Random matrices and communication systems: WEEK 3

In this lecture and in the subsequent ones, we adopt the following notations: small letters refer to scalars (deterministic or random) and deterministic vectors, while capital letters refer to matrices (deterministic or random) and random vectors. In some cases, this rule will not be followed strictly, but what is actually meant will be clear from the context.

Summary of last lecture and single-letter characterization

- We saw first that when the fading coefficient h_0 is fixed over time and deterministic, the capacity of the channel is given by

$$C = \log(1 + P\,|h_0|^2)$$

This result can also be seen as the solution of the following single-letter characterization of the channel capacity (that remains valid in the more general context of multiple antenna systems):

$$C = \sup_{p_X \,:\, \mathbb{E}(|X|^2) \le P} I(X;Y)$$

- Then, we saw that when the fading coefficients H_k are random and i.i.d. over time, known at both the transmitter and the receiver, the ergodic capacity of the channel is given by

$$C_{\text{erg}} = \sup_{\substack{Q(\cdot) \ge 0 \\ \mathbb{E}_H(Q(H)) \le P}} \mathbb{E}_H(\log(1 + Q(H)\,|H|^2))$$

while when they are known at the receiver but not at the transmitter,

$$C_{\text{erg}} = \mathbb{E}_H(\log(1 + P\,|H|^2))$$

Again, in this second case (which we will mainly focus on in the following), this ergodic capacity expression may be found as the result of the more general single-letter characterization:

$$C_{\text{erg}} = \sup_{p_X \,:\, \mathbb{E}(|X|^2) \le P} I(X;Y,H)$$

- Finally, when the fading coefficient H is random and fixed over time, known at both the transmitter and the receiver, the capacity of the channel is a random variable given by

$$C = \log(1 + P\,|H|^2)$$

while when H is not known at the transmitter, the capacity is equal to zero and the outage probability is given by

$$P_{\text{out}}(R) = \mathbb{P}_H(\log(1 + P\,|H|^2) < R)$$

for a target rate $R > 0$. Again, in this second case, this expression may be viewed as the result of the more general single-letter characterization:

$$P_{\text{out}}(R) = \inf_{p_X \,:\, \mathbb{E}(|X|^2) \le P} \mathbb{P}_H(I(X;Y) < R)$$

(notice that in this case, $I(X;Y)$ is a random variable depending on the realization of H).

Preliminaries for this lecture

- An n-variate complex-valued random vector $X = (X_1, \ldots, X_n)$ is *(jointly) continuous* if it admits a joint pdf $p_X = p_{X_!,\ldots,X_n}$, i.e.

$$\mathbb{P}(X \in B) = \int_B dx\, p_X(x) \quad \forall B \subset \mathbb{C}^n \text{ Borel set}$$

Its *mean vector* is defined as $\mu = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n))$ and its *covariance matrix* is defined as $(Q_X)_{jk} = \mathbb{E}(X_j \, \overline{X_k})$ (when they exist).

- Let X be a complex *Gaussian* random vector with mean 0 and positive definite covariance matrix Q_X (notation: $X \sim \mathcal{N}_\mathbb{C}(0, Q_X)$). This random vector admits the following joint pdf:

$$p_X(x) = \frac{1}{\pi^n \det(Q_X)} \, \exp\left(-x^* \, (Q_X)^{-1} \, x\right), \quad x \in \mathbb{C}^n$$

Notice that $\mathbb{E}(X_j) = 0$, $\mathbb{E}(X_j \, \overline{X_k}) = (Q_X)_{jk}$ and also that $X_1, \ldots, X_n$ are independent if and only if Q_X is diagonal.

- The *differential entropy* of a continuous random vector X is defined as

$$h(X) = -\int_{\mathbb{C}^n} dx \, p_X(x) \, \log(p_X(x))$$

One can check that

$$h(X) \leq \sum_{j=1}^n h(X_j)$$

with equality if and only if the X_j are independent, and also that

$$\sup_{p_X \,:\, \mathbb{E}(XX^*) = Q_X} h(X) = \log\det(\pi e \, Q_X)$$

is achieved by taking $X \sim \mathcal{N}_\mathbb{C}(0, Q_X)$.

Multiple antenna systems

We now consider the multiple antenna system (only one time-slot is considered here):

$$Y = H \, X + Z$$

where X, Y, Z are n-variate vectors and H is an $n \times n$ matrix. More precisely,

- X is the input vector submitted to the average power constraint $\mathbb{E}(\|X\|^2) \leq P$;
  (notice that $\mathbb{E}(\|X\|^2) = \mathbb{E}(X^* X) = \mathbb{E}(\mathrm{Tr}(X X^*)) = \mathrm{Tr}(Q_X)$)
- $Z \sim \mathcal{N}_\mathbb{C}(0, I)$ is the noise vector, independent of X;
- Y is the output vector;
- H is the channel fading matrix (to be specified below).

1 $H = H_0$ is deterministic and fixed over time

In this case, the single-letter characterization of the capacity reads

$$C = \sup_{p_X \,:\, \mathrm{Tr}(Q_X) \leq P} I(X;Y) = \sup_{p_X \,:\, \mathrm{Tr}(Q_X) \leq P} h(Y) - h(Y|X)$$

Notice that $h(Y|X) = h(H_0 \, X + Z \,|\, X) = h(Z)$, so

$$C = \left(\sup_{p_X \,:\, \mathrm{Tr}(Q_X) \leq P} h(H_0 \, X + Z) \right) - h(Z)$$

The above expression is maximized when $H_0\,X+Z$ is Gaussian, which happens when X itself is Gaussian. In this case, $H_0\,X + Z \sim \mathcal{N}_{\mathbb{C}}(0, I + H_0\,Q_X\,H_0^*)$, so

$$C = \left(\sup_{Q_X\,:\,\mathrm{Tr}(Q_X)\le P} \log\det(\pi e(I + H_0\,Q_X\,H_0^*)) \right) - \log\det(\pi e I) = \sup_{Q_X\,:\,\mathrm{Tr}(Q_X)\le P} \log\det(I + H_0\,Q_X\,H_0^*)$$

In order to proceed further, we need the following inequality, whose proof is left as an exercise in the homework.

Hadamard's inequality. Let A be a positive semi-definite $n \times n$ matrix. Then $\det(A) \le \prod_{j=1}^n a_{jj}$.

A second ingredient is the *singular value decomposition* of H_0, stating that there exist unitary matrices U, V and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, with $\sigma_j \ge 0$ for all j, such that $H_0 = U\,\Sigma\,V^*$. Therefore,

$$\log\det(I + H_0\,Q_X\,H_0^*) = \log\det(I + U\,\Sigma\,V^*\,Q_X\,V\,\Sigma^*\,U^*) = \log\det(I + \Sigma\,V^*\,Q_X\,V\,\Sigma^*)$$

Let now $\widetilde{Q}_X = V^*\,Q_X\,V$. Notice that $\widetilde{Q}_X$ also satisfies the above constraints:

$$\widetilde{Q}_X \ge 0 \quad \text{and} \quad \mathrm{Tr}(\widetilde{Q}_X) = \mathrm{Tr}(Q_X) \le P$$

so

$$C = \sup_{Q_X\,:\,\mathrm{Tr}(Q_X)\le P} \log\det(I + H_0\,Q_X\,H_0^*) = \sup_{\widetilde{Q}_X\,:\,\mathrm{Tr}(\widetilde{Q}_x)\le P} \log\det(I + \Sigma\,\widetilde{Q}_X\,\Sigma^*)$$

Using now Hadamard's inequality, we obtain

$$\det(I + \Sigma\,\widetilde{Q}_X\,\Sigma^*) \le \prod_{j=1}^n \left(1 + (\Sigma\,\widetilde{Q}_X\,\Sigma^*)_{jj}\right) = \prod_{j=1}^n \left(1 + (\widetilde{Q}_X)_{jj}\,\sigma_j^2\right)$$

and the equality is met by taking $\widetilde{Q}_X$ diagonal, say $\widetilde{Q}_X = \mathrm{diag}(d_1, \ldots, d_n)$. The above expression for the capacity can therefore be rewritten as

$$C = \sup_{\substack{d_1,\ldots,d_n \ge 0 \\ \sum_{j=1}^n d_j \le P}} \sum_{j=1}^n \log(1 + d_j\,\sigma_j^2)$$

The solution of this optimization problem is again obtained via water-filling:

$$C = \sum_{j=1}^n \left(\log(\nu\,\sigma_j^2)\right)^+ \quad \text{where} \quad \sum_{j=1}^n \left(\nu - \frac{1}{\sigma_j^2}\right)^+ \le P$$

As an example, let us consider the following simple case: $(H_0)_{jk} = 1$ for all j, k. In this case, $\sigma_1 = n$, $\sigma_2 = \ldots = \sigma_n = 0$, so

$$C = \log(\nu n^2) \quad \text{such that} \quad \left(\nu - \frac{1}{n^2}\right) \le P$$

i.e. $\nu = P + \frac{1}{n^2}$ and $C = \log(1 + P\,n^2)$.

2 H is random and varying ergodically over time (fast fading)

We assume now that the matrix H admits the pdf $p_H(\cdot)$ and that its realizations over time are i.i.d. (or ergodic), known at the receiver but not at the transmitter (so that X and H are independent). In this case, the single-letter characterization of the capacity reads:

$$C_{\mathrm{erg}} = \sup_{p_X\,:\,\mathrm{Tr}(Q_X)\le P} I(X; Y, H) = \sup_{p_X\,:\,\mathrm{Tr}(Q_X)\le P} I(X; H) + I(X; Y|H)$$

by the chain rule. Because of the independence of X and H, the first term is zero, so

$$C_{\text{erg}} = \sup_{p_X \,:\, \text{Tr}(Q_X) \le P} \int_{\mathbb{C}^{n^2}} dG \, p_H(G) \, I(X;Y|H=G)$$

Notice that for any fixed matrix G,

$$I(X;Y|H=G) = I(X;G\,X+Z) \le \log\det(I + G\,Q_X\,G^*)$$

and the equality is met when $X \sim \mathcal{N}_(0, Q_X)$ (which does not depend of the specific value of G). So

$$C_{\text{erg}} = \sup_{Q_X \ge 0 \,:\, \text{Tr}(Q_X) \le P} \int_{\mathbb{C}^{n^2}} dG \, p_H(G) \, \log\det(I + G\,Q_X\,G^*)$$

which can be rewritten as

$$C_{\text{erg}} = \sup_{Q_X \ge 0 \,:\, \text{Tr}(Q_X) \le P} \mathbb{E}_H(\log\det(I + H\,Q_X\,H^*))$$

Remark. It is *not* because the realizations of H are not known at the transmitter that the optimal Q_X should be a multiple of identity; it is indeed always possible to optimize over the *distribution* of H. Notice also that the solution is *not* the water-filling solution, as the singular values and vectors of H are not known at the transmitter.

Nevertheless, under symmetry conditions on the distribution of H, something more can be said on the optimal input covariance matrix Q_X. This is illustrated in the following lemmas, whose respective proofs are left as exercises in the homework.

Lemma 2.1. If h_{jk} are i.i.d. random variables, then the optimal input covariance matrix is of the form

$$Q_X = \frac{P}{n} \begin{pmatrix} 1 & c & \cdots & c & c \\ c & 1 & c & & c \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c & & c & 1 & c \\ c & c & \cdots & c & 1 \end{pmatrix}$$

where $-\frac{1}{n-1} \le c \le 1$ is some real parameter.

Lemma 2.2. If h_{jk} are independent random variables such that $h_{jk} \sim -h_{jk}$ for all j,k, then the optimal input covariance matrix Q_X is diagonal.

Lemma 2.3. If h_{jk} are i.i.d random variables such that $h_{jk} \sim -h_{jk}$ for all j,k, then the optimal input covariance matrix $Q_X = \frac{P}{n}\,I$.

Notice that the third lemma is simply a combination of the first two. It holds true in particular when h_{jk} are i.i.d.$\sim \mathcal{N}_{\mathbb{C}}(0,1)$ random variables, in which case

$$C_{\text{erg}} = \mathbb{E}_H\left(\log\det\left(I + \frac{P}{n}\,H H^*\right)\right)$$

We will analyze this expression further in the next lecture.