Information Theory for Wireless Communication

Lecture 2: Single User Gaussian Vector Channel: Fast Fading Case

Lecture by Dr. Saif K. Mohammed

Scribe by T. V. K. Chaitanya

Consider a continuous input and continuous output memoryless vector channel as shown in Fig. 1. We define the ϵ-entropy typical set for the input as

$$T_\epsilon^n(\mathbf{X}) = \left\{ (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) \middle| \left| -\frac{\log_2 f(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)}{n} - h(\mathbf{X}) \right| < \epsilon \right\} \tag{1}$$

where

$$h(\mathbf{X}) \triangleq -\int_{\mathcal{C}^{N_t}} f(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) \log_2 f(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)\, d\mathbf{x}_1 \cdots d\mathbf{x}_n. \tag{2}$$

If $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are generated i.i.d. according to $f_{\mathbf{X}}(.)$, then

$$\Pr\{(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n) \in T_\epsilon^n(\mathbf{X})\} \geq 1 - \epsilon,$$

and

$$\text{Vol}(T_\epsilon^n(\mathbf{X})) \geq (1-\epsilon)\, 2^{n(h(\mathbf{X})-\epsilon)}$$

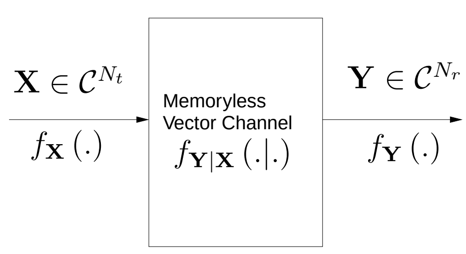

Figure 1. Memoryless vector channel.

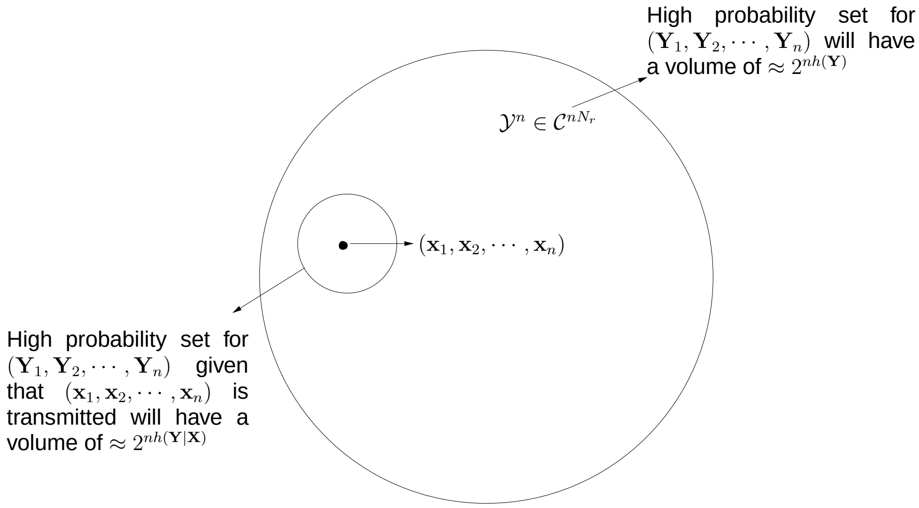

Figure 2. Capacity of a memoryless vector channel.

similarly, for the output $\mathbf{Y}$, we have

$$\mathrm{Vol}\left(T_\epsilon^n\left(\mathbf{Y}\right)\right) \geq \left(1-\epsilon\right) 2^{n(h(\mathbf{Y})-\epsilon)}$$

For a given sequence $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$, the output $(\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n)$ belongs to a high probability set of volume $\approx 2^{nh(\mathbf{Y}|\mathbf{X})}$. Which implies that the number of distinct $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ sequences that can be communicated reliably is given by

$$M = 2^{nR} \approx \frac{2^{nh(\mathbf{Y})}}{2^{nh(\mathbf{Y}|\mathbf{X})}} = 2^{nI(\mathbf{X};\mathbf{Y})}$$

and hence the rate of the code $R \approx I\left(\mathbf{X};\mathbf{Y}\right)$.

I. Deterministic Gaussian MIMO Channel

We model the MIMO communication channel as a deterministic Gaussian vector channel as shown in Fig. 3. The output vector $\mathbf{y}$ for a given input vector $\mathbf{x}$ can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}. \tag{3}$$

Here $\mathbf{y}$ is N_r-dimensional vector, and $\mathbf{x}$ is a N_t-dimensional vector. $\mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{K_z}\right)$, $\mathbf{K_z} \succeq 0$, is an N_r-dimensional noise vector. $\mathbf{H}$ is an $N_r \times N_t$ fixed channel gain matrix. The power

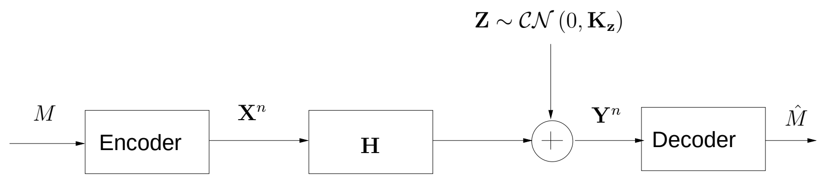

Figure 3. Deterministic Gaussian Vector Channel.

constraint on the input can be expressed as $\text{Trace}(\mathbf{K_x}) = \text{Trace}\left(\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right]\right) \leq P$. Without loss of generality, we can assume that $\mathbf{K_z} = \mathbf{I}_{N_r}$, because the channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$ with a general $\mathbf{K_z} \succeq 0$ can be transformed into the channel

$$\tilde{\mathbf{y}} = \mathbf{K_z}^{-1/2}\mathbf{y} = \mathbf{K_z}^{-1/2}\mathbf{H}\mathbf{x} + \tilde{\mathbf{z}},$$

where $\tilde{\mathbf{z}} = \mathbf{K_z}^{-1/2}\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$, and $\mathbf{K_z}^{-1/2}$ is the positive semi-definite square-root of $\mathbf{K_z}$. So assuming that $\mathbf{K_z} = \mathbf{I}_{N_r}$, we can write the mutual information between $\mathbf{x}$ and $\mathbf{y}$ as

$$\begin{aligned}
I(\mathbf{x}; \mathbf{y}) &= h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \\
&= h(\mathbf{y}) - h(\mathbf{y} - \mathbf{H}\mathbf{x}|\mathbf{x}) \\
&= h(\mathbf{y}) - h(\mathbf{z}) \\
&= h(\mathbf{y}) - \log_2\left((\pi e)^{N_r}\right) \\
&\overset{(a)}{\leq} \log_2\left((\pi e)^{N_r}|\mathbf{K_y}|\right) - \log_2\left((\pi e)^{N_r}\right) \\
&= \log_2|\mathbf{K_y}| \\
&= \log_2\left|\mathbf{I}_{N_r} + \mathbf{H}\mathbf{K_x}\mathbf{H}^H\right|.
\end{aligned}$$

where the inequality in (a) follows from the maximum differential entropy property. The capacity of the Gaussian vector channel can be expressed as

$$\begin{aligned}
C &= \max_{\mathbf{K_x} \succeq 0 : \text{Trace}(\mathbf{K_x}) \leq P} \log_2\left|\mathbf{I}_{N_r} + \mathbf{H}\mathbf{K_x}\mathbf{H}^H\right| \\
&\overset{(b)}{=} \max_{\mathbf{K_x} \succeq 0 : \text{Trace}(\mathbf{K_x}) \leq P} \log_2\left|\mathbf{I}_{N_r} + \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{K_x}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^H\right|
\end{aligned}$$

$$\overset{(c)}{=} \max_{\mathbf{K_x} \succeq 0: \text{Trace}(\mathbf{K_x}) \leq P} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{\Sigma} \underbrace{\mathbf{V}^H \mathbf{K_x} \mathbf{V}}_{\triangleq \tilde{\mathbf{K}}_\mathbf{x}} \mathbf{\Sigma} \right|$$

$$= \max_{\tilde{\mathbf{K}}_\mathbf{x} \succeq 0: \text{Trace}(\tilde{\mathbf{K}}_\mathbf{x}) \leq P} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{\Sigma} \tilde{\mathbf{K}}_\mathbf{x} \mathbf{\Sigma} \right|$$

In (b), we have used the singular value decomposition (SVD) of the matrix $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ and the relation in (c) follows from the identity that $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$. Using the Hadamard inequality (for a positive semi-definite matrix $\mathbf{A}$, $|\mathbf{A}| \leq \prod_{k=1}^{\text{rank}(\mathbf{A})} a_{k,k}$, and equality holds iff $\mathbf{A}$ is diagonal), we can write

$$C = \max_{\substack{d_i \geq 0, i=1,\cdots,N_t \\ \sum_i d_i \leq P}} \sum_{i=1}^{n_{min}} \log_2 \left(1 + \sigma_i^2 d_i\right) \tag{4}$$

with optimal $\tilde{\mathbf{K}}_\mathbf{x}^*$ being a diagonal matrix diag $(d_1, d_2, \cdots, d_{N_t})$ and $n_{min} = \min(N_t, N_r)$. Optimal $d_i^* = \left[\mu - \frac{1}{\sigma_i^2}\right]^+$, where μ is chosen such that $\sum_{i=1}^{N_t} d_i = P$. Thus the transmitter should align its signal direction with the singular vectors of the effective channel and allocate an appropriate amount of power in each direction to water-fill over the singular values. For each of the n_{min} parallel channels, independent Gaussian codebooks can be used at the transmitter with a power constraint of d_i^*. At high SNR, we have

$$d_i^* \approx \begin{cases} \frac{P}{n_{min}}, & 1 \leq i \leq n_{min} \\ 0 & \text{Otherwise} \end{cases}$$

and hence we can write

$$\begin{aligned} C_{h-SNR} &= \sum_{i=1}^{n_{min}} \log_2 \left(1 + \sigma_i^2 \frac{P}{n_{min}}\right) \\ &\approx n_{min} \log_2 P + \sum_{i=1}^{n_{min}} \log_2 \left(\frac{\sigma_i^2}{n_{min}}\right) \\ &= n_{min} \log_2 P + C' \end{aligned} \tag{5}$$

where C' is a constant term independent of P. As we can see from (5), at high SNR, by doubling the power, we increase the capacity by n_{min} bits. At low SNR, it is optimal to put all the power in the largest eigen mode and the capacity can be approximated as

$$C_{l-SNR} \approx \log_2\left(1+\sigma_{max}^2 P\right) \tag{6}$$

II. Fast Fading Gaussian MIMO Channel with Receiver CSI

Now considering a system model similar to the one in (3) but with a difference that the channel gain matrix $\mathbf{H}$ is varying with time. We assume that the receiver can track the channel matrix $\mathbf{H}$ perfectly but the transmitter does not have the knowledge about $\mathbf{H}$ to adapt its transmission. In this case, we can write

$$\begin{aligned} I\left(\mathbf{x};\mathbf{y},\mathbf{H}\right) &= h\left(\mathbf{y},\mathbf{H}\right) - h\left(\mathbf{y},\mathbf{H}|\mathbf{x}\right) \\ &= h\left(\mathbf{H}\right) + h\left(\mathbf{y}|\mathbf{H}\right) - h\left(\mathbf{H}|\mathbf{x}\right) - h\left(\mathbf{y}|\mathbf{H},\mathbf{x}\right) \\ &= h\left(\mathbf{y}|\mathbf{H}\right) - h\left(\mathbf{y}|\mathbf{H},\mathbf{x}\right) \\ &= I\left(\mathbf{x};\mathbf{y}|\mathbf{H}\right) \\ &= \mathbb{E}_{\mathbf{H}}\left[I\left(\mathbf{x};\mathbf{y}|\mathbf{H}=H\right)\right] \\ &\overset{(d)}{\leq} \mathbb{E}_{\mathbf{H}} \log_2\left|\mathbf{I}_{N_r} + \mathbf{H}\mathbf{K}_{\mathbf{x}}\mathbf{H}^H\right| \end{aligned} \tag{7}$$

where the result in (d) follows from the upper bound on $I\left(\mathbf{x};\mathbf{y}|\mathbf{H}=H\right)$. The capacity for this channel is defined as

$$C_{R-CSI} = \max_{\mathbf{K}_{\mathbf{x}}\succeq 0:\text{Trace}(\mathbf{K}_{\mathbf{x}})\leq P} \mathbb{E}_{\mathbf{H}} \log_2\left|\mathbf{I}_{N_r} + \mathbf{H}\mathbf{K}_{\mathbf{x}}\mathbf{H}^H\right| \tag{8}$$

Since the distribution of $\hat{\mathbf{H}} = \mathbf{H}\mathbf{U}$ is the same as that of $\mathbf{H}$ for any unitary matrix $\mathbf{U}$, by using the SVD of $\mathbf{K}_{\mathbf{x}}$, we can equivalently write (8) as

$$C_{R-CSI} = \max_{d_i\geq 0:\sum_{i=1}^{N_t} d_i \leq P} \mathbb{E}_{\mathbf{H}} \log_2\left|\mathbf{I}_{N_r} + \mathbf{H}\mathbf{D}\mathbf{H}^H\right| \tag{9}$$

Now considering any permutation matrix Π and defining $\mathbf{D}^{\Pi} \triangleq \Pi \mathbf{D} \Pi^{H}$, we can note that there are in total $N_t!$ permutations possible for a given $\mathbf{D}$ matrix. Let us define

$$\tilde{\mathbf{D}} = \frac{1}{N_t!} \sum_{\Pi} \Pi \mathbf{D} \Pi^{H} \tag{10}$$

we can note that $\tilde{\mathbf{D}}$ is a scaled identity matrix and that

$$\mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{H}\mathbf{D}\mathbf{H}^{H} \right| = \mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{H}\mathbf{D}^{\Pi}\mathbf{H}^{H} \right|. \tag{11}$$

If $\text{Trace}\left(\mathbf{D}\right) \leq P$, then we also have $\text{Trace}\left(\tilde{\mathbf{D}}\right) \leq P$. Now using the result that $\log|\mathbf{A}|$ is a concave function over the space of all positive semi-definite matrices (i.e., for two positive semi-definite matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ with $\mathbf{A} = \lambda \mathbf{A}_1 + (1-\lambda)\mathbf{A}_2$, we have $\log|\mathbf{A}| \geq \lambda \log|\mathbf{A}_1| + (1-\lambda)\log|\mathbf{A}_2|$ for any $\lambda \in [0,1]$), we can write

$$\begin{aligned} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{H}\tilde{\mathbf{D}}\mathbf{H}^{H} \right| &\geq \frac{1}{N_t!} \sum_{\Pi} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{H}\mathbf{D}^{\Pi}\mathbf{H}^{H} \right| \\ \mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{H}\tilde{\mathbf{D}}\mathbf{H}^{H} \right| &\geq \frac{1}{N_t!} \sum_{\Pi} \mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{H}\mathbf{D}^{\Pi}\mathbf{H}^{H} \right| \\ &= \mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \mathbf{H}\mathbf{D}\mathbf{H}^{H} \right| \end{aligned} \tag{12}$$

From (12), we can conclude that the optimal $\mathbf{D}^*$ in (9) must be scaled a identity matrix, i.e., $\mathbf{D}^* = \alpha \mathbf{I}_{N_t}$. So the optimization problem in (9) reduces to

$$C_{R-CSI} = \max_{\alpha \leq \frac{P}{N_t}} \mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \alpha \mathbf{H}\mathbf{H}^{H} \right| \tag{13}$$

and hence

$$\begin{aligned} C_{R-CSI} &= \mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \frac{P}{N_t} \mathbf{H}\mathbf{H}^{H} \right| \\ &= \mathbb{E}_{\mathbf{H}} \sum_{i=1}^{n_{min}} \log_2 \left| \mathbf{I}_{N_r} + \frac{P\sigma_i^2}{N_t} \right| \\ &= \sum_{i=1}^{n_{min}} \mathbb{E}_{\mathbf{H}} \log_2 \left| \mathbf{I}_{N_r} + \frac{P\sigma_i^2}{N_t} \right| \end{aligned} \tag{14}$$

Remark 1: For the SIMO case, we have $N_t = 1$ and $N_r > 1$, in this case

$$C_{R-CSI-SIMO} = E_{\mathbf{h}} \log_2 \left(1 + P \underbrace{\|\mathbf{h}\|^2}_{\text{Array Gain}} \right) \tag{15}$$

For MISO case, we have $N_t > 1$ and $N_r = 1$, and we have

$$C_{R-CSI-MISO} = E_{\mathbf{h}} \log_2 \left(1 + \frac{P \|\mathbf{h}\|^2}{N_t} \right) \tag{16}$$

as we can see from (15) and (16) that with SIMO system, one can achieve the array gain as compared to the MISO system ($\|\mathbf{h}\|^2 / N_t \underset{N_t \to \infty}{\longrightarrow} 1$).

Remark 2: For smaller values of x, one can write $\log_2 (1 + x) \approx x \log_e 2$. Using this, from (14), we can write:

$$C_{R-CSI-l-SNR} \approx \frac{P \log_e 2}{N_t} \mathbb{E}_{\mathbf{H}} \left(\sum_{i=1}^{n_{min}} \sigma_i^2 \right) = \left(\frac{\mathbb{E}_{\mathbf{H}} \left(\|\mathbf{H}\|_F^2 \right)}{N_t} \right) P \log_e 2 \tag{17}$$

As we can see from (17), at low SNR, only the amount of received power contributes to the capacity. For i.i.d. Rayleigh fading channel, $\mathbb{E}_{\mathbf{H}} \left(\|\mathbf{H}\|^2 \right) \propto N_t N_r$. So we have

$$C_{R-CSI-l-SNR} \approx N_r C' P \log_e 2 \tag{18}$$

where C' is a constant term independent of P. From (18), we can conclude that it is beneficial to deploy more receiving antennas to improve the capacity at low SNR values.

III. Block Fading Fading Gaussian MIMO Channel with Full CSI

Next we consider the scenario in which the channel is fading slowly (remains constant for certain duration and changes independently) and both the transmitter and receiver have the perfect knowledge about the channel gain matrix. In this case, we can write the received signal model as:

$$\mathbf{y}(t) = \mathbf{H}(t)\,\mathbf{x}(t) + \mathbf{z}(t)\,, t = 1, 2, \cdots, T.$$

For the tth time block, we can apply the principles of the deterministic channel case in Section I, and hence we can write

$$C_{Full-CSI} = \max_{\substack{\tilde{d}_i(t)\geq 0, i=1,\cdots,N_t \\ \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n_{min}} \tilde{d}_i(t)\leq P}} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n_{min}} \log_2\left(1+\sigma_i^2\left(t\right)\tilde{d}_i\left(t\right)\right) \tag{19}$$

The optimization problem in (19) corresponds to water-filling in both space and time. The optimal solution is given by

$$d_i^*\left(t\right) = \left[\mu - \frac{1}{\sigma_i^2\left(t\right)}\right]^+,$$

where μ is chosen such that the average power constraint is satisfied. However this approach might not be practical as the transmitter needs to know all $\mathbf{H}\left(t\right), t = 1, 2, \cdots, T$ a priori. However, in the limit when $T \to \infty$, the average power constraint reduces to the following

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n_{min}}\tilde{d}_i\left(t\right) \leq P \to \mathbb{E}_{\mathbf{H}}\left(\sum_{i=1}^{n_{min}}\tilde{d}_i\left(t\right)\right) \leq P = \mathbb{E}_{\mathbf{H}}\left(\sum_{i=1}^{n_{min}}\left[\mu - \frac{1}{\sigma_i^2\left(t\right)}\right]^+\right) \leq P \tag{20}$$

One can find the solution to (20) if the statistics of $\mathbf{H}$ are known at the transmitter.

Remark 3: Assuming that $N_t > N_r$ $\left(n_{min} = N_r\right)$, and that SNR is high, we can write

$$\begin{aligned}
C_{Full-CSI} &= \mathbb{E}_{\mathbf{H}}\sum_{i=1}^{n_{min}}\log_2\left(1+\sigma_i^2 d_i^*\right) \\
&\approx \mathbb{E}_{\mathbf{H}}\sum_{i=1}^{n_{min}}\log_2\left(1+\frac{\sigma_i^2 P}{n_{min}}\right) \\
&= \mathbb{E}_{\mathbf{H}}\sum_{i=1}^{n_{min}}\log_2\left(1+\frac{\sigma_i^2 P}{N_t}\underbrace{\left(\frac{N_t}{N_r}\right)}_{\triangleq(**)}\right),
\end{aligned}$$

where (**) is the array gain compared to the expression corresponding to the only R-CSI case in (14).