A Conjecture Regarding the Optimizers of Marton's Inner Bound for the Two-Receiver Broadcast Channel

Amin Gohari, Yi Liu and Chandra Nair
Department of Information Engineering
The Chinese University of Hong Kong
Sha Tin, N.T., Hong Kong
Email: {agohari, ly023, chandra}@ie.cuhk.edu.hk

Abstract

We study Marton's inner bound for a general two-receiver discrete memoryless broadcast channel. We conjecture a structural result on the optimizers of Marton's inner bound, which, if true, would greatly simplify the evaluation of the bound and provide new insights into the underlying optimization problem. We derive an equivalent characterization for the sum-rate that demonstrates a similar decoupling as one suggested by the conjecture.

I. Introduction

The communication system depicted in Figure 1 illustrates a two-receiver broadcast channel [2], wherein the sender aims to transmit a private message to each receiver reliably.[1] A detailed definition of a broadcast channel can be found in [4]. A computable characterization of the capacity region of a general two-receiver broadcast channel remains unknown. To date, the best achievable rate region is Marton's inner bound [15].

Theorem 1. *(Marton '79) The union of non-negative rate pairs (R_1, R_2) satisfying the constraints*

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

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

Given a broadcast channel T with transition probability $T(y,z|x)$, we denote the Marton's achievable region by $\mathcal{R}_M$. In [9], $\mathcal{R}_M$ is shown to be invariant if restricted to $|\mathcal{W}| \le |\mathcal{X}| + 4$, and conditioned on any $W = w$, the sizes of $\mathcal{U}$ and $\mathcal{V}$ are restricted to satisfy $|\mathcal{U}| + |\mathcal{V}| \le |\mathcal{X}| + 1$. Furthermore, we can assume X is a function of U and V. In this paper, we restrict ourselves to the dense class of broadcast channels where $T_{Y|X}(y|x) > 0$ and $T_{Z|X}(z|x) > 0$. For such classes, it is known [11, Theorem 1] that $p(u,v) > 0$ for all u, v. Further properties of the distributions corresponding to the boundary points of $\mathcal{R}_M$ can be found in [1], [7]–[11].

For any $\lambda \ge 1$, we define the weighted sum rate of Marton's inner bound as:

$$\begin{aligned} \Gamma_M^\lambda(T) &= \max_{(R_1,R_2)\in\mathcal{R}_M} R_1 + \lambda R_2, \\ &= \max_{p(u,v,w,x)} \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) + I(V;Z|W) - I(U;V|W) + (\lambda-1)I(V,W;Z). \end{aligned}$$

The weighted sum rate is known to have a simpler characterization without the auxiliary random variable W, which is derived via convex duality in [1]. For an arbitrary sequence $\{a_x \in \mathbb{R}^{|\mathcal{X}|} : x \in \mathcal{X}\}$, we define the dual characterization:

$$F(T, \{a_x\}) = \max_{p(u,v,x)} \big[-\alpha H(Y) - (\lambda-\alpha)H(Z) + I(U;Y) + \lambda I(V;Z) - I(U;V) + \sum_x p(x) a_x \big],$$

where the maximum is taken over all $p(u,v,x)$ defined on $\mathcal{U} \times \mathcal{V} \times \mathcal{X}$. Then,

$$\Gamma_M^\lambda(T) = \min_{\alpha\in[0,1]} \max_{p(x)} \Big\{ \alpha H(Y) + (\lambda-\alpha)H(Z) + \min_{a_x} \big[F(T,\{a_x\}) - \sum_x p(x) a_x \big] \Big\}. \tag{1}$$

We provide a proof of the above equation in Appendix A for completeness.

Given any two broadcast channels T_1 and T_2, with transition probabilities $T_1(y_1, z_1|x_1)$ and $T_2(y_2, z_2|x_2)$ respectively, we further define the dual characterization of Marton's inner bound for the product channel $T_1 \otimes T_2$:

$$\begin{aligned} F(T_1 \otimes T_2, \{a_{x_1} + b_{x_2}\}) = \max_{p(u,v,x_1,x_2)} \big[&-\alpha H(Y_1,Y_2) - (\lambda-\alpha)H(Z_1,Z_2) + I(U;Y_1,Y_2) + \lambda I(V;Z_1,Z_2) - I(U;V) \\ &+ \sum_{x_1,x_2} p(x_1,x_2)(a_{x_1} + b_{x_2}) \big]. \end{aligned}$$

Conjecture 1. *(Additivity Conjecture) [1] The dual characterization of Marton's inner bound is additive, i.e., for any $\{a_{x_1}\}, \{b_{x_2}\}$ we have*

$$F(T_1 \otimes T_2, \{a_{x_1} + b_{x_2}\}) = F(T_1, \{a_{x_1}\}) + F(T_2, \{b_{x_2}\}).$$

[1] For simplicity, we focus on the setting with private messages only. Extension to the setting with a common message is possible.

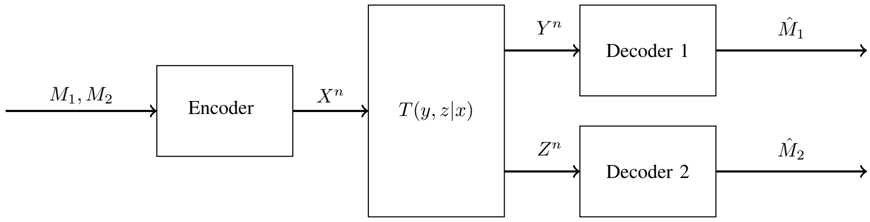

Fig. 1. Two-receiver broadcast communication system.

Conjecture 1 leads to Marton's inner bound being the capacity region, because the weighted sum rate is then sub-additive:

$$\begin{aligned}\Gamma_M^\lambda(T_1\otimes T_2) &= \min_{\alpha\in[0,1]}\max_{p(x_1,x_2)}\left\{\alpha H(Y_1,Y_2)+(\lambda-\alpha)H(Z_1,Z_2)+\min_{a_{x_1},b_{x_2}}\left[F(T_1\otimes T_2,\{a_{x_1}+b_{x_2}\})-\sum_x p(x)(a_{x_1}+b_{x_2})\right]\right\}\\ &\le \min_{\alpha\in[0,1]}\max_{p(x_1)}\left\{\alpha H(Y_1)+(\lambda-\alpha)H(Z_1)+\min_{a_{x_1}}\left[F(T_1,\{a_{x_1}\})-\sum_x p(x)a_{x_1}\right]\right\}\\ &\quad+\min_{\alpha\in[0,1]}\max_{p(x_2)}\left\{\alpha H(Y_2)+(\lambda-\alpha)H(Z_2)+\min_{b_{x_2}}\left[F(T_2,\{b_{x_2}\})-\sum_x p(x)b_{x_2}\right]\right\}\\ &=\Gamma_M^\lambda(T_1)+\Gamma_M^\lambda(T_2).\end{aligned}$$

As established in [17], Conjecture 1 is equivalent to the following: let $p^*(u_1,v_1,w_1,x_1)$ be a maximizer of $F(T_1,\{a_{x_1}\})$ and $p^\dagger(u_2,v_2,w_2,x_2)$ be a maximizer of $F(T_2,\{b_{x_2}\})$, then $p^*(u_1,v_1,w_1,x_1)\otimes p^\dagger(u_2,v_2,w_2,x_2)$, with $U=(U_1,U_2)$ and $V=(V_1,V_2)$, is a maximizer of $F(T_1\otimes T_2,\{a_{x_1}+b_{x_2}\})$.

II. A CONJECTURE AND ITS IMPLICATIONS

Conjecture 2. *(The Markovity Conjecture) When evaluating the extremal points of Marton's achievable region, it is sufficient to consider the random variable tuples (U,V,W,X) such that the Markov chain $U\to(W,X)\to V$ holds. Equivalently, for every arbitrary $\{a_x\}$, to compute $F(T,\{a_x\})$, it is sufficient to consider the random variable tuples (U,V,X) such that the Markov chain $U\to X\to V$ holds.*

If Conjecture 2 holds, the calculation of Marton's inner bound will be greatly simplified, providing insights into why it may potentially be the capacity region, as we discuss in the next subsection. The evidence for the Markovity Conjecture arises from numerical simulations. We have identified and verified maximizers of $F(T,\{a_x\})$ across various channels with different input alphabet sizes, confirming that the Markov chain $U\to X\to V$ holds. Specifically, we examined over 10,000 channels for each input alphabet size of $|\mathcal{X}|=3,4,5$, and more than 100 channels with $|\mathcal{X}|=6,7$. The conjecture is also consistent with previous results on optimal mappings for Marton maximizers:

- *Binary Input Channels*: Randomized time-division strategy is optimal. Hence, the maximizers, conditioning on each $W=w$, have either U be constant or V be constant [16]. Thus, $I(U;V|W=w,X)=0$ holds trivially in this case.
- *Semi-deterministic Channels*: The maximizers satisfy either $V=f(X)$ or $U=g(Y)$ so $I(U;V|X)=0$ follows.
- *Erasure Blackwell Channel*: To evaluate $R_1+\lambda R_2$ for the k-letter Marton, it is optimal to set $U=\hat{Y}^k$, provided that λ is greater than certain threshold [12]. Here, $\hat{Y}$ is the first layer deterministic channel output and therefore $H(U|X^k)=0$.
- *Degraded/Less-noisy/More-capable Channels*: The superposition coding region is the capacity region, which can be deduced from Marton's region by either setting U be constant or V be constant.

Remark 1. We have also verified, with limited numerical simulations, that the conjecture holds for the vector Gaussian broadcast channel. Despite knowing that Marton's bound matches the capacity region [6], [18], the Markovity conjecture seems to yield more information than what is known regarding the true maximizers. However, as was shown in [13], the associated non-convex problem has a unique local maximizer, thus enabling reliable simulation evidence.

A. Implications of the Markovity conjecture

One immediate consequence of the conjecture (if true) is that a decoupling of the optimization problem occurs in evaluating the weighted sum rate of Marton's inner bound. Note that,

$$\begin{aligned}F(T,\{a_x\}) &= \max_{p(u,v,x)}\big[-\alpha H(Y)-(\lambda-\alpha)H(Z)+I(U;Y)+\lambda I(V;Z)-I(U;V)+\sum_x p(x)a_x\big]\\ &=\max_{p(x)}\big[-\alpha H(Y)-(\lambda-\alpha)H(Z)+\max_{p(u,v|x)}\big\{I(U;Y)+\lambda I(V;Z)-I(U;V)\big\}+\sum_x p(x)a_x\big].\end{aligned}$$

Let $p^*(u,v|x)$ be a maximizer such that X is a function of (U,V) and $U\to X\to V$ (existence of such a maximizer is postulated by the conjecture) forms a Markov chain. Fix any $x\in\mathcal{X}$, and define

$$\mathcal{S}_x := \{u : p^*(u|x)>0, u\in\mathcal{U}\},$$
$$\mathcal{T}_x := \{v : p^*(v|x)>0, v\in\mathcal{V}\}.$$

Then, we have the following consequence of the Conjecture 2.

Proposition 1.

$$p^*(x|u,v)=1,\quad \forall u,v\in\mathcal{S}_x\times\mathcal{T}_x.$$

Proof. Note that $p^*(u|x) > 0$ and $p^*(v|x) > 0$. Because of the Markov chain $p^*(u,v|x) = p^*(u|x)p^*(v|x)$, we obtain that $p^*(u,v|x) > 0$. As $p^*(u,v) > 0$ for every (u,v) (for channels where $T(y|x)$ and $T(z|x) > 0$), we have $p^*(x|u,v) = \frac{p^*(u,v|x)p(x)}{p^*(u,v)} > 0$. On the other hand, as X is determined by U,V, we must have $p^*(x|u,v) = 1$. □

Proposition 1 implies the following: Consider a $|\mathcal{U}| \times |\mathcal{V}|$ matrix where the matrix entries are determined by the mapping of (u,v) to $x \in \mathcal{X}$, i.e., in the (u,v) cell of the matrix, we put the symbol x that (u,v) is mapped into. Then, for a maximizer $p^*(u,v,x)$ satisfying the Markovity conjecture, the matrix locations marked as x, for any fixed $x \in \mathcal{X}$, will form a "rectangle", i.e. of the form $S_x \times T_x$. We call such a mapping to be a *rectangular mapping*. Thus, Conjecture 2, implies the existence of a maximizer with a rectangular mapping. It turns out that the following converse statement is also true.

Lemma 1. *Suppose there is a maximizer of $F(T,\{a_x\})$ with a rectangular mapping, then for this maximizer $U \to X \to V$ is Markov.*

Proof. Let us fix p_x and a rectangular mapping (i.e. fix the mapping $(U,V) \mapsto X$). Let $\mathcal{M}$ denote the class of distributions $p(u,v|x)$ consistent with the given mapping. Consider the maximization problem (we ignore the terms in $F(T,\{a_x\})$ that depends only on $p(x)$)

$$\begin{aligned}
&\max_{p(u,v|x)\in\mathcal{M}} \{I(U;Y) + \lambda I(V;Z) - I(U;V)\}\\
=&\max_{p(u,v|x)\in\mathcal{M}} \{I(U;Y) + \lambda I(V;Z) - I(U;X) - I(V;X) + I(U,V;X) - I(U;V|X)\}\\
\le&\max_{p(u,v|x)\in\mathcal{M}} \{I(U;Y) + \lambda I(V;Z) - I(U;X) - I(V;X) + I(U,V;X)\}\\
=&\max_{p(u|x):\text{supp }(p(u|x))\in\mathcal{S}_x} \{I(U;Y) - I(U;X)\} + \max_{p(v|x):\text{supp }(p(v|x))\in\mathcal{T}_x} \{\lambda I(V;Z) - I(V;X)\} + H(X).
\end{aligned}$$

Observe that the inequality becomes an equality precisely when $U \to X \to V$ is Markov. □

Remark 2. Therefore, an alternate way to state the Markovity conjecture is to state that there is a maximizer with a rectangular mapping.

B. Rectangular mappings and their implications about the additivity of $F(T,\{a_x\})$

Observe that the function $I(U;Y) - I(U;X)$ is concave in $p(u|x)$. This concavity implies the following: let us take Marton's inner bound for a product channel $T(y_1,z_1|x_1)T(y_2,z_2|x_2)$. Assuming the conjecture holds, let $p_1(x_1)p_1(u_1|x_1)p_1(v_1|x_1)$ and $p_2(x_2)p_2(u_2|x_2)p_2(v_2|x_2)$ be optimizers of $F(T_1,\{a_{x_1}\})$ and $F(T_2,\{a_{x_2}\})$ respectively. Consider a joint distribution $p_1(x_1)p_2(x_2)$ and a rectangular mapping induced by the tensor-product of the rectangular mappings of the individual components. Due to concavity, it is easy to see that the product distribution $p_1(u_1|x_1)p_2(u_2|x_2)$ is the maximizer of $I(U;Y_1,Y_2) - I(U;X_1,X_2)$ in the space of $p_{U|X_1,X_2}$. Despite the lack of concavity in $p(v|x)$ for $\lambda I(V;Z) - I(V;X)$, one can show that $p_1(v_1|x_1)p_2(v_2|x_2)$ will be a local maximizer of $\lambda I(V;Z_1,Z_2) - I(V;X_1,X_2)$ in the space of $p_{V|X_1,X_2}$. This implies a partial "local-tensorization" property in the sense of [17].

Results demonstrating that certain non-rectangular mappings cannot be maxmizers are already established in the literature. References [8] and [11] indicate that, for any arbitrary input alphabet size $|\mathcal{X}|$, the general AND mapping and the general XOR mapping cannot appear as part of an optimal mapping. Consider any two-by-two rectangle consisting of the entries (u_1,v_1), (u_2,v_1), (u_1,v_2), and (u_2,v_2). Let $\{A,B,C\}$ be three distinct input alphabets. It has been shown that it suffices to consider $p(u,v,x)$ that do not contain the mapping $\begin{bmatrix} A & B \\ B & A \end{bmatrix}$ to obtain maximizers for $F(T,\{a_x\})$ as any such rectangular sub-matrix. This is called the XOR mapping. To complete the proof of the Markovity conjecture, we just need to show that the following mapping, $\begin{bmatrix} A & B \\ C & A \end{bmatrix}$, also does not appear in any two-by-two rectangle.[2]

A partial theoretical evidence is known, when $|\mathcal{X}| = 3$, we do not need to consider the $\begin{bmatrix} A & B \\ A & A \end{bmatrix}$ mapping (called the generalized AND). Both the generalized XOR and the generalized AND were proven using local optimality conditions. This prompted us to look for a proof that $\begin{bmatrix} A & B \\ C & A \end{bmatrix}$ also does not appear in any two-by-two rectangle, using local optimality conditions. However, we failed to do this; and in particular, we have identified an instance where such a mapping is a local maximizer but not a global maximizer.

In the next section, we will outline the general local optimality conditions that we considered in our analysis, and we will present the specific example (non-rectangular mapping) that passed our local optimality criterion.

III. Existence of local maximizers

Let us return to the main additivity conjecture (Conjecture 1). In this section, we show that the optimization problem defining $F(T,\{a_x\})$, i.e., maximum over $p(u,v,x)$ of

$$G(p(u,v,x)) = -\alpha H(Y) - (\lambda-\alpha)H(Z) + I(U;Y) + \lambda I(V;Z) - I(U;V) + \sum_x p(x)a_x. \tag{2}$$

has local maximizers (in the sense described below) that are not global maximizers.

Remark 3. This shows the difficulty in working with discrete memoryless channels. In the Gaussian setting, as [13] shows, despite the problem being non-convex, there are no non-trivial local maximizers. This enables the feasibility of local arguments

[2]Note that once we have it, the 2×2 subpattern $\begin{bmatrix} A & A \\ B & A \end{bmatrix}$ can be also discarded by considering a perturbation which adds a new input symbol C to the input, and sets $p^{\epsilon}_{YZ|C}$ where $\lim_{\epsilon\to 0} p^{\epsilon}_{YZ|C} = p_{YZ|A}$.

to show the additivity. However, as the following example shows, one cannot directly work with $F(T,\{a_x\})$ if one wants to avoid spurious local maximizers.

We say that $p(u,v,x)$ is a local maximizer if the following three types of perturbations do not increase $G(p(u,v,x))$:

- Perturbations that keep the alphabet size of U and V unchanged: $p_\epsilon(u,v,x) = (1-\epsilon)p(u,v,x) + \epsilon s(u,v,x)$ for some distribution $s(u,v,x)$.
- Perturbations that increase the alphabet size of U by one:

$$p_\epsilon(u,v,x) = \begin{cases} p(u,v,x) - \epsilon s(u,v,x), & u \in [1:\mathcal{U}|], \\ \epsilon q(v,x), & u = |\mathcal{U}|+1. \end{cases} \tag{3}$$

  for some arbitrary distributions $q(v,x)$ and $s(u,v,x)$.
- Similarly, perturbations that increase the alphabet size of V by one.

First, consider perturbations of the form $p_\epsilon(u,v,x) = (1-\epsilon)p(u,v,x) + \epsilon s(u,v,x)$: the first derivative conditions are as follows: for every (u,v,x) we have

$$\sum_{y,z} T(y,z|x) \log \frac{p(u,y)p(v,z)^\lambda e^{a_x}}{p(y)^{\bar{\alpha}} p(z)^\alpha p(u,v) p(v)^{\lambda-1}} \le F(T,\{a_x\}). \tag{4}$$

Moreover, equality holds in (4) if $x = f(u,v)$.

The second derivative conditions are as follows: consider a multiplicative perturbation $p_\epsilon(u,v,x) = p(u,v,x)(1+\epsilon L(u,v,x))$ for some function $L(u,v,x)$ satisfying $\mathbb{E}[L]=0$. The first derivative under this class of perturbations is zero; hence, the second derivative of the expression $G(p_\epsilon(u,v,x))$ should be non-positive at $\epsilon=0$. From here, we obtain

$$-\bar{\alpha}\mathbb{E}[\mathbb{E}[L|Y]^2] - \alpha\mathbb{E}[\mathbb{E}[L|Z]^2] - \mathbb{E}[\mathbb{E}[L|U,V]^2] - (\lambda-1)\mathbb{E}[\mathbb{E}[L|V]^2] + \mathbb{E}[\mathbb{E}[L|U,Y]^2] + \lambda\mathbb{E}[\mathbb{E}[L|V,Z]^2] \le 0. \tag{5}$$

Next, consider the perturbation in (3), which increases the cardinality of U by 1. The first-order condition of this perturbation implies that

$$\sum_{v,x,y,z} q(v,x)T(y,z|x) \log \frac{q(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1} q(v)} + \sum_x q(x)a_x \le F(T,\{a_x\}), \qquad \forall q(v,x). \tag{6}$$

The original first derivative conditions in (4) do not imply these inequalities. Thus, they supplement the original first derivative conditions. Note that the expression in (6) is convex in $q(x|v)$, which implies, to maximize the left-hand-side, it suffices to consider $q(v,x)$ for which $q(x|v)\in\{0,1\}$. Finally, observe that from the equality condition of equation (4), equality in (6) holds if $q(v,x) = p(v,x|u)$ for some u. Thus,

$$\sum_{v,x,y,z} q(v,x)T(y,z|x) \log \frac{q(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1} q(v)} + \max_{q(v,x)} \sum_x q(x)a_x = F(T,\{a_x\}). \tag{7}$$

A similar inequality can be obtained by expanding the alphabet of V as follows:

$$\sum_{u,x,y,z} q(u,x)T(y,z|x) \log \frac{p(u,y)q(z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha q(u)} + \max_{q(u,x)} \sum_x q(x)a_x \le F(T,\{a_x\}), \tag{8}$$

with equality if $q(u,x) = p(u,x|v)$ for some v.

A. *Alternative characterization of the conditions in* (7) *and* (8)

Consider (7). Let us fix $q(x|v)$ and consider the optimization in terms of $q(v)$. Observe that for any $r(y)$ we have

$$\sum_{v,x,y,z} q(v,x)T(y,z|x) \log \frac{q(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1} q(v)} \ge \sum_{v,x,y,z} q(v,x)T(y,z|x) \log \frac{r(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1} q(v)}.$$

This implies that for any $r(y)$ we have

$$\sum_{v,x,y,z} q(v,x)T(y,z|x) \log \frac{r(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1} q(v)} + \sum_x q(x)a_x \le F(T,\{a_x\}).$$

Given a fixed $q(x|v)$, the optimal $q(v)$ that maximizes the left hand side can be found. The first derivative conditions imply

$$\sum_{x,y,z} q(x|v)T(y,z|x) \log \frac{e^{a_x} r(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1}} + \log\frac{1}{q(v)} = K, \quad \forall v,$$

for some constant K. Thus, we find that $q(v)$ is given

$$q(v) = e^{-K} \prod_{x,y,z} \left(\frac{e^{a_x} r(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1}} \right)^{q(x|v)T(y,z|x)},$$

for some constant $K \le F(T,\{a_x\})$. Since there is a choice of $r(y)$ and $q(x|v)$ that attains $K = F(T,\{a_x\})$, we can write

$$\max_{r(y),q(x|v)} \log \left(\sum_v \prod_{x,y,z} \left(\frac{e^{a_x} r(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1}} \right)^{q(x|v)T(y,z|x)} \right) = \max_{r(y),q(x|v)} \log \sum_v e^K q(v) = F(T,\{a_x\}).$$

Equivalently,

$$\max_{r(y)} \log \left(\sum_v \max_x \prod_{y,z} \left(\frac{e^{a_x} r(y)p(v,z)^\lambda}{p(y)^{\bar{\alpha}} p(z)^\alpha p(v)^{\lambda-1}} \right)^{T(y,z|x)} \right) = F(T,\{a_x\}), \tag{9}$$

and equality holds if $r(y) = p(y|u)$ for some u.

Similarly, (8) reduces to

$$\max_{r(z)} \log \left(\sum_u \max_x \prod_{y,z} \left(\frac{e^{a_x} r(z)^\lambda p(u,y)}{p(y)^{\bar{\alpha}} p(z)^\alpha} \right)^{T(y,z|x)} \right) = F(T, \{a_x\}), \tag{10}$$

and equality holds if $r(z) = p(z|v)$ for some v.

B. A local maximizer

We hereby report a local maximizer of

$$G(p(u,v,x)) = -\alpha H(Y) - (\lambda - \alpha) H(Z) + I(U;Y) + \lambda I(V;Z) - I(U;V) + \sum_x p(x) a_x \tag{11}$$

that satisfies (4), (7) and (8), but is not a global maximizer. Let $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \{A, B, C\}$, and

$$\{a_x\} = [0, -0.36832504, -0.13005504], \alpha = 0.53699858, \lambda = 1.$$

In this example, the channel transition matrices $T_{Y|X}$ and $T_{Z|X}$ are:

$T(y\|x)$	$Y = A$	$Y = B$	$Y = C$
$X = A$	0.35332099	0.34718682	0.29949219
$X = B$	0.76824470	0.12006556	0.11168974
$X = C$	0.13810833	0.48234150	0.37955017

$T(z\|x)$	$Z = A$	$Z = B$	$Z = C$
$X = A$	0.44260251	0.21452819	0.34286930
$X = B$	0.40932411	0.00992684	0.58074905
$X = C$	0.27754304	0.56201647	0.16044049

We report the following $p^*(u,v,x)$ is a local maximizer of (11), such that $p^*(x|u,v)$ is of the $\begin{bmatrix} A & B \\ C & A \end{bmatrix}$ pattern,

$f(U,V)$	$V = 0$	$V = 1$
$U = 0$	A	B
$U = 1$	C	A

while the marginal $p^*(u,v)$ is:

$p^*(u,v)$	$V = 0$	$V = 1$
$U = 0$	0.01524814	0.22806511
$U = 1$	0.09087526	0.66581149

The corresponding value of the local maximum is -1.06956384. However, if we fix $p(u,v)$ and change the $\begin{bmatrix} A & B \\ C & A \end{bmatrix}$ mapping to rectangular patterns such as $\begin{bmatrix} A & B \\ C & B \end{bmatrix}$, $\begin{bmatrix} B & B \\ C & A \end{bmatrix}$ or $\begin{bmatrix} C & B \\ C & B \end{bmatrix}$, the expression in (11) would increase to -1.06730708, -1.06895326, and -1.06228037, respectively.

IV. Alternate Characterization for Sum-rate of Marton's Inner bound

The section derives an alternative characterization of

$$F(T, \{a_x\}) = \max_{p(u,v,x)} \left[-\alpha H(Y) - (1-\alpha) H(Z) + I(U;Y) + I(V;Z) - I(U;V) + \sum_x p(x) a_x \right].$$

The idea is similar to the one used in the Blahut-Armioto proof, and similar expressions have been obtained in [3], [14].

Theorem 2. *We have*

$$F(T, \{a_x\}) = \max_{p(u,v,x)} \log \left(\sum_{u,v} \max_x \left[e^{a_x} \prod_y (p(u|y) p(y)^\alpha)^{p(y|x)} \prod_z (p(v|z) p(z)^{\bar{\alpha}})^{p(z|x)} \right] \right) \tag{12}$$

$$= \max_{q(u,y), r(v,z)} \log \left(\sum_{u,v} \max_x \left[e^{a_x} \prod_y (q(u|y) q(y)^\alpha)^{p(y|x)} \prod_z (r(v|z) r(z)^{\bar{\alpha}})^{p(z|x)} \right] \right). \tag{13}$$

where the maximum is over all arbitrary joint distributions $q(u,y)$ and $r(v,z)$.

Proof. We first show that

$$F(T, \{a_x\}) \leq \max_{q(u,y), r(v,z)} \log \left(\sum_{u,v} \max_x \left[e^{a_x} \prod_y (q(u|y) q(y)^\alpha)^{p(y|x)} \prod_z (r(v|z) r(z)^{\bar{\alpha}})^{p(z|x)} \right] \right).$$

Note that for any arbitrary $p(u,v,x)$ we have

$$\begin{aligned} G(p(u,v,x)) &= -\alpha H(Y) - \bar{\alpha} H(Z) + I(U;Y) + I(V;Z) - I(U;V) + \sum_x p(x) a_x \\ &= -\alpha H(Y) - \bar{\alpha} H(Z) - H(U|Y) - H(V|Z) + H(U,V) + \sum_x p(x) a_x \end{aligned}$$

$$
\begin{aligned}
&= \sum_{u,v,x,y,z} p(u,v,x)T(y,z|x)\log\frac{p(u|y)p(v|z)p(y)^{\alpha}p(z)^{\bar{\alpha}}e^{a_x}}{p(u,v)}\\
&= \sum_{u,v} p(u,v)\left(\log\prod_{x,y}(p(u|y)p(y)^{\alpha})^{T(y|x)p(x|u,v)} + \log\prod_{x,z}(p(v|z)p(z)^{\bar{\alpha}})^{T(z|x)p(x|u,v)} + \log\frac{\prod_x (e^{a_x})^{p(x|u,v)}}{p(u,v)}\right)\\
&\le \log\sum_{u,v}\left(\prod_x\left(\prod_{y,z}(p(u|y)p(v|z)p(y)^{\alpha}p(z)^{\bar{\alpha}})^{T(y,z|x)}e^{a_x}\right)^{p(x|u,v)}\right)\\
&\le \log\left(\sum_{u,v}\max_x\prod_{y,z}(p(u|y)p(v|z)p(y)^{\alpha}p(z)^{\bar{\alpha}})^{T(y,z|x)}e^{a_x}\right)\\
&= \log\left(\sum_{u,v}\max_x\left[e^{a_x}\prod_y(p(u|y)p(y)^{\alpha})^{T(y|x)}\prod_z(p(v|z)p(z)^{\bar{\alpha}})^{T(z|x)}\right]\right)\\
&\le \max_{q(u,y),r(v,z)}\log\left(\sum_{u,v}\max_x\left[e^{a_x}\prod_y(q(u|y)q(y)^{\alpha})^{T(y|x)}\prod_z(r(v|z)r(z)^{\bar{\alpha}})^{T(z|x)}\right]\right).
\end{aligned}
$$

Next, we show that for any arbitrary $q(u,y), r(v,z)$ we have

$$
F(T,\{a_x\}) \ge \log\left(\sum_{u,v}\max_x\left[e^{a_x}\prod_y(q(u|y)q(y)^{\alpha})^{p(y|x)}\prod_z(r(v|z)r(z)^{\bar{\alpha}})^{p(z|x)}\right]\right). \quad (14)
$$

This would complete the proof by taking the maximum over $q(u,y), r(v,z)$. Note that for any $p(u,v,x)$ we have

$$
\begin{aligned}
-H(U|Y) &\ge \sum_{u,y} p(u,y)\log q(u|y), \quad -H(Y) \ge \sum_y p(y)\log q(y),\\
-H(V|Z) &\ge \sum_{v,z} p(v,z)\log r(v|z), \quad -H(Z) \ge \sum_z p(z)\log r(z).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
F(T,\{a_x\}) &= \max_{p(u,v,x)} -\alpha H(Y) - \bar{\alpha}H(Z) + I(U;Y) + I(V;Z) - I(U;V) + \sum_x p(x)a_x\\
&= \max_{p(u,v,x)} -\alpha H(Y) - \bar{\alpha}H(Z) + H(U,V) - H(U|Y) - H(V|Z) + \sum_x p(x)a_x\\
&\ge \max_{p(u,v,x)} \sum_{u,v,x,y,z} p(u,v,x)T(y,z|x)\log\frac{q(u|y)r(v|z)q(y)^{\alpha}r(z)^{\bar{\alpha}}}{p(u,v)e^{-a_x}} \quad (15)\\
&= \sum_{u,v} p(u,v)\left(\log\prod_y(p(u|y)p(y)^{\alpha})^{T(y|x)p(x|u,v)} + \log\prod_z(p(v|z)p(z)^{\bar{\alpha}})^{T(z|x)p(x|u,v)} + \log\frac{(e^{a_x})^{p(x|u,v)}}{p(u,v)}\right). \quad (16)
\end{aligned}
$$

For a fixed $p(x|u,v)$, the maximum of the lower bound over $p(u,v)$ (the expression is concave in $p(u,v)$) can be calculated exactly; the maximum occurs at $p(u,v)$ given by

$$
e^{-K}\left[e^{a_x}\prod_y(q(u|y)q(y)^{\alpha})^{T(y|x)}\prod_z(r(v|z)r(z)^{\bar{\alpha}})^{T(z|x)}\right]^{p(x|u,v)}.
$$

If we substitute this value of $p(u,v)$ back into (16), we get that $K \le F(T,\{a_x\})$. Since $\sum_{u,v} p(u,v) = 1$, we obtain

$$
e^K = \sum_{u,v}\left[e^{a_x}\prod_y(q(u|y)q(y)^{\alpha})^{T(y|x)}\prod_z(r(v|z)r(z)^{\bar{\alpha}})^{T(z|x)}\right]^{p(x|u,v)}.
$$

Hence, by taking logarithms and using $K \le F(T,\{a_x\})$,

$$
F(T,\{a_x\}) \ge K = \log\left(\sum_{u,v}\left[e^{a_x}\prod_y(q(u|y)q(y)^{\alpha})^{T(y|x)}\prod_z(r(v|z)r(z)^{\bar{\alpha}})^{T(z|x)}\right]^{p(x|u,v)}\right).
$$

Now taking the maximum over $p(x|u,v)$ on the right-hand-side yields a lower bound in (14) on $F(T,\{a_x\})$. □

Remark 4. In (12), the expression depends on the marginals $p(u,x)$ and $p(v,x)$ only. This decoupling is in a similar spirit as the one implied by the Markovity conjecture.

Remark 5. Note that (13) can be written as

$$
\max_{q(u,y),r(v,z)}\log\left(\sum_{u,v} q(u)r(v)\max_x\left[e^{a_x}\prod_y(q(y|u)q(y)^{-\bar{\alpha}})^{p(y|x)}\prod_z(r(z|v)r(z)^{-\alpha})^{p(z|x)}\right]\right)
$$

which can be compared with (9) and (10).

A. Properties of the maximizer

Take an arbitrary mapping $\mathcal{U}\times\mathcal{V}\to\mathcal{X}$. We can denote the mapping by x_{uv} for every pair (u,v). Therefore, a lower bound for $F(T,\{a_x\})$ (the value of $F(T,\{a_x\})$ is given by maximizing over all mappings) involves the maximization over $q(u,y), r(v,z)$ of

$$\left(\sum_{u,v}\left[e^{a_{x_{u,v}}}\prod_{y,z}(q(u|y)q(y)^{\alpha})^{T(y|x_{u,v})}(r(v|z)r(z)^{\bar{\alpha}})^{T(z|x_{u,v})}\right]\right).$$

Lemma 2. *The above expression is concave in $q(u|y)$. It is also concave in $q(y)$. Similarly, it is concave in $r(z)$. Also, concave in $r(v|z)$.*

Proof. The pmf $q(u|y)$ satisfies $\sum_u q(u|y)=1$. If we relax this constraint and consider the expression for any $q(u|y)\geq 0$, Lemma 3 shows that the expression is jointly concave in $q(u|y)$ since $\sum_y T(y|x_{uv})=1$. The concavity should also hold in the restricted space when we impose $\sum_u q(u|y)=1$. Thus, the expression is concave in $q(u|y)$; it is also concave in $q(y)$; similarly in $r(z)$, and also in $r(v|z)$. □

Lemma 3. *Take some $a_i\geq 0$. Then, the function $f(x_1,x_2,\cdots,x_n)=x_1^{a_1}x_2^{a_2}x_3^{a_3}\cdots x_n^{a_n}$ is jointly concave if $\sum_i a_i\leq 1$.*

Proof. The Hessian is proportional to a matrix whose (i,i) entry is $\frac{a_i(a_i-1)}{x_i^2}$, and whose (i,j) entry for $i\neq j$ is $\frac{a_i a_j}{x_i x_j}$. Thus, we need to show that

$$\sum_i \frac{a_i(a_i-1)}{x_i^2}\beta_i^2+\sum_{i\neq j}\frac{a_i a_j}{x_i x_j}\beta_i\beta_j\leq 0.$$

Equivalently,

$$\sum_i \frac{a_i}{x_i^2}\beta_i^2\geq\left(\sum_i\frac{a_i\beta_i}{x_i}\right)^2.$$

This follows from the Cauchy-Schwartz inequality:

$$\left(\sum_i\frac{a_i}{x_i^2}\beta_i^2\right)\left(\sum_i a_i\right)\geq\left(\sum_i\frac{a_i\beta_i}{x_i}\right)^2.$$

□

V. Discussion and Conclusion

Determining the capacity region of a two-receiver broadcast channel is a central open problem in network information theory. Marton's inner bound is potentially the capacity region, and verifying its optimality or sub-optimality relies heavily on being able to evaluate the optimization problems defining the region. In this paper, based on numerical evidence, we proposed a conjecture regarding the optimizers that greatly simplifies and decouples the optimization problem. Even if one can prove the additivity within this class (without needing to prove the conjecture), one would have shown that Marton's region is the capacity region. Further, we also demonstrated an alternate characterization for the sum rate that establishes a similar decoupling.

Acknowledgment

The authors thank Mr. Chin Wa Lau for obtaining the local maximizer in the subsection III-B. Additionally, the authors appreciate Mr. Chun Hei Shiu's efforts to develop a program to help determine the maximizers of Marton's inner bound numerically.

REFERENCES

[1] V. Anantharam, A. Gohari, and C. Nair, *On the evaluation of Marton's inner bound for two-receiver broadcast channels*, IEEE Transactions on Information Theory **65** (2019), no. 3, 1361–1371.

[2] T. Cover, *Broadcast channels*, IEEE Transactions on Information Theory **18** (1972), no. 1, 2–14.

[3] F. Dupuis, W. Yu, and F.M.J. Willems, *Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information*, International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings., 2004, pp. 179–.

[4] Abbas El Gamal and Young-Han Kim, *Network information theory*, Cambridge University Pres, 2011.

[5] Y. Geng, A. Gohari, C. Nair, and Y. Yu, *On Marton's inner bound and its optimality for classes of product broadcast channels*, IEEE Transactions on Information Theory **60** (2014), no. 1, 22–41.

[6] Y. Geng and C. Nair, *The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages*, IEEE Transactions on Information Theory **60** (2014), no. 4, 2087–2104.

[7] Yanlin Geng, Varun Jog, Chandra Nair, and Zizhou Vincent Wang, *An information inequality and evaluation of Marton's inner bound for binary input broadcast channels*, IEEE Transactions on Information Theory **59** (2013), no. 7, 4095–4105.

[8] A. Gohari, C. Nair, and V. Anantharam, *On Marton's inner bound for broadcast channels*, 2012 IEEE International Symposium on Information Theory, July 2012, pp. 581–585.

[9] A. A. Gohari and V. Anantharam, *Evaluation of Marton's inner bound for the general broadcast channel*, IEEE Transactions on Information Theory **58** (2012), no. 2, 608–619.

[10] A. A. Gohari, A. El Gamal, and V. Anantharam, *On an outer bound and an inner bound for the general broadcast channel*, 2010 IEEE International Symposium on Information Theory, June 2010, pp. 540–544.

[11] Amin Gohari, Abbas El Gamal, and Venkat Anantharam, *On Marton's inner bound for the general broadcast channel*, IEEE Transactions on Information Theory **60** (2014), no. 7, 3748–3762.

[12] Amin Gohari and Chandra Nair, *Outer bounds for multiuser settings: The auxiliary receiver approach*, IEEE Transactions on Information Theory **68** (2022), no. 2, 701–736.

[13] Chin Wa Ken Lau, Chandra Nair, and Chaorui Yao, *Uniqueness of local maximizers for some non-convex log-determinant optimization problems using information theory*, IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022, IEEE, 2022, pp. 432–437.

[14] Yanqing Liu, Yanan Dou, and Yanlin Geng, *Blahut-Arimoto algorithm for marton's inner bound*, 2023 IEEE International Symposium on Information Theory (ISIT), 2023, pp. 2159–2164.

[15] K. Marton, *A coding theorem for the discrete memoryless broadcast channel*, IEEE Transactions on Information Theory **25** (1979), no. 3, 306–311.

[16] C. Nair, H. Kim, and A. El Gamal, *On the optimality of randomized time division and superposition coding for the broadcast channel*, 2016 IEEE Information Theory Workshop (ITW), Sept 2016, pp. 131–135.

[17] Chandra Nair, *On Marton's achievable region: Local tensorization for product channels with a binary component*, 2020 Information Theory and Applications Workshop (ITA), 2020, pp. 1–7.

[18] H. Weingarten, Y. Steinberg, and S. Shamai, *On the capacity region of the multi-antenna broadcast channel with common messages*, Information Theory, 2006 IEEE International Symposium on, july 2006, pp. 2195–2199.

APPENDIX

$$\begin{aligned}
\Gamma_M^\lambda(T) &= \max_{(R_1,R_2)\in\mathcal{R}_M} R_1 + \lambda R_2 \\
&= \max_{p(u,v,w,x)} \min\{I(W;Y), I(W;Z)\} + I(U;Y|W) + I(V;Z|W) - I(U;V|W) + (\lambda-1)I(V,W;Z) \\
&= \max_{p(u,v,w,x)} \min_{\alpha\in[0,1]} \alpha I(W;Y) + (\lambda-\alpha)I(W;Z) + I(U;Y|W) + \lambda I(V;Z|W) - I(U;V|W) \\
&\overset{(a)}{=} \min_{\alpha\in[0,1]} \max_{p(u,v,w,x)} \alpha I(W;Y) + (\lambda-\alpha)I(W;Z) + I(U;Y|W) + \lambda I(V;Z|W) - I(U;V|W) \\
&= \min_{\alpha\in[0,1]} \max_{p(x)} \left[\alpha H(Y) + (\lambda-\alpha)H(Z) + \max_{p(u,v,w|x)} \big\{ -\alpha H(Y|W) - (\lambda-\alpha)H(Z|W) + I(U;Y|W) + \lambda I(V;Z|W) - I(U;V|W)\big\}\right] \\
&= \min_{\alpha\in[0,1]} \max_{p(x)} \left[\alpha H(Y) + (\lambda-\alpha)H(Z) + \mathfrak{C}\big\{ \max_{p(u,v|x)} -\alpha H(Y) - (\lambda-\alpha)H(Z) + I(U;Y) + \lambda I(V;Z) - I(U;V)\big\}\right].
\end{aligned}$$

where $\mathfrak{C}\{\}$ is the upper concave envelope, and (a) follows by a min-max exchange [5]. By convex duality, we have

$$\mathfrak{C}\{\max_{p(u,v|x)} -\alpha H(Y) - (\lambda-\alpha)H(Z) + I(U;Y) + \lambda I(V;Z) - I(U;V)\} = \min_{a_x}\left[F(T,\{a_x\}) - \sum_x p(x)a_x\right].$$

Thus, Marton's inner bound becomes

$$\Gamma_M^\lambda(T) = \min_{\alpha\in[0,1]} \max_{p(x)} \left\{\alpha H(Y) + (\lambda-\alpha)H(Z) + \min_{a_x}\left[F(T,\{a_x\}) - \sum_x p(x)a_x\right]\right\}.$$