Random matrices and communication systems: WEEK 2

In this lecture, we adopt the following (temporary) notations: small letters refer to scalar numbers and capital letters refer to scalar random variables.

Single antenna systems

Let us consider the additive white Gaussian noise (AWGN) channel:

$$Y_k = H_k \, X_k + Z_k$$

where $k \in \{1, \ldots, N\}$ is the time index and

- $(X_1, \ldots, X_N)$ is the input vector submitted to the average power constraint $\mathbb{E}\left(\frac{1}{N}\sum_{k=1}^N |X_k|^2\right) \leq P$;
- $(Z_1, \ldots, Z_N)$ is the noise vector, whose components are i.i.d.$\sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ random variables, independent of $X_1, \ldots, X_N$;
- $(Y_1, \ldots, Y_N)$ is the output vector;
- $H_1, \ldots, H_N$ are the fading coefficients (to be specified below).

The signal-to-noise ratio (SNR) of the system is defined as $\mathrm{SNR} = P/\sigma^2$. In order to simplify notation, we will assume in the following that the noise variance $\sigma^2 = 1$, so that $\mathrm{SNR} = P$.

The general question we would like to address in the present lecture is the following. Assume that two users wish to communicate over the above channel; what is then the maximum rate R at which communication can be established reliably? The answer depends of course on the specific model chosen for the fading coefficients H_k. We will review in the following various possible assumptions.

1 $H_k = h_k$, $k = 1, \ldots, N$ are deterministic coefficients

We start by considering the case of deterministic (complex-valued) fading coefficients, as an "appetizer" to the random case, which is the case of interest for this course.

In the deterministic case, the maximum rate at which one may possibly communicate over the time interval $[1, \ldots, N]$ is given by

$$\max R \leq \sup_{\substack{p_{X_1,\ldots,X_N} \\ \mathbb{E}(\frac{1}{N}\sum_{k=1}^N |X_k|^2) \leq P}} \frac{1}{N} \, I(X_1, \ldots, X_N; Y_1, \ldots, Y_N)$$

Let us compute

$$\begin{aligned}
I(X_1, \ldots, X_N; Y_1, \ldots, Y_N) &= h(Y_1, \ldots, Y_N) - h(Y_1, \ldots, Y_N \,|\, X_1, \ldots, X_N) \\
&= h(h_1 X_1 + Z_1, \ldots, h_N X_N + Z_N) - h(h_1 X_1 + Z_1, \ldots, h_N X_N + Z_N \,|\, X_1, \ldots, X_N) \\
&= h(h_1 X_1 + Z_1, \ldots, h_N X_N + Z_N) - h(Z_1, \ldots, Z_N \,|\, X_1, \ldots, X_N) \\
&= h(h_1 X_1 + Z_1, \ldots, h_N X_N + Z_N) - h(Z_1, \ldots, Z_N)
\end{aligned}$$

as $X_1, \ldots, X_N$ and $Z_1, \ldots, Z_N$ are independent by assumption. Using now the fact that for jointly continuous random variables $U_1, \ldots, U_N$,

$$h(U_1, \ldots, U_N) \leq \sum_{k=1}^N h(U_k)$$

with equality if and only if the U_k are independent, we obtain

$$I(X_1, \dots, X_N; Y_1, \dots, Y_N) \le \sum_{k=1}^{N} (h(h_k\, X_k + Z_k) - h(Z_k))$$

with equality if and only if the X_k are independent. Using then the fact

$$\sup_{p_U\,:\,\mathbb{E}(|U|^2)\le P} h(U) = \log(\pi e P)$$

where the supremum is attained for $U \sim \mathcal{N}_{\mathbb{C}}(0, P)$, we further obtain

$$I(X_1, \dots, X_N; Y_1, \dots, Y_N) \le \sum_{k=1}^{N} (\log(\pi e(P_k\, |h_k|^2 + 1)) - \log(\pi e)) = \sum_{k=1}^{N} \log(1 + P_k\, |h_k|^2)$$

by taking $X_k \sim \mathcal{N}_{\mathbb{C}}(0, P_k)$ independent with $\frac{1}{N}\sum_{k=1}^{N} P_k \le P$ in order to meet the power constraint. This finally implies

$$\max R \le \sup_{\substack{P_1,\dots,P_N \ge 0 \\ \frac{1}{N}\sum_{k=1}^{N} P_k \le P}} \frac{1}{N} \sum_{k=1}^{N} \log(1 + P_k\, |h_k|^2)$$

This optimization problem can be solved analytically; its solution is the well known "water-filling" solution, but let us not write this down explicitly at this stage. Also, without making any further assumption on the (arbitrary) sequence of fading coefficients h_k, we cannot conclude anything on the capacity of the channel in the large N limit.

1.1 $H_k \equiv h_0$ for all $k = 1, \dots, N$

In this particular case, the above optimization problem is symmetric in $P_1, \dots, P_N$ and has therefore the following simple solution:

$$\max R \le \sup_{\substack{P_1,\dots,P_N \ge 0 \\ \frac{1}{N}\sum_{k=1}^{N} P_k \le P}} \frac{1}{N} \sum_{k=1}^{N} \log(1 + P_k\, |h_0|^2) = \log(1 + P\, |h_0|^2)$$

Here, as h_0 is fixed, the above expression can also be shown to be equal to the capacity of the channel in the large N limit.

2 H_k, $k = 1, \dots, N$ are random coefficients

In this section, we consider the H_k as random, in order to take into account the uncertainty about the fading coefficients. We should specify:

- how fast do these coefficients vary over time?

- among the receiver and the transmitter, who knows the realizations of these coefficients?

In the following, we assume that these coefficients have a given distribution and that this distribution is known to everyone. This is needed in order to be able to describe the statistics of the channel between X and Y. If even the distribution itself is not known, then the channel becomes an arbitrarily varying channel, which is out of the scope of the present course.

2.1 H_k are i.i.d. random variables (fast fading assumption)

This is in some sense an extreme assumption, which could be relaxed to "the coefficients H_k vary ergodically over time". By "ergodically", we actually mean that the empirical distribution of $H_1, \ldots, H_N$ converges to a given fixed distribution p_H (this holds in particular for an i.i.d. sequence, by the law of large numbers). But present in this assumption is also the fact that the coefficients H_k should vary relatively fast with respect to the duration of communication.

We need now to specify who knows the realizations of the coefficients H_k. In the following, we assume that that receiver is able to track (perfectly) the values of the H_k (by using pilot signals first, e.g.), but we make two different assumptions regarding the transmitter.

2.1.1 The transmitter knows the realizations of the coefficients H_k

This assumption is justified when feedback is easy to obtain at the transmitter. In this case, as everyone knows the channel realizations, it is as if these were actually deterministic, so the maximum rate achievable over the time interval $[1, \ldots, N]$ is bounded above by

$$\max R \leq \sup_{\substack{P_1,\ldots,P_N \geq 0 \\ \frac{1}{N}\sum_{k=1}^N P_k \leq P}} \frac{1}{N} \sum_{k=1}^N \log(1 + P_k\,|H_k|^2)$$

This leads to the definition of *ergodic capacity*:

$$C_{\mathrm{erg}} = \lim_{N\to\infty} \sup_{\substack{P_1,\ldots,P_N \geq 0 \\ \frac{1}{N}\sum_{k=1}^N P_k \leq P}} \frac{1}{N} \sum_{k=1}^N \log(1 + P_k\,|H_k|^2) = \sup_{\substack{Q(\cdot)\geq 0 \\ \int_{\mathbb{C}} dh\, p_H(h)\, Q(h) \leq P}} \int_{\mathbb{C}} dh\, p_H(h)\, \log(1 + Q(h)\,|h|^2)$$

This can in turn be rewritten as

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

The solution of this optimization problem is given by the water-filling solution:

$$C_{\mathrm{erg}} = \mathbb{E}_H\left(\left(\log(\nu\,|H|^2)\right)^+\right) \quad \text{where } \nu \text{ satisfies} \quad \mathbb{E}_H\left(\left(\nu - \tfrac{1}{|H|^2}\right)^+\right) \leq P \tag{1}$$

and $a^+ = \max(a, 0)$ denotes the positive part of $a \in \mathbb{R}$.

2.1.2 The transmitter does not know the realizations of the coefficients H_k

This assumption is justified when feedback is difficult, or even impossible, to obtain at the transmitter. In this case, the input vector $(X_1, \ldots, X_N)$ cannot be tuned according to the channel realizations $H_1, \ldots, H_N$, which we model mathematically by saying that $X_1, \ldots, X_N$ and $H_!, \ldots, H_N$ are independent.

As we assume on the other hand that the receiver knows the H_k, the channel between the transmitter and the receiver can be seen in this case as the channel with input $(X_1, \ldots, X_N)$ and output $(Y_1, \ldots, Y_N,\, H_1, \ldots, H_N)$; it is as if a genie were revealing the channel coefficients H_k to the receiver. So the mutual information of this channel is given by

$$\begin{aligned} & I(X_1, \ldots, X_N; Y_1, \ldots, Y_N,\, H_1, \ldots, H_N) \\ &\quad = I(X_1, \ldots, X_N; H_1, \ldots, H_N) + I(X_1, \ldots, X_N; Y_1, \ldots, Y_N \,|\, H_1 \ldots, H_N) \\ &\quad = 0 + h(Y_1, \ldots, Y_N \,|\, H_1, \ldots, H_N) - h(Y_1, \ldots, Y_N \,|\, X_1, \ldots, X_N,\, H_1, \ldots, H_N) \end{aligned}$$

where the chain rule was used in the first inequality and the independence of the H_k and X_k in the second inquality. As $Y_k = H_k\, X_K + Z_K$ we further obtain

$$\begin{aligned} & I(X_1, \ldots, X_N; Y_1, \ldots, Y_N,\, H_1, \ldots, H_N) \\ &= h(Y_1, \ldots, Y_N \,|\, H_1, \ldots, H_N) - h(Z_1, \ldots, Z_N) \\ &= \int_{\mathbb{C}^N} dh_1 \cdots dh_N\, p_{H_1,\ldots,H_N}(h_1, \ldots, h_N)\, h(h_1\, X_1 + Z_1, \ldots, h_N\, X_N + Z_N) - h(Z_1, \ldots, Z_N) \end{aligned}$$

where we have used the fact that the H_k, X_k, Z_k are independent. This can in turn be bouded above by

$$\begin{aligned} I(X_1, \ldots, X_N; Y_1, \ldots, Y_N,\, H_1, \ldots, H_N) &\leq \sum_{k=1}^N \int_{\mathbb{C}} dh_k\, p_{H_k}(h_k)\; (h(h_k\, X_k + Z_k) - h(Z_k)) \\ &\leq \sum_{k=1}^N \int_{\mathbb{C}} dh_k\, p_{H_k}(h_k)\; \log(1 + P_k\, |h_k|^2) \end{aligned}$$

by chosing $X_k \sim \mathcal{N}_{\mathbb{C}}(0, P_k)$ independent with $\frac{1}{N} \sum_{k=1}^N P_k \leq P$ (and notice that this choice of X_k maximzing the mutal information does *not* depend on the particular realizations of the fading coefficients H_k). The ergodic capacity of the channel is therefore given in this case by

$$\begin{aligned} C_{\text{erg}} &= \lim_{n\to\infty} \sup_{\substack{p_{X_1,\ldots,X_N} \\ \mathbb{E}\left(\frac{1}{N}\sum_{k=1}^N |X_k|^2\right) \leq P}} \frac{1}{N}\, I(X_1, \ldots, X_N; Y_1, \ldots, Y_N,\, H_1, \ldots, H_N) \\ &= \lim_{N\to\infty} \sup_{\substack{P_1,\ldots,P_N \geq 0 \\ \frac{1}{N}\sum_{k=1}^N P_k \leq P}} \frac{1}{N} \sum_{k=1}^N \int_{\mathbb{C}} dh_k\, p_{H_k}(h_k)\; \log(1 + P_k\, |h_k|^2) = \int_{\mathbb{C}} dh\, p_H(h)\; \log(1 + P\, |h|^2) \end{aligned}$$

because again of the symmetry of the optimization problem. This can in turn be rewritten as

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

Remarks. - Comparing this expression with (1), we see the impact on the capacity of not knowing the channel coefficients at the transmitter.

- By Jensen's inequality, we obtain

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

so the ergodic capacity is less than or equal to the capacity of the channel with fixed and deterministic fading coefficient h_0 satisfying $|h_0|^2 = \mathbb{E}_H(|H|^2)$, which leads to the conclusion that fading degrades capacity in single antenna channels. We will see that the situation differs in multiple antenna channels.

2.2 $H_k \equiv H$ for all $k = 1, \ldots, N$ (slow fading assumption)

This is again an extreme assumption, which could be relaxed to "the variations of the coefficients H_k are sufficiently small over the duration of communication". We again assume that the receiver is able to track the fading coefficients and make two different assumptions on the transmitter.

2.2.1 The transmitter knows the realization of H

In this case, as H is fixed over time and known to everyone, it is as if we were in the fixed and determinisitc scenariio (see paragraph 1.1), so the capacity of the channel is given by

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

which is a random variable here, depending on the given realization of H.

2.2.2 The transmitter does not know the realization of H

In this case, let us moreover assume that the distribution of H admits a continuous pdf p_H whose support contains the point 0 (this is in particular verified for Rayleigh fading, that is, when $H \sim \mathcal{N}_{\mathbb{C}}(0,1)$). This is implying that whatever $\varepsilon > 0$, $\mathbb{P}(|H|^2 < \varepsilon) > 0$. As a consequence, whatever the rate chosen by the transmitter for communication (who does not know the value of H), there is always a non-zero probability that the chosen rate is above the actual capacity of the channel. Therefore, the capacity in this case is strictly speaking equal to zero. We therefore shift our attention to another performance measure: the *outage probability*, defined as, for a given target rate $R > 0$,

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

The outage probability is a lower bound on the error probability achieved by any scheme on this channel (exactly like the capacity is an upper bound on the rate achieved by any scheme on a given channel). As long as the assumption on p_H made at the beginning of this paragraph is verified, the outage probability is always strictly positive.

Considering the high SNR regime (i.e. $P \to \infty$), this probability can still be made vanishingly small. First, observe that as $P \to \infty$,

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

So if one wants $P_{\text{out}}(R)$ to decrease to zero as $P \to \infty$, one should not choose the target rate R higher than $\log P$. Let us therefore choose $R = r \log P$, with $0 \leq r \leq 1$. In the case where $H \sim \mathcal{N}_{\mathbb{C}}(0,1)$, we obtain

$$\begin{aligned} P_{\text{out}}(r \log P) &= \mathbb{P}_H(\log(1 + P\,|H|^2) < r \log P) = \mathbb{P}_H\left(1 + P\,|H|^2 < P^r\right) \\ &= \mathbb{P}_H\left(|H|^2 < \frac{P^r - 1}{P}\right) \simeq \mathbb{P}_H\left(|H|^2 < P^{r-1}\right) \simeq P^{r-1} \end{aligned}$$

So the decay is polynomial in P. In addition, we observe the following tradeoff: the lower the target rate $r \log P$, the higher the speed of decrease to zero of the outage probability.