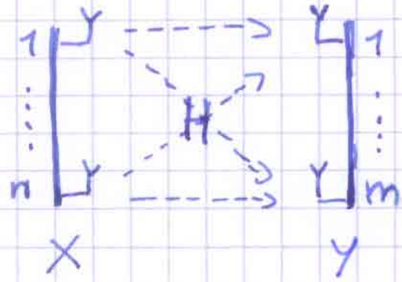
Random matrix theory: lecture 6

Capacity of multi-antenna channels [ref: Telatar 1995]

(also known as MIMO [multiple input - multiple output] channels)

Model:



n transmit antennas

m receive antennas

$$Y = HX + Z, \quad \text{i.e. } y_j = \sum_{k=1}^{n} h_{jk} x_k + z_j, \quad j = 1..m$$

- $X = (x_1 .. x_n)$ input: x_k = signal sent by antenna #k

  ↳ subject to the global power constraint $\sum_{k=1}^{m} \mathbb{E}(|x_k|^2) \leq P$

- $Y = (y_1 .. y_m)$ output: y_j = signal received by antenna #j
- $Z = (z_1 .. z_m)$ background noise: $z_1 .. z_m$ = iid r.v. $\sim N_{\mathbb{C}}(0,1)$

  notation: $Z \sim N_{\mathbb{C}}(0, I)$

- Z has i.i.d. realizations over time
- X and Z are assumed to be independent
- $H = (h_{jk})_{j,k=1}^{m,n}$ $m \times n$ channel matrix;

  h_{jk} = attenuation factor between transmit antenna #k and receive antenna #j

× We want to compute the capacity of the channel $X \longrightarrow Y$ in three different scenarios.

<u>First scenario</u> : <u>H is deterministic</u>

<u>Preliminary</u> :

- Let X be a complex random vector with density p_X.

  <u>differential entropy</u> : $h(X) = -\int_{\mathbb{C}^n} p_X(x) \log p_X(x)\, dx$

× for a given covariance matrix $Q_X = \mathbb{E}(XX^*)$,

  $h(X)$ is maximized whenever X is Gaussian

  (notation: $X \sim N_{\mathbb{C}}(0, Q_X)$), in which case we have

  $h(X) = \log \det(\pi e Q_X)$

- Let X, Y be two complex random vectors

× <u>mutual information</u> : $I(X;Y) = h(Y) - h(Y|X)$

  in the case where $Y = HX + Z$ with H deterministic,

× we therefore have $I(X;Y) = h(Y) - h(Z)$

× The <u>capacity</u> of the channel $X \to Y$ is given by

$$\boxed{C = \max_{p_X : \sum_{k=1}^{n} \mathbb{E}(|x_k|^2) \leq P} I(X, Y)}$$

<u>NB</u> : $\sum_{k=1}^{n} \mathbb{E}(|x_k|^2) = \mathbb{E}(\|X\|^2) = \mathbb{E}(X^* X)$

$= \mathbb{E}(\mathrm{Tr}(XX^*)) = \mathrm{Tr}(\mathbb{E}(XX^*)) = \mathrm{Tr}\, Q_X$

Therefore, $\boxed{C = \max_{p_X : \mathrm{Tr}\, Q_X \le P} h(Y) - h(Z)}$

Now, $h(Z) = \log\det(\pi e Q_Z) = \log\det(\pi e I)$ (indep. of p_X)

× and for a given Q_X, $Q_Y = H Q_X H^* + I$ is fixed,

so $h(Y)$ is maximum when Y is Gaussian,

and $Y = HX + Z$ is Gaussian if X is Gaussian.

In this case, we have $h(Y) = \log\det(\pi e (H Q_X H^* + I))$

× so $\boxed{C = \max_{Q_X \ge 0 : \mathrm{Tr}\, Q_X \le P} \log\det(I + H Q_X H^*)}$

(notation for "Q_X is non-negative definite" [covariance matrix])

× Recall that $\det(I + H Q_X H^*) = \det(I + H^* H Q_X)$

and that there exist U $n \times n$ unitary and $\Lambda = \mathrm{diag}(\lambda_1 .. \lambda_n)$

such that $H^* H = U \Lambda U^*$. Therefore:

× $$\det(I + H Q_X H^*) = \det(I + \Lambda^{1/2} U^* Q_X U \Lambda^{1/2})$$

Observe that $\widetilde{Q}_X := U^* Q_X U$ is non-negative definite

if and only if Q_X is, and that $\mathrm{Tr}\, \widetilde{Q}_X = \mathrm{Tr}\, Q_X$

Therefore, $\boxed{C = \max_{\widetilde{Q}_X \ge 0 ; \mathrm{Tr}\, \widetilde{Q}_X \le P} \log\det(I + \Lambda^{1/2} \widetilde{Q}_X \Lambda^{1/2})}$

(NB: $\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ [recall that $\lambda_j \ge 0$])

$\Bigg\{$ Hadamard's inequality: (one version)
If A is non-negative definite, then $\det A \leq \prod_{j=1}^{n} a_{jj}$.

(NB: recall that $A \geq 0$ implies $A = A^*$ in the complex case)

It is the case that $I + \Lambda^{1/2} \widetilde{Q}_x \Lambda^{1/2} \geq 0$, so

$$\det\left(I + \Lambda^{1/2} \widetilde{Q}_x \Lambda^{1/2}\right) \leq \prod_{k=1}^{n} \left(1 + (\widetilde{Q}_x)_{kk} \lambda_k\right)$$

with equality whenever $\widetilde{Q}_x$ is diagonal $(= D)$.

× So $$C = \max_{D \text{ diag} \geq 0 : \operatorname{Tr} D \leq P} \log\det(I + \Lambda D)$$

i.e. $$\boxed{C = \max_{d_k \geq 0 : \sum_{k=1}^{n} d_k \leq P} \sum_{k=1}^{n} \log(1 + d_k \lambda_k)}$$

where $\lambda_1 \ldots \lambda_n$ are the eigenvalues of $H^* H$.

The solution to the above maximization problem is the so-called "water-filling" solution, which can be found through Kuhn-Tucker conditions:

$$d_k = \left(\nu - \frac{1}{\lambda_k}\right)^+ \quad \text{for some parameter } \nu$$

i.e. $$\boxed{\begin{array}{l} C = \sum_{k=1}^{n} \left(\log(\nu \lambda_k)\right)^+ \\ \text{where } \sum_{k=1}^{n} \left(\nu - \frac{1}{\lambda_k}\right)^+ \leq P \end{array}}$$

PS: an alternate derivation of this result may be obtained via the singular value decomposition of H

Second scenario: H is a random matrix that varies in an ergodic manner over time ["fast fading"]

At each time instant, we assume that H has × the same stationary distribution $p(H)$, but at this point, we make no particular assumption on what $p(H)$ is.

Remark 1: if $p(H)$ is not known, then the channel transition probability $p(Y|X)$ is not known, so there is no notion of capacity in this case.

Let us therefore assume that $p(H)$ is known (to everybody).

A further question is: Who knows the realizations of the matrix H over time? This is an important question, since the communication strategy described in the preceding scenario (H deterministic) requires both the transmitter and the receiver to know H.

a) nobody knows the realizations of H a priori:

This seems a plausible assumption at first sight; the resulting analysis for computing the capacity in this case is quite difficult; we will skip that.

b) both the transmitter and the receiver know the realizations of H: apart from the "genie aided" interpretation, this assumption is reasonable when the matrix H varies slowly over time (but still fast enough so that it remains an ergodic process...), so that the receiver is able to estimate H through a sequence of pilot symbols sent by the transmitter, and then feed H back to the transmitter. In this case, it is as if the channel matrix H were deterministic, so the channel capacity is given by

$$\boxed{C = \mathbb{E}_H\left(\max_{Q_x \geq 0 \,:\, \mathrm{Tr}\, Q_x \leq P} \log\det\left(I + H Q_x H^*\right)\right)}$$

× where the expectation $\mathbb{E}_H$ comes from the fact that H varies ergodically over time.

(NB: not much more can be said in this case)
[the random water-filling solution is not explicit]

c) another interesting situation is when H varies slowly, so that the receiver can estimate H accurately, but the feedback link is weak, so that the transmitter does not get the estimate.

In this case, the channel becomes (ideally)

$$X \longrightarrow (Y, H)$$

[it is as if the receiver also "receives" the matrix H]

By the chain rule,

$$I(X; Y, H) = \underbrace{I(X; H)}_{=0 \text{ since we assume } X \,\&\, H \text{ indep.}} + I(X; Y \mid H)$$

and

$$I(X; Y \mid H) = \int dG\, \overset{\text{dist. of } H}{p(G)}\, I(X; Y \mid H = G)$$

For a given input covariance matrix Q_x and a given G, we have $I(X; Y \mid H = G) \leq \log\det(I + G Q_x G^*)$,

so

$$I(X; Y \mid H) \leq \int dG\, p(G) \log\det(I + G Q_x G^*)$$

$$= \mathbb{E}_H\left(\log\det(I + H Q_x H^*)\right)$$

And the ("ergodic") capacity is given by

$$C = \max_{p_X : \mathrm{Tr}\, Q_X \leq P} I(X; Y, H)$$

so
$$\boxed{C = \max_{Q_X \geq 0 : \mathrm{Tr}\, Q_X \leq P} \mathbb{E}_H \left(\log \det (I + H Q_X H^*) \right)}$$

Remark 2: at this point, it is not appropriate to say that "since the transmitter does not know H, the best possible input covariance
× matrix Q_X is proportional to the identity matrix" (i.i.d. signals)
Depending on the distribution of H, the true maximizing Q_X might be quite different.

On the other hand, it is clear that the solution to the above maximization problem is not the water-filling solution, because of the fact that the expectation is inside now.

(this is the penalty we have to pay for the transmitter not knowing H, actually)