An Alternative Proof for the Capacity Region of the Degraded Gaussian MIMO Broadcast Channel

Ersen Ekrem   Sennur Ulukus

Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
ersen@umd.edu   *ulukus@umd.edu*

Abstract— We provide an alternative proof for the capacity region of the degraded Gaussian multiple-input multiple-output (MIMO) broadcast channel. Our proof does not use the channel enhancement technique as opposed to the original proof of Weingertan *et. al.* and the alternative proof of Liu *et. al.*. Our proof starts with the single-letter description of the capacity region of the degraded broadcast channel, and directly evaluates it for the degraded Gaussian MIMO broadcast channel by using two main technical tools. The first one is the generalized de Bruin identity which provides a connection between the differential entropy and the Fisher information matrix. The second tool we use is an inequality due to Dembo which lower bounds the differential entropy in terms of the Fisher information matrix.

I. Introduction

The Gaussian multiple-input multiple-output (MIMO) broadcast channel consists of one transmitter and an arbitrary number of receivers, where the transmitter and receivers are equipped with multiple antennas. In this channel, each link between the transmitter and each receiver is modeled by a linear additive Gaussian channel. In general, the Gaussian MIMO broadcast channel is non-degraded, thus, we do not have a single-letter description of the capacity region. Despite this lack of a single-letter description, the capacity region of the Gaussian MIMO broadcast channel is successfully obtained in [1]. Subsequently, an alternative proof is given in [2]. In both proofs, the *channel enhancement* technique [1] is the main tool.

Reference [1] obtains the capacity region of the Gaussian MIMO broadcast channel in three main steps. As the first step, [1] finds the capacity region of the degraded Gaussian MIMO broadcast channel. To this end, [1] first shows that as opposed to the scalar Gaussian broadcast channel [3], the entropy-power inequality falls short of providing a converse proof for the degraded Gaussian vector, i.e., MIMO, broadcast channel. This insufficiency of the entropy-power inequality is alleviated by the invention of the *channel enhancement* technique [1]. Using this technique, [1] constructs a new degraded Gaussian MIMO broadcast channel for each point on the boundary of the Gaussian rate region[1] of the original degraded channel, where the boundaries of the Gaussian rate

regions of both channels intersect at that point, and the capacity region of the constructed degraded channel includes the capacity region of the original one. Then, [1] completes the first step of the proof by showing that the Gaussian rate region is the capacity region of the constructed degraded channel, for which Bergmans' converse [3] can be adapted as opposed to the original channel.

Secondly, [1] considers the aligned Gaussian MIMO broadcast channel, where the transmitter and all receivers have the same number of antennas. This channel is not degraded, thus, there is no single-letter expression for its capacity region. Reference [1] shows that the achievable rate region obtained by using dirty paper coding (DPC), i.e., the DPC region, is the capacity region of the aligned channel. For this purpose, [1] uses the channel enhancement technique one more time along with the capacity result obtained for the degraded Gaussian MIMO broadcast channel in the first step.

Finally, [1] considers the general, not necessarily degraded or aligned, Gaussian MIMO broadcast channel and shows that the DPC region again amounts to the capacity region by using some limiting arguments in conjunction with the capacity result obtained for the aligned channel.

Similar to the proof of [1], the alternative proof in [2] uses the channel enhancement technique as well. The alternative proof in [2] can be divided into two parts. In the first part, [2] considers an optimization problem which is the maximization of the difference of two differential entropy terms (see Theorem 1 of [2]) which cannot be solved by a stand-alone use of the entropy-power inequality. Next, [2] provides two proofs for the fact that the Gaussian distribution is the maximizer of this optimization problem. In both proofs provided in [2], the channel enhancement technique is used. In the second part, [2] considers Marton's outer bound [4] for the general broadcast channel, and evaluates it for the aligned Gaussian MIMO broadcast channel by using the optimization problem solved in the first part. This evaluation yields the capacity region of the two-user aligned Gaussian MIMO broadcast channel.

We note that though the proof in [2] is for the aligned, not necessarily degraded, Gaussian MIMO broadcast channel, if this proof is adapted to find the capacity region of the degraded Gaussian MIMO broadcast channel, again the channel enhancement technique will be needed. In particular, for the degraded case, the optimization problem solved in the first part

[1]The Gaussian rate region refers to the achievable rate region obtained by superposition coding and successive decoding with Gaussian codebooks.

of the proof in [2] will change slightly, however the need for the channel enhancement will remain. In fact, the optimization problem needed for the degraded case is a special case of the original optimization problem solved in the first part of the proof in [2], which is given in Corollary 4 of [2].

Here, we revisit the degraded Gaussian MIMO broadcast channel and provide an alternative proof for the capacity region of this degraded channel, without using the channel enhancement technique. Though channel enhancement is an elegant technique that finds itself diverse applications, we believe that our proof is more direct. On the other hand, our proof is limited to the degraded case and does not seem to be extendable for the general case. In other words, to obtain the capacity region for the general case after finding the capacity region for the degraded case through our proof, one needs to use the channel enhancement technique [1].

Our proof starts with the single-letter description of the capacity region of the degraded broadcast channel and directly evaluates it for the Gaussian MIMO case by using two main tools. The first one is the generalized de Bruin identity due to [5] that states a connection between the differential entropy and the Fisher information matrix. The second tool we use is an inequality due to [6], [7] that gives a lower bound for the differential entropy in terms of the Fisher information matrix.

Finally, our technique used in this alternative proof can be useful in other vector Gaussian multi-terminal information theory problems when proving the optimality of Gaussian random vectors. In fact, we have used a variant of this technique in proving the secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel in [9].

II. CHANNEL MODEL AND MAIN RESULT

The degraded K-user Gaussian MIMO broadcast channel is defined by

$$\mathbf{Y}_k = \mathbf{X} + \mathbf{N}_k, \quad k = 1, \ldots, K \tag{1}$$

where $\mathbf{N}_k$ is Gaussian with covariance matrix $\boldsymbol{\Sigma}_k$, $k = 1, \ldots, K$. The covariance matrices $\{\boldsymbol{\Sigma}_k\}_{k=1}^{K}$ satisfy the following order

$$\mathbf{0} \prec \boldsymbol{\Sigma}_1 \preceq \ldots \preceq \boldsymbol{\Sigma}_K \tag{2}$$

The channel input is subject to a covariance constraint

$$E\left[\mathbf{X}\mathbf{X}^\top\right] \preceq \mathbf{S} \tag{3}$$

where we assume $\mathbf{S} \succ \mathbf{0}$. The covariance constraint in (3) subsumes many other constraints including the trace constraint in the sense that once the capacity region is found for the constraint in (3), the capacity region arising from the use of another constraint subsumed by (3) can be obtained by using this result [1].

The capacity region of the Gaussian MIMO broadcast channel is established in [1] for the most general case. For the degraded case, it is given as follows.

Theorem 1 ([1], Theorem 2) *The capacity region of the K-user degraded Gaussian MIMO broadcast channel is given by*

the union of rate tuples $(R_1, \ldots, R_K)$ satisfying

$$R_k \leq \frac{1}{2} \log \frac{|\sum_{i=1}^{k} \mathbf{K}_i + \boldsymbol{\Sigma}_k|}{|\sum_{i=1}^{k-1} \mathbf{K}_i + \boldsymbol{\Sigma}_k|} \tag{4}$$

where the union is over all positive semi-definite matrices $\{\mathbf{K}_i\}_{i=1}^{K}$ such that $\sum_{i=1}^{K} \mathbf{K}_i = \mathbf{S}$.

In the next section, we provide an alternative proof for this theorem for $K = 2$, and in Section IV we extend this proof to the case $K > 2$. In both cases, we directly evaluate the capacity region of the degraded broadcast channel which is stated in the following theorem, for the Gaussian MIMO channel at hand.

Theorem 2 ([8], Theorem 15.6.2) *The capacity region of the degraded broadcast channel is given by the union of rate tuples $(R_1, \ldots, R_K)$ satisfying*

$$R_k \leq I(U_k; Y_k | U_{k+1}), \quad k = 1, \ldots, K \tag{5}$$

where $U_{K+1} = \phi, U_1 = X$, and the union is over all $\{U_k\}_{k=2}^{K}, X$ such that

$$U_K \rightarrow \ldots U_2 \rightarrow X \rightarrow Y_1 \rightarrow \ldots \rightarrow Y_K \tag{6}$$

III. PROOF OF THEOREM 1 FOR $K = 2$

A. Background

We need some properties of the Fisher information and the differential entropy, which are provided next.

Definition 1 ([9], Definition 3) *Let $(\mathbf{U}, \mathbf{X})$ be an arbitrarily correlated length-n random vector pair with well-defined densities. The conditional Fisher information matrix of $\mathbf{X}$ given $\mathbf{U}$ is defined as*

$$\mathbf{J}(\mathbf{X}|\mathbf{U}) = E\left[\boldsymbol{\rho}(\mathbf{X}|\mathbf{U})\boldsymbol{\rho}(\mathbf{X}|\mathbf{U})^\top\right] \tag{7}$$

where the expectation is over the joint density $f(\mathbf{u}, \mathbf{x})$, and the conditional score function $\boldsymbol{\rho}(\mathbf{x}|\mathbf{u})$ is

$$\boldsymbol{\rho}(\mathbf{x}|\mathbf{u}) = \nabla \log f(\mathbf{x}|\mathbf{u}) \tag{8}$$

$$= \left[\begin{array}{ccc} \frac{\partial \log f(\mathbf{x}|\mathbf{u})}{\partial x_1} & \ldots & \frac{\partial \log f(\mathbf{x}|\mathbf{u})}{\partial x_n} \end{array}\right]^\top \tag{9}$$

We first present the conditional form of the Cramer-Rao inequality, which is proved in [9].

Lemma 1 ([9], Lemma 13) *Let $\mathbf{U}, \mathbf{X}$ be arbitrarily correlated random vectors with well-defined densities. Let the conditional covariance matrix of $\mathbf{X}$ be $\mathrm{Cov}(\mathbf{X}|\mathbf{U}) \succ \mathbf{0}$, then we have*

$$\mathbf{J}(\mathbf{X}|\mathbf{U}) \succeq \mathrm{Cov}(\mathbf{X}|\mathbf{U})^{-1} \tag{10}$$

which is satisfied with equality if $(\mathbf{U}, \mathbf{X})$ is jointly Gaussian with conditional covariance matrix $\mathrm{Cov}(\mathbf{X}|\mathbf{U})$.

The following lemma will be used in the upcoming proof. The unconditional version of this lemma, i.e., the case $\mathbf{T} = \phi$, is proved in Lemma 6 of [9].

Lemma 2 ([9], Lemma 6) *Let* $\mathbf{T}, \mathbf{U}, \mathbf{V}_1, \mathbf{V}_2$ *be random vectors such that* $(\mathbf{T}, \mathbf{U})$ *and* $(\mathbf{V}_1, \mathbf{V}_2)$ *are independent. Moreover, let* $\mathbf{V}_1, \mathbf{V}_2$ *be Gaussian random vectors with covariance matrices* $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ *such that* $\mathbf{0} \prec \mathbf{\Sigma}_1 \preceq \mathbf{\Sigma}_2$. *Then, we have*

$$\mathbf{J}^{-1}(\mathbf{U} + \mathbf{V}_2|\mathbf{T}) - \mathbf{\Sigma}_2 \succeq \mathbf{J}^{-1}(\mathbf{U} + \mathbf{V}_1|\mathbf{T}) - \mathbf{\Sigma}_1 \quad (11)$$

The following lemma will also be used in the upcoming proof.

Lemma 3 ([9], Lemma 8) *Let* $\mathbf{K}_1, \mathbf{K}_2$ *be positive semi-definite matrices satisfying* $\mathbf{0} \preceq \mathbf{K}_1 \preceq \mathbf{K}_2$, *and* $\mathbf{f}(\mathbf{K})$ *be a matrix-valued function such that* $\mathbf{f}(\mathbf{K}) \succeq \mathbf{0}$ *for* $\mathbf{K}_1 \preceq \mathbf{K} \preceq \mathbf{K}_2$. *Then, we have*

$$\int_{\mathbf{K}_1}^{\mathbf{K}_2} \mathbf{f}(\mathbf{K}) d\mathbf{K} \geq 0 \quad (12)$$

The following generalization of the de Bruin identity [10], [11] is due to [5], where the unconditional form of this identity, i.e., $\mathbf{U} = \phi$, is proved. Its generalization to this conditional form for an arbitrary $\mathbf{U}$ is rather straightforward, and is given in Lemma 16 of [9].

Lemma 4 ([9], Lemma 16) *Let* $(\mathbf{U}, \mathbf{X})$ *be an arbitrarily correlated random vector pair with finite second order moments, and also be independent of the random vector* $\mathbf{N}$ *which is zero-mean Gaussian with covariance matrix* $\mathbf{\Sigma}_N \succ \mathbf{0}$. *Then, we have*

$$\nabla_{\mathbf{\Sigma}_N} h(\mathbf{X} + \mathbf{N}|\mathbf{U}) = \frac{1}{2} \mathbf{J}(\mathbf{X} + \mathbf{N}|\mathbf{U}) \quad (13)$$

The following lemma is due to [6], [7] which lower bounds the differential entropy in terms of the Fisher information matrix.

Lemma 5 ([6], [7]) *Let* $(U, \mathbf{X})$ *be an* $(n+1)$-*dimensional random vector, where the conditional Fisher information matrix of* $\mathbf{X}$, *conditioned on* U, *exists. Then, we have*

$$h(\mathbf{X}|U) \geq \frac{1}{2} \log(2\pi e)^n |\mathbf{J}^{-1}(\mathbf{X}|U)| \quad (14)$$

In [6], [7], the unconditional version of this lemma, i.e., $U = \phi$, is provided. Its generalization to this conditional form is rather straightforward and the proof is omitted.

B. Proof for $K = 2$

We first rewrite the capacity region of the degraded broadcast channel given in Theorem 2 for two users as a union of rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(X; Y_1|U) \quad (15)$$
$$R_2 \leq I(U; Y_2) \quad (16)$$

where we dropped the subscript of the auxiliary random variable U_2 and denoted it simply as U. The involved random variables satisfy the Markov chain $U \rightarrow X \rightarrow Y_1 \rightarrow Y_2$. To obtain the capacity region of the degraded Gaussian

MIMO broadcast channel, we need to evaluate this region. In particular, we will show that the optimal random vector $(U, \mathbf{X})$ that exhausts this region is Gaussian, and the corresponding capacity region is given by the union of rate pairs (R_1, R_2) satisfying

$$R_1 \leq \frac{1}{2} \log \frac{|\mathbf{K} + \mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_1|} \quad (17)$$

$$R_2 \leq \frac{1}{2} \log \frac{|\mathbf{S} + \mathbf{\Sigma}_2|}{|\mathbf{K} + \mathbf{\Sigma}_2|} \quad (18)$$

where the union is over all $\mathbf{K}$ such that $\mathbf{0} \preceq \mathbf{K} \preceq \mathbf{S}$. We note that the region described by (17)-(18) comes from Theorem 1 by dropping the subscript of $\mathbf{K}_1$ and denoting it simply as $\mathbf{K}$.

We begin with the bound on R_2. Starting from (16), we get

$$R_2 \leq I(U; \mathbf{Y}_2) = h(\mathbf{Y}_2) - h(\mathbf{Y}_2|U) \quad (19)$$
$$\leq \frac{1}{2} \log(2\pi e)^n |\mathbf{S} + \mathbf{\Sigma}_2| - h(\mathbf{Y}_2|U) \quad (20)$$

where the inequality in (20) comes from the maximum entropy theorem [8]. We now bound $h(\mathbf{Y}_2|U)$ in (20). We first get an upper bound as

$$h(\mathbf{Y}_2|U) \leq h(\mathbf{Y}_2) \leq \frac{1}{2} \log(2\pi e)^n |\mathbf{S} + \mathbf{\Sigma}_2| \quad (21)$$

where the first inequality comes from the fact that conditioning cannot increase entropy, and the second inequality is due to the maximum entropy theorem [8]. Furthermore, using Lemma 5, we can get the following lower bound for $h(\mathbf{Y}_2|U)$

$$h(\mathbf{Y}_2|U) \geq \frac{1}{2} \log(2\pi e)^n |\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_2|U)| \quad (22)$$

We next define the following function

$$r(t) = \frac{1}{2} \log(2\pi e)^n |\mathbf{A}(t) + \mathbf{\Sigma}_2|, \quad 0 \leq t \leq 1 \quad (23)$$

where $\mathbf{A}(t)$ is given as

$$\mathbf{A}(t) = (1 - t) \left[\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_2|U) - \mathbf{\Sigma}_2 \right] + t\mathbf{S} \quad (24)$$

We first note that

$$\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_2|U) - \mathbf{\Sigma}_2 \preceq \mathrm{Cov}(\mathbf{X} + \mathbf{N}_2|U) - \mathbf{\Sigma}_2 \quad (25)$$
$$= \mathrm{Cov}(\mathbf{X}|U) \quad (26)$$
$$\preceq \mathrm{Cov}(\mathbf{X}) \quad (27)$$
$$\preceq \mathbf{S} \quad (28)$$

where (25) is a consequence of Lemma 1, and (27) comes from the fact that the conditional covariance matrix is smaller than the unconditional one in the positive semi-definite ordering sense. This implies that for any $0 \leq t \leq 1$, $\mathbf{A}(t)$ satisfies

$$\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_2|U) - \mathbf{\Sigma}_2 \preceq \mathbf{A}(t) \preceq \mathbf{S} \quad (29)$$

Using $r(t)$, bounds in (21) and (22) can be rewritten as

$$r(0) \leq h(\mathbf{Y}_2|U) \leq r(1) \quad (30)$$

Since $r(t)$ is continuous in t [9], due to the intermediate value

theorem, there exists a t^* such that

$$r(t^*) = h(\mathbf{Y}_2|U) = \frac{1}{2}\log(2\pi e)^n|\mathbf{A}(t^*) + \mathbf{\Sigma}_2| \qquad (31)$$

where $\mathbf{A}(t^*)$ satisfies (29). Plugging (31) into (20) yields

$$R_2 \leq \frac{1}{2}\log\frac{|\mathbf{S} + \mathbf{\Sigma}_2|}{|\mathbf{A}(t^*) + \mathbf{\Sigma}_2|} \qquad (32)$$

which is the desired bound on R_2 given in (18).

We now obtain the desired bound on R_1. To this end, using (29) and Lemma 2, we get

$$\mathbf{A}(t^*) \succeq \mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_2|U) - \mathbf{\Sigma}_2 \qquad (33)$$
$$\succeq \mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}|U) - \mathbf{\Sigma}_N \qquad (34)$$

for any Gaussian random vector $\mathbf{N}$ with covariance matrix $\mathbf{\Sigma}_N$ where $\mathbf{\Sigma}_N \preceq \mathbf{\Sigma}_2$. The order in (34) is equivalent to

$$(\mathbf{A}(t^*) + \mathbf{\Sigma}_N)^{-1} \preceq \mathbf{J}(\mathbf{X} + \mathbf{N}|U) \qquad (35)$$

Next, starting from (15), we get

$$R_1 \leq I(\mathbf{X}; \mathbf{Y}_1|U) = h(\mathbf{Y}_1|U) - \frac{1}{2}\log(2\pi e)^n|\mathbf{\Sigma}_1| \qquad (36)$$

$$= h(\mathbf{Y}_1|U) - h(\mathbf{Y}_2|U) + h(\mathbf{Y}_2|U) - \frac{1}{2}\log(2\pi e)^n|\mathbf{\Sigma}_1| \qquad (37)$$

$$= h(\mathbf{Y}_1|U) - h(\mathbf{Y}_2|U) + \frac{1}{2}\log|\mathbf{A}(t^*) + \mathbf{\Sigma}_2| - \frac{1}{2}\log|\mathbf{\Sigma}_1| \qquad (38)$$

$$= -\frac{1}{2}\int_{\mathbf{\Sigma}_1}^{\mathbf{\Sigma}_2} \mathbf{J}(\mathbf{X} + \mathbf{N}|U)d\mathbf{\Sigma}_N + \frac{1}{2}\log\frac{|\mathbf{A}(t^*) + \mathbf{\Sigma}_2|}{|\mathbf{\Sigma}_1|} \qquad (39)$$

$$\leq -\frac{1}{2}\int_{\mathbf{\Sigma}_1}^{\mathbf{\Sigma}_2} (\mathbf{A}(t^*) + \mathbf{\Sigma}_N)^{-1}d\mathbf{\Sigma}_N + \frac{1}{2}\log\frac{|\mathbf{A}(t^*) + \mathbf{\Sigma}_2|}{|\mathbf{\Sigma}_1|} \qquad (40)$$

$$= \frac{1}{2}\log\frac{|\mathbf{A}(t^*) + \mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_1|} \qquad (41)$$

where (38) is due to (31), (39) is obtained by using Lemma 4, and (40) is due to (35) and Lemma 3. Since (41) is the desired bound on R_1 given in (17), this completes the proof.

IV. EXTENSION TO THE K-USER CASE

We now extend our alternative proof presented in the previous section to the case $K > 2$. For that purpose, we need the following lemma due to [9] in addition to the tools introduced in Section III-A.

Lemma 6 ([9], Lemma 17) *Let* $(\mathbf{V}, \mathbf{U}, \mathbf{X})$ *be length-n random vectors with well-defined densities. Moreover, assume that the partial derivatives of* $f(\mathbf{u}|\mathbf{x}, \mathbf{v})$ *with respect to* x_i, $i = 1, \ldots, n$ *exist and satisfy*

$$\max_{1 \leq i \leq n} \left| \frac{\partial f(\mathbf{u}|\mathbf{v}, \mathbf{x})}{\partial x_i} \right| \leq g(\mathbf{u}) \qquad (42)$$

for some integrable function $g(\mathbf{u})$. *Then, if* $(\mathbf{V}, \mathbf{U}, \mathbf{X})$ *satisfy the Markov chain* $\mathbf{V} \rightarrow \mathbf{U} \rightarrow \mathbf{X}$, *we have*

$$\mathbf{J}(\mathbf{X}|\mathbf{U}) \succeq \mathbf{J}(\mathbf{X}|\mathbf{V}) \qquad (43)$$

First, following the proof in Section III-B, we can show the existence of a covariance matrix $\mathbf{A}_K$ such that

$$\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_K|U_K) - \mathbf{\Sigma}_K \preceq \mathbf{A}_K \preceq \mathbf{S} \qquad (44)$$

$$h(\mathbf{Y}_K|U_K) = \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_K + \mathbf{\Sigma}_K| \qquad (45)$$

$$h(\mathbf{Y}_{K-1}|U_K) \leq \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_K + \mathbf{\Sigma}_{K-1}| \qquad (46)$$

Since we have

$$h(\mathbf{Y}_K) \leq \frac{1}{2}\log(2\pi e)^n|\mathbf{S} + \mathbf{\Sigma}_K| \qquad (47)$$

from the maximum entropy theorem [8], we can get the desired bound on R_K as follows

$$R_K \leq I(U_K; \mathbf{Y}_K) \leq \frac{1}{2}\log\frac{|\mathbf{S} + \mathbf{\Sigma}_K|}{|\mathbf{A}_K + \mathbf{\Sigma}_K|} \qquad (48)$$

When $K = 2$ as in Section III-B, showing the existence of an $\mathbf{A}_K$ having the properties listed in (44)-(46) is sufficient to conclude the proof. However, when $K > 2$, we need an additional tool, which is Lemma 6, and using this tool we need to repeat this step until we are left with showing the desired bound on the first user's rate R_1. We now present the basic step that needs to be repeated. In particular, we now show the existence of a covariance matrix $\mathbf{A}_{K-1}$ such that

$$\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_{K-1}|U_{K-1}) - \mathbf{\Sigma}_{K-1} \preceq \mathbf{A}_{K-1} \preceq \mathbf{A}_K \qquad (49)$$

$$h(\mathbf{Y}_{K-1}|U_{K-1}) = \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_{K-1} + \mathbf{\Sigma}_{K-1}| \qquad (50)$$

$$h(\mathbf{Y}_{K-2}|U_{K-1}) \leq \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_{K-1} + \mathbf{\Sigma}_{K-2}| \qquad (51)$$

We first note the following order

$$\mathbf{A}_K \succeq \mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_K|U_K) - \mathbf{\Sigma}_K \qquad (52)$$
$$\succeq \mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_{K-1}|U_K) - \mathbf{\Sigma}_{K-1} \qquad (53)$$
$$\succeq \mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_{K-1}|U_{K-1}) - \mathbf{\Sigma}_{K-1} \qquad (54)$$

where (52) is due to (44), (53) comes from Lemma 2, and (54) follows from Lemma 6 as we can get

$$\mathbf{J}(\mathbf{X} + \mathbf{N}_{K-1}|U_{K-1}) \succeq \mathbf{J}(\mathbf{X} + \mathbf{N}_{K-1}|U_K) \qquad (55)$$

by noting the Markov chain $U_K \rightarrow U_{K-1} \rightarrow \mathbf{X} + \mathbf{N}_{K-1}$.

Next, we consider the following lower bound on $h(\mathbf{Y}_{K-1}|U_{K-1})$ which is due to Lemma 5

$$\frac{1}{2}\log(2\pi e)^n|\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_{K-1}|U_{K-1})| \leq h(\mathbf{Y}_{K-1}|U_{K-1}) \qquad (56)$$

Moreover, we can get the following upper bound

$$h(\mathbf{Y}_{K-1}|U_{K-1}) = h(\mathbf{Y}_{K-1}|U_{K-1}, U_K) \qquad (57)$$
$$\leq h(\mathbf{Y}_{K-1}|U_K) \qquad (58)$$
$$\leq \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_K + \mathbf{\Sigma}_{K-1}| \qquad (59)$$

where (57) is due to the Markov chain $U_K \rightarrow U_{K-1} \rightarrow \mathbf{Y}_{K-1}$, (58) comes from the fact that conditioning cannot increase entropy, and (59) is due to (46).

We now define the following function

$$r_{K-1}(t) = \frac{1}{2}\log(2\pi e)^n|\tilde{\mathbf{A}}_{K-1}(t) + \boldsymbol{\Sigma}_{K-1}|, \ 0 \le t \le 1 \quad (60)$$

where $\tilde{\mathbf{A}}_{K-1}(t)$ is given by

$$\tilde{\mathbf{A}}_{K-1}(t) = (1-t)\left[\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_{K-1}|U_{K-1}) - \boldsymbol{\Sigma}_{K-1}\right]$$
$$+ t\mathbf{A}_K \quad (61)$$

Using $r_{K-1}(t)$, we can recast bounds on $h(\mathbf{Y}_{K-1}|U_{K-1})$ in (56) and (59) as

$$r_{K-1}(0) \le h(\mathbf{Y}_{K-1}|U_{K-1}) \le r_{K-1}(1) \quad (62)$$

Since $r_{K-1}(t)$ is continuous in t [9], due to the intermediate value theorem, there exists a t^* such that $h(\mathbf{Y}_{K-1}|U_{K-1}) = r_{K-1}(t^*)$, i.e.,

$$h(\mathbf{Y}_{K-1}|U_{K-1}) = \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_{K-1} + \boldsymbol{\Sigma}_{K-1}| \quad (63)$$

where we define $\mathbf{A}_{K-1} = \tilde{\mathbf{A}}_{K-1}(t^*)$. Thus, we established (50). Furthermore, it is clear that we also have (49) because of (54), (61) and $0 \le t^* \le 1$.

We now show (51). To this end, we note the following order

$$\mathbf{A}_{K-1} \succeq \mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_{K-1}|U_{K-1}) - \boldsymbol{\Sigma}_{K-1} \quad (64)$$
$$\succeq \mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}|U_{K-1}) - \boldsymbol{\Sigma}_N \quad (65)$$

for any Gaussian $\mathbf{N}$ with covariance matrix $\boldsymbol{\Sigma}_N \preceq \boldsymbol{\Sigma}_{K-1}$ due to Lemma 2. The order in (65) is equivalent to

$$\mathbf{J}(\mathbf{X} + \mathbf{N}|U_{K-1}) \succeq (\mathbf{A}_{K-1} + \boldsymbol{\Sigma}_N)^{-1}, \ \ \boldsymbol{\Sigma}_N \preceq \boldsymbol{\Sigma}_{K-1} \quad (66)$$

We now consider $h(\mathbf{Y}_{K-2}|U_{K-1})$ as follows

$$h(\mathbf{Y}_{K-2}|U_{K-1}) = h(\mathbf{Y}_{K-2}|U_{K-1}) - h(\mathbf{Y}_{K-1}|U_{K-1})$$
$$+ h(\mathbf{Y}_{K-1}|U_{K-1}) \quad (67)$$
$$= -\frac{1}{2}\int_{\boldsymbol{\Sigma}_{K-2}}^{\boldsymbol{\Sigma}_{K-1}} \mathbf{J}(\mathbf{X} + \mathbf{N}|U)d\boldsymbol{\Sigma}_N$$
$$+ \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_{K-1} + \boldsymbol{\Sigma}_{K-1}| \quad (68)$$
$$\le -\frac{1}{2}\int_{\boldsymbol{\Sigma}_{K-2}}^{\boldsymbol{\Sigma}_{K-1}} (\mathbf{A}_{K-1} + \boldsymbol{\Sigma}_N)^{-1}d\boldsymbol{\Sigma}_N$$
$$+ \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_{K-1} + \boldsymbol{\Sigma}_{K-1}| \quad (69)$$
$$= \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_{K-1} + \boldsymbol{\Sigma}_{K-2}| \quad (70)$$

where (68) comes from Lemma 4 and (50), and (69) is due to (66) and Lemma 3. Thus, we showed (51) as well. Also, we can establish the desired bound R_{K-1} as follows

$$R_{K-1} \le I(U_{K-1}; \mathbf{Y}_{K-1}|U_K) \quad (71)$$
$$= h(\mathbf{Y}_{K-1}|U_K) - h(\mathbf{Y}_{K-1}|U_{K-1}) \quad (72)$$
$$\le \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_K + \boldsymbol{\Sigma}_{K-1}| - h(\mathbf{Y}_{K-1}|U_{K-1}) \quad (73)$$
$$= \frac{1}{2}\log\frac{|\mathbf{A}_K + \boldsymbol{\Sigma}_{K-1}|}{|\mathbf{A}_{K-1} + \boldsymbol{\Sigma}_{K-1}|} \quad (74)$$

where (72) comes from the Markov chain $U_K \to U_{K-1} \to Y_{K-1}$, (73) comes from (46), and (74) is due to (50).

As of now, we outlined the basic step that needs to be repeated until we are left with getting the desired bound on the first user's rate R_1. Following the analysis from (49) to (70), we can show the existence of covariance matrices $\mathbf{A}_k$, for $k = 2, \dots, K$, such that

$$\mathbf{J}^{-1}(\mathbf{X} + \mathbf{N}_k|U_k) - \boldsymbol{\Sigma}_k \preceq \mathbf{A}_k \preceq \mathbf{A}_{k+1} \quad (75)$$
$$h(\mathbf{Y}_k|U_k) = \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_k + \boldsymbol{\Sigma}_k| \quad (76)$$
$$h(\mathbf{Y}_{k-1}|U_k) \le \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_k + \boldsymbol{\Sigma}_{k-1}| \quad (77)$$

where we set $\mathbf{A}_{K+1} = \mathbf{S}$. Using these relations, we can get the bound on R_k for any $k = 2, \dots, K$, as

$$R_k \le I(U_k; \mathbf{Y}_k|U_{k+1}) = h(\mathbf{Y}_k|U_{k+1}) - h(\mathbf{Y}_k|U_k) \quad (78)$$
$$\le \frac{1}{2}\log(2\pi e)^n|\mathbf{A}_{k+1} + \boldsymbol{\Sigma}_k| - h(\mathbf{Y}_k|U_k) \quad (79)$$
$$= \frac{1}{2}\log\frac{|\mathbf{A}_{k+1} + \boldsymbol{\Sigma}_k|}{|\mathbf{A}_k + \boldsymbol{\Sigma}_k|} \quad (80)$$

where we set $U_{K+1} = \phi$. The equality in (78) comes from the Markov chain $U_{k+1} \to U_k \to \mathbf{Y}_k$, (79) is obtained by using (77), and (80) is due to (76). For $k = 1$, we can get the bound on R_1 as

$$R_1 \le I(\mathbf{X}; \mathbf{Y}_1|U_2) = h(\mathbf{Y}_1|U_2) - \frac{1}{2}\log(2\pi e)^n|\boldsymbol{\Sigma}_1| \quad (81)$$
$$\le \frac{1}{2}\log\frac{|\mathbf{A}_2 + \boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_1|} \quad (82)$$

where (82) comes from (77). Finally, we define $\mathbf{K}_k = \mathbf{A}_{k+1} - \mathbf{A}_k$ for $k = 1, \dots, K$ where $\mathbf{A}_1 = \phi, \mathbf{A}_{K+1} = \mathbf{S}$, and plug these into (80)-(82) which yields expressions in Theorem 1.

REFERENCES

[1] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inf. Theory*, 52(9):3936–3964, Sep. 2006.

[2] T. Liu and P. Viswanath. An extremal inequality motivated by multiterminal information theoretic problems. *IEEE Trans. Inf. Theory*, 53(5):1839–1851, May 2007.

[3] P. Bergmans. A simple converse for broadcast channels with additive white Gaussian noise. *IEEE Trans. Inf. Theory*, 20(3):279–280, Mar. 1974.

[4] K. Marton. A coding theorem for the discrete memoryless channels. *IEEE Trans. Inf. Theory*, 25(1):306–311, May 1979.

[5] D. P. Palomar and S. Verdu. Gradient of mutual information in linear vector Gaussian channels. *IEEE Trans. Inf. Theory*, 52(1):141–154, Jan. 2006.

[6] A. Dembo. Information inequalities and uncertainty principles. Tech. Rep., Dept. Statist., Stanford Univ., Stanford, CA., 1990.

[7] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Trans. Inf. Theory*, 37(6):1501–1518, Nov. 1991.

[8] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley & Sons, 2006. 2nd edition.

[9] E. Ekrem and S. Ulukus. The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel. Submitted to *IEEE Trans. Inf. Theory*, Mar. 2009. Also available at [arXiv:0903.3096].

[10] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, Jun. 1959.

[11] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Inf. Theory*, IT-11(2):267–271, Apr. 1965.