The Algebra of MIMO Channels

Emmanuel Abbe* Emre Telatar† Lizhong Zheng*

∗ Massachusetts Institute of Technology
Laboratory for Information and Decision Systems
Cambridge, MA 02139, USA
eabbe@mit.edu, lizhong@mit.edu

† École Polytechnique Fédérale de Lausanne
Laboratory of Information Theory
1015 Lausanne, Switzerland
emre.telatar@epfl.ch

Abstract

We consider ergodic coherent MIMO channels. We characterize the optimal input distribution for various fading matrix distributions. First, we describe how symmetries in the fading matrix distribution are preserved as symmetries in the optimal input covariance and thus yield to specification of the optimal input. We will see that group structures and notion of commutant appear as key elements. Second, we investigate the Kronecker model, in this case we will show how an asymmetric structure in the problem is also preserved in the optimal input structure.

Notations:
We define the following subsets of the $n \times n$ complex matrices $M_n(\mathbb{C})$, $n \geq 1$:
$H(n)$: the hermitian matrices,
$H_+(n)$: the hermitian positive semidefinite matrices,
$H_+^*(n)$: the hermitian positive definite matrices,
$U(n)$: the unitary matrices,
$S(n)$: the symmetric group of all permutations matrices,
$C(n)$: the group of cyclic permutations matrices,
$Z_2(n)$: the group of diagonal matrices with $\{-1, +1\}$ elements.

1 Introduction

We consider a channel in which a vector input $x \in \mathcal{X} = \mathbb{C}^t$, $t \geq 1$, is received as a vector output $y \in \mathcal{Y} = \mathbb{C}^r$, $r \geq 1$, under the following assumptions. At each use $i \geq 1$ of the channel:

- an $r \times t$ matrix H_i is drawn from an ergodic process having marginal probability measure μ_H,
- an $r \times 1$ vector n_i is drawn i.i.d from a complex circularly-symmetric gaussian (C.C.S.G.) random variable of covariance matrix K, independently from the H_i's,
- the transmitter, without knowing the H_i's and n_i's, sends x_i,
- the receiver gets $y_i = H_i x_i + n_i$ together with H_i (and hence the term "coherent").

Moreover, the inputs $\{x_i\}$ are constrained in the following way. If we want to use a code $\mathcal{C} = \{c(1), \ldots, c(M)\} \subset \mathcal{X}^n$, $n \geq 1$, then the code has to satisfy that for every $1 \leq m \leq M$: $\frac{1}{n} b(c(m)) \in \mathrm{D}_t$, where $b(c(m)) = \sum_{i=1}^{n} c(m)_i c(m)_i^*$ and where $\mathrm{D}_t \subset H_+^*(t)$ is a given compact set. Let C be the capacity of this channel under this constraint. Then, denoting by X a random vector (r.ve.) in $\mathbb{C}^t$, we know from standard information theoretic arguments that

$$C(\mu_H, \mathrm{C}_t) = \max_{X \in \mathrm{C}_t} I(X; Y, H)$$

where

1. $\mathrm{C}_t = \{X | \mathbb{E}XX^* \in \mathrm{D}_t\}$ and $\mathrm{D}_t \subset H_+^*(t)$ a compact set
2. $N \overset{(d)}{\sim} \mathcal{N}_{\mathbb{C}^r}(K)$ with $K \in H_+^*(r)$
3. H is a $\mathbb{C}^{r \times t}$-random matrix with probability measure μ_H,
4. (X, H, N) are mutually independent,
5. $Y = HX + N$.

When $t = 1$, we maximize the mutual information over random variable (r.v.) X having variance in a compact set of $\mathbb{R}_+$, with maximal value, say, $P \in \mathbb{R}_+$. In this case, the optimal input is known to be a C.C.S.G. r.v. with variance P, no matter what μ_H is. More generally, one can show with usual arguments that in the vector settings the gaussian distribution is still optimal, but an optimization remains to be done on the covariance matrices in D_t, the result of which may depend on the distribution μ_H.
In the case where H has i.i.d. C.C.S.G. entries and D_t is the set of covariance matrices with trace bounded by a given value $P \in \mathbb{R}$, it has been shown in [1] that the optimal covariance matrix is $\frac{P}{t} I_t$ and the capacity is linearly increasing with $\min(t, r)$.
Questions:
1. The solution found when H has i.i.d C.C.S.G. entries is not surprising, in the sense that there are enough symmetries in the problem so that we expect a symmetric solution. But what does *enough symmetry* mean? What can we say when we have different **symmetric structures**, such as for example when we only have i.i.d entries? What are the relevant concepts of symmetry and how can we convert them into specification of the solution?
2. What can we do when we have **asymmetric structures**?

We will develop some algebraic tools and present a result that gives an answer to the first question. We will see how group structure and notion of commutant comes into the picture as key features. This result is also applicable to other functionals than the capacity of MIMO channels.
We will then investigate the Kronecker model (defined later) to get an understanding of the second question.

2 General expression of the capacity

Definition 1. We define the optimal inputs by

$$X_c(\mu_H, \mathrm{C}_t) = \arg\max_{X \in \mathrm{C}_t} I(X; HX + N, H),$$

where $\arg\max_{X\in \mathrm{C}_t} f(X)$, for a real function f, denotes the set of the elements x satisfying $f(x) \geq f(y)$, $\forall y \in \mathrm{C}_t$.

We now use the assumptions we made on the channel to give a more specific expression for the capacity and the optimal inputs. The fact that the gaussian distribution maximizes the entropy under a covariance constraint lead to the following result.

Proposition 1. *According to previous definitions and assumptions, we have*

$$C(\mu_H, \mathrm{C}_t) = \max_{Q\in \mathrm{D}_t} \psi(Q),$$

where

$$\psi(Q) = \mathbb{E}^{\mu_H}\phi(HQH^*), \quad \phi(A) = \log\det(I + K^{-1}A) \tag{1}$$

and

$$X_c(\mu_H, \mathrm{C}_t) \sim \mathcal{N}_{\mathbb{C}^t}(Q_c),$$

where

$$Q_c = \arg\max_{Q\in \mathrm{D}_t} \psi(Q).$$

Back to our questions: To develop an understanding of the question of symmetry, we consider the following situations. Imagine that there is not difference of output probability when sending the input x or a permuted version of it, say, Px, where P is a permutation matrix. This is to say that the distribution of H is the same as the distribution of HP. What if we had another symmetry, such as when the channel behaves the same whenever x is sent or a modified version of it in which we changed the sign of some of its components, for example Zx, where Z is a diagonal matrix with 1 and -1 on the diagonal.
The first observation is that if H has same the distribution as HP, it also has same distribution as HP^n, $n \geq 1$. More generally, if H has same distribution as Hs, for all s in a set $S \subset GL_n(\mathbb{C})$ (the set of invertible matrices in $M_n(\mathbb{C})$), then this is still true for the group generated by S. This motivates the following definition.

Definition 2. For a group $G \subset M_n(\mathbb{C})$, if $Hg \overset{(d)}{\simeq} H$, $\forall g \in G$, then H is said to be **G-invariant on the right**.

Examples of groups in $M_n(\mathbb{C})$ are $U(n)$, $Z_2(n)$, $C(n)$ and $S(n)$ (with the usual matrix multiplication). We now gave a definition to quantify symmetries in the problem, through **the group of invariance** of H, the question is then: how do we use this invariance in order to get knowledge on the optimal input? In the next section we will see that this is done through **the commutant**.

3 Invariance in Conjugation and Commutant

Definition 3. Let G be a group in $M_n(\mathbb{C})$ and $D \subset M_n(\mathbb{C})$.

- **Invariance in conjugation:** we say that $\Psi : D \to \mathbb{R}$ is invariant in G-conjugation if

1. D is invariant in G-conjugation in the sense $gDg^{-1} = D, \forall g \in G$.
2. $\Psi(gdg^{-1}) = \Psi(d), \forall d \in D, g \in G$.

- We denote by $\mathbb{G}$ a random variable with values in G and probability measure $P_G \in M_1(G)$. We define $d^{P_G} := \mathbb{E}^{P_G}\mathbb{G}D\mathbb{G}^{-1}$ and U_G the normalized Haar measure on the right on G.

- **The commutant algebra** of G is defined by $\mathrm{Comm}(G) = \{a \in M_n(\mathbb{C}) | ag = ga, \forall g \in G\}$

Interpretation and examples: Examples of functions which are invariant in G-conjugation are all functions of the form $x \rightsquigarrow \mathbb{E}f(MxM^{-1})$ where f is any measurable function and M is a random matrix that is G-invariant on the right, which, we recall, is defined by $Mg \stackrel{(d)}{=} M$, $\forall g \in G$. The reason for which a "conjugation" invariance is interesting in our MIMO settings is a consequence of the fact that we are working with second order moment constraint and implying that the mutual information has precisely the above described form (cf. (1)). We will see in the next proposition why the commutant notion is relevant in our problem, here are first some examples of commutant algebras:

$$\mathrm{Comm}(Z_2(n)) \text{ is the set diagonal matrices in } M_n(\mathbb{C}),$$

$$\mathrm{Comm}(S(n)) = \{\alpha I_n + \beta J_n | \alpha, \beta \in \mathbb{C}\}, \text{ where } J_n = 1^{n\times n},$$

$$\mathrm{Comm}(U(n)) = \{\alpha I_n | \alpha \in \mathbb{C}\}.$$

Proposition 2. *Let $G \subset M_n(\mathbb{C})$ be a compact group and $D \subset M_n(\mathbb{C})$ a convex set. Let $\Psi : D \stackrel{C^0}{\rightarrow} \mathbb{R}$ a concave function which is invariant in G-conjugation. Then*

$$\Psi(d) \leq \Psi(d^{P_G}) \leq \Psi(d^{U_G}), \ \forall d \in D,$$

in particular, among the maximizers of the function Ψ in D there is one that belongs to

$$\mathrm{Comm}(G).$$

Proof. Let $d \in D$ and $P_G, Q_G \in M_1(G)$. Because D is convex and $GDG^{-1} = D$, $d^{P_G} \in D$. By Jensen's inequality, we have

$$\mathbb{E}^{P_G}\Psi(\mathbb{G}d\mathbb{G}^{-1}) \leq \Psi(d^{P_G}),$$

but by hypothesis, $\Psi(GdG^{-1}) = \Psi(d), \forall d \in D$, thus last inequality becomes

$$\Psi(d) \leq \Psi(d^{P_G}).$$

Note that

$$\left(d^{P_G}\right)^{Q_G} = d^{P_G \star Q_G},$$

with $P_G \star Q_G = \int_G (\tau_h)_* Q_G P_G(dh)$, where $(\tau_h)_* Q_G(\Gamma) = Q_G(\Gamma h^{-1})$, for $\Gamma \in \mathcal{B}_G$. Furthermore,

$$P_G \star U_G = U_G \star P_G = U_G.$$

Therefore we have

$$\Psi(d) \leq \Psi(d^{P_G}) \leq \Psi(\left(d^{P_G}\right)^{U_G}) = \Psi(d^{U_G}),$$

which proves the first part of the proposition. Finally, observe that for $d \in D$, $d^{U_G} \in \mathrm{Comm}(G) \cap D$, proving the second assertion.

□

4 Symmetries in MIMO

We consider the MIMO channel functional $\psi(Q) = \mathbb{E}^{\mu_H} \log\det(I + K^{-1}HQH^*)$, described in proposition (1), with D_t assumed to be convex.

Corollary 1. *Let G_1, G_2 be two groups in $U(t)$. If D_t is invariant in G_1-conjugation and μ_H is such that $HG_2 \overset{(d)}{=} H$, then*

$$Q_c \in \mathrm{Comm}(G_1 \cap G_2) \cap \mathrm{D}_t.$$

Proof. Observe that the function ψ in (1) is strictly concave and $(G_1 \cap G_2)$-invariant, also note that $u^{-1} = u^*$ for $u \in U(t)$, then use proposition 2. □

Corollary 1a: Total power constraint: for a given $P \in \mathbb{R}_+$, we consider $Q \in \mathcal{D}_t = \{Q \in H_+^*(t) | \mathrm{tr}(Q) \leq P\}$. If μ_H is invariant in G-conjugation for a subgroup G of $U(t)$, then $Q_c \in \mathrm{Comm}(G) \cap \mathcal{D}_t$.
Simply observe that $\mathcal{D}_t$ is invariant in $U(t)$-conjugation. Two interesting cases of subgroups of $U(t)$ are $S(t)$ and $Z_2(t)$. From what we saw in the examples of the commutant, if we consider a distribution μ_H invariant under $Z_2(t)$, then Q_c is diagonal and if it is invariant under $S(t)$, then Q_c will have the same value for all components inside the diagonal ($\frac{P}{t}$ if one works in $\mathcal{D}_t$) and also the same value for all elements outside the diagonal, as long as it stays a positive definite matrix. Examples of $Z_2(t)$-invariant random matrices are matrices with **independent symmetric entries** (symmetric means that $H_{ij} \overset{(d)}{\simeq} -H_{ij}$) and examples of $S(t)$-invariant ones are matrices with **i.i.d. entries** or **jointly gaussian entries having a covariance matrix of the form** $\alpha I_t + \beta J_t$.

Corollary 1b: Considering $X \in \mathcal{C}_t$, if H is $S(t)Z_2(t)$-invariant, which is for example the case when H_{ij} are i.i.d. and $H_{ij} \overset{(d)}{\simeq} -H_{ij}$, $\forall 1 \leq i \leq r, 1 \leq j \leq t$, then $Q_c = \frac{P}{t} I_t$.

This is a particular case of corollary 1a, where we consider the product group $S(t)Z_2(t) \subset U(t)$ containing all permutations matrices with $+1$ and -1. In this case we have that $\mathrm{Comm}(S(t)Z_2(t))$ contains only multiples of the identity and since $Q \in \mathcal{D}_t$ has normalized trace, the result follows. Note that we did not assume that the entries of H are gaussian (which is a particular case of the above case) in order to get $\frac{P}{t} I_t$ as a maximizer. Also note that the group $S(t)$ could be replaced by $C(t)$ and we would get the same conclusion. Generally, this will be true as long as we have a group G of invariance where $\mathrm{Comm}(G)$ contains only the multiple of the identity.

Corollary 1c: Local power constraint: if X is constrained by $\mathbb{E}|X_i|^2 \leq P_i$ for given $P_i \in \mathbb{R}_+$, $\forall 1 \leq i \leq t$, and if H is $Z_2(t)$-invariant, then $Q_c = \mathrm{diag}(P_1, \ldots, P_t)$.

The constraint $\mathbb{E}|X_i|^2 \leq P_i$ implies that $Q \in \widetilde{\mathcal{D}}_t = \{Q \in H_+(t) | Q_{ii} \leq P_i, \forall 1 \leq i \leq t\}$, now we no longer have that $\widetilde{\mathcal{D}}_t$ is invariant in $U(t)$-conjugation, but we still have, for example, invariance in $Z_2(t)$-conjugation. Therefore, if H is $Z_2(t)$-invariant, the optimal covariance matrix will be commuting with this group, which means it is diagonal and thus the optimal diagonal elements are the corresponding P_i's.

Conclusion: As it has been illustrated in previous example, the problem of symmetry should be generally approached in the following way: first identify the invariance property of the domain D_t in which we are working (we saw examples of total (see corollary 1a) and local (see 1c) power constrain, several intermediary cases are possible), then identify the invariance property of the fading matrix distribution μ_H, once we have these two groups of invariance, we know that we can restrict our search of Q_c to matrices commuting with these groups and staying in D_t. Which means that **the commutant is summarizing the set exploiting the information given by the symmetries in the problem**. We saw that in some cases (see corollary 1b) this allows us to fully specify the optimal input covariance matrix, whereas in other cases, it only reduces the dimension of the optimization problem (such as for example in corollary 1b, when we have a $Z_2(t)$-invariance, we are left with t degrees of freedom for Q_c instead of $\frac{t^2+t}{2}$ at the beginning).

In other words, the final picture is the following: what is quantifying the symmetry is **the group of invariances** of H and how it is transformed into a specification of the optimal input is done via **the commutant** of this group.

5 Asymmetries in MIMO

Suppose now that $H = WB$, where W has $r \times t$ i.i.d C.C.S.G entries and $B = \mathrm{diag}(b)$, with $b \in \mathbb{R}_+^t$ and that $b_1 \leq \ldots \leq b_n$. Can we then expect that the optimal covariance matrix should preserve this ordering in some sense?

The Kronecker model: We consider $X \in \mathcal{C}_t = \{X | \mathbb{E}XX^* \in \mathcal{D}_t\}$, $\mathcal{D}_t = \{Q \in H_+^*(t) | \mathrm{tr}(Q) \leq P\}$ $H = AWB$, where W is a $r \times t$ random matrix, $A \in M_r(\mathbb{C})$ non-zero, $B \in M_t(\mathbb{C})$ non-zero with SVD $B = U_B \,\mathrm{diag}(b) V_B^*$ and μ_W is $U(t)$-invariant on the right.

We will now present two propositions that will contribute to describe the optimal input for such a channel. If the random matrix W was replaced by the identity, we know that the optimal input covariance is diagonal with diagonal q^c given via "water-filing" on the singular values of B (cf. [1]). Two particular properties of the "water-filing" solution are the following. First, if $b_i \geq b_j$ then $q_i^c \geq q_j^c$ (with equality if $b_i = b_j$). Then, if $b_i < b_{i+1}$ are two consecutive values in b, and if b_{i+1} is bigger enough than b_i, then we might end up by sharing the whole power P on the $t - i$ biggest values of b. We will see in the next two propositions that this two properties are preserved.

Proposition 3. *We have*

$$Q_c = V_B \,\mathrm{diag}(q^c) V_B^*$$

where q^c satisfies

$$q_i^c \geq 0, \sum_{i=1}^{t} q_i^c = P, \quad q_i^c \geq q_j^c \text{ if } b_i > b_j, \text{ and } q_i^c = q_j^c \text{ if } b_i = b_j.$$

Note: *If $B = I_t$ and μ_W is G-invariant on the right with $G \leq U(t)$, then $Q_c \in \mathrm{Comm}(G) \cap \mathcal{D}_t$.*

Remark: This proposition says that the optimal covariance matrix has for eigenvectors, the eigenvectors on the right of B, and has eigenvalues which are monotonically distributed with respect to the singular values of B.

Let $\lambda_1(M) \leq \ldots \leq \lambda_n(M)$ denote the ordered eigenvalues of any matrix $M \in H(n)$.

Lemma 1. *Let $n \geq 1$, $P \in H_+^*(n)$ and $H \in H(n)$. We then have,*

$$\lambda_k(H+P) > \lambda_k(H), \ \ \forall k = 1, 2, \ldots, n.$$

Proof of proposition 4. From proposition 1, we know that $X_c \sim \mathcal{N}_{\mathbb{C}^t}(Q_c)$ where

$$Q_c = \arg\max_{Q \in \mathcal{D}_t} \psi_{A,B}(Q)$$

and

$$\psi_{A,B}(Q) = \mathbb{E}^{\mu} \log\det(I + K^{-1}AWBQB^*W^*A^*).$$

(Note that A affects the function ψ in the same way as K, in other words, we could consider one of this two matrix to be the identity, for example, assume i.i.d. components for the noise and set $\tilde{A} = K^{-\frac{1}{2}}A$.)

If $B = I_r$ any invariance properties on the right for μ_W will be preserved for AW, thus the note after the proposition is a direct consequence of corollary 1.

For general B, we have from the SVD theorem that there exist $U_B, V_B \in U(t)$ and $b \in \mathbb{R}_+^t$ such that $B = U_B \operatorname{diag}(b) V_B^*$. Assuming W to be unitary invariant allow us to write

$$\psi_{A,B}(Q) = \psi_{A,\operatorname{diag}(b)}(V_B^* Q V_B).$$

Therefore, as $V_B^* Q V_B \in \mathcal{D}_t$, we can focus on

$$Q_c^{'} = \arg\max_{Q \in \mathcal{D}_t} \psi_{A,\operatorname{diag}(b)}(Q).$$

Now, $\psi_{A,\operatorname{diag}(b)}$ is $Z_2(t)$-invariant, thus from proposition 1, we have

$$Q_c^{'} = \arg\max_{\operatorname{diag}(q) \in \mathcal{D}_t} \psi_{A,\operatorname{diag}(b)}(\operatorname{diag}(q)).$$

Let q^c be the the vector such that $\operatorname{diag}(q^c)$ is the maxima of $\psi_{A,\operatorname{diag}(b)}$ (which is a concave function on a compact set). First note that if $b_i = b_j$ then $\psi_{A,\operatorname{diag}(b)}$ is $S(t)_{ij}$-invariant, where $S(t)_{ij}$ is the subgroup of permutations keeping the diagonal elements different than i and j invariant (transposition), thus we get from proposition 1 that $q_i^c = q_j^c$.
Now, let $P^{'} = P - \sum_{i=3}^t q_i^c$, such that $q_1^c + q_2^c = P^{'}$. We will show that if $b_1 > b_2$, then for any $0 \leq P^{'} \leq P$,

$$\partial_{q_1} \psi_{A,\operatorname{diag}(b)}(\operatorname{diag}(q))\Big|_{(\frac{P'}{2},\frac{P'}{2},q_3^c,\ldots,q_t^c)} > \partial_{q_2} \psi_{A,\operatorname{diag}(b)}(\operatorname{diag}(q))\Big|_{(\frac{P'}{2},\frac{P'}{2},q_3^c,\ldots,q_t^c)},$$

which, by the concavity of $\psi_{\operatorname{diag}(q)}$, implies that

$$q_1^c > q_2^c.$$

By symmetry of the problem, this clearly implies the result for any components i and j (other than 1 and 2).

We have

$$\psi_{A,\operatorname{diag}(b)}(\operatorname{diag}(q)) = \mathbb{E} \log\det(I + \sum_{i=1}^t q_i b_i^2 A w_i (A w_i)^*)$$

where w_i is the i-th column of W. For an invertible matrix M, we have the formula

$$\partial_{m_{ij}} \log\det(M) = (M^{-1})_{ji},$$

therefore we have

$$\partial_{q_j}\psi_{A,\mathrm{diag}(b)}(\mathrm{diag}(q)) = b_j^2 \mathbb{E}\,\mathrm{tr}\left(I + \sum_{i=1}^{t} q_i b_i^2 A w_i (A w_i)^*\right)^{-1} A w_j (A w_j)^*.$$

Let us denote $X_i = Aw_i(Aw_i)^*$, which are hermitian positive semidefinite matrices, as well as $\left(I + \sum_{i=1}^t q_i b_i^2 X_i\right)$ which is in addition positive definite and invertible. We define $Z = \sum_{i=3}^t q_i b_i^2 X_i$ and $Z_c = \sum_{i=3}^t q_i^c b_i^2 X_i$, we then rewrite

$$\partial_{q_1}\psi_{A,\mathrm{diag}(b)}(\mathrm{diag}(q)) = b_1^2 \mathbb{E}\,\mathrm{tr}\big(I + q_1 b_1^2 X_1 + q_2 b_2^2 X_2 + Z\big)^{-1} X_1 \tag{2}$$

and

$$\begin{aligned}\partial_{q_2}\psi_{A,\mathrm{diag}(b)}(\mathrm{diag}(q)) &= b_2^2 \mathbb{E}\,\mathrm{tr}\big(I + q_1 b_1^2 X_1 + q_2 b_2^2 X_2 + Z\big)^{-1} X_2 \\ &= b_2^2 \mathbb{E}\,\mathrm{tr}\big(I + q_1 b_1^2 X_2 + q_2 b_2^2 X_1 + Z\big)^{-1} X_1 \end{aligned} \tag{3}$$

where in the last line we interchanged the random matrices X_1 and X_2, as W is $S(t)$-invariant. To conclude the proof, we have to show that if $b_1 > b_2$

$$b_1^2 \mathbb{E}\,\mathrm{tr}\left(I + \frac{P'}{2} b_1^2 X_1 + \frac{P'}{2} b_2^2 X_2 + Z_c\right)^{-1} X_1 > b_2^2 \mathbb{E}\,\mathrm{tr}\left(I + \frac{P'}{2} b_1^2 X_2 + \frac{P'}{2} b_2^2 X_1 + Z_c\right)^{-1} X_1,$$

for any $0 \leq P' \leq 1$. This is clearly verified in the scalar case ($r = 1$). In the matrix case, a couple of more steps (using the previous lemma) are required to show that the result hold.

We now define

$$\begin{aligned} \chi_1 :\ [0,1] &\rightarrow \mathbb{R} \\ \varepsilon &\mapsto \chi_1(\varepsilon) = b_1^2 \mathrm{tr}\Big(I + \tfrac{P'}{2} b_1^2 X_1(\varepsilon) + \tfrac{P'}{2} b_2^2 X_2(\varepsilon) + Z_c\Big)^{-1} X_1(\varepsilon) \end{aligned}$$

and

$$\begin{aligned} \chi_2 :\ [0,1] &\rightarrow \mathbb{R} \\ \varepsilon &\mapsto \chi_2(\varepsilon) = b_2^2 \mathrm{tr}\Big(I + \tfrac{P'}{2} b_1^2 X_2(\varepsilon) + \tfrac{P'}{2} b_2^2 X_1(\varepsilon) + Z_c\Big)^{-1} X_1(\varepsilon) \end{aligned}$$

where

$$X_i(\varepsilon) = X_i + \varepsilon I_r.$$

Note that for $i = 1, 2$, χ_i are continuous function. Therefore, $\lim_{\varepsilon \searrow 0} \chi_i(\varepsilon) = \chi_i(0)$. Moreover, from (2), we have

$$\mathbb{E}\chi_1(0) = \partial_{q_1}\psi_{A,\mathrm{diag}(b)}(\mathrm{diag}(q))\Big|_{(\frac{P'}{2}, \frac{P'}{2}, q_3^c, \ldots, q_t^c)}$$

and from (3)

$$\mathbb{E}\chi_2(0) = \partial_{q_2}\psi_{A,\mathrm{diag}(b)}(\mathrm{diag}(q))\Big|_{(\frac{P'}{2}, \frac{P'}{2}, q_3^c, \ldots, q_t^c)}.$$

Now, let us consider $\varepsilon \in (0,1]$, we have that $X_i(\varepsilon)$ is surely in $H_+^*(r)$, thus surely invertible, so we can write

$$\chi_i(\varepsilon) = \operatorname{tr}\left(\frac{P'}{2}I_r + M_i\right)^{-1}$$

or equivalently

$$\chi_j(\varepsilon) = \sum_{i=1}^{r} \frac{1}{\frac{P'}{2} + \lambda_i(M_j)}, \quad j = 1, 2$$

where

$$M_1 := X_1^{-1}(\varepsilon)\left(b_1^{-2}I_r + \frac{P'}{2}b_1^{-2}b_2^2X_2(\varepsilon) + b_1^{-2}Z_c\right)$$

and

$$M_2 := X_1^{-1}(\varepsilon)\left(b_2^{-2}I_r + \frac{P'}{2}b_2^{-2}b_1^2X_2(\varepsilon) + b_2^{-2}Z_c\right).$$

If we try to directly insert $X_1^{-1}(\varepsilon)$ in the parenthesis of above expressions, we will not be able to apply lemma 1 part *(ii)*, as $X_1^{-1}(\varepsilon)X_2(\varepsilon)$ may not be hermitian, even though $X_1^{-1}(\varepsilon) \in H_+^*(r)$ surely and $X_2(\varepsilon) \in H_+(n)$ surely. However, from lemma 1 part *(i)*, we have that the non-zero eigenvalues of M_1 are the same as the ones of

$$X_1^{-\frac{1}{2}}(\varepsilon)\left(b_1^{-2}I_r + \frac{P'}{2}b_1^{-2}b_2^2X_2(\varepsilon) + b_1^{-2}Z_c\right)X_1^{-\frac{1}{2}}(\varepsilon)$$

which is equal to

$$b_1^{-2}X_1^{-1}(\varepsilon) + \frac{P'}{2}b_1^{-2}b_2^2X_1^{-\frac{1}{2}}(\varepsilon)X_2(\varepsilon)X_1^{-\frac{1}{2}}(\varepsilon) + b_1^{-2}X_1^{-\frac{1}{2}}(\varepsilon)Z_cX_1^{-\frac{1}{2}}(\varepsilon) =: N_1$$

and that the non-zero eigenvalues of M_2 are the same as the ones of

$$X_1^{-\frac{1}{2}}(\varepsilon)\left(b_2^{-2}I_r + \frac{P'}{2}b_2^{-2}b_1^2X_2(\varepsilon) + b_2^{-2}Z_c\right)X_1^{-\frac{1}{2}}(\varepsilon)$$

which is equal to

$$b_2^{-2}X_1^{-1}(\varepsilon) + \frac{P'}{2}b_2^{-2}b_1^2X_1^{-\frac{1}{2}}(\varepsilon)X_2(\varepsilon)X_1^{-\frac{1}{2}}(\varepsilon) + b_2^{-2}X_1^{-\frac{1}{2}}(\varepsilon)Z_cX_1^{-\frac{1}{2}}(\varepsilon) =: N_2.$$

And now we have

$$N_1 - N_2 \in H_+^*(r) \text{ surely},$$

therefore, we conclude from lemma 1 that

$$\chi_1(\varepsilon) > \chi_2(\varepsilon) \text{ surely, } \forall \varepsilon \in (0,1].$$

Thus, by the continuity of χ_i on $[0,1]$ and monotony of the expectation, we have

$$\chi_1(0) \geq \chi_2(0) \Longrightarrow q_1^c \geq q_2^c$$

and we conclude the proof. □

Proposition 4. *Let $b_1 \leq b_2 \leq \ldots \leq b_t$. We assume that $r = 1$, $w_{1j} \overset{i.i.d.}{\sim} \mathcal{N}_{\mathbb{C}}(1)$, $\forall 1 \leq j \leq t$. Then, for all $j = 1, \ldots, t$, there exists $\bar{b} = \bar{b}(b_j) \geq 0$ such that*

$$\text{if } b_{j+1} \geq \bar{b} \text{ then } q_i^c = 0, \ \forall i = 1, \ldots, j.$$

Proof. In this setting we have $C(\mu_H, \mathcal{D}_t) = \log(1 + P\sum_{i=1}^t q_i d_i X_i) =: f(q)$, with $b_i^2 = d_i$, $X_i \overset{i.i.d.}{\sim} \mathcal{E}(1)$, $\forall 1 \leq i \leq t$ and $q \in \Theta(t)$. Let $Z_j = 1 + P\sum_{i \neq j,j+1} q_i X_i$. We then have

$$\partial_{q_j} f(q) = \mathbb{E}\frac{Pd_jX_j}{Z_j + Pd_jX_j + Pd_{j+1}X_{j+1}}.$$

Let $0 < T \leq 1$ and $p^{(j)}$ be a vector with $p_{j+1}^{(j)} = T$, $p_j^{(j)} = 0$ and thus $\sum_{i \neq j,j+1} p_i^{(j)} = 1 - T$. From the concavity of f, if

$$\partial_{q_j} f(q)|_{q=p^{(j)}} < \partial_{q_{j+1}} f(q)|_{q=p^{(j)}}, \quad \forall 0 < T \leq 1, \tag{4}$$

then $q_i^c = 0$, $\forall i = 1, \ldots, j$. Now, (4) becomes

$$\mathbb{E}\frac{Z + TPd_jX_j}{Z + TPd_{j+1}X_{j+1}} < 1, \quad \forall 0 < T \leq 1,$$

so if

$$\mathbb{E}\frac{z + TPd_jX_j}{z + TPd_{j+1}X_{j+1}} = \mathbb{E}\frac{z/T + Pd_jX_j}{z/T + Pd_{j+1}X_{j+1}} < 1, \quad \forall z \geq 1, 0 < T \leq 1,$$

we are done. Last inequality is equivalent to

$$\mathbb{E}\frac{1}{z + Pd_{j+1}X_{j+1}} < \frac{1}{z + Pd_j}, \quad \forall z \geq 1.$$

Let $F(a) = \mathbb{E}\frac{1}{z+aX}$, $a_{j+1} = Pd_{j+1}$ and $a_j = Pd_j$, we now wonder when we have

$$F(a_{j+1}/z) < \frac{1}{1 + a_j/z}, \quad \forall z \geq 1.$$

For a given $\beta \in \mathbb{R}_+$, let $\alpha(\beta)$ be the smallest number satisfying $F(\alpha(\beta)) < \frac{1}{1+\beta}$. Then if for any possible values of a_j,

$$\bar{a}(a_j) = \sup_{z \geq 1} z\alpha(a_j/z) < +\infty,$$

we deduce that for $a_j \geq \bar{a}(a_j)$, we satisfy $F(a_{j+1}/z) < \frac{1}{1+a_j/z}$, $\forall z \geq 1$. One can show that α is a continuous increasing concave function with $\alpha(0) = 0$ and because of that, $\bar{a}(a_j) = \sup_{z \geq 1} z\alpha(a_j/z) = \alpha(a_j)$. And we conclude by setting $\bar{b}(b_j) = \sqrt{\frac{\bar{a}(Pb_j^2)}{P}}$.

□

Comments:
The function F is also known as the Ei or exponential integral. The knowledge of the function α gives the one of $\bar{b}$. So we are interested in the reciprocal of the function $\frac{1}{F(\cdot)} - 1$. From that we get that the function $\bar{b}^2(b_j)$ is continuous convex and increasing with $\bar{b}(0) = 0$, having a derivative of 1 at 0 and of 0 at infinity. One can compute it to evaluate the ranges where the power allocation is zero.
Conclusion: We have a **new water-filling** situation, in the sense that we verify the two properties exposed in the beginning of this section. However the optimal power allocation is not the same one as in the case of a deterministic fading matrix B.

References

[1] Telatar E. *Capacity of Multi-antenna Gaussian Channels*, European Trans. on Telecommunications, Vol. 10, No 6, 1999, pp. 585-595.