The Capacity of a MIMO Ricean Channel Is Monotonic in the Singular Values of the Mean

Daniel Hösli[1], Amos Lapidoth[2]

[1] Signal and Information Processing Laboratory, Swiss Federal Institute of Technology (ETH) Zurich, hoesli@isi.ee.ethz.ch

[2] Signal and Information Processing Laboratory, Swiss Federal Institute of Technology (ETH) Zurich, lapidoth@isi.ee.ethz.ch

Abstract

We consider a discrete-time memoryless Multiple-Input Multiple-Output (MIMO) fading channel where the fading matrix can be written as the sum of a deterministic (line-of-sight) matrix D and a random matrix $\widetilde{\mathbb{H}}$ whose entries are IID zero-mean unit-variance complex circularly-symmetric Gaussian random variables. It is demonstrated that if the realization of the fading matrix is known at the receiver but not at the transmitter, then the capacity of this channel under an average power constraint is monotonically non-decreasing in the singular values of D. This complements a recent result of Kim and Lapidoth [1] demonstrating the monotonicity of the mutual information corresponding to isotropically distributed Gaussian input vectors. We also show that the optimal covariance matrix of the Gaussian input vector has the same eigenvectors as $\mathsf{D}^\dagger\mathsf{D}$.

1 Introduction

We consider a discrete-time memoryless channel whose output $\mathbf{Y}$ takes value in the m-dimensional complex Euclidean space $\mathbb{C}^m$ and is given by

$$\mathbf{Y} = \mathbb{H}\mathbf{x} + \mathbf{Z} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^n$ is the channel input; the random vector $\mathbf{Z}$ has a $\mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \sigma^2\mathsf{I}_m\big)$ distribution for some $\sigma^2 > 0$; and the random matrix $\mathbb{H} \in \mathbb{C}^{m\times n}$ can be written as

$$\mathbb{H} = \mathsf{D} + \widetilde{\mathbb{H}} \tag{2}$$

where $\mathsf{D} \in \mathbb{C}^{m\times n}$ is a deterministic $m \times n$ matrix and where the $m \cdot n$ random components of the random matrix $\widetilde{\mathbb{H}} \in \mathbb{C}^{m\times n}$ are IID $\mathcal{N}_{\mathbb{C}}(0,1)$. Here $\mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathsf{K})$ denotes the zero-mean circularly-symmetric multivariate Gaussian distribution of covariance matrix K, and I_m denotes the $m\times m$ identity matrix. (In the scalar case we write $\mathcal{N}_{\mathbb{C}}(0,1)$ rather than $\mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathsf{I}_1)$.) It is assumed that $\widetilde{\mathbb{H}}$ and $\mathbf{Z}$ are independent, and that their joint law does not depend on the input $\mathbf{x}$.

This channel model represents a narrowband fading multiple-input multiple-output (MIMO) system with perfect channel state information at the receiver.

We shall consider the capacity of this channel when the realization of the fading matrix $\mathbb{H}$ is known to the receiver, but only its probability law is known at the transmitter. We shall assume that the transmitted signal is subject to an average power constraint

$$\mathsf{E}\big[\mathbf{X}^\dagger\mathbf{X}\big] \le \mathcal{E}_{\mathrm{s}} \tag{3}$$

where we use $\mathsf{A}^\dagger$ to denote the Hermitian conjugate of A. Making $\mathcal{E}_{\mathrm{s}}$ and σ^2 implicit, we denote the capacity of this channel by $C(\mathsf{D})$:

$$C(\mathsf{D}) = \sup_{\mathbf{X}} I\left(\mathbf{X}; \mathbb{H}\mathbf{X} + \mathbf{Z}, \mathbb{H}\right) \tag{4}$$

where $I(\mathbf{X}; \mathbb{H}\mathbf{X}+\mathbf{Z}, \mathbb{H})$ denotes the mutual information between $\mathbf{X}$ and the pair $(\mathbb{H}\mathbf{X} + \mathbf{Z}, \mathbb{H})$, and where the supremum is over all random vectors $\mathbf{X} \in \mathbb{C}^n$ that are independent of $(\mathbb{H},\mathbf{Z})$ and that satisfy (3).

Expressing mutual information as a difference between differential entropies and noting that of all random vectors $\mathbf{Y}$ of a given second moment matrix $\mathsf{E}\big[\mathbf{Y}\mathbf{Y}^\dagger\big]$ differential entropy is maximized by the zero-mean multivariate circularly-symmetric Gaussian distribution, one can show that

$$I\left(\mathbf{X}; \mathbb{H}\mathbf{X} + \mathbf{Z}, \mathbb{H}\right) \le I\left(\mathbf{X}_{\mathrm{G}}; \mathbb{H}\mathbf{X}_{\mathrm{G}} + \mathbf{Z}, \mathbb{H}\right) \tag{5}$$

where $\mathbf{X}_{\mathrm{G}} \sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \mathsf{E}\big[\mathbf{X}\mathbf{X}^\dagger\big]\big)$, and where equality holds only if $\mathbf{X}$ is zero-mean circularly-symmetric Gaussian. That is, of all inputs to the channel of a given second moment matrix, the zero-mean circularly-symmetric Gaussian input vector maximizes mutual information.

Evaluating the mutual information for such a Gaussian input explicitly we obtain that

$$I\left(\mathbf{X}_{\mathrm{G}}; \mathbb{H}\mathbf{X}_{\mathrm{G}} + \mathbf{Z}, \mathbb{H}\right) = \mathsf{E}_{\mathbb{H}}\left[\log\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\mathbb{H}\mathsf{K}\mathbb{H}^\dagger\right)\right] \tag{6}$$

where $\mathsf{K} = \mathsf{E}\Big[\mathbf{X}_{\mathrm{G}}\mathbf{X}_{\mathrm{G}}^\dagger\Big]$.

Denoting the the mutual information corresponding to a $\mathcal{N}_{\mathbb{C}}(\mathbf{0},\mathsf{K})$ input and to a mean matrix D by $J(\mathsf{K},\mathsf{D})$, we obtain from (6)

$$J(\mathsf{K},\mathsf{D}) = \mathsf{E}_{\widetilde{\mathbb{H}}}\left[\log\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}(\widetilde{\mathbb{H}} + \mathsf{D})\mathsf{K}(\widetilde{\mathbb{H}} + \mathsf{D})^\dagger\right)\right] \tag{7}$$

and we can express the capacity $C(\mathsf{D})$ as

$$C(\mathsf{D}) = \sup_{\mathsf{K}} J(\mathsf{K}, \mathsf{D}) \tag{8}$$

where the supremum is over all positive semi-definite matrices K satisfying the trace constraint

$$\mathrm{tr}\,(\mathsf{K}) \leq \mathcal{E}_{\mathrm{s}}. \tag{9}$$

In fact, the set of all positive semi-definite matrices K satisfying (9) is compact and $J(\mathsf{K}, \mathsf{D})$ is (strictly) concave in K so that the supremum in (8) is achieved. Furthermore, it can be shown that the maximizing K achieves (9) with equality and is unique.

It has recently been shown by Kim and Lapidoth [1] that if K in (7) is proportional to the identity matrix, i.e., if one only considers isotropic Gaussian inputs, then the resulting mutual information is monotonic in the singular values of the matrix D. At ISIT'03, following the presentation of that result, H. Boche asked [2] whether a similar result holds for channel capacity. Here we shall answer this affirmatively.

Theorem 1: Consider two line-of-sight matrices $\mathsf{D}^\star$ and D and let $\{\sqrt{d_i^\star}\}$ and $\{\sqrt{d_i}\}$ be the decreasingly ordered singular values of the matrices $\mathsf{D}^\star$ and D, respectively. If $d_i^\star \geq d_i$, for every $i \in \{1, \ldots, \min\{m, n\}\}$, then $C(\mathsf{D}^\star) \geq C(\mathsf{D})$.

The proof of this theorem is given in Section 4. It is simplified by the following theorem demonstrating that the capacity-achieving covariance matrix of the input vector $\mathbf{X}_{\mathrm{G}}$ has the same eigenvectors as $\mathsf{D}^\dagger\mathsf{D}$.

Theorem 2: If the line-of-sight matrix D is "diagonal"[1], then $C(\mathsf{D})$ is achieved by a Gaussian input vector of a diagonal covariance matrix.

More generally, the set of eigenvectors of the capacity-achieving covariance matrix must coincide with those of $\mathsf{D}^\dagger\mathsf{D}$.

The proof of Theorem 2 is given in Section 3. It has recently been brought to our attention that this result was independently derived by Venkatesan *et al.* [3].

2 Preliminary Results

For Theorem 1 to be meaningful we must show that the capacity $C(\mathsf{D})$ depends on the mean matrix D only via its singular values. This, (see Lemma 2) as well as some other preliminary results, will be proved in this section.

We begin by noting that the law of the additive isotropic Gaussian noise $\mathbf{Z}$ is invariant to deterministic rotations

$$\mathsf{U}\mathbf{Z} \stackrel{\mathscr{L}}{=} \mathbf{Z}, \quad \mathsf{U}\mathsf{U}^\dagger = \mathsf{I}_m \tag{10}$$

and that the law of $\widetilde{\mathbb{H}}$ is similarly invariant under left and right deterministic rotations:

$$\mathsf{U}\widetilde{\mathbb{H}}\mathsf{V}^\dagger \stackrel{\mathscr{L}}{=} \widetilde{\mathbb{H}}, \quad \mathsf{U}\mathsf{U}^\dagger = \mathsf{I}_m, \mathsf{V}\mathsf{V}^\dagger = \mathsf{I}_n. \tag{11}$$

[1]By a "diagonal" $m \times n$ matrix we refer to a matrix whose (i, j) component is zero whenever $i \neq j$, for $1 \leq i \leq m$ and $1 \leq j \leq n$.

Here $\stackrel{\mathscr{L}}{=}$ denotes equality in law.

These symmetries imply the following.

Lemma 1: The mutual information $I(\mathbf{X}; \mathbb{H}\mathbf{X}+\mathbf{Z}, \mathbb{H})$ of the MIMO Ricean channel (1) with perfect side-information at the receiver and with line-of-sight matrix D induced by an input $\mathbf{X}$ equals the mutual information of a MIMO Ricean channel with line-of-sight matrix $\hat{\mathsf{D}} = \mathsf{U}\mathsf{D}\mathsf{V}^\dagger$ induced by the input $\hat{\mathbf{X}} = \mathsf{V}\mathbf{X}$, where $\mathsf{U} \in \mathbb{C}^{m\times m}$ and $\mathsf{V} \in \mathbb{C}^{n\times n}$ are arbitrary deterministic unitary matrices.

Proof: Let $\hat{\mathsf{D}} = \mathsf{U}\mathsf{D}\mathsf{V}^\dagger$ where $\mathsf{U} \in \mathbb{C}^{m\times m}$ and $\mathsf{V} \in \mathbb{C}^{n\times n}$ are deterministic unitary matrices. Then

$$\begin{aligned}
I\big(\ &\mathbf{X}\ ; (\widetilde{\mathbb{H}} + \mathsf{D})\mathbf{X} + \mathbf{Z}, \widetilde{\mathbb{H}}\big) \\
&= I\left(\mathbf{X}; (\widetilde{\mathbb{H}} + \mathsf{U}^\dagger\hat{\mathsf{D}}\mathsf{V})\mathbf{X} + \mathbf{Z}, \widetilde{\mathbb{H}}\right) \\
&= I\left(\mathbf{X}; \mathsf{U}\big((\widetilde{\mathbb{H}} + \mathsf{U}^\dagger\hat{\mathsf{D}}\mathsf{V})\mathbf{X} + \mathbf{Z}\big), \widetilde{\mathbb{H}}\right) \qquad (12) \\
&= I\left(\mathbf{X}; (\mathsf{U}\widetilde{\mathbb{H}}\mathsf{V}^\dagger + \hat{\mathsf{D}})\mathsf{V}\mathbf{X} + \mathsf{U}\mathbf{Z}, \widetilde{\mathbb{H}}\right) \\
&= I\left(\mathsf{V}\mathbf{X}; (\mathsf{U}\widetilde{\mathbb{H}}\mathsf{V}^\dagger + \hat{\mathsf{D}})\mathsf{V}\mathbf{X} + \mathsf{U}\mathbf{Z}, \mathsf{U}\widetilde{\mathbb{H}}\mathsf{V}^\dagger\right) \qquad (13) \\
&= I\left(\hat{\mathbf{X}}; (\widetilde{\mathbb{H}} + \hat{\mathsf{D}})\hat{\mathbf{X}} + \mathbf{Z}, \widetilde{\mathbb{H}}\right) \qquad (14)
\end{aligned}$$

where (12) (respectively (13)) follows because rotating the output (respectively input) does not change mutual information, and where (14) follows by defining $\hat{\mathbf{X}} = \mathsf{V}\mathbf{X}$ and from (10) and (11). ■

Applying Lemma 1 to the zero-mean Gaussian inputs we obtain:

Corollary 1: The mutual information $J(\mathsf{K}, \mathsf{D})$ of the MIMO Ricean channel (1) with line-of-sight matrix D induced by a $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{K})$ input equals the mutual information of the MIMO Ricean channel with line-of-sight matrix $\hat{\mathsf{D}} = \mathsf{U}\mathsf{D}\mathsf{V}^\dagger$ induced by a $\mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \mathsf{V}\mathsf{K}\mathsf{V}^\dagger\big)$ input, i.e.,

$$J(\mathsf{K}, \mathsf{D}) = J(\mathsf{V}\mathsf{K}\mathsf{V}^\dagger, \mathsf{U}\mathsf{D}\mathsf{V}^\dagger). \tag{15}$$

Proof: Follows from the above lemma by considering the input $\mathbf{X}$ to be $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{K})$ distributed whence $\hat{\mathbf{X}} = \mathsf{V}\mathbf{X}$ is $\mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \mathsf{V}\mathsf{K}\mathsf{V}^\dagger\big)$ distributed. ■

The preceding corollary will now be used to prove the following lemma, which makes Theorem 1 meaningful:

Lemma 2: If D can be written in the form $\mathsf{D} = \mathsf{U}\hat{\mathsf{D}}\mathsf{V}^\dagger$ where $\mathsf{U} \in \mathbb{C}^{m\times m}$ and $\mathsf{V} \in \mathbb{C}^{n\times n}$ are deterministic unitary matrices then

$$C(\mathsf{D}) = C(\hat{\mathsf{D}}). \tag{16}$$

In particular, by the singular value decomposition theorem, $C(\mathsf{D})$ depends on D only via its singular values.

Moreover, the covariance matrix $\hat{\mathsf{K}}^*$ achieves $C(\hat{\mathsf{D}})$ if, and only if, $\mathsf{K}^* = \mathsf{V}\hat{\mathsf{K}}^*\mathsf{V}^\dagger$ achieves $C(\mathsf{D})$. Likewise, K^* achieves $C(\mathsf{D})$ if, and only if, $\hat{\mathsf{K}}^* = \mathsf{V}^\dagger\mathsf{K}^*\mathsf{V}$ achieves $C(\hat{\mathsf{D}})$.

Proof: Assume that D can be written as $\mathsf{D} = \mathsf{U}\hat{\mathsf{D}}\mathsf{V}^\dagger$ where $\mathsf{U} \in \mathbb{C}^{m\times m}$ and $\mathsf{V} \in \mathbb{C}^{n\times n}$ are deterministic unitary matrices. By (8) capacity can be expressed as

$$
\begin{aligned}
C(\mathsf{D}) &= \sup_{\mathsf{K}} J(\mathsf{K}, \mathsf{D}) \\
&= \sup_{\mathsf{K}} J(\mathsf{K}, \mathsf{U}\hat{\mathsf{D}}\mathsf{V}^\dagger) \\
&= \sup_{\mathsf{K}} J(\mathsf{V}^\dagger\mathsf{K}\mathsf{V}, \hat{\mathsf{D}}) \qquad (17) \\
&= \sup_{\hat{\mathsf{K}}} J(\hat{\mathsf{K}}, \hat{\mathsf{D}}) \qquad (18) \\
&= C(\hat{\mathsf{D}})
\end{aligned}
$$

where (17) follows from (15). Here (18) follows by noting that since V is unitary it follows that for any K,

$$\operatorname{tr}\left(\mathsf{V}^\dagger\mathsf{K}\mathsf{V}\right) = \operatorname{tr}(\mathsf{K}) \qquad (19)$$

so that the mapping $\mathsf{K} \mapsto \mathsf{V}^\dagger\mathsf{K}\mathsf{V}$ is a bijective mapping from the set of all positive semi-definite matrices K with $\operatorname{tr}(\mathsf{K}) \leq \mathcal{E}_\mathrm{s}$ onto itself.

The relationship between the input that achieves $C(\mathsf{D})$ and the one that achieves $C(\hat{\mathsf{D}})$ follows from (15). ■

3 Structure of the Optimal Input Covariance Matrix

In this section we prove Theorem 2. We only consider the case where D is “diagonal”. The more general case follows from the second part of Lemma 2. Starting from any covariance matrix $\mathsf{K} \in \mathbb{C}^{n\times n}$ satisfying (9) and picking any index $1 \leq k \leq n$, we will show that one can null all the off-diagonal elements in the k-th row and k-th column to produce a positive semi-definite matrix $\bar{\mathsf{K}} \in \mathbb{C}^{n\times n}$ satisfying (9) and for which the corresponding mutual information is at least as high as for K, and is in fact strictly higher if $\bar{\mathsf{K}} \neq \mathsf{K}$. Repeating this procedure n times will prove that an optimal covariance matrix must be diagonal. The details follow.

Let $\mathsf{D} \in \mathbb{C}^{m\times n}$ be “diagonal”, and let $\mathsf{K} \in \mathbb{C}^{n\times n}$ be an arbitrary positive semi-definite matrix fulfilling the power constraint $\operatorname{tr}(\mathsf{K}) \leq \mathcal{E}_\mathrm{s}$. Fix some $1 \leq k \leq n$ and let the $n \times n$ unitary matrix Π be a diagonal matrix all of whose diagonal entries are 1 except for the k-th entry, which is -1. Let $\tilde{\Pi}$ be an $m \times m$ unitary matrix such that

$$\tilde{\Pi}^\dagger \mathsf{D} \Pi = \mathsf{D}. \qquad (20)$$

For example, in the case where $m \geq k$ the matrix $\tilde{\Pi}$ can be chosen as a diagonal matrix all of whose diagonal entries are 1 except for the k-th entry, which is -1. In the case where $m < k$ the matrix $\tilde{\Pi}$ can be chosen as the identity matrix.

Consider now the positive semi-definite matrix $\tilde{\mathsf{K}} = \Pi\mathsf{K}\Pi^\dagger$ whose entries are identical to those of K except that its off-diagonal entries in the k-th row and in the k-th column have changed sign. Since Π is unitary, the trace of $\tilde{\mathsf{K}}$ is identical to that of K and thus, in particular, $\operatorname{tr}\left(\tilde{\mathsf{K}}\right) \leq \mathcal{E}_\mathrm{s}$. We shall next show that $J(\tilde{\mathsf{K}}, \mathsf{D}) = J(\mathsf{K}, \mathsf{D})$. Indeed,

$$
\begin{aligned}
J(\tilde{\mathsf{K}}, \mathsf{D}) &= J(\Pi\mathsf{K}\Pi^\dagger, \mathsf{D}) \qquad (21) \\
&= J(\mathsf{K}, \tilde{\Pi}^\dagger\mathsf{D}\Pi) \qquad (22) \\
&= J(\mathsf{K}, \mathsf{D}) \qquad (23)
\end{aligned}
$$

where we have used the definition of $\tilde{\mathsf{K}}$ in (21); Corollary 1 in (22); and (20) in (23).

Consider now the matrix

$$\bar{\mathsf{K}} = \frac{1}{2}\left(\mathsf{K} + \tilde{\mathsf{K}}\right). \qquad (24)$$

Its entries are identical to those of K except that its off-diagonal elements in the k-th row and in the k-th column are zero. In particular, its trace is identical to that of K and thus satisfies $\operatorname{tr}\left(\bar{\mathsf{K}}\right) \leq \mathcal{E}_\mathrm{s}$. Moreover, by the (strict) concavity of $J(\mathsf{K}, \mathsf{D})$ in K it follows that the covariance matrix $\bar{\mathsf{K}}$ achieves at least as high a mutual information as K while still meeting the power constraint $\operatorname{tr}\left(\bar{\mathsf{K}}\right) \leq \mathcal{E}_\mathrm{s}$:

$$
\begin{aligned}
J(\bar{\mathsf{K}}, \mathsf{D}) &\geq \frac{1}{2}\left(J(\mathsf{K}, \mathsf{D}) + J(\tilde{\mathsf{K}}, \mathsf{D})\right) \qquad (25) \\
&= J(\mathsf{K}, \mathsf{D}) \qquad (26)
\end{aligned}
$$

where (25) holds with equality if, and only if, $\bar{\mathsf{K}} = \mathsf{K}$ and where (26) follows from (23).

We have thus shown that nulling the off-diagonal elements in any row k and column k of any covariance matrix increases (strictly if they were not all zero to start with) the mutual information, and Theorem 2 follows.

4 Outline of the Proof of Theorem 1

In this section we prove the main result of this paper, namely, the monotonicity of capacity as a function of the singular values of the line-of-sight matrix D. Since $C(\mathsf{D})$ depends on D only via its singular values, it suffices to concentrate on the case where D is “diagonal”. The proof of the theorem hinges on the following two lemmas:

Lemma 3 (Anderson’s Theorem, [4]): Let $\mathcal{E} \subseteq \mathbb{C}^m$ be convex and symmetric about the origin. Let

$$f:\ \mathbb{C}^m \to \mathbb{R}_+,\ \boldsymbol{\xi} \mapsto f(\boldsymbol{\xi}) \qquad (27)$$

be a non-negative function such that

- $f(-\boldsymbol{\xi}) = f(\boldsymbol{\xi})$;
- the set $\{\boldsymbol{\xi} :\ f(\boldsymbol{\xi}) \geq s\}$ is convex for every $s > 0$;
- $\int_{\mathcal{E}} f(\boldsymbol{\xi})\, \mathrm{d}\boldsymbol{\xi} < \infty$.

Then for any vector $\boldsymbol{\chi} \in \mathbb{C}^m$ and any $0 \leq c \leq 1$

$$\int_{\mathcal{E}} f(\boldsymbol{\xi} + c\boldsymbol{\chi})\, \mathrm{d}\boldsymbol{\xi} \geq \int_{\mathcal{E}} f(\boldsymbol{\xi} + \boldsymbol{\chi})\, \mathrm{d}\boldsymbol{\xi}. \qquad (28)$$

Lemma 4: Let $m, n \geq 1$ be given. Let $1 \leq k \leq n$. Let the diagonal matrix $\mathsf{L} = \operatorname{diag}(()\, l_1, \ldots, l_n) \in \mathbb{C}^{n\times n}$ have non-negative diagonal elements and also be fixed. For any $\boldsymbol{\xi} \in \mathbb{C}^m$ and $n-1$ vectors $\mathbf{a}_i \in \mathbb{C}^m, i = 1, \ldots, k-1, k+1, \ldots, n$ let $\mathsf{A}[\boldsymbol{\xi} : (\mathbf{a}_i)]$ denote the $m \times n$ complex matrix whose columns are

$$\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \boldsymbol{\xi}, \mathbf{a}_{k+1}, \ldots \mathbf{a}_n.$$

Then for any fixed vectors $\{\mathbf{a}_i\}$ the function

$$g(\boldsymbol{\xi}) = \det\Big(\mathsf{I}_m + \mathsf{A}[\boldsymbol{\xi} : (\mathbf{a}_i)]\ \mathsf{L}\ \mathsf{A}^\dagger[\boldsymbol{\xi} : (\mathbf{a}_i)]\Big) \quad (29)$$

is symmetric and convex in $\boldsymbol{\xi} \in \mathbb{C}^m$.

In other words, $\det(\mathsf{I}_m + \mathsf{ALA}^\dagger)$ is a symmetric convex function of each of the columns of A when the other columns are held fixed.

Proof: We begin by noting that it suffices to treat the case where L is the identity matrix. To see this, note that if we define $\mathsf{L}^{1/2} = \operatorname{diag}\left(()\sqrt{l_1}, \ldots, \sqrt{l_n}\right)$ then

$$\begin{aligned}\det(\mathsf{I}_m + \mathsf{ALA}^\dagger) &= \det(\mathsf{I}_m + \mathsf{AL}^{1/2}\mathsf{L}^{1/2}\mathsf{A}^\dagger)\\ &= \det\big(\mathsf{I}_m + (\mathsf{AL}^{1/2})(\mathsf{AL}^{1/2})^\dagger\big)\end{aligned}$$

and

$$\mathsf{A}\left[\boldsymbol{\xi} : (\mathbf{a}_i)\right] \mathsf{L}^{1/2} = \mathsf{A}\left[\sqrt{l_k}\boldsymbol{\xi} : \left(\sqrt{l_i}\mathbf{a}_i\right)\right].$$

Thus, the presence of the diagonal matrix L is tantamount to a mere scaling of the columns of A, a scaling that does not affect the symmetry and the convexity.

We shall thus show that if $\mathsf{M} \in \mathbb{C}^{m\times n}$ is a complex matrix and k is an integer such that $1 \le k \le n$, then

$$\det(\mathsf{I}_m + \mathsf{MM}^\dagger) \quad (30)$$

is a symmetric and convex function of $\mathbf{m}_k$, the k-th column of M. Expressing $\mathsf{MM}^\dagger$ as $\mathsf{MM}^\dagger = \sum_{i=1}^n \mathbf{m}_i\mathbf{m}_i^\dagger$ where $\mathsf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_n]$ we have

$$\begin{aligned}\det\left(\mathsf{I}_m + \mathsf{MM}^\dagger\right) &= \det\left(\mathbf{m}_k\mathbf{m}_k^\dagger + \mathsf{I}_m + \sum_{\substack{i=1\\ i\neq k}}^{n} \mathbf{m}_i\mathbf{m}_i^\dagger\right)\\ &= \det\left(\mathbf{m}_k\mathbf{m}_k^\dagger + \mathsf{V\Lambda V}^\dagger\right) && (31)\\ &= \det\left(\mathsf{V}^\dagger(\mathbf{m}_k\mathbf{m}_k^\dagger + \mathsf{V\Lambda V}^\dagger)\mathsf{V}\right) && (32)\\ &= \det\left((\mathsf{V}^\dagger\mathbf{m}_k)(\mathsf{V}^\dagger\mathbf{m}_k)^\dagger + \mathsf{\Lambda}\right)\\ &= \det\left(\boldsymbol{\chi}\boldsymbol{\chi}^\dagger + \mathsf{\Lambda}\right) && (33)\\ &= \det\left((\mathsf{\Lambda}^{-1/2}\boldsymbol{\chi})(\mathsf{\Lambda}^{-1/2}\boldsymbol{\chi})^\dagger + \mathsf{I_m}\right)\det(\mathsf{\Lambda})\\ &= \left(1 + \|\mathsf{\Lambda}^{-1/2}\boldsymbol{\chi}\|^2\right)\det(\mathsf{\Lambda}) && (34)\\ &= \left(1 + \sum_{i=1}^{m}\frac{1}{\lambda_i}|\chi_i|^2\right)\prod_{j=1}^{m}\lambda_j\\ &= \prod_{i=1}^{m}\lambda_i + \sum_{i=1}^{m}|\chi_i|^2\prod_{\substack{j=1\\ j\neq i}}^{m}\lambda_j\end{aligned}$$

where in (31) we have used that $\mathsf{I}_m + \sum_{\substack{i=1\\ i\neq k}}^{n}\mathbf{m}_i\mathbf{m}_i^\dagger$ is a positive definite matrix that is independent of $\mathbf{m}_k$ and that can be written as $\mathsf{V\Lambda V}^\dagger$ with $\mathsf{V} \in \mathbb{C}^{m\times m}$ unitary and $\mathsf{\Lambda} = \operatorname{diag}\left(()\lambda_1, \ldots, \lambda_\mathsf{m}\right) \in \mathbb{R}^{\mathsf{m}\times\mathsf{m}}$ a positive definite diagonal matrix with decreasingly ordered diagonal elements $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_m > 0$; equality (32) follows from the fact that the determinant of a square matrix is unitarily invariant, i.e., $\det(\mathsf{V}^\dagger\mathsf{BV}) = \det(\mathsf{B})$, where $\mathsf{B} \in \mathbb{C}^{m\times m}$ is any complex square matrix and $\mathsf{V} \in \mathbb{C}^{m\times m}$ is unitary; and in (33) we have defined $\boldsymbol{\chi} = \mathsf{V}^\dagger\mathbf{m}_k$ where $\boldsymbol{\chi} = [\chi_1, \ldots, \chi_m]^\mathsf{T} \in \mathbb{C}^m$. The above chain of equalities shows that $\det(\mathsf{I}_m + \mathsf{MM}^\dagger)$ is a symmetric and convex function of $\boldsymbol{\chi}$. But since $\boldsymbol{\chi}$ is a linear function of $\mathbf{m}_k$ we conclude that it is also symmetric and convex in $\mathbf{m}_k$.

■

We now sketch the proof of Theorem 1 using Lemma 3 and Lemma 4. Let $\mathsf{D}^\star$ and D be two "diagonal" matrices with decreasingly ordered "diagonal" elements $\{\sqrt{d_i^\star}\}$ and $\{\sqrt{d_i}\}$, respectively, with $d_i^\star \ge d_i$ for $i \in \{1, \ldots, \min\{m, n\}\}$. To prove the theorem it suffices to consider the case

$$d_k^\star > d_k \quad \text{and} \quad d_i^\star = d_i,\ i \neq k \quad (35)$$

since the general case follows from the fact that

$$\begin{aligned}\left(\sqrt{d_1^\star}, \sqrt{d_2^\star}, \ldots, \sqrt{d_l^\star}\right) &\ge \left(\sqrt{d_1}, \sqrt{d_2^\star}, \ldots, \sqrt{d_l^\star}\right)\\ &\ge \left(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_l^\star}\right)\\ &\ge \ldots\\ &\ge \left(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_l}\right)\end{aligned}$$

where $l = \min\{m, n\}$ and where the inequalities hold elementwise, so that we have to prove only one of the series of inequalities. From Theorem 2 we know that if D is "diagonal" the capacity-achieving K is diagonal. We will therefore only consider diagonal matrices K, and we will show that in this case

$$\begin{aligned}&\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)^\dagger\right)\\ &\quad \ge^{\text{st}} \det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)^\dagger\right) \quad (36)\end{aligned}$$

for any diagonal positive semi-definite matrix K, i.e., that the LHS of (36) is stochastically larger [5] than its RHS:

$$\begin{aligned}&\Pr\left[\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)^\dagger\right) \le s\right]\\ &\le \Pr\left[\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)^\dagger\right) \le s\right]\end{aligned} \quad (37)$$

for every $s > 0$. From this it will follow by the monotonicity of the logarithm function that

$$\begin{aligned}&\mathsf{E}_{\widetilde{\mathbb{H}}}\left[\log\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)^\dagger\right)\right]\\ &\ge \mathsf{E}_{\widetilde{\mathbb{H}}}\left[\log\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)^\dagger\right)\right]\end{aligned} \quad (38)$$

i.e., that $J(\mathsf{K}, \mathsf{D}^\star) \ge J(\mathsf{K}, \mathsf{D})$, and the theorem will follow by (8).

We now proceed to outline the proof of (37). For any $s > 0$ define the set of matrices

$$\mathcal{F}_s = \left\{ \mathsf{A} \in \mathbb{C}^{m\times n} :\ \det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\mathsf{AKA}^\dagger\right) \le s \right\} \tag{39}$$

and the parameterized set of vectors

$$\mathcal{F}_s\left((\mathbf{a}_i)_{i\neq k}\right) = \left\{ \boldsymbol{\xi} \in \mathbb{C}^m :\ \mathsf{A}\left[\boldsymbol{\xi} : (\mathbf{a}_i)\right] \in \mathcal{F}_s \right\}. \tag{40}$$

(Recall the notation $\mathsf{A}[\boldsymbol{\xi} : (\mathbf{a}_i)]$, previously introduced in Lemma 4.)

Using Lemma 4 one can show that for any $(\mathbf{a}_i)_{i\neq k}$ and $s > 0$ the set $\mathcal{F}_s((\mathbf{a}_i)_{i\neq k})$ is symmetric and convex. Thus, it meets the requirements imposed on $\mathcal{E}$ in Lemma 3.

Let $\tilde{\mathbf{H}}_k$ be the k-th column of $\widetilde{\mathbb{H}}$ and $p_{\tilde{\mathbf{H}}_k}(\cdot)$ its probability density. Recall that the columns of $\widetilde{\mathbb{H}}$ are independent $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{I}_m)$. Conditioning on the columns $\{\tilde{\mathbf{H}}_i\}_{i\neq k}$ we have

$$\begin{aligned}
&\Pr\left[\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}^\star\right)^\dagger\right) \le s \,\middle|\, \{\tilde{\mathbf{H}}_i = \tilde{\mathbf{h}}_i\}_{i\neq k}\right] \\
&\qquad = \int_{\mathcal{F}_s((\tilde{\mathbf{h}}_i + \sqrt{d_i^\star}\mathbf{e}_i)_{i\neq k})} p_{\tilde{\mathbf{H}}_k}(\tilde{\mathbf{h}} - \sqrt{d_k^\star}\mathbf{e}_k)\,\mathrm{d}\tilde{\mathbf{h}} \\
&\qquad \le \int_{\mathcal{F}_s((\tilde{\mathbf{h}}_i + \sqrt{d_i^\star}\mathbf{e}_i)_{i\neq k})} p_{\tilde{\mathbf{H}}_k}(\tilde{\mathbf{h}} - \sqrt{d_k}\mathbf{e}_k)\,\mathrm{d}\tilde{\mathbf{h}} && (41) \\
&\qquad = \int_{\mathcal{F}_s((\tilde{\mathbf{h}}_i + \sqrt{d_i}\mathbf{e}_i)_{i\neq k})} p_{\tilde{\mathbf{H}}_k}(\tilde{\mathbf{h}} - \sqrt{d_k}\mathbf{e}_k)\,\mathrm{d}\tilde{\mathbf{h}} && (42) \\
&= \Pr\left[\det\left(\mathsf{I}_m + \frac{1}{\sigma^2}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)\mathsf{K}\left(\widetilde{\mathbb{H}} + \mathsf{D}\right)^\dagger\right) \le s \,\middle|\, \{\tilde{\mathbf{H}}_i = \tilde{\mathbf{h}}_i\}_{i\neq k}\right]
\end{aligned}$$

where $\mathbf{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^\mathsf{T}$ denotes the i-th unit vector of length m. This establishes (37) by multiplying both sides with the joint density of $\{\tilde{\mathbf{H}}_i\}_{i\neq k}$ and integrating over their domain. Here we have used Lemma 3 with $f(\cdot) = p_{\tilde{\mathbf{H}}_k}(\cdot)$ and $c = \sqrt{d_k/d_k^\star} < 1$ to obtain (41) and used (35) to obtain (42).

References

[1] Y.-H. Kim and A. Lapidoth, "On the log determinant of non-central Wishart matrices," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Yokohama, Japan, June 29 – July 4, 2003, p. 54.

[2] H. Boche, "Private communication," July 2003.

[3] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *Proceedings of the IEEE Semiannual Vehicular Technology Conference*, Orlando, FL, October 6–9 2003.

[4] T. Anderson, "The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities," *Proceedings of the Americal Mathematical Society*, vol. 6, no. 2, pp. 170–176, May 1955.

[5] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. Academic Press, 1979.