;Z|W{=}1) \\ &\quad = I(X;Y) + \mathrm{P}(W{=}1)\big(I(X;Z|W{=}1) - I(X;Y|W{=}1)\big) \\ &\quad = C + (1-a)(f_1(s_1) - f_2(s_1)) \end{aligned}$$

where the last inequality holds since $s_1 \in I$.

Similarly

$$I(W;Z) + \mathrm{P}(W{=}0)I(X;Y|W{=}0) + \mathrm{P}(W{=}1)I(X;Z|W{=}1)$$
$$= C + a(f_2(s_2) - f_1(s_2)).$$

Therefore the sum rate of RTD (eq. MIB) for this choice of (W, X) is given by

$$C + \min\{(1-a)(f_1(s_1) - f_2(s_1)), a(f_2(s_2) - f_1(s_2))\}. \quad (3.3)$$

Therefore if (c) is satisfied, i.e. there exists $s_1 \in I, s_2 \in J$, then there exists a (W, X) so that equation (3.3) gives a sum rate strictly larger than C.

Remark 10. A careful reader will notice that the above argument only requires $s_1 \in I, s_2 \in J$ and does not even require $f_1(s_1) + f_2(s_2) > C$. But existence of any $s_a \in I, s_b \in J$ will imply that (a) holds and hence (c) holds.

(d) $\Rightarrow$ (e): Since $TD \subset MIB$, to compute the maximum sum rate of MIB it suffices to maximize over $s_1 \in I, s_2 \in J, 0 < a < 1$ the term

$$C + \min\{(1-a)(f_1(s_1) - f_2(s_1)), a(f_2(s_2) - f_1(s_2))\}.$$

Consider any triple $s_1 \in I, s_2 \in J, 0 < a < 1$. Pick any $\varepsilon > 0$ small enough (will show later how small we require it).

Define $(U, X) = (Q, U_1, X)$ where $\mathrm{P}(Q{=}0) = 1 - a + \varepsilon$, and $\mathrm{P}(Q{=}1) = a - \varepsilon$; and $U_1 \mapsto X \sim BSC(s_1)$ conditioned on $Q = 0$, and $U_1 \mapsto X \sim BSC(0)$ conditioned on $Q = 1$. Further take $\mathrm{P}(U_1{=}0) = \mathrm{P}(U_1{=}1) = \frac{1}{2}$. Observe that this induces $\mathrm{P}(X{=}0) = \mathrm{P}(X{=}1) = \frac{1}{2}$.

Similarly define $(V, X) = (Q', V_1, X)$ where $\mathrm{P}(Q'{=}0) = a + \varepsilon$, $\mathrm{P}(Q'{=}1) = 1 - a - \varepsilon$; and $V_1 \mapsto X \sim BSC(s_2)$ conditioned on $Q' = 0$, and $V_1 \mapsto X \sim BSC(0)$ conditioned on $Q' = 1$. Further take $\mathrm{P}(V_1{=}0) = \mathrm{P}(V_1{=}1) = \frac{1}{2}$. Observe that this also induces $\mathrm{P}(X{=}0) = \mathrm{P}(X{=}1) = \frac{1}{2}$.

Since the distribution of X is consistent there exists a triple (U, V, X) with the same pairwise marginals (U, X) and (V, X) as described earlier. With this choice, UVOB reduces to

$$\begin{aligned}
R_1 &\leq I(U;Y) = (1 - a + \varepsilon) f_1(s_1) + (a - \varepsilon) C \\
R_2 &\leq I(V;Z) = (a + \varepsilon) f_2(s_2) + (1 - a - \varepsilon) C \\
R_1 + R_2 &\leq I(U;Y) + I(X;Z|U) \\
&= C + (1 - a + \varepsilon)(f_1(s_1) - f_2(s_1)) \\
R_1 + R_2 &\leq I(V;Z) + I(X;Y|V) \\
&= C + (a + \varepsilon)(f_2(s_2) - f_1(s_2)).
\end{aligned}$$

Clearly the maximum sum rate of the above region is minimum of the terms

$$C + (1 - a + \varepsilon)(f_1(s_1) - f_2(s_1)), \quad C + (a + \varepsilon)(f_2(s_2) - f_1(s_2)),$$
$$(1 - 2\epsilon)C + (1 - a + \varepsilon)f_1(s_1) + (a + \varepsilon)f_2(s_2)$$

We pick $\varepsilon > 0$ to satisfy

$$\begin{aligned} &(1 - 2\epsilon)C + (1 - a + \varepsilon)f_1(s_1) + (a + \varepsilon)f_2(s_2) \\ &\qquad > C + (1 - a)(f_1(s_1) - f_2(s_1)) \\ \iff &\qquad (1 - a)f_2(s_1) + af_2(s_2) > \varepsilon(2C - f_1(s_1) - f_2(s_2)), \end{aligned}$$

and

$$af_1(s_2) + (1 - a)f_2(s_1) > \varepsilon(2C - f_1(s_1) - f_2(s_2)),$$

then the maximum sum rate of the UVOB expression will be strictly bigger than that of MIB region. Since this is possible for every $s_1 \in I, s_2 \in J, 0 < a < 1$, the maximum sum rate of UVOB is strictly larger than that of MIB. Therefore $UVOB \supset MIB$ or (e) holds.

(e) $\Rightarrow$ (a): Since $MIB \subset UVOB$ clearly implies the channels are not more capable comparable. This is because when the channels are more capable comparable we know from [2] that superposition coding is optimal and that $MIB = CR = UVOB$.

$\square$

Chapter 4

Extended binary inequality

Marton's inner bound refers to union of rate pairs satisfying

$$
\begin{aligned}
R_1 &\leq I(U,W;Y) \\
R_2 &\leq I(V,W;Z) \\
R_1 + R_2 &\leq \min\{I(W;Y), I(W;Z)\} \\
&\qquad + I(U;Y|W) + I(V;Z|W) - I(U;V|W),
\end{aligned}
$$

for all $p(w,u,v,x)$.

To evaluate the sum-rate, the following inequality helps

$$I(U;Y) + I(V;Z) - I(U;V) \leq \max\{I(X;Y), I(X;Z)\}$$

However to evaluate the inner bound, we need a bit more.

Since Marton's inner bound is a convex region, for the boundary points, we consider optimizing the supporting hyperplanes. Notice that the sum-rate constraint is always effective, for $\alpha \geq 1$

we seek to maximize the following

$$
\begin{aligned}
&(\alpha-1)I(U,W;Y)+\min\{I(W;Y),I(W;Z)\}\\
&\quad+I(U;Y|W)+I(V;Z|W)-I(U;V|W)\\
&=(\alpha-1)I(W;Y)+\min\{I(W;Y),I(W;Z)\}\\
&\quad+\alpha I(U;Y|W)+I(V;Z|W)-I(U;V|W)
\end{aligned}
$$

One may guess the following inequality holds for all $\alpha \geq 1$

$$
\begin{aligned}
&\alpha I(U;Y)+I(V;Z)-I(U;V)\\
&\leq \max\{\alpha I(X;Y),I(X;Z)\}
\end{aligned}
\tag{4.1}
$$

It can be shown that this inequality holds for BSSC. However this inequality is not true for AND case and a counter example is provided in [8].

For the other cases, we can prove this inequality using similar way to the proof in Chapter 2. We just state the proof for XOR case in the following section. One may refer to Section 2.3 for notations and Chapter 2 for more details.

4.1 Proof of XOR case

We will show that XOR case cannot attain non-trivial local maximum for the left hand side of (4.1).

Consider an additive perturbation $q(u,v,x) = p(u,v,x) + \varepsilon\lambda(u,v,x)$ for some $\varepsilon \geq 0$. For a valid perturbation, we require $\lambda_{uvx} \geq 0$ if $p(u,v,x) = 0$, which is

$$\lambda_{uvx} \geq 0, \quad \text{if } f(u,v) \neq x$$

Further let the perturbation maintains $p(x)$, that is

$$\sum_{uv} \lambda_{uvx} = 0, \quad \forall x \in \mathcal{X}$$

For any perturbation that satisfies the above conditions at any local maximum $p(u,v,x)$, it must be true that the first derivative cannot be positive. This implies that $\sum_{xuv} \lambda_{uvx} C_{uvx} \leq 0$, where

$$C_{uvx} = -(\alpha-1)\log p_u - \log p_{uv} + \alpha \sum_y p_{y|x} \log p_{uy} + \sum_z p_{z|x} \log p_{vz}$$

Express $\lambda_{u_x v_x x}$ in the term of other λ_{uvx} variables, that is

$$\lambda_{u_x v_x x} = -\sum_{uv \neq u_x v_x} \lambda_{uvx}$$

and substituting into $\sum_{xuv} \lambda_{uvx} C_{uvx} \leq 0$, we have

$$\sum_{xuv: uv \neq u_x v_x} \lambda_{uvx}(C_{uvx} - C_{u_x v_x x}) \leq 0$$

Above holds for any signed $\{\lambda_{uvx} : f(u,v) = x, (u,v) \neq (u_x, v_x)\}$ and any nonnegative $\{\lambda_{uvx} : f(u,v) \neq x\}$, it implies

$$C_{uvx} = C_{u_x v_x x}, \quad \text{if } f(u,v) = x$$

$$C_{uvx} \leq C_{u_x v_x x}, \quad \text{if } f(u,v) \neq x$$

So now we have the following claims:

Claim 6. *Let $f(u,v) = x$, for any (u_1, v_1) we have $C_{u_1v_1x} \le C_{uvx}$, that is*

$$(\alpha - 1)\log\frac{p_{u_1}}{p_u} + \log\frac{p_{u_1v_1}}{p_{uv}} \ge \alpha\sum_y p_{y|x}\log\frac{p_{u_1y}}{p_{uy}} + \sum_z p_{z|x}\log\frac{p_{v_1z}}{p_{vz}}$$

Claim 7. *If $f(u_1, v_1) = f(u_2, v_2) = x$, then*

$$p_{u_1v_1}p_{u_2v_2} \le p_{u_1v_2}p_{u_2v_1}$$

where the equality holds iff $C_{u_1v_2x} = C_{u_2v_1x} = C_{u_1v_1x}(= C_{u_2v_2x})$.

Proof. The proof is finished by noticing that $C_{u_1v_1x} + C_{u_2v_2x} \ge C_{u_1v_2x} + C_{u_2v_1x}$. □

Now for XOR case, $f(0,0) = f(1,1) = 0$, hence from Claim 7 we have $p_{00}p_{11} \le p_{01}p_{10}$; also $f(0,1) = f(1,0) = 1$, hence $p_{00}p_{11} \ge p_{01}p_{10}$. Thus we have

$$p_{00}p_{11} = p_{01}p_{10}$$

and by Claim 7, this holds iff $C_{010} = C_{100} = C_{000} = C_{110}$ and $C_{001} = C_{111} = C_{011} = C_{101}$. In particular, $C_{000} = C_{010}$ and $C_{001} = C_{011}$ imply that

$$\log\frac{p_{00}}{p_{01}} = \sum p_{z|0}\log\frac{p_{0z}}{p_{1z}} = \sum p_{z|1}\log\frac{p_{0z}}{p_{1z}}.$$

Notice $p_{0z} = p_{00}p_{z|0} + p_{10}p_{z|1}$, take a weighted sum, we get

$$(p_{00} + p_{10})\log\frac{p_{00}}{p_{01}} = \sum p_{0z}\log\frac{p_{0z}}{p_{1z}}$$

From above and using K-L divergence, we have

$$\log \frac{p_{00}}{p_{01}} \geq \log \frac{p_{00}+p_{10}}{p_{11}+p_{01}} = \log \frac{p_{00}}{p_{01}}$$

where the last step holds since $p_{00}p_{11} = p_{01}p_{10}$. Now that the K-L divergence inequality is indeed an equality, we require

$$\frac{p_{00}}{p_{01}} \equiv \frac{p_{0z}}{p_{1z}} = \frac{p_{00}p_{z|0} + p_{10}p_{z|1}}{p_{11}p_{z|0} + p_{01}p_{z|1}}$$

From the above we obtain

$$(p_{01} - p_{11})(p_{z|0} - p_{z|1}) \equiv 0.$$

Similarly from $C_{100} = C_{110}$ and $C_{101} = C_{111}$, we can obtain

$$(p_{10} - p_{11})(p_{y|0} - p_{y|1}) \equiv 0.$$

Now we have two cases

1: $p_{y|0} \equiv p_{y|1}$, or $p_{z|0} \equiv p_{z|1}$. In this case the Theorem holds (similar to special setting SS1).

2: $p_{01} = p_{11}$, $p_{10} = p_{11}$. Combining this with $p_{00}p_{11} = p_{01}p_{10}$ one obtains that $p_{uv} = 1/4$, and as a result U, V and X are mutually independent. The Theorem holds (similar to special setting SS2).

If neither of these two cases is satisfied, there would be no local maximum.

4.2 A conjecture on extending the inequality

We propose a conjecture on the binary inequality. To this end we need the notion of concave envelope.

Definition 13. The concave envelope of $f(x)$ is defined as

$$\mathfrak{C}[f] = \inf\{g(x) : g \geq f \text{ and } g \text{ is concave}\}$$

If we are considering a functional defined on the space of probability distributions with finite alphabets, say $f(X) = I(X;Y)$, since the space is a convex set with finite dimensions, one can argue from Fenchel–Eggleston–Caratheodory that

$$\mathfrak{C}[f](X) = \max_{p(u|x), |\mathcal{U}| \leq |\mathcal{X}|} I(X;Y|U)$$

Conjecture 1. *For $(U,V) \to X \to (Y,Z)$, the following*

$$\alpha I(U;Y) + I(V;Z) - I(U;V) \leq \mathfrak{C}[\max\{\alpha I(X;Y), I(X;Z)\}]$$

holds for all $\alpha \geq 1$ and $p(u,v,x)$ over binary X.

Remark 11. As stated earlier, one only need to prove this conjecture when AND case achieves the maximum of left hand side.

Chapter 5

Conclusion

In Chapter 2, an information theoretic inequality is established for binary input broadcast channels. And it can be used to show that the sum-rate given by Marton's inner bound is equivalent to that given by randomized time-division strategy. In the proof, we borrow and generalize the perturbation method used by Gohari and Ananthram.

In Chapter 3, we look at more capable and essentially less noisy partial orders in BISO broadcast channels. We establish the capacity regions for a class of them and also show some other results related to the evaluations of various bounds. Some of the results are contrary to popular intuition and hence BISO broadcast channels can serve as a simple class from which we can improve our understanding of various relations. We hope that some of the results presented can invoke a careful rethinking of those notions of dominance between receivers.

In Chapter 4, we make a conjecture which extends the binary inequality proved in Chapter 2. This conjecture can help evaluate Marton's inner bound for binary input broadcast channels.

Bibliography

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

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

[3] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[4] Y. Geng, A. Gohari, C. Nair, and Y. Yu. The capacity region of classes of product broadcast channels. *International Symposium on Info. Theory*, pages 1549–1553, 2011.

[5] Y. Geng, V. Jog, C. Nair, and Z. Wang. An information inequality and evaluation of Marton's inner bound for binary input broadcast channels. *submitted to IEEE Trans. Info. Theory*.

[6] Y. Geng, C. Nair, S. Shamai, and Z. Wang. On broadcast channels with binary inputs and symmetric outputs.

International Symposium on Info. Theory, pages 545–549, 2010.

[7] A. Gohari and V. Anantharam. Evaluation of Marton's inner bound for the general broadcast channel. *International Symposium on Info. Theory*, pages 2462–2466, 2009.

[8] A. Gohari, C. Nair, and V. Anantharam. On Marton's inner bound for broadcast channels. *International Symposium on Info. Theory*, 2012.

[9] B. Hajek and M. Pursley. Evaluation of an achievable rate region for the broadcast channel. *IEEE Trans. Info. Theory*, IT-25:36–46, Jan 1979.

[10] V. Jog and C. Nair. An information inequality for the bssc channel. *Proceedings of the ITA Workshop*, 2010.

[11] J. Körner and K. Marton. Images of a set via two channels and their role in multi-user communication. *IEEE Trans. Info. Theory*, IT-23:751–761, Nov 1977.

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

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

[14] 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, Jan 2007.

[15] C. Nair and Z. V. Wang. On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels. *Proceedings of the ITA Workshop*, 2008.

[16] A. Wyner and J. Ziv. A theorem on the entropy of certain binary sequences and applications: Part I. *IEEE Trans. Info. Theory*, IT-19(6):769–772, Nov 1973.