Capacity-achieving Input Covariance for Correlated Multi-Antenna Channels

Antonia M. Tulino
Universita Degli Studi di Napoli "Federico II"
80125 Napoli, Italy
atulino@ee.princeton.edu

Angel Lozano
Bell Labs (Lucent Technologies)
Holmdel, NJ 07733
aloz@lucent.com

Sergio Verdú
Princeton University
Princeton, NJ 08540
verdu@ee.princeton.edu

Abstract

This paper characterizes the capacity-achieving input covariance for multi-antenna channels with (zero-mean) arbitrary distribution. The solution accommodates a wide range of correlation structures, not necessarily separate transmit-receive. Our characterization of the covariance encompasses both its eigenvectors, which are found explicitly, and its eigenvalues, for which we present necessary and sufficient conditions as well as an iterative algorithm. In addition, we identify the correlation structures for which an isotropic input achieves capacity.

1 Introduction

While, in most instances of wireless communication, the receiver can accurately track the state of the fading channel, the transmitter is often unable to perform such tracking. Statistical information about the channel, on the other hand, is virtually always accessible since the periods over which the fading process is basically stationary are several orders of magnitude larger than the duration of the fades. As a result, the most typical operating regime in mobile systems is that in which (*i*) the receiver has access to the instantaneous state of the channel, and (*ii*) the transmitter has only access to its distribution. The input cannot therefore depend on the state of the channel, but only on its distribution.

In multi-antenna channels impaired by additive Gaussian noise and with perfect knowledge of the channel at the receiver, the capacity-achieving input is zero-mean Gaussian and thus characterizing its distribution boils down to determining its spatial covariance. The earliest statement about this covariance was made in [1], where it was shown that, for channels with zero-mean IID (independent identically distributed) Gaussian entries, the capacity-achieving input is isotropic. Posterior findings have expanded this initial result in several directions:

- For channels with zero-mean Gaussian entries, correlated only at the transmitter side, the eigenvectors of the capacity-achieving input covariance were established in [2, 3]. (The result was first proved for the multi-transmit single-receive case in [2], then extended to multiple uncorrelated receive antennas in [3].) In both cases, the eigenvalues were left to numerical optimization.[1]

- For multi-transmit single-receive channels with non-zero-mean IID Gaussian entries, the eigenvectors of the input covariance were characterized in [2] (see also [5]). Some relationships between the eigenvalues were also put forth, but the need for a numerical optimization was not fully eliminated.

- For the region of low signal-to-noise ratio (SNR), a complete characterization of the limiting covariance, valid for arbitrary channels, was given in [6].

This paper generalizes the results of [1]–[4] removing the various constrains imposed therein. Specifically, the contributions are as follows:

- We obtain the capacity-achieving input covariance for channels with (*i*) arbitrary numbers of transmit and receive antennas, (*ii*) zero-mean but otherwise arbitrary fading distribution, and (*iii*) very general correlations—not necessarily *separable*—between the entries of the channel matrix.

- Our characterization of the input covariance encompasses both the eigenvectors, which are found explicitly, and the eigenvalues, for which we present necessary and sufficient conditions as well as an explicit iterative algorithm.

- We identify the correlation structures for which the result in [1] holds, i.e., for which an isotropic input achieves capacity.

2 Definitions

Throughout the paper, $(\cdot)_{i,j}$ indicates the (*i*,*j*)-th entry of a matrix, $(\cdot)_j$ indicates its j-th column, and $(\cdot)_{-j}$ indicates the submatrix obtained by removing the j-th column.

Given n_{T} transmit and n_{R} receive antennas and frequency-flat fading, the baseband model we consider is

$$\mathbf{y} = \sqrt{g}\,\mathbf{H}\,\mathbf{x} + \mathbf{n}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the input and output vectors while $\mathbf{n}$ is white Gaussian noise. The channel is represented by the ($n_{\mathrm{R}} \times n_{\mathrm{T}}$) zero-mean random matrix $\sqrt{g}\,\mathbf{H}$ where the scalar g is such that

$$E[\mathrm{Tr}\{\mathbf{H}\mathbf{H}^\dagger\}] = n_{\mathrm{R}} n_{\mathrm{T}}.$$

The covariance of the input, conveniently normalized, is denoted by

$$\mathbf{\Phi} \triangleq \frac{n_{\mathrm{T}}}{E[\|\mathbf{x}\|^2]} E[\mathbf{x}\mathbf{x}^\dagger]$$

where the normalization ensures that $E[\mathrm{Tr}\{\mathbf{\Phi}\}]{=}n_{\mathrm{T}}$. The input is isotropic when $\mathbf{\Phi}{=}\mathbf{I}$. The ergodic capacity is

$$C = \max_{\mathbf{\Phi}:\mathrm{Tr}\{\mathbf{\Phi}\}=n_{\mathrm{T}}} E\left[\log_2 \det\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\mathrm{T}}}\,\mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger\right)\right] \tag{1}$$

[1]For the multi-transmit single-receive case, an iterative scheme to compute these eigenvalues was proposed in [4]. This scheme, however, requires that an implicit equation be solved at each iteration.

with

$$\mathsf{SNR} \triangleq g \frac{E[\|\mathbf{x}\|^2]}{\frac{1}{n_{\mathrm{R}}} E[\|\mathbf{n}\|^2]}$$

which corresponds to the average signal-to-noise ratio per receive antenna when either the input is isotropic or the channel entries are IID.[2]

The correlation between the (i,j)-th and (i',j')-th entries of $\mathbf{H}$ is denoted by

$$R_{\mathbf{H}}(i,j;i',j') \triangleq E\left[(\mathbf{H})_{i,j}(\mathbf{H})^*_{i',j'}\right].$$

Most of the multi-antenna literature, however, deals only with *separable* (also termed *kronecker* or *product*) correlations [7], constrained as follows:

Definition 1 *The correlation of* $\mathbf{H}$ *is said to be* separable *if*

$$R_{\mathbf{H}}(i,j;i',j') = (\mathbf{\Theta}_{\mathrm{R}})_{i,i'} (\mathbf{\Theta}_{\mathrm{T}})_{j,j'}$$

where $\mathbf{\Theta}_{\mathrm{R}}$ *and* $\mathbf{\Theta}_{\mathrm{T}}$ *are* ($n_{\mathrm{R}} \times n_{\mathrm{R}}$) *and* ($n_{\mathrm{T}} \times n_{\mathrm{T}}$) *correlation matrices whose entries indicate the correlation between receive antennas and between transmit antennas, respectively.*

While simple and analytically friendly, the separable correlation model has clear limitations. It usually suffices to represent the correlation that arises with spatial diversity, due to antenna proximity, but it cannot accommodate other diversity mechanisms such as those that rely on polarization or pattern differences [8]. The $n_{\mathrm{R}} n_{\mathrm{T}}$ eigenvalues of a separable correlation function are determined by those of $\mathbf{\Theta}_{\mathrm{R}}$ and $\mathbf{\Theta}_{\mathrm{T}}$. This restriction on the number of degrees of freedom in the eigenvalues of $R_{\mathbf{H}}$, which ultimately govern the capacity impact of correlation, precludes the representation of many channels of interest.

In this paper, we consider more general correlations whose eigenvalues have full $n_{\mathrm{R}} n_{\mathrm{T}}$ degrees of freedom. Specifically, our analysis encompasses any channel that can be expressed as

$$\mathbf{H} = \mathbf{U} \tilde{\mathbf{H}} \mathbf{V}^\dagger \tag{2}$$

where $\tilde{\mathbf{H}}$ has uncorrelated entries while $\mathbf{U}$ and $\mathbf{V}$ are ($n_{\mathrm{R}} \times n_{\mathrm{R}}$) and ($n_{\mathrm{T}} \times n_{\mathrm{T}}$) deterministic unitary matrices. The expansion in (2) corresponds to the Karhunen-Loève transform of $\mathbf{H}$ and thus [9]:

- The columns of $\mathbf{U}$ and $\mathbf{V}$ must contain the eigenfunctions of $R_{\mathbf{H}}$. From (2), such columns are respectively the eigenvectors of $E[\mathbf{H}\mathbf{H}^\dagger]$ and $E[\mathbf{H}^\dagger\mathbf{H}]$.

- Denoted by $\lambda_{k,\ell}(\cdot)$ the (k,ℓ)-th eigenvalue, the variances of the entries of $\tilde{\mathbf{H}}$ are

$$E\left[|(\tilde{\mathbf{H}})_{k,\ell}|^2\right] = \lambda_{k,\ell}(R_{\mathbf{H}}).$$

- If the correlation is separable, then $\mathbf{U}$ and $\mathbf{V}$ correspond, respectively, with the eigenvectors of $\mathbf{\Theta}_{\mathrm{R}}$ and $\mathbf{\Theta}_{\mathrm{T}}$. In addition, $\lambda_{k,\ell}(R_{\mathbf{H}}) = \lambda_k(\mathbf{\Theta}_{\mathrm{R}})\lambda_\ell(\mathbf{\Theta}_{\mathrm{T}})$.

Note that $\mathbf{H}$ and $\tilde{\mathbf{H}}$ have the same singular values and hence the same capacity. Moreover, if $\mathbf{H}$ is Gaussian (i.e. its entries are jointly Gaussian), then $\tilde{\mathbf{H}}$ is also Gaussian.

[2]In general, the average receive signal-to-noise is $\frac{E[\|g\mathbf{H}\mathbf{x}\|^2]}{E[\|\mathbf{n}\|^2]} = \frac{\mathrm{Tr}\{E[\mathbf{H}^\dagger\mathbf{H}]\mathbf{\Phi}\}}{n_{\mathrm{R}} n_{\mathrm{T}}} \mathsf{SNR}$, which depends on $\mathbf{\Phi}$.

Definition 2 *A* ($n_{\mathrm{R}} \times n_{\mathrm{T}}$) *matrix* $\mathbf{B}$ *taking values in* $\mathcal{B} \subset \Re^+$ *is column-regular if the entries of every column exhibit the same empirical distribution, i.e.*

$$\frac{1}{n_{\mathrm{R}}} \sum_{i=1}^{n_{\mathrm{R}}} 1\{(\mathbf{B})_{i,j} < \xi\}$$

does not depend on j*, with* $1\{\cdot\}$ *the indicator function.*

3 Capacity-achieving Input Covariance

3.1 Eigenvectors and Conditions on the Eigenvalues

Let $\mathbf{\Phi}_{\mathsf{c}}$ be the input covariance that achieves capacity. Its characterization boils down to determining (*i*) its eigenvectors, i.e. the *directions* on which signalling should take place, and (*ii*) its eigenvalues, i.e. the power that should be allocated onto each such direction. The following central result—proved in the Appendix—identifies the former and lays down necessary and sufficient conditions to be satisfied by the latter.

Theorem 1 *Consider a channel with zero-mean but otherwise arbitrarily distributed entries that can be expressed as* $\mathbf{H}{=}\mathbf{U}\tilde{\mathbf{H}}\mathbf{V}^\dagger$ *with* $\tilde{\mathbf{H}}$ *having uncorrelated entries while* $\mathbf{U}$ *and* $\mathbf{V}$ *contain, respectively, the eigenvectors of* $E[\mathbf{H}\mathbf{H}^\dagger]$ *and* $E[\mathbf{H}^\dagger\mathbf{H}]$*. The capacity-achieving input covariance is* $\mathbf{\Phi}_{\mathsf{c}}{=}\mathbf{V}\mathbf{P}_{\mathsf{c}}\mathbf{V}^\dagger$ *with* $\mathbf{P}_{\mathsf{c}}$ *a diagonal matrix whose entries, constrained such that* $\mathrm{Tr}\{\mathbf{P}_{\mathsf{c}}\}{=}n_{\mathrm{T}}$*, satisfy*

$$\frac{1}{n_{\mathrm{R}}} E\left[\mathrm{Tr}\left\{\left(\mathbf{I} + \mathsf{SNR}\,(\tilde{\mathbf{H}})_j(\tilde{\mathbf{H}})_j^\dagger\right)\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}}\tilde{\mathbf{H}}\mathbf{P}_{\mathsf{c}}\tilde{\mathbf{H}}^\dagger\right)^{-1}\right\}\right] \begin{array}{lll} = 1 & \textit{if} & (\mathbf{P}_{\mathsf{c}})_{j,j} > 0 \\ \leq 1 & \textit{if} & (\mathbf{P}_{\mathsf{c}})_{j,j} = 0 \end{array} \quad (3)$$

Note that $\mathbf{U}$ *is immaterial in (3) and thus it does not affect* $\mathbf{P}_{\mathsf{c}}$*.*

In general, the power allocation $\mathbf{P}_{\mathsf{c}}$ does not admit a waterfill interpretation. An iterative algorithm to find this power allocation, which depends on the SNR, is provided in Section 3.2. At low and high SNR, however, the conditions in (3) simplify drastically:

- For $\mathsf{SNR}{\to}0$, the entire power budget should be allocated to the eigenspace within $\mathbf{V}$ corresponding to the maximal eigenvalue of $E[\mathbf{H}^\dagger\mathbf{H}]$ to achieve second-order optimality [6, Theorem 12]. If the multiplicity of such eigenvalue is plural, the power should be evenly divided between the corresponding eigenvectors.

- For $\mathsf{SNR}{\to}\infty$, the power should be evenly divided among the eigevectors within $\mathbf{V}$ whose eigenvalues in $E[\mathbf{H}^\dagger\mathbf{H}]$ are nonzero.

In the special case of separable correlations:

- $\mathbf{V}$ is defined by the eigenspaces of the transmit correlation, $\mathbf{\Theta}_{\mathrm{T}}$, as claimed (for the special case of Gaussian channels with $\mathbf{\Theta}_{\mathrm{R}}{=}\mathbf{I}$) in [3].

- The receive correlation, $\mathbf{\Theta}_{\mathrm{R}}$, enters only the computation of the powers in $\mathbf{P}_{\mathsf{c}}$. Moreover, in terms of the low- and high-SNR asymptotes, it plays no role at all.

3.2 Power Allocation: An Iterative Algorithm

For arbitrary SNR and number of antennas, the set of powers that satisfy the conditions in (3) can be found via the algorithm below, which is derived in the Appendix directly from those conditions. It is useful to introduce, as an auxiliary quantity, the mean-square error exhibited at the output of a linear MMSE receiver by the signal transmitted along the j-th eigenvector in $\mathbf{V}$. Defining

$$\mathbf{B}_j \triangleq \left(\mathbf{I} + \frac{\mathsf{SNR}}{n_\mathrm{T}}(\tilde{\mathbf{H}})_{-j}(\mathbf{P})_{-j}(\tilde{\mathbf{H}})^{\dagger}_{-j}\right)^{-1} \tag{4}$$

such error is given by [10]

$$\mathsf{MSE}_j = \frac{1}{1 + (\mathbf{P})_{j,j}\frac{\mathsf{SNR}}{n_\mathrm{T}}(\tilde{\mathbf{H}})^{\dagger}_j \mathbf{B}_j (\tilde{\mathbf{H}})_j}. \tag{5}$$

In order to accommodate the iterative nature of the algorithm, we use $(\cdot)^{(k)}$ to index the succession of values taken by the power allocation $\mathbf{P}$ and by other quantities that depend on $\mathbf{P}$. If no prior information about $\mathbf{P}_\mathsf{c}$ is available, the most reasonable initial allocation is $\mathbf{I}$, after which each recursion consists of two steps:

1. For j=$1, \ldots, n_\mathrm{T}$, let

$$p_j^{(k+1)} = \max\left\{ \frac{E\left[\mathrm{Tr}\left\{\mathbf{B}_j^{(k)}\right\}\right] + n_\mathrm{T}\frac{1-E\left[\mathsf{MSE}_j^{(k)}\right]}{(\mathbf{P})_{j,j}^{(k)}} - n_\mathrm{R}}{\frac{\mathsf{SNR}}{n_\mathrm{T}} E\left[\mathsf{MSE}_j^{(k)}(\tilde{\mathbf{H}})^{\dagger}_j \left(\mathbf{B}_j^{(k)}\right)^2 (\tilde{\mathbf{H}})_j\right]}, (\mathbf{P})_{j,j}^{(k)} \right\}. \tag{6}$$

2. Obtain the new power allocation as

$$\mathbf{P}^{(k+1)} = \frac{n_\mathrm{T}}{\sum_{j=1}^{n_\mathrm{T}} p_j^{(k+1)}} \mathrm{diag}\{p_1^{(k+1)}, p_2^{(k+1)}, \ldots, p_{n_\mathrm{T}}^{(k+1)}\}. \tag{7}$$

The scaling performed by (7) simply ensures that the total transmitted power is held at the correct value throughout the recursions. It can be verified that the capacity-achieving allocation, $\mathbf{P}_\mathsf{c}$, is the only fixed point of the algorithm.

3.3 Isotropic Input: When does it Achieve Capacity ?

If $\mathbf{P}_\mathsf{c}$=$\mathbf{I}$, then $\mathbf{\Phi}_\mathsf{c}$=$\mathbf{I}$ and thus an isotropic input suffices. There are relevant channel structures—beyond the IID Gaussian case reported in [1]—for which this is the case.

Proposition 1 *Let the entries of* $\tilde{\mathbf{H}}$ *in (2) be Gaussian. Denote by* $\tilde{\mathbf{G}}$ *a matrix such that* $(\tilde{\mathbf{G}})_{i,j}$=$E[|(\tilde{\mathbf{H}})_{i,j}|^2]$. *If* $\tilde{\mathbf{G}}$ *is column-regular as per Definition 2, then* $\mathbf{\Phi}_\mathsf{c}$=$\mathbf{I}$ *achieves capacity.*

The proof, as well as a more general—but less intuitive—condition for non-Gaussian channels, can be found in the Appendix.

4 Examples

The algorithm presented in Section 3.2 exhibits remarkable convergence speed. Moreover, convergence appears to take place irrespective of the (nonzero) allocation used to initialize the recursions. In this final section, we illustrate this process.

Example 1 *Consider a* 3*-antenna uniform linear transmit array with* 1*-wavelength antenna spacing and a broadside (truncated) Gaussian power azimuth spectrum with a* $2°$ *root-mean-square spread. The corresponding transmit correlation is* $(\mathbf{\Theta}_{\mathrm{T}})_{i,j} \approx e^{-0.05\,(i-j)^2}$ *[11]. Further consider* 4 *uncorrelated receive antennas. Signalling over the eigenvectors of* $\mathbf{\Theta}_{\mathrm{T}}$*, the performance of the power allocation algorithm at* $\mathsf{SNR}{=}-3$ dB *and* 6 dB *is depicted in Fig. 1.*

Example 2 *Consider a* $(4{\times}2)$ *channel* $\tilde{\mathbf{H}}$ *whose entries are Gaussian and independent with variances*

$$\tilde{\mathbf{G}} = \begin{pmatrix} 5/3 & 1/3 & 5/3 & 1/3 \\ 1/3 & 5/3 & 1/3 & 5/3 \end{pmatrix}$$

The performance of the power allocation algorithm at $\mathsf{SNR}{=}3$ dB *is depicted in Fig. 2.*

For both examples, the expectations in (6) are computed as averages of 10000 independent realizations of $\mathbf{H}$.

Appendix

Proof of Theorem 1

Let $\mathbf{\Phi}_{\mathsf{c}}$ be the covariance that maximizes (1). Define $\mathbf{P}_{\mathsf{c}} \triangleq \mathbf{V}^\dagger \mathbf{\Phi}_{\mathsf{c}} \mathbf{V}$ with $\mathbf{V}$ the eigenvector matrix of $E\left[\mathbf{H}^\dagger \mathbf{H}\right]$. Since $\mathrm{Tr}\{\mathbf{P}_{\mathsf{c}}\}{=}\mathrm{Tr}\{\mathbf{\Phi}_{\mathsf{c}}\}$, the capacity in terms of $\mathbf{P}_{\mathsf{c}}$ is

$$C = E\left[\log_2 \det\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}} \tilde{\mathbf{H}}^\dagger\right)\right] \tag{8}$$

where $\tilde{\mathbf{H}} \triangleq \mathbf{H}\mathbf{V}$. Let $\mathbf{P}_{\mathsf{c}}{=}\mathbf{P}_{\mathsf{c}}^{\mathsf{d}}{+}\mathbf{P}_{\mathsf{c}}^{\mathsf{off}}$ with $\mathbf{P}_{\mathsf{c}}^{\mathsf{d}} \triangleq \mathrm{diag}\{\mathbf{P}_{\mathsf{c}}\}$ and $\mathbf{P}_{\mathsf{c}}^{\mathsf{off}}$ having zeros along its diagonal. Further defining

$$\mathbf{Q} \triangleq \tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{off}} \tilde{\mathbf{H}}^\dagger \left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{d}} \tilde{\mathbf{H}}^\dagger\right)^{-1}$$

the capacity in (8) can be expanded as

$$C = E\left[\log_2 \det\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{d}} \tilde{\mathbf{H}}^\dagger\right)\right] + E\left[\log_2 \det\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \mathbf{Q}\right)\right]. \tag{9}$$

Next, we show that the second term in the right-hand side of (9) is negative and, hence, that $\mathbf{P}_{\mathsf{c}}$ must diagonal:

$$\begin{aligned}
E\left[\log_2 \det\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \mathbf{Q}\right)\right] &= E\left[\log_2 \det\left(\begin{bmatrix} \mathbf{I} & -\tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{off}} \tilde{\mathbf{H}}^\dagger \\ \left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{d}} \tilde{\mathbf{H}}^\dagger\right)^{-1} & \mathbf{I} \end{bmatrix}\right)\right] && (10) \\
&\leq \log_2 \det\left(E\begin{bmatrix} \mathbf{I} & -\tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{off}} \tilde{\mathbf{H}}^\dagger \\ \left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{d}} \tilde{\mathbf{H}}^\dagger\right)^{-1} & \mathbf{I} \end{bmatrix}\right) && (11) \\
&= \log_2 \det\left(\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ E\left[\left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \tilde{\mathbf{H}} \mathbf{P}_{\mathsf{c}}^{\mathsf{d}} \tilde{\mathbf{H}}^\dagger\right)^{-1}\right] & \mathbf{I} \end{bmatrix}\right) && (12) \\
&= 0 && (13)
\end{aligned}$$

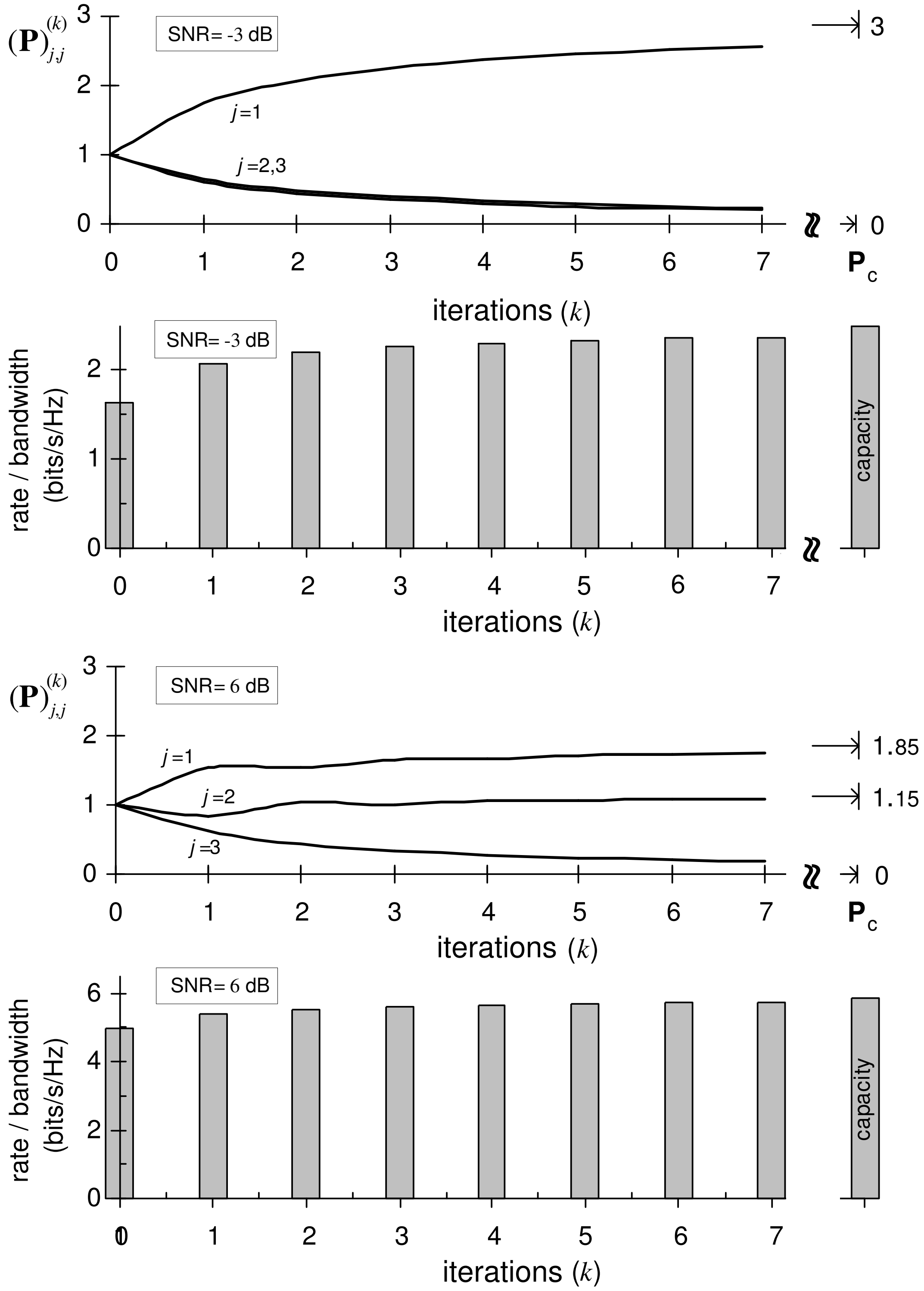

Figure 1: With $n_{\mathrm{T}}=3$ and $n_{\mathrm{R}}=4$ as per Example 1, values taken by $\mathbf{P}^{(k)}$ for $k=1,\ldots,7$ at SNR$=-3$ dB and 6 dB given the initialization $\mathbf{P}^{(0)}=\mathbf{I}$. Also shown are the corresponding rates per unit bandwidth. The rightmost values within each chart specify the actual $\mathbf{P}_{\mathrm{c}}$ (obtained numerically) and the corresponding capacity.

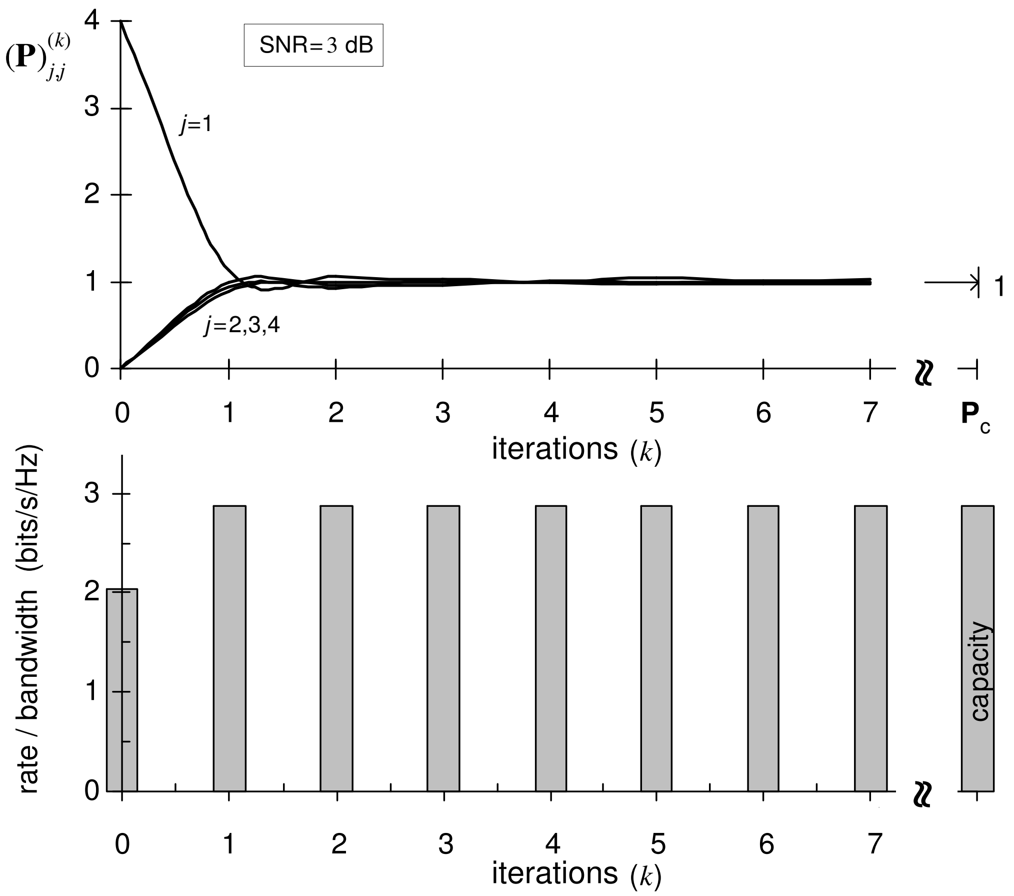

Figure 2: With n_T=4 and n_R=2 as per Example 2, values taken by $\mathbf{P}^{(k)}$ for k=1, ..., 7 at SNR=3 dB with initialization $\mathbf{P}^{(0)}$=diag$\{4, 0, 0, 0\}$. Also shown are the corresponding rates per unit bandwidth. The rightmost values specify the actual $\mathbf{P}_c$=$\mathbf{I}$ (given by Proposition 1) and the corresponding capacity.

where (10) follows from

$$\det\left(\left[\begin{array}{cc}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{array}\right]\right) = \det\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)\det\left(\mathbf{A}_{22}\right),$$

(11) applies Jensen's inequality, (12) follows from the diagonal structure of $E[\tilde{\mathbf{H}}^\dagger\tilde{\mathbf{H}}]$ and the off-diagonal structure of $\mathbf{P}_c^{\text{off}}$, and (13) is given by the unit determinant of a triangular matrix whose diagonal elements are 1.

Altogether then, the capacity-achieving covariance is $\mathbf{\Phi}_c$=$\mathbf{V}\mathbf{P}_c\mathbf{V}^\dagger$ where the eigenvectors in $\mathbf{V}$ equal those of $E[\mathbf{H}^\dagger\mathbf{H}]$ while $\mathbf{P}_c$ is the diagonal eigenvalue matrix that maximizes (8). The determination of $\mathbf{P}_c$ is thus a maximization problem for the concave function

$$I(\mathbf{P}) = E\left[\log_2 \det\left(\mathbf{I} + \frac{\text{SNR}}{n_T}\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^\dagger\right)\right] \tag{14}$$

over the convex set of diagonal real nonnegative matrices $\mathbf{P}$ whose trace equals n_T.

This maximum is characterized by a set of Kuhn-Tucker conditions, which we derive following the footsteps of [12, Section 15.2]. Hence, we impose that the derivative of (14) in the direction from $\mathbf{P}_{\mathsf{c}}$ to any alternative matrix $\mathbf{P}$ be nonpositive. Letting

$$\mathbf{P}_{\mu} = (1-\mu)\mathbf{P}_{\mathsf{c}} + \mu\mathbf{P}$$

for $0 \le \mu \le 1$, the one-side derivative of (14) with respect to μ at $\mu = 0^{+}$ is

$$\frac{d}{d\mu} E\left[\log_2 \det\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\mathrm{T}}}\tilde{\mathbf{H}}\mathbf{P}_{\mu}\tilde{\mathbf{H}}^{\dagger}\right)\right] = E\left[\mathrm{Tr}\left\{\left(\mathbf{I} + \tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^{\dagger}\right)\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\mathrm{T}}}\tilde{\mathbf{H}}\mathbf{P}_{\mathsf{c}}\tilde{\mathbf{H}}^{\dagger}\right)^{-1} - \mathbf{I}\right\}\right] \tag{15}$$

and, therefore, we impose that

$$E\left[\mathrm{Tr}\left\{\left(\mathbf{I} + \tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^{\dagger}\right)\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\mathrm{T}}}\tilde{\mathbf{H}}\mathbf{P}_{\mathsf{c}}\tilde{\mathbf{H}}^{\dagger}\right)^{-1} - \mathbf{I}\right\}\right] \le 0$$

for every $\mathbf{P}$ in the set. Following considerations analogous to those in [12], the Kuhn-Tucker conditions that characterize $\mathbf{P}_{\mathsf{c}}$ become equivalent to the necessary and sufficient conditions in (3).

Derivation of Iterative Algorithm

Using (4) and (5) and with the aid of the matrix inversion lemma,

$$E\left[\mathrm{Tr}\left\{\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\mathrm{T}}}\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^{\dagger}\right)^{-1}\right\}\right] = E\left[\mathrm{Tr}\left\{\mathbf{B}_j\right\}\right] - (\mathbf{P})_{j,j}\tfrac{\mathsf{SNR}}{n_{\mathrm{T}}} E\left[\mathsf{MSE}_j(\tilde{\mathbf{H}})_j^{\dagger}\mathbf{B}_j^2(\tilde{\mathbf{H}})_j\right]$$

which can be used to expand (3) into

$$\frac{E\left[\mathrm{Tr}\left\{\mathbf{B}_j\right\}\right] + \frac{n_{\mathrm{T}}}{(\mathbf{P})_{j,j}}\left(1 - E[\mathsf{MSE}_j]\right) - n_{\mathrm{R}}}{\frac{\mathsf{SNR}}{n_{\mathrm{T}}} E\left[\mathsf{MSE}_j(\tilde{\mathbf{H}})_j^{\dagger}\mathbf{B}_j^2(\tilde{\mathbf{H}})_j\right]} \le (\mathbf{P})_{j,j} \tag{16}$$

where we have further used

$$(\mathbf{P})_{j,j}\tfrac{\mathsf{SNR}}{n_{\mathrm{T}}}(\tilde{\mathbf{H}})_j^{\dagger}\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\mathrm{T}}}\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^{\dagger}\right)^{-1}(\tilde{\mathbf{H}})_j = 1 - \mathsf{MSE}_j.$$

The algorithm is based on selecting, at every recursion, the largest of the two quantities at either side of (16), with an additional scaling step that ensures that $\mathrm{Tr}\{\mathbf{P}\} = n_{\mathrm{T}}$.

Proof of Proposition 1

Given any diagonal matrix $\mathbf{P}$ such that $\mathrm{Tr}\{\mathbf{P}\} = n_{\mathrm{T}}$, denote by $\mathbf{P}^{\{m\}}$ its cyclic shift by m positions, i.e. another diagonal matrix such that

$$(\mathbf{P})_{j,j}^{\{m\}} = (\mathbf{P})_{j',j'}$$

with $j' = (j-m)_{\mathrm{mod}\, n_{\mathrm{T}}}$. Clearly, $\mathrm{Tr}\{\mathbf{P}^{\{m\}}\} = n_{\mathrm{T}}$ for any shift $m = 1, \ldots, n_{\mathrm{T}}$ and

$$\tfrac{1}{n_{\mathrm{T}}}\sum_{m=1}^{n_{\mathrm{T}}} \mathbf{P}^{\{m\}} = \mathbf{I}$$

Denote by $\text{vec}(\tilde{\mathbf{H}})$ the $n_{\text{R}}n_{\text{T}}$-dimensional vector obtained by stacking up the columns of $\tilde{\mathbf{H}}$. Invoking Jensen's inequality,

$$\begin{aligned} E\left[\log_2 \det\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\text{T}}}\tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\dagger}\right)\right] &\geq \frac{1}{n_{\text{T}}}\sum_{m=1}^{n_{\text{T}}} E\left[\log_2 \det\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\text{T}}}\tilde{\mathbf{H}}\mathbf{P}^{\{m\}}\tilde{\mathbf{H}}^{\dagger}\right)\right] \\ &= E\left[\log_2 \det\left(\mathbf{I} + \tfrac{\mathsf{SNR}}{n_{\text{T}}}\tilde{\mathbf{H}}\mathbf{P}^{\{m\}}\tilde{\mathbf{H}}^{\dagger}\right)\right] \qquad (17) \end{aligned}$$

where (17) holds if the joint distribution of $\text{vec}(\tilde{\mathbf{H}})$ equals that of $\text{vec}(\tilde{\mathbf{H}}^{\{m\}})$, with the columns of $\tilde{\mathbf{H}}^{\{m\}}$ being cyclic shifted versions of the columns of $\tilde{\mathbf{H}}$. It follows that $\mathbf{I}$ is the capacity-achieving input covariance whenever

$$f_{\text{vec}(\tilde{\mathbf{H}})}(\cdot) = f_{\text{vec}(\tilde{\mathbf{H}}^{\{m\}})}(\cdot). \qquad (18)$$

Sufficient condition for (18) is that the columns of $\tilde{\mathbf{H}}$ be independent and marginally identically distributed. If $\tilde{\mathbf{H}}$ is Gaussian, then this condition reverts to the column-regularity of $\tilde{\mathbf{G}}$.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585-595, Nov. 1999.

[2] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632-2639, Sep. 2001.

[3] S. A. Jafar, S. Vishwanath and A. J. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," *IEEE Intern. Conf. Communications (ICC'01)*, vol. 7, pp. 2266-2270, 2001.

[4] H. Boche and E. Jorswieck, "Optimal power allocation for MISO systems and complete characterization of the impact of correlation on the capacity," *IEEE Intern. Conf. on Acoustics, Speech and Signal Proc. (ICCASP'03)*, vol. 4, pp. 373-376, Apr. 2003.

[5] A. Narula, M. J. Lopez, M. D. Trott and G. W. Wornell, "Efficient Use of Side Information in Multiple-Antenna Data Transmission over Fading Channels," *IEEE J. Select. Areas in Commun.*, vol. 16, No. 8, pp. 1423-1436, Oct. 1998.

[6] S. Verdú, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inform. Theory*, vol. 48, No. 6, pp. 1319–1343, June 2002.

[7] D.-S. Shiu, G. J. Foschini, M. J. Gans and J. M. Kahn, "Fading Correlation and Its Effects on the Capacity of Multielement Antenna Systems," *IEEE Trans. Commun.*, vol. 48, No. 3, Mar. 2000.

[8] A. M. Tulino, S. Verdú and A. Lozano "Capacity of antenna arrays with space, polarization and pattern diversity," *IEEE Inform. Theory Workshop (ITW'03)*, Apr. 2003.

[9] W. K. Pratt, *Digital Image Processing*, New York: Wiley, 1978.

[10] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.

[11] T.-S. Chu, L. Greenstein, "A Semiempirical Representation of Antenna Diversity Gain at Cellular and PCS Base Stations," *IEEE Trans. Commun.*, vol. 45, June 1997.

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