Improved cardinality bounds on the auxiliary random variables in Marton's inner bound

Venkat Anantharam
Department of EECS
University of California
Berkeley, CA, USA
Email: ananth@eecs.berkeley.edu

Amin Gohari
Information Systems and Security Lab
Sharif University of Technology
Tehran, Iran
aminzadeh@sharif.edu

Chandra Nair
Department of Information Engineering
Chinese University of Hong Kong
Sha Tin, N.T., Hong Kong
Email: chandra@ie.cuhk.edu.hk

***Abstract*—Marton's region is the best known inner bound for a general discrete memoryless broadcast channel. We establish improved bounds on the cardinalities of the auxiliary random variables. We combine the perturbation technique along with a representation using concave envelopes to achieve this improvement. As a corollary of this result, we show that a randomized time division strategy achieves the entire Marton's region for binary input broadcast channels, extending the previously known result for the sum-rate and validating a previous conjecture due to the same authors.**

I. Introduction

In this paper we consider a discrete-memoryless two user broadcast channel consisting of a sender X and two receivers Y, Z where $|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}| < \infty$. The sender maps two private messages M_1, M_2 to a transmit sequence $X^n(m_1, m_2)$. The receivers each get a noisy version of the transmitted codeword over their individual channel. We assume that the channel is memoryless and there is no feedback implying $p(y^n, z^n|x^n) = \prod_{i=1}^n p(y_i, z_i|x_i)$. We borrow most of our notation from Chapters 5 and 8 in [1] where the classical results on broadcast channels are reviewed.

The best known achievable rate region for a broadcast channel is the following inner bound [2].

Bound 1. *(Marton) The union of non-negative rate pairs R_1, R_2 satisfying the constraints*

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

for any triple of random variables (U, V, W) such that $(U, V, W) \to X \to (Y, Z)$ is achievable.

Denote this achievable region due to Marton as $\mathcal{R}_M$. It was shown in [3], using a novel perturbation based approach, that it suffices to consider $|\mathcal{W}| \leq |\mathcal{X}| + 4, |\mathcal{U}| \leq |\mathcal{X}|, |\mathcal{V}| \leq |\mathcal{X}|$ to obtain $\mathcal{R}_M$.

In this paper we improve the cardinality bounds on the auxiliary random variables used in Marton's inner bound. We had made two conjectures related to Marton's inner bound in [4]. One of them is resolved in this paper. The other, if true, would have established the optimality of Marton's inner bound for broadcast channels. The second conjecture involves comparing a quantity related to the two-letter version of a given broadcast channel with the same quantity obtained using the single-letter description. The improved cardinality bounds allow one to numerically verify the second conjecture for binary input broadcast channels. Given the page limitations, we focus on giving a proof of our results while leaving out the broader implications.

The perturbation technique, introduced in [3], was also used to show that the extremal points of Marton's inner bound were structured [5], [6], [4]. As an example, the following much simpler characterization of the sum-rate than the one given by the third inequality in Bound 2, for binary input broadcast channels, was obtained in [5].

Lemma 1. *[5] The maximum sum-rate achievable by Marton's inner bound for any binary input broadcast channel is given by*

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

Here $\mathcal{W} = \{0, 1\}$.

Lemma 1 is equivalent to stating that setting $U = X, V = 0$ or $V = X, U = 0$ in Marton's inner bound suffices to compute the maximum sum-rate for any binary input broadcast channel. Our new cardinality bounds immediately extend this result to the entire $\mathcal{R}_M$ for any binary input broadcast channel.

A. The main result and its proof outline

The main result of the paper is the following

Theorem 1. *To compute Marton's inner bound, $\mathcal{R}_M$, it suffices to consider union over $p(u, v, w, x)$ such that $|\mathcal{U}| + |\mathcal{V}| \leq |\mathcal{X}| + 1$ and $|\mathcal{W}| \leq |\mathcal{X}| + 4$.*

Remark 1. The cardinality bound $|\mathcal{U}| + |\mathcal{V}| \leq |\mathcal{X}| + 1$ represents a trade-off between the sizes of alphabets of U and V which has a striking correspondence to the broadcast channel setting of having to trade-off an increase in communication rate to one receiver at the expense of the other receiver.

The reasoning (and intuition) for the proof of this theorem is built on the previous works by the authors, in particular [7], [5], [6], [8], [9], [4]. Since Marton's inner bound, $\mathcal{R}_M$, is a convex compact subset of $\mathbb{R}^3_+$, it is the intersection of the various supporting hyperplanes when we think of it as a subset of $\mathbb{R}^3$. The cardinality bound of $|\mathcal{U}|+|\mathcal{V}| \leq |\mathcal{X}|+1$ is first developed for the *exposed points* of $\mathcal{R}_M$. By Straszewicz's theorem [10] we know that the exposed points are dense in the set of *extreme points*. By continuity of the mutual information terms w.r.t. probability distributions, we obtain the same cardinality bound on $|\mathcal{U}|+|\mathcal{V}|$ for the extreme points. Finally, we use convexification using W to show that the bound goes through for all points on the boundary of $\mathcal{R}_M$.

B. Preliminaries

An *exposed* point of a closed convex set $\mathcal{C}$ is a point $x \in \mathcal{C}$ such that there is a supporting hyperplane that intersects the convex set only at x. An *extreme* point of a closed convex set $\mathcal{C}$ is a point $x \in \mathcal{C}$ in the set such that if $x=\lambda y+(1-\lambda)z$ for some $y, z \in \mathcal{C}$ and $\lambda \in [0,1]$, then either $x=y$ or $x=z$. For a closed compact set $\mathcal{C}$ in a finite dimensional Euclidean space, Straszewicz's theorem states that exposed points are a dense subset of extreme points.

Let us partition the exposed points of $\mathcal{R}_M$ into $(0,0)$, $(C_1,0)$, $(0,C_2)$, and finally a set $\mathcal{E}$ consisitng of exposed points (R_1,R_2) with $R_1,R_2>0$. The first three exposed points can be obtained by setting $U=0, V=0, U=X, V=0$, and $U=0, V=X$ respectively in Bound 2.

For every exposed point in $\mathbf{x} \in \mathcal{E}$ let us *associate a supporting hyperplane* of the form $\gamma_1^{\mathbf{x}} R_1 + \gamma_2^{\mathbf{x}} R_2 = c^{\mathbf{x}}$ such that the only point of intersection of the supporting hyperplane and $\mathcal{R}_M$ is $\mathbf{x}$. It is clear that if $\mathbf{x} \in \mathcal{E}$ then we can assume $\gamma_1^{\mathbf{x}}, \gamma_2^{\mathbf{x}} > 0$. By symmetry assume $\gamma_1^{\mathbf{x}} \geq \gamma_2^{\mathbf{x}}$. Let $\alpha_1^{\mathbf{x}} = \gamma_1^{\mathbf{x}} - \gamma_2^{\mathbf{x}} \geq 0$ and $\alpha_2^{\mathbf{x}} = \gamma_2^{\mathbf{x}} > 0$, then we can express the hyperplane as $\alpha_1^{\mathbf{x}} R_1 + \alpha_2^{\mathbf{x}}(R_1+R_2) = c$.

We now present the following lemma about $\mathcal{R}_M$.

Lemma 2. *Given any $\mathbf{x} \in \mathcal{E}$ with $\alpha_1^{\mathbf{x}} R_1 + \alpha_2^{\mathbf{x}}(R_1+R_2)$ its associated supporting hyperplane, the following holds:*

$$\begin{aligned}
&\max_{(R_1,R_2)\in\mathcal{R}_M} \alpha_1^{\mathbf{x}} R_1 + \alpha_2^{\mathbf{x}}(R_1+R_2)\\
&= \max_{\substack{(U,V,W):\\(U,V,W)\to X\to(Y,Z)}} \alpha_1^{\mathbf{x}} I(UW;Y)\\
&\quad + \alpha_2^{\mathbf{x}}\big(\min\{I(W;Y), I(W;Z)\} + I(U;Y|W)\\
&\quad + I(V;Z|W) - I(U;V|W)\big).
\end{aligned}$$

Proof: From the first and third inequalities in Bound 2 it is clear that the left hand side is smaller than or equal to the right hand side. Let $p_*(u,v,w,x)$ be a maximizer of the right hand side. Clearly, one must have $I_{p^*}(V;Z|W) - I_{p^*}(U;V|W) \geq 0$. Otherwise set $V=\emptyset$ to get a contradiction.

Suppose $I_{p^*}(UW;Y) \leq I_{p^*}(UW;Z)$, then by setting $W'=(U,W), U'=\emptyset, V'=X$ we see that the right hand side does not decrease. Further, $R_1 = I(UW;Y), R_2 = I(X;Z|UW) \in \mathcal{R}_M$. Hence under this setting the two sides match. Suppose, on the other hand, $I_{p^*}(VW;Z) \leq I_{p^*}(VW;Y)$, then we can see that setting $(W'=(V,W), U'=X, V'=0)$ does not decrease the right hand side. Note that in this case the right hand side is upper bounded by $(\alpha_1^{\mathbf{x}}+\alpha_2^{\mathbf{x}})I(X;Y)$. However the pair $(R_1 = I(X;Y), R_2=0) \in \mathcal{R}_M$. This implies the reverse inequality as desired.

Finally assume that $I_{p^*}(UW;Y) > I_{p^*}(UW;Z)$ and $I_{p^*}(VW;Z) > I_{p^*}(VW;Y)$. If $I(W;Y) < I(W;Z)$, then by choosing Q to be Bernoulli(ϵ) and setting $W'=(U,W), V'=X, U'=0$ when $Q=1$ and $W'=W, V'=V, U'=U$ when $Q=0$ we can increase the right hand side. Here $\epsilon \in (0,1)$ is chosen to make $I(QW';Y) = I(QW';Z)$. (Similar arguments have been employed earlier [11], [12] and hence we are being terse.) Thus we can assume that $I_{p^*}(W;Y) = I_{p^*}(W;Z)$. Then, note that $R_1 = I_{p^*}(UW;Y)$, $R_2 = I_p^*(V;Z|W) - I_p^*(V;U|W)$ belongs to $\mathcal{R}_M$. This implies that the two maximizations yield identical values. Further note that the $\mathbf{x} \in \mathcal{E}$ can be attained by a suitable choice of a maximizing distribution of the right hand side. ■

Note $\min\{I(W;Y), I(W;Z)\} = \min_{\lambda\in[0,1]} \lambda I(W;Y) + (1-\lambda)I(W;Z)$. The following min-max equality follows from the min-max theorem and its corollary in [9].

Claim 1. *[9] Let $\alpha_1 \geq 0, \alpha_2 > 0$. We have the following min-max relation:*

$$\begin{aligned}
&\max_{\substack{(U,V,W):\\(U,V,W)\to X\to(Y,Z)}} \min_{\lambda\in[0,1]} \alpha_1 I(UW;Y)\\
&\quad + \alpha_2\big(\lambda I(W;Y) + (1-\lambda)I(W;Z) + I(U;Y|W)\\
&\quad + I(V;Z|W) - I(U;V|W)\big)\\
&= \min_{\lambda\in[0,1]} \max_{\substack{(U,V,W):\\(U,V,W)\to X\to(Y,Z)}} \alpha_1 I(UW;Y)\\
&\quad + \alpha_2\big(\lambda I(W;Y) + (1-\lambda)I(W;Z) + I(U;Y|W)\\
&\quad + I(V;Z|W) - I(U;V|W)\big)
\end{aligned}$$

Further we can also infer that there is a $p^*(u,v,w,x)$ that achieves the common value in both problems. To see this: let $\lambda^* \in [0,1]$ and $q^*(u,v,w,x)$ achieve the value of the $\min$-$\max$ formulation and $p^*(u,v,w,x)$ achieve the common value using $\max$-$\min$ formulation. Then observe that

$$\lambda^*\text{-}SR_\alpha(q^*) \geq \lambda^*\text{-}SR_\alpha(p^*) \geq SR_\alpha(p^*),$$

where

$$\begin{aligned}
\lambda\text{-}SR_\alpha(p) &= \alpha_1 I_p(UW;Y) + \alpha_2\big(\lambda I_p(W;Y)\\
&\quad + (1-\lambda)I_p(W;Z) + I_p(U;Y|W)\\
&\quad + I_p(V;Z|W) - I_p(U;V|W)\big),
\end{aligned}$$

and

$$\begin{aligned}
SR_\alpha(p) &= \alpha_1 I_p(UW;Y) + \alpha_2\big(\min\{I_p(W;Y), I_p(W;Z)\}\\
&\quad + I_p(U;Y|W) + I_p(V;Z|W) - I_p(U;V|W)\big).
\end{aligned}$$

Equality holds since the values at both ends are same by Claim 1 and hence we have that $p^*(u,v,w,x)$ also achieves the value of the $\min$-$\max$ formulation.

Starting with any maximizer $p^*(u,v,w,x)$ of the $\min$-$\max$ formulation, we will find another maximizer, using Caratheodory's theorem, Lemma 4, and Theorem 2, $r^*(u,v,w,x)$ with $|\mathcal{W}| \leq |\mathcal{X}|+2, |\mathcal{U}|+|\mathcal{V}| \leq |\mathcal{X}|+1$ such that $\lambda^*\text{-}SR_\alpha(p^*) = \lambda^*\text{-}SR_\alpha(r^*), I_{p^*}(W;Y) =$

$I_{r^*}(W;Y), I_{p^*}(W;Z) = I_{r^*}(W;Z)$. Thus we will have $\lambda^*\text{-}SR_\alpha(r^*) = SR_\alpha(r^*)$.

For any fixed $\lambda \in [0,1]$ consider maximizing $\lambda\text{-}SR_\alpha(p)$ over $p(u,v,w,x)$. For simplicity of notation, let $\bar\lambda = 1-\lambda$. We can express $\lambda\text{-}SR_\alpha(p)$ as

$$\begin{aligned}&(\alpha_1+\alpha_2\lambda)H(Y)+\bar\lambda\alpha_2 H(Z)\\&+\Big(-(\alpha_1+\alpha_2\lambda)H(Y|W)-\bar\lambda\alpha_2H(Z|W)\\&+(\alpha_1+\alpha_2)I(U;Y|W)+\alpha_2(I(V;Z|W)-I(U;V|W))\Big)\end{aligned}$$

Using an upper concave envelope[1] representation the above problem is equivalent to: compute $\max_{p(x)}$

$$\begin{aligned}&(\alpha_1+\alpha_2\lambda)H(Y)+\bar\lambda\alpha_2H(Z)\\&+\mathfrak{C}\Big[-(\alpha_1+\alpha_2\lambda)H(Y)-\bar\lambda\alpha_2H(Z)\\&+\max_{p(u,v|x)}\big((\alpha_1+\alpha_2)I(U;Y)+\alpha_2(I(V;Z)-I(V;U))\big)\Big]\end{aligned} \tag{1}$$

We know that associated with any point on the upper concave envelope w.r.t. $p(x)$, there exists a supporting hyperplane $\sum_x c_x p_x = c$ such that $p(u,v,x)$ attains a global maximum of

$$\begin{aligned}&\max_{p(u,v,x)} -(\alpha_1+\alpha_2\lambda)H(Y)-\bar\lambda\alpha_2H(Z)\\&+(\alpha_1+\alpha_2)I(U;Y)+\alpha_2(I(V;Z)-I(V;U))-\sum_x c_x p(x).\end{aligned} \tag{2}$$

Our main results stem from the formulation of the optimization problem given by (2).

The proof uses perturbation ideas. Here we present a preliminary lemma that will be used later on.

Lemma 3. *Consider three random variables X,Y,Z and a perturbation of their probabilities defined according to*

$$p_\epsilon(x,y,z)=p(x,y,z)(1+\epsilon L(x,y,z)).$$

Then we have that

$$\frac{d^2}{d\epsilon^2}\left[H(Y|Z)\right]_{\epsilon=0}=0$$

implies that $H_{p_\epsilon}(Y|Z)$ is linear in ϵ.

Proof: Routine calculations yield that

$$\frac{d^2}{d\epsilon^2}\left[H(Y|Z)\right]_{\epsilon=0}=-\mathrm{E}(\mathrm{E}(L|Y,Z)^2)+\mathrm{E}(\mathrm{E}(L|Z)^2)$$

and hence the second derivative being zero (at $\epsilon=0$) implies that

$$\mathrm{E}(\mathrm{E}(L|Y,Z)^2)-\mathrm{E}(\mathrm{E}(L|Z)^2)=0.$$

Since $\mathrm{E}(L|Z)=\mathrm{E}_Y(\mathrm{E}(L|Y,Z))$ the above equality can be written as $\mathrm{E}\left[\mathrm{E}(L|Y,Z)-\mathrm{E}(L|Z)\right]^2=0$, thus $\mathrm{E}(L|Y=y,Z=z)=\mathrm{E}(L|Z=z)$ whenever $p(y,z)>0$.

Observe that $H_{p_\epsilon}(Y|Z)$ is equal to

$$\begin{aligned}&-\sum_{yz:p(y,z)>0}p(y,z)(1+\epsilon\,\mathrm{E}(L|Y=y,Z=z))\\&\qquad\times\log\frac{p(y,z)(1+\epsilon\,\mathrm{E}(L|Y=y,Z=z))}{p(z)(1+\epsilon\,\mathrm{E}(L|Z=z))}\\&=-\sum_{yz:p(y,z)>0}p(y,z)(1+\epsilon\,\mathrm{E}(L|Y=y,Z=z))\\&\qquad\times\log\frac{p(y,z)(1+\epsilon\,\mathrm{E}(L|Z=z))}{p(z)(1+\epsilon\,\mathrm{E}(L|Z=z))}\\&=-\sum_{yz:p(y,z)>0}p(y,z)(1+\epsilon\,\mathrm{E}(L|Y=y,Z=z))\log p(y|z).\end{aligned}$$

Note $\mathrm{E}(L|Y=y,Z=z)=\mathrm{E}(L|Z=z)$ implies the second equality. Thus $H_{p_\epsilon}(Y|Z)$ is linear in ϵ. ■

II. Main: New Ideas, Results, and Proofs

Consider the following maximization problem: Given $\alpha_1\geq 0, \alpha_2>0$ and $\lambda\in[0,1]$ determine $\max_{p(u,v,x)} T(p(u,v,x)$ where

$$\begin{aligned}T(p(u,v,x)):=&-(\alpha_1+\alpha_2\lambda)H(Y)-\bar\lambda\alpha_2H(Z)\\&+(\alpha_1+\alpha_2)I(U;Y)+\alpha_2(I(V;Z)-I(V;U))-\sum_x c_xp(x).\end{aligned} \tag{3}$$

Theorem 2. *Let (U,V) be a cardinality minimal pair (in the sense of $|\mathcal{U}|+|\mathcal{V}|$) such that $p(u,v,x)$ is a maximizer of* (3). *Then one cannot find $\delta_1(u)$ and $\delta_2(v)$ such that $\delta_1(u)$ and $\delta_2(v)$ are not simultaneously zero for all u and v and further*

$$\sum_u p(u)\delta_1(u)=0,\quad \sum_v p(v)\delta_2(v)=0,$$

$$\sum_{uv}p(u,v,x)\delta_1(u)-\sum_{uv}p(u,v,x)\delta_2(v)=0\quad\forall x.$$

Proof: For the main part of the proof, we will assume that $\lambda\in(0,1)$. The extreme cases are rather easy and will be taken care of in Remark 2.

Suppose are given that $p(u,v,x)$ is a (cardinality-minimal) global maximizer of $T(p(u,v,x))$. Let us first consider perturbations of the form $p_\epsilon^{(1)}(u,v,x)=p(u,v,x)(1+\epsilon\delta_1(u))$ such that $\sum_u p(u)\delta_1(u)=0$. In this case we are preserving $p(v,x|u)$ and perturbing the marginal distribution of U. Express $T(p(x,y,z))$ as follows:

$$\begin{aligned}&=\alpha_2\bar\lambda(H(Y)-H(Z))-(\alpha_1+\alpha_2)H(Y|U)\\&\quad+\alpha_2(H(V|U)-H(V|Z))-\sum_x c_xp(x).\end{aligned}$$

The terms $H(V|U), H(Y|U), \sum_x c_xp_x$ are linear in ϵ since we are preserving $p(v,x|u)$, and the terms $-H(V|Z), -H(Y)$ are convex in ϵ. The first derivative for this perturbation has to be zero, and the second derivative has to be non-positive. Since $\alpha_2\bar\lambda>0$ for the second derivative to be non-positive we must have

$$\frac{d^2}{d\epsilon^2}\left[H(Y)-H(Z)\right]_{\epsilon=0}\leq 0. \tag{4}$$

Note that the above second derivative depends solely on $p_\epsilon^{(1)}(x)=\sum_{uv}p(u,v,x)(1+\epsilon\delta_1(u))$.

Next consider perturbations of the form $p_\epsilon^{(2)}(u,v,x)=p(u,v,x)(1+\epsilon\delta_2(v))$ such that $\sum_v p(v)\delta_2(v)=0$. In this

[1] The upper concave envelope of a function $f(x)$ denoted by $\mathfrak{C}[f](x)$ is the smallest concave function that dominates $f(x)$.

case we are preserving $p(u,x|v)$ and perturbing the marginal distribution of V. Express $T(p(x,y,z))$ as follows:

$$\alpha_2\lambda(H(Z)-H(Y)) - \alpha_1 H(Y|U) - \alpha_2 H(U|Y) + \alpha_2 H(U|V) - \alpha_2 H(Z|V) - \sum_x c_x p(x).$$

The terms $H(U|V), H(Z|V), \sum_x c_x p_x$ are linear in ϵ since we are preserving $p(u,x|v)$ and the terms $-H(Y|U), -H(U|Y), -H(Z)$ are convex in ϵ. The first derivative for this perturbation has to be zero, and the second derivative has to be non-positive. Since we assumed that $\lambda\alpha_2 > 0$ we can conclude that

$$\frac{d^2}{d\epsilon^2}\left[H(Z)-H(Y)\right]_{\epsilon=0} \leq 0, \tag{5}$$

and the above second derivative depends solely on $p_\epsilon^{(2)}(x) = \sum_{uv} p(u,v,x)(1+\epsilon\delta_2(v))$.

Now let us assume that one can find $\delta_1(u)$ and $\delta_2(v)$ such that $\delta_1(u)$ and $\delta_2(v)$ are not simultaneously zero for all u and v and further

$$\sum_u p(u)\delta_1(u) = 0, \sum_v p(v)\delta_2(v) = 0,$$
$$\sum_{uv} p(u,v,x)\delta_1(u) - \sum_{uv} p(u,v,x)\delta_2(v) = 0 \quad \forall x.$$

We will arrive at a contradiction.

Considering the two perturbations induced by the non-trivial pair $\delta_1(u)$ and $\delta_2(v)$. Since we are at a global maximum, the second derivatives with respect to both of these perturbations have to be non-positive. By our choice we know that they induce a common perturbation in $p(x)$, defined according to

$$\begin{aligned} p_\epsilon(x) &= \sum_{uv} p(u,v,x)(1+\epsilon\delta_2(v)) \\ &= \sum_{uv} p(u,v,x)(1+\epsilon\delta_1(u)). \end{aligned}$$

Hence, for both perturbations, from (4), (5) we have,

$$\frac{d^2}{d\epsilon^2}\left[H(Z)-H(Y)\right]_{\epsilon=0} = 0.$$

Further for the perturbation using $\delta_1(u)$, we must have

$$\frac{d^2}{d\epsilon^2}\left[H(V|Z)\right]_{\epsilon=0} = 0,$$

and for the perturbation using $\delta_2(v)$, we must have

$$\frac{d^2}{d\epsilon^2}\left[H(U|Y)\right]_{\epsilon=0} = \frac{d^2}{d\epsilon^2}\left[\alpha_1 H(Y|U)\right]_{\epsilon=0} = 0.$$

Lemma 3 implies that for the perturbation using $\delta_1(u)$ the term $H_{p_\epsilon^{(1)}}(V|Z)$ is linear in ϵ, and for the perturbation using $\delta_2(v)$, the terms $H_{p_\epsilon^{(2)}}(U|Y)$ and $\alpha_1 H_{p_\epsilon^{(2)}}(Y|U)$ are linear in ϵ.

Now express $H(Y)-H(Z)$ as a function of ϵ induced by the common $p_\epsilon(x)$. Note that in both perturbations we have the term $H(Y)-H(Z)$ multiplied by a constant, plus a linear term in ϵ that may depend on the perturbations. In other words, we can write $T(p_\epsilon^{(1)}(u,v,x))$ as follows:

$$\alpha_2\bar{\lambda}(H(Y)-H(Z))_{p_\epsilon(x)} + a_1\epsilon + b_1,$$

and we can write $T(p_\epsilon^{(2)}(u,v,x))$ as follows:

$$\lambda\alpha_2(H(Z)-H(Y)) + a_2\epsilon + b_2.$$

Now, by pulling out the linear term and the constant terms from $(H(Y)-H(Z))_{p_\epsilon(x)}$, we can write this difference of the entropies as $g(\epsilon)+a\epsilon+b$, i.e. $g(0)=0$ and $\frac{d}{d\epsilon}g(\epsilon)\mid_{\epsilon=0}= 0$. Hence we can write $T(p_\epsilon^{(1)}(u,v,x))$ as

$$(1-\lambda)(\alpha_0+\alpha_2)g(\epsilon) + a_1'\epsilon + b_1',$$

and $T(p_\epsilon^{(2)}(u,v,x))$ as

$$-\lambda(\alpha_0+\alpha_2)g(\epsilon) + a_2'\epsilon + b_2',$$

for some appropriately defined constants a_1', a_2', b_1', b_2'.

Since $p(u,v,x)$ attains a global maximum and hence w.r.t. to ϵ for both perturbations, we have that the first derivative is zero in each perturbation, i.e. $a_1' = a_2' = 0$. Further at $\epsilon = 0$ we have $g(0) = 0$ and $p_\epsilon^{(1)}(u,v,x) = p_\epsilon^{(2)}(u,v,x) = p(u,v,x)$. This implies that $b_1' = b_2'$, say equal to b'.

Since $T(p_\epsilon^{(1)}(u,v,x))$ has a global maximum at $\epsilon = 0$ and $\lambda \in (0,1)$, we must have $g(\epsilon) \leq 0$; similarly since $T(p_\epsilon^{(2)}(u,v,x))$ has a global maximum at $\epsilon = 0$, we must have $g(\epsilon) \geq 0$; thus we have $g(\epsilon) = 0$.

Thus, $T(p(u,v,x))$ remains constant under *both perturbations*. Now if we take a non-trivial perturbation, say $\delta_1(u)$ and take ϵ to its upper or lower limit (attained when one of the probabilities reaches zero), we get a distribution on U whose support is smaller than that for $\epsilon = 0$ while the value of $T(p(u,v,x))$ has been preserved. Thus we have been able to reduce $|\mathcal{U}|+|\mathcal{V}|$, which is a contradiction. ■

Remark 2. The case when $\lambda = 0$ or $\lambda = 1$ is much simpler. For $\lambda = 0$ observe that for any perturbation of the type $p(u,v,x)(1+\epsilon\delta_2(v))$, the resultant expression is convex in ϵ, and hence we can take the limiting epsilon (either the upper or lower limit) and reduce the cardinality of V. Indeed, it is not hard to see that an optimal choice is $V = 0$. Thus combining with the previously known bound of $|\mathcal{U}| \leq |\mathcal{X}|$; we have the desired claim.

Lemma 4. *For any $p(u,v,x)$ where $|\mathcal{U}|+|\mathcal{V}| > |\mathcal{X}|+1$, one can find $\delta_1(u)$ and $\delta_2(v)$ such that $\delta_1(u)$ and $\delta_2(v)$ are not simultaneously zero for all u and v and further*

$$\sum_u p(u)\delta_1(u) = 0, \quad \sum_v p(v)\delta_2(v) = 0,$$
$$\sum_{uv} p(u,v,x)\delta_1(u) - \sum_{uv} p(u,v,x)\delta_2(v) = 0 \quad \forall x.$$

Proof: These are $|\mathcal{X}|+2$ equations in total but one of the equations is redundant since

$$\begin{aligned} \sum_x \sum_{uv} p(u,v,x)\delta_1(u) &= \sum_u p(u)\delta_1(u) = 0 \\ &= \sum_v p(v)\delta_2(v) = \sum_x \sum_{uv} p(u,v,x)\delta_2(v). \end{aligned}$$

Thus

$$\sum_x \left(\sum_{uv} p(u,v,x)\delta_1(u) - \sum_{uv} p(u,v,x)\delta_2(v) \right) = 0.$$

Thus, there are $|\mathcal{X}|+1$ independent equations at most. The choice $\delta_1(u)=\delta_2(v)=0$ for all u,v solves the above system of linear equations. Therefore the system of linear equations is not inconsistent. Since the total number of free variables is $|\mathcal{U}|+|\mathcal{V}|$ which is strictly larger than the number of equations, i.e. $|\mathcal{X}|+1$ we must have a non-trivial solution for $\delta_1(u)$ and $\delta_2(v)$. ■

A. Proof of Theorem 1

Choose a $p^*(u,v,w,x)$ such that it achieves the maximum value of the $\min\text{-}\max$ and $\max\text{-}\min$ simultaneously in Claim 1. Theorem 2 and Lemma 4 together implies that one can assume $|\mathcal{U}|+|\mathcal{V}|\leq|\mathcal{X}|+1$ to compute the upper concave envelope at any $p(x)$ denoted in (1), hence it suffices to consider $|\mathcal{U}|+|\mathcal{V}|\leq|\mathcal{X}|+1$ to achieve any exposed point in $\mathcal{E}\subseteq\mathcal{R}_M$. Hence by continuity of the terms, it suffices to consider $|\mathcal{U}|+|\mathcal{V}|\leq|\mathcal{X}|+1$ to achieve any extreme point of $\mathcal{R}_M$. Note that the exposed points not belonging to $\mathcal{E}$ already satisfy this cardinality constraint. Since any point on the boundary is a convex combination of extreme points, and this convex combination can be attained by choosing $W'=(W,Q)$, this does not alter that cardinality bounds un U and V. To get the cardinality bound on W, we use the Fenchel-Bunt extension of Caratheodory's theorem. Choose points $p(u,v,x|w)$ with probabilities $p(w)$ so as to preserve $p(x),H(Y|W),H(Z|W),I(U;Y|W),I(V;Z|W),I(U;V|W)$. This can be done using a W of size at most $|\mathcal{X}|+4$. This choice preserves all the terms that appear in Bound 2 and hence these cardinality bounds suffice to compute $\mathcal{R}_M$.

B. Binary input broadcast channels

Theorem 2 establishes Conjecture 1 in [4]. In particular when $\mathcal{X}=\{0,1\}$ we have that $|\mathcal{U}|+|\mathcal{V}|\leq 3$. Thus, it is immediate that it suffices to consider $U=X,V=0$ or $V=X,U=0$. Thus Marton's region for binary input broadcast channels reduce to:

Bound 2. *The union of non-negative rate pairs R_1,R_2 satisfying the constraints*

$$R_1\leq I(W;Y)+\sum_{i=1}^{k}p_kI(X;Y|W=i)$$
$$R_2\leq I(W;Z)+\sum_{i=k+1}^{l}p_kI(X;Z|W=i),$$
$$R_1+R_2\leq\min\{I(W;Y),I(W;Z)\}+\sum_{i=1}^{k}p_kI(X;Y|W=i),$$
$$\sum_{i=k+1}^{l}p_kI(X;Z|W=i),$$

over random variable W, with $|\mathcal{W}|=l\leq 5$ such that $W\to X\to(Y,Z)$ is Markov.

Note that the reduction to $|\mathcal{W}|=l\leq 5$ is possible since $I(U;V|W)=0$ whenever U or V is a constant. Thus we do not need to preserve this average when using Caratheodory's theorem. It may be further possible to reduce the cardinality of W in the binary setting. The main objective of this paper was to obtain sharper bounds on U and V.

Conclusion

In this paper we improved the cardinality bounds on the auxiliary random variables used in Marton's inner bound. We also resolved a conjecture, made by us, in [4] relating to binary input broadcast channels. Exhaustive numerical simulations, made possible by the results in this paper, indicate that the second conjecture in [4] is valid for binary input broadcast channels. This yields a stronger evidence towards the potential validity of the second conjecture in [4], implying optimality of Marton's inner bound for the broadcast channel. The techniques introduced here can be used potentially to extract more properties of auxiliary random variables that achieve the boundary of Marton's inner bound beyond yielding improved cardinality bounds.

Acknowledgements

V. Anantharam gratefully acknowledges research support from the ARO MURI grant W911NF-08-1-0233, "Tools for the Analysis and Design of Complex Multi-Scale Networks", from the NSF grant CNS-0910702, and from the NSF Science & Technology Center grant CCF-0939370, "Science of Information". Amin Gohari's work was supported by Iran-NSF grant No. 88114.46.The work of Chandra Nair was supported by an AoE grant (Project No. AoE/E-02/08) and two GRF grants (Project Nos. 415810 and 415612) from the UGC of the Hong Kong Special Administrative Region, China.

References

[1] A. El Gamal and Y.-H. Kim, Network Information Theory. Cambridge University Press, 2012.

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

[3] A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," International Symposium on Information Theory, pp. 2462–2466, 2009.

[4] A. A. Gohari, C. Nair, and V. Anantharam, "On Marton's inner bound for broadcast channels," in International Symposium on Information Theory, Cambridge, Massachusetts, USA, Jul. 2012, pp. 586–590.

[5] 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.

[6] A. Gohari, A. El Gamal, and V. Anantharam, "On an outer bound and an inner bound for the general broadcast channel," International Symposium on Information Theory, 2010.

[7] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," IEEE Trans. Info. Theory, vol. IT-58, no. 2, pp. 608–619, 2012.

[8] Y. Geng, A. Gohari, C. Nair, and Y. Yu, "On Marton's inner bound for two receiver broadcast channels," Presented at ITA Workshop, 2011.

[9] ——, "The capacity region of classes of product broadcast channels," International Symposium on Information Theory, pp. 1549–1553, 2011.

[10] B. Grünbaum and G. Ziegler, Convex Polytopes, ser. Pure and applied mathematics. Springer, 2003. [Online]. Available: http://books.google.com.hk/books?id=5iV75P9gIUgC

[11] S. I. Gelfand and M. S. Pinsker, "Capacity of a broadcast channel with one deterministic component," Probl. Inform. Transm., vol. 16(1), pp. 17–25, Jan. - Mar., 1980.

[12] Y. Liang, G. Kramer, and H. V. Poor, "On the equivalence of two achievable regions for the broadcast channel," IEEE Trans. Info. Theory, vol. IT-57, no. 1, pp. 95–100, 2011.