CCIT Report #491 June 2004[1]

The Capacity Region of the Gaussian MIMO Broadcast channel

Hanan Weingarten Yossef Steinberg and Shlomo Shamai (Shitz)

Abstract

The Gaussian MIMO broadcast channel is considered and the dirty paper coding rate region is shown to coincide with the capacity region. To that end, a new notion of an enhanced broadcast channel is introduced and is used jointly with the entropy power inequality, to show that Gaussian coding is optimal for the degraded vector broadcast channel. Furthermore, the capacity region is characterized under a wide range of input constraints, as long as the input has a covariance matrix which is limited to a compact set, accounting, as special cases, the total power and the per-antenna power constraints.

1. Introduction

We consider a Gaussian Multiple Input Multiple Output (MIMO) Broadcast Channel (BC) and find the capacity region for that channel. The transmitter is required to send independent messages to m receivers. We assume that the transmitter has t transmit antennas and each of the users is equipped with r_i , $i = 1, 2, \ldots, m$ receive antennas. Initially, we assume that there is an average total power limitation, P, at the transmitter. However, as will be made clear in the following section, our capacity results can be easily extended to a much broader set of input constraints and in general, we can consider any input constraint such that the input covariance matrix belongs to some compact set of positive semi-definite matrices. The BC is an additive noise channel and each time sample can be represented using the following expression:

$$\mathbf{y}_i = H_i\mathbf{x} + \mathbf{n}_i \ , \qquad i = 1, 2, \ldots, m \tag{1-1}$$

where

- $\mathbf{x}$ is a real input vector of size $t \times 1$. Under an average total power limitation, P, at the transmitter, we require that $E[\mathbf{x}^T\mathbf{x}] \leq P$. Under an input covariance constraint, we will require that $E[\mathbf{x}\mathbf{x}^T] \preceq S$ for some $S \succeq 0$ (where $\prec$, $\preceq$, $\succ$ and $\succeq$ denote partial ordering between symmetric matrices such that if $B \succeq A$, then $(B - A)$ is a positive semi-definite matrix).
- $\mathbf{y}_i$, $i = 1, 2, \ldots, m$ - are real output vectors received by user i. This is a vector of size $r_i \times 1$ (the vectors, $\mathbf{y}_i$, are not necessarily of the same size).
- H_i is a fixed, real gain matrix imposed on user i. This is a matrix of size $r_i \times t$. The gain matrices are fixed and are perfectly known at the transmitter and at all receivers.
- and $\mathbf{n}_i$ is a real Gaussian random vector with zero mean and a covariance matrix $N_i = E[\mathbf{n}_i\mathbf{n}_i^T] \succ 0$. No special structure is imposed on H_i or N_i, except that N_i must be positive definite.

Note that complex MIMO BCs can be easily accommodated by representing all complex vectors and matrices using real vectors and matrices having twice the number of elements, corresponding to real and imaginary entries.

In this paper, we find the capacity region of the BC in expression (1-1) under various input constraints. In particular, we will consider the average total input power and the input covariance constraints. This model and our results are quite relevant to many applications in wireless communications such as cellular systems [15], [20], [13].

In general, the capacity region of the BC is still unknown. There is a single-letter expression for an achievable rate region due to Marton [12] but it is unknown if it coincides with the capacity region. Nevertheless, for some special cases, a single letter formula for the capacity region does exist [7], [11], [8] and coincides with the Marton region [12]. One such case is the degraded BC [7] where the channel input and outputs form a Markov chain. It turns out that when the BC given by (1-1) is scalar ($t = r_1 = \cdots = r_m = 1$), it is a degraded BC. Furthermore, it was shown by Bergmans [1] that for the scalar Gaussian BC, coding using a superposition of random Gaussian codes is optimal. Interestingly, the optimality of Gaussian coding for this case was not deduced directly from the informational formula, as the optimization over all input distributions is not trivial. Instead, the Entropy Power Inequality (EPI) [3], [1] came into bare. Unfortunately, in general, the channel given by expression (1-1) is not degraded. To make matters worse, even when this channel is degraded but not scalar, Bergmans' proof does not directly extend to the vector case. In Section 3 we recount Bermans' proof for the vector channel and show that this extension fails due to Minkowski inequality.

Recent years have seen intensive work on the Gaussian MIMO BC. Caire and Shamai [5] were among the first to pay attention to this channel and were the first to suggest using Dirty Paper Coding (DPC) for transmitting over this channel. DPC is based on costa's [9], [6], [24] results for coding for a channel with both additive Gaussian noise and additive Gaussian interference, where the interference is non-causally known at the transmitter but not at the receiver. Costa observed that under an input power constraint, the effect of the interference can be completely cancelled out. Using this coding technique, each user can pre-code its own information, based on the signals of the users following it [5], [19], [17], [23] (for some arbitrary ordering of the users) and treating their respective signals as non-causally known interference.

Caire and Shamai [5] investigated a two user MIMO BC with an arbitrary number of antennas at the transmitter and one antenna at each of the receivers. For that channel, it was shown through direct calculation that DPC achieves the sum rate capacity (or maximum throughput as it was called in [5]). Hence, it was shown that at least for rate pairs that obtain the channel's sum capacity, DPC is optimal. Their main tool was the Sato upper bound [14] (or the cooperative upper bound).

Following the work by Caire and Shamai, new papers appeared [19], [17], [23] which expanded the maximum throughput claim in [5] to the case of any number of users and arbitrary number of antennas at each receiver. Again, the Sato upper bound played a major role. In [23], Yu and Cioffi construct a coding and decoding scheme for the cooperative channel relying on the ideas of DPC and generalized decision feedback equalization. A different approach is taken in [19], [17]. To overcome the difficulty of expanding the calculations of the maximum throughput in [5] to more than two users, with more than one antenna, the idea of BC MAC duality [19], [10], [17] has been introduced.

Another step towards characterizing the capacity region of the MIMO BC is reported in [18] and [16]. Both works show that if Gaussian coding is optimal for the Gaussian degraded vector BC, then the DPC rate region is also the capacity region. To substantiate this claim, the Sato upper-bound was replaced with the Degraded Same Marginals (DSM) bound in [18]. Furthermore, the authors conjecture that indeed, Gaussian coding is optimal for the Gaussian degraded vector BC. However, as indicated above, the proof of this conjecture is not trivial, as Bergmans' [1] proof can not be directly applied.

In a conference version of this work [21], we reported a proof of this conjecture and thus, along with the DSM bound [18], [16], we have shown that the DPC rate region is indeed the capacity region of the MIMO BC. Here, we provide a more cohesive view and are able to derive the results on first principle properties and we do not hinge on any of the above existing results (the MAC-BC duality [19], [17] and the DSM upper bound [18], [16]). This not only provides a more complete and self-contained view, but it significantly enhances the insight into this problem and its associated properties.

The main contribution of this work is the notion of an *enhanced channel*. The introduction of an enhanced channel allowed us to use the entropy power inequality, as done for the scalar case by Bergmans', to prove that Gaussian coding is optimal for the vector degraded BC, despite the Minkowski inequality. We show that instead of proving the optimality of Gaussian coding for the degraded vector channel, we can prove it for a set of enhanced degraded vector channels, for which, unlike the original degraded vector channel, we can directly extend Bergmans' [1] proof, and in this setting, the Minkowski inequality is satisfied with equality.

Another unique result is our ability to characterize the capacity region of the MIMO BC under channel input constraints other than the total power constraint. In particular, in Section 2, we show that we can characterize the capacity region of the MIMO BC which input covariance matrix is limited to some compact set. This parallels a result reported in [22] for the sum capacity of the MIMO BC under per-antenna power constraints.

In Section 2 we give preliminary results and show that instead of characterizing the capacity region of the MIMO BC under a total power constraint, we can characterize the capacity region under an input covariance matrix constraint. In Section 3, we find the capacity region of a sub class of the degraded vector BC, which we refer to as the aligned degraded MIMO BC. In Section 4 we broaden our scope to a sub-class of the MIMO BC similar to that treated in Section 3 but which is not degraded. Finally, in Section 5 we find the capacity region of the general Gaussian MIMO BC as given in (1-1).

2. Preliminaries

2.1 Subclasses of the Gaussian MIMO BC (the ADBC and AMBC)

Equation (1-1) represents the Gaussian MIMO broadcast channel. We shall refer to that channel as the General MIMO BC (GMBC). However, we will initially find it simpler to contend with two subclasses of this channel and then broaden the scope of our results and write the capacity for the GMBC.

The first subclass we will deal with is the Aligned Degraded MIMO Broadcast Channel (ADBC). We say that the MIMO BC is aligned if the number of transmit antennas is equal to the number of receive antennas at each of the receivers ($t = r_1 = \cdots = r_m$) and if the gain matrices are all identity matrices ($H_1 = \cdots = H_m = I$). Furthermore, we require that this subclass will be degraded and assume that the additive noise vectors covariance matrices at each of the receivers are ordered such that $0 \prec N_1 \preceq N_2 \preceq \cdots \preceq N_m$ (where $\prec$, $\preceq$, $\succ$ and $\succeq$ denote partial ordering between symmetric matrices such that if $B \succeq A$, then $(B - A)$ is a positive semi-definite matrix). Taking into account the fact that the capacity region of BCs (in general) depend only on the marginal distributions, $P_{Y_i|X}$, we may assume without loss of generality that a time sample of an ADBC is represented by the following expression:

$$\mathbf{y}_i = \mathbf{x} + \sum_{j=1}^{i} \tilde{\mathbf{n}}_j \ , \qquad i = 1, 2, \ldots, m \tag{2-1}$$

where $\mathbf{y}_i$ and $\mathbf{x}$ are real vectors of size $t \times 1$ and where $\tilde{\mathbf{n}}_i$, $i = 1, \ldots, m$ are independent and memoryless real Gaussian noise increments such that $\tilde{N}_i = E[\tilde{\mathbf{n}}_i \tilde{\mathbf{n}}_i^T] = N_i - N_{i-1}$ (where we define $N_0 = 0$).

The second subclass we address is a generalization of the ADBC which we will refer to as the Aligned MIMO BC (AMBC). The AMBC is also aligned (that is, $t = r_1 = \cdots = r_m$ and $H_1 = \cdots = H_m = I$), but we do not require that the channel will be degraded. In other words, we no longer require that the additive noise vectors covariance matrices, N_i, will exhibit any order between them. A time sample of an AMBC is represented by the following expression:

$$\mathbf{y}_i = \mathbf{x} + \mathbf{n}_i \ , \qquad i = 1, 2, \ldots, m \tag{2-2}$$

where $\mathbf{y}_i$ and $\mathbf{x}$ are real vectors of size $t \times 1$ and where $\mathbf{n}_i$, $i = 1, \ldots, m$ are memoryless real

Gaussian noise vectors such that $E[\mathbf{n}_i\mathbf{n}_i^T] = N_i \succ 0$.

In the case where the gain matrices of the GMBC (the channel given in (1-1)), H_i, are square and invertible, it is readily shown that the capacity region of the GMBC can be inferred from that of the AMBC by multiplying each of the channel outputs, $\mathbf{y}_i$, by H_i^{-1}. However, a problem arises when the gain matrices are no longer square or invertible. In other words, when the number of transmit and receive antennas are not the same. In Section 5 (proof of Theorem 5.1), we show that the capacity region of the GMBC with non-square or non-invertible gain matrices can be obtained from a limit process on the capacity region of an AMBC along the following steps. First, we decompose the channel using singular value decomposition (SVD), into a channel with square gain matrices. Then, we add a small perturbation to some of the gain entries and take a limit process on those entries. Thus, we simulate the fact that the gain matrices of the original channel were not square or not invertible. After adding the perturbations, the gain matrices are invertible and an equivalent AMBC can be readily established. Therefore, the greater part of this paper (Sections 3 and 4) will be devoted to characterizing the capacity regions of the ADBC and AMBC.

2.2 The covariance matrix constraint

2.2.1) Substituting the total power constraint with the matrix covariance constraint

As mentioned in the introduction, our final goal is to give a characterization of the capacity region of the GMBC under a total power constraint, P. Nonetheless, our initial characterization of the capacity region will be given for a covariance matrix constraint, $S \succeq 0$, such that $E[\mathbf{x}\mathbf{x}^T] \preceq S$. For a given total power constraint $E[\mathbf{x}^T\mathbf{x}] \leq P$, we denote by $\mathcal{C}(P, N_{1\ldots m}, H_{1\ldots m})$ the capacity region of the GMBC given in (1-1) under a total power constraint, P. For a given covariance matrix constraint S, we denote by $\mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$ the capacity region of the same broadcast channel under a covariance matrix constraint, S.

The justification for characterizing the capacity regions under a covariance matrix constraint instead of a total power constraint is made clear by the next proposition. However, prior to giving this proposition we must first give the following definition:

Definition 2.1 (Contiguity of the capacity region with respect to S) We say that the capacity region, $\mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$, is contiguous w.r.t S if for every $\epsilon > 0$, we can find a $\delta > 0$ such that every ϵ-ball around a rate vector $\overline{R} \in \mathcal{C}(S + \delta' I, N_{1\ldots m}, H_{1\ldots m})$ also contains a rate vector $\overline{R}^* \in \mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$ for all $0 \leq \delta' \leq \delta$.

As one might expect, the capacity regions that will be defined in the following sections, will all be contiguous in S.

Proposition 2.1: Assume that P is a non-negative scalar and S is a positive semi-definite matrix. If $\mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$ is contiguous in S, then

$$\mathcal{C}(P, N_{1\ldots m}, H_{1\ldots m}) = \bigcup_{\mathrm{tr}\{S\} \preceq P} \mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$$

Proof: We will prove that every rate vector $\overline{R} = (R_1, \ldots, R_m) \in \mathcal{C}(P, N_{1\ldots m}, H_{1\ldots m})$ also lies in $\mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$ for some S such that $\mathrm{tr}\{S\} \preceq P$. We use $\mathbf{C}(n, \overline{R}, \epsilon)$ to denote a codebook that maps a set of m message indexes, $W_1, \ldots, W_m$, $W_i \in \{1, \ldots, e^{nR_i}\}$, $i = 1, \ldots, m$, onto a channel input word (a real matrix of size $t \times n$) and allows the ML decoder at each of the receivers to decode the appropriate message index with an average probability of decoding error, no greater than ϵ. If indeed $\overline{R} \in \mathcal{C}(P, N_{1\ldots m}, H_{1\ldots m})$, then there exists an infinite sequence of codebooks, $\mathbf{C}(n_i, \overline{R}, \epsilon_i)$, $i = 1, 2, \ldots \infty$, with increasing lengths, n_i, rate vectors $\overline{R}$ and decreasing probabilities of error, ϵ_i, such that $\epsilon_i \to 0$ and $n_i \to \infty$ as $i \to \infty$.

Denote by S_i, the average codeword covariance matrix of the i'th codebook such that $S_i = \frac{1}{n_i \cdot e^{n_i \sum R_i}} \sum_{\overline{\overline{X}} \in \mathbf{C}(n_i, \overline{R}, \epsilon_i)} \overline{\overline{X}} \cdot \overline{\overline{X}}^T$. As a result of the total power constraint, P, it is clear that $\mathrm{tr}\{S_i\} \leq P$, $\forall i$. We now make use of the fact that the set of positive semi-definite matrices, $\mathcal{S}(P) = \{S | S \succeq 0, \mathrm{tr}\{S\} \leq P\}$, is compact and therefore, for any infinite sequence of points in $\mathcal{S}(P)$ there must be a subsequence that converges to a point $S_0 \in \mathcal{S}(P)$. Hence, for any arbitrarily small $\delta > 0$, we can find an increasing subsequence, $i(k)$, such that $S_{i(k)} \preceq S_0 + \delta I$, $\forall k > 0$.

In other words, If $\overline{R} \in \mathcal{C}(P, N_{1\dots m}, H_{1\dots m})$, for any $\delta > 0$, we can find a sequence of codebooks, $\mathbf{C}(n_k, \overline{R}, \epsilon_k)$, which achieve arbitrarily small error probabilities and follow a covariance matrix constraint, $S_0 + \delta I$, for some $S_0 \in \mathcal{S}(P)$. Therefore, for every $\delta > 0$, $\overline{R} \in \mathcal{C}(S_0 + \delta I, H_i, N_i)$. By the contiguity of $\mathcal{C}(S, N_{1\dots m}, H_{1\dots m})$ with respect to S, we conclude that every ϵ-ball around $\overline{R}$ contains vectors which lie in $\mathcal{C}(S_0, H_i, N_i)$ and therefore, $\overline{R}$ must be a convergence point of $\mathcal{C}(S_0, H_i, N_i)$. As $\mathcal{C}(S_0, H_i, N_i)$ is a closed set by definition, we conclude that $\overline{R} \in \mathcal{C}(S_0, H_i, N_i)$. ■

With minor modifications to the above proof, Proposition 2.1 can be extended to a more general constraint on the channel inputs, as stated bellow:

Corollary 2.1: Assume that there exists a power constraint on the average codeword covariance matrix, $S = \frac{1}{n \cdot e^{n \sum R_i}} \sum_{\overline{\overline{X}} \in \mathbf{C}(n, \overline{R}, \epsilon)} \overline{\overline{X}} \cdot \overline{\overline{X}}^T$, such that $S \in \mathcal{S}$ and where $\mathcal{S}$ is a compact set. If $\mathcal{C}(S, N_{1\dots m}, H_{1\dots m})$ is contiguous in S, then the capacity region of the GMBC under this power constraint is given by:

$$\mathcal{C}(\mathcal{S}, H_i, N_i) = \bigcup_{S \in \mathcal{S}} \mathcal{C}(S, N_{1\dots m}, H_{1\dots m})$$

2.2.2) The capacity region under a non-invertible covariance matrix constraint

We differentiate between the case where the covariance matrix constraint, S, is strictly positive definite (and hence invertible and full ranked), and the case where it is positive semi-definite but non-invertible, $|S| = 0$. It turns out that for an aligned MIMO BC (either an ADBC or an AMBC) with a non-invertible covariance matrix constraint, $|S| = 0$, we can define an equivalent aligned MIMO BC (either an ADBC or an AMBC), with a smaller number of transmit and receive antennas and with a covariance matrix constraint which is strictly positive definite. Thus, when proving the converse of the capacity regions of the ADBC and AMBC, we will need only to concentrate on the cases where S is strictly positive. A formal presentation of the above argument is given in the following proposition:

Proposition 2.2: Consider an AMBC with t transmit antennas, noise covariance matrices, $N_1, \ldots, N_m$, and a covariance matrix constraint, S, such that $\text{rank}(S) = r_S < t$. Then,

1. There exists an AMBC, with r_S transmit antennas (and r_S receive antennas at each of the receivers) and with noise covariance matrices $\hat{N}_1, \ldots, \hat{N}_m$, which has the same capacity region under a covariance matrix constraint, $\hat{S}$, of rank r_S. Furthermore, if $N_1 \preceq \cdots \preceq N_m$ (the channel is an ADBC), then $\hat{N}_1 \preceq \cdots \preceq \hat{N}_m$ (i.e. the equivalent channel is also an ADBC).
2. In the equivalent channel we have:

$$\hat{S} = E' \Lambda_S E'^T \quad \text{and} \quad \hat{N}_i = N_i^C - \left(N_i^B\right)^T \left(N_i^A\right)^{-1} \left(N_i^B\right), \; i = 1, 2, \ldots, m$$

where Λ_S and U_S are defined as the eigenvalues and eigenvectors matrices (unitary matrices) of $S = U_S \Lambda_S U_S^T$ such that Λ_S has all its non-zero values on the bottom right of the diagonal. N_i^A, N_i^B and N_i^C are matrices of sizes $(t - r_S) \times (t - r_S)$, $(t - r_S) \times r_S$ and $r_S \times r_S$ such that $U_S^T N_i U_S = \begin{pmatrix} N_i^A & N_i^B \\ (N_i^B)^T & N_i^C \end{pmatrix}$ and $E' = \left[0_{r_S \times (t - r_S)} \;\; I_{r_S \times r_S}\right]$.

The proof is deferred to Appendix II.

3. The capacity region of the ADBC

In this section we characterize the capacity region of the ADBC. As mentioned in the introduction, even though this channel is degraded, and we have a single letter formula for the capacity region of this channel, proving that Gaussian inputs are optimal is not trivial. In the following subsections we will state and prove our result for the capacity region of the ADBC.

3.1 ADBC - Main Result

We begin by defining the achievable rate region due to Gaussian coding under a covariance matrix constraint, $S \succeq 0$. Note that as the channel is degraded, there is no point in using Dirty Paper Coding (DPC) [6], [5], [23], [19], [17].

It is well known that for any covariance matrix input constraint, S, and a set of semi-definite matrices $B_i \succeq 0,\ i = 1, \dots m$ such that $\sum_{i=1}^{m} B_i \preceq S$, it is possible to achieve the following rates:

$$R_k \leq R_k^G(B_{1\dots m}, \tilde{N}_{1\dots m})\ , \qquad \forall k \in 1, \dots, m \tag{3-1}$$

where

$$\begin{aligned} R_1^G(B_{1\dots m}, \tilde{N}_{1\dots m}) &= \frac{1}{2} \log |\tilde{N}_1^{-1}(B_1 + \tilde{N}_1)|, \\ R_k^G(B_{1\dots m}, \tilde{N}_{1\dots m}) &= \frac{1}{2} \log \frac{\left|\left(\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} \tilde{N}_i\right)\right|}{\left|\left(\sum_{i=1}^{k-1} B_i + \sum_{i=1}^{k} \tilde{N}_i\right)\right|},\ k = 2, 3, \dots, m \end{aligned} \tag{3-2}$$

The coding scheme that achieves the above rates uses a super-position of Gaussian codes with covariance matrices, B_i, and successive decoding at the receivers. The Gaussian rate region is defined as follows:

Definition 3.1 (Gaussian rate region of an ADBC) Let S be a positive semi-definite matrix. Then, the Gaussian rate region of an ADBC under a covariance matrix constraint, S, is given by:

$$\mathcal{R}^G(S, \tilde{N}_{1\dots m}) = \left\{ \begin{array}{c|} \left(R_1^G(B_{1\dots m}, \tilde{N}_{1\dots m}), \dots, R_m^G(B_{1\dots m}, \tilde{N}_{1\dots m})\right) \\ \text{s.t.} \quad S - \sum_{i=1}^{m} B_i \succeq 0\ , \quad B_i \succeq 0 \ \ \forall i = 1, \dots, m \end{array} \right\} \tag{3-3}$$

We now give the main result of this section:

Theorem 3.1: Let $\mathcal{C}(S, \tilde{N}_{1\dots m})$ denote the capacity region of the ADBC under a covariance matrix constraint, $S \succeq 0$, then $\mathcal{C}(S, \tilde{N}_{1\dots m}) = \mathcal{R}^G(S, \tilde{N}_{1\dots m})$.

Note that $\mathcal{C}(S, \tilde{N}_{1\ldots m})$ is an operational definition of a rate region while $\mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ is a functional definition of a rate region.

As $\mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ is contiguous w.r.t S (for proof, see Proposition III.1 in Appendix III), the following corollary follows immediately by Proposition 2.1.

Corollary 3.1: Let $\mathcal{C}(P, \tilde{N}_{1\ldots m})$ denote the capacity region of the ADBC under a total power constraint, $P \geq 0$, then

$$\mathcal{C}(P, \tilde{N}_{1\ldots m}) = \bigcup_{\substack{S \succeq 0 \\ \text{s.t.} \\ \text{tr}\{S\} \leq P}} \mathcal{R}^G(S, \tilde{N}_{1\ldots m})$$

3.2 Direct application of Bergmans' proof to the ADBC and its pitfalls

Before plunging into the proof of Theorem 3.1, we explain why Bergman's [1] proof for the scalar Gaussian BC, does not directly extend to the MIMO channel. This subsection is intended to give the reader an idea of why and where the direct application of Bergmans' proof fails and how we intend to overcome this problem. As this subsection is only intended to motivate the reader, only a sketch (with no proofs) of Bergman's result will be given, in contrast to the rest of this paper.

We consider a t-antenna ADBC with two users and noise increment covariance matrices $\tilde{N}_1, \tilde{N}_2$ and with a covariance matrix constraint, S. We wish to show that $\mathcal{R}^G(S, \tilde{N}_{1,2})$ is also the capacity region.

Assume, in contrast, that there is an achievable rate pair, (R_1, R_2), which lies outside $\mathcal{R}^G(S, \tilde{N}_{1,2})$. Then, we can find a set of matrices B_1 and $B_2 = S - B_1$ such that

$$R_1 \geq R_1^G(B_{1,2}, \tilde{N}_{1,2}) \quad \text{and} \quad R_2 > R_2^G(B_{1,2}, \tilde{N}_{1,2}). \tag{3-4}$$

Let W_1 and W_2 denote the massage indexes of the users. Furthermore, let $\overline{\overline{X}}$, $\overline{\overline{Y}}_1$ and $\overline{\overline{Y}}_2$ denote the channel input and channel outputs matrices over a block of n samples. Then, by the fact that W_1 and W_2 are independent and by Fano's inequality, we may write:

$$R_1 \leq \frac{1}{n} I(W_1; \overline{\overline{Y}}_1 | W_2) \quad \text{and} \quad R_2 \leq \frac{1}{n} I(W_2; \overline{\overline{Y}}_2) \tag{3-5}$$

where, for the sake of brevity, we ignored the contribution of the decoding errors to the right hand side of the equations. In the rigorous proof given in the following subsections we will not ignore these contributions.

Therefore, by Fano's inequality (3-5) and by (3-4), we may write:

$$\begin{aligned}\frac{1}{n} I(W_1; \bar{\bar{Y}}_1 | W_2) &= \frac{1}{n} H(\bar{\bar{Y}}_1 | W_2) - \frac{1}{n} H(\bar{\bar{Y}}_1 | W_1, W_2) = \frac{1}{n} H(\bar{\bar{Y}}_1 | W_2) - \frac{1}{2} \log\left(\left|2\pi e \tilde{N}_1\right|\right) \\ &\geq R_1 \geq R_1^G(B_{1,2}, \tilde{N}_{1,2}) = \frac{1}{2} \log\left(\left|2\pi e \left(B_1 + \tilde{N}_1\right)\right|\right) - \frac{1}{2} \log\left(\left|2\pi e \tilde{N}_1\right|\right)\end{aligned}$$

and hence,

$$\frac{1}{n} H(\bar{\bar{Y}}_1 | W_2) \geq \frac{1}{2} \log\left(\left|2\pi e \left(B_1 + \tilde{N}_1\right)\right|\right)$$

Next, we lower bound $\frac{1}{n} H(\bar{\bar{Y}}_2 | W_2)$ using the Entropy Power Inequality. The EPI lower bounds the entropy of the sum of two independent vectors with a function of the entropies of each of the vectors. Bergmans slightly generalized the EPI to allow conditioning on the message indexes (see Corollary I.1 in Appendix I). As $\bar{\bar{Y}}_2 = \bar{\bar{Y}}_1 + \bar{\bar{Z}}_2$, where $\bar{\bar{Z}}_2$ is a $t \times n$ random matrix which columns are independent Gaussian random vectors with covariance matrices, $\tilde{N}_2$, we can use Bergmans' version of the EPI (Corollary I.1) to bound $\frac{1}{n} H(\bar{\bar{Y}}_2 | W_2)$, in the following manner:

$$\begin{aligned}\frac{1}{t \cdot n} H(\bar{\bar{Y}}_2 | W_2) = \frac{1}{t \cdot n} H(\bar{\bar{Y}}_1 + \bar{\bar{Z}}_2 | W_2) &\geq \frac{1}{2} \log\left(e^{\frac{2}{t \cdot n} H(\bar{\bar{Y}}_1 | W_2)} + e^{\frac{2}{t \cdot n} H(\bar{\bar{Z}}_2)}\right) \\ &\geq \frac{1}{2} \log\left(\left|2\pi e \left(B_1 + \tilde{N}_1\right)\right|^{\frac{1}{t}} + |2\pi e \tilde{N}_2|^{\frac{1}{t}}\right)\end{aligned}$$

As $|K_1|^{\frac{1}{t}} + |K_2|^{\frac{1}{t}} \leq |K_1 + K_2|^{\frac{1}{t}}$ for any $t \times t$ positive semi-definite matrices, K_1, K_2 (Minkowski's inequality - see Proposition I.5), we have

$$\frac{1}{t \cdot n} H(\bar{\bar{Y}}_2 | W_2) \geq \frac{1}{2} \log\left(\left|2\pi e \left(B_1 + \tilde{N}_1\right)\right|^{\frac{1}{t}} + |2\pi e \tilde{N}_2|^{\frac{1}{t}}\right) \leq \frac{1}{2} \log\left(\left|2\pi e \left(B_1 + \tilde{N}_1 + \tilde{N}_2\right)\right|^{\frac{1}{t}}\right)$$

with equality, if and only if $B_1 + \tilde{N}_1$ and $\tilde{N}_2$ are proportional.

If $B_1 + \tilde{N}_1$ is proportional to $\tilde{N}_2$, we can rewrite the above result such that

$$\frac{1}{n}H(\bar{\bar{Y}}_2|W_2) \geq \frac{1}{2}\log\left(\left|2\pi e\left(B_1 + \tilde{N}_1 + \tilde{N}_2\right)\right|\right) \tag{3-6}$$

and by (3-5), (3-4) and (3-6), we can write:

$$\begin{aligned}\frac{1}{n}H(\bar{\bar{Y}}_2) &= \frac{1}{n}I(W_2;\bar{\bar{Y}}_2) + \frac{1}{n}H(\bar{\bar{Y}}_2|W_2) \geq R_2 + \frac{1}{n}H(\bar{\bar{Y}}_2|W_2) \\ &> R_2^G(B_{1,2}, \tilde{N}_{1,2}) + \frac{1}{n}H(\bar{\bar{Y}}_2|W_2) \\ &= \frac{1}{2}\log\left(\left|2\pi e\left(S + \tilde{N}_1 + \tilde{N}_2\right)\right|\right) - \frac{1}{2}\log\left(\left|2\pi e(B_1 + \tilde{N}_1 + \tilde{N}_2)\right|\right) + \frac{1}{n}H(\bar{\bar{Y}}_2|W_2) \\ &\geq \frac{1}{2}\log\left(\left|2\pi e\left(S + \tilde{N}_1 + \tilde{N}_2\right)\right|\right)\end{aligned}$$

However, the above result contradicts the upper bound on the entropy of a power limited random variable. Therefore, if we can find matrices B_1 and $B_2 = S - B_1$ such that $B_1 + \tilde{N}_1$ is proportional to $\tilde{N}_2$ and such that $R_1^G \leq R_1$ and $R_2^G < R_2$, then (R_1, R_2) can not be an achievable pair. Yet, we can't always find such matrices, B_1 and B_2.

In the scalar case, $B_1 + \tilde{N}_1$ is always proportional to $\tilde{N}_2$ and therefore, by Bergmans' proof we see that all points that lie outside the Gaussian rate region can not be attained. Yet, for the MIMO case this is no longer true and therefore, we can not directly apply Bergmans' proof in the MIMO case.

In order to circumvent this problem, we will introduce a new ADBC that we will refer to as the *enhanced channel*. For every rate pair (R_1, R_2) which lies outside the Gaussian rate region of the original channel, we will define a different enhanced channel. The enhanced channel will be defined such that its capacity region will contain that of the original channel (hence the name) and such that (R_1, R_2) will also lie outside the Gaussian region of the enhanced channel. Yet, we will show that for the enhanced channel, we can find B_1 and B_2 such that the proportionality condition will hold. Therefore, we will be able to show that every rate pair, (R_1, R_2), that lies outside the Gaussian rate region, is not achievable in its respective enhanced channel, and therefore, neither in the original channel.

3.3 ADBC - Definitions and Observations

Before turning to prove Theorem 3.1, we will first need to give some definitions and intermediate results. We begin with the definition of an optimal Gaussian rate vector and solution point.

Definition 3.2: We say that the rate vector $\overline{R} = (R_1, \ldots, R_m)$ is an *optimal Gaussian rate vector* under a covariance matrix constraint, S, if $\overline{R} \in \mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ and if there is **no** other rate vector $\overline{R}^* \in \mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ such that $R_i^* \geq R_i$, $\forall i = 1, \ldots, m$ and such that at least one of the inequalities is strict. We say that the set of positive semi-definite matrices $B_1, B_2, \ldots, B_m$ such that $\sum_{i=1}^m B_i \preceq S$, is an *optimal Gaussian solution point* if $\left(R_1^G(B_{1\ldots m}, \tilde{N}_{1\ldots m}), R_2^G(B_{1\ldots m}, \tilde{N}_{1\ldots m}), \ldots, R_m^G(B_{1\ldots m}, \tilde{N}_{1\ldots m})\right)$ is an optimal Gaussian rate vector.

The following proposition allows us to associate points that do not lie in the Gaussian rate region with optimal Gaussian rate vectors.

Proposition 3.1: The following statements hold.

1. The set of boundary points of $\mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ coincides with the set of optimal Gaussian rate vectors.
2. Let $\overline{R} = (R_1, R_2, \ldots, R_m)$ be a rate vector satisfying:

$$\overline{R} \notin \mathcal{R}^G(S, \tilde{N}_{1\ldots m})$$

$$R_i \geq 0,\ \forall i = 1, \ldots, m-1$$

$$R_m > 0.$$

Then there is a strictly positive scalar, $b > 0$, and an optimal Gaussian solution point

$B_1^*, \ldots, B_m^*$ such that

$$\begin{aligned} R_i &\geq R_i^G(B_{1\ldots m}^*, \tilde{N}_{1\ldots m}), \quad i = 1, \ldots, m-1 \\ R_m &\geq R_m^G(B_{1\ldots m}^*, \tilde{N}_{1\ldots m}) + b \end{aligned} \tag{3-7}$$

The proof of the proposition is deferred to Appendix IV. It should be noted that when we refer to the boundary points of $\mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ we refer to those points that do not lie strictly on one of the axes or to be more precise, we refer to the non-negative (element-wise) boundary points of the set $\mathcal{R}^G(S, \tilde{N}_{1\ldots m}) + \mathbb{R}_-^m$ (where $\mathbb{R}_-^m = \{(R_1, \ldots, R_m) \in \mathcal{R}^m \mid R_i \leq 0, \forall i = 1, \ldots, m\}$ and where $\mathcal{A} + \mathcal{B}$ is the set of all possible additions between vectors in $\mathcal{A}$ and vectors $\mathcal{B}$).

To get a better understanding of the first statement of Proposition 3.1, we can consider the two user channel. The first statement implies that **no** section of the boundary of the Gaussian rate region lies parallel to either the horizontal or vertical axes.

In general, there is no known closed form solution for the optimal Gaussian solution points. However, we can state the following simple result:

Proposition 3.2: Let $B_1^*, \ldots, B_m^*$ be an optimal Gaussian solution point under a covariance matrix constraint, $S \succeq 0$. Then, $S = \sum_{i=1}^m B_i^*$.

Proof: We note that of all users, only the Gaussian achievable rate of user m, $R_m^G(B_{1\ldots m}, \tilde{N}_{1\ldots m})$, is a function of B_m such that

$$R_m^G(B_{1\ldots m}, \tilde{N}_{1\ldots m}) = \frac{1}{2} \log \frac{\left| B_m + \sum_{i=1}^{m-1} B_i + \sum_{i=1}^{m-1} \tilde{N}_i \right|}{\left| \sum_{i=1}^{m-1} B_i + \sum_{i=1}^{m-1} \tilde{N}_i \right|}.$$

Therefore, as $B_1^*, \ldots, B_m^*$ is an optimal Gaussian solution point, given the set of $m-1$ matrices $B_i = B_i^*$, $i = 1, \ldots, m-1$, B_m^* is the choice of B_m which maximizes $R_m^G(B_{1\ldots m}, \tilde{N}_{1\ldots m})$. However, as $B_m \preceq S - \sum_{i=1}^{m-1} B_i^*$ and as $|B| > |A|$ when $B \succeq A \succ 0$ and $B \neq A$ (see Proposition I.2), R_m^G is maximized when $B_m = S - \sum_{i=1}^{m-1} B_i^*$. ■

Next, we introduce the notion of the enhanced channel. This definition and the observation following it, are the main contributions of this paper.

Definition 3.3: (Enhanced Channel) We say that an ADBC with noise increment covariance matrices $(\tilde{N}'_1, \tilde{N}'_2, \dots, \tilde{N}'_m)$ is an enhanced version of another ADBC with noise increment covariance matrices $(\tilde{N}_1, \tilde{N}_2, \dots, \tilde{N}_m)$ if

$$\sum_{j=1}^{i} \tilde{N}'_j \preceq \sum_{j=1}^{i} \tilde{N}_j \quad , \qquad \forall i = 1, \dots, m$$

Similarly, we say that an AMBC with noise covariance matrices $(N'_1, N'_2, \dots, N'_m)$ is an enhanced version of another AMBC with noise covariance matrices $(N_1, N_2, \dots, N_m)$ if

$$N'_i \preceq N_i \ , \qquad \forall i = 1, \dots, m$$

Note that our definition holds for both AMBCs and ADBCs. Clearly, the capacity region of the original channel is contained within that of the enhanced channel. Furthermore, as $\frac{|A+B+\Delta|}{|B+\Delta|} \leq \frac{|A+B|}{|B|}$ when $A \succeq 0$, $\Delta \succeq 0$ and $B \succ 0$ (Proposition I.4), $R_i^G(B_{1\dots m}, \tilde{N}_{1\dots m}) \leq R_i^G(B_{1\dots m}, \tilde{N}'_{1\dots m})$ $\forall i$ and therefore, $\mathcal{R}^G(S, \tilde{N}_{1\dots m}) \subseteq \mathcal{R}^G(S, \tilde{N}'_{1\dots m})$.

We can now state a crucial observation, connecting between the definition of the optimal Gaussian rate vector and the enhanced channel.

Theorem 3.2: Consider an ADBC with positive semi-definite noise increment covariance matrices $(\tilde{N}_1, \dots, \tilde{N}_m)$ such that $\tilde{N}_1 \succ 0$. Let $B_1, \dots, B_m$ be an optimal Gaussian solution point under an average transmit covariance matrix constraint, $S \succ 0$. Then, there exists an enhanced ADBC with noise increment covariances $(\tilde{N}'_1, \dots, \tilde{N}'_m)$ such that the following properties hold:

1. *Enhanced channel:*

$$0 \prec \sum_{i=1}^{k} \tilde{N}'_i \preceq \sum_{i=1}^{k} \tilde{N}_i, \qquad \forall k = 1, \dots, m \tag{3-8}$$

2. *Proportionality:* There exit $\alpha_k \geq 0$ $k = 1, \dots, m-1$ such that

$$\alpha_k \left(\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} \tilde{N}'_i \right) = \tilde{N}'_{k+1}, \qquad \forall k = 1, \dots, m-1 \tag{3-9}$$

3. *Rate preservation and Optimality preservation:*

$$R_k^G(B_{1\dots m}, \tilde{N}_{1\dots m}) = R_k^G(B_{1\dots m}, \tilde{N}'_{1\dots m}), \qquad \forall k = 1, \dots, m \tag{3-10}$$

and $B_1, \dots, B_m$ is an optimal Gaussian solution point for the enhanced channel as well.

The proof is deferred to Subsection 3.5.

In words, the above theorem states that for every optimal Gaussian rate vector, $(R_1^O, \dots, R_m^O)$, we can find an enhanced channel whose Gaussian rate region upper bounds $R^G(S, \tilde{N}_{1\dots m})$ but at the same time is tangential to $(R_1^O, \dots, R_m^O)$ (it might be tangential to more than a single point). Moreover, at the point where the two regions intersect, a unique property of proportionality holds (3-9). An illustrative example appears in Subsection 3.6.

In the next subsection, we will prove Theorem 3.1 using the above observation. Our approach will be similar to Bergmans' but this time we will be able to circumvent the pitfalls that we encountered in the direct application of Bergmans' proof (Subsection 3.2) by applying his proof to the enhanced channel, and utilizing the proportionality property which holds for that channel. That is, we will no longer be limited by the Minkowski inequality since, due to the proportionality property, it will hold with equality for the enhanced channel.

3.4 Proof of Theorem 3.1

Proof: As $\mathcal{R}^G(S, \tilde{N}_{1\dots m})$ is a set of achievable rates, then $\mathcal{R}^G(S, \tilde{N}_{1\dots m}) \subseteq \mathcal{C}(S, \tilde{N}_{1\dots m})$. Therefore, we need to show that $\mathcal{C}(S, \tilde{N}_{1\dots m}) \subseteq \mathcal{R}^G(S, \tilde{N}_{1\dots m})$. We will treat separately the cases where the covariance matrix constraint, S, is strictly positive definite, $S \succ 0$, and the case where S is positive semi-definite ($S \succeq 0$) such that $|S| = 0$. We will first consider the case, $S \succ 0$.

<u>$S \succ 0$:</u>

We shall use a contradiction argument and assume that there exists an achievable rate vector $\overline{R} = (R_1, \dots, R_m) \notin \mathcal{R}^G(S, \tilde{N}_{1\dots m})$. We will initially assume that $R_i > 0$, $\forall i = 1 \dots m$ and later, we will use a simple argument to augment the proof for all non-negative rates $R_i \geq 0$, $\forall i = 1 \dots m$.

$\underline{S \succ 0,\ R_i > 0\,\forall i:}$

Since $\overline{R} \notin \mathcal{R}^G(S, \tilde{N}_{1\dots m})$ and $R_m > 0$ (by our assumption), we know by Proposition 3.1 that there exists an optimal Gaussian solution point, $B_1^*, \dots, B_m^*$, such that

$$R_1 \geq R_1^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}), \quad i = 1, \dots, m-1$$
$$R_m \geq R_m^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}) + b$$

for some $b > 0$. Since we assume that $S \succ 0$ we know by Theorem 3.2 that for every optimal Gaussian solution point, $B_1, \dots, B_m$, there exists an enhanced ADBC, $\tilde{N}_1', \dots, \tilde{N}_m'$, such that the proportionality and rate preservation properties hold. By the rate preservation property we have $R_i^G(B_{1\dots m}, \tilde{N}_{1\dots m}) = R_i^{G'}(B_{1\dots m}, \tilde{N}_{1\dots m})$. Therefore, we can rewrite the above expression as follows,

$$\begin{aligned} R_1 &\geq R_1^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}) = R_1^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}') \\ &\vdots \\ R_{m-1} &\geq R_{m-1}^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}) = R_{m-1}^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}') \\ R_m &\geq R_m^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}) + b = R_m^G(B_{1\dots m}^*, \tilde{N}_{1\dots m}') + b \end{aligned} \tag{3-11}$$

Let W_i , $i = 1, \dots, m$ denote the index of the message sent to user i and let $\bar{\bar{Y}}_i$ be a matrix of size $t \times n$ denoting the signal received by user i (in n time samples). By Fano's inequality and the fact that the W_i's are independent, we know that there is a sufficiently large n such that we can find a code-book of length-n codewords and for which

$$\begin{aligned} R_i &\leq \frac{1}{n} I(W_i; \bar{\bar{Y}}_i) + \frac{1}{2 \cdot m} b \leq \frac{1}{n} I(W_i; \bar{\bar{Y}}_i | W_{i+1}, \dots, W_m) + \frac{1}{2 \cdot m} b\ , \ \ i = 1, \dots, m-1 \\ R_m &\leq \frac{1}{n} I(W_m; \bar{\bar{Y}}_m) + \frac{1}{2 \cdot m} b \end{aligned} \tag{3-12}$$

Let $\bar{\bar{Y}}_i'$, $i = 1, \ldots m$ denote the enhanced channel outputs of each of the receiving users. $\bar{\bar{Y}}_i'$ is a matrix of size $t \times n$. By the information processing theorem we can rewrite (3-12) as follows:

$$\begin{aligned} R_i &\leq \frac{1}{n} I(W_i; \bar{\bar{Y}}_i' | W_{i+1}, \ldots, W_m) + \frac{1}{2 \cdot m} b\,, \ \ i = 1, \ldots, m-1 \\ R_m &\leq \frac{1}{n} I(W_m; \bar{\bar{Y}}_m') + \frac{1}{2 \cdot m} b \end{aligned} \tag{3-13}$$

Thus, in (3-11) and (3-13), we have shifted the problem from the original channel to the enhanced channel. However, as the proportionality property holds for the enhanced channel, we can use Bergmans' approach to prove a contradiction.

By (3-13) and (3-11) we can write

$$\begin{aligned} \frac{1}{n} I(W_1; \bar{\bar{Y}}_1' | W_2, \ldots, W_m) + \frac{1}{2 \cdot m} b &= \frac{1}{n} H(\bar{\bar{Y}}_1' | W_2, \ldots, W_m) - \frac{1}{n} H(\bar{\bar{Y}}_1' | W_1, \ldots, W_m) + \frac{1}{2 \cdot m} b \\ &= \frac{1}{n} H(\bar{\bar{Y}}_1' | W_2, \ldots, W_m) - \frac{1}{2} \log \left(\left| 2\pi e \tilde{N}_1' \right| \right) + \frac{1}{2 \cdot m} b \\ &\geq R_1^G (B^*_{1\ldots m-1}, S, \tilde{N}'_{1\ldots m}) \\ &= \frac{1}{2} \log \left(\left| 2\pi e \left(B_1^* + \tilde{N}_1' \right) \right| \right) - \frac{1}{2} \log \left(\left| 2\pi e \tilde{N}_1' \right| \right) \end{aligned}$$

Thus,

$$\frac{1}{n} H(\bar{\bar{Y}}_1' | W_2, \ldots, W_m) \geq \frac{1}{2} \log \left(\left| 2\pi e \left(B_1^* + \tilde{N}_1' \right) \right| \right) - \frac{1}{2 \cdot m} b \tag{3-14}$$

We may write $\bar{\bar{Y}}_2' = \bar{\bar{Y}}_1' + \bar{\bar{Z}}_2'$ where $\bar{\bar{Z}}_2'$ is a random Gaussian matrix with independent columns and independent of both $\bar{\bar{Y}}_1'$ and the messages $(W_1, \ldots, W_m)$. Each column in $\bar{\bar{Z}}_2'$ has Normal distribution with zero mean and a covariance matrix, $\tilde{N}_2'$. Next, we use the Entropy Power Inequality to lower bound the entropy of the sum of two independent vectors with a function of the entropies of each of the vectors. Thus, using a slightly modified version of the EPI (Corollary I.1 in Appendix I), modified to allow conditioning on the message indexes, we may write (relying on (3-14)):

$$
\begin{aligned}
\frac{1}{t\cdot n}H(\bar{\bar{Y}}_2'|W_2,\dots,W_m) &= \frac{1}{t\cdot n}H(\bar{\bar{Y}}_1'+\bar{\bar{Z}}_2'|W_2,\dots,W_m)\\
&\geq \frac{1}{2}\log\left(e^{\frac{2}{t\cdot n}H(\bar{\bar{Y}}_1|W_2,\dots,W_m)}+e^{\frac{2}{t\cdot n}H(\bar{\bar{Z}}_2)}\right)\\
&\geq \frac{1}{2}\log\left(\left|2\pi e\left(B_1^*+\tilde{N}_1'\right)\right|^{\frac{1}{t}}e^{-\frac{2}{2\cdot m\cdot t}b}+|2\pi e\tilde{N}_2'|^{\frac{1}{t}}\right)\\
&\geq \frac{1}{2}\log\left(\left|2\pi e\left(B_1^*+\tilde{N}_1'\right)\right|^{\frac{1}{t}}+|2\pi e\tilde{N}_2'|^{\frac{1}{t}}\right)-\frac{1}{2\cdot m\cdot t}b
\end{aligned}
\tag{3-15}
$$

But by the proportionality property in Theorem 3.2, we know that $\tilde{N}_2'$ is proportional to $(B_1^*+\tilde{N}_1')$ and therefore,

$$
\left|2\pi e\left(B_1^*+\tilde{N}_1'\right)\right|^{\frac{1}{t}}+|2\pi e\tilde{N}_2'|^{\frac{1}{t}}=\left|2\pi e\left(B_1^*+\tilde{N}_1'+\tilde{N}_2'\right)\right|^{\frac{1}{t}}
$$

Thus, we may write,

$$
\frac{1}{n}H(\bar{\bar{Y}}_2'|W_2,\dots,W_m)\geq\frac{1}{2}\log\left(\left|2\pi e\left(B_1^*+\tilde{N}_1'+\tilde{N}_2'\right)\right|\right)-\frac{1}{2\cdot m}b \tag{3-16}
$$

Again, we can use (3-13) and (3-11) and write:

$$
\begin{aligned}
\frac{1}{n}I(W_2;\bar{\bar{Y}}_2'|W_3,\dots,W_m)+\frac{1}{2\cdot m}b &= \frac{1}{n}H(\bar{\bar{Y}}_2'|W_3,\dots,W_m)-\frac{1}{n}H(\bar{\bar{Y}}_2'|W_2,\dots,W_m)+\frac{1}{2\cdot m}b\\
&\geq R_2^G(B_{1\dots m-1}^*,S,\tilde{N}_{1\dots m}')\\
&=\frac{1}{2}\log\left(\left|2\pi e\left(B_1^*+B_2^*+\tilde{N}_1'+\tilde{N}_2'\right)\right|\right)-\frac{1}{2}\log\left(\left|2\pi e\left(B_1^*+\tilde{N}_1'+\tilde{N}_2'\right)\right|\right)
\end{aligned}
$$

Combining the expression above and (3-16) we get:

$$
\frac{1}{n}H(\bar{\bar{Y}}_2'|W_3,\dots,W_m)\geq\frac{1}{2}\log\left(\left|2\pi e\left(B_1^*+B_2^*+\tilde{N}_1'+\tilde{N}_2'\right)\right|\right)-\frac{2}{2\cdot m}b
$$

We can continue to calculate $\frac{1}{n}H(\bar{\bar{Y}}_i'|W_{i+1},\dots,W_m)$ for $i>2$ by using the above argumentation and by alternating between the EPI and Fano's inequality we will get in the m'th iteration:

$$
\begin{aligned}
\frac{1}{n}H(\bar{\bar{Y}}_m') &\geq \frac{1}{2}\log\left(\left|2\pi e\left(\sum_{i=1}^m B_i^*+\sum_{i=1}^m\tilde{N}_i'\right)\right|\right)-\frac{m}{2\cdot m}b+b\\
&=\frac{1}{2}\log\left(\left|2\pi e\left(S+\sum_{i=1}^m\tilde{N}_i'\right)\right|\right)+\frac{1}{2}b
\end{aligned}
$$

where the equality follows from the fact that $\sum_{i=1}^{m} B_i^* = S$ (Proposition 3.2). However, the above expression can not be true, because:

$$\frac{1}{n}H(\bar{\bar{Y}}'_m) \leq \frac{1}{n}\sum_{i=1}^{n} H(\bar{\bar{Y}}'_{m,i}) \leq \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(\left|2\pi e \cdot E[\bar{\bar{Y}}'_{m,i}\bar{\bar{Y}}'_{m,i}{}^T]\right|\right)$$
$$\leq \frac{1}{2}\log\left(\left|2\pi e \cdot \frac{1}{n}E[\bar{\bar{Y}}'_m\bar{\bar{Y}}'_m{}^T]\right|\right) \leq \frac{1}{2}\log\left(\left|2\pi e \cdot \left(S + \sum_{i=1}^{m}\tilde{N}'_i\right)\right|\right)$$

where $\bar{\bar{Y}}'_{m,i}$ is the i'th column of the random matrix $\bar{\bar{Y}}'_m$. The second inequality is due to the optimality of the Entropy of the Gaussian distribution. The third inequality is due to the concavity ($\cap$) of the $\log\det$ function and the last inequality is due to the fact that $\frac{1}{n}E[\bar{\bar{Y}}'_m\bar{\bar{Y}}'^T_m] \preceq S + \sum_{i=1}^{m}\tilde{N}'_i$ and the fact that $|B| \geq |A|$ if $B \succeq A \succ 0$ (Proposition I.2 in Appendix I).

Thus, we have contradicted our initial assumption and proved that all rate vectors $\overline{R} = (R_1, \ldots, R_m) \notin \mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ such that $R_i > 0$, $\forall i = 1, \ldots, m$, are not achievable. To complete the proof for $S \succ 0$, we now treat the case where the requirements on the rates are $R_i \geq 0$, $i = 1, \ldots, m$ in place of strict inequalities.

<u>$S \succ 0, \;\; R_i \geq 0 \, \forall i$:</u>

Assume that $\overline{R} \notin \mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ such that $R_i \geq 0$, $\forall i = 1, \ldots, m$ is an achievable rate. Let $n < m$ denote the number of strictly positive elements in $\overline{R}$ and let $k(i)$, $i = 1, \ldots, n$ be the index function of those elements such that $k(i+1) > k(i)$, $i = 1 \ldots n-1$. We define the compact rate vector $\overline{R}^C = (R_{k(1)}, \ldots, R_{k(n)})$. Similarly, we define a compact, n-user ADBC, with noise increment covariance matrices $\hat{\tilde{N}}_i = \sum_{j=k(i-1)+1}^{k(i)} \tilde{N}_j$, $\forall i = 1, \ldots, n$ (where we assign $k(0) = 0$) and the same covariance matrix constraint. Clearly, as $\overline{R}$ was achievable in the original ADBC, so is the compact rate vector, $\overline{R}^C$, achievable in the compact ADBC. Furthermore, since $\overline{R} \notin \mathcal{R}^G(S, \tilde{N}_{1\ldots m})$, then also $\overline{R}^C \notin \mathcal{R}^G(S, \hat{\tilde{N}}_{1\ldots n})$. Therefore, in the compact channel we have an "alleged" achievable rate vector which lies outside the Gaussian rate region. However, in the compact channel, this vector is element-wise strictly positive and we can apply the above proof to contradict our initial argument.

To complete the proof, we proceed and consider the case where the covariance matrix constraint, S, is such that $S \succeq 0$ and $|S| = 0$.

$S \succeq 0$, $|S| = 0$:

If S is not (strictly) positive definite, by Proposition 2.2 we know that there exists an equivalent ADBC with less transmit antennas, noise increment covariance matrices $\hat{\tilde{N}}_1, \dots, \hat{\tilde{N}}_m$, and an input covariance matrix constraint, $\hat{S} \succ 0$, with the exact same capacity region. Because $\hat{S}$ is strictly positive definite, the above proof could be applied to the equivalent channel to show that its capacity region coincides with its Gaussian rate region, i.e.

$$\mathcal{C}(S, \tilde{N}_{1\dots m}) = \mathcal{C}(\hat{S}, \hat{\tilde{N}}_{1\dots m}) = \mathcal{R}^G(\hat{S}, \hat{\tilde{N}}_{1\dots m}).$$

Moreover, it is possible to show that when S is not strictly positive definite,

$$\mathcal{R}^G(S, \tilde{N}_{1\dots m}) = \mathcal{R}^G(\hat{S}, \hat{\tilde{N}}_{1\dots m}).$$

Therefore, we have shown that the functional rate region, $\mathcal{R}^G(S, \tilde{N}_{1\dots m})$, coincides with the operational rate region, $\mathcal{C}(S, \tilde{N}_{1\dots m})$ for all $S \succeq 0$. ■

3.5 Proof of Theorem 3.2

In order to prove the theorem, we investigate the properties that the gradients of the Gaussian rate functions, R_i^G, must possess at an optimal Gaussian solution point. Therefore, before proceeding to prove Theorem 3.2, we present a generalized optimization problem and a proposition which states KKT-like necessary conditions on the gradients of the problem's functions. The optimization problem that follows, is slightly more general than is needed for the proof of Theorem 3.2. This will allow us to use the same results in the proof of the capacity region of the AMBC, as well.

In the following, we reuse the letters $B_1, \dots, B_{m-1}$ and assume that these are symmetric matrices of size $t \times t$. We define the following optimization problem:

$$\begin{array}{lll} \text{maximize} & f_0(B_1, \dots, B_{m-1}) & \\ B_1, \dots, B_{m-1} & & \\ \text{such that} & f_i(B_1, \dots, B_{m-1}) \geq 0, & i = 1, \dots, k \\ & B_j \succeq 0, & j = 1, \dots, m-1 \\ & \sum_{l=1}^{m-1} B_l \preceq S & \end{array} \tag{3-17}$$

where $f_i(B_1, \dots, B_{m-1})$, $i = 1, \dots, k$ are differentiable real functions over the region

$$\mathcal{D} = \left\{ (B_1, \dots, B_{m-1}) \big| \; B_1 \succeq 0, \dots, B_{m-1} \succeq 0, \; \sum_{j=1}^{m-1} B_j \preceq S \right\} \tag{3-18}$$

Before stating the KKT-like conditions, we give some notations and definitions. We denote the partial gradient of a function $f_i(B_1, \dots, B_{m-1})$ with respect to a matrix B_j by $\nabla_{B_j} f_i$. The elements of this gradient are the partial derivatives of the function f_i with respect to each of the elements in B_j and are organized in a matrix such that $(\nabla_{B_j} f_i)_{m,n} = \frac{\partial f_i}{\partial (B_j)_{m,n}}$. The gradient of the function f_i is simply the matrix created by the concatenation of the partial gradients such that $\nabla f_i = \left[\nabla_{B_1} f_i, \dots, \nabla_{B_{m-1}} f_i \right]$. A direction is given by a set of $m-1$ matrices of size $t \times t$, $(L_1, \dots, L_{m-1})$. The directional derivative is defined by:

$$\dot{f}_i(B_{1\dots m-1}, L_{1\dots m-1}) = \lim_{\alpha \downarrow 0} \frac{f_i(B_1 + \alpha L_1, \dots, B_{m-1} + \alpha L_{m-1}) - f_i(B_1, \dots, B_{m-1})}{\alpha}$$

As the functions, f_i, are differentiable, the directional derivative is equal to $\mathrm{tr}\{\nabla f_i \cdot (L_1, \dots, L_{m-1})^T\}$.

Definition 3.4 (Strict Direction) Let r_{B_i} denote the number of null eigenvalues in B_i (i.e. $r_{B_i} = t - \mathrm{rank}(B_i)$) and let V_{B_i}, $i = 1, \dots, m-1$ be matrices of size $t \times r_{B_i}$ which columns are orthogonal unit-length eigenvectors of B_i that correspond to null eigenvalues. Similarly, V_{B_m} and r_{B_m} are defined for the matrix, $B_m = S - \sum_{i=1}^{m-1} B_i$. We say that $(L_1, \dots, L_{m-1})$ is a *strict direction* at point $B_1, \dots, B_{m-1}$ if the following conditions hold:

1. $V_{B_i}^T L_i V_{B_i} \succ 0$, $\forall i \in \{1, \dots, m-1\}$ for which $r_{B_i} > 0$.

2. If $r_{B_m} > 0$, then $V_{B_m}^T \left(\sum_{j=1}^{m-1} L_j\right) V_{B_m} \prec 0$.
3. $\mathrm{tr}\{\nabla f_i(B_1, \ldots, B_m) \cdot (L_1, \ldots, L_{m-1})^T\} > 0, \quad \forall i = 1, \ldots, k$.

The intuition behind the definition of a *strict direction*, and the mathematical reason for defining it, will become evident in the proof of the next proposition.

Proposition 3.3: Let $B_1^*, \ldots, B_{m-1}^*$ be an optimal solution of the optimization problem given by (3-17). If the functions f_i, $i = 0, \ldots, k$ are differentiable, and there exists a *strict direction* at $B_1^*, \ldots, B_{m-1}^*$, the following KKT-like necessary conditions must hold:

$$0_{t\times t} = \nabla_{B_j} f_0(B_1^*, \ldots, B_{m-1}^*) + \sum_{i=1}^{k} \left(\gamma_i \cdot \nabla_{B_j} f_i(B_1^*, \ldots, B_{m-1}^*)\right) + O_j - O_m, \qquad \forall j = 1 \ldots m-1 \tag{3-19}$$

for some $\gamma_i \geq 0$, $i = 1, \ldots, k$ and some O_i such that if $r_{B_i^*} > 0$, $O_i = V_{B_i} \Lambda_i V_{B_i}^T$ for some $\Lambda_i \succeq 0_{r_{B_i^*} \times r_{B_i^*}}$, and if $r_{B_i^*} = 0$, then $O_i = 0_{t\times t}$.

The proof is deferred to Appendix VI. Note that we did not assume that the functions, f_i, are convex. Therefore, standard results that appear in [4] and [2] were not used and a proof was needed.

We can now turn to prove Theorem 3.2.

The idea behind the proof is to utilize Equation (3-19) in Proposition 3.3 to investigate the properties of the optimal Gaussian solution point. Assume that $B_1^*, \ldots, B_m^*$ is an optimal Gaussian solution point under a strictly positive definite covariance matrix constraint, $S \succ 0$. If $B_1^*, \ldots, B_m^*$ are all strictly positive ($B_i^* \succ 0$), it is possible to show, using Equation (3-19), that the proportionality property already holds for our original channel and hence the proof of Theorem 3.2 is almost trivial for this case.

For the more general case, where $B_i^* \succeq 0$, we can create an enhanced channel by subtracting a matrix $\Delta_i \succeq 0$ from the noise increment covariance matrix of that user. However, Δ_i is chosen such $\Delta_i \cdot B_i = 0_{t\times t}$. This choice of Δ_i ensures that using the same power allocation matrices, B_i^*, the rate of user i remains the same in the enhanced channel as well, since the noise increment covariance was only modified in those directions where no information is being sent to that user. To make sure that the rates to the following users remain the same, we add Δ_i to the noise increment covariance of user $i+1$. This process is repeated recursively for the following users. Using Equation (3-19), it is possible to show that there exists a choice of matrices Δ_i such that the resultant channel is a valid ADBC and such that the proportionality property holds. In the following we present a rigorous proof of Theorem 3.2.

Proof of Theorem 3.2: Assume that $B_1^*,\ldots,B_m^*$ is an optimal Gaussian solution point under a strictly positive definite covariance matrix constraint, $S \succ 0$. We will divide our proof into two. Initially, we will assume that $B_i^* \neq 0$, $\forall i = 1,\ldots,m$ (note that the matrices B_i^* can still have null eigenvalues). In this case, we actually assume that there is a strictly positive rate to each of the users (this is a simple observation from the Gaussian rate functions, (3-2) and Proposition I.2). Later, we give a simple argument that will allow us to expand the proof to any possible optimal Gaussian solution point.

<u>$B_i^* \neq 0$, $\forall i = 1,\ldots,m$:</u>

We start with the assumption that $B_i^* \neq 0$, $\forall i = 1,\ldots,m$. By Proposition 3.2, we know that $B_m^* = S - \sum_{i=1}^{m-1} B_i^*$. Therefore, the Gaussian rates of all optimal Gaussian solution points may be written as a function of only $m-1$ variable matrices, $B_1,\ldots,B_{m-1}$, and the covariance matrix constraint, S. For this reason, we modify our notation of the rate functions R_i^G, and write $R_i^G(B_{1\ldots m-1}, S, \tilde{N}_{1\ldots m})$ instead of $R_i^G(B_{1\ldots m}, \tilde{N}_{1\ldots m})$. The rates are calculated using the same equations, (3-2), and assigning $B_m = S - \sum_{i=1}^{m-1} B_i$.

Given an optimal Gaussian rate vector $\overline{R}^* = (R_1^*,\ldots,R_m^*) = (R_1^G(B_{1\ldots m-1}^*, S, \tilde{N}_{1\ldots m}),\ldots,$ $R_m^G(B_{1\ldots m-1}^*, S, \tilde{N}_{1\ldots m}))$, we can easily see that the optimal Gaussian solution point optimizes the following problem:

$$\begin{array}{lll} \underset{B_1,\ldots,B_{m-1}}{\text{maximize}} & R_m^G(B_{1\ldots m-1},S,\tilde{N}_{1\ldots m}) & \\ \text{such that} & R_i^G(B_{1\ldots m-1},S,\tilde{N}_{1\ldots m}) - R_i^* \geq 0, & i=1,\ldots,m \\ & B_j \succeq 0, & j=1,\ldots,m-1 \\ & \sum_{l=1}^{m-1} B_l \preceq S & \end{array} \tag{3-20}$$

The above problem corresponds to the more general one presented in (3-17) with m functions altogether and with $R_i^G - R_i^*$, $i=1,\ldots,m$ replacing equations f_i, $i=1,\ldots,k$ and R_m^G replacing f_0. The functions, R_i^G, are differentiable w.r.t $B_i \succeq 0$, $\forall i$ and by the fact that $\nabla_B \log\det(B) = B^{-1}$ for all B such that $\det(B)>0$, it follows that the partial gradients are given by

$$\nabla_{B_j} R_l^G(B_{1\ldots m-1},S,\tilde{N}_{1\ldots m})$$
$$= \begin{cases} \frac{1}{2}\left(\sum_{i=1}^{l}(B_i+\tilde{N}_i)\right)^{-1} - \frac{1}{2}\left(\sum_{i=1}^{l-1}B_i + \sum_{i=1}^{l}\tilde{N}_i\right)^{-1} & 1 \leq j < l < m \\ \frac{1}{2}\left(\sum_{i=1}^{l}(B_i+\tilde{N}_i)\right)^{-1} & 1 \leq j = l < m \\ 0 & 1 \leq l < j < m, \\ -\frac{1}{2}\left(\sum_{i=1}^{l-1}B_i + \sum_{i=1}^{l}\tilde{N}_i\right)^{-1} & 1 \leq j < m,\ l=m \end{cases} \tag{3-21}$$

In Appendix VII we show that there is a strict direction at $B_1^*,\ldots,B_{m-1}^*$ for this optimization problem (Proposition VII.1) and thus, by Proposition 3.3, we can write the following $m-1$ equations on the gradients of the rates:

$$0_{t\times t} = \sum_{i=1}^{m}\left(\gamma_i \cdot \nabla_{B_k} R_i^G(B_{1\ldots m-1}^*,S,\tilde{N}_{1\ldots m})\right) + \frac{1}{2}O_k - \frac{1}{2}O_m, \quad \forall k=1\ldots m-1$$

where $\gamma_i \geq 0$, $i=1,\ldots,m-1$ and $\gamma_m = 1$ and where O_k are such that if $r_{B_k^*}>0$, $O_k = V_{B_k}\Lambda_k V_{B_k}^T$ for some $\Lambda_k \succeq 0_{r_{B_k^*}\times r_{B_k^*}}$, and if $r_{B_k^*}=0$, then $O_k = 0_{t\times t}$. Note that in the above expression we have written $\frac{1}{2}O_k$ instead of just O_k, as it appears in Proposition 3.3. Clearly, this doesn't affect the result of the proposition but allows us to cancel out the constant, $\frac{1}{2}$, which appears in the expression for the partial gradients (3-21). Furthermore, note that O_k and B_k^* are related by:

1. $\text{rank}(O_k) \leq t - \text{rank}(B_k^*)$, $\forall k = 1, \ldots, m$.
2. $B_k^* \cdot O_k = 0_{t \times t}$, $\forall k = 1, \ldots, m$

By plugging in the expression for the partial gradients (3-21) into the above $m-1$ equations and subtracting equation $k+1$ from the k'th equation (except for $k = m-1$ where the expression is taken as is), we obtain the following $m-1$ equations

$$\gamma_{k+1} \left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}_i \right)^{-1} + O_{k+1} = \gamma_k \left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i \right)^{-1} + O_k \;, \;\; k = 1, \ldots, m-1 \quad \text{(3-22)}$$

We now use the assumption that $B_i^* \neq 0$, $i = 1, \ldots, m$ to show that γ_i are strictly positive and that $0 < \frac{\gamma_k}{\gamma_{k+1}} \leq 1 \;, \;\; \forall k = 1, \ldots, m-1$. Since we defined $\gamma_m = 1 > 0$, we can see that for $k = m-1$, the left hand side of (3-22) is strictly positive definite. Furthermore, since we assume that $B_{m-1}^* \neq 0$, $\text{rank}(O_{m-1}) < t$ (see note above) and therefore, the matrix O_{m-1} must have a zero eigenvalue. In that case, γ_{m-1} must be strictly positive because, as the left hand side and the right hand side of (3-22) are equal and as the left hand side has no zero eigenvalues, we must have $\gamma_{m-1} > 0$ to compensate for the fact that O_{m-1} has a zero eigenvalue. We can repeat the above arguments for all $k = 1, \ldots, m-1$ in expression (3-22) and conclude that $\gamma_i > 0 \;, \;\; \forall i = 1, \ldots, m-1$. Next, let $\mathbf{z}$ be an eigenvector corresponding to a zero eigenvalue of O_{k+1} (which has some zero eigenvalues because we assume that $B_i^* \neq 0 \;, \;\; \forall i = 1, \ldots, m$). We have,

$$\begin{aligned} \mathbf{z}^T \left(\gamma_{k+1} \left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}_i \right)^{-1} + O_{k+1} \right) \mathbf{z} &= \gamma_{k+1} \mathbf{z}^T \left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}_i \right)^{-1} \mathbf{z} \\ &= \mathbf{z}^T \left(\gamma_k \left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i \right)^{-1} + O_k \right) \mathbf{z} \end{aligned} \quad \text{(3-23)}$$

But as $B^{-1} \preceq A^{-1}$ when $B \succeq A \succ 0$ (Proposition I.3), we know that

$$\left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}_i\right)^{-1} \preceq \left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i\right)^{-1}$$

which means that (3-23) can hold only if $0 < \frac{\gamma_k}{\gamma_{k+1}} \leq 1$, $\forall k = 1, \ldots, m-1$.

Taking into account the fact that γ_i in (3-22) are strictly positive such that $0 < \frac{\gamma_k}{\gamma_{k+1}} \leq 1$, $\forall k = 1, \ldots, m-1$, and the fact the O_i are defined such that $B_i^* \cdot O_i = 0$ (see note above regarding the relation between B_i^* and O_i), we can prove by induction (see Proposition VIII.1 in Appendix VIII), the existence of an enhanced channel with noise increment covariance matrices $\tilde{N}_1', \ldots, \tilde{N}_m'$, such that

$$\begin{aligned} \gamma_{k+1}\left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}_i\right)^{-1} + O_{k+1} &= \gamma_{k+1}\left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}_i'\right)^{-1} \\ &= \gamma_k\left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i'\right)^{-1} = \gamma_k\left(\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i\right)^{-1} + O_k \end{aligned} \tag{3-24}$$

and such that

$$\frac{|\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i|}{|\sum_{i=1}^{k-1} B_i^* + \sum_{i=1}^{k} \tilde{N}_i|} = \frac{|\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i'|}{|\sum_{i=1}^{k-1} B_i^* + \sum_{i=1}^{k} \tilde{N}_i'|}, \qquad \forall k = 1, \ldots, m \tag{3-25}$$

The proof for expressions (3-24) and (3-25) is found in Appendix VIII.

By equation (3-24) we can write:

$$\tilde{N}_{k+1}' = \frac{\gamma_{k+1} - \gamma_k}{\gamma_k}\left(\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} \tilde{N}_i'\right), \ k = 1, \ldots, m-1,$$

and therefore, the proportionality property holds for the enhanced channel. By equation (3-25), we may write,

$$\begin{aligned} R_i^G(B_{1\ldots m}^*, \tilde{N}_{1\ldots m}) = \frac{1}{2}\log\frac{|\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i|}{|\sum_{i=1}^{k-1} B_i^* + \sum_{i=1}^{k} \tilde{N}_i|} &= \frac{1}{2}\log\frac{|\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}_i'|}{|\sum_{i=1}^{k-1} B_i^* + \sum_{i=1}^{k} \tilde{N}_i'|} \\ &= R_i^G(B_{1\ldots m}^*, \tilde{N}_{1\ldots m}'), \ \forall i = 1, \ldots, m \end{aligned}$$

and therefore, the rate preservation property holds.

To complete the proof for the case where $B_i \succ 0$, $\forall i$, we still need to show that $B_1^*, \ldots, B_m^*$ is also an optimal Gaussian solution point for the enhanced channel. For that purpose, we observe that it is sufficient to show that

$$R_m^* = \sup\{R_m \in \mathbb{R} | (R_1^*, \ldots, R_{m-1}^*, R_m) \in \mathcal{R}^G(S, \tilde{N}'_{1\ldots m})\}. \tag{3-26}$$

To show that this is indeed a sufficient condition for optimality, we note that if there was another vector $(R_1, \ldots, R_{m-1}, R_m) \in \mathcal{R}^G(S, \tilde{N}'_{1\ldots m})$ such that $R_i \geq R_i^*\ \forall i = 1, \ldots, m$ where the inequality is strict for at least one of the elements, then by Proposition IV.1 we could find a rate vector $(R_1^*, \ldots, R_{m-1}^*, R_m^* + \epsilon) \in \mathcal{R}^G(S, \tilde{N}'_{1\ldots m})$ for some $\epsilon > 0$, contradicting (3-26).

We proceed to show that indeed (3-26) holds. As $\sum_{i=1}^m B_i^* = S$ and as

$$R_m^G(B_{1\ldots m}, \tilde{N}'_{1\ldots m}) = \frac{1}{2} \log \frac{\left|\sum_{i=1}^m B_i + \sum_{i=1}^m \tilde{N}_i'\right|}{\left|\sum_{i=1}^{m-1} B_i + \sum_{i=1}^m \tilde{N}_i'\right|} \leq \frac{1}{2} \log \frac{\left|S + \sum_{i=1}^m \tilde{N}_i'\right|}{\left|\sum_{i=1}^{m-1} B_i + \sum_{i=1}^m \tilde{N}_i'\right|},$$

it is enough to show that given $R_1^*, \ldots, R_{m-1}^*$, $B_1^*, \ldots, B_m^*$ minimizes $\left|\sum_{i=1}^{m-1} B_i + \sum_{i=1}^m \tilde{N}_i'\right|$ over all sequences of semi-definite matrices, $B_1, \ldots, B_m$ such that $\sum_{i=1}^m B_i \preceq S$. Using the

Minkowski inequality (Proposition I.5 in Appendix I), we may write:

$$
\begin{aligned}
\left|\sum_{i=1}^{m-1} B_i + \sum_{i=1}^{m} \tilde{N}_i'\right|^{\frac{1}{t}} &\geq \left|\sum_{i=1}^{m-1} B_i + \sum_{i=1}^{m-1} \tilde{N}_i'\right|^{\frac{1}{t}} + \left|\tilde{N}_m'\right|^{\frac{1}{t}} \\
&= e^{\frac{2}{t}R^*_{m-1}} \left|\sum_{i=1}^{m-2} B_i + \sum_{i=1}^{m-1} \tilde{N}_i'\right|^{\frac{1}{t}} + \left|\tilde{N}_m'\right|^{\frac{1}{t}} \\
&\geq e^{\frac{2}{t}R^*_{m-1}} \left(\left|\sum_{i=1}^{m-2} B_i + \sum_{i=1}^{m-2} \tilde{N}_i'\right|^{\frac{1}{t}} + \left|\tilde{N}_{m-1}'\right|^{\frac{1}{t}} \right) + \left|\tilde{N}_m'\right|^{\frac{1}{t}} \\
&= e^{\frac{2}{t}R^*_{m-1}} \left(e^{\frac{2}{t}R^*_{m-2}} \left|\sum_{i=1}^{m-3} B_i + \sum_{i=1}^{m-2} \tilde{N}_i'\right|^{\frac{1}{t}} + \left|\tilde{N}_{m-1}'\right|^{\frac{1}{t}} \right) + \left|\tilde{N}_m'\right|^{\frac{1}{t}} \\
&\geq e^{\frac{2}{t}R^*_{m-1}} \left(e^{\frac{2}{t}R^*_{m-2}} \left(\left|\sum_{i=1}^{m-3} B_i + \sum_{i=1}^{m-3} \tilde{N}_i'\right|^{\frac{1}{t}} + \left|\tilde{N}_{m-2}'\right|^{\frac{1}{t}} \right) + \left|\tilde{N}_{m-1}'\right|^{\frac{1}{t}} \right) + \left|\tilde{N}_m'\right|^{\frac{1}{t}} \\
&\vdots \\
&\geq \sum_{i=1}^{m} \exp\left\{ \frac{2}{t} \sum_{j=i}^{m-1} R^*_j \right\} \left|\tilde{N}_i'\right|^{\frac{1}{t}}
\end{aligned}
$$

where the inequalities hold with equality if and only if $\tilde{N}_{k+1}'$ is proportional to $\left(\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} \tilde{N}_i'\right)$ for all $k = 1, \ldots, m-1$. However, this is impled by the proportionality property, for the sequence $B_1^*, \ldots, B_m^*$, and therefore, we have shown that indeed $B_1^*, \ldots, B_m^*$ minimizes $\left|\sum_{i=1}^{m-1} B_i + \sum_{i=1}^{m} \tilde{N}_i{}'\right|$ and therefore, is an optimal Gaussian solution point of the enhanced channel.

$B_i^* \succeq 0,\ \forall i = 1, \ldots, m$:

Finally, we expand the proof to all possible choices of optimal Gaussian solution points, $B_1^*, \ldots, B_m^*$, some of which matrices may be zero. Let $n < m$ denote the number of non-zero matrices, B_i^*, $i = 1, \ldots, m$, and let $k(i)$, $i = 1, \ldots, n$ be the index function of those elements such that $k(i+1) > k(i)$, $i = 1 \ldots n-1$. We can define a compact channel which is also an ADBC with the same covariance matrix constraint, S, and with noise increment covariance matrices, $\hat{N}_1, \ldots, \hat{N}_n$ such that:

$$\hat{N}_i = \sum_{j=k(i-1)+1}^{k(i)} \tilde{N}_j \ , \quad i = 1, \ldots, n$$

where we define, $k(0) = 0$. Similarly, we define a compact solution point $\hat{B}_1, \ldots, \hat{B}_n$ such that

$$\hat{B}_i = B^*_{k(i)} \ , \quad i = 1, \ldots, n$$

Note that $\hat{B}_1, \ldots, \hat{B}_n$ must be an optimal Gaussian solution point of the compact channel and achieves the same rates (non-zero rates) as in the original channel. Since $\hat{B}_i \neq 0 \ , \ i = 1, \ldots, n$ and $S \succ 0$, we can use the above proof to show that Theorem 3.2 holds for the compact channel.

We can now define an enhanced channel for the original channel using the enhanced channel of the compact channel implied by Theorem 3.2. Because Theorem 3.2 holds for the compact channel, we know that there exists an enhanced channel $\hat{N}'_1, \ldots, \hat{N}'_n$ for which the results of the theorem hold. We now define an enhanced channel $\tilde{N}'_1, \ldots, \tilde{N}'_m$ for the original channel as follows:

$$\tilde{N}'_i = \begin{cases} \beta_i \hat{N}'_1 & i \leq k(1) \\ \hat{N}'_{k^{-1}(i)} & \forall i \in \{k(2), \ldots, k(n)\} \\ 0 & \text{Otherwise} \end{cases}$$

where $k(k^{-1}(i)) = i$ (if $i \in \{k(j), j = 1, \ldots, n\}$) and where $\beta_i \geq 0, \ i = 1 \ldots k(1)$ are chosen such that $\sum_{i=1}^{k(1)} \beta_i = 1$ and such that $0 \prec \left(\sum_{i=1}^{j} \beta_i\right) \cdot \hat{N}'_1 \preceq \sum_{i=1}^{j} \tilde{N}_i, \ \forall j = 1 \ldots k(1) - 1$. As $\sum_{i=1}^{j} \tilde{N}_i \succ 0, \ \forall j = 1, \ldots, m$, it is possible to find such β_i. One can verify that the results of Theorem 3.2 hold for the enhanced channel we just defined.

■

3.6 ADBC Example

The following example illustrates the result stated in Theorem 3.2. In this example, we consider a two user ADBC under a covariance matrix input constraint, where the transmitter and each of the receivers have two antennas such that:

$$\tilde{N}_1 = \begin{pmatrix} 0.5 & 0.18 \\ 0.18 & 0.7 \end{pmatrix}, \qquad \tilde{N}_2 = \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 10 \end{pmatrix}$$

$$\text{and} \quad S = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 2 \end{pmatrix}.$$

The boundary of the Gaussian region, $\mathcal{R}^G(S, \tilde{N}_{1,2})$ is plotted by a solid curve in Figure 1. Two additional curves ($\mathcal{R}^G(S, \tilde{N}'_{1,2})$ and $\mathcal{R}^G(S, \tilde{N}''_{1,2})$) were plotted, each corresponding to boundaries of Gaussian regions of different enhanced channels, which were obtained for two different points on the solid curve.

The first curve, the boundary of $\mathcal{R}^G(S, \tilde{N}'_{1,2})$ illustrated by the dashed curve, was calculated with respect to point $(R'_1, R'_2) \approx (0.08992,\ 0.51314)$ on the solid curve. The power allocation that achieves this point is given by,

$$B'_1 \approx \begin{pmatrix} 0.00004 & -0.00217 \\ -0.00217 & 0.12353 \end{pmatrix} \quad \text{and} \quad B'_2 = S - B'_1.$$

The dashed line corresponds to the boundary of a Gaussian region of an enhanced version of the original channel, $\mathcal{R}^G(S, \tilde{N}'_{1,2})$, which noise increment covariances are given by

$$\tilde{N}'_1 \approx \begin{pmatrix} 0.04896 & 0.00762 \\ 0.00762 & 0.63412 \end{pmatrix} \quad \text{and} \quad \tilde{N}'_2 = \tilde{N}_1 + \tilde{N}_2 - \tilde{N}'_1.$$

As predicted by Theorem 3.2, we can see that for the point $(R'_1, R'_2) \approx (0.08992,\ 0.51314)$ on the boundary of $\mathcal{R}^G(S, \tilde{N}_{1,2})$, there exists an enhanced version of the original channel which is also an ADBC and such that the boundary of its Gaussian region is tangential to that

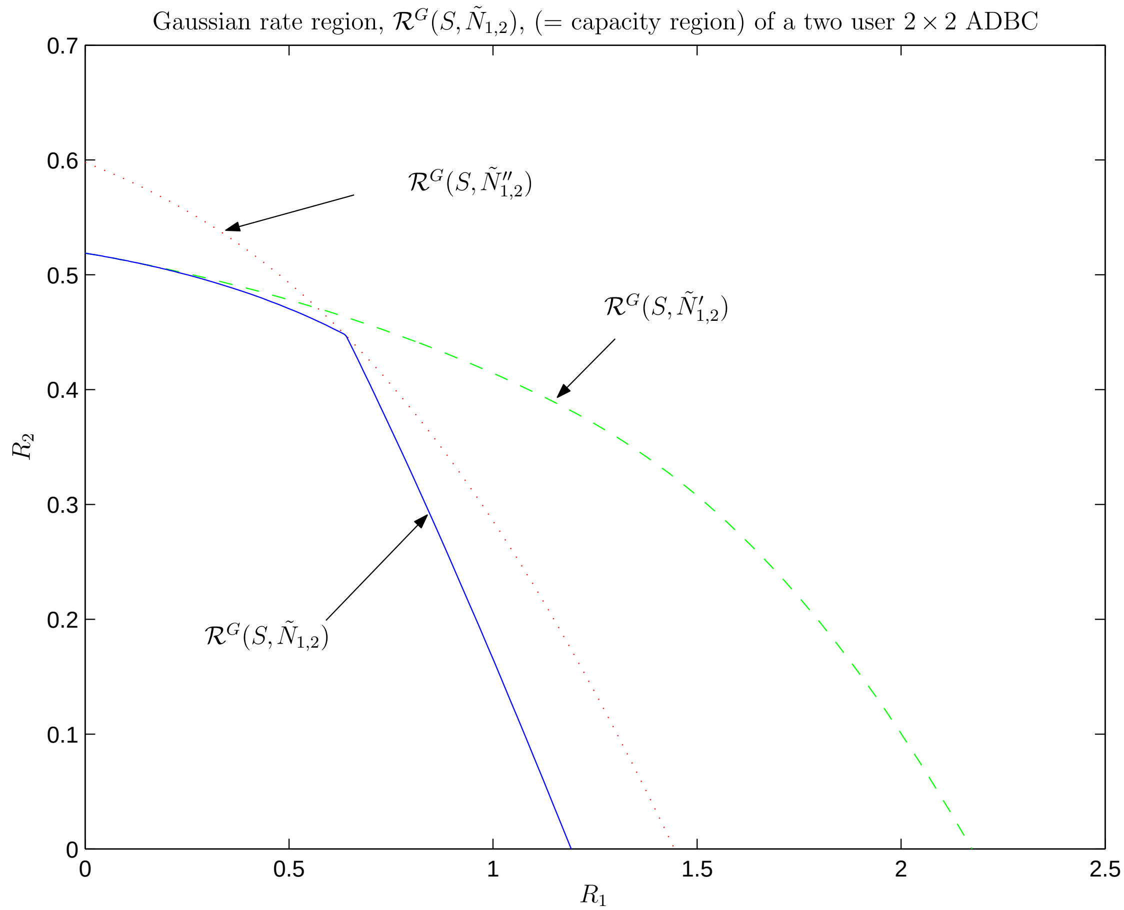

Fig. 1. Illustration of the results in Theorem 3.2

of the original channel at (R'_1, R'_2). In fact, in this case, the dashed line intersects with the solid line for all rates $R_1 \le R'_1$ and upper bounds the solid line for all $R_1 > R'_1$. Furthermore, one can easily check that the proportionality property holds, such that $(B'_1 + \tilde{N}'_1) \propto \tilde{N}'_2$.

The second curve, the boundary of $\mathcal{R}^G(S, \tilde{N}''_{1,2})$ given by the dotted line, was computed with respect to point $(R''_1, R''_2) \approx (0.63773, 0.44765)$ on the solid curve. The power allocation that achieves this point is given by,

$$B''_1 \approx \begin{pmatrix} 0.00000 & -0.00196 \\ -0.00196 & 1.63764 \end{pmatrix} \quad \text{and} \quad B''_2 = S - B''_1.$$

The dotted line corresponds to the boundary of a Gaussian region of an enhanced version of the original channel, $\mathcal{R}^G(S, \tilde{N}''_{1,2})$, which noise increment covariances are given by

$$\tilde{N}''_1 \approx \begin{pmatrix} 0.25033 & 0.08974 \\ 0.08974 & 0.66737 \end{pmatrix} \quad \text{and} \quad \tilde{N}''_2 \approx \begin{pmatrix} 0.44504 & 0.15605 \\ 0.15605 & 4.09780 \end{pmatrix}.$$

Again, we see that the prediction of Theorem 3.2 holds. The solid and the dotted curves are tangential at$(R''_1, R''_2) \approx (0.63773, 0.44765)$ and one can easily check that $(B''_1 + \tilde{N}''_1) \propto \tilde{N}''_2$.

4. The capacity region of the AMBC

In this section, we build on Theorem 3.1 in order to characterize the capacity region of the aligned (not necessarily degraded) MIMO BC. This result is particularly interesting in light of the fact that there is no single letter formula for the capacity region, as the AMBC is not necessarily degraded. In addition, a coding scheme consisting of a superposition of Gaussian codes along with successive decoding can not work when the channel is not degraded. Therefore, following the work of Caire and Shamai [5], we suggest an achievable rate region based on dirty paper coding. In [5], [23], [19], [17], it was shown that DPC achieves the sum capacity of the channel. In this section we show that DPC along with time sharing, covers the entire capacity region of the AMBC.

In [18], [16], it was shown that if Gaussian coding is optimal for the vector degraded BC, then the DPC rate region is also the capacity region of the GMBC. In the previous section, we have shown that Gaussian coding is optimal for the ADBC. This result can be generalized to the general vector degraded BC using a limit process on the noise variances of some of the receive antennas of some of the users, in a similar manner to what is done in the next section. In [21], we presented this approach to prove the converse of the GMBC. However, in this paper, we take a different approach which is a natural extension of Theorem 3.1, and which does not rely on the degraded same marginals bound presented in [18], [16].

In the following subsection, we characterize the capacity region of the AMBC and in later subsections we give a proof of the converse as well as some intermediate results.

4.1 AMBC - Main Result

The dirty paper encoder, performs successive precoding of the users information in a predetermined order, which allows a given set of rates. In order to define this order, we use π as a permutation function which permutes the set $\{1,\ldots,m\}$ such that $\pi(i)\in\{1,\ldots,m\}$, $\forall i=1,\ldots,m$ and $\pi(i)\neq\pi(j)$, $\forall i\neq j$.

Given an average transmit covariance matrix limitation, $E[\mathbf{x}\mathbf{x}^T]\preceq S$, and a set of positive semi-definite matrices, $B_1,B_2,\ldots,B_m$ ($B_k\succeq 0\ \forall k\in 1\ldots m$), such that $\sum_{i=1}^m B_i\preceq S$, and a permutation function π, the following rates are achievable in the AMBC using a DPC scheme [5], [23], [19], [17]:

$$R_k\leq R^{DPC}_{\pi^{-1}(k)}(\pi,B_{1\ldots m},N_{1\ldots m})\ ,\qquad \forall k\in 1,\ldots,m$$

where

$$R^{DPC}_l(\pi,B_{1\ldots m},N_{1\ldots m})=\frac{1}{2}\log\frac{\left|\left(\left(\sum_{i=1}^{l}B_{\pi(i)}\right)+N_{\pi(l)}\right)\right|}{\left|\left(\left(\sum_{i=1}^{l-1}B_{\pi(i)}\right)+N_{\pi(l)}\right)\right|}\quad l=1,2,\ldots,m \tag{4-1}$$

and where π^{-1} is the inverse permutation such that $\pi^{-1}(\pi(i))=i$. We can now define the DPC achievable rate region of an AMBC :

Definition 4.1 (DPC rate-region of an AMBC) Let S be a positive semi-definite matrix, then, we define the DPC rate region, $\mathcal{R}^{DPC}(S,N_{1\ldots m})$, of the AMBC with a covariance matrix constraint, S, as the following convex closure:

$$\mathcal{R}^{DPC}(S,N_{1\ldots m})=\mathbf{conv}\left\{\bigcup_{\pi\in\Pi}\mathcal{R}^{DPC}(\pi,S,N_{1\ldots m})\right\} \tag{4-2}$$

where Π is the set of all possible permutations of the set $\{1,\ldots,m\}$, **conv** is the convex closure operator and

$$\mathcal{R}^{DPC}(\pi, S, N_{1\ldots m}) = \left\{ \begin{pmatrix} R_1^{DPC}(\pi, B_{1\ldots m}, N_{1\ldots m}), \ldots, \\ R_m^{DPC}(\pi, B_{1\ldots m}, N_{1\ldots m}) \end{pmatrix} \Bigg| \quad \text{s.t.} \quad S - \sum_{i=1}^{m} B_i \succeq 0, \; B_i \succeq 0 \;\; \forall i = 1, \ldots, m \right\} \tag{4-3}$$

Note that for an ADBC, $\mathcal{R}^{DPC}(\pi_I, S, N_{1\ldots m}) = \mathcal{R}^{G}(S, \tilde{N}_{1\ldots m})$, where π_I is the identity permutation such that $\pi_I(i) = i$ and where $\tilde{N}_i = N_i - N_{i-1}, i = 1, \ldots, m$ (and where $N_0 = 0$).

Theorem 4.1: Let $\mathcal{C}(S, N_{1\ldots m})$ denote the capacity region of the AMBC under a covariance matrix constraint, $S \succeq 0$, then $\mathcal{C}(S, N_{1\ldots m}) = \mathcal{R}^{DPC}(S, N_{1\ldots m})$.

As $\mathcal{R}^{DPC}(S, N_{1\ldots m})$ is contiguous w.r.t S (see Proposition III.2 in Appendix III), the following corollary follows immediately by Proposition 2.1.

Corollary 4.1: Let $\mathcal{C}(P, N_{1\ldots m})$ denote the capacity region of the AMBC under a total power constraint, $P \geq 0$, then

$$\mathcal{C}(P, N_{1\ldots m}) = \bigcup_{\substack{S \succeq 0 \\ \text{s.t.} \\ \text{tr}\{S\} \leq P}} \mathcal{R}^{DPC}(S, N_{1\ldots m})$$

4.2 Proof of Theorem 4.1

To prove Theorem 4.1, we will show that for every rate vector, $\overline{R}^a$, which lies outside $\mathcal{R}^{DPC}(S, N_{1\ldots m})$, we can find an enhanced ADBC, whose capacity region doesn't contain $\overline{R}^a$. However, the capacity region of the enhanced channel outer bounds that of the original channel, and therefore, $\overline{R}^a$, can not be an achievable rate vector. In the following, we give some important intermediate results before proving the theorem.

Given a sequence $\gamma = (\gamma_1, \ldots \gamma_m)$ and a scalar b, we say that $\{\overline{R} = (R_1, \ldots, R_m) \mid \sum_{i=1}^{m} \gamma_i R_i = b\}$ is a supporting hyperplane of a closed and bounded set $\mathcal{X} \subset \mathbb{R}^m$, if $\sum_{i=1}^{m} \gamma_i R_i \leq b$, $\forall (R_1, \ldots, R_m) \in \mathcal{X}$ where for at least one rate vector, $(R_1, \ldots, R_m) \in \mathcal{X}$, we have $\sum_{i=1}^{m} \gamma_i R_i =$

b. Note that as $\mathcal{X}$ is closed and bounded, $\max_{\overline{R}\in\mathcal{X}} \sum_{i=1}^{m} \gamma_i R_i$ exists for all vectors $\gamma = (\gamma_1, \ldots, \gamma_m)$. Hence, we can find a supporting hyperplane for the set $\mathcal{X}$ for all vectors γ.

Proposition 4.1: Let $\overline{R}^a = (R_1^a, \ldots, R_m^a)$ be a rate vector which lies outside the DPC rate region, $\mathcal{R}^{DPC}(S, N_{1\ldots m})$. Since $R_i^a \geq 0$, $\forall i = 1, \ldots, m$, then, there exists a constant $b \geq 0$ and a vector $\gamma = (\gamma_1, \ldots, \gamma_m)$ such that $\gamma_i \geq 0$, $\forall i = 1, \ldots, m$ and where not all γ_i are zero, such that the hyperplane $\{(R_1, \ldots, R_m) \mid \sum_{i=1}^{m} \gamma_i R_i = b\}$ is a supporting and separating hyperplane for which

$$\sum_{i=1}^{m} \gamma_i R_i \leq b, \qquad \forall (R_1, \ldots, R_m) \in \mathcal{R}^{DPC}(S, N_{1\ldots m}) \tag{4-4}$$

and for which

$$\sum_{i=1}^{m} \gamma_i R_i^a > b \tag{4-5}$$

where (4-4) holds with equality for at least one point in $\mathcal{R}^{DPC}(S, N_{1\ldots m})$.

Proof: As $\mathcal{R}^{DPC}(S, N_{1\ldots m})$ is a closed and convex set and $(R_1^a, \ldots, R_m^a)$ is a point which lies outside that set, we can use the separating hyperplane theorem (see Section 2, Chapter 1 in [4]) to show that there exists a supporting hyperplane which strictly separates $\mathcal{R}^{DPC}(S, N_{1\ldots m})$ and $(R_1^a, \ldots, R_m^a)$. In other words, there is a vector $\gamma = (\gamma_1, \ldots, \gamma_m)^T$, and a constant b such that equations (4-4) and (4-5) hold. We need to show that we can find a vector γ with non-negative elements and a non-negative scalar, b.

Assume, in contrast, that $\gamma_k < 0$ for some $k \in \{1, \ldots, m\}$ and that $\overline{R}^* = (R_1^*, \ldots, R_m^*)$ maximizes $\sum_{i=1}^{m} \gamma_i R_i$ over all $\overline{R} = (R_1, \ldots, R_m) \in \mathcal{R}^{DPC}(S, N_{1\ldots m})$. We note that for all vectors $\overline{R} \in \mathcal{R}^{DPC}(S, N_{1\ldots m})$, also $(R_1, \ldots, R_{k-1}, 0, R_{k+1}, \ldots, R_m) \in \mathcal{R}^{DPC}(S, N_{1\ldots m})$ (see Corollary V.1). Therefore, as $\gamma_k < 0$, the vector which optimizes $\sum_{i=1}^{m} \gamma_i R_i$ must be such that $R_k = 0$. Thus, we conclude that $R_k^* = 0$. Consequently, choosing $\gamma_k = 0$, will also lead to a supporting and separating hyperplane since it can only increase $\sum_{i=1}^{m} \gamma_i R_i^a$ (as R_i^a are all non-negative), and does not affect the optimization of $\sum_{i=1}^{m} \gamma_i R_i$ over all $(R_1, \ldots, R_m) \in \mathcal{R}^{DPC}(S,\ N_{1\ldots m})$, as otherwise, by Corollary V.1 we know that $\overline{R}^*$ wouldn't have been optimal for the original choice of γ_k. Moreover, we know that at least one of the elements of γ is strictly positive because of the strict separation. Finally, as we have shown that we can find a supporting and separating hyperplane with a vector γ that contains only non-negative elements, and as $\mathcal{R}^{DPC}(S, N_{1\dots m})$ contains only non-negative vectors, $b = \max_{\overline{R}\in\mathcal{R}^{DPC}} \sum_{i=1}^m \gamma_i R_i \geq 0$. ■

The following proposition brings to bare a relation between the ideas of a supporting hyperplane and the enhanced channel. This proposition is a natural extension of Theorem 3.2 to the AMBC.

Proposition 4.2: Consider an AMBC with noise covariance matrices, $(N_1, \dots, N_m)$, and an average transmit covariance matrix constraint, $S \succ 0$. Define π_I to be the identity permutation, $\pi_I(i) = i$, $i = 1 \dots m$. If $\{(R_1, \dots, R_m) \mid \sum_{i=1}^m \gamma_i R_i = b\}$ is a supporting hyperplane of the rate region, $\mathcal{R}^{DPC}(\pi_I, S, N_{1\dots m})$, such that $0 \leq \gamma_1 \leq \dots \leq \gamma_m$, $\gamma_m > 0$ and $b \geq 0$, then, there exists an enhanced ADBC with noise increment covariances, $(\tilde{N}'_1, \dots, \tilde{N}'_m)$, such that the following properties hold:

1. *Enhanced channel:*
$$0 \prec \sum_{i=1}^k \tilde{N}'_i \preceq N_k, \qquad \forall k = 1, \dots, m$$
2. *Supporting hyperplane preservation:*
$\{(R_1, \dots, R_m) \mid \sum_{i=1}^m \gamma_i R_i = b\}$ is also a supporting hyperplane of the rate region $\mathcal{R}^G(S, \tilde{N}'_{1\dots m})$.

The proof is deferred to Subsection 4.3. We can now turn to prove Theorem 4.1.

Proof of Theorem 4.1: Just as in the proof of Theorem 3.1, we treat separately the cases $S \succ 0$ and $S \succeq 0$, $|S| = 0$. We first treat the case where $S \succ 0$ and then broaden the scope of the proof to all $S \succeq 0$.

<u>$S \succ 0$:</u>

Let $(R_1^a, \dots, R_m^a)$ be a rate vector with non-negative elements which lies outside the rate region $\mathcal{R}^{DPC}(S, N_{1\dots m})$. By Proposition 4.1, we know that there is a supporting and separating hyperplane $\{(R_1, \dots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b\}$ where γ_i, $i = 1, \dots, m$ are non-negative and at least one of the elements is positive.

Let π_γ be a permutation on the set $\{1, \dots, m\}$ that orders the elements of γ such that $\gamma_{\pi_\gamma(i+1)} \geq \gamma_{\pi_\gamma(i)}$ $\forall i = 1, \dots, m-1$. We observe that as $\mathcal{R}^{DPC}(\pi_\gamma, S, N_{1\dots m}) \subseteq \mathcal{R}^{DPC}(S, N_{1\dots m})$ and as $\{(R_1, \dots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b\}$ is a supporting hyperplane of $\mathcal{R}^{DPC}(S, N_{1\dots m})$, we can write

$$b' = \max_{\substack{(R_1,\dots,R_m)\in \\ \mathcal{R}^{DPC}(\pi_\gamma,S,N_{1\dots m})}} \sum_{i=1}^m \gamma_i R_i \leq \max_{\substack{(R_1,\dots,R_m)\in \\ \mathcal{R}^{DPC}(S,N_{1\dots m})}} \sum_{i=1}^m \gamma_i R_i = b$$

Furthermore, as $\{(R_1, \dots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b\}$ is also a separating hyperplane, we can also write

$$\sum_{i=1}^m \gamma_i R_i^a > b \geq b'$$

and therefore, $\{(R_1, \dots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b'\}$ is a supporting and separating hyperplane for the rate region $\mathcal{R}^{DPC}(\pi_\gamma, S, N_{1\dots m})$.

Note that in general, π_γ, may be any one of the possible permutations. Therefore, to prove the last statement, we exploited the fact that $\mathcal{R}^{DPC}(S, N_{1\dots m})$ is the convex hull of the union over all DPC rate regions $\mathcal{R}^{DPC}(\pi, S, N_{1\dots m})$, where the union is taken over all possible DPC pre-coding orders.

For brevity, we will assume in the following that $\gamma_1 \leq \dots \leq \gamma_m$ or alternatively, that $\pi_\gamma = \pi_I$. If that is not the case, we can always reorder the users such that this relation will hold. From the above, we know that $\{(R_1, \dots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b'\}$ is a supporting and separating hyperplane of $\mathcal{R}^{DPC}(\pi_\gamma = \pi_I, S, N_{1\dots m})$. By Proposition 4.2 we know that there exists an enhanced ADBC whose Gaussian rate region, $\mathcal{R}^G(S, \tilde{N}'_{1\dots m})$, lies under the supporting hyperplane and hence, $(R_1^a, \ldots, R_m^a) \notin \mathcal{R}^G(S, \tilde{N}'_{1\ldots m})$. However, by Theorem 3.1 we know that the Gaussian rate region of the enhanced ADBC is also the capacity region. Therefore, we conclude that $(R_1^a, \ldots, R_m^a)$ must lie outside the capacity region of the enhanced ADBC.

To complete the proof for the case $S \succ 0$, we recollect that the capacity region of the enhanced ADBC contains that of the original channel and therefore, $(R_1^a, \ldots, R_m^a)$, must lie outside the capacity region of the original AMBC. As this is true for all rate vectors which lie outside $\mathcal{R}^{DPC}(S, N_{1\ldots m})$, we conclude that $\mathcal{C}(S, N_{1\ldots m}) \subseteq \mathcal{R}^{DPC}(S, N_{1\ldots m})$. However, $\mathcal{R}^{DPC}(S, N_{1\ldots m})$, is a set of achievable rates and therefore, $\mathcal{C}(S, N_{1\ldots m}) \supseteq \mathcal{R}^{DPC}(S, N_{1\ldots m})$.

$S \succeq 0$:

Following the same ideas that appeared in the proof of Theorem 3.1, we broaden the scope of our proof to the case of any $S \succeq 0$. If S is not (strictly) positive definite, by Proposition 2.2 we know that there exists an equivalent AMBC with less transmit antennas, noise covariance matrices $\hat{N}_1, \ldots, \hat{N}_m$, and an input covariance matrix constraint, $\hat{S} \succ 0$, with the exact same capacity region. Because $\hat{S}$ is strictly positive definite, the above proof (for the $S \succ 0$ case) could be applied to the equivalent channel to show that its capacity region is equal to its DPC rate region, $\mathcal{R}^{DPC}(\hat{S}, \hat{N}_{1\ldots m})$. Moreover, it is possible to show that when S is not strictly positive definite, $\mathcal{R}^{DPC}(S, N_{1\ldots m}) = \mathcal{R}^{DPC}(\hat{S}, \hat{N}_{1\ldots m})$. Hence, $\mathcal{R}^{DPC}(S, N_{1\ldots m})$ coincides the the capacity region of the AMBC for all $S \succeq 0$. ■

4.3 Proof of Proposition 4.2

Before turning to prove Proposition 4.2, we first prove an auxiliary proposition.

Proposition 4.3: Let $\gamma = (\gamma_1, \ldots, \gamma_m)$ be a vector of non-negative scalars and let n, $1 \leq n < m$ be the number of strictly positive elements in the vector. Furthermore, let $k(i)$, $i =$

$1,\ldots,n$ denote the index function of those elements such that $k(i)$, $i=1,\ldots,n$ points to strictly positive entries of the vector γ and such that $k(i+1)>k(i)\ \forall i=1,\ldots,n-1$. Consider an AMBC with m users and noise covariance matrices $N_1,\ldots,N_m$ and define a compact AMBC with n users and noise covariance matrices given by $\hat{N}_i=N_{k(i)}$, $i=1,\ldots,n$. Then,

$$\max_{B_{1\ldots m}\in\mathcal{D}}\sum_{i=1}^{m}\gamma_i R_i^{DPC}(\pi_I,B_{1\ldots m},N_{1\ldots m})=\max_{\hat{B}_{1\ldots n}\in\hat{\mathcal{D}}}\sum_{i=1}^{n}\gamma_{k(i)}R_i^{DPC}(\hat{\pi}_I,\hat{B}_{1\ldots n},\hat{N}_{1\ldots n}) \qquad \text{(4-6)}$$

where $\mathcal{D}=\{(B_1,\ldots,B_m)|\ B_i\succeq 0,\ i=1,\ldots,m \text{ and } \sum_{i=1}^{m}B_i\preceq S\}$ and $\hat{\mathcal{D}}=\{(\hat{B}_1,\ldots,\hat{B}_n)|\ \hat{B}_i\succeq 0,\ i=1,\ldots,n \text{ and } \sum_{i=1}^{n}\hat{B}_i\preceq S\}$ and where $\hat{\pi}_I$ is an identity permutation function over the set $\{1,\ldots,n\}$.

Proof: Let B_i^*, $i=1,\ldots,m$ be the optimizing matrices, B_i, of the optimization problem on the left hand side of (4-6). Furthermore, let $R_i^*=R_i^{DPC}(\pi_I,B_{1\ldots m}^*,N_{1\ldots m})$, $i=1,\ldots,m$. We claim that R_i^* and B_i^* are such that $R_i^*=0$ and $B_i^*=0_{t\times t}\ \forall i$ s.t. $\gamma_i=0$. To show this we observe that we can always modify the choice of B_i's to increase the DPC rates, $R_i^{DPC}(\pi_I,B_{1\ldots m},N_{1\ldots m})$, for i's which correspond to $\gamma_i>0$ at the expense of DPC rates for users, i, which correspond to $\gamma_i=0$, without violating the matrix constraint, S (see Proposition V.1 in Appendix V). This will only increase the target function, $\sum_{i=1}^{m}\gamma_i R_i^{DPC}(\pi_I,B_{1\ldots m},N_{1\ldots m})$, since the rates we reduce are multiplied by $\gamma_i=0$ and the rates we increase are multiplied by $\gamma_i>0$. Thus, we have shown that $R_i^*=0\ \forall i$ s.t. $\gamma_i=0$. Furthermore, by the definition of the rates $R_i^{DPC}(\pi_I,B_{1\ldots m},N_{1\ldots m})$, it is easy to show that as $B_j\succeq 0\ \forall j$, $R_i^{DPC}(\pi_I,B_{1\ldots m},N_{1\ldots m})=0$ if an only if $B_i=0$. Therefore, $B_i^*=0_{t\times t}\ \forall i$ s.t. $\gamma_i=0$.

We can now define, $\hat{B}_i^*=B_{k(i)}^*$, $i=1,\ldots,n$. It is easy to see that $R_i^{DPC}(\pi_I,B_{1\ldots m}^*,N_{1\ldots m})=R_i^{DPC}(\hat{\pi}_I,\hat{B}_{1\ldots n}^*,\hat{N}_{1\ldots n})$ and therefore,

$$\max_{B_{1\ldots m}\in\mathcal{D}}\sum_{i=1}^{m}\gamma_i R_i^{DPC}(\pi_I,B_{1\ldots m},N_{1\ldots m})\le\max_{\hat{B}_{1\ldots n}\in\hat{\mathcal{D}}}\sum_{i=1}^{n}\gamma_{k(i)}R_i^{DPC}(\hat{\pi}_I,\hat{B}_{1\ldots n},\hat{N}_{1\ldots n})$$

On the other hand, we can show that the opposite inequality also holds for the above expression. Let $\hat{B}_i^{**}$, $i=1,\ldots,n$ be the optimizing solution of the optimization problem on the right hand side of (4-6). Define

$$B_i^{**} = \begin{cases} \hat{B}_{k^{-1}(i)}^{**} & i \in \{k(1), \ldots, k(n)\} \\ 0 & \text{otherwise} \end{cases}$$

where k^{-1} is the inverse of the index function such that $k^{-1}(k(i)) = i$. It is easy to see that $R_i^{DPC}(\pi_I, B_{1\ldots m}^{**}, N_{1\ldots m}) = R_i^{DPC}(\hat{\pi}_I, \hat{B}_{1\ldots n}^{**}, \hat{N}_{1\ldots n})$ and therefore,

$$\max_{B_{1\ldots m} \in \mathcal{D}} \sum_{i=1}^{m} \gamma_i R_i^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m}) \geq \max_{\hat{B}_{1\ldots n} \in \hat{\mathcal{D}}} \sum_{i=1}^{n} \gamma_{k(i)} R_i^{DPC}(\hat{\pi}_I, \hat{B}_{1\ldots n}, \hat{N}_{1\ldots n})$$

Thus, the proof is completed. ∎

Proof of Proposition 4.2: To prove the proposition, we will investigate the properties of the gradients of the DPC rates, $R_i^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m})$, at the point where the rate region $\mathcal{R}^{DPC}(\pi_I, S, N_{1\ldots m})$ and the hyperplane, $\{(R_1, \ldots, R_m) \mid \sum_{i=1}^m \gamma_i R_i = b\}$, touch. By the assumptions of the proposition, $\{(R_1, \ldots, R_m) \mid \sum_{i=1}^m \gamma_i R_i = b\}$ is a supporting hyperplane and therefore, there must be such a common rate vector. Let $\overline{R}^* = (R_1^*, \ldots, R_m^*)$ be that rate vector and let $B_1^*, \ldots, B_m^*$ be a sequence of positive semi-definite matrices such that $\sum_{i=1}^m B_i^* \preceq S$ and such that $R_i^{DPC}(\pi_I, B_{1\ldots m}^*, N_{1\ldots m}) = R_i^*,\ i = 1, \ldots, m$. By the definition of the supporting hyperplane, we know that the scalar b and the sequence of matrices, $B_1^*, \ldots, B_m^*$, are the solution and optimizing variables of the following optimization problem

$$\begin{array}{lll} \underset{B_1, \ldots, B_m}{\text{maximize}} & \sum_{i=1}^m \gamma_i R_i^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m}) & \\ \text{such that} & B_i \succeq 0, & \forall i = 1, \ldots, m \\ & \sum_{i=1}^m B_i \preceq S & \end{array}$$

We now note that of all $i = 1, \ldots, m$, only $R_m^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m})$ is a function of B_m and is given by:

$$R_m^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m}) = \frac{1}{2} \log \frac{\left| B_m + \left(\sum_{i=1}^{m-1} B_i\right) + N_m \right|}{\left| \left(\sum_{i=1}^{m-1} B_i\right) + N_m \right|}$$

Therefore, for any given sequence of $m-1$ matrices, $B_1, \ldots, B_{m-1}$, in order to maximize the weighted sum, $\sum_{i=1}^{m} \gamma_i R_i^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m})$, we need to maximize $R_m^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m})$ over all choices of B_m such that $B_m \preceq S - \sum_{i=1}^{m-1} B_i$. From the explicit expression for $R_m^{DPC}(\pi_I, B_{1\ldots m},\, N_{1\ldots m})$, given above, and the fact that $|B| > |A|$ if $B \succeq A \succ 0$ and $B \neq A$ (see Proposition I.2), it is clear that choosing $B_m = S - \sum_{i=1}^{m-1} B_i$ is the best choice. Therefore, $B_m^* = S - \sum_{i=1}^{m-1} B_i^*$ and the matrices $B_1^*, \ldots, B_{m-1}^*$ are the optimizing variables of the following optimization problem,

$$
\begin{array}{lll}
\underset{B_1,\ldots,B_{m-1}}{\text{maximize}} & \sum_{i=1}^{m} \gamma_i R_i^{DPC}(\pi_I, B_{1\ldots m-1}, S, N_{1\ldots m}) & \\
\text{such that} & B_i \succeq 0, & \forall i = 1, \ldots, m-1 \\
& \sum_{i=1}^{m-1} B_i \preceq S &
\end{array}
\tag{4-7}
$$

where $R_i^{DPC}(\pi_I, B_{1\ldots m-1}, S, N_{1\ldots m}) = R_i^{DPC}(\pi_I, B_{1\ldots m}, N_{1\ldots m})$ and where $B_m = S - \sum_{i=1}^{m-1} B_i$. Note that the above optimization problem differs from the previous one in that the optimization is done over $m-1$ matrices instead of m matrices.

Optimization problem (4-7) corresponds to the more general one presented in (3-17) with only one function such that $\sum_{i=1}^{m} \gamma_i R_i^{DPC}(\pi_I, B_{1\ldots m-1}, S, N_{1\ldots m})$ replaces f_0. This function, is differentiable over $B_i \succeq 0$, $\forall i$ and its partial gradients are given by

$$
\begin{aligned}
&\nabla_{B_k}\left(\sum_{i=1}^{m} \gamma_i R_i^{DPC}(\pi_I, B_{1\ldots m-1}, S, N_{1\ldots m})\right) \\
&\quad = \frac{1}{2}\sum_{j=k}^{m-1} \gamma_j \cdot \left(\sum_{i=1}^{j} B_i + N_j\right)^{-1} - \frac{1}{2}\sum_{j=k+1}^{m} \gamma_j \cdot \left(\sum_{i=1}^{j-1} B_i + N_j\right)^{-1}, \quad \forall k = 1, \ldots, m-1
\end{aligned}
\tag{4-8}
$$

In Appendix VII, Proposition VII.2, we show that there is a strict direction for this optimization problem and thus, by proposition 3.3 and equation (4-8), we can write the following $m-1$ equations on the gradients of the DPC rates at $B_1^*,\dots,B_{m-1}^*$:

$$0_{t\times t} = \frac{1}{2}\sum_{j=k}^{m-1}\gamma_j\cdot\left(\sum_{i=1}^{j}B_i^*+N_j\right)^{-1} - \frac{1}{2}\sum_{j=k+1}^{m}\gamma_j\cdot\left(\sum_{i=1}^{j-1}B_i^*+N_j\right)^{-1}+\frac{1}{2}O_k-\frac{1}{2}O_m, \quad \forall k=1\dots m-1$$

where O_k are such that if $r_{B_k^*}>0$, $O_k=V_{B_k}\Lambda_k V_{B_k}^T$ for some $\Lambda_k\succeq 0_{r_{B_k^*}\times r_{B_k^*}}$ ($r_{B_k^*}$ is the number of zero eigenvalues in B_k^*), and if $r_{B_k^*}=0$, then $O_k=0_{t\times t}$. Recall that V_{B_k} is a matrix of size $(t\times r_{B_k^*})$ which columns are orthogonal unit-length eigenvectors of B_k^* that correspond to null eigenvalues. By subtracting equation $k+1$ from the k'th equation (except for $k=m-1$ where the expression is taken as is), we obtain the following $m-1$ equations

$$\gamma_{k+1}\left(\sum_{i=1}^{k}B_i^*+N_{k+1}\right)^{-1}+O_{k+1}=\gamma_k\left(\sum_{i=1}^{k}B_i^*+N_k\right)^{-1}+O_k\ ,\ \ k=1,\dots,m-1 \tag{4-9}$$

We treat separately the case where $\gamma_i>0$ for all $i=1,\dots,m$ and the case where for some $i=1,\dots,m$, $\gamma_i=0$. Once we have proved the proposition under the assumption that $\gamma_i>0$ for all $i=1,\dots,m$ we will be able to use a simple argument to generalize this result to the more general case.

<u>$\gamma_i>0$, $\forall i=1,\dots,m$:</u>

We now assume that $\gamma_i>0$, $\forall i=1,\dots,m$. Therefore, since we also assumed that $\gamma_1\le\cdots\le\gamma_m$ (this assumption was made in Proposition 4.2) and since O_i are defined such that $B_i\cdot O_i=0$, we can prove by induction (see Proposition VIII.1 in Appendix VIII), the existence of an enhanced ADBC with noise increment covariance matrices $\tilde{N}_1',\dots,\tilde{N}_m'$, such that

$$\begin{aligned}\gamma_{k+1}\left(\sum_{i=1}^{k}B_i+N_{k+1}\right)^{-1}+O_{k+1} &= \gamma_{k+1}\left(\sum_{i=1}^{k}B_i+\sum_{i=1}^{k+1}\tilde{N}_i'\right)^{-1}\\ &= \gamma_k\left(\sum_{i=1}^{k}B_i+\sum_{i=1}^{k}\tilde{N}_i'\right)^{-1}=\gamma_k\left(\sum_{i=1}^{k}B_i+N_k\right)^{-1}+O_k\end{aligned} \tag{4-10}$$

and such that

$$\frac{|\sum_{i=1}^{k} B_i + N_k|}{|\sum_{i=1}^{k-1} B_i + N_k|} = \frac{|\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} \tilde{N}'_i|}{|\sum_{i=1}^{k-1} B_i + \sum_{i=1}^{k} \tilde{N}'_i|}, \qquad \forall k = 1, \ldots, m \tag{4-11}$$

The proof of expressions (4-10) and (4-11) is found in Appendix VIII.

By the expressions of R_k^{DPC} (4-1) and R_k^G (3-2), and (4-11), we can see that

$$R_k^* = R_k^{DPC}(\pi_I, B_{1\ldots m}^*, N_{1\ldots m}) = R_k^G(B_{1\ldots m}^*, \tilde{N}'_{1\ldots m}), \quad \forall k = 1, \ldots, m. \tag{4-12}$$

We now complete the proof for the case where $\gamma_i > 0, \forall i = 1, \ldots, m$, by showing that $\{(R_1, \ldots, R_m) \mid \sum_{i=1}^{m} \gamma_i R_i = b\}$ is also a supporting hyperplane of the rate region, $\mathcal{R}^G(S, \tilde{N}'_{1\ldots m})$, and intersects with it at $\overline{R}^*$. For that purpose, it is sufficient to show that $\sum_{i=1}^{m} \gamma_i \cdot R_i^G(B_{1\ldots m}, \tilde{N}'_{1\ldots m})$ is maximized when $B_i = B_i^*$, $i = 1, \ldots, m$. For that reason, we define:

$$\Delta_k = \sum_{i=1}^{k} (B_i - B_i^*), \quad K_k = \sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}'_i, \quad T_k = \sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}'_i$$

and rewrite the difference between the weighted sums as follows,

$$\begin{aligned}
&\sum_{i=1}^{m} \gamma_i R_i^G(B_{1\ldots m}, \tilde{N}'_{1\ldots m}) - \sum_{i=1}^{m} \gamma_i R_i^G(B_{1\ldots m}^*, \tilde{N}'_{1\ldots m}) \\
&= \sum_{k=1}^{m} \gamma_k \cdot \frac{1}{2} \log \frac{\left|\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} \tilde{N}'_i\right|}{\left|\sum_{i=1}^{k-1} B_i + \sum_{i=1}^{k} \tilde{N}'_i\right|} - \sum_{k=1}^{m} \gamma_k \cdot \frac{1}{2} \log \frac{\left|\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}'_i\right|}{\left|\sum_{i=1}^{k-1} B_i^* + \sum_{i=1}^{k} \tilde{N}'_i\right|} \\
&\le \sum_{k=1}^{m-1} \left(\gamma_k \cdot \frac{1}{2} \log \frac{\left|\sum_{i=1}^{k} B_i + \sum_{i=1}^{k} \tilde{N}'_i\right|}{\left|\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k} \tilde{N}'_i\right|} - \gamma_{k+1} \cdot \frac{1}{2} \log \frac{\left|\sum_{i=1}^{k} B_i + \sum_{i=1}^{k+1} \tilde{N}'_i\right|}{\left|\sum_{i=1}^{k} B_i^* + \sum_{i=1}^{k+1} \tilde{N}'_i\right|} \right) \qquad (4\text{-}13) \\
&= \sum_{k=1}^{m-1} \left(\gamma_k \cdot \frac{1}{2} \log \frac{|\Delta_k + K_k|}{|K_k|} - \gamma_{k+1} \cdot \frac{1}{2} \log \frac{|\Delta_k + T_k|}{|T_k|} \right) \\
&= \sum_{k=1}^{m-1} \left(\gamma_k \cdot \frac{1}{2} \log \left| K_k^{-\frac{1}{2}} \Delta_k K_k^{-\frac{\dagger}{2}} + I \right| - \gamma_{k+1} \cdot \frac{1}{2} \log \left| T_k^{-\frac{1}{2}} \Delta_k T_k^{-\frac{\dagger}{2}} + I \right| \right)
\end{aligned}$$

where the inequality results from the fact that $|\sum_{i=1}^{m} B_i + \sum_{i=1}^{m} \tilde{N}'_i| \le |\sum_{i=1}^{m} B_i^* + \sum_{i=1}^{m} \tilde{N}'_i|$ because $\sum_{i=1}^{m} B_i \preceq \sum_{i=1}^{m} B_i^* = S$ (Proposition I.2). By (4-10), we know that $T_k^{-1} = \frac{\gamma_k}{\gamma_{k+1}} K_k^{-1}$.

Therefore, we may write,

$$\begin{aligned}
&\sum_{i=1}^{m}\gamma_i R_i^G(B_{1\ldots m},\tilde{N}'_{1\ldots m})-\sum_{i=1}^{m}\gamma_i R_i^G(B^*_{1\ldots m},\tilde{N}'_{1\ldots m})\\
&\leq \sum_{k=1}^{m-1}\left(\gamma_k\cdot\frac{1}{2}\log\left|K_k^{-\frac{1}{2}}\Delta_k K_k^{-\frac{\dagger}{2}}+I\right|-\gamma_{k+1}\cdot\frac{1}{2}\log\left|\frac{\gamma_k}{\gamma_{k+1}}K_k^{-\frac{1}{2}}\Delta_k K_k^{-\frac{\dagger}{2}}+I\right|\right)\\
&=\sum_{k=1}^{m-1}\frac{1}{2}\cdot\gamma_{k+1}\cdot\left(\begin{array}{r}\left(\frac{\gamma_k}{\gamma_{k+1}}\log\left|K_k^{-\frac{1}{2}}\Delta_k K_k^{-\frac{\dagger}{2}}+I\right|+\left(1-\frac{\gamma_k}{\gamma_{k+1}}\right)\log|I|\right)\\
-\left(\log\left|\frac{\gamma_k}{\gamma_{k+1}}(K_k^{-\frac{1}{2}}\Delta_k K_k^{-\frac{\dagger}{2}}+I)+\left(1-\frac{\gamma_k}{\gamma_{k+1}}\right)\cdot I\right|\right)\end{array}\right)
\end{aligned} \tag{4-14}$$

As $0<\frac{\gamma_k}{\gamma_{k+1}}\leq 1$, by the concavity ($\bigcap$) of the $\log\det$ function and Jensen's inequality, each of the summands in the last equality are negative and therefore, we may write:

$$\sum_{i=1}^{m}\gamma_i R_i^G(B_{1\ldots m},\tilde{N}'_{1\ldots m})-\sum_{i=1}^{m}\gamma_i R_i^G(B^*_{1\ldots m},\tilde{N}'_{1\ldots m})\leq 0$$

for all positive semi-definite matrices B_i with power covariance constraint, S. This completes the proof of the proposition for the case $\gamma_i>0\ \forall i$.

$\gamma_i\geq 0,\ \forall i=1,\ldots,m$:

Finally, we expand the proof to the case of any set of non-negative scalars, $\gamma_1,\ldots,\gamma_m$ where at least one of them is strictly positive. The method we will apply here is very similar to the one used in the proof of Theorem 3.2. Let n, $1\leq n<m$ denote the number of strictly positive scalars, γ_i, $i=1,\ldots,m$. Since we assume that $\gamma_1\leq\cdots\leq\gamma_m$, it is clear that $\gamma_i=0$, $\forall i=1,\ldots,m-n$ and $\gamma_i>0$, $\forall i=m-n+1,\ldots,m$. We can now define a compact channel which is also an AMBC with the same covariance matrix constraint, S, and with noise covariance matrices, $\hat{N}_1,\ldots,\hat{N}_n$ such that:

$$\hat{N}_i=N_{i+m-n}\ ,\quad i=1,\ldots,n$$

Similarly, we define a compact hyperplane, $\{(\hat{R}_1,\ldots,\hat{R}_n)\in\mathbb{R}^n\mid\sum_{i=1}^{n}\gamma_{i+m-n}\hat{R}_i=b\}$. By Proposition 4.3, we conclude that

$$\left\{(\hat{R}_1,\ldots,\hat{R}_n)\in\mathbb{R}^n\ \Big|\ \sum_{i=1}^{n}\gamma_{i+m-n}\hat{R}_i=b\right\}$$ is a supporting hyperplane of $\mathcal{R}^{DPC}(\hat{\pi}_I,S,\hat{N}_{1\ldots n})$

where $\hat{\pi}_I$ is now the identity permutation over the set $\{1,\ldots,n\}$. Therefore, we can use the above proof for the case where γ_i are strictly positive to show that the proposition holds for the compact channel. That is, we can find an enhanced ADBC, with noise increment covariance matrices, $\hat{\tilde{N}}'_1,\ldots,\hat{\tilde{N}}'_n$, such that

$$\left\{(\hat{R}_1,\ldots,\hat{R}_n)\in\mathbb{R}^n \mid \sum_{i=1}^{n}\gamma_{i+m-n}\hat{R}_i=b\right\} \text{ is a supporting hyperplane of } \mathcal{R}^G(S,\hat{\tilde{N}}'_{1\ldots n})$$

We now define an enhanced ADBC for the original channel using the enhanced ADBC of the compact channel. The noise increment covariance matrices, $\tilde{N}'_1,\ldots,\tilde{N}'_m$, are defined as follows:

$$\tilde{N}'_i=\begin{cases}\beta_i\hat{\tilde{N}}'_1 & i\le m-n\\ \hat{\tilde{N}}'_{i-(m-n)} & \forall i>m-n\end{cases}$$

where $\beta_i\ge 0$, $i=1\ldots k(1)$ are chosen such that $\sum_{i=1}^{k(1)}\beta_i=1$ and such that $0\prec\left(\sum_{i=1}^{j}\beta_i\right)\cdot\hat{\tilde{N}}'_1\preceq N_j$, $\forall j=1,\ldots,m-n$. As $N_j\succ 0$, $\forall j=1,\ldots,m$, it is possible to find such β_i. Clearly, we have defined an enhanced ADBC for the original channel. Furthermore, as $\{(\hat{R}_1,\ldots,\hat{R}_n)\in\mathbb{R}^n\mid\sum_{i=1}^{n}\gamma_{i+m-n}\hat{R}_i=b\}$ is a supporting hyperplane of $\mathcal{R}^G(S,\hat{\tilde{N}}'_{1\ldots n})$, we can use Proposition 4.3 to show that

$$\left\{(R_1,\ldots,R_m)\in\mathbb{R}^m \mid \sum_{i=1}^{m}\gamma_i R_i=b\right\} \text{ is a supporting hyperplane of } \mathcal{R}^G(S,\tilde{N}'_{1\ldots m}).$$

■

4.4 AMBC Example

The following example illustrates the statements of Proposition 4.2. In this example, we consider a two user non-degraded AMBC under a covariance matrix input constraint, where the transmitter and each of the receivers have four antennas such that:

$$N_1 \approx \begin{pmatrix} 5.20425 & 2.96515 & 3.31237 & -0.67287 \\ 2.96515 & 8.40350 & 2.15397 & 0.51274 \\ 3.31237 & 2.15397 & 4.01902 & -1.23370 \\ -0.67287 & 0.51274 & -1.23370 & 1.15827 \end{pmatrix},$$

$$N_2 \approx \begin{pmatrix} 5.24564 & -3.80767 & -1.49311 & 3.69896 \\ -3.80767 & 8.28347 & 2.97520 & -5.87397 \\ -1.49311 & 2.97520 & 2.39828 & -4.53655 \\ 3.69896 & -5.87397 & -4.53655 & 10.10308 \end{pmatrix}$$

$$\text{and } S \approx \begin{pmatrix} 4.93783 & -0.78695 & -3.16956 & 3.35222 \\ -0.78695 & 3.32543 & -2.51664 & -0.12348 \\ -3.16956 & -2.51664 & 5.08241 & -2.40199 \\ 3.35222 & -0.12348 & -2.40199 & 2.92163 \end{pmatrix}.$$

The boundary of the dirty paper coding region, $\mathcal{R}^{SPC}(S, N_{1,2})$, is plotted by a solid curve in Figure 2. The dotted line, in the same graph, corresponds to the hyperplane $\{(R_1, R_2) \mid \frac{1}{2} \cdot R_1 + 1 \cdot R_2 = 2.01574\}$. In addition, we plotted a dashed curve, corresponding to the boundary of the Gaussian region of an enhanced and degraded version of the above AMBC with noise increment covariances

$$\tilde{N}_1' \approx \begin{pmatrix} 1.41956 & -0.68733 & 0.01387 & 0.03833 \\ -0.68733 & 1.66279 & -0.03823 & 0.29121 \\ 0.01387 & -0.03823 & 0.25974 & -0.27691 \\ 0.03833 & 0.29121 & -0.27691 & 0.76268 \end{pmatrix}$$

$$\text{and} \quad \tilde{N}_2' \approx \begin{pmatrix} 3.37933 & -1.71534 & -0.64685 & 1.92184 \\ -1.71534 & 2.20204 & 0.30836 & -0.69681 \\ -0.64685 & 0.30836 & 0.48250 & -0.91193 \\ 1.92184 & -0.69681 & -0.91193 & 2.57291 \end{pmatrix}.$$

As predicted by Proposition 4.2, we can see from Figure 2 that for the given hyperplane, we can find an enhanced and degraded version of the channel such that its Gaussian region is

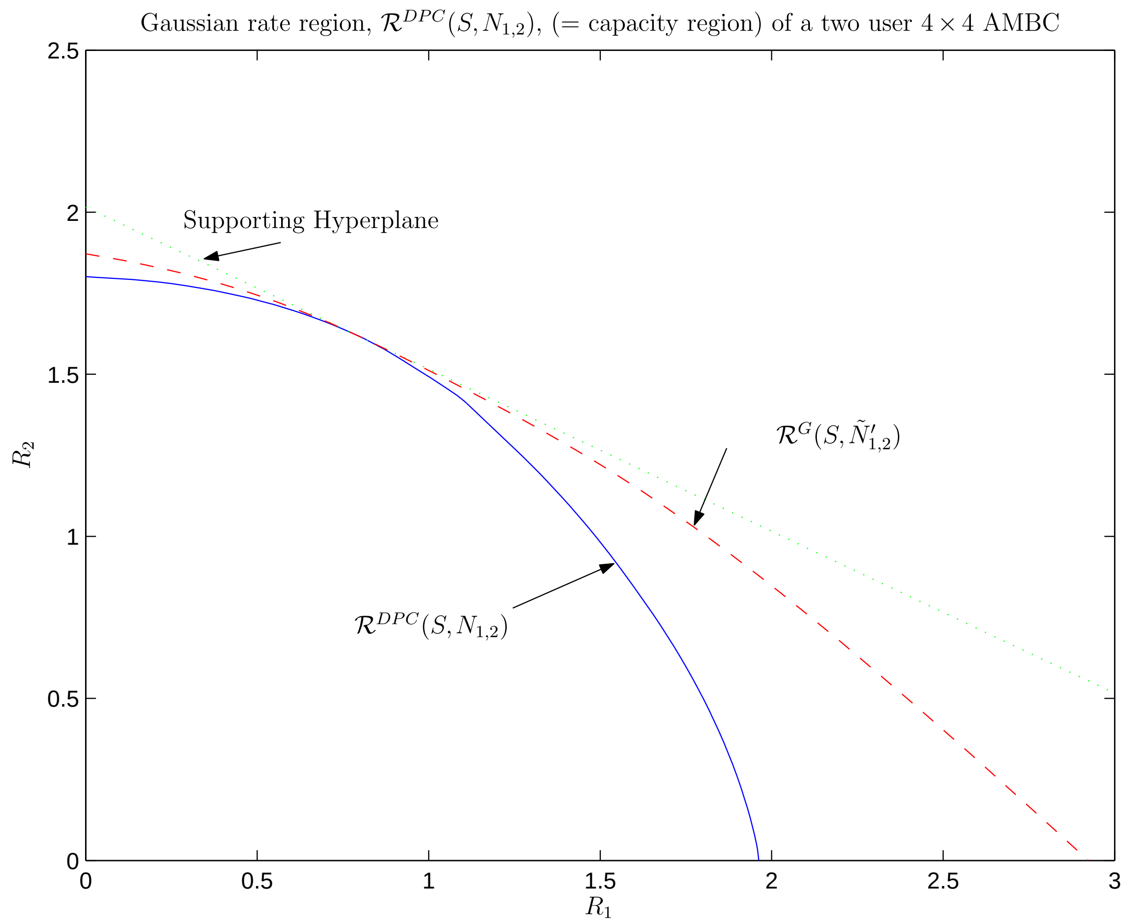

Fig. 2. Illustration of the results of Proposition 4.2

supported by the same hyperplane. Furthermore, in relation to the proof of this proposition, we note that the hyperplane and both curves intersect at $(R_1, R_2) \approx (0.79656, 1.61746)$.

5. Extension to the General MIMO BC

We now consider the GMBC (expression (1-1)) which, unlike the ADBC and AMBC, is characterized by both the noise covariance matrices, $N_1, \ldots, N_m$, and gain matrices, $H_1, \ldots, H_m$. We will prove that the DPC rate region [5], [23], [19], [17] of this channel, coincides with the capacity region.

5.1 GMBC - Main Results

We begin by characterizing the DPC rate region of the GMBC. Given an average transmit covariance matrix limitation, $E[\mathbf{x}\mathbf{x}^T] \preceq S$, and a set of positive semi-definite matrices, $B_1, B_2, \ldots, B_m$ ($B_k \succeq 0\, \forall k \in 1 \ldots m$), such that $\sum_{i=1}^{m} B_i \preceq S$, and a permutation function π, the following rates are achievable in the GMBC using a DPC scheme [5], [23], [19], [17]:

$$R_k \leq R^{DPC}_{\pi^{-1}(k)}(\pi, B_{1\ldots m}, N_{1\ldots m}, H_{1\ldots m}) \,, \qquad \forall k \in 1, \ldots, m$$

where

$$R^{DPC}_l(\pi, B_{1\ldots m}, N_{1\ldots m}, H_{1\ldots m}) = \frac{1}{2} \log \frac{\left| \left(H_{\pi(l)} \left(\sum_{i=1}^{l} B_{\pi(i)} \right) H^T_{\pi(l)} + N_{\pi(l)} \right) \right|}{\left| \left(H_{\pi(l)} \left(\sum_{i=1}^{l-1} B_{\pi(i)} \right) H^T_{\pi(l)} + N_{\pi(l)} \right) \right|} \quad l = 2, 3, \ldots, m \tag{5-1}$$

We now define the DPC achievable rate region of a GMBC:

Definition 5.1 (DPC rate-region of a GMBC) Let S be a positive semi-definite matrix, then, we define the DPC rate region, $\mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m})$, for the GMBC with a covariance matrix constraint, S, as the following convex closure:

$$\mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m}) = \mathbf{conv} \left\{ \bigcup_{\pi \in \Pi} \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m}, H_{1\ldots m}) \right\} \tag{5-2}$$

where

$$\mathcal{R}^{DPC}(\pi, S, N_{1\ldots m}, H_{1\ldots m}) = \left\{ \begin{array}{c} \left(\begin{array}{l} R^{DPC}_1(\pi, B_{1\ldots m}, N_{1\ldots m}, H_{1\ldots m}), \ldots, \\ \qquad R^{DPC}_m(\pi, B_{1\ldots m}, N_{1\ldots m}, H_{1\ldots m}) \end{array} \right) \Bigg| \\ \text{s.t.} \quad S - \sum_{i=1}^{m} B_i \succeq 0, \;\; B_i \succeq 0 \;\; \forall i = 1, \ldots, m \end{array} \right\} \tag{5-3}$$

Theorem 5.1: Let $\mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$ denote the capacity region of the GMBC under a covariance matrix constraint, $S \succeq 0$, then $\mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m}) = \mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m})$.

As $\mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m})$ is contiguous in S (for proof, see Proposition III.3 in Appendix III), the following corollary follows immediately by Proposition 2.1.

Corollary 5.1: Let $\mathcal{C}(P, N_{1\ldots m}, H_{1\ldots m})$ denote the capacity region of the GMBC under a total power constraint, $P \geq 0$, then

$$\mathcal{C}(P, N_{1\ldots m}, H_{1\ldots m}) = \bigcup_{\substack{S \succeq 0 \\ \text{s.t.} \\ \text{tr}\{S\} \leq P}} \mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m})$$

5.2 Proof of Theorem 5.1

As $\mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m}) \subseteq \mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m})$, we need only to show that $\mathcal{C}(S, N_{1\ldots m}, H_{1\ldots m}) \subseteq \mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m})$. We do that in two steps. In the first step, we use SVD to rewrite the original MIMO BC (1-1) as a MIMO BC with square gain matrices of sizes $t \times t$. In the second step, we add a small perturbation to some of the entries of the gain matrices such that they will be invertible and such that their capacity region is enlarged. As the gain matrices are invertible, we will be able to infer the capacity region of the new channel from that of an equivalent AMBC. We will complete the proof by showing that the capacity region of the original MIMO BC can be obtained by a limit process on the capacity region of the perturbed MIMO BC.

For the first step of our proof, we will consider a sequence of variants of the GMBC which preserve both the DPC rate region and the capacity region of the original GMBC.

As a reminder, the channel we are considering is given by:

$$\mathbf{y}_i = H_i \mathbf{x} + \mathbf{n}_i = U_i \Lambda_i V_i \mathbf{x} + \mathbf{n}_i \;, \qquad i = 1, 2, \ldots, m \tag{5-4}$$

where $U_i \Lambda_i V_i$ is the singular value decomposition of H_i and where U_i and V_i are unitary matrices of sizes $r_i \times r_i$ and $t \times t$ and Λ_i is a diagonal matrix of size $r_i \times t$. We assume without loss of generality that the singular values of Λ_i are arranged in rising order such that all non-zero singular values are located on the lower right corner of Λ_i. That is, if we define $r_{\Lambda_i} = \text{rank}(\Lambda_i)$, then $\Lambda_{i(m-t+r_i,m)} > 0$, $\forall m = t - r_{\Lambda_i} + 1, \ldots, t$ and all other elements of Λ_i are zero. Thus, the first $(r_i - r_{\Lambda_i})$ rows of Λ_i are zero.

We multiply the channel outputs at each of the users by U_i^T. As this manipulation is a reversible one, it has no effect on the capacity region of the channel. Furthermore, the DPC rate-region remains the same. We now get the following channel:

$$\mathbf{y}'_i = \Lambda_i V_i \mathbf{x} + \mathbf{n}'_i\ , \qquad i = 1, 2, \ldots, m \tag{5-5}$$

where $E[\mathbf{n}'_i \mathbf{n}'^T_i] = U_i^T N_i U_i \ = N'_i$. Next, we rewrite N'_i such that

$$N'_i = \begin{pmatrix} N'^A_i & N'^B_i \\ (N'^B_i)^T & N'^C_i \end{pmatrix}\ , \quad i = 1, \ldots, m$$

where N'^A_i, N'^B_i and N'^C_i are of sizes $(r_i - r_{\Lambda_i}) \times (r_i - r_{\Lambda_i})$, $r_{\Lambda_i} \times (r_i - r_{\Lambda_i})$ and $r_{\Lambda_i} \times r_{\Lambda_i}$. We define,

$$D_i = \begin{pmatrix} I_{(r_i - r_{\Lambda_i}) \times (r_i - r_{\Lambda_i})} & 0_{(r_i - r_{\Lambda_i}) \times r_{\Lambda_i}} \\ -(N'^B_i)^T (N'^A_i)^{-1} & I_{r_{\Lambda_i} \times r_{\Lambda_i}} \end{pmatrix}$$

We now create a new variation of the channel by multiplying the received vector of each user by its appropriate D_i.

$$\mathbf{y}''_i = D_i \Lambda_i V_i \mathbf{x} + D_i \mathbf{n}'_i = \Lambda_i V_i \mathbf{x} + D_i \mathbf{n}'_i = H''_i \mathbf{x} + \mathbf{n}''_i\ , \qquad i = 1, 2, \ldots, m \tag{5-6}$$

where the second equality is a consequence of having $(r_i - r_{\Lambda_i})$ zero rows at the top of Λ_i. As D_i is a reversible transformation, the capacity and DPC rate regions remain unchanged for this new channel. Furthermore,

$$N''_i = E[\mathbf{n}''_i \mathbf{n}''^T_i] = D_i N'_i D_i^T = \begin{pmatrix} N'^A_i & 0 \\ 0 & N''^C_i \end{pmatrix} \tag{5-7}$$

is block diagonal with N'^A_i and N''^C_i of sizes $(r_i - r_{\Lambda_i}) \times (r_i - r_{\Lambda_i})$ and $r_{\Lambda_i} \times r_{\Lambda_i}$ on the diagonal and therefore, as $\mathbf{n}''_i$ is a Gaussian vector, the noise at the first $(r_i - r_{\Lambda_i})$ antennas is independent of the noise at the other r_{Λ_i}-antennas.

As the first $r_i - r_{\Lambda_i}$ rows in H''_i are all zero, $H''_i = \begin{pmatrix} 0_{(r_i - r_{\Lambda_i}) \times t} \\ H''^B_i \end{pmatrix}$, it is clear that the first $r_i - r_{\Lambda_i}$ receive antennas at each user are not affected by the transmitted signal. Furthermore, as the noise vector at these antennas is independent of the noise at the other r_{Λ_i}-antennas (by the structure of the covariance matrix N_i'' (5-7)), it is clear that the first $r_i - r_{\Lambda_i}$ antennas (in (5-6)) do not play a role in the ML receiver. Hence, we may remove these antennas altogether, without any effect on the capacity region. In addition, due to the block diagonal structure of the matrix, N_i'', removing the first $r - r_{\Lambda_i}$ antennas will not affect the DPC rate region as well, as is shown in the following equation:

$$
\begin{aligned}
&R_i^{DPC}(\pi, B_{1\ldots m}, N''_{1\ldots m}, H''_{1\ldots m}) \\
&\quad = \frac{1}{2}\log \frac{\left| H''_{\pi(i)} \left(\sum_{j=1}^{i} B_{\pi(j)}\right) H''^{\,T}_{\pi(i)} + N''_{\pi(i)} \right|}{\left| H''_{\pi(i)} \left(\sum_{j=1}^{i-1} B_{\pi(j)}\right) H''^{\,T}_{\pi(i)} + N''_{\pi(i)} \right|} \\
&\quad = \frac{1}{2}\log \frac{\left| N'^{\,A}_{\pi(i)} \right|}{\left| N'^{\,A}_{\pi(i)} \right|} \cdot \frac{\left| H''^{\,B}_{\pi(i)} \left(\sum_{j=1}^{i} B_{\pi(j)}\right) H''^{\,BT}_{\pi(i)} + N''^{\,C}_{\pi(i)} \right|}{\left| H''^{\,B}_{\pi(i)} \left(\sum_{j=1}^{i-1} B_{\pi(j)}\right) H''^{\,BT}_{\pi(i)} + N''^{\,C}_{\pi(i)} \right|} = R_i^{DPC}(\pi, B_{1\ldots m}, N''^{C}_{1\ldots m}, H''^{B}_{1\ldots m})
\end{aligned}
$$

Alternatively, adding zero rows to H_i'' at its top (or alternatively, adding zero rows at the top of Λ_i) and appropriately adding receive antennas with independent (of the other antennas) Gaussian noise will also preserve the capacity and DPC rate regions.

Therefore, we may write yet another variant of the channel, this time with t transmit antennas and t receive antennas for each user.

$$
\hat{\mathbf{y}}_i = \hat{H}_i \mathbf{x} + \hat{\mathbf{n}}_i \,, \qquad i = 1, 2, \ldots, m \tag{5-8}
$$

where $\hat{H}_i = \hat{\Lambda}_i V_i$ and where $\hat{\Lambda}_i$ is a $t \times t$ diagonal matrix with the first $t - r_{\Lambda_i}$ elements equal to zero and the other r_{Λ_i} equal to the r_{Λ_i}-elements on the lower right diagonal of Λ_i such that

$$
\hat{\Lambda}_{i(m,m)} = \begin{cases} 0 & 1 \le m \le t - r_{\Lambda_i} \\ \Lambda_{i(m-t+r_i,m)} & t - r_{\Lambda_i} < m \le t \end{cases}
$$

Again, $\hat{N}_i = \begin{pmatrix} \hat{N}^A & 0 \\ 0 & N_i''^{C} \end{pmatrix}$ is block diagonal where $\hat{N}^A$ is of size $(t - r_{\Lambda_i}) \times (t - r_{\Lambda_i})$ and again both the capacity and DPC regions are preserved.

To complete the proof of Theorem 5.1, it is sufficient to show that

$$\mathcal{C}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m}) \subseteq \mathcal{R}^{DPC}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m}).$$

To that end, we proceed with the second step of our proof and define a new channel, which this time, does not preserve the capacity region.

$$\breve{\mathbf{y}}_i = \breve{H}_i \mathbf{x} + \breve{\mathbf{n}}_i \ , \qquad i = 1, 2, \ldots, m \tag{5-9}$$

where $\breve{H}_i = (\hat{\lambda}_i + \alpha \breve{I}_i) V_i$ and $\breve{I}$ is a $t \times t$ diagonal matrix such that

$$\breve{I}_i(m, n) = \begin{cases} 1 & m = n \leq t - r_{\Lambda_i} \\ \\ 0 & \text{Otherwise} \end{cases}$$

and some $\alpha > 0$.

Note that the last r_{Λ_i} rows of $\breve{H}_i$ are identical to those of $\hat{H}_i$. Therefore, as the ML receiver of the channel in (5-8) only observes that lower r_{Λ_i} antennas, any code-book and ML receiver, designed for the channel in (5-8) will achieve the exact same results in the channel given in (5-9). Thus, it is clear that:

$$\mathcal{C}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m}) \subseteq \mathcal{C}(S, \breve{N}_{1\ldots m}, \breve{H}_{1\ldots m}) \quad \forall \alpha > 0$$

Additionally, note that $\breve{H}_i$ is invertible and multiplying each receive vector by $\breve{H}_i^{-1}$ will yield an AMBC with the same DPC and capacity regions as that of the channel in (5-9). As the DPC and capacity regions of an AMBC coincide (Theorem 4.1), we can write:

$$\mathcal{C}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m}) \subseteq \mathcal{R}^{DPC}(S, \breve{N}_{1\ldots m}, \breve{H}_{1\ldots m}) \quad \forall \alpha > 0 \tag{5-10}$$

We now make use of the fact that we can choose α arbitrarily and note that due to the contiguity of the $\log\det$ function over positive definite matrices, we have

$$\lim_{\alpha \to 0} R_k^{DPC}(\pi, B_{1\ldots m}, \breve{N}_{1\ldots m}, \breve{H}_{1\ldots m}) = R_k^{DPC}(\pi, B_{1\ldots m}, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m}) \quad k = 1 \ldots m$$

where the convergence is uniform over all positive semi-definite matrices B_i s.t. $\sum B_i \preceq S$. Thus, by the uniform convergence property and by (5-10), it is clear that for every $\epsilon > 0$ and every rate vector $\overline{R} \in \mathcal{C}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m})$, the ϵ-ball around $\overline{R}$ contains a vector $\overline{R}' \in \mathcal{R}^{DPC}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m})$. Therefore, $\mathcal{C}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m}) \subseteq \text{closure}\{\mathcal{R}^{DPC}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m})\}$. However, as $\mathcal{R}^{DPC}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m})\}$ is closed, $\mathcal{C}(S, \hat{N}_{1\ldots m},\ \hat{H}_{1\ldots m}) \subseteq \mathcal{R}^{DPC}(S, \hat{N}_{1\ldots m}, \hat{H}_{1\ldots m})$.

6. Summary

We have characterized the capacity region of the Gaussian MIMO broadcast channel (BC) and proved that it coincides with the dirty paper coding (DPC) rate region. We have shown this for a wide range of input constraints such as the average total power and the input covariance constraints. In general, our results apply to any input constraint such that the input covariance matrix lies in a compact set of positive semi-definite matrices.

For that purpose we have introduced a new notion of an enhanced channel. Using the enhanced channel, we were able to modify Bergmans' proof [1] to give a converse for the capacity region of an aligned and degraded Gaussian vector BC (ADBC). The modification was based on the fact that Bergmans' proof could be directly extended to the vector case when the enhanced channel was considered instead of the original one. By associating an ADBC with points on the boundary of the DPC region of an aligned (and not necessarily degraded) MIMO BC (AMBC), we were able to extend our converse to the AMBC and then, to the general Gaussian MIMO BC.

Appendix

I. Basic Inequalities

A The Entropy Power Inequality

Proposition I.1 (Entropy Power Inequality (EPI) [3]) Let $\mathbf{x}$ and $\mathbf{y}$ be statistically independent vectors of length n. Then, we may write the following inequality:

$$e^{\frac{2}{n}H(\mathbf{x}+\mathbf{y})} \geq e^{\frac{2}{n}H(\mathbf{x})} + e^{\frac{2}{n}H(\mathbf{y})}$$

with equality if, and only if, $\mathbf{x}$ and $\mathbf{y}$ are normally distributed with proportional covariances.

Corollary I.1: Let $\mathbf{x}$ and $\mathbf{y}$ be statistically independent vectors of length n. in addition, let W be a random variable which takes it values from a finite set. Further more assume that $\mathbf{y}$ is independent of the pair $(\mathbf{x}, W)$. Then, we can write:

$$H(\mathbf{x}+\mathbf{y}|W) \geq \frac{n}{2}\log\left(e^{\frac{2}{n}H(\mathbf{x}|W)} + e^{\frac{2}{n}H(\mathbf{y})}\right)$$

Proof:

$$\begin{aligned} H(\mathbf{x}+\mathbf{y}|W) &= \sum_w P(W=w)H(\mathbf{x}+\mathbf{y}|W=w) \\ &\geq \sum_w P(W=w)\frac{n}{2}\log\left(e^{\frac{2}{n}H(\mathbf{x}|W=w)} + e^{\frac{2}{n}H(\mathbf{y})}\right) \\ &\geq \frac{n}{2}\log\left(e^{\frac{2}{n}H(\mathbf{x}|W)} + e^{\frac{2}{n}H(\mathbf{y})}\right) \end{aligned}$$

where the first inequality is due to the EPI and the fact that $\mathbf{y}$ is independent of both $\mathbf{x}$ and W. The second inequality is due to Jensen inequality and the fact that $\log(e^x + c)$ is a convex (U) function in x. ■

B Matrix Inequalities

Proposition I.2: Let $A \succ 0$ and $B \succ 0$ be positive definite Hermitian matrices such that $B \succeq A$ and such that $B \neq A$. Then,

$$|B| > |A|$$

Proof: B can be written as $A + \Delta$ where $\Delta \succeq 0$ such that $\Delta \neq 0$. Therefore, $|B| = |A| \cdot |I + A^{-\frac{\dagger}{2}} \Delta A^{-\frac{1}{2}}|$. Since $A^{-\frac{\dagger}{2}} \Delta A^{-\frac{1}{2}} \neq 0$ is positive semi-definite, we have $|I + A^{-\frac{\dagger}{2}} \Delta A^{-\frac{1}{2}}| > 1$. ■

Proposition I.3: Let $A \succ 0$ and $B \succ 0$ be positive definite Hermitian matrices such that $B \succ A$ ($B \succeq A$). Then,

$$B^{-1} \prec A^{-1} \qquad \big(B^{-1} \preceq A^{-1}\big)$$

Proof: B can be written as $A + \Delta$ where $\Delta \succeq 0$. Therefore,

$$B^{-1} = A^{-\frac{1}{2}}(I + A^{-\frac{\dagger}{2}} \Delta A^{-\frac{1}{2}})^{-1} A^{-\frac{\dagger}{2}}$$

Hence,

$$A^{-1} - B^{-1} = A^{-\frac{1}{2}}\big(I - (I + A^{-\frac{\dagger}{2}} \Delta A^{-\frac{1}{2}})^{-1}\big) A^{-\frac{\dagger}{2}}$$

Since the eigenvalues of $(I + A^{-\frac{\dagger}{2}} \Delta A^{-\frac{1}{2}})^{-1}$ are all smaller than (or equal to) 1, $I - (I + A^{-\frac{\dagger}{2}} \Delta A^{-\frac{1}{2}})^{-1} \succ 0$ (or $\succeq$) and therefore, $A^{-1} - B^{-1} \succ 0$ (or $\succeq$). ■

Proposition I.4: Let A, B and Δ be Hermitian matrices such that $A \succeq 0$, $B \succ 0$ and $\Delta \succeq 0$, then

$$\frac{|A + B + \Delta|}{|B + \Delta|} \leq \frac{|A + B|}{|B|}$$

Proof:

$$\begin{aligned}\frac{|A+B+\Delta|}{|B+\Delta|} &= |(B+\Delta)^{-\frac{\dagger}{2}}A(B+\Delta)^{-\frac{1}{2}}+I| \\ &= |A^{\frac{1}{2}}(B+\Delta)^{-1}A^{\frac{\dagger}{2}}+I| \\ &\leq |A^{\frac{1}{2}}B^{-1}A^{\frac{\dagger}{2}}+I| \\ &= |B^{-\frac{\dagger}{2}}AB^{-\frac{1}{2}}+I| = \frac{|A+B|}{|B|}\end{aligned}$$

where the inequality is a result of Proposition I.3 and I.2 (the inequality is not strict since we might have $A^{\frac{1}{2}}(B+\Delta)^{-1}A^{\frac{\dagger}{2}} = A^{\frac{1}{2}}B^{-1}A^{\frac{\dagger}{2}}$). ■

C The Minkowski Inequality

Proposition I.5 (Minkowski's Inequality) Let K_1 and K_2 be two $n \times n$ positive semi-definite matrices. Then,

$$|K_1 + K_2|^{\frac{1}{n}} \geq |K_1|^{\frac{1}{n}} + |K_2|^{\frac{1}{n}}$$

with equality iff K_1 and K_2 are proportional.

II. Proof of Proposition 2.2

Proof: Let us first define an intermediate and equivalent channel with t antennas at the transmitter and each of the receivers by multiplying each of the receive vectors, $\mathbf{y}_i,\ i = 1, \ldots, m$ by U_S^T. The new channel takes the following form:

$$U_S^T\mathbf{y}_i = U_S^T\mathbf{x} + U_S^T\mathbf{n}_i = \mathbf{x}' + \mathbf{n}_i' = \mathbf{y}_i', \qquad i = 1, \ldots, m \tag{II-1}$$

where there is a matrix power constraint, $S' = U_S^T S U_S = \Lambda_S$ on the input vector, $\mathbf{x}'$, and additive real Gaussian noise vectors with a covariance matrices, $N_i' = E[\mathbf{n}_i'\mathbf{n}_i'^T] = U_S^T N_i U_S$. Since the transformation that was done on channel outputs is reversible, any codebook designed for the channel in (II-1) can be modified to work in the original AMBC (and vice-versa) with the same probability of error. Therefore, the capacity region of the intermediate channel is exactly the same as that of the original BC.

To accommodate future calculations, we define the sub matrices: N_A, N_B and N_C of sizes $(t-r_S)\times(t-r_S)$, $(t-r_S)\times r_S$ and $r_S \times r_S$, such that

$$N_i' = U_S^T N_i U_S = \begin{pmatrix} N_i^A & N_i^B \\ (N_i^B)^T & N_i^C \end{pmatrix}, \quad i = 1,\ldots,m$$

Note that N_i^A and N_i^C are symmetric and positive semi-definite.

We now recollect that we assumed that S is not full ranked and that the first $t-r_S$ values on the diagonal of Λ_S are zero and the rest are strictly positive. Therefore, no signal will be transmitted through the first $t-r_S$ input elements of the intermediate channel (II-1). Notice that the users on the receiving ends, receive pure channel noise on the first $t-r_S$ receiving antennas. Hence, each user can use the signal on first $t-r_S$ antennas to cancel out the effect of the noise on the rest of the antennas such that the resultant accumulated noise in the first $t-r_S$ antennas will be de-correlated from that of the other r_S antennas. Hence, we can define a second intermediate channel such that:

$$\mathbf{y}_i'' = D_i\mathbf{y}_i' = D_i\mathbf{x}' + D_i\mathbf{n}_i' = \mathbf{x}' + \mathbf{n}_i'' = \mathbf{x}'' + \mathbf{n}_i'', \qquad i = 1,\ldots,m \tag{II-2}$$

where

$$D_i = \begin{pmatrix} I_{(t-r_S)\times(t-r_S)} & 0_{(t-r_S)\times r_S} \\ -(N_i^B)^T(N_i^A)^{-1} & I_{r_S\times r_S} \end{pmatrix}$$

and where the third equality follows from the fact that no signal is sent through the first $t-r_S$ transmit antennas. Again, since the transformation was reversible, the capacity region of the channel in (II-2) is identical to that of (II-1) and to that of the original AMBC. Furthermore, the noise is still Gaussian and the matrix power constraint remains Λ_S.

One can verify that the resultant noise covariance at output, i, of channel (II-2) is given by

$$N_i'' = D_i N_i' D_i^T = \begin{pmatrix} N_i^A & 0 \\ 0 & N_i^C - \left(N_i^B\right)^T \left(N_i^A\right)^{-1} \left(N_i^B\right) \end{pmatrix}$$

However, because the noise vectors are Gaussian, the fact that the noise at the first $t - r_S$ receive antennas (at each user) is uncorrelated with the noise at the other r_S antennas means that the first $t - r_S$ channel output signals are statistically independent of the other r_S output signals (since the former outputs carry no information signal). Moreover, since only the latter r_S carry the information signal, the decoder can disregard the first $1, \ldots, t - r_S$ receive antenna signals for each user without suffering any degradation in the code's performance (because these signal have no effect on the decision made by the ML or MAP decoder).

Thus, we define a new equivalent AMBC with only r_S transmit antennas such that

$$\hat{\mathbf{y}}_i = E'\mathbf{y}_i'' = E'\mathbf{x}'' + E'\mathbf{n}_i'' = \hat{\mathbf{x}} + \hat{\mathbf{n}}_i, \quad i = 1, \ldots, m \tag{II-3}$$

where $E' = \left[0_{r_S \times (t - r_S)} \ I_{r_S \times r_S}\right]$ and where $\hat{\mathbf{x}}$ is a vector of size $r_S \times 1$ and a full ranked input covariance constraint $\hat{S} = E_i' \Lambda_S E_i'^T$. The additive noise $\hat{\mathbf{n}}_i$ is a real Gaussian vector with a covariance matrix, $\hat{N}_i = N_i^C - \left(N_i^B\right)^T \left(N_i^A\right)^{-1} \left(N_i^B\right)$.

As removing the first $t - r_S$ receive antennas did not cause any degradation in performance, it is clear that any codebook that was designed for channel (II-2) can be modified (chop off the first r_S channel inputs) to work in channel (II-3) with the same results. It is also clear that every codebook that was designed to work in (II-3), will also work, with some minor modifications (pad with zeros the first $t - r_S$ inputs), in the intermediate channel (II-2) with the exact same results. Therefore (II-3) and (II-2) have the same capacity region and hence our equivalent channel (II-3) will have the same capacity region as the original AMBC.

Finally, we need to show that if $N_1 \preceq \cdots \preceq N_m$, then $\hat{N}_1 \preceq \cdots \preceq \hat{N}_m$ (noise covariance matrices of the channel in (II-3)). Assume that indeed $N_1 \preceq \cdots \preceq N_m$. It is easily shown that $N'_1 \preceq \cdots \preceq N'_m$ where N'_i are the noise covariances of the intermediate channel (II-1). We will need to show that $\hat{N}_{i+1} - \hat{N}_i \succeq 0$, $i = 1, \ldots, m-1$. For that purpose, we define $D'_i = E'D_i$. Note that $D'_i\mathbf{n}'_i$ is the estimation error of the optimal MMSE estimator of the last r_S elements of the noise vector given the first $t - r_S$ elements of the vector $\mathbf{n}'_i$. We now write $\hat{N}_{i+1} - \hat{N}_i$ as follows:

$$\begin{aligned}\hat{N}_{i+1} - \hat{N}_i &= D'_i N'_i (D'_i)^T - D'_{i-1} N'_{i-1} (D'_{i-1})^T \\ &= D'_i N'_{i-1} (D'_i)^T - D'_{i-1} N'_{i-1} (D'_{i-1})^T + D'_i (N'_i - N'_{i-1})(D'_i)^T\end{aligned}$$

As $N'_1 \preceq \cdots \preceq N'_m$, the last summand in the last equality is semi-definite positive. Furthermore, $D'_{i-1} N'_{i-1} (D'_{i-1})^T$ is the covariance matrix of the estimation error of the optimal estimator of the last r_S elements of the noise vector $\mathbf{n}'_{i-1}$ given the first $t - r_S$ elements of the vector $\mathbf{n}'_{i-1}$ while $D'_i N'_{i-1} (D'_i)^T$ is a covariance matrix of an estimation error of a non-optimal estimator. Therefore, $D'_i N'_{i-1} (D'_i)^T - D'_{i-1} N'_{i-1} (D'_{i-1})^T \succeq 0$ and we conclude that if $N_1 \preceq \cdots \preceq N_m$, then $\hat{N}_{i+1} - \hat{N}_i \succeq 0$, $i = 1, \ldots, m-1$. ■

III. Contiguity of the capacity region with respect to S

Proposition III.1: The rate region $\mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ is contiguous with respect to S.

Proof: It is sufficient to show that if $N \succ 0$, then, for every $\epsilon > 0$, we can find a $\delta > 0$ such that $(\log|B + \delta' I + N| - \log|B + N|) < \epsilon$ for any given positive semi-definite matrix $B \succeq 0$ and every $0 \le \delta' \le \delta$. We write the following equality:

$$\log|B + \delta' I + N| - \log|B + N| = \log|I + \delta'(B + N)^{-1}| \le n \log(1 + \delta' \frac{1}{\psi}) \le n\delta' \frac{1}{\psi}$$

where n is the dimension of the matrices B and N, and ψ is the smallest eigenvalue of N.

The first inequality is due to the fact that $(B+N)^{-1} \preceq N^{-1}$ and the fact that $\frac{1}{\psi}$ is the largest eigenvalue of N^{-1}. The second inequality follows from $\log(1+x) \leq x$. As $N \succ 0$, $\psi > 0$. This completes our proof since for any $\epsilon > 0$, we can find a $\delta > 0$ such that $n\delta' \frac{1}{\psi} \leq \epsilon$ for all $0 \leq \delta' \leq \delta$. ■

Proposition III.2: The rate region $\mathcal{R}^{DPC}(S, N_{1\ldots m})$ is contiguous with respect to S.

Proof: The proof follows the exact same lines as in the proof of Proposition III.1 ■

Proposition III.3: The rate region $\mathcal{R}^{DPC}(S, N_{1\ldots m}, H_{1\ldots m})$ is contiguous with respect to S.

Proof: Just as in the proof of Proposition III.1, it is sufficient to show that if $N \succ 0$, then, for every $\epsilon > 0$, we can find a $\delta > 0$ such that $(\log|HBH^T + \delta' I + N| - \log|HBH^T + N|) < \epsilon$ for any given positive semi-definite matrix $B \succeq 0$ and every $0 \leq \delta' \leq \delta$. Replacing HBH^T with $B' \succeq 0$, we get the exact same expression as in the proof of Proposition III.1, and the proof follows immediately. ■

IV. Proof of Proposition 3.1

We first note that both statements will hold only if $\mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ is a closed set. However, as the space of sequences of matrices $B_1, \ldots, B_m$ such that $B_i \succeq 0$, $i = 1, \ldots, m$ and such that $\sum_{i=1}^m B_i \preceq S$ is compact and the functions $R_i^G(B_{1\ldots m}, \tilde{N}_{1\ldots m})$ are contiguous, $\mathcal{R}^G(S, \tilde{N}_{1\ldots m})$ is indeed a closed set.

Before turning to prove the proposition, we shall give two supporting propositions and corollaries. But first, for the sake of clarity, we shall restate the definition of the Gaussian region of an ADBC but with slight notational modifications. This allows us to extend some of the results in this appendix to the AMBC as well. For this reason, we replace the covariance matrices of the noise increments, $\tilde{N}_k$, with those of the accumulated noise at each of the receivers, N_k, such that $N_k = \sum_{i=1}^{k} \tilde{N}_i$.

For a given set of positive semi-definite matrices, $B_1, \dots, B_m$, the following rates are achievable:

$$R_k \leq R_k^G(B_{1\dots m}, N_{1\dots m}) = \frac{1}{2} \log \frac{|N_k + \sum_{i=1}^{k} B_i|}{|N_k + \sum_{i=1}^{k-1} B_i|}, \qquad k = 1, \dots, m$$

The Gaussian rate region is given by

$$\mathcal{R}^G(S, N_{1\dots m}) = \left\{ \begin{array}{c|} \Big(R_1^G(B_{1\dots m}, N_{1\dots m}), \dots, R_m^G(B_{1\dots m}, N_{1\dots m}) \Big) \\ \text{s.t.} \quad S - \sum_{i=1}^{m} B_i \succeq 0\,, \quad B_i \succeq 0 \;\; \forall i = 1, \dots, m \end{array} \right\} \tag{IV-1}$$

The following proposition suggests that we can increase the achievable rate of user $k+1$ at the expense of the user k.

Proposition IV.1: Assume that $(R_1, \dots, R_k, R_{k+1}, \dots, R_m) \in \mathcal{R}^G(S, N_{1\dots m})$ and that $R_k > 0$, then

1. if $1 \leq k < m$, then for every $0 < \delta \leq R_k$, there exists an $\epsilon > 0$ such that

$$(R_1, \dots, R_{k-1}, R_k - \delta, R_{k+1} + \epsilon, R_{k+2}, \dots, R_m) \in \mathcal{R}^G(S, N_{1\dots m})$$

2. if $k = m$, then for every $0 < \delta \leq R_m$,

$$(R_1, \dots, R_{m-1}, R_m - \delta) \in \mathcal{R}^G(S, N_{1\dots m})$$

Proof: If $(R_1, \dots, R_k, R_{k+1}, \dots, R_m) \in \mathcal{R}^G(S, N_{1\dots m})$, then there exists a set of positive semi-definite matrices $B_1, \dots, B_m$ such that $\sum_{i=1}^{m} B_i \preceq S$ and such that

$$R_i = R_i^G(B_{1\dots m}, N_{1\dots m}) \;\; \forall i \in 1, \dots, m$$

We define a new set of positive semi-definite matrices $B'_1, \dots, B'_m$ such that

1. If $1 \le k < m$, then

$$B'_i = \begin{cases} B_i & i \neq k, k+1 \\ \alpha B_k & i = k \\ B_{k+1} + (1-\alpha) B_k & i = k+1 \end{cases}$$

2. If $k = m$, then

$$B'_i = \begin{cases} B_i & 1 \le i < m \\ \alpha B_m & i = m \end{cases}$$

where $0 \le \alpha < 1$. Note that as $R_k > 0$, $B_k \neq 0$ (as otherwise, $R_k = R_k^G(B_{1\ldots m}, N_{1\ldots m}) = 0$).

Since the determinant function is contiguous, for every $0 < \delta \le R_k$ it is possible to choose ($\forall k = 1, \ldots, m$) an $0 \le \alpha < 1$ such that

$$R_k^G(B'_{1\ldots m}, N_{1\ldots m}) = R_k - \delta$$

and as $|B| > |A|$ if $B \succeq A \succ 0$ and $B \neq A$ (Proposition I.2), we can write the following expression for all k such that $1 \le k < m$,

$$\begin{aligned} R_{k+1}^G(B'_{1\ldots m}, N_{1\ldots m}) &= \frac{1}{2} \log \frac{|N_k + \sum_{i=1}^{k+1} B'_i|}{|N_k + \sum_{i=1}^{k} B'_i|} = \frac{1}{2} \log \frac{|N_k + \sum_{i=1}^{k+1} B_i|}{|N_k + \sum_{i=1}^{k-1} B_i + \alpha B_k|} \\ &= \frac{1}{2} \log \frac{|N_k + \sum_{i=1}^{k+1} B_i|}{|N_k + \sum_{i=1}^{k} B_i|} + \epsilon = R_{k+1}^G(B_{1\ldots m}, N_{1\ldots m}) + \epsilon = R_{k+1} + \epsilon \end{aligned}$$

for some $\epsilon > 0$. However, since $\sum_{j=1}^{i} B_j = \sum_{j=1}^{i} B'_j \quad \forall i \neq k$, then

$$R_i^G(B'_{1\ldots m}, N_{1\ldots m}) = R_i^G(B_{1\ldots m}, N_{1\ldots m}) = R_i \qquad \forall i \neq k, k+1$$

Thus, we have found a set of positive semi-definite matrices $B'_1, \ldots, B'_m$ such that $\sum_{j=1}^{m} B'_j \preceq S$ and which achieve the required rates and therefore, these rates must lie in $\mathcal{R}^G(S, N_{1\ldots m})$. ■

An interesting result from the above proposition is that if a rate vector $\overline{R}$ lies inside the Gaussian rate region, then all rate vectors which are element-wise smaller than $\overline{R}$ also lie inside the Gaussian rate region. This is formalized in the following corollary.

Corollary IV.1: If $(R_1, \dots, R_m) \in \mathcal{R}^G(S, N_{1 \dots m})$, then every rate vector $(R'_1, \dots, R'_m)$ such that $0 \le R'_i \le R_i \quad \forall i = 1, \dots, m$ also belongs to the set $\mathcal{R}^G(S, N_{1 \dots m})$.

Proof: This is a simple result of recursively applying Proposition IV.1. ■

While Proposition IV.1 proved that we can increase the achievable rate of one user at the expense of the preceding user, the following proposition shows that we could do the same in the opposite direction.

Proposition IV.2: Assume that $(R_1, \dots, R_{k-1}, R_k, \dots, R_m) \in \mathcal{R}^G(S, N_{1 \dots m})$ and that $R_k > 0$, then

1. if $1 < k \le m$, then for every $0 < \delta \le R_k$, there exists an $\epsilon > 0$ such that

$$(R_1, \dots, R_{k-2}, R_{k-1} + \epsilon, R_k - \delta, R_{k+1}, \dots, R_m) \in \mathcal{R}^G(S, N_{1 \dots m})$$

2. if $k = 1$, then for every $0 < \delta \le R_1$,

$$(R_1 - \delta, R_2, \dots, R_m) \in \mathcal{R}^G(S, N_{1 \dots m})$$

The proof of the above proposition is very similar to that of Proposition IV.1 and therefore, is omitted here.

An important result from the previous two propositions is given in the following corollary.

Corollary IV.2: If $(R_1, \dots, R_m) \in \mathcal{R}^G(S, N_{1 \dots m})$ and $(R'_1, \dots, R'_m) \in \mathcal{R}^G(S, N_{1 \dots m})$ such that $R'_i \ge R_i \quad \forall i = 1, \dots, m$ and such that for some $k \in \{1, \dots, m\}$, $R'_k > R_k$, then there exists a rate vector $(R''_1, \dots, R''_m) \in \mathcal{R}^G(S, N_{1 \dots m})$ such that $R''_i > R_i \quad \forall i = 1, \dots, m$.

Proof: Because $R'_k > R_k$, it is possible to recursively use Proposition IV.1 to show that there is an $\epsilon > 0$ for which

$$(R'_1, \dots, R'_{k-1}, R_k, R'_{k+1}, \dots, R'_m + \epsilon) \in \mathcal{R}^G(S, N_{1 \dots m})$$

We can now recursively use Proposition IV.2 to show that there are $\epsilon_i > 0$, $i = 1, \dots, m$ such that

$$(R'_1+\epsilon_1,\ldots,R'_{k-1}+\epsilon_{k-1},R_k+\epsilon_k,R'_{k+1}+\epsilon_{k+1},\ldots,R'_m+\epsilon_m)\in\mathcal{R}^G(S,N_{1\ldots m})$$

■

Proof of Proposition 3.1

Proof: We begin with the first statement of the proposition. Since every ϵ-ball around any optimal Gaussian rate vector $(R_1,\ldots,R_m)$ will always contain a point $(R'_1,\ldots,R'_m)\notin\mathcal{R}^G(S,N_{1\ldots m})$ (e.g. $R'_i>R_i\ \forall i$), it is clear that all optimal Gaussian rate vectors are boundary points. Therefore, we only need to show that all boundary points are also optimal Gaussian rate vectors. Assume, in contrast, that $(R_1,\ldots,R_m)\in\mathcal{R}^G(S,N_{1\ldots m})$ is not an optimal Gaussian rate vector but is still a boundary point. Then, there exists a rate vector $(R'_1,\ldots,R'_m)\in\mathcal{R}^G(S,N_{1\ldots m})$ such that $R'_i\geq R_i\quad\forall i=1,\ldots,m$ and such that for some $k\in\{1,\ldots,m\}$, $R'_k>R_k$. By Corollary IV.2 we know that there exists a rate vector $(R''_1,\ldots,R''_m)\in\mathcal{R}^G(S,N_{1\ldots m})$ such that $R''_i>R_i\quad\forall i=1,\ldots,m$ and from Corollary IV.1 it is clear that $(R_1,\ldots,R_m)$ must be an inner point of $\mathcal{R}^G(S,N_{1\ldots m})$ and hence not a boundary point.

We proceed to prove the second statement of Proposition 3.1. We use induction on the number of receiving users, m. The case of $\underline{m=1}$ is a single user channel with a noise covariance matrix, N_1. If, in the single user channel, $(R_1)\notin\mathcal{R}^G(S,N_1)$, then $R_1>C$ (channel capacity) and therefore, there is a scalar $b>0$ such that $R_1=C+b$.

Next, we assume that the statement is correct for $k<m$ users where $\underline{m>1}$ and prove that it must be true for $k=m$ users. That is, we assume that for all ADBCs with $k<m$ users and with noise covariance matrices, $N_1,\ldots,N_k$, that if there is a rate vector $(R_1,R_2,\ldots,R_k)$

satisfying:

$$\begin{aligned} &(R_1, R_2, \ldots, R_k) \notin \mathcal{R}^G(S, N_{1\ldots k}) \\ &R_j \geq 0, \, \forall j = 1, \ldots, k-1 \\ &R_k > 0, \end{aligned} \tag{IV-2}$$

then there is a strictly positive scalar, $b > 0$, and an optimal Gaussian solution point $B_1^*, \ldots, B_k^*$ such that

$$\begin{aligned} R_j &\geq R_j^G(B_{1\ldots k}^*, N_{1\ldots k}), \quad j = 1, \ldots, k-1 \\ R_k &\geq R_i^G(B_{1\ldots k}^*, N_{1\ldots k}) + b \end{aligned} \tag{IV-3}$$

and want to prove that for the m-users ADBC with noise covariance matrices, $N_1, \ldots, N_m$, that if there is a vector $(R_1, R_2, \ldots, R_m)$ satisfying the conditions in (IV-2) for $k = m$, then (IV-3) will hold for $k = m$.

We consider the $m - 1$ user channel and distinguish between two cases regarding the rate vector $(R_1, \ldots, R_{m-1})$: Either $(R_1, \ldots, R_{m-1}) \notin \mathcal{R}^G(S, N_{1\ldots m-1})$ or $(R_1, \ldots, R_{m-1}) \in \mathcal{R}^G(S, N_{1\ldots m-1})$. We start with the first case. Then, necessarily, $(R_1, \ldots, R_{m-1}) \neq 0$. Let k be the index of the last nonzero entry in $(R_1, \ldots, R_{m-1})$. Then, we must have $(R_1, \ldots, R_k) \notin \mathcal{R}^G(S, N_{1\ldots k})$, $R_k > 0$, as otherwise

$$(R_1, \ldots, R_k, R_{k+1}, \ldots, R_{m-1}) = (R_1, \ldots, R_k, 0, \ldots, 0) \in \mathcal{R}^G(S, N_{1\ldots m-1}),$$

contradicting our assumption that $(R_1, \ldots, R_{m-1}) \notin \mathcal{R}^G(S, N_{1\ldots m-1})$. Now, applying our induction assumption, we can find an optimal Gaussian rate vector for the k-user BC, $(R_1^*, \ldots, R_k^*)$, such that

$$\begin{aligned} &R_i \geq R_i^* \quad i = 1, 2, \ldots, k-1 \\ &R_k \geq R_k^* + b \;\text{ for some } b > 0. \end{aligned}$$

However, $(R_1^*, \ldots, R_k^*, 0, \ldots, 0)$ is also an optimal Gaussian rate vector of the m-user ADBC. To see this, observe that if not, there must exist some non-negative vector $(\delta_1, \ldots, \delta_m)$ such that at least one of the entries is strictly positive and such that

$$(R_1^* + \delta_1, \ldots, R_k^* + \delta_k, \delta_{k+1}, \ldots, \delta_m) \in \mathcal{R}^G(S, N_{1\ldots m})$$

and now, applying Proposition IV.2 we get

$$(R_1^* + \epsilon, \ldots, R_k^*, 0, \ldots, 0) \in \mathcal{R}^G(S, N_{1\ldots m})$$

for some $\epsilon > 0$, which implies

$$(R_1^* + \epsilon, \ldots, R_k^*) \in \mathcal{R}^G(S, N_{1\ldots k})$$

contradicting the assumption that $(R_1^*, \ldots, R_k^*)$ is an optimal Gaussian rate vector for the k-users ADBC. Now, since $R_m > 0$, and $R_m^* = 0$, the second statement of the proposition follows for the case $(R_1, \ldots, R_{m-1}) \notin \mathcal{R}^G(S, N_{1\ldots m-1})$.

Next, we consider the second case where $(R_1, \ldots, R_{m-1}) \in \mathcal{R}^G(S, N_{1\ldots m-1})$. Define $\overline{R}^* = (R_1^*, \ldots, R_{m-1}^*, R_m^*)$ such that $R_i^* = R_i$, $\forall i = 1, \ldots, m-1$ and

$$R_m^* = \sup\{R_m | (R_1^*, \ldots, R_{m-1}^*, R_m) \in \mathcal{R}^G(S, N_{1\ldots m})\}.$$

As $\mathcal{R}^G(S, N_{1\ldots m})$ is a closed set, $\overline{R}^* \in \mathcal{R}^G(S, N_{1\ldots m})$. Furthermore, $\overline{R}^*$ is an optimal Gaussian rate vector, as otherwise, using Proposition IV.1 we could write $(R_1^*, \ldots, R_{m-1}^*, R_m^* + \epsilon) \in \mathcal{R}^G(S, N_{1\ldots m})\}$, contradicting the definition of R_m^*. However, as we know that all rate vectors that are element-wise smaller then $\overline{R}^*$ lie in the Gaussian region (Corollary IV.1) and as

$$(R_1, \ldots, R_m) = (R_1^*, \ldots, R_{m-1}^*, R_m) \notin \mathcal{R}^G(S, N_{1\ldots m}),$$

there must exist a $b > 0$ such that $R_m = b + R_m^*$. ■

V. Extension of Propositions IV.1 and IV.2 to the AMBC

As our proofs of Propositions IV.1 and IV.2 do not depend on the degradedness of the Noise covariance matrices, N_i, we can automatically extend the intermediate results of the previous section to the AMBC. The following proposition and corollary formalize this extension.

Proposition V.1: Let π be an m-user permutation function. If $(R_1, \ldots, R_m) \in \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$, and $R_{\pi(k)} > 0$ for some $k \in 1 \ldots m$, then, for every $\delta \in (0, R_{\pi(k)}]$, there exists an $\epsilon > 0$, such that $(R_1^*, \ldots, R_m^*) \in \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$ and $(R_1^{**}, \ldots, R_m^{**}) \in \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$ where

$$R^*_{\pi(i)} = \begin{cases} R_{\pi(i)} - \delta & i = k \\ R_{\pi(i)} + \epsilon & i = k+1 \\ R_{\pi(i)} & \text{otherwise} \end{cases} \qquad R^{**}_{\pi(i)} = \begin{cases} R_{\pi(i)} - \delta & i = k \\ R_{\pi(i)} + \epsilon & i = k-1 \\ R_{\pi(i)} & \text{otherwise} \end{cases}$$

Proof: The proof is identical to the proofs of Propositions IV.1 and IV.2 where the functions $R_i^G(B_{1\ldots m}, N_{1\ldots m})$ are replaced by $R_i^{DPC}(\pi, B_{1\ldots m}, N_{1\ldots m})$. ■

Corollary V.1: Define the rate vector $\overline{R} = (R_1, \ldots, R_m)$ and let $\overline{R}' = (R_1', \ldots, R_m')$ be any rate vector such that $0 \le R_i' \le R_i \; \forall i = 1, \ldots, m$. Then,

1. If $\overline{R} \in \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$, then $\overline{R}' \in \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$.
2. If $\overline{R} \in \mathcal{R}^{DPC}(S, N_{1\ldots m})$, then $\overline{R}' \in \mathcal{R}^{DPC}(S, N_{1\ldots m})$.

Proof: As in the case of corollary IV.1, the proof of the first statement is a simple result of a recursive application of Proposition V.1. To prove the second statement, we note that as $\mathcal{R}^{DPC}(S, N_{1\ldots m})$ is convex, it is sufficient to show that for all $k = 1 \ldots m$, $(R_1, \ldots, R_{k-1}, 0, R_{k+1}, \ldots, R_m) \in \mathcal{R}^{DPC}(S, N_{1\ldots m})$. Moreover, as every point in $\mathcal{R}^{DPC}(S, N_{1\ldots m})$ is a convex combination of points in $\mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$, we need only to show that if

$(R_1^\pi, \ldots, R_m^\pi) \in \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$, then, $(R_1^\pi, \ldots, R_{k-1}^\pi, 0, R_{k+1}^\pi, \ldots, R_m^\pi) \in \mathcal{R}^{DPC}(\pi, S, N_{1\ldots m})$. However, this is a trivial case of the first statement. ∎

VI. Proof of Proposition 3.3

This section relies heavily on the concept of a feasible direction which appears in [2] and on dual functions and generalized inequalities as they appear in Chapter I, Subsections 2.4 and 5.9 of [4]. The reader is referred to these references for more insight. The proof of Proposition 3.3 is based on a sequence of definitions and Propositions.

Definition VI.1 (Feasible Direction (chapter 4, [2])) Given a subset $\mathcal{X} \subseteq \mathbb{R}^n$ and a vector $\mathbf{x} \in \mathcal{X}$, a vector $\mathbf{y} \in \mathbb{R}^n$ is said to be a feasible direction of $\mathcal{X}$ at $\mathbf{x}$ if there exists an $\alpha > 0$ such that $\mathbf{x} + \alpha'\mathbf{y} \in \mathcal{X}$, $\forall \alpha' \in [0, \alpha]$.

Proposition VI.1: Consider the set of positive semi-definite real matrices of size $t \times t$, $\mathcal{B} = \{B' \in \mathbb{R}^{t\times t} \mid B' \succeq 0\}$. For a matrix $B \in \mathcal{B}$, we define r_B as the number of zero eigenvalues ($r_B = t - \text{rank}(B)$) and let $\mathbf{v}_i^B$, $i = 1 \ldots t$ denote orthogonal unit-length eigenvectors of B, ordered so that the eigenvectors $\mathbf{v}_i^B$ $i \in 1 \ldots r_B$ have null eigenvalues. Furthermore, let V_B be a $t \times r_B$ matrix such that $V_B = [\mathbf{v}_1^B \ldots \mathbf{v}_{r_B}^B]$ and let L be a $t \times t$ symmetric matrix. Then, if $r_B = 0$, then any symmetric matrix L will be a feasible direction. Otherwise, L is a feasible direction if

$$V_B^T L V_B \succ 0$$

Note that the proposition only states a sufficient condition.

Proof: We need to show that there exists a $\delta > 0$ such that $B + \delta' L \succeq 0$, $\forall \delta' \in [0, \delta]$. In other words, we need to show that for every arbitrary vector, $\mathbf{v}$, of size $t \times 1$,

$$\mathbf{v}^T (B + \delta' L)\mathbf{v} \geq 0, \quad \forall \delta' \in [0, \delta]$$

When $B = 0_{t\times t}$, V_B is a unitary matrix and the condition, $V_B^T L V_B \succ 0$, implies that $L \succ 0$. Therefore, when $B = 0_{t\times t}$, clearly, the above expression holds. Hence, we will consider the case where $B \neq 0$.

Assume that $B \neq 0$ and that $r_B > 0$. Any arbitrary vector, $\mathbf{v}$, can be represented as an arbitrary linear combination of $\mathbf{v}_i^B$, $i = 1 \ldots t$ such that

$$\mathbf{v} = \sum_{i=1}^{t} \beta_i \mathbf{v}_i^B = \sum_{i=1}^{r_B} \beta_i \mathbf{v}_i^B + \sum_{i=r_B+1}^{t} \beta_i \mathbf{v}_i^B = \mathbf{v}_\perp + \mathbf{v}_{||}$$

where $\mathbf{v}_\perp = \sum_{i=1}^{r_B} \beta_i \mathbf{v}_i^B$ and $\mathbf{v}_{||} = \sum_{i=r_B+1}^{t} \beta_i \mathbf{v}_i^B$. By taking into account that $B\mathbf{v}_\perp = 0$, we can write

$$\mathbf{v}^T(B + \delta' L)\mathbf{v} = \mathbf{v}_{||}^T B \mathbf{v}_{||} + \delta' \mathbf{v}_{||}^T L \mathbf{v}_{||} + 2 \cdot \delta' \mathbf{v}_{||}^T L \mathbf{v}_\perp + \delta' \mathbf{v}_\perp^T L \mathbf{v}_\perp \qquad \text{(VI-1)}$$

Let m_B denote the smallest, non-zero eigenvalue of B and m_L denote the smallest eigenvalue of $V_B^T L V_B$. As $B \neq 0$ and as $V_B^T L V_B \succ 0$, both m_B and m_L are strictly positive. Furthermore, let M_L denote the largest absolute value of an eigenvalue of L. Note that as $L \neq 0$, $M_L > 0$.

Since $\mathbf{v}_\perp = V_B \overline{\beta}_\perp$ where $\overline{\beta}_\perp = (\beta_1, \ldots, \beta_{r_B})^T$, we can write the following lower bound on the last summand of (VI-1):

$$\mathbf{v}_\perp^T L \mathbf{v}_\perp = \overline{\beta}_\perp^T (V_B^T L V_B) \overline{\beta}_\perp \geq m_L |\overline{\beta}_\perp|^2 = m_L |\mathbf{v}_\perp|^2,$$

where the inequality is due to m_L being the smallest eigenvalue of $V_B^T L V_B$ and the last equality is due to the fact that $\mathbf{v}_i^B$, $i = 1 \ldots t$ are orthonormal and hence, $|v_\perp| = |\overline{\beta}_\perp|$. Furthermore, we can lower bound the third summand of (VI-1) as follows:

$$2 \cdot \delta' \mathbf{v}_{||}^T L \mathbf{v}_\perp \geq -2 \cdot \delta' |\mathbf{v}_{||}^T L \mathbf{v}_\perp| \geq -2 \cdot \delta' |\mathbf{v}_{||}^T| \cdot |L \mathbf{v}_\perp| \geq -2 \cdot \delta' |\mathbf{v}_{||}^T| \cdot M_L |\mathbf{v}_\perp|$$

where the second inequality is due to the Cauchy-Schwartz inequality. Therefore, we can now write,

$$\begin{aligned}\mathbf{v}^T(B+\delta' L)\mathbf{v} &\geq m_B|\mathbf{v}_{||}|^2 - \delta' M_L|\mathbf{v}_{||}|^2 - 2\cdot\delta' M_L|\mathbf{v}_{||}|\cdot|\mathbf{v}_\perp| + \delta' m_L|\mathbf{v}_\perp|^2 \\ &= \left(m_B - \delta'\left(M_L + \frac{M_L^2}{m_L}\right)\right)|\mathbf{v}_{||}|^2 + \delta'\left(\frac{M_L}{\sqrt{m_L}}|\mathbf{v}_{||}| - \sqrt{m_L}|\mathbf{v}_\perp|\right)^2\end{aligned}$$

Clearly, for any $0 \leq \delta' \leq \frac{m_B}{\left(M_L + \frac{M_L^2}{m_L}\right)}$ the above expression will be non-negative (regardless of $\mathbf{v}$). We now complete the proof and consider the case where $r_B = 0$. As, B has no zero eigenvalues, m_B is the smallest eigenvalue of B and is strictly positive. Therefore, we may write,

$$\mathbf{v}^T(B+\delta' L)\mathbf{v} \geq m_B|\mathbf{v}|^2 - \delta' M_L|\mathbf{v}|^2.$$

Clearly, for any $0 \leq \delta' \leq \frac{m_B}{M_L}$ the above expression will be non-negative. ■

We can now state an extension of Proposition VI.1, relevant to the proof of Proposition 3.3.

Corollary VI.1: The set of matrices $L_1, \ldots, L_{m-1}$ is a feasible direction of

$$\mathcal{D} = \left\{ (B_1, \ldots, B_{m-1}) \big| \; B_1 \succeq 0, \ldots, B_{m-1} \succeq 0, \; \sum_{j=1}^{m-1} B_j \preceq S \right\}$$

at $B_1, \ldots, B_{m-1}$ if the following conditions hold:

1. If $r_{B_i} > 0$, then $V_{B_i}^T L_i V_{B_i} \succ 0$, where $i = 1, \ldots, m-1$.
2. If $r_{B_m} > 0$, then $V_{B_m}^T \left(\sum_{j=1}^{m-1} L_j\right) V_{B_m} \prec 0$

where $B_m = S - \sum_{i=1}^{m-1} B_i$ and where r_{B_i} and V_{B_i} are defined in a similar manner to r_B and V_B in Proposition VI.1 with respect to the matrices B_i, $i = 1, \ldots, m$.

Proof: This is a trivial result of Proposition VI.1. ■

Note that when $r_{B_i} = 0$, we do not have any (individual) constraint on L_i. In addition, note that the above two conditions are identical to the first two conditions in the definition of a *strict direction* (definition 3.4). We now turn to prove Proposition 3.3.

Proof: As $B_1^*, \ldots, B_{m-1}^*$ is an optimal solution, there can not be a feasible direction of $\mathcal{D}$ at $B_1^*, \ldots, B_{m-1}^*$ such that the directional derivatives,

$$\dot{f}_i(B_{1\ldots m-1}^*, L_{1\ldots m-1}) = \operatorname{tr}\{\nabla f_i(B_1^*, \ldots, B_{m-1}^*) \cdot (L_1, \ldots, L_{m-1})^T\}$$

are strictly positive for all $i = 0, \ldots, k$, as this would suggest that $B_1^*, \ldots, B_{m-1}^*$ was not optimal to begin with.

By corollary VI.1, we observe that the first two conditions in the definition of a strict direction are sufficient conditions for a feasible direction in $\mathcal{D}$. Furthermore, the third (and last) condition in the same definition states that the directional derivatives of f_i, $i = 1, \ldots, k$ are all strictly positive. Therefore, we conclude that if $L_1, \ldots, L_{m-1}$ is a strict direction, the directional derivative (in the same direction) of f_0 can not be strictly positive.

To summarize, define $\mathcal{G} = \{(L_1, \ldots, L_{m-1}) |\, (L_1, \ldots, L_{m-1}) \text{ is a strict direction}\}$. Then,

$$\operatorname{tr}\{\nabla f_0(B_1^*, \ldots, B_{m-1}^*) \cdot (L_1, \ldots, L_{m-1})^T\} \leq 0, \quad \forall (L_1, \ldots, L_{m-1}) \in \mathcal{G}$$

Next, we show that we can use the fact that we know of some *strict direction*, $L_1^*, \ldots, L_m^*$, to extend the range of directions over which the directional derivative of f_0 is non-positive.

Let $L_1, \ldots, L_m$ be any point on the closure of $\mathcal{G}$. It is easily shown that if $L_1^*, \ldots, L_m^*$ is a strict direction, so is $L_1 + \alpha L_1^*, \ldots, L_m + \alpha L_m^*$, for every $\alpha > 0$. Therefore,

$$\operatorname{tr}\{\nabla f_0(B_1^*, \ldots, B_{m-1}^*) \cdot (L_1 + \alpha L_1^*, \ldots, L_{m-1} + \alpha L_m^*)^T\} \leq 0, \quad \forall \alpha > 0.$$

However, as the above expression is linear and contiguous with α, we have

$$\begin{aligned} &\operatorname{tr}\{\nabla f_0(B_1^*, \ldots, B_{m-1}^*) \cdot (L_1, \ldots, L_{m-1})^T\} \\ &\qquad = \lim_{\alpha \downarrow 0} \operatorname{tr}\{\nabla f_0(B_1^*, \ldots, B_{m-1}^*) \cdot (L_1 + \alpha L_1^*, \ldots, L_{m-1} + \alpha L_m^*)^T\} \leq 0 \end{aligned}$$

Thus, we can write,

$$\operatorname{tr}\{\nabla f_0(B_1^*, \ldots, B_{m-1}^*) \cdot (L_1, \ldots, L_{m-1})^T\} \leq 0, \quad \forall (L_1, \ldots, L_{m-1}) \in \operatorname{cl}(\mathcal{G}) \qquad \text{(VI-2)}$$

We now use (VI-2) to write a bounded, convex optimization problem with conic inequalities [4] and then use the dual problem to derive the necessary conditions on the gradients as they appear in the proposition.

We begin with the primal problem:

$$
\begin{array}{lll}
\underset{L_1,\dots,L_{m-1}}{\text{minimize}} & -\mathrm{tr}\{\nabla f_0(B_1^*,\dots,B_{m-1}^*)\cdot(L_1,\dots,L_{m-1})^T\} & \\
\text{such that} & -\mathrm{tr}\{\nabla f_i(B_1^*,\dots,B_{m-1}^*)\cdot(L_1,\dots,L_{m-1})^T\}\le 0, & i=1,\dots,k \\
 & -V_{B_i}^T L_i V_{B_i}\preceq 0, & \forall i\in\{i=1,\dots,m-1 \mid r_{B_i^*}>0\} \\
 & V_{B_m}^T(\sum_{j=1}^{m-1}L_j)V_{B_m}\preceq 0 & \text{if } r_{B_m^*}>0
\end{array}
\tag{VI-3}
$$

One can easily verify that the feasible region defined by the inequalities in (VI-3), forms the closure of $\mathcal{G}$. By (VI-2), it is clear that the solution of (VI-3) is bounded by 0. In fact, (VI-3) is actually a trivial problem and the optimum is achieved by $L_i=0$, $i=1,\dots,m-1$. However, we are not interested in the solution of the problem but rather in the conditions on the gradients of the functions, f_i, which are implied by the existence of this solution.

We now turn to write the dual problem. We begin with the Lagrangian:

$$
\begin{aligned}
\mathcal{L}=&-\mathrm{tr}\{\nabla f_0(B_1^*,\dots,B_{m-1}^*)\cdot(L_1,\dots,L_{m-1})^T\}\\
&-\sum_{i=1}^{k}\gamma_i\cdot\mathrm{tr}\{\nabla f_i(B_1^*,\dots,B_{m-1}^*)\cdot(L_1,\dots,L_{m-1})^T\}\\
&-\sum_{i=1}^{m-1}\mathrm{tr}\{\Lambda_i V_{B_i}^T L_i V_{B_i}\}+\mathrm{tr}\left\{\Lambda_m V_{B_m}^T\left(\sum_{i=1}^{m-1}L_i\right)V_{B_m}\right\}
\end{aligned}
$$

where, for the sake of brevity, we assigned $V_{B_i}=0_{t\times t}$ if $r_{B_i^*}=0$. In addition, $\gamma_i\ge 0$, $i=1,\dots,k$ are scalars and $\Lambda_i\succeq 0$, $i=1,\dots,m$ are matrices of size $r_{B_i^*}\times r_{B_i^*}$ if $r_{B_i^*}>0$ and $t\times t$ otherwise. Note that Λ_i play a similar role to the scalar Lagrange multipliers in a constrained program with normal inequalities. However, here these scalar multipliers are replaced by matrices in-order to accommodate conic inequalities.

We can rewrite the Lagrangian as follows:

$$\mathcal{L} = -\sum_{i=1}^{m-1} \operatorname{tr}\left\{ \left(\begin{array}{r} \nabla_{B_i} f_0(B_1^*, \ldots, B_{m-1}^*) + \sum_{j=1}^{k} \gamma_j \cdot \nabla_{B_i} f_j(B_1^*, \ldots, B_{m-1}^*) \\ + V_{B_i} \Lambda_i V_{B_i}^T - V_{B_m} \Lambda_m V_{B_m}^T \end{array} \right) L_i^T \right\}$$

The dual function, $g(\gamma_{1\ldots k}, \Lambda_{1\ldots m})$, is given by

$$g(\gamma_{1\ldots k}, \Lambda_{1\ldots m}) = \min_{L_1,\ldots,L_m} \mathcal{L}$$

$$= \begin{cases} 0 & \text{if} \quad \left(\begin{array}{r} \nabla_{B_i} f_0(B_1^*, \ldots, B_{m-1}^*) + \sum_{j=1}^{k} \gamma_j \cdot \nabla_{B_i} f_j(B_1^*, \ldots, B_{m-1}^*) \\ + V_{B_i} \Lambda_i V_{B_i}^T - V_{B_m} \Lambda_m V_{B_m}^T \end{array} \right) = 0, \quad \forall i = 1 \ldots m-1 \\ -\infty & \text{otherwise} \end{cases} \tag{VI-4}$$

The dual problem takes the following form:

$$\begin{array}{lll} \text{maximize} & g(\gamma_{1\ldots k}, \Lambda_{1\ldots m}) & \\ \gamma_{1\ldots k}, \Lambda_{1\ldots m} & & \\ \text{such that} & \gamma_i \geq 0, & i = 1, \ldots, k \\ & \Lambda_i \succeq 0, & i = 1, \ldots, m \end{array} \tag{VI-5}$$

We use p^* and d^* to denote the optimal values of the primal problem (VI-3) and the dual problem (VI-5) respectively. In general, $d^* \leq p^*$. However, as the primal problem is convex, and since the existence of a *strict direction* implies the existence of a point $\in$ relint $\text{cl}(\mathcal{G})$, Slater's conditions for strong duality are met. Therefore, $d^* = p^* = 0$. By (VI-4), one can verify that strong duality implies that there must be scalars $\gamma_i \geq 0$, $i = 1, \ldots, k$ and matrices $\Lambda_i \succeq 0$, $i = 1, \ldots, m$ such that

$$\nabla_{B_i} f_0(B_1^*, \ldots, B_{m-1}^*) + \sum_{j=1}^{k} \gamma_j \cdot \nabla_{B_i} f_j(B_1^*, \ldots, B_{m-1}^*) + V_{B_i} \Lambda_i V_{B_i}^T - V_{B_m} \Lambda_m V_{B_m}^T = 0,$$

$$\forall i = 1, \ldots, m-1$$

■

VII. The existence of a *strict direction*

Proposition VII.1: Let $B_1^*, \ldots, B_{m-1}^*$ be an optimal solution of the problem in equation (3-20). If $S \succ 0$ (strictly positive) and $B_m^* = S - \sum_{i=1}^{m-1} B_i^* \neq 0$, there exists a *strict direction* at $B_1^*, \ldots, B_{m-1}^*$.

Proof: Let $r_{B_k^*}$, $k = 1, \ldots, m$ denote the number of null eigenvalues in B_k^* (where $B_m^* = S - \sum_{i=1}^{m-1} B_i^*$) and $\mathbf{v}_i^{B_k^*}$, $i = 1, \ldots, r_{B_k^*}$ denote orthogonal unit-length eigenvectors of B_k^* which correspond to zero eigenvalues. Let $\mathbf{z}_k = \sum_{i=1}^{r_{B_k^*}} \beta_i \mathbf{v}_i^{B_k^*}$ for some choice of scalars β_i. We need to show that there exists a set of matrices $L_1, \ldots, L_m$ such that

1. $\mathbf{z}_k^T L_k \mathbf{z}_k > 0$, for all $k \in \{j = 1, \ldots, m-1 \mid r_{B_j^*} > 0\}$ and for every choice of scalars, β_i, such that $\mathbf{z}_k \neq 0$.
2. If $r_{B_m^*} > 0$, then $\mathbf{z}_m^T \left(\sum_{j=1}^{m-1} L_j \right) \mathbf{z}_m^T < 0$ for every choice of scalars, β_i, such that $\mathbf{z}_m \neq 0$.
3. $\mathrm{tr}\{\nabla R_i^G(B_{1\ldots m-1}^*, S, \tilde{N}_{1\ldots m-1}) \cdot (L_1, \ldots, L_{m-1})^T\} > 0, \quad \forall i = 1, \ldots, m-1$.

Note that the above conditions are equivalent to those in definition 3.4.

For that purpose, we suggest a direction, $L_1^*, \ldots, L_{m-1}^*$ of the form

$$L_i^* = -B_i^* + \frac{1}{m} \sum_{j \neq i \,,\, j < m} B_j^* + \alpha_i B_m^* \,, \quad i = 1, \ldots, m-1 \tag{VII-1}$$

for some $\alpha_i > 0$, $i = 1, \ldots, m-1$. We begin by checking the first condition. As $\mathbf{z}_k$ is a linear combination of orthogonal eigenvectors corresponding to null eigenvalues of B_k^*, we can write:

$$\mathbf{z}_k^T L_k^* \mathbf{z}_k = \frac{1}{m} \sum_{j \neq i \,,\, j < m} \mathbf{z}_k^T B_j^* \mathbf{z}_k + \alpha_i \mathbf{z}_k^T B_m^* \mathbf{z}_k$$

As $B_i^* \succeq 0$, $i = 1, \ldots, m$, it is clear from the above equation that $\mathbf{z}_k^T L_k^* \mathbf{z}_k \geq 0$. Assume that for some $\mathbf{z}_k \neq 0$ we get $\mathbf{z}_k^T L_k^* \mathbf{z}_k = 0$. This can only occur if $\mathbf{z}_k^T B_j^* \mathbf{z}_k = 0, \quad \forall j \neq k$ which implies that $\mathbf{z}_k^T (\sum_{j=1}^m B_j^*) \mathbf{z}_k = \mathbf{z}_k^T S \mathbf{z}_k = 0$. However, we assume that $S \succ 0$, and therefore, for any $\mathbf{z}_k \neq 0$, $\mathbf{z}_k^T S \mathbf{z}_k > 0$. Thus we have proved that the first condition holds for any choice of

$\alpha_i > 0$.

To prove that the second condition holds for our choice of direction, $L_1^*, \ldots, L_{m-1}^*$, and when $r_{B_m} > 0$, we write:

$$\mathbf{z}_m^T \left(\sum_{i=1}^{m-1} L_i^* \right) \mathbf{z}_m = \mathbf{z}_m^T \left(-\frac{2}{m}(S - B_m^*) + \sum_{i=1}^{m-1} \alpha_i B_m^* \right) \mathbf{z}_m = -\frac{2}{m} \mathbf{z}_m^T S \mathbf{z}_m < 0$$

for every choice of β_i such that $\mathbf{z}_m \neq 0$.

Note that to show that the first two requirements hold, we did not use the assumption that $B_m^* \neq 0$. That is, we have shown that if $S \succ 0$, we can find a direction for which the first two conditions hold. This fact is of importance in the proof of the following proposition (VII.2).

Next, we need to show that we can choose the scalars, $\alpha_i > 0$, such that the third condition is met. Using the explicit expression for the rate gradients (3-21), we can rewrite the directional gradient of the i'th rate as follows:

$$\begin{aligned}
&\mathrm{tr}\{\nabla R_i^G(B_{1\ldots m-1}^*, S, \tilde{N}_{1\ldots m-1}) \cdot (L_1^*, \ldots, L_{m-1}^*)^T\} \\
&\quad = \sum_{j=1}^{m-1} \mathrm{tr}\{\nabla_{B_j} R_i^G(B_{1\ldots m-1}^*, S, \tilde{N}_{1\ldots m-1}) \cdot L_j^*\} \\
&\quad = \sum_{j=1}^{i-1} \mathrm{tr}\{\nabla_{B_j} R_i^G(B_{1\ldots m-1}^*, S, \tilde{N}_{1\ldots m-1}) \cdot L_j^*\} + \mathrm{tr}\left\{ \frac{1}{2} \left(\sum_{l=1}^{i} (B_l + \tilde{N}_l) \right)^{-1} L_i^* \right\} \\
&\quad = \sum_{j=1}^{i-1} \mathrm{tr}\{\nabla_{B_j} R_i^G(B_{1\ldots m-1}^*, S, \tilde{N}_{1\ldots m-1}) \cdot L_j^*\} \\
&\qquad + \mathrm{tr}\left\{ \frac{1}{2} \left(\sum_{l=1}^{i} (B_l + \tilde{N}_l) \right)^{-1} \left(-B_i + \frac{1}{m} \sum_{j \neq i\,,\, j<m} B_j \right) \right\} \\
&\qquad + \alpha_i \cdot \mathrm{tr}\left\{ \frac{1}{2} \left(\sum_{l=1}^{i} (B_l + \tilde{N}_l) \right)^{-1} B_m \right\}
\end{aligned}$$

In general, we observe that the directional gradient of the i'th rate depends only on $L_1, \ldots, L_i$, or alternatively, on $\alpha_1, \ldots, \alpha_i$. Therefore, we will set the values of α_i recursively such that α_i will be a function of α_j, $j = 1, \ldots, i-1$. We observe that for $i = 1$, the first summand in the last equality in the above expression does not exist. Furthermore, note that because, $B_m \neq 0$ is positive semi-definite and $\left(\sum_{l=1}^{i}(B_l + \tilde{N}_l)\right)^{-1}$ is strictly positive definite, the last summand is always positive and could be made arbitrarily large by choosing α_1 to be large enough. Therefore, we choose α_1 to be large enough such that the entire expression is positive for $i = 1$. For $i = 2$, again, the last summand can be made arbitrarily large by choosing α_2 to be large enough. Thus, we can recursively choose α_i, such that the directional derivative is positive for all $i = 1, \ldots, m-1$. ■

Proposition VII.2: Let $B_1^*, \ldots, B_{m-1}^*$ be an optimal solution of the problem in equation (4-7). If $S \succ 0$ (strictly positive), there exists a *strict direction* at $B_1^*, \ldots, B_{m-1}^*$.

Proof: We suggest a direction, $L_1^*, \ldots, L_{m-1}^*$ of the form

$$L_i^* = -B_i^* + \frac{1}{m} \sum_{j \neq i \, , \, j < m} B_j^* + \alpha_i B_m^* \; , \quad i = 1, \ldots, m-1 \tag{VII-2}$$

for some $\alpha_i > 0$, $i = 1, \ldots, m-1$.

As a byproduct of the proof of the previous proposition (VII.1), we have shown that if $S \succ 0$, this direction is also a strict direction for the problem in (4-7).

■

Appendix VIII

Proposition VIII.1: Let $N_i \succ 0$, $i = 1, \ldots, m$, $B_i \succeq 0$, $i = 1, \ldots, m$ and $O_i \succeq 0$, $i = 1, \ldots, m$ be $t \times t$ matrices such that $B_i \cdot O_i = 0_{t \times t}$, $i = 1, \ldots, m$ and such that for all $k = 1, \ldots, m$,

$$\gamma_{k+1} \left(\sum_{i=1}^{k} B_i + N_{k+1} \right)^{-1} + O_{k+1} = \gamma_k \left(\sum_{i=1}^{k} B_i + N_k \right)^{-1} + O_k \tag{VIII-3}$$

for some $0 < \gamma_1 \leq \ldots \leq \gamma_m$. Then, we can find a set of $t \times t$ matrices, $0 \preceq N'_1 \preceq \cdots \preceq N'_m$ such that $N'_i \preceq N_i$, $\forall i = 1, \ldots, m$ and such that for all $k = 1, \ldots, m$,

$$\begin{aligned}\gamma_{k+1}\left(\sum_{i=1}^{k} B_i + N_{k+1}\right)^{-1} + O_{k+1} &= \gamma_{k+1}\left(\sum_{i=1}^{k} B_i + N'_{k+1}\right)^{-1} \\ &= \gamma_k\left(\sum_{i=1}^{k} B_i + N'_k\right)^{-1} = \gamma_k\left(\sum_{i=1}^{k} B_i + N_k\right)^{-1} + O_k\end{aligned} \tag{VIII-4}$$

Furthermore,

$$\frac{|\sum_{i=1}^{k} B_i + N_k|}{|\sum_{i=1}^{k-1} B_i + N_k|} = \frac{|\sum_{i=1}^{k} B_i + N'_k|}{|\sum_{i=1}^{k-1} B_i + N'_k|}, \qquad \forall k = 1, \ldots, m$$

Before proving the proposition we give two intermediate results.

Proposition VIII.2: Let $B \succeq 0$, $X \succ 0$ and $O \succeq 0$ be $t \times t$ symmetric matrices such that $B \cdot O = 0_{t \times t}$ and let α be a strictly positive scalar. Then, the following statements hold:

1. $\alpha(B+X)^{-1} + O = \alpha(B+X')^{-1}$ where $X \succeq X' = (X^{-1} + \frac{1}{\alpha}O)^{-1} \succ 0$
2. $\frac{|B+X|}{|X|} = \frac{|B+X'|}{|X'|}$

Proof: To prove the first statement, we will show that $\left((B+X)^{-1} + \frac{1}{\alpha}O\right)^{-1} = B + X'$ where X' is as defined in the statement. For that purpose we write,

$$\begin{aligned}\left((B+X)^{-1} + \frac{1}{\alpha}O\right)^{-1} &= \left((B+X)^{-1}\left(I + (B+X)\frac{1}{\alpha}O\right)\right)^{-1} - B + B \\ &\overset{(a)}{=} \left(I + \frac{1}{\alpha}XO\right)^{-1}(B+X) - B + B \\ &\overset{(b)}{=} \left(I + \frac{1}{\alpha}XO\right)^{-1}\left((B+X) - \left(I + \frac{1}{\alpha}XO\right)B\right) + B \\ &= \left(I + \frac{1}{\alpha}XO\right)^{-1} X + B = \left(X^{-1}\left(I + \frac{1}{\alpha}XO\right)\right)^{-1} + B \\ &= B + \left(X^{-1} + \frac{1}{\alpha}O\right)^{-1}\end{aligned}$$

where in (a) and (b) we used $B \cdot O = 0_{t \times t}$.

We can now prove the second statement. As $(B+X')^{-1} = (B+X)^{-1} + \frac{1}{\alpha}O$, we can write:

$$\begin{aligned}\frac{|B+X'|}{|X'|} &= \frac{|I|}{|X'(B+X')^{-1}|} = \frac{|I|}{|(B+X'-B)(B+X')^{-1}|}\\ &= \frac{|I|}{|I-B(B+X')^{-1}|} = \frac{|I|}{\left|I-B\left((B+X)^{-1}+\frac{1}{\alpha}O\right)\right|}\\ &\overset{(a)}{=} \frac{|I|}{|I-B(B+X)^{-1}|} = \frac{|B+X|}{|X|}\end{aligned}$$

where, once again, in (a) we used $B \cdot O = 0_{t\times t}$. ■

Proposition VIII.3: Let $X_1 \succ 0$, $X_2 \succ 0$, $B_1 \succeq 0$, $B_2 \succeq 0$ and $O_2 \succeq 0$ be $t \times t$ symmetric matrices such that $B_2 \cdot O_2 = 0_{t\times t}$. Given that for some scalar $0 < \alpha \leq 1$, $(B_1+X_2)^{-1}+O_2 = \alpha(B_1+X_1)^{-1}$, the following two statements hold:

1. $(B_1+X_2)^{-1}+O_2 = (B_1+X_2')^{-1}$ for some X_2' satisfying $X_2 \succeq X_2' \succeq X_1$.
2. $\frac{|B_1+B_2+X_2|}{|B_1+X_2|} = \frac{|B_1+B_2+X_2'|}{|B_1+X_2'|}$

Proof: Define $K = (B_1+X_2)^{-1}+O_2$. As $O_2 \succeq 0$, we know that $K^{-1} \preceq (B_1+X_2)$ and since $0 < \alpha \leq 1$ and $(B_1+X_2)^{-1}+O_2 = \alpha(B_1+X_1)^{-1}$, we know that $(B_1+X_1) \preceq K^{-1} \preceq (B_1+X_2)$. Therefore, by choosing $X_2' = K^{-1}-B_1$, we have $X_2 \succeq X_2' \succeq X_1$ and we can write:

$$(B_1+X_2')^{-1} = K = (B_1+X_2)^{-1}+O_2$$

By the above equation we may write:

$$\begin{aligned}\frac{|B_1+B_2+X_2'|}{|B_1+X_2'|} &= \frac{|B_2(B_1+X_2')^{-1}+I|}{|I|} = \frac{|B_2\left((B_1+X_2)^{-1}+O_2\right)+I|}{|I|}\\ &= \frac{|B_2(B_1+X_2)^{-1}+I|}{|I|} = \frac{|B_1+B_2+X_2|}{|B_1+X_2|}\end{aligned}$$

■

Proof of Proposition VIII.1: For brevity, we rewrite expression (VIII-3) with some notational modifications. Define

$$\begin{aligned} L_k &= \left(\sum_{i=1}^{k} B_i + N_{k+1}\right)^{-1} + \frac{1}{\gamma_{k+1}} O_{k+1} \\ R_k &= \gamma_k \left(\sum_{i=1}^{k} B_i + N_k\right)^{-1} + O_k \qquad k = 1 \ldots m-1, \end{aligned} \tag{VIII-5}$$

By (VIII-3), we have

$$L_k = \frac{1}{\gamma_{k+1}} R_k, \quad \forall k = 1 \ldots m-1. \tag{VIII-6}$$

We need to show that

$$L_k = \left(\sum_{i=1}^{k} B_i + N'_{k+1}\right)^{-1} = \frac{\gamma_k}{\gamma_{k+1}} \left(\sum_{i=1}^{k} B_i + N'_k\right)^{-1} = \frac{1}{\gamma_{k+1}} R_k, \quad \forall k = 1, \ldots, m-1 \tag{VIII-7}$$

for some matrices $N'_1, \ldots, N'_m$ such that $N'_i \preceq N_i$, $\forall i$ and such that $0 \prec N'_1 \preceq \cdots \preceq N'_m$. Furthermore, we need to show that for these matrices, $N'_1, \ldots, N'_m$, we have

$$\frac{|\sum_{i=1}^{k} B_i + N_k|}{|\sum_{i=1}^{k-1} B_i + N_k|} = \frac{|\sum_{i=1}^{k} B_i + N'_k|}{|\sum_{i=1}^{k-1} B_i + N'_k|}, \qquad \forall k = 1, \ldots, m. \tag{VIII-8}$$

We use induction on k and begin by exploring (VIII-5) and (VIII-6) for $k = 1$. As $O_1 \succeq 0$ and as $B_1 \cdot O_1 = 0$, by Proposition VIII.2, we can replace R_1 with $\gamma_1 (B + N'_1)^{-1}$ where $0 \prec N'_1 \preceq N_1$. Furthermore, by the same proposition we have $\frac{|B_1 + N_1|}{|N_1|} = \frac{|B_1 + N'_1|}{|N'_1|}$. Since $\gamma_2 \geq \gamma_1$, by Proposition VIII.3, L_1 can be replaced by $(B_1 + N'_2)^{-1}$ such that $N'_1 \preceq N'_2 \preceq N_2$ and in addition $\frac{|B_1 + B_2 + N_2|}{|B_2 + N_2|} = \frac{|B_1 + B_2 + N'_1|}{|B_1 + N'_2|}$ and therefore, (VIII-7) holds for k=1 and (VIII-8) holds for $k = 1, 2$.

Next, we assume that (VIII-7) holds for $k = 1, \ldots, n$ for some $1 \leq n < m$ with matrices $N'_1, \ldots, N'_{n+1}$ such that $N'_i \preceq N_i$, $i = 1, \ldots, n+1$ and such that $0 \prec N'_1 \preceq \cdots \preceq N'_{n+1}$ and prove that (VIII-7) must hold for $k = 1, \ldots, n+1$ and (VIII-8) holds for $k = 1, \ldots, n+2$ with matrices $N'_1, \ldots, N'_{n+2}$ such that $N'_i \preceq N_i$, $i = 1, \ldots, n+2$ and such that $0 \prec N'_1 \preceq \cdots \preceq N'_{n+2}$. For that purpose, we define

$$Q = \sum_{i=1}^{n} B_i + N_{n+1}. \tag{VIII-9}$$

As $O_{n+1} \succeq 0$ and as $B_{n+1} \cdot O_{n+1} = 0_{t\times t}$, we can use Proposition VIII.2 to rewrite the expression for R_{n+1} in (VIII-5) as follows,

$$R_{n+1} = \gamma_{n+1}(B_{n+1} + Q)^{-1} + O_{n+1} = \gamma_{n+1}(B_{n+1} + Q')^{-1} \tag{VIII-10}$$

where $Q' \preceq Q$ and where $Q' = (Q^{-1} + \frac{1}{\gamma_{n+1}} O_{n+1})^{-1}$. However, by (VIII-9) and (VIII-5), we may write $Q' = L_n^{-1}$. Furthermore, by our induction assumption, (VIII-7) holds for $k = n$ and therefore, we may write

$$Q' = L_n^{-1} = \sum_{i=1}^{n} B_i + N'_{n+1}. \tag{VIII-11}$$

Thus, by (VIII-11), (VIII-10), (VIII-5) and (VIII-6) we may write

$$L_{n+1} = \left(\sum_{i=1}^{n+1} B_i + N_{n+2}\right)^{-1} + \frac{1}{\gamma_{n+2}} O_{n+2} = \frac{\gamma_{n+1}}{\gamma_{n+2}} \left(\sum_{i=1}^{n+1} B_i + N'_{n+1}\right)^{-1} = \frac{1}{\gamma_{n+2}} R_{n+1}.$$

As $O_{n+2} \succeq 0$ and as $\frac{\gamma_{n+1}}{\gamma_{n+2}} \leq 1$, we can use Proposition VIII.3 to rewrite the above expression as follows:

$$L_{n+1} = \left(\sum_{i=1}^{n+1} B_i + N'_{n+2}\right)^{-1} = \frac{\gamma_{n+1}}{\gamma_{n+2}} \left(\sum_{i=1}^{n+1} B_i + N'_{n+1}\right)^{-1} = \frac{1}{\gamma_{n+2}} R_{n+1}.$$

for some $N'_{n+2} \succeq N'_{n+1}$ and such that $N'_{n+2} \preceq N_{n+2}$. Furthermore, by the same proposition we have

$$\frac{|\sum_{i=1}^{n+2} B_i + N_{n+2}|}{|\sum_{i=1}^{n+1} B_i + N_{n+2}|} = \frac{|\sum_{i=1}^{n+2} B_i + N'_{n+2}|}{|\sum_{i=1}^{n+1} B_i + N'_{n+2}|},$$

and thus we have shown that (VIII-7) holds for $k = 1, \ldots, n+1$ and that (VIII-8) holds for $k = 1, \ldots, n+2$. ■

References

[1] Patrick P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise," *IEEE Trans. on IT.*, pp. 279–280, Mar. 1974.

[2] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Belmont, Massachusetts, 2003.

[3] Nelson M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. on IT.*, pp. 267–271, Apr. 1965.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[5] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Trans. on IT.*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[6] M. Costa, "Writing on dirty paper," *IEEE Trans. on IT.*, vol. 29, pp. 439–441, May 1983.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, 1991.

[8] A. El Gamal, "A capacity of a class of broadcast channels," *IEEE Trans. on IT.*, vol. 25, no. 2, pp. 166–169, Mar. 1979.

[9] S. Gelfand and M. Pinsker, "Coding for channels with random parameters," *Probl, Control Inform. Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.

[10] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of gaussian multiple-access and broadcast channels," *IEEE Trans. on IT.*, vol. 50, no. 5, pp. 768–783, May 2004.

[11] J. Körner and K. Marton, "General brodcast channels with degraded message sets,"

IEEE Trans. on IT., vol. 23, no. 1, pp. 60–64, Jan. 1977.

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

[13] B.M. Zaidel S. Shamai (Shitz), "Enhancing the cellular down-link capacity via co-processing at the transmitting end," in *Proc. IEEE Semiannual Vehicular Technology, VTC2001 Spring Conference*, Rhodes, Greece, May 6–9, 2001.

[14] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. on IT.*, vol. 24, pp. 374–377, May 1978.

[15] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channel with partial side information," *submitted to IEEE Trans. on IT.*

[16] D. Tse and P. Viswanath, "On the capacity of the miltiple antenna broadcast channel," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G. J. Foschini and S. Verdú, Eds. American Mathematical Society, DIMACS, Providence, RI, 2003.

[17] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of gaussian mimo broadcast channels," *IEEE Trans. on IT.*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[18] S. Vishwanath, G. Kramer, S. Shamai (Shitz), S. Jafar, and A. Goldsmith, "Capacity bounds for gaussian vector broadcast channels," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G. J. Foschini and S. Verdú, Eds. American Mathematical Society, DIMACS, Providence, RI, 2003.

[19] P. Viswanath and D. Tse, "Sum capacity of the vector gaussian channel and uplink-downlink duality," *IEEE Trans. on IT.*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[20] P. Viswanathan, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. on IT.*, vol. 48, no. 6, pp. 1276–1294, June 2002.

[21] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the gaussian mimo broadcast channel," in *Proc. CISS 2004*, Princeton, NJ, March, 2004.

[22] W. Yu, "Uplink-downlink duality via minimax duality," in *proc. of ISIT 2004*, Chicago, IL, June 27 – July 2, 2004, p. 177.

[23] W. Yu and J. Cioffi, "The sum capacity of a gaussian vector broadcast channel," in *proc. of ISIT 2002*, Lausanne, Switzerland, June 30 – July 5, 2002, p. 498.

[24] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang, "Writing on colored paper," in *proc. of ISIT 2001*, Washington, DC, June 24 – 29, 2001, p. 309.