An information inequality and evaluation of Marton's inner bound for binary input broadcast channels

Yanlin Geng Varun Jog Chandra Nair* Vincent Wang.

January 5, 2011

Abstract

We establish an information inequality that is intimately connected to the evaluation of the sum rate given by Marton's inner bound for two receiver broadcast channels with a binary input alphabet. The inequality implies that randomized time-division strategy indeed achieves the sum rate of Marton's inner bound for all binary input broadcast channels. To deduce this, we produce a new cardinality bound for evaluating the sum-rate for Marton's inner bound for all broadcast channels. Using these tools we explicitly evaluate the inner and outer bounds for the binary skew-symmetric broadcast channel and demonstrate a gap between the bounds; in the process correcting the evaluation of the outer bound to an earlier published result. This is the first example where the gaps between the best inner and outer bounds are explicitly determined.

1 Introduction

A two-receiver broadcast channel models the communication scenario where two (independent) messages are to be transmitted from a sender X to two receivers Y, Z. Each receiver is interested in decoding his/her message. A transition probability matrix given by $p(y, z|x)$ models the stochastic nature of the errors introduced during the communication. For formal definitions and early results the reader can refer to [1, 2].

1.1 Background

The following region obtained by Marton[3] represents the best-known achievable region to-date:

Bound 1. *[3] (Marton's inner bound) The set of rate-pairs (R_1, R_2) satisfying the following constraints:*

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

for any set of random variables (U, V, W) such that $(U, V, W) \to X \to (Y, Z)$ forms a Markov chain is achievable.

Recently Gohari and Ananthram [4] used a remarkable perturbation-based argument to establish that it suffices to consider (U, V, W) with alphabet sizes bounded by $|\mathcal{U}| \le |\mathcal{X}|, |\mathcal{V}| \le |\mathcal{X}|, |\mathcal{W}| \le |\mathcal{X}| + 4$ to compute the extreme points of Bound 1. In general the computation of Marton's inner bound is difficult, and prior to [4], this bound was not strictly evaluatable. Even with these bounds on cardinalities, explicit evaluation of the bounds is still a difficult task.

The following region represents the best-known outer-bound to the capacity region of the broadcast channel with private messages.

*This work was partially supported by a grant from the University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08), and by a direct grant for Research at the Chinese University of Hong Kong. Part of this work was done when Varun Jog was a summer intern at CUHK.

Bound 2. *[5] (UV outer bound) The union of rate-pairs* (R_1, R_2) *satisfying the following constraints:*

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

over all pairs of random variables (U, V) *such that* $(U, V) \to X \to (Y, Z)$ *forms a Markov chain forms an outer-bound to the capacity region of the broadcast channel.*

The capacity regions of special classes of broadcast channels have been established and in every case it turns out that Bounds 1 and 3 agree. Further, there have been attempts to improve the bounds, but it has been shown [6] that the UV outer bound is still the best (for private messages). In order to study whether the Bounds 1 and 3 are indeed different or whether they are different representations of the same region, the authors [7] studied a particular channel called the binary skew-symmetric broadcast channel (BSSC) shown in Figure 1.

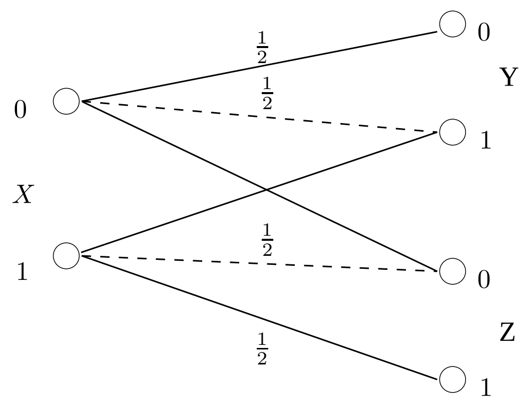

Figure 1: Binary Skew-Symmetric Channel(BSSC)

The authors conjectured that for BSSC the following inequality holds:

$$I(U;Y) + I(V;Z) - I(U;V) \leq \max\{I(X;Y), I(X;Z)\}, \tag{1}$$

for all $(U, V) \to X \to (Y, Z)$ that forms a Markov chain. The authors further showed that, assuming (1) holds, the Bounds 1 and 3 *differed* for BSSC.

In [4], the authors established that Bounds 1 and 3 were indeed different for BSSC without actually establishing that (1) was true. They verified that (1) was indeed plausible by confirming it for a large number of (randomly-generated) samples from the cardinality constrained space.

In this paper we establish that (1) is true and also that this is true for any binary input broadcast channel. This paper unifies the results in two papers([8], [9]) and adds a few other results, along with providing complete details of proofs.

1.2 Summary of results

The main results of the paper is the following:

Theorem 1. *Consider a five tuple of random variables* (U, V, X, Y, Z) *such that* $(U, V) \to X \to (Y, Z)$ *forms a Markov chain and further let* $|\mathcal{X}| = 2$*. Then the following inequality holds:*

$$I(U;Y) + I(V;Z) - I(U;V) \leq \max\{I(X;Y), I(X;Z)\}. \tag{2}$$

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

Theorem 1 generalizes (1) to be true for *every* binary-input broadcast channel. Combining this result with the cardinality bounds in Theorem 2, we also establish that the maximum sum-rate given by Marton's coding strategy indeed matches that given via the randomized time-division strategy [5], 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 $|\mathcal{W}| = 2$.

Theorem 3. *For the BSSC channel shown in Figure 1 the following is true*

- *The maximum sum-rate of Marton's inner bound evaluates to 0.36164288...*
- *The maximum sum-rate of UV outer bound evaluates to 0.3725562...*
- *The maximum sum-rate of the Korner-Marton outer bound evaluates to 0.3743955..*

Remark: This constitutes the first example where the gap between the inner and outer bounds have been quantified.

1.2.1 Randomized time-division strategy

Randomized time-division (R-TD) strategy [5] corresponds to an achievable strategy for the following setting of (U, V, W) in Bound 1: $W = 0$ implies that $U = X, V = \emptyset$; and $W = 1$ implies that $V = X, U = \emptyset$ (where $\emptyset$ refers to the trivial random variable). Observe that this 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.

1.2.2 Relationship between Theorem 1 and Γ_5^*

Recently there has been a lot of interest in information inequalities and the study of the structure of the entropic space Γ_N^*. Theorem 1 refers to a subset, $\mathcal{S}$, of points in Γ_5^*: those corresponding to a five tuple of random variables (U, V, X, Y, Z) such that $(U, V) \to X \to (Y, Z)$ forms a Markov chain and with a binary constraint on the cardinality of X, i.e. $|\mathcal{X}| = 2$. It shows that the points in $\mathcal{S}$ have to lie in the union of two half-spaces induced by the two hyperplanes:

$$\begin{aligned} I(U;Y) + I(V;Z) - I(U;V) &\leq I(X;Y) \\ I(U;Y) + I(V;Z) - I(U;V) &\leq I(X;Z). \end{aligned}$$

Since the inequalities are tight, $\mathcal{S}$ is not a convex region in general. Consider the Blackwell channel shown in Figure 2.

For this channel, consider $U = Y, V = Z$ and $X \sim [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$. Observe that $(U, V) \to X \to Y, Z$ is still Markov and that $I(U;Y) + I(V;Z) - I(U;V) = H(Y,Z) = \log_2 3 > 1 \geq \max\{I(X;Y), I(X;Z)\}$. This shows that in particular $\mathcal{S} \subsetneq \Gamma_5^*$, and hence this is an example[1] of an inequality that cannot be deduced by even the knowledge of Γ_5^*.

Organization of the paper: In the next section we will prove Theorem 2, and show how Corollary 1 follows from Theorems 1 and 2. In the subsequent section we will prove Theorem 1; and finally we will establish Theorem 3.

[1] This is the first known such example to the best of the knowledge of the authors.

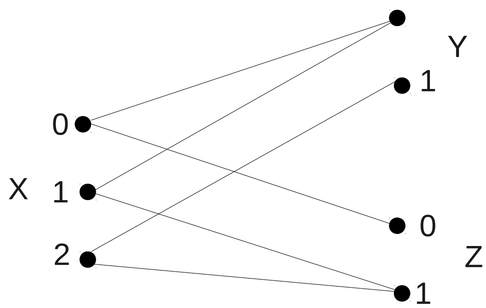

Figure 2: Blackwell broadcast channel

2 Proof of Theorem 2 and Corollary 1

2.1 Proof of Theorem 2

From [4] we know that it suffices to consider $|\mathcal{U}| \leq |\mathcal{X}|$, $|\mathcal{V}| \leq |\mathcal{X}|$ and $|\mathcal{W}| \leq |\mathcal{X}|+4$ to obtain the maximum of Marton's inner bound. Indeed from the standard Fenchel-Bunt extension to Caratheodory's theorem, it will follow that we can restrict ourselves to $|\mathcal{W}| \leq |\mathcal{X}|+1$, if we are just interested in computing the inner bound. The purpose of this Theorem is to show how one may reduce the cardinality further to $|\mathcal{W}| \leq |\mathcal{X}|$. This (mild but non-trivial) improvement is very useful when one needs to explicitly evaluate the bounds, as we are doing in this paper.

Consider the term

$$\max_{p(u,v,w,x)} \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$$

subject to a fixed $p(x)$.

It is known that the maximum sum-rate is one of $I(X;Y)$, $I(X;Z)$, or can be attained at a $p^*(w,u,v,x)$ that satisfies: $I(W;Y) = I(W;Z)$, $I(U,W;Y) > I(U,W;Z)$, $I(V,W;Z) > I(V,W;Y)$. The proof is presented here for completeness.

Proof. If $I(U,W;Y) \leq I(U,W;Z)$ then observe that

$$SR \leq I(U,W;Y) + I(V;Z|W) - I(U;V|W) \leq I(U,W;Y) + I(V;Z|U,W) \leq I(V,U,W;Z) \leq I(X;Z).$$

Similarly if $I(V,W;Z) \leq I(V,W;Y)$ then observe that $I(X;Y)$ will be the sum-rate.

Thus, one may assume that $I(U,W;Y) > I(U,W;Z)$ and $I(V,W;Z) > I(V,W;Y)$. If this holds, then we claim that we can choose $I(W;Y) = I(W;Z)$.

Suppose that at a maximizing distribution we have $I(W;Y) > I(W;Z)$. Let Q be Bernoulli(a) and independent of (U,V,W,X). When $Q=0$, set $(\tilde{U},\tilde{V},\tilde{W}) = (U,V,W)$ as before; and when $Q=1$ set $(\tilde{U},\tilde{V},\tilde{W}) = (U,\emptyset,(V,W))$. Let $W' = (\tilde{W},Q)$. Now $a \in (0,1)$ is chosen such that

$$(1-a)I(W;Z) + aI(V,W;Z) = (1-a)I(W;Y) + aI(V,W;Y)$$

For the new triple $(W',\tilde{U},\tilde{V})$, we have $I(W';Y) = I(W';Z)$, and the sum-rate does not decrease since

$$\begin{aligned}
SR_{new} &= (1-a)\left(I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W)\right) \\
&\quad + a\left(I(V,W;Z) + I(U;Y|V,W)\right) \\
&= (1-a)SR_{old} + a\left(I(W;Z) + I(U;Y|W) + I(V;Z|W) - I(U;V|W) + I(U;V|W,Y)\right) \\
&= SR_{old} + aI(U;V|WY) \geq SR_{old}.
\end{aligned}$$

As a corollary of the above manipulation, we also have that if there is a distribution that maximizes the sum-rate, which is strictly larger than $I(X;Y), I(X;Z)$, and has $I(W;Y) > I(W;Z)$, then it must be that $I(U;V|WY) = 0$. Notice that here $|\mathcal{W}'| = 2|\mathcal{W}|$ is still bounded and satisfies $I(W';Y) = I(W';Z)$. $\square$

Starting from this $p^*(w,u,v,x)$, we can find a W of cardinality at most $|\mathcal{X}|+1$, and a $q^*(w,u,v,x)$ (standard application of Fenchel-Butt extension to Caratheodory) such that $p(X), -H(Y|W)+I(U;Y|W)+I(V;Z|W)-I(U;V|W), -H(Z|W)+I(U;Y|W)+I(V;Z|W)-I(U;V|W)$ is preserved. Note that this implies that in particular $I_{q^*}(W;Y) = I_{q^*}(W;Z)$, and this $q^*(w,u,v,x)$ also attains a global maximum of

$$\max_{p(u,v,w,x)} \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) + I(V;Z|W) - I(U;V|W)$$

subject to a fixed $p(x)$.

Now if the maximum is $I(X;Y)$ or $I(X;Z)$, then it can be attained by $(U,V,W) = (X,\emptyset,\emptyset)$ or $(U,V,W) = (\emptyset,X,\emptyset)$, hence $|\mathcal{W}| = 1$; otherwise we have $I(W;Y) = I(W;Z)$ and the cardinality of W is bounded by $|\mathcal{X}|+1$. Here below we prove Theorem 2 by showing that for the maximum pmf such that $I(W;Y) = I(W;Z)$, we can reduce the cardinality of W by 1 so long as $|\mathcal{W}| > |\mathcal{X}|$.

Proof. Let $r(w,u,v,x) = q^*(w,u,v,x)(1+\epsilon L(w))$, satisfying $\sum_{w,u,v} q^*(w,u,v,x)L(w) = 0, \forall x \in \mathcal{X}$. A non-trivial $L(w)$ exists whenever $|\mathcal{W}| > |\mathcal{X}|$.

Observe that under this perturbation the term

$$\begin{aligned}
H_r(A,W) &= -\sum_{a,w} r(a,w)\log_2 r(a,w) \\
&= -\sum_{a,w} q^*(a,w)(1+\epsilon L(w))\log_2 q^*(a,w)(1+\epsilon L(w)) \\
&= H_{q^*}(A,W) + \epsilon H_L(A,W) - \sum_{a,w} q^*(a,w)(1+\epsilon L(w))\log_2(1+\epsilon L(w)) \\
&= H_{q^*}(A,W) + \epsilon H_L(A,W) - \sum_{w} q^*(w)(1+\epsilon L(w))\log_2(1+\epsilon L(w))
\end{aligned}$$

Hence the higher order epsilon terms only depends on W, thus will cancel out in mutual information terms below. Observe that

$$\begin{aligned}
I_r(W;Y) &= I_{q^*}(W;Y) + \epsilon(H_L(W) - H_L(W,Y)) \\
I_r(W;Z) &= I_{q^*}(W;Z) + \epsilon(H_L(W) - H_L(W,Z)) \\
I_r(U;Y|W) &= I_{q^*}(U;Y|W) + \epsilon(H_L(U,W) + H_L(Y,W) - H_L(UYW) - H_L(W)) \\
I_r(V;Z|W) &= I_{q^*}(V;Z|W) + \epsilon(H_L(V,W) + H_L(Z,W) - H_L(VZW) - H_L(W)) \\
I_r(U;V|W) &= I_{q^*}(U;V|W) + \epsilon(H_L(U,W) + H_L(V,W) - H_L(UVW) - H_L(W))
\end{aligned}$$

In the first two equalities, we also used that $H_r(Y) = H_{q^*}(Y)$ since distribution of X is preserved. Therefore the sum-rate corresponding to the distribution $r(w,u,v,x)$ is given by

$$\begin{aligned}
&\min\{I_{q^*}(W;Y) + \epsilon(H_L(W) - H_L(W,Y)), I_{q^*}(W;Z) + \epsilon(H_L(W) - H_L(W,Z))\} \\
&\quad + I_{q^*}(U;Y|W) + \epsilon(H_L(U,W) + H_L(Y,W) - H_L(UYW) - H_L(W)) \\
&\quad + I_{q^*}(V;Z|W) + \epsilon(H_L(V,W) + H_L(Z,W) - H_L(VZW) - H_L(W)) \\
&\quad - I_{q^*}(U;V|W) - \epsilon(H_L(U,W) + H_L(V,W) - H_L(UVW) - H_L(W)) \\
&= I_{q^*}(W;Y)(= I_{q^*}(W;Z)) + I_{q^*}(U;Y|W) + I_{q^*}(V;Z|W) - I_{q^*}(U;V|W) \\
&\quad + \epsilon\Big(\min\{H_L(W) - H_L(W,Y), H_L(W) - H_L(W,Z)\} + H_L(Y,W) \\
&\quad - H_L(UYW) - H_L(W) + H_L(Z,W) - H_L(VZW) + H_L(UVW)\Big)
\end{aligned}$$

Since $q^*(w,u,v,x)$ is a global maximum of the sum-rate it implies that the factor multiplying ϵ must be zero or that the sum-rate corresponding to the distribution $r(w,u,v,x)$ matches that of $q^*(w,u,v,x)$.

Choose ϵ large enough such that $\min_w r^*(w) = 0$, hence we can reduce the cardinality of $\mathcal{W}$ to $|\mathcal{X}|$. $\square$

2.2 Proof of Corollary 1

Proof. Let $\bar{R}$ be the maximum sum-rate obtained by the randomized time-division strategy and R be that by Marton's inner bound. We need to show $R = \bar{R}$. Clearly, we have $R \geq \bar{R}$ 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. W.l.o.g. we consider two cases:

Case 1:

$$I(X;Y|W=0) \geq I(X;Z|W=0) \text{ and } I(X;Y|W=1) \geq I(X;Z|W=1). \tag{3}$$

Clearly

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

where (a) follows from Theorem 1 and (3).

Case 2:

$$I(X;Y|W=0) \geq I(X;Z|W=0) \text{ and } I(X;Y|W=1) \leq I(X;Z|W=1). \tag{4}$$

Observe that

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

where (b) follows from Theorem 1 and (4).

The other two cases follow similarly. This implies $R \leq \bar{R}$ thus completes the proof of Corollary 1. □

3 Proof of Theorem 1

The idea of the proof is to fix a $p(y, z|x)$ (i.e. a particular broadcast channel) and show that for all $p_o(x)$ we have that

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

Denote LHS and RHS as the left-hand side and right-hand side of the inequality (2), respectively. We use the notation: $U \wedge V$ (and), $U \vee V$ (or), $U \oplus V$ (xor), $\bar{U}$ (not). Let

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

Similarly p_u refers to the margin $\mathrm{P}(U=u)$. For binary X, for brevity let

$$\begin{aligned}
a_i &= \mathrm{P}(Y=i|X=0), \quad \hat{a}_i = \mathrm{P}(Y=i|X=1), \quad i = 1, \ldots, |\mathcal{Y}| \\
b_i &= \mathrm{P}(Z=i|X=0), \quad \hat{b}_i = \mathrm{P}(Z=i|X=1), \quad i = 1, \ldots, |\mathcal{Z}|.
\end{aligned}$$

Normally we use $p_{y|x}$ for general purpose and a_i for a binary input broadcast channel.

Remark 1. As LHS and RHS are continuous in $p_{y|x}$ and $p_{z|x}$, it suffices to prove the inequality when $p_{y|x}$ and $p_{z|x}$ are positive.

Remark 2. From [4] (or see Fact 1 and Claim 1 in [8] for a self-contained shorter proof) it suffices to establish Theorem 1 for the scenario $|\mathcal{U}| \le |\mathcal{X}|, |\mathcal{V}| \le |\mathcal{X}|$ and $X = f(U,V)$, a deterministic function of (U,V).

Claim 1. *For 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 $v_2 \ne v_1$ 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)

$$\begin{aligned} LHS(q) - LHS(p) &= H_q(U,V) - H_q(U,Y) - H_q(V,Y) - H_p(U,V) + H_p(U,Y) + H_p(V,Y) \\ &= \epsilon\{-\ln\epsilon + \ln p_{u_1v_2} + \sum_z p_{z|x_1} \ln\frac{p_{v_1z}}{p_{v_2z}}\} + o(\epsilon) \end{aligned}$$

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

The outline of the proof is as follows:

1. We first prove the inequality for some special settings, or "trivial" cases. (Section 3.1)
2. We show that it suffices to prove for the nontrivial cases $X = U \wedge V$ and $X = U \oplus V$. (Section 3.2)
3. For $X = U \oplus V$, nontrivial maximum of LHS is not achievable when $p(U,V) > 0$. (Section 3.3)
4. For $X = U \wedge V$, nontrivial maximum of LHS is not achievable when $p(U,V) > 0$. (Section 3.4)

3.1 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

$$\begin{aligned} I(U;Y) \le I(X;Y), \quad I(U;Y) \le I(U;X), \\ I(V;Z) \le I(X;Z), \quad I(V;Z) \le I(V;X). \end{aligned} \tag{5}$$

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

SS1. $a_i \equiv \hat{a}_i$. Then $X \perp Y$, thus $I(U;Y) = I(X;Y) = 0$. From (5) and the non-negativity of $I(U;V)$ we have $I(V;Z) - I(V;U) \le I(X;Z)$, i.e. Theorem 1 holds. Similarly Theorem 1 holds when $b_i \equiv \hat{b}_i$.

SS2. $U \perp X$. Then $I(U;Y) = I(U;X) = 0$. Again from (5) and the non-negativity of $I(U;V)$ Theorem 1 holds. Similarly when $V \perp X$, Theorem 1 also holds.

3.2 Two Nontrivial Cases

According to Remark 2, to prove the inequality (2), 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 (equivalence is 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 the groups except G_4.

The first two cases are trivial. For $X = 0$, the theorem is reduced to $-I(U; V) \le 0$. For $X = U$, i.e. $I(U; Y) = I(X; Y)$, the theorem follows from the data processing inequality, $I(V; Z) \le I(V; U) = I(V; X)$ (see Eqn.(5)). 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.

3.3 Proof of XOR Case

Just as in [8] 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 $X = U \oplus V$.

Let the level sets for $X = f(U, V)$ be $E_x = \{(u, v) : f(u, v) = x\}$ and the complement sets be $N_x = E_x^c = \{(u, v) : f(u, v) \ne x\}$. For $x \in \mathcal{X}$, choose one pair (u_x, v_x) in E_x (E_x is nonempty as $p(x) > 0$).

Consider an additive perturbation $q(u, v, x) = p(u, v, x) + \varepsilon\lambda(u, v, x)$ for some $\varepsilon \ge 0$. 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 $\lambda_{uvx} \ge 0$ if the corresponding $p(u, v, x)$ is zero, which is

$$\lambda_{uvx} \ge 0, \quad \forall uvx : uv \in N_x$$

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

$$\sum_{uv} \lambda_{uvx} = 0, \quad \forall x \in \mathcal{X} \tag{6}$$

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

$$-\sum_{xuv} \lambda_{uvx} \log p_{uv} + \sum_{xuvy} \lambda_{uvx} p_{y|x} \log p_{uy} + \sum_{xuvz} \lambda_{uvx} p_{z|x} \log p_{vz} \le 0 \tag{7}$$

Inequality (7) is equivalent to $\sum_{xuv} \lambda_{uvx} C_{uvx} \le 0$, where

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

From (6), we express $\lambda_{u_x v_x x}$ in the term of other λ_{uvx} variables, that is

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

Substituting the above equations into (7), we have

$$\sum_{xuv:uv\neq u_xv_x} \lambda_{uvx}(C_{uvx} - C_{u_xv_xx}) \leq 0 \tag{8}$$

Since (8) holds for any signed $\{\lambda_{uvx} : uv \in E_x\backslash\{(u_x, v_x)\}\}$ and any nonnegative $\{\lambda_{uvx} : uv \in N_x\}$, it implies

$$uv \in E_x, \quad C_{uvx} = C_{u_xv_xx} \tag{9}$$

$$uv \in N_x, \quad C_{uvx} \leq C_{u_xv_xx} \tag{10}$$

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 $C_{u_1v_1x} + C_{u_2v_2x} \geq C_{u_1v_2x} + C_{u_2v_1x}$. □

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

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

and 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

$$\log \frac{p_{00}}{p_{01}} = \sum b_i \log \frac{b_ip_{00} + \hat{b}_ip_{10}}{b_ip_{11} + \hat{b}_ip_{01}} = \sum \hat{b}_i \log \frac{b_ip_{00} + \hat{b}_ip_{10}}{b_ip_{11} + \hat{b}_ip_{01}}.$$

Taking a weighted sum, we get

$$(p_{00} + p_{10}) \log \frac{p_{00}}{p_{01}} = \sum (b_ip_{00} + \hat{b}_ip_{10}) \log \frac{b_ip_{00} + \hat{b}_ip_{10}}{b_ip_{11} + \hat{b}_ip_{01}} \tag{12}$$

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

$$\frac{b_ip_{00} + \hat{b}_ip_{10}}{b_ip_{11} + \hat{b}_ip_{01}} \equiv \frac{p_{00}}{p_{01}}.$$

From the above we obtain

$$(p_{01} - p_{11})(b_i - \hat{b}_i) \equiv 0. \tag{13}$$

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

$$(p_{10} - p_{11})(a_i - \hat{a}_i) \equiv 0. \tag{14}$$

Now we have two cases

1. $b_i \equiv \hat{b}_i$, or $a_i \equiv \hat{a}_i$. In this case the Theorem holds (special setting SS1).
2. $p_{01} = p_{11},\ p_{10} = p_{11}$. Combining this with $p_{00}p_{11} = p_{01}p_{10}$ (Eqn.(11)) 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.

3.4 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

$$\begin{aligned} J(p_{10}, p_{01}) &= H(U,V) - H(U,Y) - H(V,Z) \\ &= -p_{00}\log p_{00} - p_{01}\log p_{01} - p_{10}\log p_{10} - p_{11}\log p_{11} \\ &\quad + \sum a_i(p_{00}+p_{01})\log[a_i(p_{00}+p_{01})] + \sum (a_i p_{10} + \hat{a}_i p_{11})\log[a_i p_{10} + \hat{a}_i p_{11}] \\ &\quad + \sum b_i(p_{00}+p_{10})\log[b_i(p_{00}+p_{10})] + \sum (b_i p_{01} + \hat{b}_i p_{11})\log[b_i p_{01} + \hat{b}_i p_{11}]. \end{aligned}$$

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

$$\nabla J = \vec{0}, \ \nabla^2 J \preceq \mathbf{0}, \tag{15}$$

where $\nabla^2 J \preceq \mathbf{0}$ 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 w.r.t. the free variables we obtain:

$$\begin{aligned} \frac{\partial J}{\partial p_{10}} &= \log\frac{p_{00}}{p_{10}} - \sum a_i \log\frac{a_i(p_{00}+p_{01})}{a_i p_{10} + \hat{a}_i p_{11}} \\ \frac{\partial J}{\partial p_{01}} &= \log\frac{p_{00}}{p_{01}} - \sum b_i \log\frac{b_i(p_{00}+p_{10})}{b_i p_{01} + \hat{b}_i p_{11}}. \end{aligned}$$

The condition $\nabla J = \vec{0}$ implies that

$$\log\frac{p_{00}}{p_{10}} = \sum a_i \log\frac{a_i(p_{00}+p_{01})}{a_i p_{10} + \hat{a}_i p_{11}} \tag{16}$$

$$\log\frac{p_{00}}{p_{01}} = \sum b_i \log\frac{b_i(p_{00}+p_{10})}{b_i p_{01} + \hat{b}_i p_{11}}. \tag{17}$$

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

$$\frac{p_{00}}{p_{10}} \leq \sum \frac{a_i^2(p_{00}+p_{01})}{a_i p_{10} + \hat{a}_i p_{11}}, \qquad \frac{p_{00}}{p_{01}} \leq \sum \frac{b_i^2(p_{00}+p_{10})}{b_i p_{01} + \hat{b}_i p_{11}}, \tag{18}$$

where the equalities hold iff. (using Remark 1)

$$a_i \equiv c_a \hat{a}_i, \ b_i \equiv c_b \hat{b}_i,$$

for some constants c_a, c_b respectively. However since $\sum_i a_i = \sum_i \hat{a}_i = 1$ we obtain that $c_a = 1$ (similarly $c_b = 1$). Thus equalities hold iff.

$$a_i \equiv \hat{a}_i, \ b_i \equiv \hat{b}_i. \tag{19}$$

2. Second Derivative:
We now compute the Hessian $G \equiv \nabla^2 J$, The second derivatives are

$$\begin{aligned} G_{11} &= \frac{\partial^2 J}{\partial p_{10}^2} = -\frac{1}{p_{00}} - \frac{1}{p_{10}} + \frac{1}{p_{00}+p_{01}} + \sum \frac{a_i^2}{a_i p_{10} + \hat{a}_i p_{11}} \\ G_{12} &= G_{21} = -\frac{1}{p_{00}} \\ G_{22} &= \frac{\partial^2 J}{\partial p_{01}^2} = -\frac{1}{p_{00}} - \frac{1}{p_{01}} + \frac{1}{p_{00}+p_{10}} + \sum \frac{b_i^2}{b_i p_{01} + \hat{b}_i p_{11}}. \end{aligned}$$

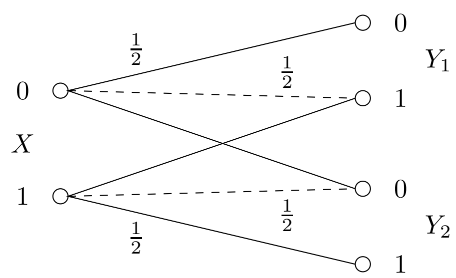

Figure 3: binary skew-symmetric broadcast channel

As $p_{01} > 0$, we have $G_{11} \leq -\frac{1}{p_{00}} - \frac{1}{p_{10}} + \frac{1}{p_{00}+p_{01}} + \frac{1}{p_{10}} < 0$. Similarly we have $G_{22} < 0$. For G with $G_{11} < 0$ and $G_{22} < 0$ to be negative semi-definite, it is necessary and sufficient that $\det(G) \geq 0$.

From (18) we have

$$G_{11} \geq -\frac{1}{p_{00}} - \frac{1}{p_{10}} + \frac{1}{p_{00}+p_{01}} + \frac{p_{00}}{p_{10}(p_{00}+p_{01})} = -\frac{p_{01}(p_{00}+p_{10})}{p_{00}p_{10}(p_{00}+p_{01})}.$$

And similarly

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

It is clear that equalities in the above two inequalities hold iff. (19) holds.

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

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

with equality holding only if (19) holds.

Thus $\det(G) < 0$ or there is no local maximum when $p(U,V) > 0$ unless the channel parameters satisfy (19). However when (19) holds, the inequality is true from the special setting SS1.

This completes the proof of Theorem 1.

4 Sum-rate evaluations of inner and outer bounds for BSSC

We shall evaluate the inner and outer bounds for the BSSC (Figure 3) from [5] and [7]. Apart from completeness, this section serves some purposes:

- We correct a minor typo in the evaluation of the maximum sum-rate of the outer bound[5].
- We also explicitly compute the maximum sum-rate obtained via the Korner-Marton outer bound for the BSSC.

4.1 Sum-rate evaluation of Marton's inner bound

Clearly from Theorem 2 we can assume that $|\mathcal{W}| = 2$. Observe that for the BSSC, $I(X;Y_1) \geq I(X;Y_2)$ iff $\mathrm{P}(X=0) \leq \frac{1}{2}$. If $0 \leq \mathrm{P}(X=0|W=0), \mathrm{P}(X=0|W=1) \leq \frac{1}{2}$ it is easy to see, using the established inequality that

$$\begin{aligned} SR &= \min(I(W;Y_1), I(W;Y_2)) + I(X;Y_1|W) \\ &\leq I(X;Y_1) \\ &\leq C \end{aligned}$$

where C is the single channel capacity given by $h(0.2) - 0.4 \approx 0.321928..$ Similarly if $\frac{1}{2} \leq \mathrm{P}(X = 0|W = 0), \mathrm{P}(X = 0|W = 1) \leq 1$ again the sum-rate will be bounded by C. Hence we can assume that $0 \leq \mathrm{P}(X = 0|W = 0) \leq \frac{1}{2} \leq \mathrm{P}(X = 0|W = 1) \leq 1$.

Let $d = \max_{p(x)} I(X;Y_1) - I(X;Y_2)$. Then we can solve for $d = 0.10072952..$ and the optimizing choice for $\mathrm{P}(X = 0) = 0.5 - \sqrt{105}/30 = 0.15843497...$ Now observe that

$$\begin{aligned} SR &\leq I(W;Y_1) + \mathrm{P}(W = 0)I(X;Y_1|W = 0) \\ &\quad + \mathrm{P}(W = 1)I(X;Y_2|W = 1) \\ &= I(X;Y_1) + \mathrm{P}(W = 1)(I(X;Y_2|W = 1) \\ &\quad - I(X;Y_1|W = 1)) \\ &\leq I(X;Y_1) + \mathrm{P}(W = 1)d. \end{aligned}$$

Similarly

$$\begin{aligned} SR &\leq I(W;Y_2) + \mathrm{P}(W = 0)I(X;Y_1|W = 0) \\ &\quad + \mathrm{P}(W = 1)I(X;Y_2|W = 1) \\ &= I(X;Y_2) + \mathrm{P}(W = 0)(I(X;Y_1|W = 0) \\ &\quad - I(X;Y_2|W = 0)) \\ &\leq I(X;Y_2) + \mathrm{P}(W = 0)d. \end{aligned}$$

From these two (by adding them) we can deduce that

$$2SR \leq I(X;Y_1) + I(X;Y_2) + d.$$

The maximum of $I(X;Y_1) + I(X;Y_2) = 0.6225562..$ occurs when $\mathrm{P}(X = 0) = \frac{1}{2}$ and hence substituting we obtain that $SR \leq 0.36164288...$

To show that it is indeed on the boundary of the achievable region consider the joint distribution on X and W as follows:

$$\begin{aligned} \mathrm{P}(W = 0) = \mathrm{P}(W = 1) &= \frac{1}{2} \\ \mathrm{P}(X = 0|W = 0) = \mathrm{P}(X = 1|W = 1) &= 0.5 - \sqrt{105}/30 = 0.15843497.. \end{aligned}$$

For this distribution, all inequalities reduce to equalities and SR of 0.3616.. is achieved.

4.2 Sum-rate evaluations of the outer bounds for BSSC

4.2.1 Case 1: Bound 3

To evaluate maximum of the sum-rate of the outer bound (Bound 3) it was shown [5] that it suffices to consider $\mathrm{P}(X = 0) = \frac{1}{2}$. (It is immediate using the skew-symmetry of the channel and the inherent symmetry of the outer bound expressions.)

The sum-rate maximum is hence given by

$$\max_{p(u,x),\mathrm{P}(x=0)=\frac{1}{2}} I(U;Y_1) + I(X;Y_2|U)$$

or in other words maximize

$$\max_{p(u,x),\mathrm{P}(x=0)=\frac{1}{2}} I(X;Y_1) + I(X;Y_2|U) - I(X;Y_1|U)$$

Let $\mathrm{P}(X = 0) = x$. In [7] it was shown that the curve $f(x) = I(X;Y_1) - I(X;Y_2) = h(\frac{x}{2}) - h(\frac{1-x}{2}) + 1 - 2x$ is concave when $x \in [0, \frac{1}{2}]$ and convex when $x \in [\frac{1}{2}, 1]$. Further it was also shown that the lower convex envelope[2] was given by

$$g(x) = \begin{cases} \frac{5x}{4} f(\frac{4}{5}) & 0 \leq x \leq \frac{4}{5} \\ f(x) & \frac{4}{5} \leq x \leq 1 \end{cases}.$$

[2]more precisely, in [7] the upper concave envelope was characterized, and the characterization of the lower convex envelope follows by symmetry.

From the definition of the lower convex envelope, we know that

$$I(X;Y_1|U) - I(X;Y_2|U) \geq g(\frac{1}{2})$$

and it easy to see that the equality is indeed achieved for a binary U.

Therefore

$$\max_{p(u,x),\mathrm{P}(X=0)=\frac{1}{2}} I(X;Y_1) + I(X;Y_2|U) - I(X;Y_1|U)$$
$$= h(\frac{1}{4}) - 0.5 + g(0.5) \approx 0.3725562...$$

This is a correction to the implicit error we made in [5] while calculating the lower convex envelope and obtained a bound of 0.37111....

4.2.2 Case 2: Korner-Marton Bound

Bound 3. *[3] (Korner-Marton outer bound) The union of rate-pairs (R_1, R_2) satisfying the following constraints:*

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

over all random variablessuch that $U \to X \to (Y,Z)$ forms a Markov chain forms an outer-bound to the capacity region of the broadcast channel.

Similarly one can interchange the roles of the receivers Y and Z and will lead to yet another outer bound. The intersection of these two regions is normally termed as the Korner-Marton outer bound.

To show that this sum-rate is still strictly inside the Korner-Marton [3] outer bound observe that we need to evaluate the union over $p(u,x)$

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

Further if a point $(R_1, R_2) = (a, a)$ belongs to this region, by the skew-symmetry of BSSC, it will also belong to the union over $p(v,x)$

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

and hence to the intersection of the two regions. The key difference between the bounds is that while the former takes the intersection before the union, the latter takes the union prior to the intersection.

Suppose we wish to compute

$$\max_{p(u,x)} I(X;Y_1) + I(X;Y_2|U) - I(X;Y_1|U)$$

then from the earlier discussion, this will be the maximum over $x \in [0,1]$ of

$$h(\frac{x}{2}) - x - g(x)$$

It is easy to see that the global maximum will lie when $x \in [0, \frac{4}{5}]$ (otherwise maximum occurs when U is trivial and equals $I(X;Y_2)$). Taking derivatives we obtain that maximum occurs when either

$$\frac{1}{2}\log_2 \frac{2-x}{x} - 1 - \frac{5}{4} f(\frac{4}{5}) = 0$$

or

$$x^* = \frac{2}{1+2^c} \approx 0.4571429.$$

where $c = 2(1 + \frac{5}{4} f(\frac{4}{5})) \approx 1.7548875$.

Thus the maximum sum rate given by

$$\max_{p(u,x)} I(X;Y_1) + I(X;Y_2|U) - I(X;Y_1|U) \approx 0.3743955.$$

The pair (U, X) that achieves the maximum can be characterized by

$$\begin{aligned} \mathrm{P}(U=0) = 1-a, \mathrm{P}(X=0|U=0) = 0, \\ \mathrm{P}(U=1) = a, \mathrm{P}(X=0|U=0) = \frac{4}{5} \end{aligned}$$

where $0.8 * a = x^*$ or $a \approx 0.5714286$.

Observe that for this choice

$$\begin{aligned} I(U;Y_1) &= h(\frac{x^*}{2}) - ah(0.4) \approx 0.2206837... \\ I(X;Y_2|U) &\approx 0.1537118.. \\ I(X;Y_2) &\approx 0.3006499 \end{aligned}$$

Therefore the symmetric rate-pair $(R_1, R_2) = (0.1871978.., 0.1871978..)$ lies on the boundary of the Korner-Marton outer bound. In summary, the maximum sum rate given by Korner-Marton outer bound for the BSSC is 0.3743955....

Historical remarks

Perturbation method as a tool in computing bounds on cardinalities of auxiliary random variables were used by Amin Gohari and Venkat Anantharam [4]. The perturbations used in their work were support-preserving (or multiplicative) in nature. In [8], the authors used the perturbation technique (and here the additive perturbation was also required) to compute the local maximas of $I(U;Y) + I(V;Z) - I(U;V)$ for the BSSC channel. Using this technique they established the inequality in Theorem 1 for the BSSC channel. Working on a related problem, one of the authors realized that the inequality may be more generally true for all binary input broadcast channels. A (non-trivial) modification of the arguments in [8] yielded a proof for this fact, which was then presented in [9]. To present a complete picture to the community, it was decided to combine the related proofs in [8], [9] into a single paper.

5 Conclusion

An information theoretic inequality is established for binary input broadcast channels. This is used to show that the sum-rate given by Marton's inner bound is indeed equivalent to that given by randomized time-division strategy. The inequality fails when $|X| \geq 3$ so a natural question is whether there is a correct generalization for higher cardinality input-alphabets. It would also be useful to find a more intuitive (geometric) argument to shed more light into the actual counting of the sizes of typical sets. Here is an equivalent formulation which is related to the sizes of certain typical sets. It can be shown that the information inequality is equivalent to showing that

$$H(U|Y) + H(V|Z) \geq \min\{H(UV|Y), H(UV|Z)\}$$

whenever $(U,V) \to X \to (Y,Z)$ forms a Markov chain, $X = f(U,V)$ and $|\mathcal{X}| = 2$. It would also be interesting to see which of the inner and outer bounds (possibly both) are strictly weak and find examples where one could demonstrate this.

Acknowedgements

The authors are grateful to a number of comments on this work from Abbas El Gamal and Amin Gohari.

References

[1] T. Cover, "Broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 2–14, January, 1972.

[2] ——, "Comments on broadcast channels," *IEEE Trans. Info. Theory*, vol. IT-44, pp. 2524–2530, October, 1998.

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

[4] A. A. Gohari and V. Anantharam, "Evaluation of marton's inner bound for the general broadcast channel," *CoRR*, vol. abs/0904.4541, 2009.

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

[6] C. Nair, "A note on outer bounds for broadcast channel," *CoRR*, vol. abs/1101.0640, 2011.

[7] C. Nair and V. W. Zizhou, "On the inner and outer bounds for 2-receiver discrete memoryless broadcast channels," *Proceedings of the ITA Workshop*, 2008.

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

[9] C. Nair, Z. V. Wang, and Y. Geng, "An information inequality and evaluation of marton's inner bound for binary input broadcast channels," *International Symposium on Information Theory*, 2010.