Mathematical methods in communication June 16th, 2011

Lecture 12

Lecturer: Haim Permuter *Scribe: Eynan Maydan and Asaf Aharon*

I. MIMO - Multiple Input Multiple Output

MIMO is the use of multiple antennas at both the transmitter and reviver to improve communication performance. It is one of serveral forms of smart antenna technology. MIMO technology has attracted attention in wireless communication, because it offers significant increases in data throughput and link range without additional bandwidth or transmit power.

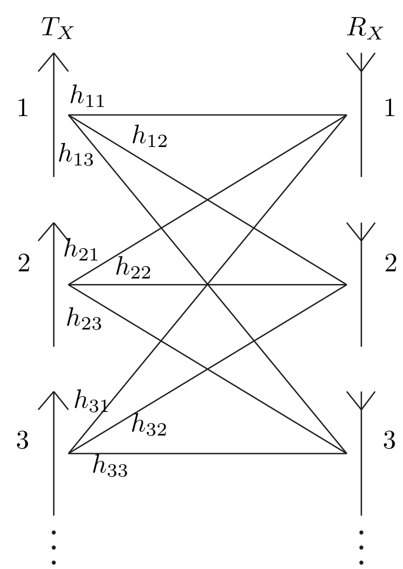

Fig. 1. MIMO illustration

We have seen in last lectures that for a Gaussian channel with power constraint, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P$, and Gaussian noise ,$Z \sim \mathcal{N}(0,\sigma^2)$, the capacity is defined by

$$C = \max_{P(X),E(X^2)\leq P} I(X,Y) = \frac{1}{2}\log(1+\frac{P}{\sigma^2}).$$

Let's check the case of parallel several channels for independent noise vectors $Z_1, Z_2, .., Z_m$.
In the parallel case, the power constraint is given by $P_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}^2$, where, $\sum_{i=1}^{m} P_i \leq P$ with probability one, and the capacity is defined by

$$C = \max_{f(X_1,X_2,...,X_K),E(\sum_{i=1}^{k} X_i^2)\leq P} I(X_1,X_2,...,X_K;Y_1,Y_2,...,Y_K)$$

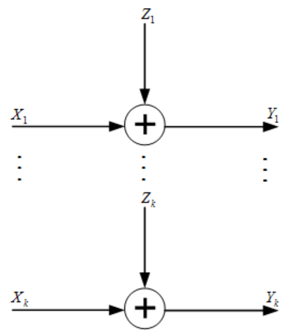

Fig. 2. Parallel Gaussian channels.

After some calculations (described in lecture 8), we get that

$$C = \sum_{i=1}^{m} \frac{1}{2} \log(1 + \frac{P_i}{\sigma_i^2}).$$

According to 'Water-filling' algorithm, P_i can be found by $P_i = [\frac{1}{\nu} - \sigma_i^2]^+ = \max(0, \frac{1}{\nu} - \sigma_i^2)$, where $P = \sum P_i$. In MIMO systems, each output is affected by all channel inputs. The power constraint for this model is given by $\frac{1}{r} \sum_{i=1}^{t} \sum_{j=1}^{r} X_{i,j}^2 \leq P$, with probability 1. The channel model can be presented as following:

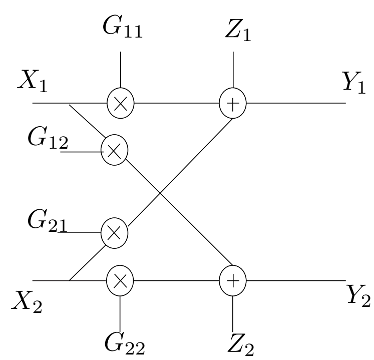

Fig. 3. MIMO

Note: In this section we suppose that G is deterministic, therefore we can assume that G is known at the decoder and encoder.

Mathematically, $Y_{r\times 1}$ is defined by $Y_{r\times 1} = G_{r\times t} X_{t\times 1} + Z_{r\times 1}$, where $X_{t\times 1}$ is the channel input, $G_{r\times t}$ is a constant channel gain matrix and $Z \sim \mathcal{N}(0, K_z)$ is the Gaussian noise. The element G_{jk} representing the gain of the channel from the transmitter antenna j to the receiver antenna k.

Lets assume, without loss of generality, that $Z \sim \mathcal{N}(0, 1)$ (We will see later that this assumption do not lead to any kind of losses). Indeed, the channel $Y = GX + Z$ with a general $K_z \succ 0$ can be transformed into the channel

$$\widetilde{Y} = {K_z}^{-\frac{1}{2}} Y = {K_z}^{-\frac{1}{2}} GX + \widetilde{Z}$$

Where, $\widetilde{Z} = {K_z}^{-\frac{1}{2}} Z \sim \mathcal{N}(0, \mathbf{1}_r)$. Since K_z is a covariance matrix, it is a symmetric matrix, i.e $K_z = K_z^T$. Therefore, we can write K_z as:

$$K_z = Q \Lambda Q^T$$

Where, Q is unitary matrix ($QQ^T = \mathbf{1}_r$) and Λ is the eigen values matrix which can be described as:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & & & \\ . & & . & & \\ . & & & . & \\ 0 & . & . & . & \lambda_n \end{bmatrix}$$

Note that since K_z is a symmetric matrix, the transpose matrix is equal to the inverted matrix, i.e. $Q^T = Q^{-1}$. Similarly, $K_z^{\frac{1}{2}} = Q \Lambda^{\frac{1}{2}} Q^T$, where, $\Lambda^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & . & . & 0 \\ 0 & \sqrt{\lambda_2} & & & \\ . & & . & & \\ . & & & . & \\ 0 & . & . & . & \sqrt{\lambda_n} \end{bmatrix}$

As we defined above:

$$\widetilde{Y} = {K_z}^{-\frac{1}{2}} Y = {K_z}^{-\frac{1}{2}} GX + {K_z}^{-\frac{1}{2}} Z$$

Now, let's calculate the covariance matrix of $\widetilde{Z} = K_Z^{-\frac{1}{2}} Z$

$$K_{\widetilde{z}} = E[\widetilde{Z}\widetilde{Z}^T] = E[K_z^{-\frac{1}{2}} Z Z^T K_z^{-\frac{1}{2}}] = k_z^{-\frac{1}{2}} K_z K_z^{-\frac{1}{2}} \stackrel{(a)}{=} k_z^{-\frac{1}{2}} K_z^{\frac{1}{2}} K_z^{\frac{1}{2}} K_z^{-\frac{1}{2}} = \mathbf{1}_{r \times r}$$

Where,

(a) follows from $K_z = Q \Lambda Q^T = Q \Lambda^{\frac{1}{2}} Q^T Q \Lambda^{\frac{1}{2}} Q^T = K_z^{\frac{1}{2}} K_z^{\frac{1}{2}}$.

The notation $\mathbf{1}_{r \times r}$ is the 'identity matrix' of dimension $r \times r$. Accordingly, we can conclude that the noise covariance can be assumed, without loss of generality, to be the unit matrix.

Now, let's find the capacity of the model:

$$\begin{aligned} C &= \max_{\sum P_i \leq P} I(X;Y) \\ &\stackrel{(a)}{=} \max_{\text{tr}(K_x) \leq P} I(X;Y) \\ &= \max_{\text{tr}(K_x \leq P} h(Y) - h(Y|X) \end{aligned}$$

$$
\begin{aligned}
&= \max_{\mathrm{tr}(K_x)\leq P} h(Y) - h(Z) \\
&\overset{(b)}{\leq} \frac{1}{2}\log[(2\Pi e)^r |K_Y|] - \frac{1}{2}\log[(2\Pi e)^r] \\
&= \frac{1}{2}\log|K_Y|
\end{aligned}
$$

Where,

(a) follows from $P_i = E[X_i^2]$, where, X is a vector of dimension $t \times 1$.

(b) follows from $h(Y) \leq \frac{1}{2}\log[(2\Pi e)^r|K_Y|]$, where, r is the dimension of Y, and from the fact that Z has Noraml distribution.

In order to find the capacity, we need to maximize the expression $\frac{1}{2}\log|K_Y|$,or actually, to maximize K_Y:

$$
\begin{aligned}
K_Y &= E[(GX+Z)(GX+Z)^T] \\
&= GK_XG^T + \mathbf{1}
\end{aligned}
$$

For that reason

$$
\max_{\mathrm{tr}(K_x))\leq P} \frac{1}{2}\log|K_Y| = \max_{\mathrm{tr}(K_x)\leq P} \frac{1}{2}\log|GK_XG^T + \mathbf{1}|
$$

Note: 'SVD' (Single Value Decomposition) - We know that every squared matrix can be decomposed as $A_{n\times n} = Q\Lambda Q^{-1}$. The SVD allows us to decompose unsquared matrix i.e. $G_{r\times t} = U_{r\times r}\Sigma_{r\times t}V^T_{t\times t}$ where U,V are unitary matrices ($UU^T = \mathbf{1}$, $VV^T = \mathbf{1}$), and Σ is a diagonal matrix. Therefore,

$$
GG^T = U\Sigma V^TV\Sigma^TU^T = U\Sigma\Sigma^TU^T
$$

$$
G^TG = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T
$$

Note that U and V can be calculated from the eigen vectors of GG^T and G^TG, respectively. Now, using the 'SVD', we can finally maximize K_Y :

$$
\begin{aligned}
\max_{\mathrm{tr}(K_x))\leq P} \frac{1}{2}\log|K_Y| &= \max_{\mathrm{tr}(K_x)\leq P} \frac{1}{2}\log|U\Sigma V^TK_XV\Sigma^TU^T + \mathbf{1}| \\
&\overset{(a)}{=} \max_{\mathrm{tr}(K_x)\leq P} \frac{1}{2}\log|U^TU\Sigma V^TK_XV\Sigma^TU^TU + U^TU| \\
&\overset{(b)}{=} \max_{\mathrm{tr}(\widetilde{K}_x)\leq P} \frac{1}{2}\log|\mathbf{1}\Sigma\widetilde{K}_X\Sigma^T\mathbf{1} + \mathbf{1}| \\
&= \max_{\mathrm{tr}(\widetilde{K}_X)\leq P} \frac{1}{2}\log|\Sigma\widetilde{K}_X\Sigma^T + \mathbf{1}|
\end{aligned}
$$

Where,

(a) follows from the property from linear algebra $|A| = |UAU^T| = |U||A||U^T|$ for square matrices.

(b) follows from the fact that U is unitary matrix, therefore, $U^TU = \mathbf{1}$. In addition, $\widetilde{K_x}$ is defined by

$\widetilde{K_x} = V^T K_x V$. Note that according to the 'trace' properties tr$(K_X) =$ tr$(VV^T K_X) =$ tr$(V^T K_X V) =$ tr$(\widetilde{K}_X)$

Remark1 - Assuming K is a nonnegative definite symmetric $n \times n$ matrix and $|K|$ denote the determinant of K, then it follows $|K| \leq \Pi_{i=1}^n |k_{ii}|$

Using the fact that X has a Noraml distribution ($X^n \sim \mathcal{N}(0, K)$) and the remark above, we can continue maximize $\frac{1}{2} \log |K_Y|$ as following:

$$\begin{aligned}
\max_{\text{tr}(K_X) \leq P} \frac{1}{2} \log |\Sigma \widetilde{K}_X \Sigma^T + 1| \quad &\overset{(a)}{\leq} \quad \max_{\text{tr}(\widetilde{K}_X) \leq P} \sum_{i=1}^{min(r,t)} \log(\gamma_i^2 \widetilde{P}_i + 1) \\
&\overset{(b)}{=} \quad \max_{\text{tr}(\widetilde{K}_X) \leq P} \sum_{i=1}^{\min(r,t)} \log(\gamma_i^2 (\widetilde{K}_X)_{ii} + 1) \\
&\overset{(c)}{=} \quad \max_{\sum \widetilde{P}_i \leq P} \sum_{i=1}^{\min(r,t)} \log(\gamma_i^2 \widetilde{P}_i + 1)
\end{aligned}$$

Where:

(a) follows from the fact that X has Normal distribution $X^n \sim \mathcal{N}(0, K)$ and from **Remark1**. In addition $\gamma_i = \Sigma_{ii}$.

(b) follows from $\widetilde{P}_i = (\widetilde{K}_X)_{ii}$

(c) follows from $\widetilde{P}_i = (\widetilde{K}_X)_{ii}$ which follows that $\Sigma \widetilde{P}_i \leq P$

Note that knowing the γ_i's let us solve the problem using 'Water-filling' algorithm with the given constraints. **Example** - Let's assume

$$G_{4x2} = \begin{bmatrix} 2 & 1 \\ 3 & 0 \\ 5 & 8 \\ 3 & 2 \end{bmatrix}$$

Since$G = U \Sigma V^T$ the G matrix can be written by:

$$G_{4x2} = U \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} V^T$$

In this example, the 'SVD' decomposition gives us tow eigenvalues K_1 and K_2, which can be easily found by 'Matlab' ($K_1 = 10.3559, K_2 = 2.9590$). These tow eigenvalues let us find P_i using 'Water-filling' algorithm, and then the diagonal matrix $\widetilde{K}_X$. Then it's easy to extract K_X by $K_X = V \widetilde{K}_X V^T$.

A. Alternative proof of capacity

There is another way to calculate the capacity by transforming the MIMO channel into a parallel channel. The MIMO channel is given by

$$Y = U\Sigma V^T X + Z$$

Now, define

$$\widetilde{X} = V^T X$$

Note that $E[\widetilde{X}^T \widetilde{X}] = E[X^T X]$. In addition, let's define

$$\begin{aligned} \widetilde{Y} &= U^T Y \\ &\stackrel{(a)}{=} U^T U \Sigma V^T X + U^T Z \\ &\stackrel{(b)}{=} \Sigma V^T X + U^T Z \\ &= \Sigma \widetilde{X} + \widetilde{Z} \end{aligned}$$

Where,

(a) $Y = U\Sigma V^T X + Z$.

(b) Since U is unitary, $U^T U = \mathbf{1}$.

Remark - Since U and V are unitar matrices, multiplying in U^T or V^T does not add any power to channel.

Lemma 1 The MIMO channel in figure 4 has the same capacity as in figure 5

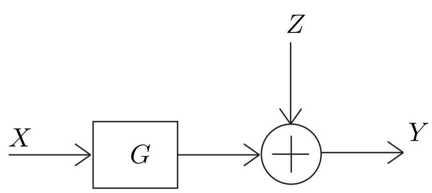

Fig. 4. MIMO Channel

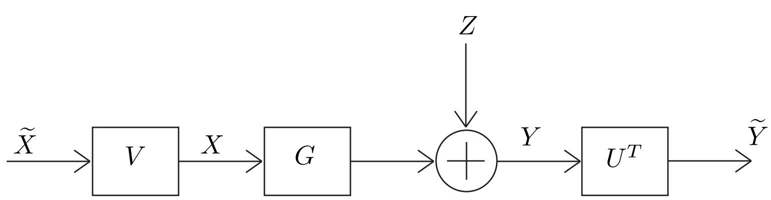

Fig. 5. Parallel Channel

Solution to the MIMO channel can be done by the following:

1) Solve the parallel channel in figure 5 using the water-filling.
2) The input to the MIMO is obtain by $X = V\widetilde{X}$.

Having this definitions let us analyze the channel from $\widetilde{X}$ to $\widetilde{Y}$ (parallel channel), where $\widetilde{Y} = U^T Y$, instead of analize the MIMO channel.

II. MIMO WITH FADING

Now, we assume that G is random. This assumption follows 3 cases:

1) G is not known to the transmitter and the receiver.
2) G is known only to the receiver.
3) G is known to transmitter and receiver.

Case I :

In this case, it's not necessarily that G will get the maximum capacity because it doesn't depend on G's distribution.

$$C = \max_{\mathrm{tr}(K_X)\le P} I(\overline{X};\overline{Y})$$

Case II :

$$\begin{aligned} C &= \max_{\mathrm{tr}(K_X))\le P} I(\overline{X};\overline{Y}|G) \\ &= \max_{\mathrm{tr}(K_X))\le P} (h(\overline{Y}|G) - h(\overline{Y}|G,\overline{X})) \\ &= \max_{\mathrm{tr}(K_X))\le P} \Sigma P(g) h(Y|G=g) - h(Z) \\ &= \max_{\mathrm{tr}(K_X))\le P} \Sigma P(g) h(Xg+Z) \end{aligned}$$

Case III :

$$C = \max_{P(x|g):tr(X))\le P} I(\overline{X};\overline{Y}|G)$$