Lecture 2: Capacity of MIMO Channels

Jun Chen

1 Water-filling and Water-pouring:

Theorem 1.1. *Let the random vector* $\mathbf{X} \in \mathcal{R}^n$ *have zero mean and covariance* $K = \mathbb{E}\mathbf{X}\mathbf{X}^T$*. Then* $h(\mathbf{X}) \leq \frac{1}{2}\log(2\pi e)^n det(K)$*, with equality iff* $\mathbf{X} \sim \mathcal{N}(0, K)$*.*

Theorem 1.2 (Hadamard's inequality). *Let K be a non-negative definite symmetric $n \times n$ matrix. We have* $det(K) \leq \prod_{i=1}^{n} K_{ii}$*, with equality iff* $K_{ij} = 0$*,* $i \neq j$*.*

Rate Distortion Function: Water-filling in the eigenspace.

$\mathbf{X} = (X_1, \cdots, X_n)^T \sim \mathcal{N}(0, K_{\mathbf{X}})$. Let $K_{\mathbf{X}} = U\Lambda U^T$ be the eigenvalue decomposition of $K_{\mathbf{X}}$, where U is an $n \times n$ unitary matrix and $\Lambda = \text{diag}\{\lambda_1, \cdots, \lambda_n\}$.

$$
\begin{aligned}
\min_{\mathbb{E}\|\mathbf{X}-\hat{\mathbf{X}}\|^2 \leq D} I(\mathbf{X}; \hat{\mathbf{X}}) &= \min_{\mathbb{E}\|U^T\mathbf{X}-U^T\hat{\mathbf{X}}\|^2 \leq D} I(U^T\mathbf{X}; U^T\hat{\mathbf{X}}) \\
&= \min_{\mathbb{E}\|U^T\mathbf{X}-U^T\hat{\mathbf{X}}\|^2 \leq D} \left[h(U^T\mathbf{X}) - h(U^T\mathbf{X}|U^T\hat{\mathbf{X}})\right] \\
&= \min_{\mathbb{E}\|U^T\mathbf{X}-U^T\hat{\mathbf{X}}\|^2 \leq D} \left[\frac{1}{2}\sum_{i=1}^{n}\log(2\pi e\lambda_i) - h(U^T\mathbf{X}|U^T\hat{\mathbf{X}})\right] \\
&\geq \min_{\mathbb{E}\|U^T\mathbf{X}-U^T\hat{\mathbf{X}}\|^2 \leq D} \left[\frac{1}{2}\sum_{i=1}^{n}\log(2\pi e\lambda_i) - h(U^T\mathbf{X} - U^T\hat{\mathbf{X}})\right],
\end{aligned}
$$

where the inequality becomes equality if $U^T\mathbf{X} - U^T\hat{\mathbf{X}}$ is independent of $\hat{\mathbf{X}}$.

Let $\text{diag}\left(\mathbb{E}(U^T\mathbf{X} - U^T\hat{\mathbf{X}})(U^T\mathbf{X} - U^T\hat{\mathbf{X}})^T\right) = \text{diag}\{D_1, \cdots, D_n\}$. We have $h(U^T\mathbf{X} - U^T\hat{\mathbf{X}}) \leq \frac{1}{2}\sum_{i=1}^{n}(\log 2\pi e D_i)$, where the equality holds when the components of $U^T\mathbf{X} - U^T\hat{\mathbf{X}}$ are independent and jointly Gaussian with $\mathbb{E}(U^T\mathbf{X} - U^T\hat{\mathbf{X}})(U^T\mathbf{X} - U^T\hat{\mathbf{X}})^T = \text{diag}\{D_1, \cdots, D_n\}$. Therefore,

$$
\begin{aligned}
I(\mathbf{X}; \hat{\mathbf{X}}) &\geq \min_{\sum_{i=1}^{n} D_i \leq D} \frac{1}{2}\sum_{i=1}^{n}\log\frac{\lambda_i}{D_i} \\
&= \frac{1}{2}\sum_{i=1}^{n}\log\frac{\lambda_i}{\min(D^*, \lambda_i)}
\end{aligned}
$$

with D^* given by the equation

$$
\sum_{i=1}^{n}\min(D^*, \lambda_i) = D.
$$

Channel Capacity: Water-filling in the eigenspace.

$\mathbf{Y} = H\mathbf{X} + \mathbf{N}$, where $\mathbf{Y}, \mathbf{N} \in \mathcal{R}^n$, $\mathbf{X} \in \mathcal{R}^m$, $H \in \mathcal{R}^{n\times m}$, and $\mathbf{N} \sim \mathcal{N}(0, K_{\mathbf{N}})$. Let $K_{\mathbf{N}} = U\Sigma U^T$ be the eigenvalue decomposition of $K_{\mathbf{N}}$, where $\Sigma = \text{diag}\{\sigma_1^2, \cdots, \sigma_n^2\}$. Let $\mathbf{N}' = \Sigma^{-1/2}U^T\mathbf{N}$, $H' = \Sigma^{-1/2}\mathbf{U}^T H$, $\mathbf{Y}' = \Sigma^{-1/2}U^T\mathbf{Y}$, and $\mathbf{X}' = \mathbf{X}$. We define an equivalent channel model $\mathbf{Y}' = H'\mathbf{X}' + \mathbf{N}'$, where $\mathbf{N}' \sim \mathcal{N}(0, I_n)$. Let $H' = V\Lambda W^T$ be the singular value decomposition of H', where $\Lambda = \text{diag}\{\lambda_1, \cdots, \lambda_r, 0, \cdots, 0\} \in \mathcal{R}^{n\times m}$ ($r = \text{rank}H'$), and $V \in \mathcal{R}^{n\times n}$, $W \in \mathcal{R}^{m\times m}$ are unitary matrices. Let $\mathbf{X}'' = W^T\mathbf{X}'$, $\mathbf{Y}'' = V^T\mathbf{Y}'$, and $\mathbf{N}'' = V^T\mathbf{N}'$. Now we get another equivalent channel model $\mathbf{Y}'' = \Lambda\mathbf{X}'' + \mathbf{N}''$, where $\mathbf{N}'' \sim \mathcal{N}(0, I_n)$. We have

$$\max_{\text{trace}(\mathbb{E}\mathbf{X}'\mathbf{X}'^T)\le P} I(\mathbf{X}';\mathbf{Y}') = \max_{\text{trace}(\mathbb{E}\mathbf{X}''\mathbf{X}''^T)\le P} I(\mathbf{X}'';\mathbf{Y}'').$$

Let the components of $\mathbf{X}''$ be independent[1] and jointly Gaussian with $\mathbb{E}\mathbf{X}''\mathbf{X}''^T = \text{diag}\{\mu_1, \cdots, \mu_r, 0, \cdots, 0\}$. We have

$$\begin{aligned}
\max_{\text{trace}(\mathbb{E}\mathbf{X}''\mathbf{X}''^T)\le P} I(\mathbf{X}'';\mathbf{Y}'') &= \max_{\sum_{i=1}^r \mu_i \le P} \frac{1}{2}\sum_{i=1}^r \log\left(1 + \mu_i\lambda_i^2\right) \\
&= \sum_{i=1}^r \max\left(\frac{1}{2}\log\left(\nu\lambda_i^2\right), 0\right)
\end{aligned}$$

where the maximizer μ_i^* is

$$\mu_i^* = \max\left(\nu - \frac{1}{\lambda_i^2}\right)$$

and ν is given by the equation

$$\sum_{i=1}^r \max\left(\nu - \frac{1}{\lambda_i^2}, 0\right) = P.$$

2 Scalar Fading Channels: Ergodic Capacity

$\mathbf{y}(t) = \mathbf{h}(t)\mathbf{x}(t) + \mathbf{n}(t)$, where $\{\mathbf{h}(t)\}_t$ is a stationary and ergodic process with stationary distribution $p(h)$, $\{\mathbf{n}(t)\}_t$ is an i.i.d. Gaussian process with $\mathbf{n}(t) \sim \mathcal{N}(0, \sigma_N^2)$ for all t.

1. $\{\mathbf{h}(t)\}_t$ is known at the receiver: (coherent channel capacity)

$$\begin{aligned}
C &= \max_{\mathbb{E}\mathbf{x}^2\le P} I(\mathbf{x};\mathbf{y}|\mathbf{h}) \\
&= \max_{\mathbb{E}\mathbf{x}^2\le P} \mathbb{E}_{\mathbf{h}} I(\mathbf{x};\mathbf{y}|\mathbf{h} = h) \\
&= \frac{1}{2}\mathbb{E}_{\mathbf{h}} \log\left(1 + \frac{\mathbf{h}^2 P}{\sigma_N^2}\right).
\end{aligned}$$

Define the average SNR at the receiver as

$$\mathsf{SNR} = \frac{\mathbb{E}_{\mathbf{h}}\mathbf{h}^2 P}{\sigma_N^2}.$$

[1] For the independent parallel channels, the input should also be independent over the parallel channels in order to maximize the sum channel capacity.

By Jensen's inequality,

$$\frac{1}{2}\mathbb{E}_{\mathbf{h}}\log\left(1+\frac{\mathbf{h}^2P}{\sigma_N^2}\right)\leq\frac{1}{2}\log\left(1+\frac{\mathbb{E}_{\mathbf{h}}\mathbf{h}^2P}{\sigma_N^2}\right)=\frac{1}{2}\log\left(1+\mathsf{SNR}\right).$$

Bad news: the capacity of fading channel is smaller than that of the AWGN channel with the same receiver SNR. In the low SNR regime, we have

$$\frac{1}{2}\mathbb{E}_{\mathbf{h}}\log\left(1+\frac{\mathbf{h}^2P}{\sigma_N^2}\right)\approx\frac{1}{2}\mathbb{E}_{\mathbf{h}}\left(\frac{\mathbf{h}^2P}{\sigma_N^2}\right)\log_2 e=\frac{1}{2}\mathsf{SNR}\log_2 e\approx C_{\text{AWGN}}.$$

At high SNR,

$$\begin{aligned}\frac{1}{2}\mathbb{E}_{\mathbf{h}}\log\left(1+\frac{\mathbf{h}^2P}{\sigma_N^2}\right) &\approx \frac{1}{2}\mathbb{E}_{\mathbf{h}}\log\left(\frac{\mathbf{h}^2P}{\sigma_N^2}\right)\\ &= \frac{1}{2}\log\mathsf{SNR}+\frac{1}{2}\mathbb{E}_{\mathbf{h}}\log\left(\frac{\mathbf{h}^2}{E_{\mathbf{h}}\mathbf{h}^2}\right)\\ &\approx C_{\text{AWGN}}+\frac{1}{2}\mathbb{E}_{\mathbf{h}}\log\left(\frac{\mathbf{h}^2}{E_{\mathbf{h}}\mathbf{h}^2}\right).\end{aligned}$$

2. $\{\mathbf{h}(t)\}_t$ is known at the transmitter and the receiver: (dynamic power control and opportunistic communication)

$$\begin{aligned}C &= \max_{\mathbb{E}\mathbf{x}^2(\mathbf{h})\leq P} I(\mathbf{x};\mathbf{y}|\mathbf{h})\\ &= \max_{\mathbb{E}\mathbf{x}^2(\mathbf{h})\leq P}\mathbb{E}_{\mathbf{h}}\max_{\mathbf{x}(h)} I(\mathbf{x};\mathbf{y}|\mathbf{h}=h)\end{aligned}$$

Define $\mathbb{E}\mathbf{x}^2(h)=P(h)$. We have

$$\max_{\mathbb{E}\mathbf{x}^2(\mathbf{h})\leq P}\mathbb{E}_{\mathbf{h}}\max_{\mathbf{x}(h)} I(\mathbf{x};\mathbf{y}|\mathbf{h}=h) = \max_{\mathbb{E}_{\mathbf{h}}P(\mathbf{h})\leq P}\frac{1}{2}\mathbb{E}_{\mathbf{h}}\log\left(1+\frac{\mathbf{h}^2P(\mathbf{h})}{\sigma_N^2}\right).$$

The maximizer $P^*(h)$ is

$$P^*(h)=\max\left(\lambda-\frac{\sigma_N^2}{h^2},0\right)$$

with λ given by

$$\mathbb{E}_{\mathbf{h}}P^*(\mathbf{h})=P.$$

Therefore,

$$C=\frac{1}{2}\int_{|h|=\sigma_N\lambda^{-1/2}}^{\infty}p(h)\log\left(\frac{h^2\lambda}{\sigma_N^2}\right)\mathrm{d}h$$

In the high SNR regime, λ is very large and $P^*(h)\approx\lambda$. Therefore the capacity converges to that of the case with no state information at transmitter.

Suppose the channel gain $\mathbf{h}^2$ has a peak value $G_{\max}$. At low SNR, we have

$$\mathbb{P}(h^2\approx G_{\max})\left(\lambda-\frac{\sigma_N^2}{G_{\max}}\right)=P$$

and thus

$$\begin{aligned}
\frac{1}{2}\mathbb{E}_{\mathbf{h}} \log\left(1+\frac{\mathbf{h}^2 P^*(\mathbf{h})}{\sigma_N^2}\right) &\approx \frac{1}{2}\mathbb{P}(h^2 \approx G_{\max}) \log\left(1+\frac{G_{\max}P}{\sigma_N^2 \mathbb{P}(h^2 \approx G_{\max})}\right) \\
&\approx \frac{G_{\max}P}{2\sigma_N^2}\log_2 e \\
&= \frac{1}{2}\mathsf{SNR}\frac{G_{\max}}{\mathbb{E}_{\mathbf{h}}\mathbf{h}^2}\log_2 e \\
&\approx \frac{G_{\max}}{\mathbb{E}_{\mathbf{h}}\mathbf{h}^2} C_{\text{AWGN}}.
\end{aligned}$$

Good news: the capacity gain over the AWGN channel depends on the peakness of the fading process, which can be unbounded.

Since the capacity scales linearly with the power at low SNR ($\frac{1}{2}\log(1+\mathsf{SNR}) \approx \frac{1}{2}\mathsf{SNR}\log_2 e$) and logrithmically at high SNR ($\frac{1}{2}\log(1+\mathsf{SNR}) \approx \frac{1}{2}\log(\mathsf{SNR})$), power control is more effective in the low SNR regime. We will see that at high SNR, a more efficient method is to increase the degree of freedom, say, using multiple antennas.

When proving the capacity formula for the case where the state process is know at both the transmitter and receiver, we adopted the multiplexing technique, i.e., using different coding schemes for different channel states. But such a scheme may cause long decoding delay. Furthermore, for fading processes with continuous amplitude, there are essentially infinite states, which make the multiplexing method infeasible. Fortunately, for fading channels with additive Gaussian noise, the multiplexing method can be replaced by dynamic power control. That is because we if assume the fading process is $\{\sqrt{P^*(\mathbf{h}(t))}\mathbf{h}(t)\}_t$ and the power constraint is $\mathbb{E}\mathbf{x}^2 \leq 1$, then the capacity with the state information at the receiver is exactly

$$\frac{1}{2}\mathbb{E}_{\mathbf{h}} \log\left(1+\frac{P^*(\mathbf{h})\mathbf{h}^2}{\sigma_N^2}\right).$$

Since $P^*(\mathbf{h}(t))$ is a deterministic function of $\mathbf{h}(t)$, the receiver can compute $\{\sqrt{P^*(\mathbf{h}(t))}\mathbf{h}(t)\}_t$ from the fading process $\{\mathbf{h}(t)\}_t$. Out assumption is thus valid. So in order to achieve the capacity, we only need use the power control policy $P^*(h)$ and a coding scheme with power constraint $\mathbb{E}\mathbf{x}^2 \leq 1$ designed for the effective fading process $\{\sqrt{P^*(\mathbf{h}(t))}\mathbf{h}(t)\}_t$. Note: the resulting transmission power is $\mathbb{E}_{\mathbf{h}}P^*(\mathbf{h}) = P$.

3. $\{\mathbf{h}(t)\}_t$ is unknown: (noncoherent channel capacity)

   For this case, it is insufficient to determine the channel capacity with only the knowledge of the stationary distribution of $\{\mathbf{h}(t)\}_t$. We need more complete statistic characterization of the fading process. It is in general very hard to get a closed form expression of the noncoherent channel capacity, and the results rely heavily on the assumption of the fading process. People usually resort to the asymptotic analysis in the high SNR regime or the low SNR regime. We will return to this topic by the end of this semester.

3 MIMO: Ergodic Capacity

$\mathbf{Y}(t) = \mathbf{H}(t)X(t) + \mathbf{N}(t)$, where $\{\mathbf{H}(t) \in \mathcal{R}^{n\times m}\}_t$ is stationary and ergodic, and $\{\mathbf{N}(t)\}_t$ is i.i.d. with $\mathbf{N}(t) \sim \mathcal{N}(0, \sigma_N^2 I_n)$ for all t.

1. $\{\mathbf{H}(t)\}_t$ is known at the receiver: Let $K_{\mathbf{X}} = \mathbb{E}\mathbf{X}\mathbf{X}^T$. We have

$$\begin{aligned} C &= \max_{\text{trace}(K_{\mathbf{X}})\le P} I(\mathbf{X};\mathbf{Y}|\mathbf{H}) \\ &= \max_{\text{trace}(K_{\mathbf{X}})\le P} \mathbb{E}_{\mathbf{H}} I(\mathbf{X};\mathbf{Y}|\mathbf{H}=H) \\ &= \max_{\text{trace}(K_{\mathbf{X}})\le P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}K_{\mathbf{X}}\mathbf{H}^T\right). \end{aligned}$$

Raleigh fading: the entries of $\mathbf{H}$ are i.i.d. $\mathcal{N}(0,1)$.

Let $K_{\mathbf{X}} = UDU^T$ be the eigenvalue decomposition of $K_{\mathbf{X}}$, where U is a unitary matrix and D is a diagonal with nonnegative entries. Note: $\text{trace}(K_{\mathbf{X}}) = \text{trace}(D)$.

Lemma 3.1. *The distribution of* $\mathbf{H}U$ *is the same as that of* $\mathbf{H}$.

Proof. Let $\mathbf{g}_{i,j}$ be the (i,j) entry of $\mathbf{H}U$. Let $\mathbf{H}_i^T$ be the ith row of $\mathbf{H}$, and U_j be the jth column of U.

We have

$$\begin{aligned} \mathbb{E}\mathbf{g}_{i,j}\mathbf{g}_{k,l} &= \mathbb{E}\left(\mathbf{H}_i^T U_j \mathbf{H}_k^T U_l\right) \\ &= \mathbb{E}\left(\mathbf{H}_i^T U_j U_l^T \mathbf{H}_k\right) \\ &= \mathbb{E}\left(U_l^T \mathbf{H}_k \mathbf{H}_i^T U_j\right) \\ &= U_l^T \mathbb{E}\left(\mathbf{H}_k\mathbf{H}_i^T\right) U_j. \end{aligned}$$

If $i \neq k$, we have $\mathbb{E}\left(\mathbf{H}_k\mathbf{H}_i^T\right) = 0$ and thus $\mathbb{E}\mathbf{g}_{i,j}\mathbf{g}_{k,l} = 0$. For $i = k$, we have $\mathbb{E}\left(\mathbf{H}_k\mathbf{H}_i^T\right) = I_m$ and thus $\mathbb{E}\mathbf{g}_{i,j}\mathbf{g}_{k,l} = 0$ if $j \neq l$ and 1 otherwise. □

By the above lemma, we have

$$\max_{\text{trace}(K_{\mathbf{X}})\le P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}K_{\mathbf{X}}\mathbf{H}^T\right) = \max_{\text{trace}(D)\le P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}D\mathbf{H}^T\right).$$

By symmetry, if D_π is a diagonal matrix obtained via permutating the diagonal entries of D, then

$$\frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}D_\pi\mathbf{H}^T\right) = \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}D\mathbf{H}^T\right).$$

Hence by the concavity of $\log\det$ and the power constraint, the optimal D must be $\frac{P}{m}I_m$. Therefore, we have

$$C = \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{P}{m\sigma_N^2}\mathbf{H}\mathbf{H}^T\right).$$

Let $\mathbf{H} = \mathbf{V}\mathbf{\Lambda}\mathbf{W}^T$ be the singular value decomposition of $\mathbf{H}$. Let $\lambda_1, \lambda_2, \cdots, \lambda_{\min(m,n)}$ be the singular values of $\mathbf{H}$. Note: $\lambda_1^2, \lambda_2^2, \cdots, \lambda_{\min(m,n)}^2$ are the non-zero eigenvalues of $\mathbf{H}\mathbf{H}^T$

and $\mathbf{H}^T\mathbf{H}$. We have

$$\begin{aligned}
C &= \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{P}{m\sigma_N^2}\mathbf{H}\mathbf{H}^T\right) \\
&= \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{P}{m\sigma_N^2}\mathbf{V}\mathbf{\Lambda}\mathbf{\Lambda}^T\mathbf{V}^T\right) \\
&= \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{P}{m\sigma_N^2}\mathbf{\Lambda}\mathbf{\Lambda}^T\right) \\
&= \frac{1}{2}\mathbb{E}_{\mathbf{H}} \sum_{i=1}^{\min(m,n)} \log\left(1 + \frac{P\lambda_i^2}{m\sigma_N^2}\right).
\end{aligned}$$

Note: $\mathrm{trace} AA^T = \mathrm{trace} A^TA = \sum_{i,j} a_{i,j}^2$. Law of energy (/power) conservation:

$$\sum_{i,j}\mathbf{h}_{i,j}^2 = \mathrm{trace}(\mathbf{H}\mathbf{H}^T) = \mathrm{trace}(\mathbf{V}\mathbf{\Lambda}\mathbf{\Lambda}^T\mathbf{V}^T) = \mathrm{trace}(\mathbf{\Lambda}^T\mathbf{\Lambda}) = \sum_{i=1}^{\min(m,n)}\lambda_i^2.$$

Define $\mathsf{SNR} = \frac{P}{\sigma_N^2}$. In the low SNR regime, we have

$$\begin{aligned}
C &\approx \frac{1}{2m}\mathsf{SNR}\log_2 e \sum_{i=1}^{\min(m,n)} \mathbb{E}_{\mathbf{H}}\lambda_i^2 \\
&= \frac{1}{2m}\mathsf{SNR}\log_2 e \sum_{i=1}^{\min(m,n)} \mathbb{E}_{\mathbf{H}}\mathrm{trace}(\mathbf{H}\mathbf{H}^T) \\
&= \frac{1}{2m}\mathsf{SNR}\log_2 e \sum_{i,j} \mathbb{E}\mathbf{h}_{i,j}^2 \\
&= \frac{n}{2}\mathsf{SNR}\log_2 e
\end{aligned}$$

Thus at low SNR, an m by n system yields a power gain of n over a single antenna system. This is due to the fact that the multiple receive antennas can coherently combine their received signals to get a power boost (Note: at low SNR, the capacity scales linearly with the power). Note that increasing the number of transmit antennas does not increase the power gain since, unlike the case when the channel is known at the transmitter, transmit beamforming cannot be done to constructively add signals from the different antennas. Thus, at low SNR and without channel knowledge at the transmitter, multiple transmit antennas are not very useful: the performance of an m by n channel is comparable with that of a 1 by n channel.

At high SNR, we have

$$\begin{aligned}
C &\approx \frac{1}{2}\mathbb{E}_{\mathbf{H}} \sum_{i=1}^{\min(m,n)} \log\left(\frac{1}{m}\mathsf{SNR}\lambda_i^2\right) \\
&= \frac{\min(m,n)}{2}\log\mathsf{SNR} + \frac{1}{2}\mathbb{E}_{\mathbf{H}} \sum_{i=1}^{\min(m,n)} \log\frac{\lambda_i^2}{m}.
\end{aligned}$$

Hence the capacity gain is $\min(m,n)$ over a single antenna system. $\min(m,n)$ is referred to as the degree of freedom.

2. $\{\mathbf{H}(t)\}_t$ is known at both the transmitter and receiver:

$$\begin{aligned} C &= \max_{\text{trace}(\mathbb{E}K_{\mathbf{X}}(\mathbf{H}))\leq P} I(\mathbf{X};\mathbf{Y}|\mathbf{H}) \\ &= \max_{\text{trace}(\mathbb{E}K_{\mathbf{X}}(\mathbf{H}))\leq P} \mathbb{E}_{\mathbf{H}} I(\mathbf{X};\mathbf{Y}|\mathbf{H}=H) \\ &= \max_{\text{trace}(\mathbb{E}K_{\mathbf{X}}(\mathbf{H}))\leq P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}K_{\mathbf{X}}(\mathbf{H})\mathbf{H}^T\right) \\ &= \max_{\text{trace}(\mathbb{E}K_{\mathbf{X}}(\mathbf{H}))\leq P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}K_{\mathbf{X}}(\mathbf{H})\mathbf{H}^T\right). \end{aligned}$$

Let $\mathbf{H} = \mathbf{V}\mathbf{\Lambda}\mathbf{W}^T$ be the singular decomposition of $\mathbf{H}$. We have

$$\begin{aligned} & \max_{\text{trace}(\mathbb{E}K_{\mathbf{X}}(\mathbf{H}))\leq P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}K_{\mathbf{X}}(\mathbf{H})\mathbf{H}^T\right) \\ = & \max_{\text{trace}(\mathbb{E}K_{\mathbf{X}}(\mathbf{\Lambda}))\leq P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{\Lambda}K_{\mathbf{X}}(\mathbf{\Lambda})\mathbf{\Lambda}^T\right) \\ = & \max_{\sum_{i=1}^{\min(m,n)}\mathbb{E}P(\lambda_i)\leq P} \frac{1}{2}\mathbb{E}_{\mathbf{H}} \sum_{i=1}^{\min(m,n)} \log\left(1 + \frac{P(\lambda_i)\lambda_i^2}{\sigma_N^2}\right). \end{aligned}$$

The optimal power control policy $P(\lambda_i)$ is given by

$$P^*(\lambda_i) = \max\left(\mu - \frac{\sigma_N^2}{\lambda_i^2}, 0\right),$$

where μ is determined by the power constraint

$$\sum_{i=1}^{\min(m,n)} \mathbb{E}P^*(\lambda_i) = P.$$

Note that this is waterfilling over time and space (the eigenmodes).

4 Outage Capacity

If $\mathbf{H}$ is chosen randomly at the beginning of all time and is held fixed for all the uses of the channel, the maximum mutual information is in general not equal to the channel capacity. For the case when entries of $\mathbf{H}$ are i.i.d. Gaussian with $\mathbb{E}\mathbf{h}_{i,j} = 1$, the Shannon capacity of the channel is zero: however small the rate we attempt to communicate at, there is a non-zero probability that the realized $\mathbf{H}$ is incapable of supporting it no matter how long we take our code length. On the other hand one can talk about a tradeoff between outage probability and supportable rate. Namely, given a rate R, and power P, one can find $\mathbb{P}_{\text{out}}(R,P)$ such that for any rate less than R and any δ there exists a code satisfying the power constraint P for which the error probability is less than δ for all but a set of $\mathbf{H}$ whose total probability is less than $P_{\text{out}}(R,P)$:

$$P_{\text{out}}(R,P) = \inf_{\text{trace}(K_{\mathbf{X}})\leq P} \mathbb{P}\left(\frac{1}{2}\log\det\left(I_n + \frac{1}{\sigma_N^2}\mathbf{H}K_{\mathbf{X}}\mathbf{H}^T\right) < R\right).$$

For finite code length, the outage probability serves as a lower bound of the real error probability over all coding schemes. As the code length goes to infinity, this lower bound becomes tight. We can view such a model as a compound channel indexed by the fading state H.

Clearly, for fixed SNR (or P), there exists a tradeoff between $\mathbb{P}_{\text{out}}$ and R. But it is hard to get an explicit characterization of this tradeoff. A common approach is to let SNR go to infinity and study the tradeoff between the increasing rate of R and the decreasing rate of $\mathbb{P}_{\text{out}}$ with respect to SNR. This is also called the diversity-multiplexing tradeoff. We will discuss this in detail later this semester.

References

[1] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 1991.

[2] J. Wolfowitz, Coding Theorems of Information Theory. Berlin, Prentice-Hall, 1961.

[3] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Tran. Inform. Theory*, vol. 43, pp.1986-1992, Nov. 1997.

[4] R. G. Gallager, Information Theory and Reliable Communication. New York, Wiley, 1968.

[5] G. J. Foschini, "Layered spaceCtime architecture for wireless communication in a fading environment when using multi-element antennas," *AT&T Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41-59, 1996.

[6] G. Foschini and M. Gans, "On limits of wireless communications in fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311-335, 1998.

[7] L. Ozarow, S. Shamai (Shitz), and A. Wyner, "Information-theoretic considerations in cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359-378, May 1994.

[8] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585-595, Nov./Dec. 1999.

[9] T. Marzetta and B. Hochwald, "Capacity of mobile multiple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Jan. 1999.

[10] S. Verdú, "Spectral efficiency in the wide-band regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319-1343, June 2002

[11] L. Zheng and D. N. C. Tse,"Communicating on the Grassmann manifold: A geometric approach to the noncoherent multiple antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359-383, Feb. 2002.

[12] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095-3110, Dec. 2004.

[13] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426-2467, Oct. 2003.

[14] D. Tse and P. Viswanath, Fundamentals of Wireless Communication, to be published by Cambridge University Press. Online: http://degas.eecs.berkeley.edu/dtse.

[15] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels", *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684-702, June 2003.

[16] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. IT-44, pp. 2148-2177, Oct. 1998.

[17] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[18] E. Biglieri, J. Proakis, and S. Shamai "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619-2692, Oct. 1998.

[19] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073-1096, May 2003.