(Article begins on next page)

Capacity Pre-Log of SIMO Correlated Block-Fading Channels

Wei Yang[1], Giuseppe Durisi[1], Veniamin I. Morgenshtern[2], Erwin Riegler[3]

[1]Chalmers University of Technology, 41296 Gothenburg, Sweden
[2]ETH Zurich, 8092 Zurich, Switzerland
[3]Vienna University of Technology, 1040 Vienna, Austria

***Abstract*—We establish an upper bound on the noncoherent capacity pre-log of temporally correlated block-fading single-input multiple-output (SIMO) channels. The upper bound matches the lower bound recently reported in Riegler *et al.* (2011), and, hence, yields a complete characterization of the SIMO noncoherent capacity pre-log, provided that the channel covariance matrix satisfies a mild technical condition. This result allows one to determine the optimal number of receive antennas to be used to maximize the capacity pre-log for a given block-length and a given rank of the channel covariance matrix.**

I. Introduction

A crucial step in the design of wireless communication systems operating over fading channels is to determine the optimal amount of resources to be used for channel estimation. A fruitful approach to address this problem in a fundamental fashion is to characterize the channel capacity *pre-log* (i.e., the asymptotic ratio between capacity and the logarithm of the signal-to-noise ratio (SNR) as SNR goes to infinity) in the *noncoherent setting* where neither transmitter nor receiver are aware of the realization of the fading process, but both know its statistics perfectly.[1] While a capacity pre-log characterization for single-input single-output (SISO) systems is available for several fading models of practical interest [1]–[4], the multiple-input multiple-output (MIMO) case is still largely open.

The impact of multiple antennas on the capacity pre-log has been characterized in [5] for the Rayleigh-fading *constant block-fading* model. According to this model, the channel stays constant over a block of N channel uses and changes in an independent fashion from block to block. The approach used in [5] to characterize the capacity pre-log is based on an apposite change of variables, which reveals the geometry in the problem. One interesting consequence of the analysis in [5] is that the SISO capacity pre-log of constant block-fading channels coincides with the single-input multiple-output (SIMO) capacity pre-log. Hence, using multiple antennas at the receiver only does not yield a larger capacity pre-log.

A more accurate yet simple way to capture channel variations in time is to assume that the channel is correlated (but not necessarily constant) in each block, with the rank of the corresponding $N \times N$ correlation matrix given by Q. We shall refer to this model as *correlated block-fading*. For this channel model, the SISO capacity pre-log was determined in [3], whereas the MIMO case is still open. A lower bound on the SIMO capacity pre-log was recently reported in [6] and refined in [7]. The results in [6], [7] are surprising, as they imply that, when $Q > 1$, the SIMO pre-log can be larger than the SISO pre-log.

Contributions: In this paper, we provide an upper bound on the SIMO capacity pre-log that matches the lower bound reported in [7]. Hence, the SIMO capacity pre-log is fully characterized. Our result allows us to establish that the optimal number of receive antennas to be used to maximize the capacity pre-log for a given block-length N and rank $Q < N$ of the channel correlation matrix is $\lceil (N-1)/(N-Q) \rceil$.

For the constant block-fading case, we provide an alternative and much simpler derivation of the SIMO capacity pre-log than the one provided in [5]. Our proof is based on *duality* [4] and fully exploits the geometry in the problem unveiled in [5].

Notation: Uppercase boldface letters denote matrices and lowercase boldface letters designate vectors. The superscripts ^{T} and ^{H} stand for transposition and Hermitian transposition, respectively. For a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$, we write $\mathbf{a}_i$ for its ith column, $\mathrm{tr}\{\mathbf{A}\}$ for its trace, and $\sigma_i(\mathbf{A})$ for its ith largest singular value. For a vector $\mathbf{a}$, $\mathrm{diag}\{\mathbf{a}\}$ denotes the diagonal matrix that has the entries of $\mathbf{a}$ on its main diagonal and a_i denotes the ith entry of $\mathbf{a}$. We use a combination of superscripts and subscripts to indicate sequences of random variables or vectors. For example, $\mathbf{a}_m^n$ denotes the sequence of random vectors $\mathbf{a}_m, \mathbf{a}_{m+1} \dots, \mathbf{a}_n$. We use $|\mathcal{I}|$ to denote the cardinality of the set $\mathcal{I}$. We denote expectation by $\mathbb{E}[\cdot]$ and use the notation $\mathbb{E}_{\mathbf{x}}[\cdot]$ or $\mathbb{E}_{\mathsf{Q}}[\cdot]$ to stress that expectation is taken with respect to $\mathbf{x}$ with probability distribution Q. The relative entropy between two probability distributions Q and R is denoted by $\mathsf{D}(\mathsf{Q}\|\mathsf{R})$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = \mathcal{O}(g(x))$, $x \to \infty$, means that $\limsup_{x\to\infty} \bigl|f(x)/g(x)\bigr| < \infty$, and $f(x) = o(g(x))$, $x \to \infty$, means that $\lim_{x\to\infty} \bigl|f(x)/g(x)\bigr| = 0$. For two random matrices $\mathbf{A}$ and $\mathbf{B}$, we write $\mathbf{A} \stackrel{d}{=} \mathbf{B}$ to indicate that $\mathbf{A}$ and $\mathbf{B}$ have the same distribution. Finally, $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ stands for the distribution of a circularly-symmetric complex Gaussian random vector with covariance matrix $\mathbf{R}$.

II. System Model

We consider a Rayleigh-fading correlated block-fading SIMO channel with block-length N and M receive antennas. The main

The work of Erwin Riegler was supported by the WWTF project NOWIRE.

[1]Capacity in the noncoherent setting is often referred to as noncoherent capacity. In the remainder of this paper, it will be referred to simply as capacity.

feature of the correlated block-fading model is that the fading in each component channel between the transmit antenna and each receive antenna is independent across blocks of N channel uses, but is correlated within each block, with the rank of the corresponding covariance matrix given by $Q \leq N$. We shall also assume that the fading is independent and identically distributed (i.i.d.) across component channels. The input-output (IO) relation within a block of N channel uses can be conveniently expressed in matrix form as follows:

$$\mathbf{Y} = \mathbf{S}\mathbf{P}^T \operatorname{diag}\{\mathbf{x}\} + \mathbf{W}. \tag{1}$$

Here, $\mathbf{x} \in \mathbb{C}^N$ contains the input symbols transmitted within the block. We assume that $\mathbf{x}$ is subject to the following average-power constraint:

$$\mathbb{E}\left[\|\mathbf{x}\|^2\right] \leq N\rho. \tag{2}$$

The *whitened* fading matrix $\mathbf{S}$ is of size $M \times Q$ and has i.i.d. $\mathcal{CN}(0,1)$ entries. The $N \times Q$ matrix $\mathbf{P}$, which is deterministic and of full rank $Q \leq N$, describes the correlation structure within a block. We shall assume that the rows of $\mathbf{P}$ have unit norm, and, hence, that the entries of the matrix $\mathbf{S}\mathbf{P}^T$ are identically distributed. Finally, the $M \times N$ Gaussian noise matrix $\mathbf{W}$ has i.i.d. $\mathcal{CN}(0,1)$ entries, and the $M \times N$ matrix $\mathbf{Y}$ collects the signals from the M receive antennas during N channel uses. The model just described is of practical relevance, because it captures channel variation in time in an accurate but simple way: large Q corresponds to fast channel variation. Furthermore, (1) models accurately the IO relation in the frequency domain of a cyclic-prefix orthogonal frequency-division multiplexing system that operates over a multipath channel with Q uncorrelated taps. Note that, when $Q = 1$, the correlated block-fading model reduces to the constant block-fading model.

The capacity of the channel (1) is given by

$$C(\rho) \triangleq \frac{1}{N} \sup_{\mathsf{Q}} I(\mathbf{x};\mathbf{Y})$$

where $I(\mathbf{x};\mathbf{Y})$ denotes the mutual information between $\mathbf{x}$ and $\mathbf{Y}$ in (1), and the supremum is over all probability distributions Q on $\mathbf{x}$ that satisfy (2). As the noise has unit variance, ρ denotes the SNR. The capacity pre-log χ is defined as

$$\chi = \lim_{\rho \to \infty} C(\rho)/\log\rho.$$

III. Known Results

In the noncoherent setting where the realizations of the fading process $\mathbf{S}$ are not known to transmitter and receiver (but $\mathbf{P}$ and the statistics of $\mathbf{S}$ are perfectly known), an analytic characterization of $C(\rho)$ is not available. As we shall review next, pre-log expressions are available for some values of N, Q, and M.

For the SISO case ($M = 1$), Liang and Veeravalli [3] proved that the pre-log is equal to $1 - Q/N$. This result can be interpreted as follows: channel uncertainty yields a penalty of Q/N compared to the case when the channel is perfectly known to the receiver (in this case, capacity grows logarithmically with SNR and the capacity pre-log is one [8]). Alternatively, we can interpret Q/N as the fraction of channel uses in which *pilot symbols* need to be transmitted to learn the channel at the receiver [9]. When $Q = N$, learning the channel requires to transmit pilot symbols in each channel use; hence, $\chi = 0$. In this case, capacity turns out to grow double-logarithmically with SNR, independently of the number of receive antennas [4, Thm. 4.2].

For the special case $Q = 1$ (i.e., constant block-fading), the SISO capacity can actually be characterized up to a $o(1)$ term [2], [5] (see [9] for a simple proof). For the SIMO case, such a characterization is available only when $N \geq M+1$ [5, Lem. 13]. However, a pre-log characterization is available for all block-length values N. In particular, it follows from [5, Eq. (27)] that the SIMO capacity pre-log for the $Q = 1$ case is equal to $1-1/N$, i.e., it coincides with the SISO capacity pre-log. This result implies that, when $Q = 1$, using multiple antennas at the receiver only is not beneficial from a pre-log point of view.

This statement turns out to be no longer valid when $Q > 1$. More precisely, the following result was recently proven in [7]:

Theorem 1 ([7, Thm. 1]): Suppose that $\mathbf{P}$ in (1) satisfies the following *Property (A)*: There exists a subset of indices $\mathcal{K} \subset \{1,\dots,N\}$ with cardinality

$$|\mathcal{K}| \triangleq \min(\lceil (QM-1)/(M-1)\rceil, N)$$

such that every Q row vectors of the submatrix of $\mathbf{P}$ obtained by retaining the rows in $\mathbf{P}$ with indices in $\mathcal{K}$ are linearly independent. Then the pre-log of the channel (1) is lower-bounded as

$$\chi \geq \min\left\{M\left(1 - Q/N\right), 1 - 1/N\right\}.$$

Theorem 1 implies that the pre-log penalty of Q/N incurred in the SISO case by not knowing the channel at the receiver can be reduced to $1/N$ by deploying multiple antennas at the receiver side, as long as the block-length is sufficiently large and $\mathbf{P}$ satisfies Property (A). In other words, one pilot symbol per block suffices to learn the channel at the receiver. Intuitively, Property (A) ensures that one can recover both $\mathbf{S}$ and $N-1$ entries of $\mathbf{x}$ from the noiseless receive signal $\mathbf{S}\mathbf{P}^T \operatorname{diag}\{\mathbf{x}\}$, once one entry of $\mathbf{x}$ is fixed [7].

IV. A Matching Pre-Log Upper Bound

The main result of this paper is the following theorem:

Theorem 2: The capacity pre-log of the channel (1) is upper-bounded by

$$\chi \leq \min\left\{M\left(1 - Q/N\right), 1 - 1/N\right\}. \tag{3}$$

Remarks: Theorem 2, combined with Theorem 1, yields a complete characterization of the SIMO capacity pre-log for the case when $\mathbf{P}$ satisfies Property (A). The SIMO capacity pre-log is given by the minimum between the number of receive antennas M times the SISO capacity pre-log of a rank-Q channel, and the SISO capacity pre-log of a rank-1 channel. Note that the pre-log upper bound in (3) holds independently of whether $\mathbf{P}$ satisfies Property (A) or not. We expect the upper bound to be loose if Property (A) is not satisfied. Assume now that every $Q \times Q$ submatrix of $\mathbf{P}$ has full rank (a condition slightly stronger than Property (A)). Then, (3) implies that the optimal number of receive antennas to be used to maximize the capacity pre-log for a given block-length N and rank $Q < N$ of the channel correlation matrix is $\lceil (N-1)/(N-Q)\rceil$.

Outline of the proof: The proof consists of two parts. We first prove that $\chi \le M(1-Q/N)$ by generalizing to the SIMO case the approach used in [3, Prop. 4] to establish a tight upper bound on the SISO capacity pre-log. Then, we prove that $\chi \le 1-1/N$ by showing that the capacity of a rank-Q channel with M receive antennas can be upper-bounded by the capacity of a rank-1 channel with MQ receive antennas. The desired result then follows by [5, Eq. (27)]. As the proof of [5, Eq. (27)] is rather involved, we provide an alternative, much simpler proof of this result (for the SIMO case) in Section V-A.

V. Proof of Theorem 2

First Part: $\chi \le M(1-Q/N)$: Without loss of generality, we assume that the first Q rows of $\mathbf{P}$ are linearly independent. This can always be achieved by rearranging the columns of $\mathbf{Y}$ in (1). We start by manipulating $I(\mathbf{x};\mathbf{Y})$ as follows (we use the notation convention introduced in Section I):

$$\begin{aligned} I(\mathbf{x};\mathbf{Y}) &= I(x_1^N;\mathbf{y}_1^N) \\ &\overset{(a)}{=} I(x_1^N;\mathbf{y}_1^Q) + I(x_1^N;\mathbf{y}_{Q+1}^N \,|\, \mathbf{y}_1^Q) \\ &\overset{(b)}{=} I(x_1^Q;\mathbf{y}_1^Q) + I(x_1^N;\mathbf{y}_{Q+1}^N \,|\, \mathbf{y}_1^Q). \end{aligned} \tag{4}$$

Here, in (a) we used chain rule for mutual information and (b) follows because $\mathbf{y}_1^Q$ and x_{Q+1}^N are conditionally independent given x_1^Q. We next upper-bound each term on the right-hand side (RHS) of (4) separately. The assumption that the first Q rows of $\mathbf{P}$ are linearly independent implies that the first term on the RHS of (4) grows at most double-logarithmically with SNR. More precisely, we have that [4, Thm. 4.2]:

$$I(x_1^Q;\mathbf{y}_1^Q) \le \log\log\rho + \mathcal{O}(1), \quad \rho\to\infty. \tag{5}$$

For the second term on the RHS of (4), we proceed as follows:

$$\begin{aligned} I(x_1^N;\mathbf{y}_{Q+1}^N \,|\, \mathbf{y}_1^Q) &= h(\mathbf{y}_{Q+1}^N \,|\, \mathbf{y}_1^Q) - h(\mathbf{y}_{Q+1}^N \,|\, \mathbf{y}_1^Q, x_1^N) \\ &\overset{(a)}{\le} h(\mathbf{y}_{Q+1}^N) - h(\mathbf{y}_{Q+1}^N \,|\, \mathbf{y}_1^Q, x_1^N, \mathbf{S}) \\ &= h(\mathbf{y}_{Q+1}^N) - h(\mathbf{w}_{Q+1}^N) \\ &\overset{(b)}{\le} \sum_{k=Q+1}^{N} h(\mathbf{y}_k) + \mathcal{O}(1), \quad \rho\to\infty \\ &\overset{(c)}{\le} \sum_{k=Q+1}^{N} M\log\Big(1+\mathbb{E}\big[|x_k|^2\big]\Big) + \mathcal{O}(1), \quad \rho\to\infty \\ &\overset{(d)}{\le} \sum_{k=Q+1}^{N} M\log(1+N\rho) + \mathcal{O}(1), \quad \rho\to\infty \\ &= M(N-Q)\log\rho + \mathcal{O}(1), \quad \rho\to\infty. \end{aligned} \tag{6}$$

Here, in (a) we used that conditioning reduces entropy; (b) follows by chain rule for differential entropy and because conditioning reduces entropy; (c) follows because jointly proper Gaussian random vectors are entropy-maximizers for a fixed covariance matrix and because $\mathbb{E}\big[\mathbf{y}_k\mathbf{y}_k^H\big] = (1+\mathbb{E}\big[|x_k|^2\big])\mathbf{I}_M$ (recall that we assumed that the rows of $\mathbf{P}$ have unit norm); finally, in (d) we used the average-power constraint (2). The desired upper bound on the capacity pre-log follows by substituting (5) and (6) into (4).

Second part: $\chi \le 1-1/N$: We show that the capacity of a rank-Q channel with M receive antennas is upper-bounded by the capacity of a rank-1 channel with QM receive antennas. By simple matrix manipulations, we can rewrite the IO relation (1) in the following more convenient form:

$$\mathbf{Y} = \sum_{q=1}^{Q} \mathbf{s}_q\mathbf{x}^T \operatorname{diag}\{\mathbf{p}_q\} + \mathbf{W}.$$

Let now $\mathbf{W}_1,\cdots,\mathbf{W}_Q$ be $M\times N$ independent random matrices with i.i.d. $\mathcal{CN}(0,1)$ entries. As, by assumption, the rows of $\mathbf{P}$ have unit norm, we have that

$$\mathbf{W} \overset{d}{=} \sum_{q=1}^{Q} \mathbf{W}_q \operatorname{diag}\{\mathbf{p}_q\}.$$

Hence, we can rewrite $\mathbf{Y}$ as

$$\mathbf{Y} \overset{d}{=} \sum_{q=1}^{Q} \mathbf{Y}_q \operatorname{diag}\{\mathbf{p}_q\}$$

where

$$\mathbf{Y}_q \triangleq \mathbf{s}_q\mathbf{x}^T + \mathbf{W}_q.$$

Note now that each $\mathbf{Y}_q$ is the output of a rank-1 SIMO channel with M receive antennas. By observing that $\mathbf{x}$ and $\mathbf{Y}$ are conditionally independent given $\{\mathbf{Y}_1,\cdots,\mathbf{Y}_Q\}$, we conclude that, by the data-processing inequality [10, Sec. 2.8],

$$I(\mathbf{x};\mathbf{Y}) \le I\left(\mathbf{x};\mathbf{Y}_1,\dots,\mathbf{Y}_Q\right).$$

The claim follows by noting that the $(QM)\times N$ matrix obtained by stacking the matrices $\mathbf{Y}_q$ on top of each others is the output of a rank-1 SIMO channel with QM receive antennas. As reviewed in Section III, the SIMO capacity pre-log for the rank-1 case coincides with the SISO capacity pre-log and is given by $1-1/N$. This result follows from [4, Thm. 4.2], for the case $N=1$, and from [5, Eq. (27)], for the case $N\ge 1$. This concludes the proof.

For completeness, in Lemma 3 below we restate [5, Eq. (27)] for the SIMO case, and provide an alternative, much simpler proof of this result in Section V-A below.

Lemma 3: The capacity of the SIMO channel (1) with M receive antennas, $Q=1$, and $N\ge 2$ is given by

$$C(\rho) = (1-1/N)\log\rho + \mathcal{O}(1), \quad \rho\to\infty. \tag{7}$$

A. Proof of Lemma 3

1) Geometric Intuition: When $Q=1$, we can rewrite the IO relation as

$$\mathbf{Y} = \mathbf{s}\mathbf{x}^T + \mathbf{W}$$

where $\mathbf{s}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_M)$. We next provide a geometric argument illustrating why the SIMO capacity pre-log coincides with the SISO capacity pre-log when $Q=1$. A similar argument can be found in [5]. Let $\mathbf{x}$ be an arbitrary vector in $\mathbb{C}^N$. In the absence of noise, the rows of $\mathbf{Y}$ are collinear with $\mathbf{x}$. The only information the receiver can recover (in the absence of noise) about the transmit vector $\mathbf{x}$ from any of these rows is the line on which $\mathbf{x}$ lies. A line in $\mathbb{C}^N$ is characterized by $N-1$ complex parameters. Hence, as argued in [9], the receive signal $\mathbf{Y}$ carries $N-1$

parameters describing $\mathbf{x}$. This number, divided by N, coincides with the capacity pre-log we want to establish. As one column of $\mathbf{Y}$ is sufficient to recover the $N-1$ parameters describing the line on which $\mathbf{x}$ lies, adding more receive antennas does not appear to be beneficial. We next prove this result by sandwiching capacity between a lower bound and an upper bound that are tight at high SNR.

2) A Capacity Lower Bound: The RHS of (7) is a lower-bound on capacity. This result follows directly from [3, Prop. 7].

3) A Matching Upper Bound Through Duality: Establishing an asymptotically tight capacity upper bound is more involved. Our proof is based on duality [4], a technique that allows us to obtain a tight upper bound on $I(\mathbf{x};\mathbf{Y})$ by carefully choosing a probability distribution on $\mathbf{Y}$. More precisely, let $\mathsf{W}(\cdot\,|\,\mathbf{x})$ denote the conditional distribution of $\mathbf{Y}$ given $\mathbf{x}$, and let QW denote the distribution induced on $\mathbf{Y}$ by the input distribution Q and by the channel $\mathsf{W}(\cdot\,|\,\mathbf{x})$. Finally, let R be an arbitrary distribution on $\mathbf{Y}$ with probability density function (pdf) $\mathsf{r}(\mathbf{Y})$. We use duality to upper-bound the mutual information $I(\mathbf{x};\mathbf{Y})$ as follows [4, Thm. 5.1]:

$$\begin{aligned} I(\mathbf{x};\mathbf{Y}) &\leq \mathbb{E}_{\mathsf{Q}}[\mathsf{D}(\mathsf{W}(\cdot\,|\,\mathbf{x})\|\mathsf{R}(\cdot))] \\ &= -\,\mathbb{E}_{\mathsf{QW}}[\log \mathsf{r}(\mathbf{Y})] - h(\mathbf{Y}\,|\,\mathbf{x}). \end{aligned} \tag{8}$$

To get a tight capacity upper bound, the output distribution R must be chosen appropriately. For the SISO case, this choice can be motivated as follows: the geometry unveiled in Section V-A1 suggests to use the subspace spanned by $\mathbf{x}$ to convey information. This can be achieved by choosing an input distribution that is uniformly distributed on the sphere in $\mathbb{C}^N$ with radius $\sqrt{N\rho}$. The output distribution induced by this input distribution in the absence of additive noise turns out to yield a tight capacity upper bound, as shown in [9].

Generalizing this approach to the SIMO case is not straightforward. The reason is as follows: for any choice of the input distribution, the matrix $\mathbf{s}\mathbf{x}^T$ has rank at most 1, whereas the additive noise matrix $\mathbf{W}$ has full rank with probability one. This implies that, independently of the choice of the input distribution, the induced output distribution in the absence of additive noise is not absolutely continuous [11, Def. 6.7] with respect to $\mathsf{W}(\cdot\,|\,\mathbf{x})$, and, hence, the RHS of (8) diverges. To get a tight bound, one needs to choose an output distribution for which $\mathbf{Y}$ has full rank with probability one. This implies that, differently from the SISO case, the additive noise needs to be accounted for in the choice of the output distribution.

To shed light on how this can be done, it is convenient to express $\mathbf{Y}$ in terms of its singular-value decomposition (SVD). More specifically, let $P=\min\{M,N\}$ and $L=\max\{M,N\}$; then $\mathbf{Y}$ can be written as $\mathbf{Y}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$, where $\mathbf{U}\in\mathbb{C}^{M\times P}$ and $\mathbf{V}\in\mathbb{C}^{N\times P}$ are (truncated) unitary matrices, and $\boldsymbol{\Sigma}=\operatorname{diag}\{[\sigma_1(\mathbf{Y})\cdots\sigma_P(\mathbf{Y})]\}$ contains the singular values of $\mathbf{Y}$ in descending order. To make the SVD unique, we assume that the first row of $\mathbf{U}$ is real and non-negative. We shall take an output distribution for which $\sigma_1(\mathbf{Y})$ is distributed as the nonzero singular value of the noiseless receive matrix $\mathbf{s}\mathbf{x}^T$ and the remaining singular values are distributed as the ordered singular values of a $(M-1)\times(N-1)$ random matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. More specifically, we take[2]

$$\mathsf{r}(\sigma_1,\cdots,\sigma_P)=\mathsf{r}(\sigma_1)\cdot\mathsf{r}(\sigma_2,\cdots,\sigma_P)$$

where

$$\mathsf{r}(\sigma_1)=\frac{2\sigma_1}{MN\rho}\cdot e^{-\sigma_1^2/(MN\rho)},\quad \sigma_1>0$$

and [12, Thm. 2.17]

$$\begin{aligned}\mathsf{r}(\sigma_2,\cdots,\sigma_P) &= 2^{P-1}e^{-\sum_{i=2}^P\sigma_i^2}\cdot\prod_{i=2}^P\frac{\sigma_i^{2(L-P)+1}}{(L-i)!(P-i)!} \\ &\quad\cdot\prod_{i=2}^{P-1}\prod_{j=i+1}^P\left(\sigma_i^2-\sigma_j^2\right)^2,\quad \sigma_2,\ldots,\sigma_P>0.\end{aligned}$$

Finally, we take $\mathbf{V}$ and $\mathbf{U}$ independent of the singular values and uniformly distributed (with respect to the Haar measure) on the Stiefel manifold[3] $\mathrm{S}(N,P)$, and on the submanifold of $\mathrm{S}(M,P)$ induced by the nonnegativity of the first row of $\mathbf{U}$, respectively. We next evaluate the RHS of (8) for the resulting output pdf, which we (still) denote by $\mathsf{r}(\mathbf{Y})$. The conditional differential entropy $h(\mathbf{Y}\,|\,\mathbf{x})$ in (8) can be easily computed:

$$h(\mathbf{Y}\,|\,\mathbf{x})=M\,\mathbb{E}_{\mathbf{x}}\left[\log\left(\|\mathbf{x}\|^2+1\right)\right]+MN\log(\pi e). \tag{9}$$

To evaluate the first term on the RHS of (8), it is convenient to express $\mathsf{r}(\mathbf{Y})$ in the SVD coordinate system. By the change of variables theorem [11, Thm. 7.26], we get

$$\begin{aligned}-\,\mathbb{E}_{\mathsf{QW}}[\log\mathsf{r}(\mathbf{Y})] &= -\,\mathbb{E}_{\mathsf{QW}}[\log\mathsf{r}(\mathbf{U},\boldsymbol{\Sigma},\mathbf{V})] \\ &\quad+\mathbb{E}_{\mathsf{QW}}[\log J_{M,N}(\sigma_1,\cdots,\sigma_P)]\end{aligned} \tag{10}$$

where $J_{M,N}(\sigma_1,\cdots,\sigma_P)$ is the Jacobian of the SVD, which is given by [5, App. A]

$$J_{M,N}(\sigma_1,\cdots,\sigma_P)=\prod_{i=1}^P\sigma_i^{2(L-P)+1}\cdot\prod_{i=1}^{P-1}\prod_{j=i+1}^P\left(\sigma_i^2-\sigma_j^2\right)^2.$$

By construction, we have that

$$\begin{aligned}-\,\mathbb{E}_{\mathsf{QW}}[\log\mathsf{r}(\mathbf{U},\boldsymbol{\Sigma},\mathbf{V})] &= \underbrace{-\,\mathbb{E}_{\mathsf{QW}}[\log\mathsf{r}(\mathbf{U})]-\mathbb{E}_{\mathsf{QW}}[\log\mathsf{r}(\mathbf{V})]}_{=\mathcal{O}(1),\quad \rho\to\infty} \\ &\quad-\mathbb{E}_{\mathsf{QW}}[\log\mathsf{r}(\sigma_1)]-\mathbb{E}_{\mathsf{QW}}[\log\mathsf{r}(\sigma_2,\ldots,\sigma_P)] \\ &=\log\rho-\mathbb{E}_{\mathsf{QW}}[\log\sigma_1] \\ &\quad\underbrace{+\,\mathbb{E}_{\mathsf{QW}}\left[\sigma_1^2\right]/(MN\rho)+\mathbb{E}_{\mathsf{QW}}\left[\sum_{i=2}^P\sigma_i^2\right]}_{\triangleq c_1(\rho)} \\ &\quad-\sum_{i=2}^{P-1}\sum_{j=i+1}^P\mathbb{E}_{\mathsf{QW}}\left[\log(\sigma_i^2-\sigma_j^2)^2\right] \\ &\quad-\sum_{i=2}^P\mathbb{E}_{\mathsf{QW}}\left[\log\sigma_i^{2(L-P)+1}\right]+\mathcal{O}(1),\quad\rho\to\infty.\end{aligned} \tag{11}$$

[2]We shall indicate $\sigma_i(\mathbf{Y})$ simply as σ_i whenever no ambiguity occurs.

[3]The set of complex $m\times n$ ($n\geq m$) unitary matrices form a manifold $\mathrm{S}(n,m)$ of $2mn-m^2$ real dimensions, called the Stiefel manifold [13], [5]. This manifold has volume $|\mathrm{S}(n,m)|=\prod_{i=n-m+1}^n 2\pi^i/(i-1)!$.

The expectation of the Jacobian in (10) can be rewritten as

$$\begin{aligned}&\mathbb{E}_{\mathsf{QW}}[\log J_{M,N}(\sigma_1,\cdots,\sigma_P)]\\&=\mathbb{E}_{\mathsf{QW}}\Big[\log\sigma_1^{2(L-P)+1}\Big]+\sum_{j=2}^{P}\mathbb{E}_{\mathsf{QW}}\underbrace{\big[\log(\sigma_1^2-\sigma_j^2)^2\big]}_{\le\log\sigma_1^4}\\&\quad+\sum_{i=2}^{P}\mathbb{E}_{\mathsf{QW}}\Big[\log\sigma_i^{2(L-P)+1}\Big]\\&\quad+\sum_{i=2}^{P-1}\sum_{j=i+1}^{P}\mathbb{E}_{\mathsf{QW}}\big[\log(\sigma_i^2-\sigma_j^2)^2\big]\,. \qquad (12)\end{aligned}$$

Substituting (11) and (12) into (10), we obtain

$$\begin{aligned}-\mathbb{E}_{\mathsf{QW}}[\mathsf{r}(\mathbf{Y})]\le\log\rho+(N+M-2)\,\mathbb{E}_{\mathsf{QW}}\big[\log\sigma_1^2\big]\\+c_1(\rho)+\mathcal{O}(1),\quad\rho\to\infty.\qquad(13)\end{aligned}$$

Finally, substituting (13) and (9) into (8), we get

$$\begin{aligned}&I(\mathbf{x};\mathbf{Y})\le\log\rho+(N-2)\,\mathbb{E}_{\mathsf{QW}}\big[\log\sigma_1^2\big]+c_1(\rho)\\&+M\underbrace{\left(\mathbb{E}_{\mathsf{QW}}\big[\log\sigma_1^2\big]-\mathbb{E}_{\mathbf{x}}\big[\log\left(\|\mathbf{x}\|^2+1\right)\big]\right)}_{\triangleq c_2(\rho)}+\mathcal{O}(1),\quad\rho\to\infty.\end{aligned}$$

We conclude the proof by showing that, $\mathbb{E}_{\mathsf{QW}}\big[\log\sigma_1^2\big]\le\log\rho+\mathcal{O}(1)$, $\rho\to\infty$ and that $c_1(\rho)$ and $c_2(\rho)$ can be upper-bounded by finite constants. For the first term, we have that

$$\begin{aligned}\mathbb{E}_{\mathsf{QW}}\big[\log\sigma_1^2\big]&\le\mathbb{E}_{\mathsf{QW}}\big[\log\mathrm{tr}\{\mathbf{Y}^H\mathbf{Y}\}\big]\stackrel{(a)}{\le}\log\sum_{i=1}^{N}\mathbb{E}_{\mathsf{QW}}\big[\|\mathbf{y}_i\|^2\big]\\&\stackrel{(b)}{\le}\log\rho+\mathcal{O}(1),\quad\rho\to\infty.\qquad(14)\end{aligned}$$

Here, in (a) we used Jensen's inequality and (b) follows from (2). To show that $c_1(\rho)$ and $c_2(\rho)$ are bounded, the following lemma will turn out to be useful.

Lemma 4 ([14, Sec. 7.3]): Let $\mathbf{A}$, $\mathbf{B}\in\mathbb{C}^{m\times n}$ and $p=\min\{m,n\}$. Then

$$\sigma_{i+j-1}(\mathbf{A}+\mathbf{B})\le\sigma_i(\mathbf{A})+\sigma_j(\mathbf{B}),\quad 1\le i,j\le p,\ i+j\le p+1.$$

If we choose $\mathbf{A}=\mathbf{s}\mathbf{x}^T$ and $\mathbf{B}=\mathbf{W}$, we obtain from Lemma 4 that

$$\sigma_i(\mathbf{Y})\le\begin{cases}\|\mathbf{s}\|\|\mathbf{x}\|+\sigma_1(\mathbf{W}), & i=1\\ \sigma_{i-1}(\mathbf{W}), & 2\le i\le P.\end{cases}\qquad(15)$$

By using (15), it follows that

$$\mathbb{E}_{\mathsf{QW}}\left[\sum_{i=2}^{P}\sigma_i^2(\mathbf{Y})\right]\le\mathbb{E}_{\mathbf{W}}\left[\sum_{i=1}^{P-1}\sigma_i^2(\mathbf{W})\right]\le MN.$$

This inequality, together with the inequality

$$\mathbb{E}_{\mathsf{QW}}\big[\sigma_1^2\big]\le MN(\rho+1)$$

which can be established using similar steps to the ones leading to (14), are sufficient to conclude that $c_1(\rho)$ is bounded.

To establish that $c_2(\rho)$ is bounded, we start by noting that the first term in the expression that defines $c_2(\rho)$ can be upper-bounded as follows:

$$\begin{aligned}\mathbb{E}_{\mathsf{QW}}\big[\log\sigma_1^2(\mathbf{Y})\big]&\stackrel{(a)}{\le}2\,\mathbb{E}_{\mathsf{QW}}\big[\log\big(\|\mathbf{s}\|\|\mathbf{x}\|+\sigma_1(\mathbf{W})\big)\big]\\&\stackrel{(b)}{\le}2\,\mathbb{E}_{\mathbf{x}}\big[\log\big(\mathbb{E}_{\mathbf{s},\mathbf{W}}[\|\mathbf{s}\|\|\mathbf{x}\|+\sigma_1(\mathbf{W})]\big)\big]\\&\stackrel{(c)}{\le}\mathbb{E}_{\mathbf{x}}\Big[\log\big(\sqrt{M}(\|\mathbf{x}\|+\sqrt{N})\big)^2\Big]\,.\end{aligned}$$

Here, (a) follows from (15), (b) holds because of Jensen's inequality, and in (c) we used that $\mathbb{E}[\|\mathbf{s}\|]\le\sqrt{M}$ and that

$$\begin{aligned}\left(\mathbb{E}[\sigma_1(\mathbf{W})]\right)^2\le\mathbb{E}\Big[\big(\sigma_1(\mathbf{W})\big)^2\Big]\le\mathbb{E}\big[\mathrm{tr}\{\mathbf{W}^H\mathbf{W}\}\big]\\=MN.\end{aligned}$$

Hence,

$$\begin{aligned}c_2(\rho)&\le\mathbb{E}_{\mathbf{x}}\left[\log\frac{\big(\sqrt{M}(\|\mathbf{x}\|+\sqrt{N})\big)^2}{\|\mathbf{x}\|^2+1}\right]\\&\le\sup_{\mathbf{x}}\left\{\log\frac{\big(\sqrt{M}(\|\mathbf{x}\|+\sqrt{N})\big)^2}{\|\mathbf{x}\|^2+1}\right\}=\log[M(N+1)].\end{aligned}$$

This concludes the proof.

References

[1] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[2] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[3] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[4] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[5] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[6] V. I. Morgenshtern, G. Durisi, and H. Bölcskei, "The SIMO pre-log can be larger than the SISO pre-log," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, U.S.A., Jun. 2010, pp. 320–324.

[7] E. Riegler, V. I. Morgenshtern, G. Durisi, S. Lin, B. Sturmfels, and H. Bölcskei, "Noncoherent SIMO pre-log via resolution of singularities," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2149–2153.

[8] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[9] G. Durisi and H. Bölcskei, "High-SNR capacity of wireless communication channels in the noncoherent setting: A primer," *Int. J. Electron. Commun. (AEÜ)*, vol. 65, no. 8, pp. 707–712, Aug. 2011, invited paper.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, U.S.A.: Wiley, 2006.

[11] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, U.S.A.: McGraw-Hill, 1987.

[12] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*. Delft, The Netherlands: now Publishers, 2004, vol. 1, no. 1, pp. 1–182.

[13] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd ed. Orlando, FL, U.S.A.: Academic Press, Inc., 1986.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.