ISIT'2013 1569719651

A new extremal entropy inequality with applications

Hon-Fah Chong and Ying-Chang Liang
1 Fusionopolis Way, #21-01 Connexis (South Tower),
Institute for Infocomm Research, Singapore 138632.
E-mail: chong.hon.fah@ieee.org, ycliang@i2r.a-star.edu.sg

***Abstract*—Liu et al. proved an extremal entropy inequality using a vector generalization of the Costa entropy-power inequality (EPI). The generalized Costa EPI was proved, in turn, using a perturbation approach via a fundamental relationship between the derivative of mutual information and the minimum mean-square error (MMSE) estimate in linear vector Gaussian channels. In this paper, we consider two new variations of the (Liu et al.) extremal entropy inequality. Instead of employing perturbation approaches, we employ a new method recently introduced by Geng and Nair, which was used to resolve the capacity region of the Gaussian MIMO broadcast channel (BC) with common and private messages. As an application, we use one of the extremal entropy inequalities to prove the capacity region of a class of reversely degraded Gaussian MIMO BC with three users and three-degraded message sets.**

I. Introduction

In [1], Liu et al. proved the following extremal entropy inequality:

[1, Theorem 2]: Let $\mathbf{Z}_k$, $k=0,\dots,K$, be Gaussian random t-vectors with covariance matrices $\mathbf{N}_k$, respectively. Assume that $\mathbf{N}_1 \preceq \dots \preceq \mathbf{N}_K$. If there exists a $t\times t$ positive semidefinite matrix $\mathbf{Q}^*$ such that

$$\sum_{k=1}^{K}\mu_k\left(\mathbf{Q}^*+\mathbf{N}_k\right)^{-1}+\mathbf{M}=\left(\mathbf{Q}^*+\mathbf{N}_0\right)^{-1}+\mathbf{M}_S \tag{1}$$

for some $t\times t$ positive semidefinite matrices, $\mathbf{M}$, $\mathbf{M}_S$ and $\mathbf{S}$ satisfying

$$\mathbf{Q}^*\mathbf{M}=0 \tag{2}$$
$$\left(\mathbf{S}-\mathbf{Q}^*\right)\mathbf{M}_S=0 \tag{3}$$

and real scalers $\mu_k\geq 0$ with $\sum_{k=1}^{K}\mu_k=1$, then we have

$$\begin{aligned}&\sum_{k=1}^{K}\mu_k h\left(\boldsymbol{X}+\boldsymbol{N}_k|U\right)-h\left(\boldsymbol{X}+\mathbf{Z}_0|U\right)\\&\leq\sum_{k=1}^{K}\frac{\mu_k}{2}\log\left|\mathbf{Q}^*+\mathbf{N}_k\right|-\frac{1}{2}\log\left|\mathbf{Q}^*+\mathbf{N}_0\right|\end{aligned} \tag{4}$$

for any $(\boldsymbol{X},U)$ independent of $\{\mathbf{Z}_k\}_{k=0}^{K}$ such that $\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]\preceq\mathbf{S}$.

Liu et al. first considered the case for $K=2$ assuming that the following equality holds (see [1, Corollary 2]):

$$\mu_1\left(\mathbf{Q}^*+\mathbf{N}_1\right)^{-1}+\mu_2\left(\mathbf{Q}^*+\mathbf{N}_2\right)^{-1}=\left(\mathbf{Q}^*+\mathbf{N}_0\right)^{-1} \tag{5}$$

where $\mu_1+\mu_2=1$ and $\mathbf{N}_1\preceq\mathbf{N}_2$. In fact, Liu et al. showed that (4) holds for $K=2$ as long as (5) is satisfied, without any constraint on the input covariance matrix $\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]\preceq\mathbf{S}$.

We note that the following holds from equality (5):

$$\mathbf{N}_1\preceq\mathbf{N}_0\preceq\mathbf{N}_2. \tag{6}$$

Instead of employing a Gaussian perturbation approach, which would require $\{\mathbf{N}_1,\mathbf{N}_2\}\preceq\mathbf{N}_0$ (see proof of [2, Lemma 2]), Liu et al. first proved a vector generalization of Costa's EPI [1, Theorem 1] using a different perturbation approach (via a fundamental relationship between the derivative of mutual information and the MMSE estimate in linear vector Gaussian channels). They then employed the generalized Costa's EPI to prove [1, Corollary 2] for $K=2$. Finally, by using induction as well as enhancement, they proved the extremal entropy inequality in [1, Theorem 2] for an arbitrary K.

In this paper, we consider two variations of the extremal entropy inequality given by [1, Theorem 2] for the case where $K=2$. We assume that either of the following holds:

$$\mathbf{N}_1\preceq\{\mathbf{N}_0,\mathbf{N}_2\} \tag{7}$$
$$\{\mathbf{N}_0,\mathbf{N}_1\}\preceq\mathbf{N}_2. \tag{8}$$

We wish to prove the following extremal entropy inequality:

***Theorem** 1:* Let $\mathbf{Z}_k$, $k=0,1,2$, be Gaussian random t-vectors with covariance matrices $\mathbf{N}_k$, respectively. The covariance matrices $\mathbf{N}_k$ satisfy either (7) or (8). Then the following holds:

$$\begin{aligned}&\sup_{\substack{\mathbb{E}[\boldsymbol{X}\boldsymbol{X}]\preceq\mathbf{S}\\U\to\boldsymbol{X}\to\{\boldsymbol{X}+\mathbf{Z}_k\}_{k=0}^{k=2}}}\begin{array}{l}\beta_1 h\left(\boldsymbol{X}+\mathbf{Z}_1|U\right)+\beta_2 h\left(\boldsymbol{X}+\mathbf{Z}_2|U\right)\\-h\left(\boldsymbol{X}+\mathbf{Z}_0|U\right)\end{array}\\&=\frac{\beta_1}{2}\log\left|\mathbf{Q}^*+\mathbf{N}_1\right|+\frac{\beta_2}{2}\log\left|\mathbf{Q}^*+\mathbf{N}_2\right|-\frac{1}{2}\log\left|\mathbf{Q}^*+\mathbf{N}_0\right|\end{aligned} \tag{9}$$

where $\beta_1,\beta_2\geq 0$, $\beta_1+\beta_2\leq 1$ and for some $0\preceq\mathbf{Q}^*\preceq\mathbf{S}$.

***Remark** 1:* A few comments are in order here. Comparing Theorem 1 and [1, Corollary 2], we note that both conditions (7) and (8) satisfy (6). Moreover, Theorem 1 assumes that $\beta_1+\beta_2\leq 1$, rather than the stricter condition that $\mu_1+\mu_2=1$ holds in [1, Corollary 2]. On the other hand, Liu et al. ([1, Corollary 2]) showed that (4) holds for $K=2$ as long as (5) is satisfied, without any constraint on the input covariance matrix $\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right]\preceq\mathbf{S}$.

***Remark** 2:* Comparing Theorem 1 and [1, Theorem 2] (for $K=2$), we note that the assumption on the covariance matrices $\mathbf{N}_1\preceq\mathbf{N}_2$ in [1, Theorem 2] is more general than that assumed by Theorem 1 (see (7) and (8)). On the hand, we note again that Theorem 1 assumes that $\beta_1 + \beta_2 \leq 1$, rather than the stricter condition that $\mu_1 + \mu_2 = 1$ holds in [1, Theorem 2].

The proof of [1, Theorem 2] and [1, Corollary 2], employing the generalized Costa EPI, cannot be extended to prove Theorem 1. Instead, we employ a new technique recently introduced by Geng and Nair in [3], which was used to prove the capacity region of the two-user Gaussian MIMO BC with common and private messages. On another note, our proof cannot be extended to prove [1, Theorem 2] (or [1, Corollary 2]) as well.

The paper is organized as follows: In Section II, we employ the Geng-Nair methodology to prove Theorem 1. As an application, we employ Theorem 1 to prove the capacity region of a class of reversely degraded Gaussian MIMO BC with three users and three-degraded message sets in Section III.

II. Proof of Theorem 1

We will first prove the case for (7) in detail as the proof of the case for (8) follows along the same lines with only a slight variation. Without loss of generality, we assume that the inequalities in (7) and (8) are strict. Let us first consider the following vector additive Gaussian channels:

$$\begin{aligned} \boldsymbol{Y}_1 &= \boldsymbol{X} + \boldsymbol{Z}_1, \\ \boldsymbol{Y}_2 &= \boldsymbol{Y}_1 + \overline{\boldsymbol{Z}}_2, \\ \boldsymbol{Y}_0 &= \boldsymbol{Y}_1 + \overline{\boldsymbol{N}}_0, \end{aligned} \tag{10}$$

where $\boldsymbol{Z}_1 \sim \mathcal{N}(0, \mathbf{N}_1)$, $\overline{\boldsymbol{Z}}_2 \sim \mathcal{N}\left(0, \overline{\mathbf{N}}_2 = \mathbf{N}_2 - \mathbf{N}_1\right)$, $\overline{\boldsymbol{Z}}_0 \sim \mathcal{N}\left(0, \overline{\mathbf{N}}_0 = \mathbf{N}_0 - \mathbf{N}_1\right)$. We note that channels $\boldsymbol{Y}_2$ and $\boldsymbol{Y}_0$ are degraded versions of channel $\boldsymbol{Y}_1$, which follows from (7). However, we do not assume any order of degradedness between channel $\boldsymbol{Y}_0$ and channel $\boldsymbol{Y}_2$.

We note that Theorem 1 is equivalent to the following:

$$\begin{aligned} &\sup_{\substack{\mathbb{E}[\boldsymbol{X}\boldsymbol{X}] \preceq \mathbf{S} \\ U \to \boldsymbol{X} \to \{\boldsymbol{X}+\boldsymbol{Z}_k\}_{k=0}^{k=2}}} \begin{array}{l} \beta_1 I(\boldsymbol{X}; \boldsymbol{X}+\boldsymbol{Z}_1|U) + \beta_2 I(\boldsymbol{X}; \boldsymbol{X}+\boldsymbol{Z}_2|U) \\ -I(\boldsymbol{X}; \boldsymbol{X}+\boldsymbol{Z}_0|U) \end{array} \\ &= \frac{\beta_1}{2}\log\frac{|\mathbf{Q}^* + \mathbf{N}_1|}{|\mathbf{N}_1|} + \frac{\beta_2}{2}\log\frac{|\mathbf{Q}^* + \mathbf{N}_2|}{|\mathbf{N}_2|} - \frac{1}{2}\log\frac{|\mathbf{Q}^* + \mathbf{N}_0|}{|\mathbf{N}_0|} \end{aligned} \tag{11}$$

for some $0 \preceq \mathbf{Q}^* \preceq \mathbf{S}$.

A. Mathematical Preliminaries

As we shall rely on some properties of Fisher information matrices as well as some properties of product BC, we shall first provide some definitions as well as some previously known results.

***Definition** 1:* Let $\boldsymbol{Y}$ be a random vector where the probability density function (pdf) $f(\boldsymbol{y})$ is differentiable. The Fisher information matrix of $\boldsymbol{Y}$ is then defined as

$$\mathbf{J}(\boldsymbol{Y}) = \mathbb{E}\left[(\nabla \log f(\boldsymbol{y}))(\nabla \log f(\boldsymbol{y}))^T\right].$$

***Definition** 2:* The product channels of (10) can be represented as follows:

$$\begin{bmatrix} \boldsymbol{Y}_{i,1} \\ \boldsymbol{Y}_{i,2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{i,1} \\ \boldsymbol{Z}_{i,2} \end{bmatrix}, \quad i = 0, 1, 2 \tag{12}$$

where $\boldsymbol{Z}_{i,1}$ and $\boldsymbol{Z}_{i,2}$, $i = 0, 1, 2$, are identical and independent Gaussian noise vectors and where $\boldsymbol{Z}_{1,j} \sim \mathcal{N}(0, \mathbf{N}_1)$, $\boldsymbol{Z}_{2,j} = \boldsymbol{Z}_{1,j} + \overline{\boldsymbol{Z}}_{2,j}$, where $\overline{\boldsymbol{Z}}_{2,j} \sim \mathcal{N}(0, \mathbf{N}_2 - \mathbf{N}_1)$ and $\boldsymbol{Z}_{0,j} = \boldsymbol{Z}_{1,j} + \overline{\boldsymbol{Z}}_{0,j}$, where $\overline{\boldsymbol{Z}}_{0,j} \sim \mathcal{N}(0, \mathbf{N}_0 - \mathbf{N}_1)$, $j = 1, 2$.

***Proposition** 1:* Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be zero-mean independent t-dimensional random vectors. If $\boldsymbol{X}_1 + \boldsymbol{X}_2$ and $\boldsymbol{X}_1 - \boldsymbol{X}_2$ are independent, then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent Gaussian random vectors with identical covariance matrices.

Proof: See proof of [3, Corollary 3]. ■

***Proposition** 2:* Consider the following vector additive Gaussian product channel with identical components:

$$\begin{aligned} \boldsymbol{Y}_1 &= \boldsymbol{X}_1 + \boldsymbol{N}_1 \\ \boldsymbol{Y}_2 &= \boldsymbol{X}_2 + \boldsymbol{N}_2 \end{aligned}$$

where $\boldsymbol{N}_1 \sim \mathcal{N}(0, \mathbf{N}_G)$ and $\boldsymbol{N}_2 \sim \mathcal{N}(0, \mathbf{N}_G)$ are independent. Let $\tilde{\boldsymbol{X}} = \frac{1}{\sqrt{2}}(\boldsymbol{X}_1 + \boldsymbol{X}_2)$, $\hat{\boldsymbol{X}} = \frac{1}{\sqrt{2}}(\boldsymbol{X}_1 - \boldsymbol{X}_2)$, $\tilde{\boldsymbol{Y}} = \frac{1}{\sqrt{2}}(\boldsymbol{Y}_1 + \boldsymbol{Y}_2)$, and $\hat{\boldsymbol{Y}} = \frac{1}{\sqrt{2}}(\boldsymbol{Y}_1 - \boldsymbol{Y}_2)$, we must have

$$I\left(\tilde{\boldsymbol{X}}, \hat{\boldsymbol{X}}; \tilde{\boldsymbol{Y}}, \hat{\boldsymbol{Y}}\right) = I(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_1, \boldsymbol{Y}_2).$$

Proof: See proof of [3, Claim 1]. ■

***Proposition** 3:* Consider the vector additive Gaussian product channels given by (12). The random vectors $\boldsymbol{Y}_{i,1}$ and $\boldsymbol{Y}_{j,2}$, $i, j \in \{0, 1, 2\}$, are independent if and only if $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent.

Proof: See proof of [3, Claim 2]. ■

The above propositions were used to prove the capacity region of the two-user Gaussian MIMO BC with private and common messages in [3] and will also be useful later. Next, we present some new propositions that are necessary in our present context.

***Proposition** 4:* $\operatorname{tr}(\mathbf{J}(\boldsymbol{X}))$ is a strictly convex functional in $f(\boldsymbol{x})$.

Proof: Convexity follows from [4, Prop. 11]. Strict convexity for the scalar case was proven in [5, Theorem]. Strict convexity for the vector case follows rather straightforwardly from the proof of both [4, Prop. 11] and [5, Theorem]. ■

***Proposition** 5:* For the vector additive Gaussian product channels given by (12), $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent if and only if

$$I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1}) - I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1}) = 0. \tag{13}$$

Proof: We consider the non-trivial direction. We must have $I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1}) \geq I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1})$ since $\boldsymbol{Y}_{0,2}$ is a degraded version of $\boldsymbol{Y}_{1,2}$. There are two possible cases:

1) $I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1}) = I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1}) > 0$.
2) $I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1}) = I(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1}) = 0$.

We first show that Case 1) is impossible.

Let us consider the following degraded version of $\boldsymbol{Y}_{1,2}$:

$$\boldsymbol{Y}_{1,2}^{\delta} = \boldsymbol{Y}_{1,2} + \sqrt{\delta}\boldsymbol{N}_G^I$$

where $\boldsymbol{N}_G^I \sim \mathcal{N}(0, \mathbf{I}_t)$. From the vector De Bruijn's identity (see [4, 41(a)]), we have

$$\frac{d}{d\delta} h\left(\boldsymbol{Y}_{1,2}^{\delta}\right) = \frac{1}{2} \mathrm{tr}\left(\mathbf{J}\left(\boldsymbol{Y}_{1,2}^{\delta}\right)\right).$$

Next, we note that we may exchange the order of differentiation and integration (see [6, (42)-(45)]) to obtain:

$$\frac{d}{d\delta} h\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{2,1}\right) = \frac{1}{2} \mathrm{tr}\left[\mathbf{J}\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{2,1}\right)\right], \tag{14}$$

$$\frac{d}{d\delta} h\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{1,1}\right) = \frac{1}{2} \mathrm{tr}\left[\mathbf{J}\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{1,1}\right)\right]. \tag{15}$$

From (14) and (15), we obtain

$$\begin{aligned} \frac{d}{d\delta} I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{2,1}\right) = & \frac{1}{2} \mathrm{tr}\left(\mathbf{J}\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{2,1}\right)\right) \\ & - \frac{1}{2} \mathrm{tr}\left(\mathbf{J}\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{1,1}\right)\right). \end{aligned} \tag{16}$$

Since we assume that $I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2} | \boldsymbol{Y}_{2,1}\right) > 0$, we must have $f\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{2,1}\right) \neq f\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{1,1}, \boldsymbol{y}_{2,1}\right)$ for some $\left(\boldsymbol{y}_{1,1}, \boldsymbol{y}_{2,1}, \boldsymbol{y}_{1,2}^{\delta}\right)$. Hence, we have

$$\begin{aligned} & \mathrm{tr}\left(\mathbf{J}\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{1,1}\right)\right) \\ & = \int\int \frac{\sum_{l=1}^{t} \left| f_l'\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{1,1}\right)\right|^2}{f\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{1,1}\right)} d\boldsymbol{y}_{1,2}^{\delta} f\left(\boldsymbol{y}_{1,1}\right) d\boldsymbol{y}_{1,1} \\ & \overset{(a)}{=} \int\int \frac{\sum_{l=1}^{t} \left| f_l'\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{1,1}\right)\right|^2}{f\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{1,1}\right)} f\left(\boldsymbol{y}_{1,1}\right) d\boldsymbol{y}_{1,1} d\boldsymbol{y}_{1,2}^{\delta} \\ & \overset{(b)}{=} \int\int\int \frac{\sum_{l=1}^{t} \left| f_l'\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{1,1}, \boldsymbol{y}_{2,1}\right)\right|^2}{f\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{1,1}, \boldsymbol{y}_{2,1}\right)} \\ & \qquad \times f\left(\boldsymbol{y}_{1,1}, \boldsymbol{y}_{2,1}\right) d\boldsymbol{y}_{2,1} d\boldsymbol{y}_{1,1} d\boldsymbol{y}_{1,2}^{\delta} \\ & \overset{(c)}{>} \int\int \frac{\sum_{l=1}^{t} \left| f_l'\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{2,1}\right)\right|^2}{f\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{2,1}\right)} f\left(\boldsymbol{y}_{2,1}\right) d\boldsymbol{y}_{2,1} d\boldsymbol{y}_{1,2}^{\delta} \\ & \overset{(a)}{=} \int\int \frac{\sum_{l=1}^{t} \left| f_l'\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{2,1}\right)\right|^2}{f\left(\boldsymbol{y}_{1,2}^{\delta} | \boldsymbol{y}_{2,1}\right)} d\boldsymbol{y}_{1,2}^{\delta} f\left(\boldsymbol{y}_{2,1}\right) d\boldsymbol{y}_{2,1} \\ & = \mathrm{tr}\left(\mathbf{J}\left(\boldsymbol{Y}_{1,2}^{\delta} | \boldsymbol{Y}_{2,1}\right)\right) \end{aligned}$$

where (a) follows from Tonelli's theorem, (b) follows from the fact that $\boldsymbol{Y}_{1,2}^{\delta} \to \boldsymbol{Y}_{1,1} \to \boldsymbol{Y}_{2,1}$ forms a Markov chain and (c) follows from Proposition 4 and Jensen's inequality for strictly convex functions. Hence,

$$I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}^{\delta} = \boldsymbol{Y}_{1,2} + \sqrt{\delta} \boldsymbol{N}_G^I | \boldsymbol{Y}_{2,1}\right)$$

is a strictly decreasing function of δ if $I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2} | \boldsymbol{Y}_{2,1}\right) > 0$.

Therefore, we must have

$$\begin{aligned} I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2} | \boldsymbol{Y}_{2,1}\right) & > I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2} + \sqrt{\delta} \boldsymbol{N}_G | \boldsymbol{Y}_{2,1}\right) \\ & \geq I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2} | \boldsymbol{Y}_{2,1}\right) \end{aligned}$$

for some $\delta > 0$. This is a contradiction of the original assumption that $I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2} | \boldsymbol{Y}_{2,1}\right) = I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2} | \boldsymbol{Y}_{2,1}\right)$ and hence, the impossibility of Case 1).

The second case follows from the proof of the first case as we note that

$$I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2} | \boldsymbol{Y}_{2,1}\right) = 0 \Rightarrow I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}\right) - I\left(\boldsymbol{Y}_{2,1}; \boldsymbol{Y}_{1,2}\right) = 0.$$

Again there are two possible cases:

1) $I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}\right) = I\left(\boldsymbol{Y}_{2,1}; \boldsymbol{Y}_{1,2}\right) > 0$,
2) $I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2}\right) = I\left(\boldsymbol{Y}_{2,1}; \boldsymbol{Y}_{1,2}\right) = 0$.

The first case above follows along the exact same lines as the proof for the case where

$$I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{1,2} | \boldsymbol{Y}_{2,1}\right) = I\left(\boldsymbol{Y}_{1,1}; \boldsymbol{Y}_{0,2} | \boldsymbol{Y}_{2,1}\right) > 0.$$

For the second case, the result follows readily from Proposition 3. ■

Next, we define the following functions for the vector additive Gaussian channels (10):

***Definition** 3:* We define the following function of $P_{\boldsymbol{X}}$:

$$s_{\vec{\beta}}(\boldsymbol{X}) \triangleq \beta_1 I\left(\boldsymbol{X}; \boldsymbol{Y}_1\right) + \beta_2 I\left(\boldsymbol{X}; \boldsymbol{Y}_2\right) - I\left(\boldsymbol{X}; \boldsymbol{Y}_0\right). \tag{17}$$

We define the following function of $P_{U,\boldsymbol{X}}$:

$$s_{\vec{\beta}}(\boldsymbol{X}|U) \triangleq \beta_1 I\left(\boldsymbol{X}; \boldsymbol{Y}_1 | U\right) + \beta_2 I\left(\boldsymbol{X}; \boldsymbol{Y}_2 | U\right) - I\left(\boldsymbol{X}; \boldsymbol{Y}_0 | U\right). \tag{18}$$

We define the following function of $P_{\boldsymbol{X}}$:

$$S_{\vec{\beta}}(\boldsymbol{X}) \triangleq \sup_{P_{U|\boldsymbol{X}}} s_{\vec{\beta}}(\boldsymbol{X}|U). \tag{19}$$

We also define the following function of $P_{Q,\boldsymbol{X}}$:

$$S_{\vec{\beta}}(\boldsymbol{X}|Q) \triangleq \mathbb{E}_Q\left[S_{\vec{\beta}}(\boldsymbol{X}|Q=q)\right]. \tag{20}$$

For $\mathbf{K} \succeq 0$, define

$$G_{\vec{\beta}}(\mathbf{K}) = \sup_{\boldsymbol{X}: \mathbb{E}(\boldsymbol{X}\boldsymbol{X}^T) \preceq \mathbf{K}} S_{\vec{\beta}}(\boldsymbol{X}). \tag{21}$$

We also denote the extensions of (17)-(19) to the vector additive Gaussian product channels (12) by $s_{\vec{\beta}}(\boldsymbol{X}_1, \boldsymbol{X}_2)$, $s_{\vec{\beta}}(\boldsymbol{X}_1, \boldsymbol{X}_2 | U)$ and $S_{\vec{\beta}}(\boldsymbol{X}_1, \boldsymbol{X}_2)$, respectively.

***Proposition** 6:* The following inequality holds for the vector additive Gaussian product channels of (12):

$$S_{\vec{\beta}}(\boldsymbol{X}_1, \boldsymbol{X}_2) \leq S_{\vec{\beta}}(\boldsymbol{X}_1) + S_{\vec{\beta}}(\boldsymbol{X}_2).$$

If $P_{U|\boldsymbol{X}_1,\boldsymbol{X}_2}$ attains $S_{\vec{\beta}}(\boldsymbol{X}_1, \boldsymbol{X}_2)$ and equality holds above, then all the following must be true:1) $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are conditionally independent given U; 2) $(U, \boldsymbol{X}_1)$ attains $S_{\vec{\beta}}(\boldsymbol{X}_1)$ and 3) $(U, \boldsymbol{X}_2)$ attains $S_{\vec{\beta}}(\boldsymbol{X}_2)$.

Proof: For any $P_{U|\boldsymbol{X}_1,\boldsymbol{X}_2}$, we observe the following:

$$\begin{aligned} & \beta_1 I\left(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_{1,1}, \boldsymbol{Y}_{1,2} | U\right) + \beta_2 I\left(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_{2,1}, \boldsymbol{Y}_{2,2} | U\right) \\ & - I\left(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{Y}_{0,1}, \boldsymbol{Y}_{0,2} | U\right) \\ & = \beta_1 I\left(\boldsymbol{X}_1; \boldsymbol{Y}_{1,1} | U\right) + \beta_2 I\left(\boldsymbol{X}_1; \boldsymbol{Y}_{2,1} | U\right) - I\left(\boldsymbol{X}_1; \boldsymbol{Y}_{0,1} | \boldsymbol{Y}_{0,2}, U\right) \\ & \quad + \beta_1 I\left(\boldsymbol{X}_2; \boldsymbol{Y}_{1,2} | \boldsymbol{Y}_{1,1}, U\right) + \beta_2 I\left(\boldsymbol{X}_2; \boldsymbol{Y}_{2,2} | \boldsymbol{Y}_{2,1}, U\right) \\ & \quad - I\left(\boldsymbol{X}_2; \boldsymbol{Y}_{0,2} | U\right) \\ & = \beta_1 I\left(\boldsymbol{X}_1; \boldsymbol{Y}_{1,1} | \boldsymbol{Y}_{0,2}, U\right) + \beta_2 I\left(\boldsymbol{X}_1; \boldsymbol{Y}_{2,1} | \boldsymbol{Y}_{0,2}, U\right) \end{aligned}$$

$$
\begin{aligned}
&- I\left(\boldsymbol{X}_1;\boldsymbol{Y}_{0,1}|\boldsymbol{Y}_{0,2},U\right)\\
&+ \beta_1 I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1},U\right) + \beta_2 I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{2,2}|\boldsymbol{Y}_{2,1},U\right)\\
&- I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1},U\right)\\
&+ \beta_1 I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{0,2}|U\right) - \beta_1 I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1},U\right)\\
&+ \beta_2 I\left(\boldsymbol{Y}_{0,2};\boldsymbol{Y}_{2,1}|U\right) - I\left(\boldsymbol{Y}_{2,1};\boldsymbol{Y}_{0,2}|U\right)\\
\le\ & S_{\vec{\beta}}\left(\boldsymbol{X}_1|\boldsymbol{Y}_{0,2},U\right) + S_{\vec{\beta}}\left(\boldsymbol{X}_2|\boldsymbol{Y}_{2,1},U\right)\\
&- \beta_1\left[I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1},U\right) - I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1},U\right)\right]\\
&- \left(1-\beta_1-\beta_2\right) I\left(\boldsymbol{Y}_{2,1};\boldsymbol{Y}_{0,2}|U\right)\\
\overset{(a)}{\le}\ & S_{\vec{\beta}}\left(\boldsymbol{X}_1\right) + S_{\vec{\beta}}\left(\boldsymbol{X}_2\right)
\end{aligned}
$$

where (a) follows from the fact that $\beta_1+\beta_2 \le 1$, the fact that $\boldsymbol{Y}_{0,2}$ is a degraded version of $\boldsymbol{Y}_{1,2}$ and from the definition of $S_{\vec{\beta}}\left(\boldsymbol{X}\right)$.

When equality holds, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ must be conditionally independent given U (from Proposition 5. We also have

$$
\begin{aligned}
&\beta_1 I\left(\boldsymbol{X}_1;\boldsymbol{Y}_{1,1}|\boldsymbol{Y}_{0,2},U\right) + \beta_2 I\left(\boldsymbol{X}_1;\boldsymbol{Y}_{2,1}|\boldsymbol{Y}_{0,2},U\right)\\
&- I\left(\boldsymbol{X}_1;\boldsymbol{Y}_{0,1}|\boldsymbol{Y}_{0,2},U\right) = S_{\vec{\beta}}\left(\boldsymbol{X}_1\right).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&\beta_1 I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{2,1},U\right) + \beta_2 I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{2,2}|\boldsymbol{Y}_{2,1},U\right)\\
&- I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1},U\right) = S_{\vec{\beta}}\left(\boldsymbol{X}_2\right).
\end{aligned}
$$

∎

***Remark** 3:* For the case (8), we may show that the following holds:

$$
\begin{aligned}
&\beta_1 I\left(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}_{1,1},\boldsymbol{Y}_{1,2}|U\right) + \beta_2 I\left(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}_{2,1},\boldsymbol{Y}_{2,2}|U\right)\\
&- I\left(\boldsymbol{X}_1,\boldsymbol{X}_2;\boldsymbol{Y}_{0,1},\boldsymbol{Y}_{0,2}|U\right)\\
&= \beta_1 I\left(\boldsymbol{X}_1;\boldsymbol{Y}_{1,1}|\boldsymbol{Y}_{0,2},U\right) + \beta_2 I\left(\boldsymbol{X}_1;\boldsymbol{Y}_{2,1}|\boldsymbol{Y}_{0,2},U\right)\\
&\quad - I\left(\boldsymbol{X}_1;\boldsymbol{Y}_{0,1}|\boldsymbol{Y}_{0,2},U\right)\\
&\quad + \beta_1 I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{1,2}|\boldsymbol{Y}_{1,1},U\right) + \beta_2 I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{2,2}|\boldsymbol{Y}_{1,1},U\right)\\
&\quad - I\left(\boldsymbol{X}_2;\boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{1,1},U\right)\\
&\quad + \beta_1 I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{0,2}|U\right) + \beta_2 I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{2,2}|\boldsymbol{Y}_{2,1},U\right)\\
&\quad + \beta_2 I\left(\boldsymbol{Y}_{0,2};\boldsymbol{Y}_{2,1}|U\right) - I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{0,2}|U\right)\\
&\overset{(a)}{\le} S_{\vec{\beta}}\left(\boldsymbol{X}_1|\boldsymbol{Y}_{0,2},U\right) + S_{\vec{\beta}}\left(\boldsymbol{X}_2|\boldsymbol{Y}_{1,1},U\right)\\
&\quad - \beta_2\left[I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{0,2}|\boldsymbol{Y}_{2,1},U\right) - I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{2,2}|\boldsymbol{Y}_{2,1},U\right)\right]\\
&\quad - \left(1-\beta_1-\beta_2\right) I\left(\boldsymbol{Y}_{1,1};\boldsymbol{Y}_{0,2}|U\right)\\
&\overset{(a)}{\le} S_{\vec{\beta}}\left(\boldsymbol{X}_1\right) + S_{\vec{\beta}}\left(\boldsymbol{X}_2\right)
\end{aligned}
$$

where (a) follows from the fact that for (8), $\boldsymbol{Y}_{2,j}$ is stochastically degraded with respect to both $\boldsymbol{Y}_{0,j}$ and $\boldsymbol{Y}_{1,j}$, $j=1,2$. The rest of the proof for (8) follows along the same lines as the proof for (7).

The following proposition shows that the supremum is attained:

***Proposition** 7:* There is a pair of random variables $(U_*, \boldsymbol{X}_*)$ with $|U_*| \le \frac{t(t+1)}{2}+1$ such that

$$
G_{\vec{\beta}}\left(\mathbf{S}\right) = s_{\vec{\beta}}\left(\boldsymbol{X}_*|U_*\right).
$$

Proof: The result follows using the same arguments from the proof of [3, Claim 6]. ∎

We assume that $P_{U_*,\boldsymbol{X}_*}$ attains $G_{\vec{\beta}}\left(\mathbf{S}\right)$. Let us also assume that $|\mathcal{U}| = M \le \frac{t(t+1)}{2}+1$, $\boldsymbol{X}_u$ is a zero-mean r.v. distributed according to $P_{\boldsymbol{X}_*|U_*=u_*}$, and $\mathbf{S}_u = \mathbb{E}\left(\boldsymbol{X}_u\boldsymbol{X}_u^T\right)$. Then we have $\sum_{u=1}^{M} p_*\left(u\right)\mathbf{S}_u \preceq \mathbf{S}$ and all the $\mathbf{S}_u$'s are bounded.

***Proposition** 8:* Let $(U_1,U_2,\boldsymbol{X}_1,\boldsymbol{X}_2) \sim P_{U_1,\boldsymbol{X}_1,U_2,\boldsymbol{X}_2} = P_{U_1,\boldsymbol{X}_1}P_{U_2,\boldsymbol{X}_2}$ be two i.i.d copies of $P_{U_*,\boldsymbol{X}_*}$. Let

$$
\begin{aligned}
\tilde{U} = \left(U_1,U_2\right), \quad \tilde{\boldsymbol{X}}_{u_1,u_2} &\sim \frac{1}{\sqrt{2}}\left(\boldsymbol{X}_{u_1}+\boldsymbol{X}_{u_2}\right),\\
\hat{\boldsymbol{X}}_{u_1,u_2} &\sim \frac{1}{\sqrt{2}}\left(\boldsymbol{X}_{u_1}-\boldsymbol{X}_{u_2}\right),
\end{aligned}
$$

then the following holds:1) $\tilde{\boldsymbol{X}}$ and $\hat{\boldsymbol{X}}$ are conditionally independent given $\tilde{U}$; 2) $\tilde{U}$, $\tilde{\boldsymbol{X}}$ attains $G_{\vec{\beta}}\left(\mathbf{S}\right)$; 3) $\tilde{U}$, $\hat{\boldsymbol{X}}$ attains $G_{\vec{\beta}}\left(\mathbf{S}\right)$.

Proof: The result follows using the same arguments from the proof of [3, Claim 7]. ∎

Finally, we note that $\boldsymbol{X}_{u_1}$ and $\boldsymbol{X}_{u_2}$ are independent random vectors and $\boldsymbol{X}_{u_1}+\boldsymbol{X}_{u_2}$, $\boldsymbol{X}_{u_1}-\boldsymbol{X}_{u_2}$ are also independent random vectors. Thus, from Proposition 1, $\boldsymbol{X}_{u_1}$ and $\boldsymbol{X}_{u_2}$ are Gaussian random vectors having the same distribution $\boldsymbol{X}^* \sim \mathcal{N}\left(0,\mathbf{Q}^*\right)$. Since this holds for all u_1, u_2, all $\boldsymbol{X}_{u_1}, \boldsymbol{X}_{u_2} \sim \mathcal{N}\left(0,\mathbf{Q}^*\right)$. We then have

$$
\begin{aligned}
G_{\vec{\alpha},\vec{\beta}}\left(\mathbf{S}\right) &= \sum_{u_1} p\left(u_1\right) s_{\vec{\beta}}\left(\boldsymbol{X}_{u_1}\right)\\
&= s_{\vec{\beta}}\left(\boldsymbol{X}^*\right)\sum_{u_1} p\left(u_1\right) = s_{\vec{\beta}}\left(\boldsymbol{X}^*\right).
\end{aligned}
$$

***Remark** 4:* We note the uniqueness of $\boldsymbol{X}^* \sim \mathcal{N}\left(0,\mathbf{Q}^*\right)$ from above as well as the proof of Proposition 8. We may choose $U \sim \mathcal{N}\left(0,\mathbf{S}-\mathbf{Q}^*\right)$, and $\boldsymbol{X} = U + \boldsymbol{R}$, where $\boldsymbol{R} \sim \mathcal{N}\left(0,\mathbf{Q}^*\right)$, in Theorem 1.

III. Capacity region of a class of reversely degraded Gaussian MIMO BC with three users and three-degraded message sets

In this section, we consider an application of Theorem 1 to determine the capacity region of a class of reversely degraded Gaussian MIMO BC with three users (W (receiver 1), Y (receiver 2) and Z (receiver 3)) and three-degraded message sets (M_1, M_2, M_3), where the message M_1 is intended for all three receivers, the message M_2 is intended for receivers 2 and 3 and the message M_3 is intended only for receiver 3.

We consider the product of two reversely degraded Gaussian BCs such that

$$
\begin{aligned}
\boldsymbol{W}_i &= \boldsymbol{X}_i + \boldsymbol{N}_i^{(w)}, \quad 1 \le i \le 2,\\
\boldsymbol{Y}_i &= \boldsymbol{X}_i + \boldsymbol{N}_i^{(y)}, \quad 1 \le i \le 2,\\
\mathbf{Z}_i &= \boldsymbol{X}_i + \boldsymbol{N}_i^{(z)}, \quad 1 \le i \le 2.
\end{aligned}
\tag{22}
$$

where $\boldsymbol{X}_i$ is a real input vector of size $t \times 1$ corresponding to subchannel i; $\boldsymbol{W}_i$ is a real output vector of size $t \times 1$ corresponding to subchannel i of user $\boldsymbol{W}$, $\boldsymbol{Y}_i$ is a real output vector of size $t \times 1$ corresponding to subchannel i of user $\boldsymbol{Y}$, and $\mathbf{Z}_i$ is a real output vector of size $t \times 1$ corresponding to subchannel i of user $\boldsymbol{Z}$; $\boldsymbol{N}_i^{(w)}$ is a real Gaussian noise vector with zero mean and covariance matrix $\mathbf{N}_i^{(w)}$, $\boldsymbol{N}_i^{(y)}$ is a real Gaussian noise vector with zero mean and covariance matrix $\mathbf{N}_i^{(y)}$, and $\boldsymbol{N}_i^{(z)}$ is a real Gaussian noise vector with zero mean and covariance matrix $\mathbf{N}_i^{(z)}$. Furthermore, the covariance matrices of user $\boldsymbol{W}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ satisfy the following:

$$\mathbf{N}_1^{(y)} \prec \mathbf{N}_1^{(z)} \prec \mathbf{N}_1^{(w)} \tag{23}$$

$$\mathbf{N}_2^{(w)} \prec \mathbf{N}_2^{(z)} \prec \mathbf{N}_2^{(y)}. \tag{24}$$

We consider a separate matrix constraint on the input for each subchannel, i.e., $\mathbb{E}\left[\boldsymbol{X}_i\boldsymbol{X}_i\right] \preceq \mathbf{S}_i$, $1 \leq i \leq 2$, where $\mathbf{S}_i \succeq 0$. We also assume that the inequalities in (23)-(24) are strict. Since the set of all covariance matrices are dense, by the continuity of the log det function over positive definite matrices, our result also hold for the case when the inequalities are not strict.

***Theorem** 2:* Subject to the following input matrix constraints:

$$\mathbb{E}\left[\boldsymbol{X}_i\boldsymbol{X}_i^T\right] \preceq \mathbf{S}_i, \quad 1 \leq i \leq 2, \tag{25}$$

the capacity region $\mathcal{C}\left(\mathbf{S}_1, \mathbf{S}_2\right)$ is given by the convex hull of the closure of all non-negative rate pairs (R_1, R_2, R_3) such that

$$R_1 \leq g\left(\mathbf{S}_1, \mathbf{Q}_1, \mathbf{N}_1^{(w)}\right) + g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(w)}\right) \tag{26}$$

$$R_1 + R_2 \leq g\left(\mathbf{S}_1, 0, \mathbf{N}_1^{(y)}\right) + g\left(\mathbf{S}_2, \mathbf{Q}_2, \mathbf{N}_2^{(y)}\right) \tag{27}$$

$$R_1 + R_2 \leq g\left(\mathbf{S}_1, \mathbf{Q}_1, \mathbf{N}_1^{(w)}\right) + g\left(\mathbf{Q}_1, 0, \mathbf{N}_1^{(y)}\right) + g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(w)}\right) \tag{28}$$

$$R_1 + R_2 + R_3 \leq g\left(\mathbf{S}_1, 0, \mathbf{N}_1^{(z)}\right) + g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(z)}\right) \tag{29}$$

$$R_1 + R_2 + R_3 \leq g\left(\mathbf{S}_1, \mathbf{Q}_1, \mathbf{N}_1^{(w)}\right) + g\left(\mathbf{Q}_1, 0, \mathbf{N}_1^{(z)}\right) + g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(w)}\right) \tag{30}$$

$$R_1 + R_2 + R_3 \leq g\left(\mathbf{S}_1, 0, \mathbf{N}_1^{(y)}\right) + g\left(\mathbf{S}_2, \mathbf{Q}_2, \mathbf{N}_2^{(y)}\right) + g\left(\mathbf{Q}_2, 0, \mathbf{N}_2^{(z)}\right) \tag{31}$$

$$R_1 + R_2 + R_3 \leq g\left(\mathbf{S}_1, \mathbf{Q}_1, \mathbf{N}_1^{(w)}\right) + g\left(\mathbf{Q}_1, 0, \mathbf{N}_1^{(y)}\right) + g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(w)}\right) \tag{32}$$

$$2R_1 + R_2 + R_3 \leq g\left(\mathbf{S}_1, \mathbf{Q}_1, \mathbf{N}_1^{(w)}\right) + g\left(\mathbf{S}_1, 0, \mathbf{N}_1^{(y)}\right) + g\left(\mathbf{S}_2, \mathbf{Q}_2, \mathbf{N}_2^{(y)}\right) + g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(w)}\right) \tag{33}$$

where $\mathbf{Q}_1, \mathbf{Q}_2 \succeq 0$, $\mathbf{Q}_1 \preceq \mathbf{S}_1$ and $\mathbf{Q}_2 \preceq \mathbf{S}_2$ and where $g\left(\mathbf{A}, \mathbf{B}, \mathbf{C}\right) \triangleq \frac{1}{2}\log\frac{|\mathbf{A}+\mathbf{C}|}{|\mathbf{B}+\mathbf{C}|}$.

Proof: A single characterization for the capacity region can be found in [7, Theorem 1]. The bulk of the proof consists in showing that Gaussian inputs, together with time-sharing, attains the capacity region. Since the capacity region is convex due to time-sharing, it may be described by the weighted sum rate, i.e., by the following optimization problem:

$$(P1): \sup_{\substack{P_{U,X} \\ (R_1,R_2,R_3)}} \mu_1 R_1 + \mu_2 R_2 + \mu_3 R_3$$

and $R_1, R_2, R_3 \geq 0$ satisfy [7, (12)-(19)] subject to the input matrix constraints (25). Due to space constraints, an essential step of the proof relies on showing that the following holds

$$\begin{aligned}
&\sup_{\substack{P_{U_1,X_1} \\ \text{s.t. } \mathbb{E}\left[\boldsymbol{X}_1\boldsymbol{X}_1^T\right] \preceq \mathbf{S}_1}} \begin{array}{l} \left(\alpha_1+\alpha_3+\alpha_5+\alpha_7+\alpha_8\right) I\left(U_1; \boldsymbol{W}_1\right) \\ + \left(\alpha_3+\alpha_7\right) I\left(\boldsymbol{X}_1; \boldsymbol{Y}_1 | U_1\right) \\ + \alpha_5 I\left(\boldsymbol{X}_1; \boldsymbol{Z}_1 | U_1\right) \\ + \left(\alpha_2+\alpha_6+\alpha_8\right) I\left(\boldsymbol{X}_1; \boldsymbol{Y}_1\right) \\ + \alpha_4 I\left(\boldsymbol{X}_1; \boldsymbol{Z}_1\right) \end{array} \\
&= \left(\alpha_1+\alpha_3+\alpha_5+\alpha_7+\alpha_8\right) g\left(\mathbf{S}_1, \mathbf{Q}_1, \mathbf{N}_1^{(w)}\right) \\
&\quad + \left(\alpha_3+\alpha_7\right) g\left(\mathbf{Q}_1, 0, \mathbf{N}_1^{(y)}\right) + \alpha_5 g\left(\mathbf{Q}_1, 0, \mathbf{N}_1^{(z)}\right) \\
&\quad + \left(\alpha_2+\alpha_6+\alpha_8\right) g\left(\mathbf{S}_1, 0, \mathbf{N}_1^{(y)}\right) + \alpha_4 g\left(\mathbf{S}_1, 0, \mathbf{N}_1^{(z)}\right)
\end{aligned}$$

for some $0 \preceq \mathbf{Q}_1 \preceq \mathbf{S}_1$ and the following holds

$$\begin{aligned}
&\sup_{\substack{P_{U_2,X_2} \\ \text{s.t. } \mathbb{E}\left[\boldsymbol{X}_2\boldsymbol{X}_2^T\right] \preceq \mathbf{S}_2}} \begin{array}{l} \left(\alpha_1+\alpha_3+\alpha_5+\alpha_7+\alpha_8\right) I\left(\boldsymbol{X}_2; \boldsymbol{W}_2\right) \\ + \left(\alpha_2+\alpha_6+\alpha_8\right) I\left(U_2; \boldsymbol{Y}_2\right) \\ + \alpha_4 I\left(\boldsymbol{X}_2; \boldsymbol{Z}_2\right) + \alpha_6 I\left(\boldsymbol{X}_2; \boldsymbol{Z}_2 | U_2\right) \end{array} \\
&= \left(\alpha_1+\alpha_3+\alpha_5+\alpha_7+\alpha_8\right) g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(w)}\right) \\
&\quad + \left(\alpha_2+\alpha_6+\alpha_8\right) g\left(\mathbf{S}_2, \mathbf{Q}_2, \mathbf{N}_2^{(y)}\right) \\
&\quad + \alpha_4 g\left(\mathbf{S}_2, 0, \mathbf{N}_2^{(z)}\right) + \alpha_6 g\left(\mathbf{Q}_2, 0, \mathbf{N}_2^{(z)}\right)
\end{aligned}$$

for some $0 \preceq \mathbf{Q}_2 \preceq \mathbf{S}_2$. The above follow directly from Theorem 1. Interested readers may refer to [8] for more details of the proof in a more general setting. ■

References

[1] R. Liu, T. Liu, H. V. Poor and S. Shamai, "A vector generalization of Costa's entropy-power inequality with applications," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1865–1879, 2010.

[2] H. Weingarten, T. Liu, S. Shamai, Y. Steinberg and P. Viswanath, "The capacity region of the degraded multiple-input multiple-output compound broadcast channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5011–5023, Nov. 2009.

[3] Y. Geng and C. Nair, "The capacity region of the two-receiver vector gaussian broadcast channel with private and common messages." [Online]. Available: http://arxiv.org/pdf/1202.0097v1.pdf

[4] O. Rioul, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, Jan. 2011.

[5] M. Cohen, "Fisher information and convexity," *IEEE Trans. Inf. Theory*, vol. 14, no. 4, pp. 591–592, Jul. 1968.

[6] H. F. Chong and Y.-C. Liang, "The capacity region of the class of Gaussian MIMO three-receiver multilevel broadcast channels with degraded message sets," in *Proc. 48th Annual Allerton Conf. Commun. Control Comput.*, UIUC, Illinois, USA, 29 Sept.-1 Oct. 2010 2010, pp. 94–101.

[7] ——, "The capacity region of some classes of parallel degraded broadcast channels with three receivers and three-degraded message sets," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Jul. 1–6, 2012, pp. 383–387.

[8] ——, "On the Capacity Region of the Parallel Degraded Broadcast Channel with Three Receivers and Three-Degraded Message Sets," *IEEE Trans. Inf. Theory*, submitted for publication.