The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel

Hanan Weingarten, *Student Member, IEEE,* Yossef Steinberg, *Member, IEEE,* and
Shlomo Shamai (Shitz), *Fellow, IEEE*

Abstract—The Gaussian multiple-input multiple-output (MIMO) broadcast channel (BC) is considered. The dirty-paper coding (DPC) 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 a superposition of Gaussian codes is optimal for the degraded vector broadcast channel and that DPC is optimal for the nondegraded case. Furthermore, the capacity region is characterized under a wide range of input constraints, accounting, as special cases, for the total power and the per-antenna power constraints.

Index Terms—Broadcast channel, capacity region, dirty-paper coding (DPC), enhanced channel, entropy power inequality, Minkowski's inequality, multiple-antenna.

I. INTRODUCTION

WE consider a Gaussian multiple-input multiple-output (MIMO) broadcast channel (BC) and find the capacity region of this channel. The transmitter is required to send independent messages to m receivers. We assume that the transmitter has t transmit antennas and user $i, i = 1, 2, \ldots, m$ has r_i 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 a compact set of positive semidefinite matrices. The Gaussian BC is an additive noise channel and each time sample can be represented using the following expression:

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x} + \boldsymbol{n}_i, \qquad i = 1, 2, \ldots, m \qquad (1)$$

where

- $\boldsymbol{x}$ is a real input vector of size $t \times 1$. Under an average total power limitation P at the transmitter, we require that $E[\boldsymbol{x}^T \boldsymbol{x}] \leq P$. Under an input covariance constraint, we will require that $E[\boldsymbol{x}\boldsymbol{x}^T] \preceq \boldsymbol{S}$ for some $\boldsymbol{S} \succeq 0$ (where $\prec, \preceq, \succ,$ and $\succeq$ denote partial ordering between symmetric

matrices where $\boldsymbol{B} \succeq \boldsymbol{A}$ means that $(\boldsymbol{B} - \boldsymbol{A})$ is a positive semidefinite matrix).

- $\boldsymbol{y}_i$ is a real output vector, received by user $i, i = 1, 2, \ldots, m$. This is a vector of size $r_i \times 1$ (the vectors $\boldsymbol{y}_i$ are not necessarily of the same size).
- $\boldsymbol{H}_i$ 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.
- $\boldsymbol{n}_i$ is a real Gaussian random vector with zero mean and a covariance matrix $\boldsymbol{N}_i = E[\boldsymbol{n}_i \boldsymbol{n}_i^T] \succ 0$. No additional structure is imposed on $\boldsymbol{H}_i$ or $\boldsymbol{N}_i$, except that $\boldsymbol{N}_i$ must be strictly positive definite, for $i = 1, 2, \ldots, m$.

We shall refer to the channel in (1) as the General MIMO BC (GMBC). Note that complex MIMO BCs can be easily accommodated by representing all complex vectors and matrices using real vectors and matrices having twice and four times (for matrices) the number of elements, corresponding to real and imaginary entries.

In this paper, we find the capacity region of the BC in expression (1) under various input constraints. In particular, we will consider the average total power constraint and the input covariance constraint. This model and our results are quite relevant to many applications in wireless communications such as cellular systems [17], [18], [23].

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 [15], but it is unknown whether it coincides with the capacity region. Nevertheless, for some special cases, a single-letter formula for the capacity region does exist [9]–[11], [14] and coincides with the Marton region [15]. One such case is the degraded BC (see [10, pp. 418–428]) where the channel input and outputs form a Markov chain. Since the capacity region of the BC depends only on its conditional marginal distributions $P_{Y_i|X}$ (see [10, p. 422]), it turns out that when the BC defined by (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, 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) [1], [3] is used. Unfortunately, in general, the channel given by expression (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 III, we recount Bermans' proof for the vector channel and show where the extension to the degraded vector Gaussian BC fails.

Recent years have seen intensive work on the Gaussian MIMO BC. Caire and Shamai [7] 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 [8], [12], [28] results for coding for a channel with both additive Gaussian noise and additive Gaussian interference, where the interference is noncausally 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 canceled out. Using this coding technique, each user can precode his own information, based on the signals of the users following it [7], [20], [22], [27] (for some arbitrary ordering of the users) and treating their respective signals as noncausally known interference.

Caire and Shamai [7] 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 capacity (or maximum throughput as it was called in [7]). Hence, it was shown that for rate pairs that obtain the channel's sum capacity, DPC is optimal. Their main tool was the Sato upper bound [16] (or the cooperative upper bound).

Following the work by Caire and Shamai, new papers appeared [20], [22], [27] which expanded the maximum throughput claim in [7] 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 [27], 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 [20], [22]. To overcome the difficulty of expanding the calculations of the maximum throughput in [7] to more than two users, each equipped with more than one antenna, the idea of MAC-BC duality [13], [20], [22] was used.

Another step toward characterizing the capacity region of the MIMO BC is reported in [21] and [19]. 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 [21]. Indeed, the authors conjecture that 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 cannot be directly applied.

In a conference version of this work [24], we reported a proof of this conjecture and thus, along with the DSM bound [19], [21], we have shown that the DPC rate region is indeed the capacity region of the MIMO BC. Here, we provide a more cohesive view. We derive the results on first principles, not hinging on any of the above existing results (the multiple-access channel (MAC)-BC duality [20], [22] and the DSM upper bound [19], [21]). 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. 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.

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 II, we show that we can characterize the capacity region of the MIMO BC which input is constrained to have a covariance matrix S that lies in a compact set. This allows us to characterize the capacity region for various scenarios; for example, the per-antenna power constraint. In this sense, our result parallels a result reported in [26] for the sum capacity of the MIMO BC under per-antenna power constraints.

This paper is structured as follows.

1) In Section II, we introduce two subclasses of the MIMO BC: the aligned and degraded MIMO BC (ADBC) and the aligned MIMO BC (AMBC). We also prove that the capacity region of the MIMO BC under a total power constraint on the input can be easily deduced from the capacity region of the same channel under a covariance constraint on the input.

2) Next, in Section III, we find and prove the capacity region of the ADBC under a covariance matrix constraint and total power constraint on the input.

3) Using the result for the ADBC, we extend the proof of the capacity region to AMBCs in Section IV.

4) Finally, in Section V, we further extend our results and find the capacity region of the Gaussian MIMO BC defined in (1).

II. PRELIMINARIES

A. Subclasses of the Gaussian MIMO BC (the ADBC and AMBC)

In (1), we presented the 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 consider is the 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 ($\boldsymbol{H}_1 = \cdots = \boldsymbol{H}_m = \boldsymbol{I}$). Furthermore, we require that this subclass will be degraded and assume that the additive noise vector covariance matrices at each of the receivers are ordered such that $0 \prec \boldsymbol{N}_1 \preceq \boldsymbol{N}_2 \preceq \cdots \preceq \boldsymbol{N}_m$. Taking into account the fact that the capacity region of BCs (in general) depends on the marginal distributions $P_{Y_i|X}$, we may assume without loss of generality that a time sample of an ADBC is given by the following expression:

$$\boldsymbol{y}_i = \boldsymbol{x} + \sum_{j=1}^{i} \tilde{\boldsymbol{n}}_j, \qquad i = 1, 2, \ldots, m \qquad (2)$$

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

increments such that $\tilde{N}_i = E[\tilde{n}_i \tilde{n}_i^T] = N_i - N_{i-1}$ (where we define $N_0 = 0_{t \times t}$).

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 given by the following expression:

$$y_i = x + n_i \qquad i = 1, 2, \ldots, m \qquad (3)$$

where y_i and x are real vectors of size $t \times 1$ and where n_i, $i = 1, \ldots, m$ are memoryless real Gaussian noise vectors such that $E[n_i n_i^T] = N_i \succ 0_{t \times t}$.

In the case where the gain matrices of the GMBC (the channel given in (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 y_i by H_i^{-1}. However, a problem arises when the gain matrices are no longer square or invertible. In Section V (see proof of Theorem 5), we show that the capacity region of the GMBC with non-square or noninvertible gain matrices can be obtained by 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 are 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 III and IV) will be devoted to characterizing the capacity regions of the ADBC and AMBC.

B. The Covariance Matrix Constraint

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 this region will be given for a covariance matrix constraint, $S \succeq 0$, such that $E[xx^T] \preceq S$. We show here that we can extend this constraint to the case where the input covariance matrix must lie in a compact set $\mathcal{S}$ (matrices are elements in a metric space where the metric distance between A and B is defined by the Frobenius norm: $\text{tr}\{(A-B)(A-B)^T\}$. A compact set in this space is defined in a standard manner, see [5, pp. 45–51]). The total power constraint is just one example of a compact set constraint as the set $\mathcal{S} = \{S \mid \text{tr}\{S\} \preceq P\}$ is a compact set.

Denote by $\bar{R} = (R_1, R_2, \ldots, R_m)$ a given rate vector. We use $C(n, S, \bar{R}, \epsilon)$ to denote a codebook that maps a set of m message indices $W_1, \ldots, W_m, W_i \in \{1, \ldots, e^{nR_i}\}$, $i = 1, \ldots, m$ onto an input word (a real matrix of size $t \times n$), where the maximum-likelihood (ML) decoder at each of the receivers decodes the appropriate message index with an average prob-

ability of decoding error no greater than ϵ. Furthermore, the codewords are such that

$$S = \frac{1}{e^{n \sum R_i}} \sum_{\overline{X} \in C(n, S, \bar{R}, \epsilon_i)} \overline{X} \cdot \overline{X}^T.$$

A rate vector $\bar{R}$ is said to be achievable under a matrix covariance constraint S, if there exists an infinite sequence of codebooks, $C(n_i, S_i, \bar{R}, \epsilon_i)$, $i = 1, 2, \ldots \infty$, with increasing lengths n_i, rate vectors $\bar{R}$, matrices $S_i \preceq S$, and decreasing probabilities of error ϵ_i, such that $\epsilon_i \to 0$ as $i \to \infty$. Similarly, a rate vector $\bar{R}$ is said to be achievable with a matrix covariance constraint that lies in a compact set $\mathcal{S}$ if $S_i \in \mathcal{S}$ for all i.

For a given covariance matrix constraint S, $\mathcal{C}(S, N_{1\ldots m}, H_{1,\ldots,m})$ denotes the capacity region of the GMBC under a covariance constraint and is defined by the closure of the set of all achievable rates under a covariance constraint S. Similarly, $\mathcal{C}(\mathcal{S}, N_{1,\ldots,m}, H_{1,\ldots,m})$ and $\mathcal{C}(P, N_{1,\ldots,m}, H_{1,\ldots,m})$ denote the capacity region under a compact set constraint $\mathcal{S}$ and a total power constraint P. Note that

$$\mathcal{C}(P, N_{1,\ldots,m}, H_{1,\ldots,m}) = \mathcal{C}(\mathcal{S} = \{S | \text{tr}\{S\} \leq P\}, N_{1,\ldots,m}, H_{1,\ldots,m}).$$

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

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

Lemma 1: Assume that $\mathcal{S}$ is a compact set of positive semidefinite matrices. If $\mathcal{C}(S, N_{1,\ldots,m}, H_{1,\ldots,m})$ is contiguous w.r.t. S, then

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

Proof: The fact that

$$\mathcal{C}(\mathcal{S}, N_{1,\ldots,m}, H_{1,\ldots,m}) \supseteq \bigcup_{S \in \mathcal{S}} \mathcal{C}(S, N_{1,\ldots,m}, H_{1,\ldots,m})$$

follows from the definition of those regions. Therefore, we only need to prove the inclusion in the reverse direction. We will prove that every rate vector

$$\bar{R} = (R_1, \ldots, R_m) \in \mathcal{C}(\mathcal{S}, N_{1,\ldots,m}, H_{1,\ldots,m})$$

also lies in $\mathcal{C}(S, N_{1,\ldots,m}, H_{1,\ldots,m})$ for some S such that $S \in \mathcal{S}$.

If indeed $\bar{R} \in \mathcal{C}(\mathcal{S}, N_{1,\ldots,m}, H_{1,\ldots,m})$, then there exists an infinite sequence of codebooks $C(n_i, S_i, \bar{R}, \epsilon_i)$, $i = 1, 2, \ldots, \infty$ with rate vectors $\bar{R}$ and decreasing probabilities of error, ϵ_i, such

that $\epsilon_i \to 0$ as $i \to \infty$. As $\mathcal{S}$ is a compact set in a metric space, for any infinite sequence of points in $\mathcal{S}$ there must be a subsequence that converges to a point $\boldsymbol{S}_0 \in \mathcal{S}$. Hence, for any arbitrarily small $\delta > 0$, we can find an increasing subsequence $i(k)$ such that $\boldsymbol{S}_{i(k)} \preceq \boldsymbol{S}_0 + \delta \boldsymbol{I}, \forall k > 0$.

In other words, if $\bar{R} \in \mathcal{C}(\mathcal{S}, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$, for any $\delta > 0$, we can find a sequence of codebooks $C(n_k, \boldsymbol{S}_0 + \delta \boldsymbol{I}, \bar{R}, \epsilon_k)$, with $\boldsymbol{S}_0 \in \mathcal{S}$, which achieve arbitrarily small error probabilities. Therefore, for every $\delta > 0$, $\bar{R} \in \mathcal{C}(\boldsymbol{S}_0 + \delta \boldsymbol{I}, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$. By the contiguity of $\mathcal{C}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$ with respect to $\boldsymbol{S}$, we conclude that every ϵ-ball around $\bar{R}$ contains vectors which lie in $\mathcal{C}(\boldsymbol{S}_0, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$ and therefore, $\bar{R}$ must be a limit point of $\mathcal{C}(\boldsymbol{S}_0, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$. As $\mathcal{C}(\boldsymbol{S}_0, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$ is a closed set by definition, we conclude that $\bar{R} \in \mathcal{C}(\boldsymbol{S}_0, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$. $\qquad\square$

Lemma 1 can be easily applied to the total power constraint, as stated in the following.

Corollary 1: The capacity region under a total power constraint is given by

$$\mathcal{C}(P, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m}) = \bigcup_{\text{tr}\{\boldsymbol{S}\} \leq P} \mathcal{C}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m}, \boldsymbol{H}_{1,\dots,m})$$

Proof: The proof is a direct result of Lemma 1 where the compact set is defined as $\mathcal{S} = \{\boldsymbol{S} | \text{tr}\{\boldsymbol{S}\} \leq P\}$. $\qquad\square$

2) The Capacity Region Under a Noninvertible Covariance Matrix Constraint: We differentiate between the case where the covariance matrix constraint $\boldsymbol{S}$ is strictly positive definite (and hence invertible), and the case where it is positive semidefinite but noninvertible, $|\boldsymbol{S}| = 0$. It turns out that for an aligned MIMO BC (either an ADBC or an AMBC) with a noninvertible covariance matrix constraint $|\boldsymbol{S}| = 0$, 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 $\boldsymbol{S}$ is strictly positive. A formal presentation of the above argument is given in the following lemma.

Lemma 2: Consider an AMBC with t transmit antennas, noise covariance matrices $\boldsymbol{N}_1, \dots, \boldsymbol{N}_m$, and a covariance matrix constraint $\boldsymbol{S}$, such that $\text{rank}(\boldsymbol{S}) = r_{\boldsymbol{S}} < t$. Then we have the following.

1) There exists an AMBC, with $r_{\boldsymbol{S}}$ antennas and with noise covariance matrices $\hat{\boldsymbol{N}}_i, i = 1, 2, \dots, m$ which has the same capacity region (as the original AMBC) under a covariance matrix constraint $\hat{\boldsymbol{S}}$, of rank $r_{\boldsymbol{S}}$. Furthermore, if $\boldsymbol{N}_1 \preceq \cdots \preceq \boldsymbol{N}_m$ (the channel is an ADBC), then $\hat{\boldsymbol{N}}_1 \preceq \cdots \preceq \hat{\boldsymbol{N}}_m$ (i.e., the equivalent channel is also an ADBC).

2) In the equivalent channel we have

$$\hat{\boldsymbol{S}} = \left[\boldsymbol{0}_{r_{\boldsymbol{S}} \times (t - r_{\boldsymbol{S}})} \; \boldsymbol{I}_{r_{\boldsymbol{S}} \times r_{\boldsymbol{S}}} \right] \boldsymbol{\Lambda}_{\boldsymbol{S}} \left[\boldsymbol{0}_{r_{\boldsymbol{S}} \times (t - r_{\boldsymbol{S}})} \; \boldsymbol{I}_{r_{\boldsymbol{S}} \times r_{\boldsymbol{S}}} \right]^T$$

$$\hat{\boldsymbol{N}}_i = \boldsymbol{N}_i^C - \left(\boldsymbol{N}_i^B \right)^T \left(\boldsymbol{N}_i^A \right)^{-1} \left(\boldsymbol{N}_i^B \right), \quad i = 1, 2, \dots, m$$

where $\boldsymbol{\Lambda}_{\boldsymbol{S}}$ and $\boldsymbol{U}_{\boldsymbol{S}}$ are defined as the eigenvalue and eigenvector matrices (unitary matrices) of $\boldsymbol{S} = \boldsymbol{U}_{\boldsymbol{S}} \boldsymbol{\Lambda}_{\boldsymbol{S}} \boldsymbol{U}_{\boldsymbol{S}}^T$ such that $\boldsymbol{\Lambda}_{\boldsymbol{S}}$ has all its nonzero values on the bottom right

of the diagonal. $\boldsymbol{N}_i^A, \boldsymbol{N}_i^B$, and $\boldsymbol{N}_i^C$ are matrices of sizes $(t - r_{\boldsymbol{S}}) \times (t - r_{\boldsymbol{S}}), (t - r_{\boldsymbol{S}}) \times r_{\boldsymbol{S}}$ and $r_{\boldsymbol{S}} \times r_{\boldsymbol{S}}$ such that

$$\boldsymbol{U}_{\boldsymbol{S}}^T \boldsymbol{N}_i \boldsymbol{U}_{\boldsymbol{S}} = \begin{pmatrix} \boldsymbol{N}_i^A & \boldsymbol{N}_i^B \\ (\boldsymbol{N}_i^B)^T & \boldsymbol{N}_i^C \end{pmatrix}.$$

Proof: see Appendix I. $\qquad\square$

III. THE CAPACITY REGION OF THE ADBC

In this section, we characterize the capacity region of the ADBC given in (2). As mentioned in the Introduction, even though this channel is degraded and we have a single-letter formula for its capacity region, proving that Gaussian inputs are optimal is not trivial. In the following subsections, we give motivations, present intermediate results, and finally, state and prove our result for the capacity region of the ADBC.

We begin by defining the achievable rate region due to Gaussian coding under a covariance matrix constraint $\boldsymbol{S} \succeq 0$. Note that as the channel is degraded, there is no point in using DPC [7], [8], [20], [22], [27].

It is well known that for any covariance matrix input constraint $\boldsymbol{S}$ and a set of semidefinite matrices $\boldsymbol{B}_i \succeq 0$ $(i = 1, \cdots m)$ such that $\sum_{i=1}^m \boldsymbol{B}_i \preceq \boldsymbol{S}$, it is possible to achieve the following rates:

$$R_k \leq R_k^G(\boldsymbol{B}_{1,\dots,m}, \tilde{\boldsymbol{N}}_{1,\dots,m}), \qquad \forall k \in 1, \dots, m \quad (4)$$

where

$$R_1^G(\boldsymbol{B}_{1,\dots,m}, \tilde{\boldsymbol{N}}_{1,\dots,m}) = \frac{1}{2} \log |\tilde{\boldsymbol{N}}_1^{-1}(\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1)|$$

$$R_k^G(\boldsymbol{B}_{1,\dots,m}, \tilde{\boldsymbol{N}}_{1,\dots,m}) = \frac{1}{2} \log \frac{\left| \left(\sum_{i=1}^k \boldsymbol{B}_i + \sum_{i=1}^k \tilde{\boldsymbol{N}}_i \right) \right|}{\left| \left(\sum_{i=1}^{k-1} \boldsymbol{B}_i + \sum_{i=1}^k \tilde{\boldsymbol{N}}_i \right) \right|},$$
$$k = 2, 3, \dots, m. \quad (5)$$

The coding scheme that achieves the above rates uses a superposition of Gaussian codes with covariance matrices $\boldsymbol{B}_i$, and successive decoding at the receivers. The Gaussian rate region is defined as follows.

Definition 2 (Gaussian Rate Region of an ADBC): Let $\boldsymbol{S}$ be a positive semidefinite matrix. Then, the Gaussian rate region of an ADBC under a covariance matrix constraint $\boldsymbol{S}$ is given by

$$\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\dots,m})$$
$$= \left\{ \begin{array}{l} \left(R_1^G(\boldsymbol{B}_{1,\dots,m}, \tilde{\boldsymbol{N}}_{1,\dots,m}), \dots, R_m^G(\boldsymbol{B}_{1,\dots,m}, \tilde{\boldsymbol{N}}_{1,\dots,m}) \right) \Big| \\ \text{s.t.} \quad \boldsymbol{S} - \sum_{i=1}^m \boldsymbol{B}_i \succeq 0, \quad \boldsymbol{B}_i \succeq 0, \; \forall i = 1, \dots, m \end{array} \right\}$$
$$(6)$$

Our goal is to show that $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\dots,m})$ is indeed the capacity region or the ADBC.

A. Motivation: A Direct Application of Bergmans' Proof to the ADBC and its Pitfalls

Before proving that $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\dots,m})$ is the capacity region, we explain why Bergmans' [1] proof for the scalar Gaussian BC does not directly extend to the vector case (the ADBC). 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. For the interest of conciseness, only a sketch of Bergmans' proof will be given here. Furthermore, Bergmans' approach will only be presented for the two-users case.

We wish to show, using Bergmans' line of thought, that $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,2})$ is also the capacity region. Assume, in contrast, that there is an achievable rate pair (R_1, R_2) which lies outside $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,2})$. Then, we can find a pair of matrices $\boldsymbol{B}_1$ and $\boldsymbol{B}_2 = \boldsymbol{S} - \boldsymbol{B}_1$ such that

$$R_1 \geq R_1^G(\boldsymbol{B}_{1,2}, \tilde{\boldsymbol{N}}_{1,2})$$

and

$$R_2 > R_2^G(\boldsymbol{B}_{1,2}, \tilde{\boldsymbol{N}}_{1,2}). \tag{7}$$

Let W_1 and W_2 denote the massage indices 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. Let $\overline{\overline{Z}}_1$ and $\overline{\overline{Z}}_2$ denote the additive noise such that $\overline{\overline{Y}}_1 = \overline{\overline{X}} + \overline{\overline{Z}}_1$ and $\overline{\overline{Y}}_2 = \overline{\overline{Y}}_1 + \overline{\overline{Z}}_2$. Note that $\overline{\overline{Z}}_1$ and $\overline{\overline{Z}}_2$ are $t \times n$ random matrices the columns of which are independent Gaussian random vectors with covariance matrices $\tilde{\boldsymbol{N}}_1$ and $\tilde{\boldsymbol{N}}_2$, respectively. In addition $\overline{\overline{Z}}_1$ and $\overline{\overline{Z}}_2$ are independent of $\overline{\overline{X}}, W_1$, and W_2. As W_1 and W_2 are independent we can use Fano's inequality to write

$$R_1 \leq \frac{1}{n} I(W_1; \overline{\overline{Y}}_1 | W_2) + \delta(n)$$

and

$$R_2 \leq \frac{1}{n} I(W_2; \overline{\overline{Y}}_2) + \delta(n) \tag{8}$$

where $\delta(n) \to 0$ as $n \to \infty$. For the sake of brevity, in the following we ignore $\delta(n)$ (in the rigorous proof given in the following subsections $\delta(n)$ in not ignored).

Therefore, by Fano's inequality (8) and by (7), we may write

$$\frac{1}{n} I(W_1; \overline{\overline{Y}}_1 | W_2)$$
$$= \frac{1}{n} H(\overline{\overline{Y}}_1 | W_2) - \frac{1}{n} H(\overline{\overline{Y}}_1 | W_1, W_2)$$
$$= \frac{1}{n} H(\overline{\overline{Y}}_1 | W_2) - \frac{1}{2} \log(|2\pi e \tilde{\boldsymbol{N}}_1|)$$
$$\geq R_1 \geq R_1^G(\boldsymbol{B}_{1,2}, \tilde{\boldsymbol{N}}_{1,2})$$
$$= \frac{1}{2} \log(|2\pi e(\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1)|) - \frac{1}{2} \log(|2\pi e \tilde{\boldsymbol{N}}_1|)$$

and hence,

$$\frac{1}{n} H(\overline{\overline{Y}}_1 | W_2) \geq \frac{1}{2} \log(|2\pi e(\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1)|).$$

Next, we lower-bound $\frac{1}{n} H(\overline{\overline{Y}}_2 | W_2)$ using the EPI. The EPI (see [10, pp. 496–497]) lower-bounds the entropy of the sum of two independent vectors with a function of the entropies of each of the vectors. As $\overline{\overline{Y}}_2 = \overline{\overline{Y}}_1 + \overline{\overline{Z}}_2$ where $\overline{\overline{Z}}_2$ is independent of $\overline{\overline{Y}}_1$ and W_2, we can use the EPI to bound $\frac{1}{n} H(\overline{\overline{Y}}_2 | W_2)$ in the following manner:

$$\frac{1}{t \cdot n} H(\overline{\overline{Y}}_2 | W_2) = \frac{1}{t \cdot n} H(\overline{\overline{Y}}_1 + \overline{\overline{Z}}_2 | W_2)$$
$$\geq \frac{1}{2} \log \left(e^{\frac{2}{t \cdot n} H(\overline{\overline{Y}}_1 | W_2)} + e^{\frac{2}{t \cdot n} H(\overline{\overline{Z}}_2)} \right)$$

$$\geq \frac{1}{2} \log(|2\pi e(\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1)|^{\frac{1}{t}} + |2\pi e \tilde{\boldsymbol{N}}_2|^{\frac{1}{t}}).$$

As $|\boldsymbol{K}_1|^{\frac{1}{t}} + |\boldsymbol{K}_2|^{\frac{1}{t}} \leq |\boldsymbol{K}_1 + \boldsymbol{K}_2|^{\frac{1}{t}}$ for any $t \times t$ positive semidefinite matrices, $\boldsymbol{K}_1, \boldsymbol{K}_2$ (Minkowski's inequality—see [10, p. 505]), we have

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

with equality in the second inequality, if and only if $\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1$ and $\tilde{\boldsymbol{N}}_2$ are proportional.

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

$$\frac{1}{n} H(\overline{\overline{Y}}_2 | W_2) \geq \frac{1}{2} \log(|2\pi e(\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1 + \tilde{\boldsymbol{N}}_2)|) \tag{9}$$

and by (8), (7), and (9), we can write

$$\frac{1}{n} H(\overline{\overline{Y}}_2) = \frac{1}{n} I(W_2; \overline{\overline{Y}}_2) + \frac{1}{n} H(\overline{\overline{Y}}_2 | W_2)$$
$$\geq R_2 + \frac{1}{n} H(\overline{\overline{Y}}_2 | W_2)$$
$$> R_2^G(\boldsymbol{B}_{1,2}, \tilde{\boldsymbol{N}}_{1,2}) + \frac{1}{n} H(\overline{\overline{Y}}_2 | W_2)$$
$$= \frac{1}{2} \log(|2\pi e(\boldsymbol{S} + \tilde{\boldsymbol{N}}_1 + \tilde{\boldsymbol{N}}_2)|)$$
$$\quad - \frac{1}{2} \log(|2\pi e(\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1 + \tilde{\boldsymbol{N}}_2)|) + \frac{1}{n} H(\overline{\overline{Y}}_2 | W_2)$$
$$\geq \frac{1}{2} \log(|2\pi e(\boldsymbol{S} + \tilde{\boldsymbol{N}}_1 + \tilde{\boldsymbol{N}}_2)|).$$

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

In the scalar case, $\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1$ is always proportional to $\tilde{\boldsymbol{N}}_2$ and, therefore, by Bergmans' proof we see that all points that lie outside the Gaussian rate region cannot be attained. Unfortunately, in the vector case $\boldsymbol{B}_1 + \tilde{\boldsymbol{N}}_1$ and $\tilde{\boldsymbol{N}}_2$ are not necessarily proportional and, therefore, we cannot directly apply Bergmans' proof in the ADBC case.

In order to circumvent this problem, we will introduce, later, 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 $\bar{\boldsymbol{B}}_1$ and $\bar{\boldsymbol{B}}_2 = \boldsymbol{S} - \bar{\boldsymbol{B}}_1$ 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.

B. ADBC—Definitions and Intermediate Results

Before turning to prove the main result of this section, we will first need to give some definitions and intermediate results. We begin with the definition of an optimal Gaussian rate vector and the realizing matrices of that vector.

Definition 3: We say that the rate vector $\bar{R} = (R_1, \ldots, R_m)$ is an *optimal Gaussian rate vector* under a covariance matrix constraint $\boldsymbol{S}$, if $\bar{R} \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ and if there is **no** other rate vector $\bar{R}^* \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ 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 semidefinite matrices $\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_m$ such that $\sum_{i=1}^m \boldsymbol{B}_i \preceq \boldsymbol{S}$, are *realizing matrices of an optimal Gaussian rate vector* if

$$(R_1^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}), R_2^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}), \ldots, \\ R_m^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}))$$

is an optimal Gaussian rate vector.

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

Lemma 3: Let $\bar{R} = (R_1, R_2, \ldots, R_m)$ be a rate vector satisfying

$$\bar{R} \notin \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) \\ R_i \geq 0, \qquad \forall i = 1, \ldots, m-1 \\ R_m > 0.$$

Then, there is a strictly positive scalar $b > 0$, and realizing matrices of an optimal Gaussian rate vector $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$, such that

$$R_i \geq R_i^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}), \qquad i = 1, \ldots, m-1 \\ R_m \geq R_m^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}) + b. \qquad (10)$$

Proof: see Appendix II. $\square$

In general, there is no known closed form solution for the realizing matrices of an optimal Gaussian rate vector. However, we can state the following two simple lemmas:

Lemma 4: Let $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ be realizing matrices of an optimal Gaussian rate vector under a covariance matrix constraint $\boldsymbol{S} \succeq 0$. Then, $\boldsymbol{S} = \sum_{i=1}^m \boldsymbol{B}_i^*$.

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

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

Therefore, as $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ are realizing matrices of an optimal Gaussian rate vector, given the set of $m-1$ matrices, $\bar{\boldsymbol{B}}_i = \boldsymbol{B}_i^* \, i = 1, \ldots, m-1, \bar{\boldsymbol{B}}_m^*$ is the choice of $\boldsymbol{B}_m$ which maximizes $R_m^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$. However, as $\boldsymbol{B}_m \preceq \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i^*$ and as $|\boldsymbol{B}| > |\boldsymbol{A}|$ when $\boldsymbol{B} \succeq \boldsymbol{A} \succ 0$ and $\boldsymbol{B} \neq \boldsymbol{A}$, R_m^G is maximized when $\boldsymbol{B}_m = \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i^*$. $\square$

As a result of the preceding lemma, we can see that the rates of any optimal Gaussian rate vector may be written as a function of

only $m-1$ matrices $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_{m-1}^*$, and the covariance matrix constraint $\boldsymbol{S}$. For this reason, we modify our notation of the rate functions, R_i^G, and write $R_i^G(\boldsymbol{B}_{1,\ldots,m-1}, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ instead of $R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$. The rates are calculated using (5) and assigning $\boldsymbol{B}_m = \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i$. Furthermore, given an optimal Gaussian rate vector

$$\bar{R}^* = (R_1^*, \ldots, R_m^*) \\ = (R_1^G(\boldsymbol{B}_{1,\ldots,m-1}^*, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}), \ldots, \\ R_m^G(\boldsymbol{B}_{1,\ldots,m-1}^*, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}))$$

the realizing matrices of an optimal Gaussian rate vector, $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_{m-1}^*, \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i^*$, can be represented as the solution of the following optimization problem:

$$\begin{aligned} &\text{maximize} && R_m^G(\boldsymbol{B}_{1,\ldots,m-1}, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) \\ &\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{m-1} \\ &\text{such that} && R_i^G(\boldsymbol{B}_{1,\ldots,m-1}, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) - R_i^* \geq 0, \\ &&& \qquad\qquad i = 1, \ldots, m-1 \qquad (11)\\ &&& \boldsymbol{B}_j \succeq 0_{t \times t}, \qquad j = 1, \ldots, m-1 \\ &&& \sum_{l=1}^{m-1} \boldsymbol{B}_l \preceq \boldsymbol{S}. \end{aligned}$$

The following lemma states necessary Karush–Kuhn–Tucker (KKT) conditions on the realizing matrices of an optimal Gaussian rate vector.

Lemma 5: Assume that $\boldsymbol{S} \succ 0_{t \times t}$ (strictly positive) and let $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_{m-1}^*$ solve the problem in (11). In addition, define $\boldsymbol{B}_m^* = \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i^*$ and assume that $\boldsymbol{B}_m^* \neq 0_{t \times t}$. Then, the following necessary KKT conditions hold:

$$0_{t \times t} = \sum_{i=1}^m \left(\gamma_i \cdot \nabla_{\boldsymbol{B}_k} R_i^G(\boldsymbol{B}_{1,\ldots,m-1}^*, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) \right) \\ + \frac{1}{2} \boldsymbol{O}_k - \frac{1}{2} \boldsymbol{O}_m, \qquad \forall k = 1, \ldots, m-1 \quad (12)$$

where $\gamma_i \geq 0, i = 1, \ldots, m-1$, and $\gamma_m = 1$ and where $\boldsymbol{O}_k \succeq 0$ $(k = 1, \ldots, m)$ are positive semidefinite $t \times t$ matrices such that $\mathrm{tr}\{\boldsymbol{B}_k^* \boldsymbol{O}_k\} = 0$. Furthermore, the partial gradients are given by (13) at the top of the following page.

Proof: For every l, the functions R_l^G are differentiable w.r.t. $\boldsymbol{B}_j(\boldsymbol{B}_j \succeq 0)$. Thus, the gradients in (13) are the standard gradients of the log det functions. The KKT necessary conditions that appear in (12) are the standard ones (see [2, Secs. 5.1–5.4, pp. 270–312] and [4, pp. 241–248]). However, as the problem is not convex, we need to show that some set of constraint qualifications (CQs) hold in order to prove the existence of Lagrange multipliers (see [2, Sec. 5.4, pp. 302–312,]). Hence, we demanded here that $\boldsymbol{B}_m^* = \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i^* \neq 0_{t \times t}$ in order to satisfy the CQs. As the proof that these CQs are satisfied is rather technical, we defer the rest of the proof to Appendix IV. $\square$

Next, we introduce a class of enhanced channels. This class of channels has at least one element that has several surprising and fundamental properties derived below. These properties form the basis of our capacity results.

$$\nabla_{\boldsymbol{B}_j} R_l^G(\boldsymbol{B}_{1,\ldots,m-1}, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) = \begin{cases} \frac{1}{2}\left(\sum_{i=1}^{l}(\boldsymbol{B}_i + \tilde{\boldsymbol{N}}_i)\right)^{-1} - \frac{1}{2}\left(\sum_{i=1}^{l-1}\boldsymbol{B}_i + \sum_{i=1}^{l}\tilde{\boldsymbol{N}}_i\right)^{-1}, & 1 \le j < l < m \\ \frac{1}{2}\left(\sum_{i=1}^{l}(\boldsymbol{B}_i + \tilde{\boldsymbol{N}}_i)\right)^{-1}, & 1 \le j = l < m \\ 0, & 1 \le l < j < m \\ -\frac{1}{2}\left(\sum_{i=1}^{l-1}\boldsymbol{B}_i + \sum_{i=1}^{l}\tilde{\boldsymbol{N}}_i\right)^{-1}, & 1 \le j < m, l = m. \end{cases} \quad (13)$$

Definition 4: (Enhanced Channel) We say that an ADBC with noise increment covariance matrices $(\tilde{\boldsymbol{N}}'_1, \tilde{\boldsymbol{N}}'_2, \ldots, \tilde{\boldsymbol{N}}'_m)$ is an enhanced version of another ADBC with noise increment covariance matrices $(\tilde{\boldsymbol{N}}_1, \tilde{\boldsymbol{N}}_2, \ldots, \tilde{\boldsymbol{N}}_m)$ if

$$\sum_{j=1}^{i} \tilde{\boldsymbol{N}}'_j \preceq \sum_{j=1}^{i} \tilde{\boldsymbol{N}}_j, \qquad \forall i = 1, \ldots, m.$$

Similarly, we say that an AMBC with noise covariance matrices $(\boldsymbol{N}'_1, \boldsymbol{N}'_2, \ldots, \boldsymbol{N}'_m)$ is an enhanced version of another AMBC with noise covariance matrices $(\boldsymbol{N}_1, \boldsymbol{N}_2, \ldots, \boldsymbol{N}_m)$ if

$$\boldsymbol{N}'_i \preceq \boldsymbol{N}_i, \qquad \forall i = 1, \ldots, m.$$

Clearly, the capacity region of the original channel is contained within that of the enhanced channel. Furthermore, as

$$\frac{|\boldsymbol{A} + \boldsymbol{B} + \Delta|}{|\boldsymbol{B} + \Delta|} \le \frac{|\boldsymbol{A} + \boldsymbol{B}|}{|\boldsymbol{B}|}$$

when $\boldsymbol{A} \succeq 0$, $\Delta \succeq 0$, and $\boldsymbol{B} \succ 0$,

$$R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) \le R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}'_{1,\ldots,m}), \quad \forall i$$

and therefore,

$$\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) \subseteq \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}'_{1,\ldots,m}).$$

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

Theorem 1: Consider an ADBC with positive semidefinite noise increment covariance matrices $(\tilde{\boldsymbol{N}}_1, \ldots, \tilde{\boldsymbol{N}}_m)$ such that $\tilde{\boldsymbol{N}}_1 \succ 0$. Let $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ be realizing matrices of an optimal Gaussian rate vector under an average transmit covariance matrix constraint $\boldsymbol{S} \succ 0$. Then, there exists an enhanced ADBC with noise increment covariances $(\tilde{\boldsymbol{N}}'_1, \ldots, \tilde{\boldsymbol{N}}'_m)$ such that the following properties hold.

1) *Enhanced Channel:*

$$0 \prec \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \preceq \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i, \qquad \forall k = 1, \ldots, m. \quad (14)$$

2) *Proportionality:* There exist $\alpha_k \ge 0$, $k = 1, \ldots, m-1$ such that

$$\alpha_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i\right) = \tilde{\boldsymbol{N}}'_{k+1}, \quad \forall k = 1, \ldots, m-1. \quad (15)$$

3) *Rate Preservation and Optimality Preservation:*

$$R_k^G\left(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}\right) = R_k^G\left(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}'_{1,\ldots,m}\right),$$
$$\forall k = 1, \ldots, m \quad (16)$$

and $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ are realizing matrices of an optimal Gaussian rate vector in the enhanced channel as well.

Proof: Assume that $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ are realizing matrices of an optimal Gaussian rate vector under a strictly positive-definite covariance matrix constraint $\boldsymbol{S} \succ 0$. If $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ are all strictly positive ($\boldsymbol{B}_i^* \succ 0$), by the KKT conditions in Lemma 5 we know that $\boldsymbol{O}_i = 0_{t \times t} \forall i$ (as $\text{tr}\{\boldsymbol{B}_i^* \boldsymbol{O}_i\} = 0$) and by manipulating (11) it is possible to show that the proportionality property already holds for our original channel and hence the proof of the theorem is almost trivial for this case.

Loosely speaking, for the more general case where $\boldsymbol{B}_i^* \succeq 0$, we can create an enhanced channel by subtracting a matrix $\Delta_i \succeq 0$ from the noise covariance matrix of user i such that the new noise covariance matrix is given by $\boldsymbol{N}'_i = (\sum_{j=1}^{i} \tilde{\boldsymbol{N}}_j) - \Delta_i$. However, Δ_i is chosen such that

$$\Delta_i \cdot \left(\sum_{j=1}^{i-1} \boldsymbol{B}_j^* + \sum_{j=1}^{i} \tilde{\boldsymbol{N}}_j\right)^{-1} \boldsymbol{B}_i^* = 0_{t \times t}.$$

This choice of Δ_i ensures that the noise covariance of user i is modified only in those directions where, effectively, no information is being sent to that user and hence, using the same power allocation matrices $\boldsymbol{B}_i^*$, the rate of user i remains the same in the enhanced channel as in the original channel. This process is repeated for all users. Using Lemma 5, 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 give a rigorous proof of this argument.

We divide our proof into two. Initially, we will assume that $\boldsymbol{B}_i^* \neq 0_{t \times t} \forall i = 1, \ldots, m$ (note that the matrices $\boldsymbol{B}_i^*$ 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, (5)). Later, we give a simple argument that will allow us to expand the proof to any possible optimal Gaussian rate vector.

$\underline{\boldsymbol{B}_i^* \neq 0_{t \times t}, \forall i = 1, \ldots, m}$:

As $\boldsymbol{B}_m^* \neq 0_{t \times t}$, Lemma 5 holds. Therefore, by plugging the expression for the partial gradients (13) into the $m-1$ equations in (12) and subtracting equation $k+1$ from the kth equation

(except for $k = m - 1$ where the expression is taken as is), we obtain the following $m - 1$ equations:

$$
\gamma_{k+1} \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}_i \right)^{-1} + \boldsymbol{O}_{k+1}
$$
$$
= \gamma_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \right)^{-1} + \boldsymbol{O}_k,
$$
$$
k = 1, \ldots, m - 1 \quad (17)
$$

We now use the assumption that $\bar{\boldsymbol{B}}_i^* \neq 0$, $i = 1, \ldots, m$ to show that γ_i are strictly positive and that $0 < \frac{\gamma_k}{\gamma_{k+1}} \leq 1 \ \forall k = 1, \ldots, m - 1$. As $\gamma_m = 1 > 0$, we can see that for $k = m - 1$, the left-hand side of (17) is strictly positive definite (recall that in the Introduction we assumed that $\boldsymbol{N}_k = \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \succ 0 \forall k$ in order to limit ourselves to the case where all users have finite rates). Furthermore, as $\boldsymbol{B}_{m-1}^* \neq 0$, $\mathrm{rank}(\boldsymbol{O}_{m-1}) < t$, as otherwise, we would have had $\mathrm{tr}\{\boldsymbol{B}_{m-1}^* \boldsymbol{O}_{m-1}\} > 0$ (in contradiction to Lemma 5). Therefore, the matrix $\boldsymbol{O}_{m-1}$ must have a zero eigenvalue and hence, γ_{m-1} must be strictly positive, as otherwise the right-hand side and hence the left-hand side of (17) would have had a zero eigenvalue (and the left-hand side would not be strictly positive). We can repeat the above arguments for all $k = 1, \ldots, m - 1$ in expression (17) and conclude that $\gamma_i > 0 \ \forall i = 1, \ldots, m - 1$.

Next, let $\boldsymbol{z}$ be an eigenvector corresponding to a zero eigenvalue of $\boldsymbol{O}_{k+1}$ (which has some zero eigenvalues because we assume that $\bar{\boldsymbol{B}}_i^* \neq 0 \forall i = 1, \ldots, m$). We have

$$
\boldsymbol{z}^T \left(\gamma_{k+1} \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}_i \right)^{-1} + \boldsymbol{O}_{k+1} \right) \boldsymbol{z}
$$
$$
= \gamma_{k+1} \boldsymbol{z}^T \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}_i \right)^{-1} \boldsymbol{z}
$$
$$
= \boldsymbol{z}^T \left(\gamma_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \right)^{-1} + \boldsymbol{O}_k \right) \boldsymbol{z} \quad (18)
$$

But as $\boldsymbol{B}^{-1} \preceq \boldsymbol{A}^{-1}$ when $\boldsymbol{B} \succeq \boldsymbol{A} \succ 0$, we know that

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

This means that (18) can hold only if

$$
0 < \frac{\gamma_k}{\gamma_{k+1}} \leq 1, \qquad \forall k = 1, \ldots, m - 1.
$$

As both $\boldsymbol{B}_i^*$ and $\boldsymbol{O}_i$ are positive semidefinite and as $\mathrm{tr}\{\bar{\boldsymbol{B}}_i^* \boldsymbol{O}_i\} = 0$ (Lemma 5), we must have $\bar{\boldsymbol{B}}_i^* \cdot \boldsymbol{O}_i = 0_{t \times t}$. Taking into account the fact that γ_i in (17) are strictly positive such that $0 < \frac{\gamma_k}{\gamma_{k+1}} \leq 1 \forall k = 1, \ldots, m - 1$ and that

$\boldsymbol{B}_i^* \cdot \boldsymbol{O}_i = 0_{t \times t}$, we can use Lemma 10 in Appendix V to show that there exists an enhanced channel with noise increment covariance matrices $\tilde{\boldsymbol{N}}_1', \ldots, \tilde{\boldsymbol{N}}_m'$ such that

$$
\gamma_{k+1} \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}_i \right)^{-1} + \boldsymbol{O}_{k+1}
$$
$$
= \gamma_{k+1} \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}_i' \right)^{-1}
$$
$$
= \gamma_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i' \right)^{-1}
$$
$$
= \gamma_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \right)^{-1} + \boldsymbol{O}_k \quad (19)
$$

and such that

$$
\frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \right|}{\left| \sum_{i=1}^{k-1} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \right|} = \frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i' \right|}{\left| \sum_{i=1}^{k-1} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i' \right|},
$$
$$
\forall k = 1, \ldots, m \quad (20)
$$

and where $\boldsymbol{N}_i' = \sum_{j=1}^{i} \tilde{\boldsymbol{N}}_j' \preceq \sum_{j=1}^{i} \tilde{\boldsymbol{N}}_j = \boldsymbol{N}_i \ \forall i$.

By (19) we can write

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

and, therefore, the proportionality property holds for the enhanced channel. By (20), we may write

$$
R_k^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}) = \frac{1}{2} \log \frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \right|}{\left| \sum_{i=1}^{k-1} \bar{\boldsymbol{B}}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i \right|}
$$
$$
= \frac{1}{2} \log \frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i' \right|}{\left| \sum_{i=1}^{k-1} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i' \right|}
$$
$$
= R_k^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}'), \quad \forall k = 1, \ldots, m
$$

and, therefore, the rate preservation property holds.

To complete the proof for the case where $\boldsymbol{B}_i^* \neq 0 \forall i$, we still need to show that $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ are also realizing matrices of an optimal Gaussian rate vector in the enhanced channel. For that purpose, we observe that it is sufficient to show that

$$
R_m^* = \sup \left\{ R_m \in \mathbb{R} \mid (R_1^*, \ldots, R_{m-1}^*, R_m) \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}') \right\}. \quad (21)
$$

To show that this is indeed a sufficient condition for optimality, we note that if there were another vector $(R_1, \ldots, R_{m-1}, R_m) \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}')$ such that $R_i \geq R_i^* \forall i = 1, \ldots, m$ where the inequality was strict for at least

one of the elements, then by Lemma 8 in Appendix II we would be able to find a rate vector

$$(R_1^*, \ldots, R_{m-1}^*, R_m^* + \epsilon) \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}')$$

for some $\epsilon > 0$, contradicting (21). We proceed to show that indeed (21) holds. For any set of matrices $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_m$ such that $\sum_{i=1}^m \boldsymbol{B}_i \preceq \boldsymbol{S}$ we have

$$
\begin{aligned}
R_m^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}') &= \frac{1}{2} \log \frac{\left| \sum_{i=1}^m \boldsymbol{B}_i + \sum_{i=1}^m \tilde{\boldsymbol{N}}_i' \right|}{\left| \sum_{i=1}^{m-1} \boldsymbol{B}_i + \sum_{i=1}^m \tilde{\boldsymbol{N}}_i' \right|} \\
&\leq \frac{1}{2} \log \frac{\left| \boldsymbol{S} + \sum_{i=1}^m \tilde{\boldsymbol{N}}_i' \right|}{\left| \sum_{i=1}^{m-1} \boldsymbol{B}_i + \sum_{i=1}^m \tilde{\boldsymbol{N}}_i' \right|}.
\end{aligned}
$$

Hence, as $\sum_{i=1}^m \boldsymbol{B}_i^* = \boldsymbol{S}$, it is enough to show that given $R_1^*, \ldots, R_{m-1}^*$, the set of realizing matrices $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ minimizes $|\sum_{i=1}^{m-1} \boldsymbol{B}_i + \sum_{i=1}^m \tilde{\boldsymbol{N}}_i'|$ over all sequences of semidefinite matrices $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_m$ such that $\sum_{i=1}^{m-1} \boldsymbol{B}_i \preceq \boldsymbol{S}$ and such that $R_k^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}') = R_k^* \, \forall k = 1, \ldots, m-1$. As $|\boldsymbol{A} + \boldsymbol{B}|^{\frac{1}{t}} \geq |\boldsymbol{A}|^{\frac{1}{t}} + |\boldsymbol{B}|^{\frac{1}{t}}$ for positive semidefinite $t \times t$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ (Minkowski's inequality [10, p. 505]), we may write

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

where the inequalities hold with equality if and only if $\tilde{\boldsymbol{N}}_{k+1}'$ is proportional to $(\sum_{i=1}^k \boldsymbol{B}_i + \sum_{i=1}^k \tilde{\boldsymbol{N}}_i')$ for all $k = 1, \ldots,$

$m-1$. However, we have shown that this proportionality property holds for $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$. Therefore, if we assign $\boldsymbol{B}_i^*$ instead of $\boldsymbol{B}_i$, we obtain the lower bound given by the last line in the preceding equation. Thus, we have shown that indeed $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$ minimize $|\sum_{i=1}^{m-1} \boldsymbol{B}_i + \sum_{i=1}^m \tilde{\boldsymbol{N}}_i'|$ and, therefore, are realizing matrices of an optimal Gaussian rate vector in the enhanced channel.

$\underline{\boldsymbol{B}_i^* \succeq 0, \, \forall i = 1, \ldots, m}$:

Finally, we expand the proof to all possible sets of realizing matrices of an optimal Gaussian rate vector $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$. Note that some of the matrices $\boldsymbol{B}_i^*$ may be zero. Let $n < m$ denote the number of nonzero matrices $\boldsymbol{B}_i^*$, and let $k(i)$ for $i = 1, \ldots, n$ be the index function of the nonzero matrices such that $\boldsymbol{B}_{k(i)}^* \neq 0_{t \times t} \forall i = 1, \ldots, n$ and such that $k(i+1) > k(i), \forall i = 1 \ldots n-1$. We can define a compact channel which is also an ADBC with the same covariance matrix constraint $\boldsymbol{S}$ and with noise increment covariance matrices $\hat{\boldsymbol{N}}_1, \ldots, \hat{\boldsymbol{N}}_n$ such that

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

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

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

Note that $\hat{\boldsymbol{B}}_1, \ldots, \hat{\boldsymbol{B}}_n$ are realizing matrices of an optimal Gaussian rate vector in the compact channel and achieve the same rates (nonzero rates) as in the original channel. Since $\hat{\boldsymbol{B}}_i \neq 0, i = 1, \ldots, n$, and $\boldsymbol{S} \succ 0$, we can use the above proof to show that Theorem 1 holds for the compact channel.

We can now define an enhanced version of the original channel using the construction implied by Theorem 1 for the case of nonzero $\boldsymbol{B}_i^*$. Since Theorem 1 holds for the compact channel, we know that there exists an enhanced compact channel $\hat{\boldsymbol{N}}_1', \ldots, \hat{\boldsymbol{N}}_n'$ for which the results of the theorem hold. We now define an enhanced version $\tilde{\boldsymbol{N}}_1', \ldots, \tilde{\boldsymbol{N}}_m'$ of the original channel as follows:

$$\tilde{\boldsymbol{N}}_i' = \begin{cases} \beta_i \hat{\boldsymbol{N}}_1', & i \leq k(1) \\ \hat{\boldsymbol{N}}_{k^{-1}(i)}', & \forall i \in \{k(2), \ldots, k(n)\} \\ 0, & \text{otherwise} \end{cases} \qquad (22)$$

where $k(k^{-1}(i)) = i$ (if $i \in \{k(j), j = 1, \ldots, n\}$) and where $\beta_i \geq 0 \, \forall i = 1, \ldots, k(1)$ are chosen such that $\sum_{i=1}^{k(1)} \beta_i = 1$ and such that

$$0 \prec \left(\sum_{i=1}^j \beta_i \right) \cdot \hat{\boldsymbol{N}}_1' \preceq \sum_{i=1}^j \tilde{\boldsymbol{N}}_i, \qquad \forall j = 1, \ldots, k(1) - 1.$$

As $\sum_{i=1}^j \tilde{\boldsymbol{N}}_i \succ 0 \forall j = 1, \ldots, m$, it is possible to find such β_i.

To verify that this is an enhanced version of the original channel we need to show that $\sum_{j=1}^i \tilde{\boldsymbol{N}}_j' \preceq \sum_{j=1}^i \tilde{\boldsymbol{N}}_j$, $\forall i = 1, \ldots, m$. For the case of $1 \leq i < k(1)$, we observe that by the definition of β_i this equality holds. For the case of $k(1) \leq i \leq m$, we define $d = \max\{l = 1, \ldots, n | k(l) \leq i\}$.

Due to (22) and the definitions of the compact and the enhanced compact channels we can now write

$$\sum_{j=1}^{i} \tilde{\boldsymbol{N}}'_j = \sum_{j=1}^{d} \hat{\boldsymbol{N}}'_j \preceq \sum_{j=1}^{d} \hat{\boldsymbol{N}}_j = \sum_{j=1}^{k(d)} \tilde{\boldsymbol{N}}_j \preceq \sum_{j=1}^{i} \tilde{\boldsymbol{N}}_j,$$
$$\forall i = \{k(1), \ldots, m\}.$$

To verify that proportionality property holds for the channel defined in (22), we need to show that

$$\alpha_i \left(\sum_{j=1}^{i} \boldsymbol{B}_j + \sum_{j=1}^{i} \tilde{\boldsymbol{N}}'_j \right) = \tilde{\boldsymbol{N}}'_{i+1} \qquad \forall i = 1, \ldots, m-1$$

for some nonnegative α_i's. We consider three cases.

1) If $i < k(1)$ then, as $\boldsymbol{B}_i^* = 0 \ \forall i < k(1)$ and by (22) we can write

$$\alpha_i \left(\sum_{j=1}^{i} \boldsymbol{B}_j + \sum_{j=1}^{i} \tilde{\boldsymbol{N}}'_j \right) = \alpha_i \left(\sum_{j=1}^{i} \tilde{\boldsymbol{N}}'_j \right)$$
$$= \alpha_i \left(\sum_{j=1}^{i} \beta_j \right) \hat{\boldsymbol{N}}'_1$$
$$= \beta_{i+1} \hat{\boldsymbol{N}}'_1 = \tilde{\boldsymbol{N}}'_{i+1}$$

where $\alpha_i = \frac{\beta_{i+1}}{\sum_{j=1}^{i} \beta_j}$.

2) If $i \geq k(1)$ and $(i+1) \notin \{k(1), \ldots, k(n)\}$ then, by (22)

$$0 \cdot \left(\sum_{j=1}^{i} \boldsymbol{B}_j + \sum_{j=1}^{i} \tilde{\boldsymbol{N}}'_j \right) = 0_{t \times t} = \tilde{\boldsymbol{N}}'_{i+1}.$$

3) If $i \geq k(1)$ and $i+1 \in \{k(1), \ldots, k(n)\}$ then, as $\boldsymbol{B}_i^* = 0 \ \forall i \notin \{k(1), \ldots, k(n)\}$ and by (22) we can write

$$\alpha_i \left(\sum_{j=1}^{i} \boldsymbol{B}_j + \sum_{j=1}^{i} \tilde{\boldsymbol{N}}'_j \right) = \alpha_i \left(\sum_{j=1}^{k(d)} \boldsymbol{B}_j + \sum_{j=1}^{k(d)} \tilde{\boldsymbol{N}}'_j \right)$$
$$= \alpha_i \left(\sum_{j=1}^{d} \hat{\boldsymbol{B}}_j + \sum_{j=1}^{d} \hat{\boldsymbol{N}}'_j \right)$$
$$= \hat{\boldsymbol{N}}'_{d+1} = \tilde{\boldsymbol{N}}'_{i+1}$$

for some $\alpha_i \geq 0$.

Finally, we note that the optimality of the rates in the enhanced channel is a result of the proportionality property (as was shown for the case of $\boldsymbol{B}_i^* \neq 0$ using the Minkowski inequality) and the rate preservation property holds for this version since it holds for the enhanced compact channel and since for user i with $\boldsymbol{B}_i^* = 0$ the Gaussian rate is always zero (regardless of the user's noise matrix). $\qquad \square$

C. ADBC—Main Result

We can now use Lemma 3 and Theorem 1 that were presented in the previous subsection to prove that $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ is the capacity region. 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 III-A) by applying his proof to the enhanced channel instead of the original channel, and utilizing the proportionality property which holds for that channel. Before we turn to the proof, we formally state the main result.

Theorem 2: Let $\mathcal{C}(\bar{\boldsymbol{S}}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ denote the capacity region of the ADBC under a covariance matrix constraint $\boldsymbol{S} \succeq 0$. Then $\mathcal{C}(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) = \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$.

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

$\underline{\boldsymbol{S} \succ 0}$:

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

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

Since $\bar{R} \notin \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ and $R_m > 0$ (by our assumption $R_i > 0$. The case of $R_m \geq 0$ may include the case of $R_m = 0$ and therefore it is treated separately), we know by Lemma 3 that there exist realizing matrices of an optimal Gaussian rate vector $\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_m^*$, such that

$$R_1 \geq R_1^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}), \qquad i = 1, \ldots, m-1$$
$$R_m \geq R_m^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}) + b$$

for some $b > 0$. Since we assume that $\boldsymbol{S} \succ 0$, we know by Theorem 1 that for every set of realizing matrices of an optimal Gaussian rate vector $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_m$, there exists an enhanced ADBC, $\tilde{\boldsymbol{N}}'_1, \ldots, \tilde{\boldsymbol{N}}'_m$, such that the proportionality and rate preservation properties hold. By the rate preservation property, we have $R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) = R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$. Therefore, we can rewrite the precediing expression as follows:

$$R_1 \geq R_1^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}) = R_1^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$$
$$\vdots$$
$$R_{m-1} \geq R_{m-1}^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}) = R_{m-1}^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$$
$$R_m \geq R_m^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}_{1,\ldots,m}) + b = R_m^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}'_{1,\ldots,m}) + b.$$
$$(23)$$

Let $W_i (i = 1, \ldots, m)$ denote the index of the message sent to user i and let $\overline{\overline{Y}}_i$ 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 codebook of length-n codewords and for which

$$R_i \leq \frac{1}{n} I(W_i; \overline{\overline{Y}}_i) + \frac{1}{2 \cdot m} b$$

$$\leq \frac{1}{n}I(W_i; \overline{\overline{Y}}_i | W_{i+1}, \ldots, W_m)$$
$$+ \frac{1}{2 \cdot m} b, \; i = 1, \ldots, m-1$$
$$R_m \leq \frac{1}{n}I(W_m; \overline{\overline{Y}}_m) + \frac{1}{2 \cdot m} b \qquad (24)$$

Let $\overline{\overline{Y'}}_i, \; i = 1, \ldots m$ denote the enhanced channel outputs of each of the receiving users. As $(W_i, \ldots, W_m) \to \overline{\overline{Y'}}_i \to \overline{\overline{Y}}_i$ form a Markov chain, we can use the data processing theorem to rewrite (24) as follows:

$$R_i \leq \frac{1}{n}I(W_i; \overline{\overline{Y'}}_i | W_{i+1}, \ldots, W_m) + \frac{1}{2 \cdot m}b,$$
$$i = 1, \ldots, m-1$$
$$R_m \leq \frac{1}{n}I(W_m; \overline{\overline{Y'}}_m) + \frac{1}{2 \cdot m}b. \qquad (25)$$

Thus, in (23) and (25), 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. Next, we basically repeat Bergmans' steps, as they were presented in [1] and Subsection III-A.

By (25) and (23) we can write

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

Thus,

$$\frac{1}{n}H(\overline{\overline{Y'}}_1 | W_2, \ldots, W_m)$$
$$\geq \frac{1}{2}\log\left(\left|2\pi e \left(B^*_1 + \tilde{N}'_1\right)\right|\right) - \frac{1}{2 \cdot m}b. \quad (26)$$

We may write $\overline{\overline{Y'}}_2 = \overline{\overline{Y'}}_1 + \overline{\overline{Z'}}_2$ where $\overline{\overline{Z'}}_2$ is a random Gaussian matrix with independent columns and independent of both $\overline{\overline{Y'}}_1$ and the messages $(W_1, \ldots, W_m)$. Each column in $\overline{\overline{Z'}}_2$ has Normal distribution with zero mean and a covariance matrix $\tilde{N}'_2$. Next, we use the EPI to lower-bound the entropy of the sum of two independent vectors with a function of the entropies of each of the vectors as follows:

$$\frac{1}{t \cdot n}H(\overline{\overline{Y'}}_2 | W_2, \ldots, W_m)$$
$$= \frac{1}{t \cdot n}H(\overline{\overline{Y'}}_1 + \overline{\overline{Z'}}_2 | W_2, \ldots, W_m)$$
$$\geq \frac{1}{2}\log\left(e^{\frac{2}{t \cdot n}H(\overline{\overline{Y'}}_1 | W_2, \ldots, W_m)} + e^{\frac{2}{t \cdot n}H(\overline{\overline{Z'}}_2)}\right)$$
$$\geq \frac{1}{2}\log\left(\left|2\pi e\left(B^*_1 + \tilde{N}'_1\right)\right|^{\frac{1}{t}}e^{-\frac{2}{2 \cdot m \cdot t}b} + |2\pi e \tilde{N}'_2|^{\frac{1}{t}}\right)$$

$$\geq \frac{1}{2}\log\left(\left|2\pi e\left(B^*_1 + \tilde{N}'_1\right)\right|^{\frac{1}{t}} + |2\pi e \tilde{N}'_2|^{\frac{1}{t}}\right)$$
$$- \frac{1}{2 \cdot m \cdot t}b. \qquad (27)$$

But by the proportionality property in Theorem 1, 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

$$\frac{1}{n}H(\overline{\overline{Y'}}_2 | W_2, \ldots, W_m)$$
$$\geq \frac{1}{2}\log\left(\left|2\pi e\left(B^*_1 + \tilde{N}'_1 + \tilde{N}'_2\right)\right|\right) - \frac{1}{2 \cdot m}b. \quad (28)$$

Again, we can use (25) and (23) and write

$$\frac{1}{n}I(W_2; \overline{\overline{Y'}}_2 | W_3, \ldots, W_m) + \frac{1}{2 \cdot m}b$$
$$= \frac{1}{n}H(\overline{\overline{Y'}}_2 | W_3, \ldots, W_m)$$
$$\quad - \frac{1}{n}H(\overline{\overline{Y'}}_2 | W_2, \ldots, W_m) + \frac{1}{2 \cdot m}b$$
$$\geq R_2^G(B^*_{1,\ldots,m-1}, S, \tilde{N}'_{1,\ldots,m})$$
$$= \frac{1}{2}\log\left(\left|2\pi e\left(B^*_1 + B^*_2 + \tilde{N}'_1 + \tilde{N}'_2\right)\right|\right)$$
$$\quad - \frac{1}{2}\log\left(\left|2\pi e\left(B^*_1 + \tilde{N}'_1 + \tilde{N}'_2\right)\right|\right).$$

Combining the expression above and (28) we get

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

We continue to calculate $\frac{1}{n}H(\overline{\overline{Y'}}_i | W_{i+1}, \ldots, W_m)$ for $i > 2$ by using the above argumentation and by alternating between the EPI and Fano's inequality. Thus, for the mth iteration we get

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

where the equality follows from the fact that $\sum_{i=1}^m B^*_i = S$ (Lemma 4). However, the above expression cannot hold due to the upper bound on the entropy of a covariance limited random vector, as follows:

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

$$\leq \frac{1}{2} \log \left(\left| 2\pi e \cdot \frac{1}{n} E \left[\overline{\overline{Y'}}_m \overline{\overline{Y'}}_m^T \right] \right| \right)$$

$$\leq \frac{1}{2} \log \left(\left| 2\pi e \cdot \left(\boldsymbol{S} + \sum_{i=1}^{m} \tilde{\boldsymbol{N}}_i' \right) \right| \right)$$

where $\overline{\overline{Y'}}_{m,i}$ is the ith column of the random matrix $\overline{\overline{Y'}}_m$. 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 $\frac{1}{n} E[\overline{\overline{Y'}}_m \overline{\overline{Y'}}_m^T] \preceq \boldsymbol{S} + \sum_{i=1}^{m} \tilde{\boldsymbol{N}}_i'$ and the fact that $|\boldsymbol{B}| \geq |\boldsymbol{A}|$ if $\boldsymbol{B} \succeq \boldsymbol{A} \succ 0$.

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

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

Assume that $\bar{R} \notin \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ 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 $\bar{R}$ 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 $\bar{R}^C = (R_{k(1)}, \ldots, R_{k(n)})$. Similarly, we define a compact n-user ADBC, with noise increment covariance matrices

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

(where we assign $k(0) = 0$) and the same covariance matrix constraint. Clearly, as $\bar{R}$ was achievable in the original ADBC, so is the compact rate vector $\bar{R}^C$ achievable in the compact ADBC. Furthermore, since $\bar{R} \notin \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$, then also $\bar{R}^C \notin \mathcal{R}^G(\boldsymbol{S}, \hat{\tilde{\boldsymbol{N}}}_{1,\ldots,n})$. 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 $\boldsymbol{S}$ is such that $\boldsymbol{S} \succeq 0$ and $|\boldsymbol{S}| = 0$.

$\underline{\boldsymbol{S} \succeq 0, \ |\boldsymbol{S}| = 0}$:

If $\boldsymbol{S}$ is not (strictly) positive definite, by Lemma 2 we know that there exists an equivalent ADBC with less transmit antennas, noise increment covariance matrices $\hat{\tilde{\boldsymbol{N}}}_1, \ldots, \hat{\tilde{\boldsymbol{N}}}_m$, and an input covariance matrix constraint, $\hat{\boldsymbol{S}} \succ 0$, with the exact same capacity region. Because $\hat{\boldsymbol{S}}$ 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.,

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

Moreover, it is possible to show that when $\boldsymbol{S}$ is not strictly positive definite

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

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

Next, in the following corollary we extend the result of Theorem 2 to the case of the total power constraint.

Corollary 2: Let $\mathcal{C}(P, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ denote the capacity region of the ADBC under a total power constraint $P \geq 0$. Then

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

Proof: As $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ is contiguous w.r.t. $\boldsymbol{S}$, the corollary follows immediately by Lemma 1. $\qquad \square$

D. An ADBC Example

The following example illustrates the result stated in Theorem 1. 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

$$\tilde{\boldsymbol{N}}_1 = \begin{pmatrix} 0.5 & 0.18 \\ 0.18 & 0.7 \end{pmatrix}$$

$$\tilde{\boldsymbol{N}}_2 = \begin{pmatrix} 0.2 & -0.1 \\ -0.1 & 10 \end{pmatrix}$$

and

$$\boldsymbol{S} = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 2 \end{pmatrix}.$$

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

The boundary of $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,2}')$, 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

$$\boldsymbol{B}_1' \approx \begin{pmatrix} 0.00004 & -0.00217 \\ -0.00217 & 0.12353 \end{pmatrix}$$

and

$$\boldsymbol{B}_2' = \boldsymbol{S} - \boldsymbol{B}_1'.$$

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

$$\tilde{\boldsymbol{N}}_1' \approx \begin{pmatrix} 0.04896 & 0.00762 \\ 0.00762 & 0.63412 \end{pmatrix}$$

and

$$\tilde{\boldsymbol{N}}_2' = \tilde{\boldsymbol{N}}_1 + \tilde{\boldsymbol{N}}_2 - \tilde{\boldsymbol{N}}_1'.$$

As predicted by Theorem 1, we can see that for the point $(R_1', R_2') \approx (0.08992, 0.51314)$ on the boundary of $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,2})$, there exists an enhanced version of the original
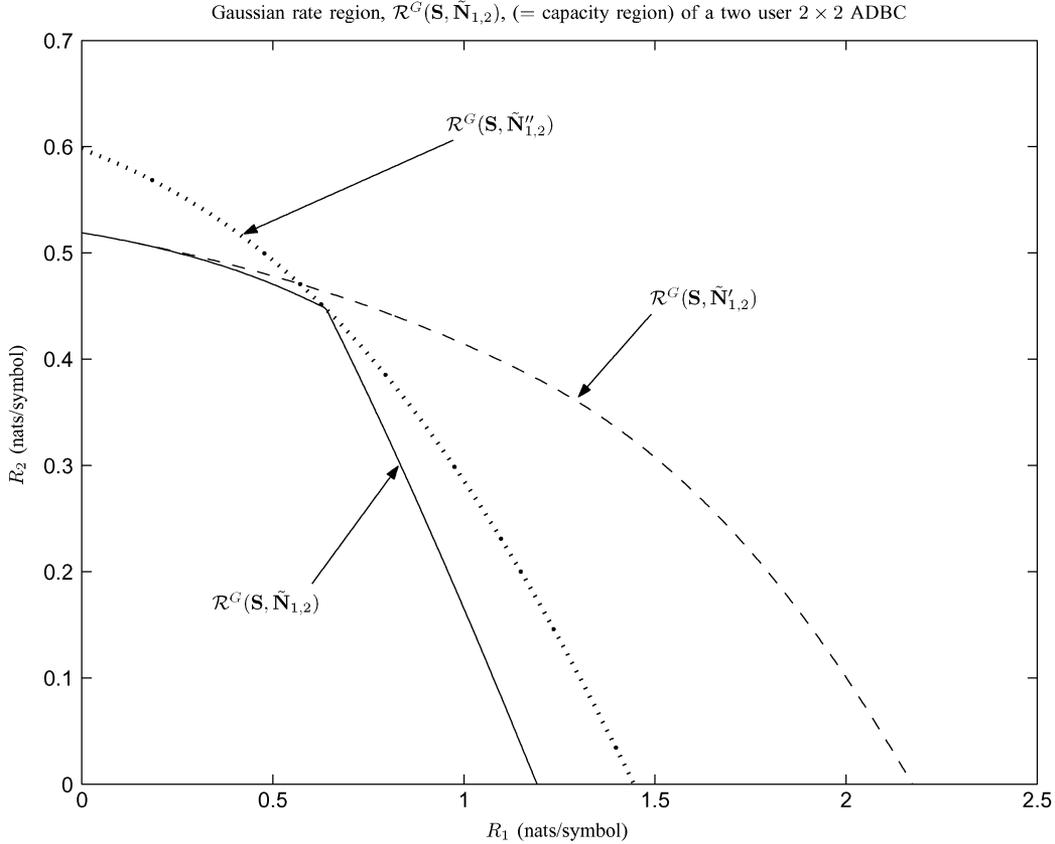
Fig. 1.   Illustration of the results of Theorem 1.

channel which is also an ADBC and such that the boundary of its Gaussian region is tangential to that 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 $R_1 \leq R_1'$ and upper bounds the solid line for all $R_1 > R_1'$. Furthermore, one can easily check that the proportionality property holds such that $(\boldsymbol{B}_1' + \tilde{\boldsymbol{N}}_1') \propto \tilde{\boldsymbol{N}}_2'$.

The second curve, the boundary of $\mathcal{R}^G(\mathbf{S}, \tilde{\boldsymbol{N}}_{1,2}'')$ 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

$$\boldsymbol{B}_1'' \approx \begin{pmatrix} 0.0000024 & -0.00196 \\ -0.00196 & 1.63764 \end{pmatrix}$$

and

$$\boldsymbol{B}_2'' = \boldsymbol{S} - \boldsymbol{B}_1''.$$

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

$$\tilde{\boldsymbol{N}}_1'' \approx \begin{pmatrix} 0.25033 & 0.08974 \\ 0.08974 & 0.66737 \end{pmatrix}$$

and

$$\tilde{\boldsymbol{N}}_2'' \approx \begin{pmatrix} 0.44504 & 0.15605 \\ 0.15605 & 4.09780 \end{pmatrix}.$$

Again, we see that the prediction of Theorem 1 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 $(\boldsymbol{B}_1'' + \tilde{\boldsymbol{N}}_1'') \propto \tilde{\boldsymbol{N}}_2''$.

IV. THE CAPACITY REGION OF THE AMBC

In this section, we build on Theorem 2 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 cannot work when the channel is not degraded. Therefore, following the work of Caire and Shamai [7], we suggest an achievable rate region based on DPC. In [7], [20], [27], [22], 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 [19], [21], the authors used the DSM bound to show that if Gaussian coding is optimal for the vector degraded BC (where we do not necessarily have $r_i = t$ and $\boldsymbol{H}_i = \boldsymbol{I}$), 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 extended 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 [24], we used the

results of [19], [21], and the result presented in the previous section to prove the converse of the GMBC. However, as the tightness of the DSM bound was established only under a total power constraint $E[\boldsymbol{x}^T\boldsymbol{x}] \leq P$, the capacity result for the GMBC in [24] only holds under these input constraints. In this paper, we take a different approach which is a natural and simple extension of Theorem 2, and which does not rely on the DSM bound presented in [19], [21]. In the following, we are able to give a general result which is true under any compact covariance set constraint on the input $E[\boldsymbol{xx}^T] \in \mathcal{S}$.

In the following subsections, we present the DPC region, give intermediate results, and provide a full characterization of the capacity region of the AMBC.

A. AMBC—DPC Rate Region

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

Given an average transmit covariance matrix limitation $E[\boldsymbol{xx}^T] \preceq \boldsymbol{S}$, a permutation function π and a set of positive semidefinite matrices $\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_m (\boldsymbol{B}_k \succeq 0 \, \forall k \in 1, \dots, m)$, such that $\sum_{i=1}^m \boldsymbol{B}_i \preceq \boldsymbol{S}$, the following rates are achievable in the AMBC using a DPC scheme [7], [20], [22], [27]

$$R_k \leq R_{\pi^{-1}(k)}^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m}), \qquad \forall k \in 1, \dots, m$$

where

$$
\begin{aligned}
&R_l^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m}) \\
&= \frac{1}{2}\log \frac{\left|\left(\left(\sum_{i=1}^l \boldsymbol{B}_{\pi(i)}\right) + \boldsymbol{N}_{\pi(l)}\right)\right|}{\left|\left(\left(\sum_{i=1}^{l-1} \boldsymbol{B}_{\pi(i)}\right) + \boldsymbol{N}_{\pi(l)}\right)\right|}, \quad l = 1, 2, \dots, m
\end{aligned}
$$
(29)

and where π^{-1} is the inverse permutation such that $\pi^{-1}(\pi(i)) = i$. Note that $R_l^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m})$ is not the rate of the (physical) lth user but rather the rate of the $m - l + 1$th user in line to be encoded. We can now define the DPC achievable rate region of an AMBC.

Definition 5 (DPC Rate Region of an AMBC): Let $\boldsymbol{S}$ be a positive semidefinite matrix. The DPC rate region, $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$, of an AMBC with a covariance matrix constraint $\boldsymbol{S}$ is defined by the convex closure:

$$\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m}) = \boldsymbol{conv}\left\{\bigcup_{\pi \in \Pi} \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})\right\}$$
(30)

where Π is the collection of all possible permutations of the ordered set $\{1, \dots, m\}$, $\boldsymbol{conv}$ is the convex closure operator and

where $\mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$ is given in (31) at the bottom of the page.

Note that for an ADBC

$$\mathcal{R}^{\mathrm{DPC}}(\pi_I, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m}) = \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\dots,m})$$

where π_I is the identity permutation such that $\pi_I(i) = i$ and where $\tilde{\boldsymbol{N}}_i = \boldsymbol{N}_i - \boldsymbol{N}_{i-1}, i = 1, \dots, m$ (and where $\boldsymbol{N}_0 = 0$). In Section IV-C, we prove that $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$ is indeed the capacity region.

B. AMBC—Intermediate Results

We first note that not all points on the boundary of $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$ can be directly obtained using a DPC scheme, but rather, only through time-sharing between rate points that can be obtained using DPC. Therefore, unlike the ADBC case, we do not use a similar notion to the optimal Gaussian rate vector, as not all boundary points can be immediately characterized as a solution of an optimization problem (such as in the ADBC case). Instead, as the DPC region $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$, is convex by definition, we use supporting hyperplanes (see [4, pp. 46–50]) in order to define this region. In this subsection, we present supporting lemmas and a crucial observation that is made in Theorem 3 that will help us in the next subsection to prove that $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$ is indeed the capacity region of the AMBC.

Given a sequence of scalars $\bar{\gamma} = (\gamma_1, \dots, \gamma_m)$ and a scalar b, we say that the set $\{\bar{R} = (R_1, \dots, R_m) \,|\, \sum_{i=1}^m \gamma_i R_i = b\}$ is a supporting hyperplane of a closed and bounded set $\mathcal{X} \subset \mathbb{R}^m$, if $\sum_{i=1}^m \gamma_i R_i \leq b \, \forall (R_1, \dots, R_m) \in \mathcal{X}$, with equality for at least one rate vector $(R_1, \dots, R_m) \in \mathcal{X}$. Note that as $\mathcal{X}$ is closed and bounded, $\max_{\bar{R} \in \mathcal{X}} \sum_{i=1}^m \gamma_i R_i$ exists for any vector $\bar{\gamma} = (\gamma_1, \dots, \gamma_m)$. Hence, for any vector $\bar{\gamma}$, we can find a supporting hyperplane for the set $\mathcal{X}$.

Before stating the main result of this subsection, we present and prove two auxiliary lemmas.

Lemma 6: Let $\bar{R}^a = (R_1^a, \dots, R_m^a) \notin \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$ be a rate vector such that $R_i^a \geq 0 \, \forall i = 1, \dots, m$. Then, there exists a constant $b \geq 0$ and a vector $\bar{\gamma} = (\gamma_1, \dots, \gamma_m)$ such that $\gamma_i \geq 0 \, \forall i = 1, \dots, m$ and where not all γ_i are zero, such that the hyperplane $\{(R_1, \dots, R_m) \,|\, \sum_{i=1}^m \gamma_i R_i = b\}$ is a supporting and separating hyperplane for which

$$\sum_{i=1}^m \gamma_i R_i \leq b \, \forall (R_1, \dots, R_m) \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\dots,m}) \quad (32)$$

and for which

$$\sum_{i=1}^m \gamma_i R_i^a > b \quad (33)$$

$$
\mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m}) = \left\{(R_1, \dots, R_m) \,\middle|\, \begin{array}{c} R_i = R_{\pi^{-1}(i)}^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m}) \, \forall i = 1, \dots, m \\ \text{for some } \boldsymbol{B}_1, \dots, \boldsymbol{B}_m \text{ such that} \\ \boldsymbol{S} - \sum_{i=1}^m \boldsymbol{B}_i \succeq 0 \text{ and } \boldsymbol{B}_i \succeq 0 \, \forall i = 1, \dots, m \end{array}\right\}. \quad (31)
$$

where (32) holds with equality for at least one point in $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$.

Proof: As $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ 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 [4, Pt. I, Ch. 2, pp. 46–50]) to show that there exists a supporting hyperplane which strictly separates $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ and $(R_1^a, \ldots, R_m^a)$. In other words, there is a vector $\bar{\gamma} = (\gamma_1, \ldots, \gamma_m)^T$, and a constant b such that (32) and (33) hold. We need to show that we can find a vector $\bar{\gamma}$ with nonnegative elements and a nonnegative scalar b.

Assume, in contrast, that $\gamma_k < 0$ for some $k \in \{1, \ldots, m\}$ and that $\bar{R}^* = (R_1^*, \ldots, R_m^*)$ maximizes $\sum_{i=1}^m \gamma_i R_i$ over all

$$\bar{R} = (R_1, \ldots, R_m) \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}).$$

We note that for all vectors $\bar{R} \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, also

$$(R_1, \ldots, R_{k-1}, 0, R_{k+1}, \ldots, R_m) \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$$

(see Corollary 5 in Appendix III). 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 nonnegative), and does not affect the optimization of $\sum_{i=1}^m \gamma_i R_i$ over all

$$(R_1, \ldots, R_m) \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$$

as otherwise, $\bar{R}^*$ would not have been optimal for the original choice of γ_k. Moreover, we know that at least one of the elements of $\bar{\gamma}$ is strictly positive due to the strict separation implied by the separating hyperplane theorem. Finally, as we have shown that we can find a supporting and separating hyperplane with a vector $\bar{\gamma}$ that contains only nonnegative elements, and as $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ contains only nonnegative vectors,

$$b = \max_{\bar{R} \in \mathcal{R}^{\mathrm{DPC}}} \sum_{i=1}^m \gamma_i R_i \geq 0. \qquad \square$$

Lemma 7: Let $\bar{\gamma} = (\gamma_1, \ldots, \gamma_m)$ be a vector with nonnegative entries 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 that $k(i), i = 1, \ldots, n$, points to strictly positive entries of the vector $\bar{\gamma}$ and such that $k(i+1) > k(i) \forall i = 1, \ldots, n-1$. Consider an AMBC with m users and noise covariance matrices $\boldsymbol{N}_1, \ldots, \boldsymbol{N}_m$ and define a compact AMBC with n users and noise covariance matrices given by $\hat{\boldsymbol{N}}_i = \boldsymbol{N}_{k(i)}, i = 1, \ldots, n$. Then

$$\max_{\boldsymbol{B}_{1,\ldots,m} \in \mathcal{D}} \sum_{i=1}^m \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$$

$$= \max_{\hat{\boldsymbol{B}}_{1,\ldots,n} \in \hat{\mathcal{D}}} \sum_{i=1}^n \gamma_{k(i)} R_i^{\mathrm{DPC}}(\hat{\pi}_I, \hat{\boldsymbol{B}}_{1,\ldots,n}, \hat{\boldsymbol{N}}_{1,\ldots,n}) \quad (34)$$

where

$$\mathcal{D} = \left\{ (\boldsymbol{B}_1, \ldots, \boldsymbol{B}_m) | \boldsymbol{B}_i \succeq 0, \forall i = 1, \ldots, m \text{ and } \sum_{i=1}^m \boldsymbol{B}_i \preceq \boldsymbol{S} \right\}$$

and

$$\hat{\mathcal{D}} = \left\{ (\hat{\boldsymbol{B}}_1, \ldots, \hat{\boldsymbol{B}}_n) | \hat{\boldsymbol{B}}_i \succeq 0, \forall i = 1, \ldots, n \text{ and } \sum_{i=1}^n \hat{\boldsymbol{B}}_i \preceq \boldsymbol{S} \right\}$$

and where $\hat{\pi}_I$ is an identity permutation function over the set $\{1, \ldots, n\}$.

Proof: Let $\boldsymbol{B}_i^*, i = 1, \ldots, m$, be the optimizing matrices of the optimization problem on the left-hand side of (34). Furthermore, let $R_i^* = R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}^*, \boldsymbol{N}_{1,\ldots,m})$ $\forall i = 1, \ldots, m$. We claim that if $\gamma_i = 0 (i \in \{1, \ldots, m\})$ then $R_i^* = 0$ and $\boldsymbol{B}_i^* = 0_{t \times t}$. To show this, we observe that we can always modify the choice of $\boldsymbol{B}_i$'s to increase the DPC rates $R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$ for i's which correspond to $\gamma_i > 0$ at the expense of DPC rates for users which correspond to $\gamma_i = 0$, without violating the matrix constraint $\boldsymbol{S}$ (see Lemma 9 in Appendix III). This will only increase the target function $\sum_{i=1}^m \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$, 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 if $\gamma_i = 0$, then $R_i^* = 0$. Furthermore, by the definition of the rates $R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$ it is easy to show that as $\boldsymbol{B}_j \succeq 0 \forall j$

$$R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m}) = 0$$

if and only if $\boldsymbol{B}_i = 0$. Therefore, if $\gamma_i = 0$, then $\boldsymbol{B}_i^* = 0_{t \times t}$.

We can now define $\hat{\boldsymbol{B}}_i^* = \boldsymbol{B}_{k(i)}^*, i = 1, \ldots, n$. It is easy to see that

$$R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}^*, \boldsymbol{N}_{1,\ldots,m}) = R_i^{\mathrm{DPC}}(\hat{\pi}_I, \hat{\boldsymbol{B}}_{1,\ldots,n}^*, \hat{\boldsymbol{N}}_{1,\ldots,n})$$

and therefore,

$$\max_{\boldsymbol{B}_{1,\ldots,m} \in \mathcal{D}} \sum_{i=1}^m \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$$

$$\leq \max_{\hat{\boldsymbol{B}}_{1,\ldots,n} \in \hat{\mathcal{D}}} \sum_{i=1}^n \gamma_{k(i)} R_i^{\mathrm{DPC}}(\hat{\pi}_I, \hat{\boldsymbol{B}}_{1,\ldots,n}, \hat{\boldsymbol{N}}_{1,\ldots,n})$$

On the other hand, we can show that the opposite inequality also holds for the above expression. Let $\hat{\boldsymbol{B}}_i^{**}, i = 1, \ldots, n$, be the optimizing solution of the optimization problem on the right-hand side of (34). Define

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

where k^{-1} is the inverse of the index function such that $k(k^{-1}(i)) = i$. It is easy to see that

$$R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}^{**}, \boldsymbol{N}_{1,\ldots,m}) = R_i^{\mathrm{DPC}}(\hat{\pi}_I, \hat{\boldsymbol{B}}_{1,\ldots,n}^{**}, \hat{\boldsymbol{N}}_{1,\ldots,n})$$

and, therefore,

$$\max_{\boldsymbol{B}_{1,\ldots,m} \in \mathcal{D}} \sum_{i=1}^m \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$$

$$\geq \max_{\hat{\boldsymbol{B}}_{1,\dots,n} \in \hat{\mathcal{D}}} \sum_{i=1}^{n} \gamma_{k(i)} R_i^{\mathrm{DPC}}(\hat{\pi}_I, \hat{\boldsymbol{B}}_{1,\dots,n}, \hat{\boldsymbol{N}}_{1,\dots,n}). \qquad \square$$

The following theorem brings to bare a relation between the ideas of a supporting hyperplane and the enhanced channel. This theorem is a natural extension of Theorem 1 to the AMBC case and plays a similar role in the proof of the capacity region of the AMBC to the role played by Theorem 1 in the ADBC case.

Theorem 3: Consider an AMBC with noise covariance matrices $(\boldsymbol{N}_1, \dots, \boldsymbol{N}_m)$ and an average transmit covariance matrix constraint $\boldsymbol{S} \succ 0$. Define π_I to be the identity permutation, $\pi_I(i) = i \; \forall i = 1, \dots, m$. If $\{(R_1, \dots, R_m) \,|\, \sum_{i=1}^{m} \gamma_i R_i = b\}$ is a supporting hyperplane of the rate region $\mathcal{R}^{\mathrm{DPC}}(\pi_I, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$ such that $0 \leq \gamma_1 \leq \cdots \leq \gamma_m, \gamma_m > 0$ and $b \geq 0$, then there exists an enhanced ADBC with noise increment covariances $(\tilde{\boldsymbol{N}}_1', \dots, \tilde{\boldsymbol{N}}_m')$ such that the following properties hold.

Enhanced Channel:

$$0 \prec \sum_{i=1}^{k} \tilde{\boldsymbol{N}}_i' \preceq \boldsymbol{N}_k, \qquad \forall k = 1, \dots, m.$$

Supporting Hyperplane Preservation:

$$\left\{ (R_1, \dots, R_m) \,\Big|\, \sum_{i=1}^{m} \gamma_i R_i = b \right\}$$

is also a supporting hyperplane of the rate region $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\dots,m}')$.

Proof: To prove the lemma, we will investigate the properties of the gradients of the DPC rates, $R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m})$, at the point where the rate region $\mathcal{R}^{\mathrm{DPC}}(\pi_I, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$ and the hyperplane

$$\left\{ (R_1, \dots, R_m) \,\Big|\, \sum_{i=1}^{m} \gamma_i R_i = b \right\}$$

touch (as $\{(R_1, \dots, R_m) \,|\, \sum_{i=1}^{m} \gamma_i R_i = b\}$ is a supporting hyperplane, there must be such a common rate vector). Let $\bar{R}^* = (R_1^*, \dots, R_m^*)$ be that rate vector and let $\boldsymbol{B}_1^*, \dots, \boldsymbol{B}_m^*$ be a sequence of positive semidefinite matrices such that $\sum_{i=1}^{m} \boldsymbol{B}_i^* \preceq \boldsymbol{S}$ and such that

$$R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m}^*, \boldsymbol{N}_{1,\dots,m}) = R_i^*, \qquad i = 1, \dots, m.$$

By the definition of the supporting hyperplane, we know that the scalar b and the sequence of matrices $\boldsymbol{B}_1^*, \dots, \boldsymbol{B}_m^*$ are the solution of the following optimization problem:

$$\text{maximize} \quad \sum_{i=1}^{m} \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m})$$
$$\boldsymbol{B}_1, \dots, \boldsymbol{B}_m$$
$$\text{such that} \quad \boldsymbol{B}_i \succeq 0, \qquad \forall i = 1, \dots, m$$
$$\sum_{i=1}^{m} \boldsymbol{B}_i \preceq \boldsymbol{S}.$$

We now note that of all $i = 1, \dots, m$, only $R_m^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m})$ is a function of $\boldsymbol{B}_m$ and is given by

$$R_m^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m})$$
$$= \frac{1}{2} \log \frac{\left| \boldsymbol{B}_m + \left(\sum_{i=1}^{m-1} \boldsymbol{B}_i \right) + \boldsymbol{N}_m \right|}{\left| \left(\sum_{i=1}^{m-1} \boldsymbol{B}_i \right) + \boldsymbol{N}_m \right|}.$$

Therefore, we can use the fact that $|\boldsymbol{B}| > |\boldsymbol{A}|$ for any positive semidefinite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ such that $\boldsymbol{B} \succeq \boldsymbol{A} \succ 0$ and $\boldsymbol{B} \neq \boldsymbol{A}$ to show that for a given sequence of $m-1$ matrices, $\bar{\boldsymbol{B}}_1, \dots, \bar{\boldsymbol{B}}_{m-1}$, the weighted sum $\sum_{i=1}^{m} \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m})$ is maximized by setting $\boldsymbol{B}_m = \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i$ (due to the constraint $\sum_{i=1}^{m} \boldsymbol{B}_i \preceq \boldsymbol{S}$). Therefore, $\boldsymbol{B}_m^* = \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i^*$ and the matrices $\boldsymbol{B}_1^*, \dots, \boldsymbol{B}_{m-1}^*$ are the solution of the following optimization problem:

$$\text{maximize} \quad \sum_{i=1}^{m} \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m-1}, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$$
$$\boldsymbol{B}_1, \dots, \boldsymbol{B}_{m-1}$$
$$\text{such that} \quad \boldsymbol{B}_i \succeq 0, \qquad \forall i = 1, \dots, m-1$$
$$\sum_{i=1}^{m-1} \boldsymbol{B}_i \preceq \boldsymbol{S} \qquad (35)$$

where

$$R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m-1}, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m})$$
$$= R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m}, \boldsymbol{N}_{1,\dots,m})$$

and where $\boldsymbol{B}_m = \boldsymbol{S} - \sum_{i=1}^{m-1} \boldsymbol{B}_i$. 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.

The objective function in (35) is differentiable over $\boldsymbol{B}_i \succeq 0 \,\forall i$ and its partial gradients are given by

$$\nabla_{\boldsymbol{B}_k} \left(\sum_{i=1}^{m} \gamma_i R_i^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\dots,m-1}, \boldsymbol{S}, \boldsymbol{N}_{1,\dots,m}) \right)$$
$$= \frac{1}{2} \sum_{j=k}^{m-1} \gamma_j \cdot \left(\sum_{i=1}^{j} \boldsymbol{B}_i + \boldsymbol{N}_j \right)^{-1}$$
$$- \frac{1}{2} \sum_{j=k+1}^{m} \gamma_j \cdot \left(\sum_{i=1}^{j-1} \boldsymbol{B}_i + \boldsymbol{N}_j \right)^{-1}$$
$$\forall k = 1, \dots, m-1. \quad (36)$$

Let

$$\mathcal{X} = \left\{ (\boldsymbol{B}_1, \dots, \boldsymbol{B}_{m-1}) \,|\, \boldsymbol{B}_i \succeq 0 \;\forall i \text{ and } \sum_{i} \boldsymbol{B}_i \preceq \boldsymbol{S} \right\}$$

denote the optimization region of the problem in (35). As the objective function in (35) is continuously differentiable over an open set that contains $\mathcal{X}$ and as $\mathcal{X}$ is nonempty and convex, the optimal solution of (35), $\boldsymbol{B}_1^*, \dots, \boldsymbol{B}_{m-1}^*$, must observe the following necessary KKT conditions (note that as the optimization problem in (35) does not contain any inequality constraints, there is no need to examine any constraint qualifications, as

in Theorem 1. See also Appendix IV and [2, Sec. 5.4, pp. 302–312])

$$0_{t \times t} = \frac{1}{2} \sum_{j=k}^{m-1} \gamma_j \cdot \left(\sum_{i=1}^{j} \boldsymbol{B}_i^* + \boldsymbol{N}_j \right)^{-1}$$
$$- \frac{1}{2} \sum_{j=k+1}^{m} \gamma_j \cdot \left(\sum_{i=1}^{j-1} \boldsymbol{B}_i^* + \boldsymbol{N}_j \right)^{-1} + \frac{1}{2} \boldsymbol{O}_k - \frac{1}{2} \boldsymbol{O}_m,$$
$$\forall k = 1, \ldots, m-1$$

where $\boldsymbol{O}_k \succeq 0$ is a positive semidefinte $t \times t$ matrix are such that $\mathrm{tr}\{\bar{\boldsymbol{B}}_k \boldsymbol{O}_k\} = 0 \forall k$. Note that as $\boldsymbol{O}_k \succeq 0$ and $\bar{\boldsymbol{B}}_k \succeq 0$, the fact that $\mathrm{tr}\{\bar{\boldsymbol{B}}_k \boldsymbol{O}_k\} = 0$ implies that $\bar{\boldsymbol{B}}_k \cdot \boldsymbol{O}_k = 0_{t \times t}$. By subtracting equation $k + 1$ from the kth equation (except for $k = m - 1$, where the expression is taken as is), we obtain the following $m - 1$

$$\gamma_{k+1} \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \boldsymbol{N}_{k+1} \right)^{-1} + \boldsymbol{O}_{k+1}$$
$$= \gamma_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i^* + \boldsymbol{N}_k \right)^{-1} + \boldsymbol{O}_k, \quad k = 1, \ldots, m-1. \tag{37}$$

We now treat separately the case where $\gamma_i > 0$ for all $i = 1, \ldots, m$ and the case where for some $i = 1, \ldots, m, \gamma_i = 0$. We start with the case of $\gamma_i > 0 \forall i$. Once we have proved the lemma under the assumption that $\gamma_i > 0$ for all $i = 1, \ldots, m$, we will be able to use a simple argument to extend this result to the more general case.

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

As we assume that $\gamma_1 \leq \cdots \leq \gamma_m$ (this assumption was made in Theorem 3) and as $\boldsymbol{B}_i \cdot \boldsymbol{O}_i = 0 \forall i$, we can use Lemma 10 in Appendix V to show that there exists an enhanced ADBC with noise increment covariance matrices $\tilde{\boldsymbol{N}}'_1, \ldots, \tilde{\boldsymbol{N}}'_m$, such that

$$\gamma_{k+1} \left(\sum_{i=1}^{k} \boldsymbol{B}_i + \boldsymbol{N}_{k+1} \right)^{-1} + \boldsymbol{O}_{k+1}$$
$$= \gamma_{k+1} \left(\sum_{i=1}^{k} \boldsymbol{B}_i + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}'_i \right)^{-1}$$
$$= \gamma_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \right)^{-1}$$
$$= \gamma_k \left(\sum_{i=1}^{k} \boldsymbol{B}_i + \boldsymbol{N}_k \right)^{-1} + \boldsymbol{O}_k \tag{38}$$

and such that

$$\frac{|\sum_{i=1}^{k} \boldsymbol{B}_i + \boldsymbol{N}_k|}{|\sum_{i=1}^{k-1} \boldsymbol{B}_i + \boldsymbol{N}_k|} = \frac{|\sum_{i=1}^{k} \boldsymbol{B}_i + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i|}{|\sum_{i=1}^{k-1} \boldsymbol{B}_i + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i|},$$
$$\forall k = 1, \ldots, m. \tag{39}$$

By the expressions of R_k^{DPC} (29), R_k^G (5), and (39), we can see that

$$R_k^* = R_k^{\mathrm{DPC}}(\pi_I, \boldsymbol{B}_{1,\ldots,m}^*, \boldsymbol{N}_{1,\ldots,m})$$
$$= R_k^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}'_{1,\ldots,m}), \quad \forall k = 1, \ldots, m. \tag{40}$$

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 $\mathcal{R}^G(\bar{\boldsymbol{S}}, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$ and intersects with it at $\bar{R}^*$. For that purpose, it is sufficient to show that

$$\sum_{i=1}^{m} \gamma_i \cdot R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$$

is maximized when $\bar{\boldsymbol{B}}_i = \boldsymbol{B}_i^*$, $i = 1, \ldots, m$. For that reason, we define

$$\Delta_k = \sum_{i=1}^{k} (\boldsymbol{B}_i - \boldsymbol{B}_i^*)$$
$$\boldsymbol{K}_k = \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i$$
$$\boldsymbol{T}_k = \sum_{i=1}^{k} \bar{\boldsymbol{B}}_i^* + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}'_i$$

and rewrite the difference between the weighted sums as follows:

$$\sum_{i=1}^{m} \gamma_i R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}'_{1,\ldots,m}) - \sum_{i=1}^{m} \gamma_i R_i^G(\boldsymbol{B}_{1,\ldots,m}^*, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$$
$$= \sum_{k=1}^{m} \gamma_k \cdot \frac{1}{2} \log \frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \right|}{\left| \sum_{i=1}^{k-1} \boldsymbol{B}_i + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \right|}$$
$$- \sum_{k=1}^{m} \gamma_k \cdot \frac{1}{2} \log \frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \right|}{\left| \sum_{i=1}^{k-1} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \right|}$$
$$\leq \sum_{k=1}^{m-1} \left(\gamma_k \cdot \frac{1}{2} \log \frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \right|}{\left| \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k} \tilde{\boldsymbol{N}}'_i \right|} \right.$$
$$\left. - \gamma_{k+1} \cdot \frac{1}{2} \log \frac{\left| \sum_{i=1}^{k} \boldsymbol{B}_i + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}'_i \right|}{\left| \sum_{i=1}^{k} \boldsymbol{B}_i^* + \sum_{i=1}^{k+1} \tilde{\boldsymbol{N}}'_i \right|} \right)$$
$$= \sum_{k=1}^{m-1} \left(\gamma_k \cdot \frac{1}{2} \log \frac{|\Delta_k + \boldsymbol{K}_k|}{|\boldsymbol{K}_k|} - \gamma_{k+1} \cdot \frac{1}{2} \log \frac{|\Delta_k + \boldsymbol{T}_k|}{|\boldsymbol{T}_k|} \right)$$
$$= \sum_{k=1}^{m-1} \left(\gamma_k \cdot \frac{1}{2} \log \left| \boldsymbol{K}_k^{-\frac{1}{2}} \Delta_k \boldsymbol{K}_k^{-\frac{\dagger}{2}} + \boldsymbol{I} \right| \right.$$
$$\left. - \gamma_{k+1} \cdot \frac{1}{2} \log \left| \boldsymbol{T}_k^{-\frac{1}{2}} \Delta_k \boldsymbol{T}_k^{-\frac{\dagger}{2}} + \boldsymbol{I} \right| \right) \tag{41}$$

where the inequality results from the fact that

$$\left| \sum_{i=1}^{m} \boldsymbol{B}_i + \sum_{i=1}^{m} \tilde{\boldsymbol{N}}'_i \right| \leq \left| \sum_{i=1}^{m} \boldsymbol{B}_i^* + \sum_{i=1}^{m} \tilde{\boldsymbol{N}}'_i \right|$$

because

$$\sum_{i=1}^{m} \boldsymbol{B}_i \preceq \sum_{i=1}^{m} \boldsymbol{B}_i^* = \boldsymbol{S}.$$

By (38), we know that $\boldsymbol{T}_k^{-1} = \frac{\gamma_k}{\gamma_{k+1}} \boldsymbol{K}_k^{-1}$. Therefore, we may write (42) at the bottom of the page. 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

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

for all positive semidefinite matrices $\boldsymbol{B}_i$ with power covariance constraint $\boldsymbol{S}$. This completes the proof of the lemma for the case $\gamma_i > 0 \forall i$.

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

Finally, we expand the proof to the case of any set of nonnegative 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 1. 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 $\boldsymbol{S}$ and with noise covariance matrices $\hat{\boldsymbol{N}}_1, \ldots, \hat{\boldsymbol{N}}_n$ such that

$$\hat{\boldsymbol{N}}_i = \boldsymbol{N}_{i+m-n}, \qquad i = 1, \ldots, n.$$

Similarly, we define a compact hyperplane

$$\left\{ (\hat{R}_1, \ldots, \hat{R}_n) \in \mathbb{R}^n \, \Big| \, \sum_{i=1}^{n} \gamma_{i+m-n} \hat{R}_i = b \right\}.$$

By Lemma 7, 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 $\mathcal{R}^{\mathrm{DPC}}(\hat{\pi}_I, \boldsymbol{S}, \hat{\boldsymbol{N}}_{1,\ldots,n})$, where $\hat{\pi}_I$ is now the identity permutation over the set $\{1, \ldots, n\}$. Therefore, we can use the preceding proof for the case where γ_i are strictly positive to show that the theorem holds for the compact

channel. That is, we can find an enhanced ADBC, with noise increment covariance matrices $\hat{\tilde{\boldsymbol{N}}}'_1, \ldots, \hat{\tilde{\boldsymbol{N}}}'_n$, such 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 $\mathcal{R}^G(\boldsymbol{S}, \hat{\tilde{\boldsymbol{N}}}'_{1,\ldots,n})$.

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

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

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

$$0 \prec \left(\sum_{i=1}^{j} \beta_i \right) \cdot \hat{\tilde{\boldsymbol{N}}}'_1 \preceq \boldsymbol{N}_j, \qquad \forall j = 1, \ldots, m-n.$$

As $\boldsymbol{N}_j \succ 0 \forall j = 1, \ldots, m$, it is possible to find such β_i. Clearly, we have defined an enhanced ADBC for the original channel. Furthermore, as $\{(\hat{R}_1, \ldots, \hat{R}_n) \in \mathbb{R}^n \, | \, \sum_{i=1}^{n} \gamma_{i+m-n} \hat{R}_i = b\}$ is a supporting hyperplane of $\mathcal{R}^G(\boldsymbol{S}, \hat{\tilde{\boldsymbol{N}}}'_{1,\ldots,n})$, we can use Lemma 7 to show that $\{(R_1, \ldots, R_m) \in \mathbb{R}^m \, | \, \sum_{i=1}^{m} \gamma_i R_i = b\}$ is a supporting hyperplane of $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$. $\qquad \square$

C. AMBC—Main Result

We can now use Theorem 3 and the capacity region result of the ADBC (Theorem 2) to prove that $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ is the capacity region of the AMBC. Before we turn to the proof, we formally state the main result of this section.

Theorem 4: Let $\mathcal{C}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ denote the capacity region of the AMBC under a covariance matrix constraint $\boldsymbol{S} \succeq 0$. Then $\mathcal{C}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}) = \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$.

Proof: To prove Theorem 4, we will use Theorem 3 to show that for every rate vector, $\bar{R}^a$, which lies outside $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, we can find an enhanced ADBC, whose capacity region does not contain $\bar{R}^a$. However, due to the first statement of Theorem 3, the capacity region of the enhanced

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

$$\leq \sum_{k=1}^{m-1} \left(\gamma_k \cdot \frac{1}{2} \log \left| \boldsymbol{K}_k^{-\frac{1}{2}} \boldsymbol{\Delta}_k \boldsymbol{K}_k^{-\frac{\dagger}{2}} + \boldsymbol{I} \right| - \gamma_{k+1} \cdot \frac{1}{2} \log \left| \frac{\gamma_k}{\gamma_{k+1}} \boldsymbol{K}_k^{-\frac{1}{2}} \boldsymbol{\Delta}_k \boldsymbol{K}_k^{-\frac{\dagger}{2}} + \boldsymbol{I} \right| \right)$$

$$= \sum_{k=1}^{m-1} \frac{1}{2} \cdot \gamma_{k+1} \cdot \left(\begin{array}{l} \left(\frac{\gamma_k}{\gamma_{k+1}} \log \left| \boldsymbol{K}_k^{-\frac{1}{2}} \boldsymbol{\Delta}_k \boldsymbol{K}_k^{-\frac{\dagger}{2}} + \boldsymbol{I} \right| + \left(1 - \frac{\gamma_k}{\gamma_{k+1}} \right) \log |\boldsymbol{I}| \right) \\ - \left(\log \left| \frac{\gamma_k}{\gamma_{k+1}} (\boldsymbol{K}_k^{-\frac{1}{2}} \boldsymbol{\Delta}_k \boldsymbol{K}_k^{-\frac{\dagger}{2}} + \boldsymbol{I}) + \left(1 - \frac{\gamma_k}{\gamma_{k+1}} \right) \cdot \boldsymbol{I} \right| \right) \end{array} \right). \tag{42}$$

channel outer bounds that of the original channel, and therefore, $\bar{R}^a$ cannot be an achievable rate vector. Just as in the proof of Theorem 2, we treat separately the cases $\boldsymbol{S} \succ 0$ and $\boldsymbol{S} \succeq 0, |\boldsymbol{S}| = 0$. We first treat the case where $\boldsymbol{S} \succ 0$ and then broaden the scope of the proof to all $\boldsymbol{S} \succeq 0$.

$\underline{\boldsymbol{S} \succ 0}$:

Let $(R_1^a, \ldots, R_m^a)$ be a rate vector with nonnegative elements which lies outside the rate region $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$. By Lemma 6, we know that there is a supporting and separating hyperplane $\{(R_1, \ldots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b\}$ where $\gamma_i, i = 1, \ldots, m$ are nonnegative and at least one of the elements is positive.

Let $\pi_{\bar{\gamma}}$ be a permutation on the set $\{1, \ldots, m\}$ that orders the elements of $\bar{\gamma}$ such that $\gamma_{\pi_{\bar{\gamma}}(i+1)} \geq \gamma_{\pi_{\bar{\gamma}}(i)} \, \forall i = 1, \ldots, m-1$. We observe that as

$$\mathcal{R}^{\mathrm{DPC}}(\pi_{\bar{\gamma}}, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}) \subseteq \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$$

and as $\{(R_1, \ldots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b\}$ is a supporting hyperplane of $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, we can write

$$
\begin{aligned}
b' &= \max_{\substack{(R_1,\ldots,R_m) \in \\ \mathcal{R}^{\mathrm{DPC}}(\pi_{\bar{\gamma}}, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})}} \sum_{i=1}^m \gamma_i R_i \\
&\leq \max_{\substack{(R_1,\ldots,R_m) \in \\ \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})}} \sum_{i=1}^m \gamma_i R_i = b.
\end{aligned}
$$

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

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

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

Note that, in general, $\pi_{\bar{\gamma}}$ may be any one of the possible permutations. Therefore, to prove the last statement, we exploited the fact that $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ is the convex hull of the union over all DPC rate regions $\mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, where the union is taken over all possible DPC precoding orders.

For brevity, we will assume in the following that $\gamma_1 \leq \cdots \leq \gamma_m$ or alternatively, that $\pi_{\bar{\gamma}} = \pi_I$. If that is not the case, we can always reorder the indices of the users such that this relation will hold. From the above, we know that $\{(R_1, \ldots, R_m) \in \mathbb{R}^m \mid \sum_{i=1}^m \gamma_i R_i = b'\}$ is a supporting and separating hyperplane of $\mathcal{R}^{\mathrm{DPC}}(\pi_{\bar{\gamma}} = \pi_I, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$. By Theorem 3, we know that there exists an enhanced ADBC whose Gaussian rate region $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$ lies under the supporting hyperplane and hence, $(R_1^a, \ldots, R_m^a) \notin \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}'_{1,\ldots,m})$. However, by Theorem 2, 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 $\boldsymbol{S} \succ 0$, we recall 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 $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, we conclude that $\mathcal{C}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}) \subseteq \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$. However, $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ is a set of achievable rates and therefore,

$$\mathcal{C}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}) \supseteq \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}).$$

$\underline{\boldsymbol{S} \succeq 0}$:

Following the same ideas that appeared in the proof of Theorem 2, we broaden the scope of our proof to the case of any $\boldsymbol{S} \succeq 0$. If $\boldsymbol{S}$ is not (strictly) positive definite, by Lemma 2, we know that there exists an equivalent AMBC with less transmit antennas, noise covariance matrices $\hat{\boldsymbol{N}}_1, \ldots, \hat{\boldsymbol{N}}_m$, and an input covariance matrix constraint $\hat{\boldsymbol{S}} \succ 0$ with the exact same capacity region. Because $\hat{\boldsymbol{S}}$ is strictly positive definite, the above proof (for the case of $\boldsymbol{S} \succ 0$) could be applied to the equivalent channel to show that its capacity region is equal to its DPC rate region, $\mathcal{R}^{\mathrm{DPC}}(\hat{\boldsymbol{S}}, \hat{\boldsymbol{N}}_{1,\ldots,m})$. Moreover, it is possible to show that when $\boldsymbol{S}$ is not strictly positive definite, $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}) = \mathcal{R}^{\mathrm{DPC}}(\hat{\boldsymbol{S}}, \hat{\boldsymbol{N}}_{1,\ldots,m})$. Hence, $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ coincides with the capacity region of the AMBC for all $\boldsymbol{S} \succeq 0$. $\square$

The following corollary extends the result of Theorem 4 to the case of the total power constraint.

Corollary 3: Let $\mathcal{C}(P, \boldsymbol{N}_{1,\ldots,m})$ denote the capacity region of the AMBC under a total power constraint $P \geq 0$. Then

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

Proof: As $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ is contiguous w.r.t. $\boldsymbol{S}$, the corollary follows immediately by Lemma 1. $\square$

D. AMBC Example

The following example illustrates the statements of Theorem 3. In this example, we consider a two-user nondegraded AMBC under a covariance matrix input constraint, where the transmitter and each of the receivers have four antennas such that

$$\boldsymbol{N}_1 \approx \begin{pmatrix} 5.20425 & 2.96515 & 3.31237 & -0.67287 \\ 2.96515 & 8.40350 & 2.15397 & 0.51274 \\ 3.31237 & 2.15397 & 4.01902 & -1.23370 \\ -0.67287 & 0.51274 & -1.23370 & 1.15827 \end{pmatrix}$$

$$\boldsymbol{N}_2 \approx \begin{pmatrix} 5.24564 & -3.80767 & -1.49311 & 3.69896 \\ -3.80767 & 8.28347 & 2.97520 & -5.87397 \\ -1.49311 & 2.97520 & 2.39828 & -4.53655 \\ 3.69896 & -5.87397 & -4.53655 & 10.10308 \end{pmatrix}$$

and

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

The boundary of the DPC region, $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,2})$ is illustrated by a solid curve in Fig. 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
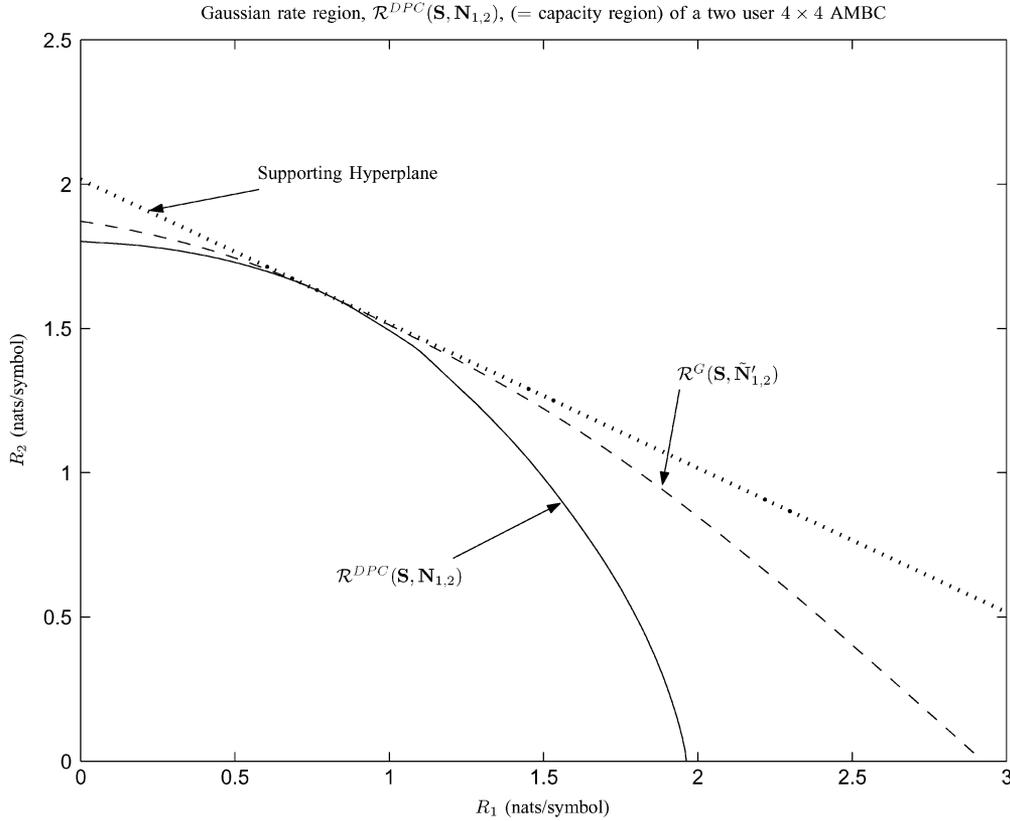
Fig. 2. Illustration of the results of Theorem 3.

and degraded version of the above AMBC with noise increment covariances

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

and

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

As predicted by Theorem 3, we can see from Fig. 2 that for the given hyperplane, we can find an enhanced and degraded version of the channel such that its Gaussian region is supported by the same hyperplane. Furthermore, in relation to the proof of this lemma, we note that the hyperplane and both curves intersect at $(R_1, R_2) \approx (0.79656, 1.61746)$.

V. EXTENSION TO THE GENERAL MIMO BC

We now consider the GMBC (expression (1)) which, unlike the ADBC and AMBC, is characterized by both the noise covariance matrices, $\boldsymbol{N}_1, \ldots, \boldsymbol{N}_m$, and gain matrices, $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_m$. We will prove that the DPC rate region [7], [20], [22], [27] of this channel coincides with the capacity region.

A. GMBC—DPC Rate Region

We begin by characterizing the DPC rate region of the GMBC. Given an average transmit covariance matrix limitation $E[\boldsymbol{x}\boldsymbol{x}^T] \preceq \boldsymbol{S}$, a permutation function π, and a set of

positive semidefinite matrices $\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_m (\boldsymbol{B}_k \succeq 0 \, \forall k \in 1, \ldots, m)$, such that $\sum_{i=1}^m \boldsymbol{B}_i \preceq \boldsymbol{S}$, the following rates are achievable in the GMBC using a DPC scheme [7], [20], [22], [27]:

$$R_k \leq R_{\pi^{-1}(k)}^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m}), \, \forall k \in 1, \ldots, m$$

where

$$R_l^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$$
$$= \frac{1}{2} \log \frac{\left| \left(\boldsymbol{H}_{\pi(l)} \left(\sum_{i=1}^l \boldsymbol{B}_{\pi(i)} \right) \boldsymbol{H}_{\pi(l)}^T + \boldsymbol{N}_{\pi(l)} \right) \right|}{\left| \left(\boldsymbol{H}_{\pi(l)} \left(\sum_{i=1}^{l-1} \boldsymbol{B}_{\pi(i)} \right) \boldsymbol{H}_{\pi(l)}^T + \boldsymbol{N}_{\pi(l)} \right) \right|},$$
$$\forall l = 1, \ldots, m. \quad (43)$$

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

Definition 6 (DPC Rate-Region of a GMBC): Let $\boldsymbol{S}$ be a positive semidefinite matrix. The DPC rate region, $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$, of the GMBC with a covariance constraint $\boldsymbol{S}$, is defined by the following convex closure:

$$\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$$
$$= \boldsymbol{conv} \left\{ \bigcup_{\pi \in \Pi} \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m}) \right\} \quad (44)$$

where $\mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$ is given in (45) at the top of the following page.

$$\mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m}) = \left\{ (R_1, \ldots, R_m) \left| \begin{array}{c} R_i = R^{\mathrm{DPC}}_{\pi^{-1}(i)}(\pi, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m}) \\ \text{for some } \boldsymbol{B}_1, \ldots, \boldsymbol{B}_m \text{ such that} \\ \boldsymbol{S} - \sum_{i=1}^{m} \boldsymbol{B}_i \succeq 0 \text{ and } \boldsymbol{B}_i \succeq 0 \; \forall i = 1, \ldots, m \end{array} \right. \right\} \qquad (45)$$

B. GMBC—Main Result

In the following, we extend the result of Theorem 4 and prove that $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$ is the capacity region. We first formalize this result in the following theorem.

Theorem 5: Let $\mathcal{C}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$ denote the capacity region of the GMBC under a covariance matrix constraint $\boldsymbol{S} \succeq 0$. Then

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

Proof: As

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

we need only to show that

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

We do that in two steps. In the first step, we use the SVD to rewrite the original MIMO BC (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

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{x} + \boldsymbol{n}_i = \boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{V}_i \boldsymbol{x} + \boldsymbol{n}_i, \qquad i = 1, 2, \ldots, m \quad (46)$$

where $\boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{V}_i$ is the singular value decomposition of $\boldsymbol{H}_i$ and where $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ are unitary matrices of sizes $r_i \times r_i$ and $t \times t$ and $\boldsymbol{\Lambda}_i$ is a diagonal matrix of size $r_i \times t$. We assume, without loss of generality, that the singular values of $\boldsymbol{\Lambda}_i$ are arranged in rising order such that all nonzero singular values are located on the lower right corner of $\boldsymbol{\Lambda}_i$. That is, $\boldsymbol{\Lambda}_{i(k-t+r_i,k)} > 0$, $\forall k \in \{t - r_{\Lambda_i} + 1, \ldots, t\}$ and all other elements of $\boldsymbol{\Lambda}_i$ are zero and where $r_{\Lambda_i} = \mathrm{rank}(\boldsymbol{\Lambda}_i)$. Therefore, the first $(r_i - r_{\Lambda_i})$ rows of $\boldsymbol{\Lambda}_i$ are zero.

We multiply the channel outputs at each of the users by $\boldsymbol{U}_i^T$. 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:

$$\boldsymbol{y}_i' = \boldsymbol{\Lambda}_i \boldsymbol{V}_i \boldsymbol{x} + \boldsymbol{n}_i', \qquad i = 1, 2, \ldots, m \qquad (47)$$

where $E[\boldsymbol{n}_i' \boldsymbol{n}_i'^T] = \boldsymbol{U}_i^T \boldsymbol{N}_i \boldsymbol{U}_i = \boldsymbol{N}_i'$. Next, we rewrite $\boldsymbol{N}_i'$ such that

$$\boldsymbol{N}_i' = \begin{pmatrix} \boldsymbol{N}_i'^A & \boldsymbol{N}_i'^B \\ \left(\boldsymbol{N}_i'^B \right)^T & \boldsymbol{N}_i'^C \end{pmatrix}, \qquad i = 1, \ldots, m$$

where $\boldsymbol{N}_i'^A, \boldsymbol{N}_i'^B$, and $\boldsymbol{N}_i'^C$ are of sizes $(r_i - r_{\Lambda_i}) \times (r_i - r_{\Lambda_i})$, $(r_i - r_{\Lambda_i}) \times r_{\Lambda_i}$, and $r_{\Lambda_i} \times r_{\Lambda_i}$. We define

$$\boldsymbol{D}_i = \begin{pmatrix} \boldsymbol{I}_{(r_i - r_{\Lambda_i}) \times (r_i - r_{\Lambda_i})} & 0_{(r_i - r_{\Lambda_i}) \times r_{\Lambda_i}} \\ - \left(\boldsymbol{N}_i'^B \right)^T \left(\boldsymbol{N}_i'^A \right)^{-1} & \boldsymbol{I}_{r_{\Lambda_i} \times r_{\Lambda_i}} \end{pmatrix}.$$

We now create a new variation of the channel by multiplying the received vector of each user by its appropriate $\boldsymbol{D}_i$

$$\begin{aligned} \boldsymbol{y}_i'' &= \boldsymbol{D}_i \boldsymbol{\Lambda}_i \boldsymbol{V}_i \boldsymbol{x} + \boldsymbol{D}_i \boldsymbol{n}_i' \\ &= \boldsymbol{\Lambda}_i \boldsymbol{V}_i \boldsymbol{x} + \boldsymbol{D}_i \boldsymbol{n}_i' = \boldsymbol{H}_i'' \boldsymbol{x} + \boldsymbol{n}_i'', \qquad i = 1, 2, \ldots, m \end{aligned} \tag{48}$$

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

$$\boldsymbol{N}_i'' = E[\boldsymbol{n}_i'' \boldsymbol{n}_i''^T] = \boldsymbol{D}_i \boldsymbol{N}_i' \boldsymbol{D}_i^T = \begin{pmatrix} \boldsymbol{N}_i'^A & 0 \\ 0 & \boldsymbol{N}_i''^C \end{pmatrix} \qquad (49)$$

is block diagonal with $\boldsymbol{N}_i'^A$ and $\boldsymbol{N}_i''^C$ of sizes $(r_i - r_{\Lambda_i}) \times (r_i - r_{\Lambda_i})$ and $r_{\Lambda_i} \times r_{\Lambda_i}$ on the diagonal. Therefore, as $\boldsymbol{n}_i''$ 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 $\boldsymbol{H}_i''$ are all zero such that

$$\boldsymbol{H}_i'' = \begin{pmatrix} 0_{(r_i - r_{\Lambda_i}) \times t} \\ \boldsymbol{H}_i''^B \end{pmatrix}$$

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 $\boldsymbol{N}_i''$ (49)), it is clear that the first $r_i - r_{\Lambda_i}$ antennas (in (48)) 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 $\boldsymbol{N}_i''$, removing the first $r_i - r_{\Lambda_i}$ antennas will not affect the DPC rate region as well, as is shown in the following equation:

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

$$= \frac{1}{2} \log \frac{\left| \boldsymbol{N}'_{\pi(i)}{}^{A} \right|}{\left| \boldsymbol{N}'_{\pi(i)}{}^{A} \right|}$$

$$\cdot \frac{\left| \boldsymbol{H}''_{\pi(i)}{}^{B} \left(\sum_{j=1}^{i} \boldsymbol{B}_{\pi(j)} \right) \boldsymbol{H}''_{\pi(i)}{}^{B^T} + \boldsymbol{N}''_{\pi(i)}{}^{C} \right|}{\left| \boldsymbol{H}''_{\pi(i)}{}^{B} \left(\sum_{j=1}^{i-1} \boldsymbol{B}_{\pi(j)} \right) \boldsymbol{H}''_{\pi(i)}{}^{B^T} + \boldsymbol{N}''_{\pi(i)}{}^{C} \right|}$$

$$= R_i^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}''{}^{C}_{1,\ldots,m}, \boldsymbol{H}''{}^{\boldsymbol{B}}_{1,\ldots,m}).$$

Alternatively, adding zero rows to $\boldsymbol{H}''_i$ at its top (or alternatively, adding zero rows at the top of $\boldsymbol{\Lambda}_i$) 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

$$\hat{\boldsymbol{y}}_i = \hat{\boldsymbol{H}}_i \boldsymbol{x} + \hat{\boldsymbol{n}}_i, \qquad i = 1, 2, \ldots, m \qquad (50)$$

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

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

Again,

$$\hat{\boldsymbol{N}}_i = \begin{pmatrix} \hat{\boldsymbol{N}}^A & 0 \\ 0 & \boldsymbol{N}''_i{}^{C} \end{pmatrix}$$

is block diagonal where $\hat{\boldsymbol{N}}^A$ 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, it is sufficient to show that

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

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

$$\breve{\boldsymbol{y}}_i = \breve{\boldsymbol{H}}_i \boldsymbol{x} + \breve{\boldsymbol{n}}_i, \qquad i = 1, 2, \ldots, m. \qquad (51)$$

where $\breve{\boldsymbol{n}}_i = \hat{\boldsymbol{n}}_i$ and $\breve{\boldsymbol{H}}_i = (\hat{\lambda}_i + \alpha \breve{\boldsymbol{I}}_i) \boldsymbol{V}_i$ for some $\alpha > 0$. $\breve{\boldsymbol{I}}$ is a $t \times t$ diagonal matrix such that

$$\breve{\boldsymbol{I}}_{i(m,n)} = \begin{cases} 1, & m = n \le t - r_{\Lambda_i} \\ 0, & \text{otherwise.} \end{cases}.$$

Note that the last r_{Λ_i} rows of $\breve{\boldsymbol{H}}_i$ are identical to those of $\hat{\boldsymbol{H}}_i$. Therefore, as the ML receiver of the channel in (50) only observes the lower r_{Λ_i} antennas, any codebook and ML receiver

designed for the channel in (50) will achieve the exact same results in the channel given in (51). Thus, it is clear that

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

Additionally, note that $\breve{\boldsymbol{H}}_i$ is invertible and multiplying each receive vector by $\breve{\boldsymbol{H}}_i^{-1}$ will yield an AMBC with the same DPC and capacity regions as that of the channel in (51). As the DPC and capacity regions of an AMBC coincide (Theorem 4), we can write

$$\mathcal{C}(\boldsymbol{S}, \hat{\boldsymbol{N}}_{1,\ldots,m}, \hat{\boldsymbol{H}}_{1,\ldots,m})$$
$$\subseteq \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \breve{\boldsymbol{N}}_{1,\ldots,m}, \breve{\boldsymbol{H}}_{1,\ldots,m}), \qquad \forall \alpha > 0. \quad (52)$$

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

$$\lim_{\alpha \to 0} R_k^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\ldots,m}, \breve{\boldsymbol{N}}_{1,\ldots,m}, \breve{\boldsymbol{H}}_{1,\ldots,m})$$
$$= R_k^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\ldots,m}, \hat{\boldsymbol{N}}_{1,\ldots,m}, \hat{\boldsymbol{H}}_{1,\ldots,m}), \qquad \forall k = 1, \ldots, m$$

where the convergence is uniform over all positive semidefinite matrices $\boldsymbol{B}_i$ such that $\sum \boldsymbol{B}_i \preceq \boldsymbol{S}$. Thus, by the uniform convergence property and by (52), it is clear that for every $\epsilon > 0$ and every rate vector $\bar{R} \in \mathcal{C}(\boldsymbol{S}, \hat{\boldsymbol{N}}_{1,\ldots,m}, \hat{\boldsymbol{H}}_{1,\ldots,m})$, the ϵ-ball around $\bar{R}$ contains a vector $\bar{R}' \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \hat{\boldsymbol{N}}_{1,\ldots,m}, \hat{\boldsymbol{H}}_{1,\ldots,m})$. Therefore,

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

However, as $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \hat{\boldsymbol{N}}_{1,\ldots,m}, \hat{\boldsymbol{H}}_{1,\ldots,m})\}$ is closed,

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

We finalize this section by extending the result of Theorem 5 to the case of the total power constraint, as stated in the following corollary.

Corollary 4: Let $\mathcal{C}(P, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$ denote the capacity region of the GMBC under a total power constraint, $P \ge 0$. Then

$$\mathcal{C}(P, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m}) = \bigcup_{\substack{\boldsymbol{S} \succeq 0 \\ \text{s.t.} \\ \mathrm{tr}\{\boldsymbol{S}\} \le P}} \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m}).$$

Proof: As $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}, \boldsymbol{H}_{1,\ldots,m})$ is contiguous w.r.t. $\boldsymbol{S}$, the corollary follows immediately by Lemma 1. $\square$

We can easily derive the capacity region under a per-antenna power constraint by replacing the constraint $\mathrm{tr}\{\boldsymbol{S}\} \le P$ with the set of constraints $\boldsymbol{S}_{(k,k)} \le P_k$ ($k = 1, \ldots, t$) in the preceding corollary. $\boldsymbol{S}_{(k,k)}$ is the kth element on the diagonal of $\boldsymbol{S}$ and P_k is the power constraint on the kth transmit antenna.

VI. SUMMARY

We have characterized the capacity region of the Gaussian multiple-input multiple-output (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 constraint and the input covariance constraint. In general, our results apply to any input constraint such that the input covariance matrix lies in a compact set of positive semidefinite 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 instead of the original one, an enhanced channel was considered. 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. We suspect that the enhanced channel might find use beyond this paper.

In this paper, we considered the case where only private messages are sent to all users. In [25], we obtained some results for the case where in addition to private messages, a common message is sent. Other aspects of the Gaussian MIMO BC are reviewed in [6].

APPENDIX I
PROOF OF LEMMA 2

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

$$\boldsymbol{U}_{\boldsymbol{S}}^T \boldsymbol{y}_i = \boldsymbol{U}_{\boldsymbol{S}}^T \boldsymbol{x} + \boldsymbol{U}_{\boldsymbol{S}}^T \boldsymbol{n}_i = \boldsymbol{x}' + \boldsymbol{n}_i' = \boldsymbol{y}_i', \qquad i = 1, \ldots, m \tag{53}$$

where there is a matrix power constraint $\boldsymbol{S}' = \boldsymbol{U}_{\boldsymbol{S}}^T \boldsymbol{S} \boldsymbol{U}_{\boldsymbol{S}} = \boldsymbol{\Lambda}_{\boldsymbol{S}}$ on the input vector $\boldsymbol{x}'$, and additive real Gaussian noise vectors with covariance matrices $\boldsymbol{N}_i' = E[\boldsymbol{n}_i' \boldsymbol{n}_i'^T] = \boldsymbol{U}_{\boldsymbol{S}}^T \boldsymbol{N}_i \boldsymbol{U}_{\boldsymbol{S}}$. As this transformation is invertible, 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: $\boldsymbol{N}^A$, $\boldsymbol{N}^B$, and $\boldsymbol{N}^C$ of sizes $(t-r_{\boldsymbol{S}}) \times (t-r_{\boldsymbol{S}})$, $(t-r_{\boldsymbol{S}}) \times r_{\boldsymbol{S}}$ and $r_{\boldsymbol{S}} \times r_{\boldsymbol{S}}$, respectively, such that

$$\boldsymbol{N}_i' = \boldsymbol{U}_{\boldsymbol{S}}^T \boldsymbol{N}_i \boldsymbol{U}_{\boldsymbol{S}} = \begin{pmatrix} \boldsymbol{N}_i^A & \boldsymbol{N}_i^B \\ (\boldsymbol{N}_i^B)^T & \boldsymbol{N}_i^C \end{pmatrix}, \qquad i = 1, \ldots, m.$$

Note that $\boldsymbol{N}_i^A$ and $\boldsymbol{N}_i^C$ are symmetric and positive semidefinite.

We now recall that we assumed that $\boldsymbol{S}$ is not full ranked and that the first $t - r_{\boldsymbol{S}}$ values on the diagonal of $\boldsymbol{\Lambda}_{\boldsymbol{S}}$ are zero and the rest are strictly positive. Therefore, no signal will be transmitted through the first $t - r_{\boldsymbol{S}}$ input elements of the intermediate channel (53). Notice that the users on the receiving ends, receive pure channel noise on the first $t - r_{\boldsymbol{S}}$ receiving antennas. Hence, each user can use the signal on first $t - r_{\boldsymbol{S}}$ antennas to cancel out the effect of the noise on the rest of the antennas such that

the resultant accumulated noise in the first $t - r_{\boldsymbol{S}}$ antennas will be de-correlated from that of the other $r_{\boldsymbol{S}}$ antennas. Hence, we can define a second intermediate channel such that

$$\begin{aligned} \boldsymbol{y}_i'' = \boldsymbol{D}_i \boldsymbol{y}_i' &= \boldsymbol{D}_i \boldsymbol{x}' + \boldsymbol{D}_i \boldsymbol{n}_i' \\ &= \boldsymbol{x}' + \boldsymbol{n}_i'' = \boldsymbol{x}'' + \boldsymbol{n}_i'', \qquad i = 1, \ldots, m \end{aligned} \tag{54}$$

where

$$\boldsymbol{D}_i = \begin{pmatrix} \boldsymbol{I}_{(t-r_{\boldsymbol{S}}) \times (t-r_{\boldsymbol{S}})} & 0_{(t-r_{\boldsymbol{S}}) \times r_{\boldsymbol{S}}} \\ -\left(\boldsymbol{N}_i^B\right)^T \left(\boldsymbol{N}_i^A\right)^{-1} & \boldsymbol{I}_{r_{\boldsymbol{S}} \times r_{\boldsymbol{S}}} \end{pmatrix}$$

and where the third equality follows from the fact that no signal is sent through the first $t - r_{\boldsymbol{S}}$ transmit antennas. Again, since the transformation is invertible, the capacity region of the channel in (54) is identical to that of (53) and to that of the original ADBC/AMBC. Furthermore, the noise is still Gaussian and the matrix power constraint remains $\boldsymbol{\Lambda}_{\boldsymbol{S}}$.

One can verify that the resultant noise covariance at the ith output of channel (54) is given by

$$\begin{aligned} \boldsymbol{N}_i'' &= \boldsymbol{D}_i \boldsymbol{N}_i' \boldsymbol{D}_i^T \\ &= \begin{pmatrix} \boldsymbol{N}_i^A & 0 \\ 0 & \boldsymbol{N}_i^C - \left(\boldsymbol{N}_i^B\right)^T \left(\boldsymbol{N}_i^A\right)^{-1} \left(\boldsymbol{N}_i^B\right) \end{pmatrix}. \end{aligned}$$

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

Thus, we define a new equivalent AMBC with only $r_{\boldsymbol{S}}$ transmit antennas such that

$$\hat{\boldsymbol{y}}_i = \boldsymbol{E}' \boldsymbol{y}_i'' = \boldsymbol{E}' \boldsymbol{x}'' + \boldsymbol{E}' \boldsymbol{n}_i'' = \hat{\boldsymbol{x}} + \hat{\boldsymbol{n}}_i, \qquad i = 1, \ldots, m \tag{55}$$

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

As removing the first $t - r_{\boldsymbol{S}}$ receive antennas did not cause any degradation in performance, it is clear that any codebook that was designed for channel (54) can be modified (chop off the first $r_{\boldsymbol{S}}$ channel inputs) to work in channel (55) with the same results. It is also clear that every codebook that was designed to work in (55), will also work, with some minor modifications (pad with zeros the first $t - r_{\boldsymbol{S}}$ inputs), in the intermediate channel (54) with the exact same results. Therefore, (55) and (54) have the same capacity region and hence our equivalent channel (55) will have the same capacity region as the original ADBC/AMBC.

Finally, we need to show that if $\boldsymbol{N}_1 \preceq \cdots \preceq \boldsymbol{N}_m$, then $\hat{\boldsymbol{N}}_1 \preceq \cdots \preceq \hat{\boldsymbol{N}}_m$. Assume that indeed $\boldsymbol{N}_1 \preceq \cdots \preceq \boldsymbol{N}_m$. It is easily shown that $\boldsymbol{N}_1' \preceq \cdots \preceq \boldsymbol{N}_m'$ where $\boldsymbol{N}_i'$ are the noise covariances of the intermediate channel (53). We will need to show that $\hat{\boldsymbol{N}}_{i+1} - \hat{\boldsymbol{N}}_i \succeq 0$, $i = 1, \ldots, m-1$. For that purpose,

we define $\boldsymbol{D}_i' = \boldsymbol{E}' \boldsymbol{D}_i$. Note that $\boldsymbol{D}_i' \boldsymbol{n}_i'$ is the estimation error of the optimal minimum mean-squared eror (MMSE) estimator of the last $r_{\boldsymbol{S}}$ elements of the noise vector given the first $t - r_{\boldsymbol{S}}$ elements of the vector $\boldsymbol{n}_i'$. We now write $\hat{\boldsymbol{N}}_i - \hat{\boldsymbol{N}}_{i-1}$ as follows:

$$\hat{\boldsymbol{N}}_i - \hat{\boldsymbol{N}}_{i-1} = \boldsymbol{D}_i' \boldsymbol{N}_i' (\boldsymbol{D}_i')^T - \boldsymbol{D}_{i-1}' \boldsymbol{N}_{i-1}' (\boldsymbol{D}_{i-1}')^T$$
$$= \boldsymbol{D}_i' \boldsymbol{N}_{i-1}' (\boldsymbol{D}_i')^T - \boldsymbol{D}_{i-1}' \boldsymbol{N}_{i-1}' (\boldsymbol{D}_{i-1}')^T$$
$$+ \boldsymbol{D}_i' (\boldsymbol{N}_i' - \boldsymbol{N}_{i-1}') (\boldsymbol{D}_i')^T.$$

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

APPENDIX II
PROOF OF LEMMA 3

In order to prove this statement, we introduce the following lemma:

Lemma 8: Assume that

$$(R_1, \ldots, R_{k-1}, R_k, R_{k+1}, \ldots, R_m) \in \mathcal{R}^G(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}).$$

1) If $R_k > 0$, $k = 1, \ldots, m-1$ then for every $0 < \delta \le R_k$, there exists an $\epsilon > 0$ such that

$$(R_1, \ldots, R_{k-1}, R_k - \delta, R_{k+1} + \epsilon, R_{k+2}, \ldots, R_m)$$
$$\in \mathcal{R}^G(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}).$$

2) If $R_k > 0$, $k = 2, \ldots, m$, then for every $0 < \delta \le R_k$, there exists an $\epsilon > 0$ such that

$$(R_1, \ldots, R_{k-1} + \epsilon, R_k - \delta, R_{k+1}, R_{k+2}, \ldots, R_m)$$
$$\in \mathcal{R}^G(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}).$$

3) $(R_1', \ldots, R_m') \in \mathcal{R}^G(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, for all $0 \le R_i' \le R_i$, $i = 1, \ldots, m$.

Proof: If $(R_1, \ldots, R_m) \in \mathcal{R}^G(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, then we can find positive semidefinite matrices $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_m$ such that indeed $R_i = R_i^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m}) \forall i$ (defined in (5)) and such that $\sum_{i=1}^m \boldsymbol{B}_i \preceq \boldsymbol{S}$. By replacing $\boldsymbol{B}_k$ and $\boldsymbol{B}_{k+1}$ with $(1 - \alpha)\boldsymbol{B}_k$ and $\boldsymbol{B}_{k+1} + \alpha\boldsymbol{B}_k$ where $0 < \alpha \le 1$ and by relying on the fact that $|\boldsymbol{B}| > |\boldsymbol{A}|$ if $\boldsymbol{B} \succeq \boldsymbol{A} \succ 0$ and $\boldsymbol{B} \ne \boldsymbol{A}$, it is easy to obtain the first result. Note that our new set of matrices still observes the covariance constraint $\boldsymbol{S}$. The proof of the second statement is almost identical but instead, we replace $\boldsymbol{B}_{k-1}$ and $\boldsymbol{B}_k$ with $\boldsymbol{B}_{k-1} + \alpha\boldsymbol{B}_k$ and $(1-\alpha)\boldsymbol{B}_k$. The last statement is easily proved by recursively using the first statement in an increasing order of users k. The last user's rate, $R_m^G(\boldsymbol{B}_{1,\ldots,m}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$, can be arbitrarily reduced by replacing $\boldsymbol{B}_m$ with $\alpha'\boldsymbol{B}_m$ where $0 \le \alpha' \le 1$. $\qquad \square$

We can now turn to prove Lemma 3.

Proof: 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 $\tilde{\boldsymbol{N}}_1$. If, in the single-user channel $(R_1) \notin \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_1)$, 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. We assume that $\bar{R} = (R_1, \ldots, R_m) \notin \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ and distinguish between two cases:

Case 1: In the $\underline{m-1}$-user channel $\underline{(R_1, \ldots, R_{m-1}) \in}$ $\underline{\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m-1})}$. In this case, we choose the rate vector $\bar{R}^* = (R_1, \ldots, R_{m-1}, R_m^*)$ where

$$R_m^* = \max\{R_m' | (R_1, \ldots, R_{m-1}, R_m') \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})\}$$

(the set is not empty because $(R_1, \ldots, R_{m-1}, 0) \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ and as $\mathcal{R}^G$ is a closed set, the maximum exists). $\bar{R}^*$ is an optimal Gaussian rate vector as otherwise, we could have found a rate vector

$$\bar{R}' = (R_1', R_2', \ldots, R_m') \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$$

with $R_i' \ge R_i^* \forall i$ and with a strict inequality for some i. By iteratively using the first statement of Lemma 8 on $\bar{R}'$, we could have shown that there is a vector $\tilde{R}_m = (R_1, \ldots, R_{m-1}, \tilde{R}_m)$ such that $\tilde{R}_m > R_m' \ge R_m^*$. However, as this implies that

$$R_m^* < \max\{R_m' | (R_1, \ldots, R_{m-1}, R_m') \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})\}$$

(in contradiction to the definition of $\bar{R}^*$), we conclude that $\bar{R}^*$ is indeed an optimal Gaussian rate vector. Finally, R_m^* is strictly smaller than $R_m (R_m^* < R_m)$ as, otherwise, by the third statement of Lemma 8 we would have had $\bar{R} = (R_1, \ldots, R_m) \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$.

Case 2: In the $\underline{m-1}$-user channel $\underline{(R_1, \ldots, R_{m-1}) \notin}$ $\underline{\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m-1})}$. By our induction assumption we know that there must exist an optimal Gaussian rate vector $(R_1', \ldots, R_{m-1}')$ in the $m-1$-user channel such that $R_i \ge R_i'$ $\forall i = 1, \ldots, m-1$. Therefore, we choose

$$\bar{R}' = (R_1', \ldots, R_{m-1}', 0) \in \mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m}).$$

The fact that $\bar{R}'$ exists in $\mathcal{R}^G(\boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ is readily shown by using the same choice of realizing matrices as in the $m-1$-user channel and $\boldsymbol{B}_m = 0_{t \times t}$. Furthermore, $\bar{R}' = (R_1', \ldots, R_{m-1}', 0)$ is an optimal Guassian rate vector in the m-user channel as, otherwise, we could have used the second statement of Lemma 8 to show that $(R_1', \ldots, R_{m-1}')$ is not an optimal Gaussian rate vector in the $m-1$-user channel. As we assume that R_m is strictly larger than 0, the proof is complete. $\qquad \square$

APPENDIX III
EXTENSION OF LEMMA 8 TO THE AMBC CASE

As our proof of Lemma 8 does not depend on the degradedness of the noise covariance matrices $\boldsymbol{N}_i$, we can automatically

extend it to the AMBC case. The following lemma and corollary formalize this extension.

Lemma 9: Let π be an m-user permutation function. If

$$(R_1, \ldots, R_m) \in \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$$

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

$$(R_1^*, \ldots, R_m^*) \in \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$$

and

$$(R_1^{**}, \ldots, R_m^{**}) \in \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$$

where

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

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

Proof: The proof is identical to the proofs of the first and second statements in Lemma 8 where the functions $R_i^G(\boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$ are replaced by $R_i^{\mathrm{DPC}}(\pi, \boldsymbol{B}_{1,\ldots,m}, \boldsymbol{N}_{1,\ldots,m})$. $\square$

Corollary 5: Define the rate vector $\bar{R} = (R_1, \ldots, R_m)$ and let $\bar{R}' = (R_1', \ldots, R_m')$ be any rate vector such that $0 \le R_i' \le R_i, \forall i = 1, \ldots, m$. Then we have the following.
1) If $\bar{R} \in \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, then $\bar{R}' \in \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$.
2) If $\bar{R} \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, then $\bar{R}' \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$.

Proof: The proof of the first statement is a simple result of a recursive application of Lemma 9. To prove the second statement, we note that as $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ is convex, it is sufficient to show that for all $k = 1, \ldots, m$

$$(R_1, \ldots, R_{k-1}, 0, R_{k+1}, \ldots, R_m) \in \mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}).$$

Moreover, as every point in $\mathcal{R}^{\mathrm{DPC}}(\boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$ is a convex combination of points in $\mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$, we need only to show that if

$$(R_1^\pi, \ldots, R_m^\pi) \in \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m})$$

then

$$(R_1^\pi, \ldots, R_{k-1}^\pi, 0, R_{k+1}^\pi, \ldots, R_m^\pi) \in \mathcal{R}^{\mathrm{DPC}}(\pi, \boldsymbol{S}, \boldsymbol{N}_{1,\ldots,m}).$$

However, this is a trivial case of the first statement. $\square$

APPENDIX IV
PROOF OF LEMMA 5

As the optimization problem in (11) is not convex, we may not assume automatically that the KKT conditions apply in this case. Therefore, we must make sure that a set of CQs hold here such that the existence of the KKT conditions is guarantied. Before we examine the appropriate CQs, we first rewrite the problem in (11) in a general form, using the same notation as in [2], and establish a weaker set of conditions (the Fritz John conditions), instead of the KKT conditions

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \le 0, \qquad i = 1, \ldots, m-1 \\ & \boldsymbol{x} \in \mathcal{X} \end{aligned}$$

where the vector $\boldsymbol{x} \in \mathbb{R}^{(m-1) \cdot t^2}$ is created by the concatenation of the rows of $\boldsymbol{B}_1$ through $\boldsymbol{B}_{m-1}$. $f(\boldsymbol{x}) = -R_m^G(\boldsymbol{B}_{1,\ldots,m-1}, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ is the objective function and $g_i(\boldsymbol{x}) = R_i^* - R_i^G(\boldsymbol{B}_{1,\ldots,m-1}, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1,\ldots,m})$ are the inequality constraints. The set $\mathcal{X}$ is the set of vectors over which the optimization is done and is given by $\mathcal{X} = \mathcal{X}_1 \cap \mathcal{X}_2 \cap \cdots \cap \mathcal{X}_{m-1} \cap \mathcal{X}_\Sigma$ where for $i = 1, \ldots, m-1$

$$\mathcal{X}_i = \{\text{row concatenation of } (\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{m-1}) | \boldsymbol{B}_i \succeq 0_{t \times t}\}$$

and where

$$\mathcal{X}_\Sigma = \{\text{row concatenation of } (\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{m-1}) | \sum_{j=1}^{m-1} \boldsymbol{B}_j \preceq \boldsymbol{S}\}.$$

As f and g_i are continuously differentiable over an open set that contains $\mathcal{X}$, the enhanced Fritz John optimality conditions [2, Sec. 5.2, pp. 281–287] hold here (in [2], f and g_i are required to be continuously differentiable over $\mathbb{R}^{(m-1) \cdot t^2}$. However, the proof of these conditions goes through with no modification if f and g_i are continuously differentiable over an open set that contains $\mathcal{X}$). Therefore, for a local minimum, $\boldsymbol{x}^*$, we can write

$$-\left(\mu_0 \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{m-1} \mu_j \nabla g_j(\boldsymbol{x}^*)\right) \in N_\mathcal{X}(\boldsymbol{x}^*) \qquad (56)$$

where $\mu_j \ge 0$ for all $j = 0, 1, \ldots, m-1$ and where $N_\mathcal{X}(\boldsymbol{x}^*)$ is the normal cone of $\mathcal{X}$ at $\boldsymbol{x}^*$ (see [2, Sec. 4.6, pp. 248–254]). As $\mathcal{X}_i \forall i$ and $\mathcal{X}_\Sigma$ are nonempty convex sets such that $\mathrm{ri}(\mathcal{X}_1) \cap \cdots \cap \mathrm{ri}(\mathcal{X}_{m-1}) \cap \mathrm{ri}(\mathcal{X}_\Sigma)$ is not empty (where $\mathrm{ri}(\mathcal{X})$ is the relative interior of the set $\mathcal{X}$ as defined in [4, p. 23]), we can write (see [2, Problem 4.23, p. 267])

$$\begin{aligned} N_\mathcal{X}(\boldsymbol{x}^*) &= N_{\mathcal{X}_1}(\boldsymbol{x}^*) + \cdots + N_{\mathcal{X}_{m-1}}(\boldsymbol{x}^*) + N_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*) \\ &= T_{\mathcal{X}_1}(\boldsymbol{x}^*)^* + \cdots + T_{\mathcal{X}_{m-1}}(\boldsymbol{x}^*)^* + T_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*)^* \end{aligned}$$

where $T_\mathcal{X}(\boldsymbol{x}^*)^*$ is the polar cone of the tangent cone of $\mathcal{X}$ at $\boldsymbol{x}^*$ denoted by $T_\mathcal{X}(\boldsymbol{x}^*)$ (see [2, Sec. 4.6, pp. 248–254]) and the second equality is due to the convexity of the sets $\mathcal{X}_i$ and $\mathcal{X}_\Sigma$ (and hence, $\boldsymbol{x}^*$ is a regular point with respect to these sets). The sum of two vector sets $\mathcal{F}_1$ and $\mathcal{F}_2$ is defined as $\mathcal{F}_1 + \mathcal{F}_2 = \{\boldsymbol{f}_1 + \boldsymbol{f}_2 | \boldsymbol{f}_1 \in \mathcal{F}_1 \text{ and } \boldsymbol{f}_2 \in \mathcal{F}_2\}$.

In order to characterize the normal cones and tangent cones in our case, we first need to define the zero eigenvalue matrix $\boldsymbol{V_B}$. Consider a positive semidefinite matrix $\boldsymbol{B}$ of size $t \times t$ and let $r_{\boldsymbol{B}}$ denote the rank of $\boldsymbol{B}$ (i.e., $r_{\boldsymbol{B}} = \mathrm{rank}(\boldsymbol{B})$). Furthermore, let $\boldsymbol{v}_i^{\boldsymbol{B}}$ be the eigenvectors

of $\boldsymbol{B}$ such that for $i = 1, 2, \ldots, t - r_{\boldsymbol{B}}$, $\boldsymbol{v}_i^{\boldsymbol{B}}$ correspond to zero eigenvalues. Then, $\boldsymbol{V_B} = (\boldsymbol{v}_1^{\boldsymbol{B}} \boldsymbol{v}_2^{\boldsymbol{B}}, \ldots, \boldsymbol{v}_{t-r_{\boldsymbol{B}}}^{\boldsymbol{B}})$. Note that $\boldsymbol{V_B}$ is a matrix of size $t \times (t - r_{\boldsymbol{B}})$. Assume that $\boldsymbol{x}^*$ is the row concatenation of the semidefinite matrices $(\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_{m-1}^*)$ and $\boldsymbol{d}$ is the row concatenation of the symmetric matrices $(\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{m-1})$. It is not difficult to show that if $\boldsymbol{V}_{\boldsymbol{B}_i^*}^T \boldsymbol{L}_i \boldsymbol{V}_{\boldsymbol{B}_i^*} \succ 0_{t \times t}$ then $\boldsymbol{d} \in T_{\mathcal{X}_i}(\boldsymbol{x}^*)$ $(i = 1, \ldots, m - 1)$ and if $\boldsymbol{V}_{\boldsymbol{B}_m^*}^T (\sum_{j=1}^{m-1} \boldsymbol{L}_j) \boldsymbol{V}_{\boldsymbol{B}_m^*} \prec 0_{t \times t}$ then $\boldsymbol{d} \in T_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*)$, where we define $\boldsymbol{B}_m^* = \boldsymbol{S} - \sum_{j=1}^{m-1} \boldsymbol{B}_j^*$.

As the sets $\mathcal{X}_i$ and $\mathcal{X}_\Sigma$ are convex, we have

$$N_{\mathcal{X}_i}(\boldsymbol{x}^*) = T_{\mathcal{X}_i}(\boldsymbol{x}^*)^*$$

and

$$N_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*) = T_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*)^*$$

(i.e., $\boldsymbol{x}^*$ is a regular point). Consequently, one can show that

$$N_{\mathcal{X}_i}(\boldsymbol{x}^*) \cap \mathcal{A}$$
$$\subseteq \{\text{row concatenation of the negative semidefinite}$$
$$t \times t \text{ matrices } (-\boldsymbol{M}_1, \ldots, -\boldsymbol{M}_{m-1}) \,|\, \boldsymbol{M}_i \succeq 0,$$
$$\text{tr}\{\boldsymbol{M}_i \boldsymbol{B}_i^*\} = 0, \boldsymbol{M}_j = 0 \, j \neq i\}$$

and

$$N_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*) \cap \mathcal{A}$$
$$\subseteq \{\text{row concatenation of the positive semidefinite}$$
$$t \times t \text{ matrices } (\boldsymbol{M}_1, \ldots, \boldsymbol{M}_{m-1}) \,|\, \boldsymbol{M}_1 = \cdots$$
$$= \boldsymbol{M}_{m-1} = \boldsymbol{M}_m \succeq 0, \text{tr}\{\boldsymbol{M}_m \boldsymbol{B}_m^*\} = 0\}$$

where $\mathcal{A}$ is defined as the set of all vectors created by the row concatenation of $m - 1$ symmetric matrices of size $t \times t$ (i.e., $\mathcal{A} = \{$ row concatenation of $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_{m-1}) \,|\, \boldsymbol{A}_j = $ a $t \times t$ symmetric matrix $\forall j = 1, \ldots, m - 1\}$).

As the right-hand side of (56) is a row concatenation of symmetric matrices, we can write

$$-\left(\mu_0 \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{m-1} \mu_j \nabla g_j(\boldsymbol{x}^*)\right) \in N_{\mathcal{X}_1}(\boldsymbol{x}^*) \cap \mathcal{A} + \cdots$$
$$+ N_{\mathcal{X}_{m-1}}(\boldsymbol{x}^*) \cap \mathcal{A} + N_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*) \cap \mathcal{A}$$

or alternatively

$$\left(\mu_0 \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{m-1} \mu_j \nabla g_j(\boldsymbol{x}^*)\right)$$
$$= \text{row concatenation of } (\boldsymbol{M}_1 - \boldsymbol{M}_m, \ldots, \boldsymbol{M}_{m-1} - \boldsymbol{M}_m)$$

for some set of positive semidefinite matrices $(\boldsymbol{M}_1, \ldots, \boldsymbol{M}_{m-1}, \boldsymbol{M}_m)$ such that $\text{tr}\{\boldsymbol{M}_i \boldsymbol{B}_i^*\} = 0$ for $i = 1, \ldots, m$. Equation (12) is simply a restatement of the preceding equation by assigning the appropriate objective function and the inequality constraints, assigning $\boldsymbol{O}_i = 2\boldsymbol{M}_i$ and where $\mu_i = \gamma_i$ for $i = 1, \ldots, m - 1$ and $\mu_0 = \gamma_m = 1$.

To complete the proof we need to show that (56) holds with $\mu_0 = 1$. For that purpose, we will use the fact that every point

in $\mathcal{X}$ is regular (as $\mathcal{X}$ is nonempty and convex, see [2, Sec. 4.6]) and we will show that the constraint qualifications denoted by CQ5a in [2, Sec. 5.4, pp. 248–254] hold. More specifically, we will show that there exists a point

$$\boldsymbol{d} \in N_{\mathcal{X}}(\boldsymbol{x}^*)^* = T_{\mathcal{X}}(\boldsymbol{x}^*) = T_{\mathcal{X}_1}(\boldsymbol{x}^*) \cap \cdots \cap T_{\mathcal{X}_{m-1}}(\boldsymbol{x}^*) \cap T_{\mathcal{X}_\Sigma}(\boldsymbol{x}^*)$$

such that $\nabla g_i(\boldsymbol{x}^*)^T \boldsymbol{d} < 0 \; \forall i = 1, \ldots, m - 1$ (where the equality $N_{\mathcal{X}}(\boldsymbol{x}^*)^* = T_{\mathcal{X}}(\boldsymbol{x}^*)$ is a direct result of Proposition 4.6.3 in [2]).

We now translate these CQs to our case. Let $\boldsymbol{z}_k = \sum_{i=1}^{t-r_{\boldsymbol{B}_k^*}} \beta_i \boldsymbol{v}_i^{\boldsymbol{B}_k^*}$ for some choice of scalars β_i and assume that the local minimum $\boldsymbol{x}^*$ is the row concatenation of the matrices $(\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_{m-1}^*)$ and $\boldsymbol{d}$ is the row concatenation of the matrices $(\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{m-1})$. We need to show that there exists a set of matrices $\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{m-1}$ such that
1) if $r_{\boldsymbol{B}_k^*} < t$ $(k = 1, \ldots, m - 1)$, then $\boldsymbol{z}_k^T \boldsymbol{L}_k \boldsymbol{z}_k > 0$ for every choice of scalars β_i such that $\boldsymbol{z}_k \neq 0$;
2) if $r_{\boldsymbol{B}_m^*} < t$, then $\boldsymbol{z}_m^T (\sum_{j=1}^{m-1} \boldsymbol{L}_j) \boldsymbol{z}_m^T < 0$ for every choice of scalars β_i such that $\boldsymbol{z}_m \neq 0$;
3)
$$\text{tr}\{\nabla R_i^G(\boldsymbol{B}_{1, \ldots, m-1}^*, \boldsymbol{S}, \tilde{\boldsymbol{N}}_{1, \ldots, m-1})$$
$$\cdot (\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{m-1})^T\} > 0, \quad \forall i = 1, \ldots, m - 1.$$

The first two items ensure that the row concatenation of $(\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{m-1})$ lie in $N_{\mathcal{X}}(\boldsymbol{x}^*)^*$ and the last ensures that $\nabla g_i(\boldsymbol{x}^*)^T \boldsymbol{d} < 0 \; \forall i$.

For that purpose, we suggest a set of matrices $\boldsymbol{L}_1^*, \ldots, \boldsymbol{L}_{m-1}^*$ of the form

$$\boldsymbol{L}_i^* = -\boldsymbol{B}_i^* + \frac{1}{m} \sum_{j \neq i, j < m} \boldsymbol{B}_j^* + \alpha_i \boldsymbol{B}_m^*, \quad i = 1, \ldots, m - 1 \tag{57}$$

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

$$\boldsymbol{z}_k^T \boldsymbol{L}_k^* \boldsymbol{z}_k = \frac{1}{m} \sum_{j \neq k, j < m} \boldsymbol{z}_k^T \boldsymbol{B}_j^* \boldsymbol{z}_k + \alpha_k \boldsymbol{z}_k^T \boldsymbol{B}_m^* \boldsymbol{z}_k.$$

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

To prove that the second condition holds for $\boldsymbol{L}_1^*, \ldots, \boldsymbol{L}_{m-1}^*$, we note that when $r_{\boldsymbol{B}_m^*} < t$

$$\boldsymbol{z}_m^T \left(\sum_{i=1}^{m-1} \boldsymbol{L}_i^*\right) \boldsymbol{z}_m = \boldsymbol{z}_m^T \left(-\frac{2}{m}(\boldsymbol{S} - \boldsymbol{B}_m^*) + \sum_{i=1}^{m-1} \alpha_i \boldsymbol{B}_m^*\right) \boldsymbol{z}_m$$
$$= -\frac{2}{m} \boldsymbol{z}_m^T \boldsymbol{S} \boldsymbol{z}_m < 0$$

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

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 (13), we can rewrite the directional gradient of the ith rate as follows:

$$
\begin{aligned}
&\mathrm{tr}\left\{\nabla R_i^G(\boldsymbol{B}_{1,\ldots,m-1}^*,\boldsymbol{S},\tilde{\boldsymbol{N}}_{1,\ldots,m-1})\cdot(\boldsymbol{L}_1^*,\ldots,\boldsymbol{L}_{m-1}^*)^T\right\}\\
&=\sum_{j=1}^{m-1}\mathrm{tr}\left\{\nabla_{\boldsymbol{B}_j}R_i^G(\boldsymbol{B}_{1,\ldots,m-1}^*,\boldsymbol{S},\tilde{\boldsymbol{N}}_{1,\ldots,m-1})\cdot\boldsymbol{L}_j^*\right\}\\
&=\sum_{j=1}^{i-1}\mathrm{tr}\left\{\nabla_{\boldsymbol{B}_j}R_i^G(\boldsymbol{B}_{1,\ldots,m-1}^*,\boldsymbol{S},\tilde{\boldsymbol{N}}_{1,\ldots,m-1})\cdot\boldsymbol{L}_j^*\right\}\\
&\quad+\mathrm{tr}\left\{\frac{1}{2}\left(\sum_{l=1}^{i}(\boldsymbol{B}_l^*+\tilde{\boldsymbol{N}}_l)\right)^{-1}\boldsymbol{L}_i^*\right\}\\
&=\sum_{j=1}^{i-1}\mathrm{tr}\left\{\nabla_{\boldsymbol{B}_j}R_i^G(\boldsymbol{B}_{1,\ldots,m-1}^*,\boldsymbol{S},\tilde{\boldsymbol{N}}_{1,\ldots,m-1})\cdot\boldsymbol{L}_j^*\right\}\\
&\quad+\mathrm{tr}\left\{\frac{1}{2}\left(\sum_{l=1}^{i}(\boldsymbol{B}_l^*+\tilde{\boldsymbol{N}}_l)\right)^{-1}\right.\\
&\qquad\qquad\left.\times\left(-\boldsymbol{B}_i^*+\frac{1}{m}\sum_{j\neq i,j<m}\boldsymbol{B}_j^*\right)\right\}\\
&\quad+\alpha_i\cdot\mathrm{tr}\left\{\frac{1}{2}\left(\sum_{l=1}^{i}(\boldsymbol{B}_l^*+\tilde{\boldsymbol{N}}_l)\right)^{-1}\boldsymbol{B}_m^*\right\}.
\end{aligned}
$$

In general, we observe that the expression above for the ith rate depends only on $\bar{\boldsymbol{L}}_1,\ldots,\bar{\boldsymbol{L}}_i$, 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 as $\boldsymbol{B}_m^*\neq 0$ is positive semidefinite and nonzero (this is one of the assumptions of the lemma) and as $(\sum_{l=1}^{i}(\boldsymbol{B}_l^*+\tilde{\boldsymbol{N}}_l))^{-1}$ 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 this expression is positive for all $i=1,\ldots,m-1$.

APPENDIX V
PROOF OF (19) AND (20)

In order to prove (19) and (20) we rely on the following lemma.

Lemma 10: Let $\boldsymbol{B}_i$ and $\boldsymbol{O}_i$ be positive semidefinite $t\times t$ matrices such that $\boldsymbol{B}_i\cdot\boldsymbol{O}_i=0_{t\times t}\,\forall i=1,\ldots,m$ and let $\boldsymbol{N}_i\succ 0$ be strictly positive definite $t\times t$ matrices $\forall i=1,\ldots,m$. If for all $k=1,\ldots,m$

$$
\gamma_{k+1}\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_{k+1}\right)^{-1}+\boldsymbol{O}_{k+1}
$$

$$
=\gamma_k\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k\right)^{-1}+\boldsymbol{O}_k \quad (58)
$$

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

$$
\begin{aligned}
&\gamma_{k+1}\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_{k+1}\right)^{-1}+\boldsymbol{O}_{k+1}\\
&=\gamma_{k+1}\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_{k+1}'\right)^{-1}\\
&=\gamma_k\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k'\right)^{-1}\\
&=\gamma_k\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k\right)^{-1}+\boldsymbol{O}_k. \quad (59)
\end{aligned}
$$

Furthermore

$$
\frac{|\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k|}{|\sum_{i=1}^{k-1}\boldsymbol{B}_i+\boldsymbol{N}_k|}=\frac{|\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k'|}{|\sum_{i=1}^{k-1}\boldsymbol{B}_i+\boldsymbol{N}_k'|},\quad\forall k=1,\ldots,m.
$$

Before proving the lemma we first present and prove two auxiliary results.

Lemma 11: Let $\boldsymbol{B}\succeq 0$, $\boldsymbol{X}\succ 0$ and $\boldsymbol{O}\succeq 0$ be $t\times t$ symmetric matrices such that $\boldsymbol{B}\cdot\boldsymbol{O}=0_{t\times t}$ and let α be a strictly positive scalar. Then the following statements hold:
1) $\alpha(\boldsymbol{B}+\boldsymbol{X})^{-1}+\boldsymbol{O}=\alpha(\boldsymbol{B}+\boldsymbol{X}')^{-1}$ where $\boldsymbol{X}\succeq\boldsymbol{X}'=(\boldsymbol{X}^{-1}+\frac{1}{\alpha}\boldsymbol{O})^{-1}\succ 0$;
2) $\frac{|\boldsymbol{B}+\boldsymbol{X}|}{|\boldsymbol{X}|}=\frac{|\boldsymbol{B}+\boldsymbol{X}'|}{|\boldsymbol{X}'|}$.

   Proof: To prove the first statement, we will show that

$$
((\boldsymbol{B}+\boldsymbol{X})^{-1}+\frac{1}{\alpha}\boldsymbol{O})^{-1}=\boldsymbol{B}+\boldsymbol{X}'
$$

where $\boldsymbol{X}'$ is as defined in the first statement. For that purpose we write

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

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

We can now prove the second statement. As

$$(\boldsymbol{B}+\boldsymbol{X}')^{-1} = (\boldsymbol{B}+\boldsymbol{X})^{-1} + \frac{1}{\alpha}\boldsymbol{O}$$

we can write

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

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

Lemma 12: Let $\boldsymbol{X}_1 \succ 0, \boldsymbol{X}_2 \succ 0, \boldsymbol{B}_1 \succeq 0, \boldsymbol{B}_2 \succeq 0$, and $\boldsymbol{O}_2 \succeq 0$ be $t \times t$ symmetric matrices such that $\boldsymbol{B}_2\cdot\boldsymbol{O}_2 = 0_{t\times t}$. If for some scalar $0 < \alpha \leq 1$

$$(\boldsymbol{B}_1+\boldsymbol{X}_2)^{-1} + \boldsymbol{O}_2 = \alpha(\boldsymbol{B}_1+\boldsymbol{X}_1)^{-1}$$

then the following two statements hold:

1) $(\boldsymbol{B}_1+\boldsymbol{X}_2)^{-1}+\boldsymbol{O}_2 = (\boldsymbol{B}_1+\boldsymbol{X}_2')^{-1}$ for some $\boldsymbol{X}_2'$ satisfying $\boldsymbol{X}_2 \succeq \boldsymbol{X}_2' \succeq \boldsymbol{X}_1$.

2) $\frac{|\boldsymbol{B}_1+\boldsymbol{B}_2+\boldsymbol{X}_2|}{|\boldsymbol{B}_1+\boldsymbol{X}_2|} = \frac{|\boldsymbol{B}_1+\boldsymbol{B}_2+\boldsymbol{X}_2'|}{|\boldsymbol{B}_1+\boldsymbol{X}_2'|}$

Proof: Define $\boldsymbol{K} = (\boldsymbol{B}_1+\boldsymbol{X}_2)^{-1}+\boldsymbol{O}_2$. As $\boldsymbol{O}_2 \succeq 0$, we know that $\boldsymbol{K}^{-1} \preceq (\boldsymbol{B}_1+\boldsymbol{X}_2)$. As $0 < \alpha \leq 1$ and $(\boldsymbol{B}_1+\boldsymbol{X}_2)^{-1}+\boldsymbol{O}_2 = \alpha(\boldsymbol{B}_1+\boldsymbol{X}_1)^{-1}$, we have $(\boldsymbol{B}_1+\boldsymbol{X}_1) \preceq \boldsymbol{K}^{-1} \preceq (\boldsymbol{B}_1+\boldsymbol{X}_2)$. Therefore, by choosing $\boldsymbol{X}_2' = \boldsymbol{K}^{-1} - \boldsymbol{B}_1$ we have $\boldsymbol{X}_2 \succeq \boldsymbol{X}_2' \succeq \boldsymbol{X}_1$ and we can write

$$(\boldsymbol{B}_1+\boldsymbol{X}_2')^{-1} = \boldsymbol{K} = (\boldsymbol{B}_1+\boldsymbol{X}_2)^{-1}+\boldsymbol{O}_2.$$

By the above equation we may write

$$
\begin{aligned}
\frac{|\boldsymbol{B}_1+\boldsymbol{B}_2+\boldsymbol{X}_2'|}{|\boldsymbol{B}_1+\boldsymbol{X}_2'|} &= \frac{|\boldsymbol{B}_2(\boldsymbol{B}_1+\boldsymbol{X}_2')^{-1}+\boldsymbol{I}|}{|\boldsymbol{I}|} \\
&= \frac{\left|\boldsymbol{B}_2\left((\boldsymbol{B}_1+\boldsymbol{X}_2)^{-1}+\boldsymbol{O}_2\right)+\boldsymbol{I}\right|}{|\boldsymbol{I}|} \\
&= \frac{|\boldsymbol{B}_2(\boldsymbol{B}_1+\boldsymbol{X}_2)^{-1}+\boldsymbol{I}|}{|\boldsymbol{I}|} \\
&= \frac{|\boldsymbol{B}_1+\boldsymbol{B}_2+\boldsymbol{X}_2|}{|\boldsymbol{B}_1+\boldsymbol{X}_2|}. \qquad\square
\end{aligned}
$$

We now turn to prove Lemma 10.

Proof: For brevity, we rewrite expression (58) with some notational modifications. For $k = 1, \ldots, m-1$ define

$$
\boldsymbol{L}_k = \left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_{k+1}\right)^{-1} + \frac{1}{\gamma_{k+1}}\boldsymbol{O}_{k+1}
$$

$$
\boldsymbol{R}_k = \gamma_k\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k\right)^{-1} + \boldsymbol{O}_k. \tag{60}
$$

By (58), we have

$$
\boldsymbol{L}_k = \frac{1}{\gamma_{k+1}}\boldsymbol{R}_k, \qquad \forall k = 1, \ldots, m-1. \tag{61}
$$

We need to show that

$$
\begin{aligned}
\boldsymbol{L}_k &= \left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_{k+1}'\right)^{-1} \\
&= \frac{\gamma_k}{\gamma_{k+1}}\left(\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k'\right)^{-1} \\
&= \frac{1}{\gamma_{k+1}}\boldsymbol{R}_k, \qquad \forall k = 1, \ldots, m-1 \tag{62}
\end{aligned}
$$

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

$$
\frac{|\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k|}{|\sum_{i=1}^{k-1}\boldsymbol{B}_i+\boldsymbol{N}_k|} = \frac{|\sum_{i=1}^{k}\boldsymbol{B}_i+\boldsymbol{N}_k'|}{|\sum_{i=1}^{k-1}\boldsymbol{B}_i+\boldsymbol{N}_k'|}, \quad \forall k = 1, \ldots, m. \tag{63}
$$

We use induction on k and begin by exploring (60) and (61) for $k = 1$. As $\boldsymbol{O}_1 \succeq 0$ and as $\boldsymbol{B}_1\cdot\boldsymbol{O}_1 = 0$, by Lemma 11, we can replace $\boldsymbol{R}_1$ with $\gamma_1(\boldsymbol{B}+\boldsymbol{N}_1')^{-1}$ where $0 \prec \boldsymbol{N}_1' \preceq \boldsymbol{N}_1$. Furthermore, by the same lemma we have

$$
\frac{|\boldsymbol{B}_1+\boldsymbol{N}_1|}{|\boldsymbol{N}_1|} = \frac{|\boldsymbol{B}_1+\boldsymbol{N}_1'|}{|\boldsymbol{N}_1'|}.
$$

Since $\gamma_2 \geq \gamma_1$, by Lemma 12, $\boldsymbol{L}_1$ can be replaced by $(\boldsymbol{B}_1+\boldsymbol{N}_2')^{-1}$ such that $\boldsymbol{N}_1' \preceq \boldsymbol{N}_2' \preceq \boldsymbol{N}_2$ and in addition

$$
\frac{|\boldsymbol{B}_1+\boldsymbol{B}_2+\boldsymbol{N}_2|}{|\boldsymbol{B}_2+\boldsymbol{N}_2|} = \frac{|\boldsymbol{B}_1+\boldsymbol{B}_2+\boldsymbol{N}_2'|}{|\boldsymbol{B}_1+\boldsymbol{N}_2'|}
$$

and, therefore, (62) holds for $k = 1$ and (63) holds for $k = 1, 2$.

Next, we assume that (62) holds for $k = 1, \ldots, n$ for some $1 \leq n < m$ with matrices $\boldsymbol{N}_1', \ldots, \boldsymbol{N}_{n+1}'$ such that $\boldsymbol{N}_i' \preceq \boldsymbol{N}_i \ \forall i = 1, \ldots, n+1$ and such that $0 \prec \boldsymbol{N}_1' \preceq \cdots \preceq \boldsymbol{N}_{n+1}'$ and prove that (62) must hold for $k = 1, \ldots, n+1$ and (63) holds for $k = 1, \ldots, n+2$ with matrices $\boldsymbol{N}_1', \ldots, \boldsymbol{N}_{n+2}'$ such that $\boldsymbol{N}_i' \preceq \boldsymbol{N}_i \ \forall i = 1, \ldots, n+2$ and such that $0 \prec \boldsymbol{N}_1' \preceq \cdots \preceq \boldsymbol{N}_{n+2}'$. For that purpose, we define

$$
\boldsymbol{Q} = \sum_{i=1}^{n}\boldsymbol{B}_i+\boldsymbol{N}_{n+1}. \tag{64}
$$

As $\boldsymbol{O}_{n+1} \succeq 0$ and as $\boldsymbol{B}_{n+1}\cdot\boldsymbol{O}_{n+1} = 0_{t\times t}$, we can use Lemma 11 to rewrite the expression for $\boldsymbol{R}_{n+1}$ in (60) as follows:

$$
\begin{aligned}
\boldsymbol{R}_{n+1} &= \gamma_{n+1}(\boldsymbol{B}_{n+1}+\boldsymbol{Q})^{-1}+\boldsymbol{O}_{n+1} \\
&= \gamma_{n+1}(\boldsymbol{B}_{n+1}+\boldsymbol{Q}')^{-1} \tag{65}
\end{aligned}
$$

where $\boldsymbol{Q}' \preceq \boldsymbol{Q}$ and where $\boldsymbol{Q}' = (\boldsymbol{Q}^{-1}+\frac{1}{\gamma_{n+1}}\boldsymbol{O}_{n+1})^{-1}$. However, by (64) and (60), we may write $\boldsymbol{Q}' = \boldsymbol{L}_n^{-1}$. Furthermore, by our induction assumption, (62) holds for $k = n$ and, therefore, we may write

$$
\boldsymbol{Q}' = \boldsymbol{L}_n^{-1} = \sum_{i=1}^{n}\boldsymbol{B}_i+\boldsymbol{N}_{n+1}'. \tag{66}
$$

Thus, by (66), (65), (60), and (61) we may write

$$\boldsymbol{L}_{n+1} = \left(\sum_{i=1}^{n+1} \boldsymbol{B}_i + \boldsymbol{N}_{n+2}\right)^{-1} + \frac{1}{\gamma_{n+2}}\boldsymbol{O}_{n+2}$$

$$= \frac{\gamma_{n+1}}{\gamma_{n+2}}\left(\sum_{i=1}^{n+1} \boldsymbol{B}_i + \boldsymbol{N}'_{n+1}\right)^{-1}$$

$$= \frac{1}{\gamma_{n+2}}\boldsymbol{R}_{n+1}.$$

As $\boldsymbol{O}_{n+2} \succeq 0$ and as $\frac{\gamma_{n+1}}{\gamma_{n+2}} \leq 1$, we can use Lemma 12 to rewrite the above expression as follows:

$$\bar{\boldsymbol{L}}_{n+1} = \left(\sum_{i=1}^{n+1} \boldsymbol{B}_i + \boldsymbol{N}'_{n+2}\right)^{-1}$$

$$= \frac{\gamma_{n+1}}{\gamma_{n+2}}\left(\sum_{i=1}^{n+1} \boldsymbol{B}_i + \boldsymbol{N}'_{n+1}\right)^{-1}$$

$$= \frac{1}{\gamma_{n+2}}\boldsymbol{R}_{n+1}.$$

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

$$\frac{|\sum_{i=1}^{n+2} \boldsymbol{B}_i + \boldsymbol{N}_{n+2}|}{|\sum_{i=1}^{n+1} \boldsymbol{B}_i + \boldsymbol{N}_{n+2}|} = \frac{|\sum_{i=1}^{n+2} \boldsymbol{B}_i + \boldsymbol{N}'_{n+2}|}{|\sum_{i=1}^{n+1} \boldsymbol{B}_i + \boldsymbol{N}'_{n+2}|}$$

and thus we have shown that (62) holds for $k = 1, \ldots, n+1$ and that (63) holds for $k = 1, \ldots, n+2$. $\square$

REFERENCES

[1] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 279–280, Mar. 1974.

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

[3] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 2, pp. 267–271, Apr. 1965.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[5] V. Bryant, *Metric Spaces*. New York: Cambridge Univ. Press, 1985.

[6] G. Caire, S. Shamai (Shitz), Y. Steinberg, and H. Weingarten, *Space-Time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2006, ch. On Information Theoretic Aspects of MIMO-Broadcast Channels.

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

[8] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[9] T. M. Cover, "Comments on broadcast channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2524–2530, Sep. 1998.

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

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

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

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

[14] J. Körner and K. Marton, "General brodcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 60–64, Jan. 1977.

[15] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.

[16] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 374–377, May 1978.

[17] M. Sharif and B. Hassibi, "On the capacity of mimo broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[18] S. Shamai (Shitz) and B. Zaidel, "Enhancing the cellular down-link capacity via co-processing at the transmitting end," in *Proc. IEEE Semi-annual Vehicular Technology, VTC2001 Spring Conf.*, Rhodes, Greece, May. 2001, ch. 3, sec. 36.

[19] D. N. C. Tse and P. Viswanath, "On the capacity of the multiple antenna broadcast channel," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G. J. Foschini and S. Verdú, Eds. Providence, RI: DIMACS, American Math. Soc., 2003, pp. 87–105.

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

[21] 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. Providence, RI: DIMACS, Amer. Math. Soc., 2003, pp. 107–122.

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

[23] P. Viswanathan, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1276–1294, Jun. 2002.

[24] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the gaussian mimo broadcast channel," in *Proc. Conf. Information Sciences and Systems (CISS 2004)*, Princeton, NJ, Mar. 2004, pp. 7–12.

[25] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "On the capacity region of the multi-antenna broadcast channel with common messages," in *Proc. IEE Int. Symp. Information Theory*, Seatle, WA, Jul. 2006, pp. 2195–2199.

[26] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory.*, vol. 52, no. 2, pp. 361–374, Feb. 2006.

[27] W. Yu and J. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[28] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang, "Writing on colored paper," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, Jun. 2001, p. 309.