

Capacity Region of Reversely Degraded Gaussian MIMO Broadcast Channel

Jun Chen

Dept. of Electrical and Computer Engr.
McMaster University
Hamilton, Ontario, Canada

Chao Tian

AT&T Labs-Research
180 Park Avenue
Florham Park, NJ 07932, USA

Abstract—We consider the problem of broadcasting a common message and two individual messages to two users on a product channel of two reversely degraded Gaussian multiple-input multiple-output (MIMO) broadcast channels. Though El Gamal provided a single letter characterization for the general discrete memoryless problem in 1980, this characterization in fact does not include a channel cost constraint, and thus does not apply directly to the Gaussian MIMO setting. We first show El Gamal’s single letter characterization can indeed be generalized to include channel cost constraints, however special care has to be taken and the characterization holds only with certain class of cost functions. This characterization has an equivalent form, and by utilizing this form, as well as the enhancement technique and an extremal inequality which were only discovered recently, we show that indeed Gaussian codebooks are optimal for this MIMO setting.

I. INTRODUCTION

The capacity region of general memoryless broadcast channels remains an open problem in multi-user information theory, despite the establishment of capacity region in many special cases such as degraded [1][2], deterministic [3][4], more powerful broadcast channels [5] and broadcast channels with degraded message sets [6] (see [7] for a review). Recently, important progress was made in this area for Gaussian multiple-input multiple-output (MIMO) channels in [8], where Weingarten et al. showed that dirty paper coding based on Gaussian codebooks is optimal when only individual messages are required, and thus completely characterized the capacity region for this case.

When common message is involved in non-degraded broadcast channels in addition to individual messages, known results in the literature are quite limited. In [9], El Gamal provided a single letter characterization of the capacity region for the product of reversely degraded broadcast channels (often simply referred to as the reversely degraded broadcast channel), and furthermore, showed that when the two individual broadcast channels are scalar Gaussian, Gaussian codebooks are optimal. In [10], the problem of Gaussian MIMO broadcast channel with common message is considered, and it was shown a coding scheme based on dirty paper coding [11] with Gaussian codebooks is optimal under certain specific conditions.

In this work, we consider the problem of broadcasting a common message and two individual messages to two users

on a reversely degraded Gaussian MIMO broadcast channel. Although a single letter characterization of the capacity region was established in [9] for general discrete memoryless channels, this characterization in fact does not include channel cost constraint, and thus can not be applied directly to the MIMO setting. We first show that the result in [9] can indeed be generalized to include channel cost constraints, however special care has to be taken and the characterization holds only with certain class of cost functions. The characterization has an equivalent form, and by utilizing this form, as well as the enhancement technique [8] and an extremal inequality [12], we prove the non-trivial fact that Gaussian codebooks are optimal for this MIMO setting under separate covariance constraints on two sub-channels, individual antenna power constraints, or a total power constraint¹. Similar to [8], the case with separate covariance constraints on two sub-channels can be understood as an intermediate step toward the more realistic constraint on total power.

The rest of the paper is organized as follows. In Section II, we describe channel model and provide necessary notations. The main result is given in Section III. The generalization of El Gamal’s characterization to include certain cost functions, and the alternative single letter characterization are discussed in Section IV, and the proof of the main result is given in Section V. Finally Section VI concludes the paper.

II. CHANNEL MODELS

We first describe the system for general discrete memoryless channels, and then introduce the specific Gaussian MIMO notations.

The broadcast channel in question is given by channel input in the alphabet $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$, and first user channel output in the alphabet $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2)$, and second user channel output in the alphabet $\mathcal{Z} = (\mathcal{Z}_1, \mathcal{Z}_2)$, where the transition probability $P(Y_1, Y_2, Z_1, Z_2 | X_1, X_2)$ factorizes as follows

$$\begin{aligned} P(Y_1, Y_2, Z_1, Z_2 | X_1, X_2) \\ = P(Y_1 | X_1) P(Z_1 | Y_1) \cdot P(Z_2 | X_2) P(Y_2 | Z_2). \end{aligned} \quad (1)$$

Thus, the overall broadcast channel consists of two parallel broadcast channels with reversely degraded outputs; note this

¹These settings do not satisfy the specific conditions under which tight characterization was given in [10].

also implies the fact that $Y_2 \leftrightarrow Z_2 \leftrightarrow X_2 \leftrightarrow X_1 \leftrightarrow Y_1 \leftrightarrow Z_1$ is a Markov chain.

For channel with cost constraints, a set of cost functions $c_i : \mathcal{X} \rightarrow \mathcal{C}_i$, $i = 1, 2, \dots, K$ are defined, and \mathcal{C}_i is a space of t_i -by- t_i positive semidefinite matrices, with the partial order \preceq defined by positive semidefiniteness.

Definition 1: An (M_0, M_1, M_2, n) broadcast code consists of an encoding function

$$f : \mathcal{I}_{M_0} \times \mathcal{I}_{M_1} \times \mathcal{I}_{M_2} \rightarrow \mathcal{X}^n, \quad (2)$$

where $\mathcal{I}_n = \{1, 2, \dots, n\}$ and two decoding functions

$$\begin{aligned} g_1 : \mathcal{Y}^n &\rightarrow \mathcal{I}_{M_0} \times \mathcal{I}_{M_1} = (\hat{W}_0, \hat{W}_1) \\ g_2 : \mathcal{Z}^n &\rightarrow \mathcal{I}_{M_0} \times \mathcal{I}_{M_2} = (\hat{W}_0, \hat{W}_2). \end{aligned} \quad (3)$$

A decoding error is defined in the usual way as either at user one $(w_0, w_1) \neq (\hat{w}_0, \hat{w}_1)$ or at user two $(w_0, w_2) \neq (\hat{w}_0, \hat{w}_2)$; the probability of decoding error P_e is averaged over the message space $\mathcal{I}_{M_0} \times \mathcal{I}_{M_1} \times \mathcal{I}_{M_2}$, i.e., assuming the messages are distributed uniformly.

Definition 2: A rate triple (R_0, R_1, R_2) is called achievable under channel cost constraints $\mathbf{C}_i \in \mathcal{C}_i$ where $\mathbf{C}_i \succcurlyeq \mathbf{0}$, $i = 1, 2, \dots, K$, if for any $\epsilon > 0$ and sufficiently large n there exists an (M_0, M_1, M_2) code such that

$$M_i \geq 2^{nR_i}, \quad i = 1, 2, 3$$

$$\mathbb{E} \frac{1}{n} \sum_{j=1}^n c_i(X(j)) \preceq \mathbf{C}_i + \epsilon \mathbf{I}, \quad i = 1, 2, \dots, K, \quad (4)$$

where $X(j)$ is the channel codeword input at j -th position, and $P_e < \epsilon$. The capacity region under cost constraints $(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)$, is the closure of the set of achievable rate triples, denoted as $\mathcal{C}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)$.

For the Gaussian MIMO problem, we consider the setting that the input X_i is in fact a vector \mathbf{X}_i for $i = 1, 2$, and

$$\begin{aligned} Y_1 = \mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{n}_1, \quad Z_1 = \mathbf{Z}_1 = \mathbf{X}_1 + \mathbf{m}_1, \\ Y_2 = \mathbf{Y}_2 = \mathbf{X}_2 + \mathbf{n}_2, \quad Z_2 = \mathbf{Z}_2 = \mathbf{X}_2 + \mathbf{m}_2, \end{aligned} \quad (5)$$

where \mathbf{n}_i is a t_i -dimensional random Gaussian vector with covariance \mathbf{N}_i , and \mathbf{m}_i is a t_i -dimensional random Gaussian vector with covariance \mathbf{M}_i , for $i = 1, 2$; moreover, $(\mathbf{n}_1, \mathbf{m}_1)$ is independent of $(\mathbf{n}_2, \mathbf{m}_2)$. The covariances are given such that $\mathbf{0} \prec \mathbf{N}_1 \preceq \mathbf{M}_1$ and $\mathbf{0} \prec \mathbf{M}_2 \preceq \mathbf{N}_2$.

We shall consider three kinds of constraints:

- Separate covariance constraints: given two positive semidefinite matrices \mathbf{S}_1 and \mathbf{S}_2 , the two cost functions over t_i -dimensional vectors are given by $c_i(\mathbf{x}) = \mathbf{x}_i \mathbf{x}_i^t$. In other words, two covariance constraints are placed separately over the two sub-channels. We denote the capacity region as $\mathcal{C}(\mathbf{S}_1, \mathbf{S}_2)$.
- Individual antenna power constraints: given a $(t_1 + t_2)$ -dimensional non-negative vector $(P_1, P_2, \dots, P_{t_1+t_2})$, the $(t_1 + t_2)$ cost functions are given by $c_i(\mathbf{x}) = x^2(i)$. We denote the capacity region as $\mathcal{C}(P_1, P_2, \dots, P_{t_1+t_2})$.
- Total power constraint: given a non-negative power constraint P , the scalar-valued cost function is given by $c_1(\mathbf{x}) = \mathbf{x}^t \mathbf{x}$. We denote the capacity region as $\mathcal{C}(P)$.

As we shall discuss in Section IV, these three kinds of cost

functions belong to the so-called class of statistically separable cost functions, which allows an explicit characterization of the capacity region.

III. MAIN RESULTS

Define the region $\mathcal{R}(\mathbf{S}_1, \mathbf{S}_2)$ as the non-negative rate triple (R_0, R_1, R_2) for which there exist some positive semidefinite matrices $(\mathbf{Q}_1, \mathbf{Q}_2)$, $\mathbf{0} \preceq \mathbf{Q}_i \preceq \mathbf{S}_i$ for $i = 1, 2$, such that

$$\begin{aligned} R_0 + R_1 &\leq R_0^* + R_1^* \\ R_0 + R_2 &\leq R_0^* + R_2^* \\ R_0 + R_1 + R_2 &\leq R_0^* + R_1^* + R_2^* \end{aligned} \quad (6)$$

where

$$R_0^* = \min \left\{ \frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{N}_1|}{|\mathbf{Q}_1 + \mathbf{N}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{N}_2|}{|\mathbf{Q}_2 + \mathbf{N}_2|}, \right. \\ \left. \frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{M}_1|}{|\mathbf{Q}_1 + \mathbf{M}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{M}_2|}{|\mathbf{Q}_2 + \mathbf{M}_2|} \right\}, \quad (7)$$

$$R_1^* = \frac{1}{2} \log \frac{|\mathbf{Q}_1 + \mathbf{N}_1|}{|\mathbf{N}_1|}, \quad (8)$$

$$R_2^* = \frac{1}{2} \log \frac{|\mathbf{Q}_2 + \mathbf{M}_2|}{|\mathbf{M}_2|}. \quad (9)$$

We have the following theorems.

Theorem 1:

$$\mathcal{C}(\mathbf{S}_1, \mathbf{S}_2) = \text{conv}(\mathcal{R}(\mathbf{S}_1, \mathbf{S}_2)),$$

where $\text{conv}(\cdot)$ is the convex hull operator.

Theorem 2:

$$\begin{aligned} \mathcal{C}(P_1, P_2, \dots, P_{t_1+t_2}) \\ = \bigcup_{(\mathbf{S}_1, \mathbf{S}_2) \in \mathcal{P}(P_1, P_2, \dots, P_{t_1+t_2})} \text{conv}(\mathcal{R}(\mathbf{S}_1, \mathbf{S}_2)), \end{aligned}$$

where $\mathcal{P}(P_1, P_2, \dots, P_{t_1+t_2})$ is the set of pair of positive semidefinite matrices such that $\mathbf{S}_1(i) \leq P_i$ and $\mathbf{S}_2(i) \leq P_{i+t_1}$, where $\mathbf{S}_i(j)$ is the j -th diagonal element of \mathbf{S}_i .

Theorem 3:

$$\mathcal{C}(P) = \bigcup_{(\mathbf{S}_1, \mathbf{S}_2) \in \mathcal{P}(P)} \text{conv}(\mathcal{R}(\mathbf{S}_1, \mathbf{S}_2)),$$

where $\mathcal{P}(P)$ is the set of pair of positive semidefinite matrices such that $\text{Tr}(\mathbf{S}_1) + \text{Tr}(\mathbf{S}_2) \leq P$, where $\text{Tr}(\cdot)$ is the trace.

Remark: We write the form of the capacity region as in (6)-(9), because it reveals a fundamental constraint in broadcast when common message is involved. Since common message can be also used to convey individual messages, and vice versa, the rate region in (6) can be taken as a minimal capacity template, or the so-called latent capacity region [13][14].

The rest of the paper is largely devoted to the proof of Theorem 1, along the direction outlined in Section I.

IV. CAPACITY CHARACTERIZATION WITH CHANNEL COST CONSTRAINT AND AN ALTERNATIVE FORM

In this section, we first show that the capacity region characterization in [9] can be generalized to include channel cost constraints, however the result only holds for certain class of cost functions, namely channel cost functions that are statistically separable. Through the proof of this capacity

result, we introduce an alternative form of the capacity region, which will become useful in the next section. In this section, we focus on the general discrete memoryless model.

Definition 3: A cost function $c(\cdot)$ is called statistically separable, if for any joint distribution $P(X_1, X_2)$, and $P'(X_1, X_2) = P(X_1)P(X_2)$, $\mathbb{E}_P(c(X_1, X_2)) = \mathbb{E}_{P'}(c(X_1, X_2))$, where the expectations are with respect to the distribution $P(X_1, X_2)$ and $P'(X_1, X_2)$, respectively.

Let $\mathcal{R}(U_1, U_2, X_1, X_2)$ be the collection of non-negative rate triples (R_0, R_1, R_2) , for some choice of probability distribution $P(U_1, U_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2)$ such that

$$R_0 \leq \min \{I(U_1; Y_1) + I(U_2; Y_2), I(U_1; Z_1) + I(U_2; Z_2)\}, \quad (10)$$

$$R_0 + R_1 \leq I(U_1; Y_1) + I(U_2; Y_2) + I(X_1; Y_1|U_1), \quad (11)$$

$$R_0 + R_2 \leq I(U_1; Z_1) + I(U_2; Z_2) + I(X_2; Z_2|U_2), \quad (12)$$

$$R_0 + R_1 + R_2 \leq I(X_1; Y_1|U_1) + I(X_2; Z_2|U_2) \\ \min \{I(U_1; Y_1) + I(U_2; Y_2), I(U_1; Z_1) + I(U_2; Z_2)\}. \quad (13)$$

The following theorem states that when the cost functions are statistically separable, the characterization in [9] can be generalized to the case with channel cost constraints.

Theorem 4: If the cost functions c_1, c_2, \dots, c_K are statistically separable, then

$$\mathcal{C}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K) \\ = \bigcup_{(U_1, U_2, X_1, X_2) \in \mathcal{P}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)} \mathcal{R}(U_1, U_2, X_1, X_2),$$

where $\mathcal{P}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)$ is the set of distributions satisfying the following conditions

- The distribution factorizes as $P(U_1)(U_2)P(X_1|U_1) \cdot P(X_2|U_2)P(Y_1, Z_1|X_1)P(Y_2, Z_2|X_2)$; in other words, (U_1, X_1) and (U_2, X_2) are independent;
- $\mathbb{E}(c_i(X_1, X_2)) \preceq \mathbf{C}_i$, for $i = 1, 2, \dots, K$.

Remark: The only difference from [9] is the condition on the cost functions, and the last condition on the distribution induced by the cost function. We can in fact similarly define the notion of statistical sub-separability when $\mathbb{E}_{P'}(c(X_1, X_2)) \preceq \mathbb{E}_P(c(X_1, X_2))$, and it can also be shown this is sufficient for Theorem 4 to hold.

Remark: It is clear for the Gaussian MIMO case, all three channel cost functions given in Section II are statistically separable.

To prove Theorem 4, we need the following theorem. Let $\check{\mathcal{R}}(U_1, U_2, X_1, X_2)$ be the collection of non-negative rate triples (R_0, R_1, R_2) , for some choice of probability distribution $P(U_1, U_2, X_1, X_2, Y_1, Y_2, Z_1, Z_2)$ such that

$$R_0 \leq \check{R}_0, \quad (14)$$

$$R_0 + R_1 \leq \check{R}_0 + I(X_1; Y_1|U_1), \quad (15)$$

$$R_0 + R_2 \leq \check{R}_0 + I(X_2; Z_2|U_2), \quad (16)$$

$$R_0 + R_1 + R_2 \leq \check{R}_0 + I(X_1; Y_1|U_1) + I(X_2; Z_2|U_2). \quad (17)$$

where

$$\check{R}_0 = \min \{I(U_1; Y_1) + I(U_2; Y_2), I(U_1; Z_1) + I(U_2; Z_2)\}$$

Theorem 5:

$$\mathcal{C}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K) \\ \supseteq \bigcup_{(U_1, U_2, X_1, X_2) \in \mathcal{P}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)} \check{\mathcal{R}}(U_1, U_2, X_1, X_2),$$

where $\mathcal{P}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)$ is the same as in Theorem 4.

The proof of this theorem is rather straightforward by using the well known super-position codebook used in degraded broadcast channels [2] and the cost function conditions are the direct consequences of typical sequences, and we thus omitted the details of the proof. Next we prove Theorem 4.

Proof of Theorem 4: The converse part follows directly from the proof in [9]. However, during the last step of proof, the dependence between (U_1, X_1) and (U_2, X_2) can not be severed arbitrarily, since though the mutual information quantities only involve the marginal distributions, the cost functions may not preserve their costs during this transformation. However when the cost functions are statistically separable, indeed we can sever the dependence between (U_1, X_1) and (U_2, X_2) without causing any change to the region.

Now we turn to the forward part of the proof. We claim that for any (U_1, U_2, X_1, X_2) such that (U_1, X_1) is independent of (U_2, X_2) and $U'_i \leftrightarrow X_i \leftrightarrow (Y_i, Z_i)$ is a Markov string for $i = 1, 2$, there exist some (U'_1, U'_2, X_1, X_2) where (U'_1, X_1) is also independent of (X_2, U'_2) which preserves the marginal distribution of (X_1, X_2) , and $U'_i \leftrightarrow X_i \leftrightarrow (Y_i, Z_i)$ is a Markov string for $i = 1, 2$, and furthermore,

$$\mathcal{R}(U_1, U_2, X_1, X_2) \subseteq \check{\mathcal{R}}(U'_1, U'_2, X_1, X_2). \quad (18)$$

Note that since the marginal distribution of (X_1, X_2) is preserved, the cost function conditions are also preserved. Once this inclusion is proved, Theorem 5 can be invoked to complete the proof of Theorem 4.

Without loss of generality, we assume $I(U_1; Y_1) + I(U_2; Y_2) \geq I(U_1; Z_1) + I(U_2; Z_2)$. If this inequality in fact holds with equality, then clearly $\mathcal{R}(U_1, U_2, X_1, X_2) = \check{\mathcal{R}}(U_1, U_2, X_1, X_2)$. If the inequality is strict, then define $U_2^{(\alpha)} \triangleq (T_\alpha, \hat{U})$, $\alpha \in [0, 1]$, where T_α is a Bernoulli random variable with $P(T_\alpha = 1) = \alpha$ and

$$\hat{U} = \begin{cases} U_2, & T_\alpha = 0 \\ X_2, & T_\alpha = 1 \end{cases} \quad (19)$$

This implies that $U_2^{(0)} = (0, U_2)$, $U_2^{(1)} = (1, X_2)$, and moreover $U_2 \leftrightarrow U_2^{(\alpha)} \leftrightarrow X_2$ is a Markov string for any $\alpha \in [0, 1]$, and the marginal distribution of (U_1, X_1, X_2) is preserved, and clearly the Markov string $U_2^{(\alpha)} \leftrightarrow X_2 \leftrightarrow (Y_2, Z_2)$ is true for any α .

We shall increase α until $I(U_1; Y_1) + I(U_2; Y_2) = I(U_1; Z_1) + I(U_2^{(\alpha)}; Z_2)$ or $\alpha = 1$ (i.e., $U_2^{(\alpha)} = X_2$); we denote the corresponding α as α^* . Next we show that in either case, $\mathcal{R}(U_1, U_2, X_1, X_2) \subseteq \check{\mathcal{R}}(U_1, U_2^{(\alpha^*)}, X_1, X_2)$.

If $\alpha^* \neq 1$, the respective comparisons between (10)-(13) and (14)-(17) are straightforward by noticing that both $I(U_1; Y_1) + I(U_2^{(\alpha)}; Y_2)$ and $I(U_1; Z_1) + I(U_2^{(\alpha)}; Z_2)$ are

monotone increasing functions of α , and thus

$$\begin{aligned} I(U_1; Z_1) + I(U_2^{\alpha^*}; Z_2) &= I(U_1; Y_1) + I(U_2; Y_2) \\ &\leq I(U_1; Y_1) + I(U_2^{(\alpha)}; Y_2). \end{aligned}$$

If $\alpha^* = 1$, then we clearly have

$$I(U_1; Z_1) + I(X_2; Z_2) \leq I(U_1; Y_1) + I(U_2; Y_2). \quad (20)$$

The comparison of other quantities are again straightforward, and we only need to consider the comparison between (11) and (15). Note that because of (20), the right hand side of (13) can be written to satisfy the following inequality

$$\begin{aligned} I(U_1; Z_1) + I(U_2; Z_2) + I(X_1; Y_1|U_1) + I(X_2; Y_2|U_2) \\ = I(U_1; Z_1) + I(X_2; Z_2) + I(X_1; Y_1|U_1) \end{aligned} \quad (21)$$

$$\leq I(U_1; Y_1) + I(U_2; Y_2) + I(X_1; Y_1|U_1), \quad (22)$$

however the quantity in (22) is exactly the right hand side of (11), and since R_0, R_1, R_2 are non-negative, the condition in (11) can be safely removed and replaced by the quantity in (21), i.e.,

$$R_0 + R_1 \leq I(U_1; Z_1) + I(X_2; Z_2) + I(X_1; Y_1|U_1),$$

which is however exactly the inequality of (15) for $\alpha^* = 1$. This completes the proof. ■

From Theorem 4 and Theorem 5, it is seen that an alternative characterization of $\mathcal{C}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)$ is as follows.

Corollary 1:

$$\begin{aligned} \mathcal{C}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K) \\ = \bigcup_{(U_1, U_2, X_1, X_2) \in \mathcal{P}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)} \tilde{\mathcal{R}}(U_1, U_2, X_1, X_2), \end{aligned}$$

where $\mathcal{P}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K)$ is the same as in Theorem 4.

V. REVERSELY DEGRADED GAUSSIAN MIMO BROADCAST CHANNEL

Now we turn to the reversely degraded Gaussian MIMO broadcast channel. We first prove Theorem 1. The proof of Theorem 2 and 3 will also be discussed briefly.

To prove the converse, a result from [12] is needed.

Theorem 6 ([12]): Let \mathbf{Q}^* be a positive semidefinite matrix such that $\mathbf{0} \preceq \mathbf{Q}^* \preceq \mathbf{S}$ and such that

$$\beta_1 \sum_{i=1}^{K_1} \lambda_i^{(1)} (\mathbf{Q}^* + \mathbf{N}_i)^{-1} = \beta_2 \sum_{j=1}^{K_2} \lambda_j^{(2)} (\mathbf{Q}^* + \mathbf{M}_j)^{-1} + \mathbf{O},$$

where $\beta_i > 0$ for $i = 1, 2$, $\lambda_i^{(1)} \geq 0$, $\lambda_j^{(2)} \geq 0$, $\mathbf{O} \succeq \mathbf{0}$, $(\mathbf{S} - \mathbf{Q}^*) \cdot \mathbf{O} = \mathbf{0}$, and there exists an \mathbf{N}^* such that $\mathbf{N}_i \preceq \mathbf{N}^* \preceq \mathbf{M}_j$ for all i and j . Then for any T independent of $\mathbf{W}_i \sim \mathcal{N}(0, \mathbf{N}_i)$ and $\mathbf{V}_j \sim \mathcal{N}(0, \mathbf{M}_j)$ and \mathbf{X} such that $\mathbb{E}(\mathbf{X}\mathbf{X}^t) \preceq \mathbf{S}$ we have

$$\begin{aligned} \beta_1 \sum_{i=1}^{K_1} \lambda_i^{(1)} h(\mathbf{X} + \mathbf{W}_i|T) - \beta_2 \sum_{j=1}^{K_2} \lambda_j^{(2)} h(\mathbf{X} + \mathbf{V}_j|T) \\ \leq \frac{\beta_1}{2} \sum_{i=1}^{K_1} \lambda_i^{(1)} \log |2\pi e(\mathbf{Q}^* + \mathbf{N}_i)| \\ - \frac{\beta_2}{2} \sum_{j=1}^{K_2} \lambda_j^{(2)} \log |2\pi e(\mathbf{Q}^* + \mathbf{M}_j)|. \end{aligned}$$

Next consider the optimization problem

$$\max_{\mathbf{0} \preceq \mathbf{Q}_1 \preceq \mathbf{S}_1, \mathbf{0} \preceq \mathbf{Q}_2 \preceq \mathbf{S}_2} \mu_0 R_0^* + \mu_1 R_1^* + \mu_2 R_2^*,$$

where the clearly achievable rate triple (R_0^*, R_1^*, R_2^*) is defined in (7)-(9), and $\mu_0 \geq (\mu_1, \mu_2) \geq 0$ (since otherwise one can set $R_0 = 0$ and the problem reduces to the case with only private messages). Clearly, we can assume $\mu_0 > 0$ since the objective function is zero otherwise. It is possible to show using Fritz John necessary conditions [16] that the following lemma is true.

Lemma 1: There exist $\lambda_i \geq 0$ ($i = 1, 2$), $\lambda_1 + \lambda_2 = 1$ and \mathbf{Q}_1^* and \mathbf{Q}_2^* such that

$$\begin{aligned} \max_{\mathbf{0} \preceq \mathbf{Q}_1 \preceq \mathbf{S}_1, \mathbf{0} \preceq \mathbf{Q}_2 \preceq \mathbf{S}_2} \mu_0 R_0^* + \mu_1 R_1^* + \mu_2 R_2^* \\ = \mu_0 \lambda_1 \left(\frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{N}_1|}{|\mathbf{Q}_1^* + \mathbf{N}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{N}_2|}{|\mathbf{Q}_2^* + \mathbf{N}_2|} \right) \\ + \mu_0 \lambda_2 \left(\frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{M}_1|}{|\mathbf{Q}_1^* + \mathbf{M}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{M}_2|}{|\mathbf{Q}_2^* + \mathbf{M}_2|} \right) \\ + \frac{\mu_1}{2} \log \frac{|\mathbf{Q}_1^* + \mathbf{N}_1|}{|\mathbf{N}_1|} + \frac{\mu_2}{2} \log \frac{|\mathbf{Q}_2^* + \mathbf{M}_2|}{|\mathbf{M}_2|}, \end{aligned}$$

and furthermore there exist positive semidefinite matrices \mathbf{O}_1 and \mathbf{O}_2 such that

- 1) $\mu_1(\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{O}_1 = \mu_0 \lambda_1 (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mu_0 \lambda_2 (\mathbf{Q}_1^* + \mathbf{M}_1)^{-1} + \mathbf{O}'_1$ and $\mu_2(\mathbf{Q}_2^* + \mathbf{M}_2)^{-1} + \mathbf{O}_2 = \mu_0 \lambda_1 (\mathbf{Q}_2^* + \mathbf{N}_2)^{-1} + \mu_0 \lambda_2 (\mathbf{Q}_2^* + \mathbf{M}_2)^{-1} + \mathbf{O}'_2$.
- 2) $\mathbf{Q}_i^* \cdot \mathbf{O}_i = \mathbf{0}$ and $(\mathbf{S}_i - \mathbf{Q}_i^*) \cdot \mathbf{O}'_i = \mathbf{0}$, $i = 1, 2$.
- 3) We have $\lambda_1 = 1$ and $\lambda_2 = 0$ if

$$\begin{aligned} \frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{N}_1|}{|\mathbf{Q}_1^* + \mathbf{N}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{N}_2|}{|\mathbf{Q}_2^* + \mathbf{N}_2|} \\ < \frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{M}_1|}{|\mathbf{Q}_1^* + \mathbf{M}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{M}_2|}{|\mathbf{Q}_2^* + \mathbf{M}_2|} \end{aligned}$$

and $\lambda_1 = 0$ and $\lambda_2 = 1$ if

$$\begin{aligned} \frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{N}_1|}{|\mathbf{Q}_1^* + \mathbf{N}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{N}_2|}{|\mathbf{Q}_2^* + \mathbf{N}_2|} \\ > \frac{1}{2} \log \frac{|\mathbf{S}_1 + \mathbf{M}_1|}{|\mathbf{Q}_1^* + \mathbf{M}_1|} + \frac{1}{2} \log \frac{|\mathbf{S}_2 + \mathbf{M}_2|}{|\mathbf{Q}_2^* + \mathbf{M}_2|}. \end{aligned}$$

Now we are ready to prove Theorem 1.

Proof of Theorem 1:

The forward part is clearly true, and we only need to prove the converse. Without loss of generality, we shall assume $\mu_0 \geq \mu_1 \geq \mu_2 \geq 0$, and subsequently choose λ_1 and λ_2 as specified in Lemma 1. We start by writing the following inequality using Corollary 1,

$$\begin{aligned} \mu_0 R_0 + \mu_1 R_1 + \mu_2 R_2 \\ = \mu_2 (R_0 + R_1 + R_2) + (\mu_1 - \mu_2) (R_0 + R_1) + (\mu_0 - \mu_1) R_0 \\ \leq \mu_0 \min \{ I(U_1; \mathbf{Y}_1) + I(U_2; \mathbf{Y}_2), I(U_1; \mathbf{Z}_1) + I(U_2; \mathbf{Z}_2) \} \\ + \mu_1 I(\mathbf{X}_1; \mathbf{Y}_1|U_1) + \mu_2 I(\mathbf{X}_2; \mathbf{Z}_2|U_2) \\ \leq \mu_0 \lambda_1 (I(U_1; \mathbf{Y}_1) + I(U_2; \mathbf{Y}_2)) + \mu_0 \lambda_2 I(U_1; \mathbf{Z}_1) \\ + \mu_0 \lambda_2 I(U_2; \mathbf{Z}_2) + \mu_1 I(\mathbf{X}_1; \mathbf{Y}_1|U_1) + \mu_2 I(\mathbf{X}_2; \mathbf{Z}_2|U_2). \end{aligned}$$

Note that

$$\begin{aligned}
& \mu_0 \lambda_1 (I(U_1; \mathbf{Y}_1) + I(U_2; \mathbf{Y}_2)) \\
& + \mu_0 \lambda_2 (I(U_1; \mathbf{Z}_1) + I(U_2; \mathbf{Z}_2)) \\
& + \mu_1 I(\mathbf{X}_1; \mathbf{Y}_1|U_1) + \mu_2 I(\mathbf{X}_2; \mathbf{Z}_2|U_2) \\
& = \mu_0 \lambda_1 (h(\mathbf{Y}_1) + h(\mathbf{Y}_2)) + \mu_0 \lambda_2 (h(\mathbf{Z}_1) + h(\mathbf{Z}_2)) \\
& - \mu_1 h(\mathbf{Y}_1|\mathbf{X}_1) - \mu_2 h(\mathbf{Z}_2|\mathbf{X}_2) + (\mu_1 - \mu_0 \lambda_1) h(\mathbf{Y}_1|U_1) \\
& - \mu_0 \lambda_2 h(\mathbf{Z}_1|U_1) + (\mu_2 - \mu_0 \lambda_2) h(\mathbf{Z}_2|U_2) - \mu_0 \lambda_1 h(\mathbf{Y}_2|U_2) \\
& \leq \mu_0 \lambda_1 \left(\frac{1}{2} \log(|2\pi e(\mathbf{S}_1 + \mathbf{N}_1)|) \right) + \frac{1}{2} \log(|2\pi e(\mathbf{S}_2 + \mathbf{N}_2)|) \\
& + \mu_0 \lambda_2 \left(\frac{1}{2} \log(|2\pi e(\mathbf{S}_1 + \mathbf{M}_1)|) \right) + \frac{1}{2} \log(|2\pi e(\mathbf{S}_2 + \mathbf{M}_2)|) \\
& - \mu_1 h(\mathbf{Y}_1|\mathbf{X}_1) - \mu_2 h(\mathbf{Z}_2|\mathbf{X}_2) \\
& + (\mu_1 - \mu_0 \lambda_1) h(\mathbf{Y}_1|U_1) - \mu_0 \lambda_2 h(\mathbf{Z}_1|U_1) \\
& + (\mu_2 - \mu_0 \lambda_2) h(\mathbf{Z}_2|U_2) - \mu_0 \lambda_1 h(\mathbf{Y}_2|U_2).
\end{aligned}$$

To complete the proof, we need to show that

$$\begin{aligned}
& (\mu_1 - \mu_0 \lambda_1) h(\mathbf{Y}_1|U_1) - \mu_0 \lambda_2 h(\mathbf{Z}_1|U_1) - \mu_1 h(\mathbf{Y}_1|\mathbf{X}_1) \\
& \leq \frac{\mu_1 - \mu_0 \lambda_1}{2} \log(|2\pi e(\mathbf{Q}_1^* + \mathbf{N}_1)|) \\
& - \frac{\mu_0 \lambda_2}{2} \log(|2\pi e|\mathbf{Q}_1^* + \mathbf{M}_1|) - \frac{\mu_1}{2} \log(|2\pi e\mathbf{N}_1|), \tag{23}
\end{aligned}$$

$$\begin{aligned}
& (\mu_2 - \mu_0 \lambda_2) h(\mathbf{Z}_2|U_2) - \mu_0 \lambda_1 h(\mathbf{Y}_2|U_2) - \mu_2 h(\mathbf{Z}_2|\mathbf{X}_2) \\
& \leq \frac{\mu_2 - \mu_0 \lambda_2}{2} \log(|2\pi e(\mathbf{Q}_2^* + \mathbf{M}_2)|) \\
& - \frac{\mu_0 \lambda_1}{2} \log(|2\pi e(\mathbf{Q}_2^* + \mathbf{N}_2)|) - \frac{\mu_2}{2} \log(|2\pi e\mathbf{M}_2|), \tag{24}
\end{aligned}$$

where $\mathbf{Q}_1^*, \mathbf{Q}_2^*$ are as specified by Lemma 1. By symmetry, we only need to show that (23) is true.

If $\mathbf{O}_1 = \mathbf{0}$ in Lemma 1, then we must have $\mu_1 - \mu_0 \lambda_1 \geq 0$, and it follows from Theorem 6 that (23) is true. Now consider the case $\mathbf{O}_1 \succneq \mathbf{0}$. If $\mu_1 - \mu_0 \lambda_1 < 0$, then $\mathbf{O}_1 \succneq \mathbf{0}$ since $\mu_1(\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{O}_1 = \mu_0 \lambda_1(\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mu_0 \lambda_2(\mathbf{Q}_1^* + \mathbf{M}_1)^{-1} + \mathbf{O}_1'$, which further implies $\mathbf{Q}_1^* = \mathbf{0}$ due to the fact that $\mathbf{Q}_1^* \cdot \mathbf{O}_1 = \mathbf{0}$. Then it is clear that (23) is true. If $\mu_1 - \mu_0 \lambda_1 = 0$, then again we have either $\mathbf{Q}_1^* = \mathbf{0}$ or $\lambda_2 = 0$. It can be seen that (23) is true in both cases.

The remaining case is $\mu_1 - \mu_0 \lambda_1 > 0$ and $\mathbf{O}_1 \succneq \mathbf{0}$. Define $\tilde{\mathbf{N}}_1 \succneq \mathbf{0}$ such that $(\mu_1 - \mu_0 \lambda_1)(\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1)^{-1} = (\mu_1 - \mu_0 \lambda_1)(\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{O}_1$. It is clear that $\tilde{\mathbf{N}}_1 \preceq \mathbf{N}_1 \preceq \mathbf{M}_1$. Define an enhanced receiver $\tilde{\mathbf{Y}}_1$ associated with $\tilde{\mathbf{N}}_1$. It is important to note that if we replace the overall sub-channel \mathbf{Y}_1 by $\tilde{\mathbf{Y}}_1$, then the rate of the common message will also be affected. Instead, we shall only enhance the component related to the private messages. More specifically, let \mathbf{Y}_1 be physically degraded version of $\tilde{\mathbf{Y}}_1$, then it follows

$$I(\mathbf{X}_1; \mathbf{Y}_1|U_1) \leq I(\mathbf{X}_1; \tilde{\mathbf{Y}}_1, \mathbf{Y}_1|U_1) = I(\mathbf{X}_1; \tilde{\mathbf{Y}}_1|U_1).$$

It follows that we can now bound the left hand side of (23)

as given next

$$\begin{aligned}
& (\mu_1 - \mu_0 \lambda_1) h(\mathbf{Y}_1|U_1) - \mu_0 \lambda_2 h(\mathbf{Z}_1|U_1) - \mu_1 h(\mathbf{Y}_1|\mathbf{X}_1) \\
& \leq \mu_1 I(\mathbf{X}_1; \tilde{\mathbf{Y}}_1|U_1) - \mu_0 \lambda_1 h(\mathbf{Y}_1|U_1) - \mu_0 \lambda_2 h(\mathbf{Z}_1|U_1) \\
& = \mu_1 h(\tilde{\mathbf{Y}}_1|U) - \mu_0 \lambda_1 h(\mathbf{Y}_1|U_1) \\
& \quad - \mu_0 \lambda_2 h(\mathbf{Z}_1|U_1) - \mu_1 h(\tilde{\mathbf{Y}}_1|\mathbf{X}_1) \\
& \stackrel{(a)}{\leq} \frac{\mu_1}{2} \log(|2\pi e(\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1)|) - \frac{\mu_0 \lambda_1}{2} \log(|2\pi e(\mathbf{Q}_1^* + \mathbf{N}_1)|) \\
& \quad - \frac{\mu_0 \lambda_2}{2} \log(|2\pi e(\mathbf{Q}_1^* + \mathbf{M}_1)|) - \frac{\mu_1}{2} \log(|2\pi e(\tilde{\mathbf{N}}_1)|),
\end{aligned}$$

where (a) is due to Theorem 6. Since $\mathbf{Q}_1^* \cdot \mathbf{O}_1 = \mathbf{0}$ and $(\mu_1 - \mu_0 \lambda_1)(\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1)^{-1} = (\mu_1 - \mu_0 \lambda_1)(\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{O}_1$, we have

$$\begin{aligned}
& (\mu_1 - \mu_0 \lambda_1) \mathbf{Q}_1^* \cdot (\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1)^{-1} \\
& = (\mu_1 - \mu_0 \lambda_1) \mathbf{Q}_1^* \cdot (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} + \mathbf{Q}_1^* \cdot \mathbf{O}_1 \\
& = (\mu_1 - \mu_0 \lambda_1) \mathbf{Q}_1^* \cdot (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1},
\end{aligned}$$

i.e., $\mathbf{Q}_1^* \cdot (\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1)^{-1} = \mathbf{Q}_1^* \cdot (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1}$. This implies $\tilde{\mathbf{N}}_1 \cdot (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1} = \mathbf{N}_1 \cdot (\mathbf{Q}_1^* + \mathbf{N}_1)^{-1}$. Therefore, we have

$$\frac{|\mathbf{Q}_1^* + \tilde{\mathbf{N}}_1|}{|\tilde{\mathbf{N}}_1|} = \frac{|\mathbf{Q}_1^* + \mathbf{N}_1|}{|\mathbf{N}_1|}.$$

Thus (23) is proved. Since both regions in the statement of Theorem 1 are convex, and we have essentially shown that their supporting hyperplanes match, the proof is now complete. \blacksquare

The proof of Theorem 2 can be outlined as follows. Due to the single letter characterization of Theorem 4, there exist independent random variables (U_1, \mathbf{X}_1) and (U_2, \mathbf{X}_2) to satisfy the individual antenna power constraints. However, such random variables induce separate covariances $\mathbb{E}(\mathbf{X}_1 \mathbf{X}_1^t)$ and $\mathbb{E}(\mathbf{X}_2 \mathbf{X}_2^t)$, the diagonal elements of which satisfy the individual antenna power constraint. But this rate triple must belong to a certain capacity region $\mathcal{C}(\mathbb{E}(\mathbf{X}_1 \mathbf{X}_1^t), \mathbb{E}(\mathbf{X}_2 \mathbf{X}_2^t))$, again due to Theorem 4. Thus the union of capacity region over all possible covariance pairs $(\mathbf{S}_1, \mathbf{S}_2)$ gives the complete capacity region $\mathcal{C}(P_1, P_2, \dots, P_{t_1+t_2})$. The proof of Theorem 3 follows a similar argument.

VI. CONCLUSION

We considered the problem of broadcasting on reversely degraded Gaussian MIMO channels. The characterization given in [9] is generalized to include channel cost constraints for statistically separable cost functions. By utilizing an alternative characterization, an extremal inequality and the enhancement technique, we showed that Gaussian codebooks are optimal for the reversely degraded Gaussian MIMO channels under various power constraints.

REFERENCES

- [1] P. Bergmans, "A simple converse proof for the broadcast channels with additive white Gaussian noise," *IEEE Trans. Information Theory*, vol. 20, no. 2, pp. 279–280, Mar. 1974.
- [2] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Pered. Inform.*, vol. 10, no. 3, pp. 3–14, Jul.-Sep. 1974; translated *Probl. Inform. Transm.*, vol. 10, no. 3, pp. 185–193, Jul.-Sep. 1974.

- [3] K. Marton, "The capacity region of deterministic broadcast channels," in *Trans. Int. Symp. Inform. Theory*, pp. 243-248, Cachan, France, 1977
- [4] M. S. Pinsker, "The capacity region of noiseless broadcast channels," *Probl. Pered. Inform.*, vol. 14, no. 2, pp. 28-34, Apr.-Jun. 1978; translated *Probl. Inform. Transm.*, vol. 14, no. 2, pp. 97-102, Apr.-Jun. 1978.
- [5] A. A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Information Theory*, vol. 25, no. 2, pp. 166-169, Mar. 1979.
- [6] J. Korner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Information Theory*, vol. 23, no. 1, pp. 60-64, Jan. 1977.
- [7] T. Cover, "Comments on broadcast channels," *IEEE Trans. Information Theory*, vol. 44, no. 6, pp. 2524-2530, Oct. 1998.
- [8] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Information Theory*, vol. 52, no. 9, pp. 3936-3964, Sep. 2006.
- [9] A. A. El Gamal, "The capacity of the product and sum of two reversely degraded broadcast channels," *Probl. Pered. Inform.*, vol. 16, no. 1, pp. 3-23, Jan.-Mar. 1980; translated *Probl. Inform. Transm.*, vol. 16, no. 1, pp. 1-16, Jan.-Mar. 1980.
- [10] H. Weingarten, Y. Steinberg, and S. Shamai, "On the capacity region of the multi-antenna broadcast channel with common messages," *IEEE Int. Symp. Inform. Th.*, pp. 2195-2199, Jul. 2006, Seattle, USA.
- [11] N. Jindal and A. Goldsmith, "Capacity and dirty paper coding for Gaussian broadcast channels with common information," *IEEE Int. Symp. Inform. Th.*, p. 215, Jun.-Jul. 2004, Chicago, USA.
- [12] H. Weingarten, T. Liu, S. Shamai, Y. Steinberg and P. Viswanath, "Capacity region of the degraded MIMO compound broadcast channel," *IEEE Trans. Information Theory*, to appear.
- [13] L. Gropop and D. N. C. Tse, "Fundamental constraints on multicast capacity regions," *IEEE Trans. Information Theory*, submitted for publication. Arxiv:0809.2835v1.
- [14] C. Tian, "Latent capacity region: a case study on symmetric broadcast with common messages," *IEEE Trans. Information Theory*, submitted for publication.
- [15] T. Berger, "Multiterminal source coding," in *Lecture notes at CISM summer school on the information theory approach to communications*, 1977.
- [16] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.