Multi-antenna Gaussian Channel (MIMO)

Wei Zhang

Department of Electrical Engineering
University of Notre Dame

I. Introduction

Nowadays, the research on multi-input multi-output (MIMO) systems becomes a hot spot for the reason that it can greatly increase the spectral efficiency (capacity) over a limited bandwidth. Since it has additional dimension to carry information, the capacity gain of MIMO systems over the single input single output (SISO) system is a remedy to the fast increasing demands of higher data rates in wireless communications. The definition of MIMO itself doesn't give much specification of the channel, but we will limit our discussion only on multi-antenna Gaussian channel, which is defined as a single user Gaussian channel with multiple transmitter/receiver antennas. The single user scenario gives us the convenience to process the transmitted and received information jointly. The system is shown in Figure 1.

The summary presents some major results in Telatar's paper [1]. If the channel transfer matrix is fixed and both known to the transmitter/receiver, through some clever joint processing, we can achieve the capacity of this type of channel. In the fading environment, under the assumption of independent fades and noises at the different antennas, the capacity gain of multi-antenna systems over single-antenna systems can be very large. For example, if the number of receiving antenna equals the number of transmitting antenna, the capacity grows approximate linearly with this number when it is large, which does not occur in the fixed transfer matrix situation.

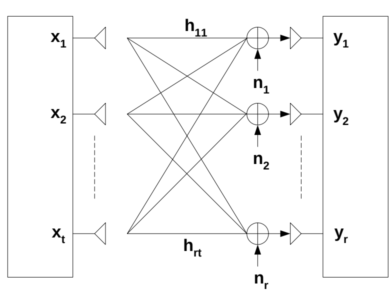

Figure 1. MIMO channel

II. Model of Multi-antenna Gaussian Channel

As shown in the Figure 1, suppose there are t transmitting antennas and r receiving antennas, we consider a linear superposition channel model. The received vector signal $\mathbf{y} \in \mathbb{C}^r$ is

$$\mathbf{y} = H\mathbf{x} + \mathbf{n} \tag{1}$$

where H is a $r \times t$ complex matrix, $\mathbf{x} \in \mathbb{C}^t$ is the transmitted vector signal and $\mathbf{n}$ is zero-mean complex Gaussian noise with independent real and imaginary part. The matrix H characterizes the channel transition property, in which each entry H_{ij} represents the path gain from j-th transmitter antenna to i-th receiver antenna. Reasonably, we assume the noise of each receiving antenna is independent, i.e., the covariance matrix of $\mathbf{n}$ is $\mathcal{E}[\mathbf{n}\mathbf{n}^\dagger] = \mathbf{I}_r$[1].

Given a total power constraint P to the transmitter antennas,

$$\mathcal{E}[\mathbf{x}^\dagger \mathbf{x}] \leq P, \tag{2}$$

naturally we want to know the channel capacity C. For different realization of the channel transition matrix, we can divide it into 3 typical situations:

- H is a deterministic matrix.
- H is a random matrix, each use of the channel corresponding to an independent realization.
- H is random, but is fixed once it is chosen.

In [1], the author first solve the capacity problem of the first case and provide some useful skill to the later cases.

III. Gaussian Channel with Fixed Transfer Function

If the channel transfer function is constant and known, which is the common case in the fixed wireless service such as Wireless Local Loop (WLL), then we can define the channel capacity as

$$C(H, P) = \max_{p(\mathbf{x}):\, \mathcal{E}[\mathbf{x}^\dagger \mathbf{x}] \leq P} I(\mathbf{x}, \mathbf{y}) \tag{3}$$

From the observation, we can see that for each input of the receiver antenna 1, if it only cares about the output of the transmitter antenna 1, then we can look the output of other transmitter antenna as crosstalks. So we could call it correlated parallel Gaussian channel. If using some method to cancel the interference from other transmitter antenna, then we get a classical independent parallel Gaussian channel. For every $H \in \mathbb{C}^{r\times t}$, by the singular value decomposition theorem, it can be written as

$$H = UDV^\dagger \tag{4}$$

where $U \in \mathbb{C}^{r\times r}$ and $V \in \mathbb{C}^{t\times t}$ are unitary, and $D \in \mathbb{R}^{r\times t}$ is a non-negative and diagonal with these entries being the non-negative square roots of eigenvalues of $HH^\dagger$. Then the channel model (1) becomes

$$\mathbf{y} = UDV^\dagger \mathbf{x} + \mathbf{n}. \tag{5}$$

After orthogonal transform of $\tilde{\mathbf{y}} = U^\dagger \mathbf{y}$,$\tilde{\mathbf{x}} = V^\dagger \mathbf{x}$ and $\tilde{\mathbf{n}} = U^\dagger \mathbf{n}$, the channel changes to

$$\tilde{\mathbf{y}} = D\tilde{\mathbf{x}} + \tilde{\mathbf{n}}. \tag{6}$$

It's the property of orthogonal transform that the power constraint reserves, the distribution of noise not being changed and more important, information lossless. Note that D is diagonal, thus we decompose the correlated parallel channel into independent parallel channels, see Figure 2,

$$\tilde{y}_i = \lambda_i^{1/2} \tilde{x}_i + \tilde{n}_i, \quad 1 \leq i \leq \min(r, t) \tag{7}$$

[1] $\mathbf{A}^\dagger$ means the Hermitian (conjugate transpose) transform of a matrix $\mathbf{A}$.

To maximize the mutual information $I(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$, we need to choose $\tilde{x}_i$'s to be independent, and each has independent Gaussian, zero-mean real and imaginary part. Then allocating power to each virtual channel via water filling,

$$\mathcal{E}[\mathcal{R}e(\tilde{x}_i)^2] = \mathcal{E}[\mathcal{I}m(\tilde{x}_i)^2] = \frac{1}{2}(\mu - \lambda_i^{-1})^+ \tag{8}$$

where μ is chosen to meet power constraint [2]. Then the power and capacity can be parameterized as

$$P(\mu) = \sum_i (\mu - \lambda_i^{-1})^+, \quad C(\mu) = \sum_i [\ln(\mu\lambda_i)]^+. \tag{9}$$

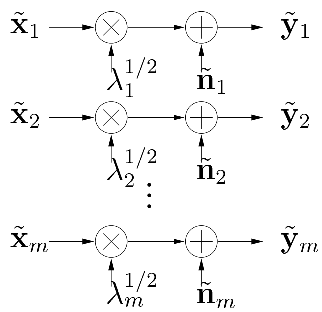

Figure 2. Independent parallel Gaussian channel

We can also derive the capacity formula through a way more useful in the fading case. The deduction needs some knowledge on the circularly symmetric complex Gaussian random vector.

Definition 1: A Gaussian random vector $\mathbf{x}$ is circularly symmetric, if for $\hat{\mathbf{x}} = [\mathcal{R}e[\mathbf{x}]\ \mathcal{I}m[\mathbf{x}]]^\tau$,

$$cov[\hat{\mathbf{x}}] = \frac{1}{2}\begin{bmatrix} \mathcal{R}e[Q] & -\mathcal{I}m[Q] \\ \mathcal{I}m[Q] & \mathcal{R}e[Q] \end{bmatrix}, \quad \text{where } Q = cov[\mathbf{x}]. \tag{10}$$

The most important property of this kind of random vectors is that they are entropy maximizer,

$$H(\gamma_Q) = \log\det(\pi e Q). \tag{11}$$

The mutual information can be written as

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) = h(\mathbf{y}) - h(\mathbf{n}). \tag{12}$$

We only need to maximize $h(\mathbf{y})$, because $h(\mathbf{n})$ is fixed. If $\mathbf{x}$ is zero-mean, $\mathcal{E}[\mathbf{x}\mathbf{x}^\dagger] = Q$, then $\mathcal{E}[\mathbf{y}\mathbf{y}^\dagger] = I_r + HQH^\dagger$. For any Q, $h(\mathbf{y})$ is maximized when $\mathbf{y}$ is circularly symmetric Gaussian. In this case,

$$I(\mathbf{x};\mathbf{y}) = \log\det[I_r + HQH^\dagger] = \log\det[I_r + QH^\dagger H]. \tag{13}$$

Using eigenvalue decomposition, $H^\dagger H = U^\dagger \Lambda U$, with $\Lambda = diag[\lambda_1, \lambda_2, \cdots, \lambda_t]$,

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= \log\det[I_t + \Lambda^{1/2} U Q U^\dagger \Lambda^{1/2}] \\ &= \log\det[I_t + \Lambda^{1/2} \tilde{Q} \Lambda^{1/2}] \quad \left(\tilde{Q} = UQU^\dagger\right) \\ &\leq \log\prod_i (1 + \tilde{Q}_{ii}\lambda_i), \end{aligned} \tag{14}$$

[2] a^+ means $\max(0, a)$.

with equality holds when $\tilde{Q}$ is also diagonal. Then we can use the same water filling to the diagonal elements of $\tilde{Q}$ and get the same capacity formula.

From the decomposition procedure, we can see that this capacity can be achieved by the following precoding technique. First we form the data stream into a vector signal $\tilde{\mathbf{x}}$ with $m = min(r,t)$ substreams, code and modulate according to the water-filling of the power to the different substreams, and then rotate the signal $\tilde{\mathbf{x}}$ by multiplying V. At the receiver, we rotate the received signal $\tilde{\mathbf{y}}$ by $U^\dagger$ to get the equivalent parallel channel, and then we can decode the substreams. In the process of coding, we can either jointly or independently encode the substreams. If we do encoding jointly, from deduction of the error exponents of parallel channels in [2], there will be a gain of m to the log of the error probability, which is called antenna diversity gain in space-time codes.

IV. Gaussian Channel with Rayleigh Fading

On the fading channels, the path gains are characterized as random variables. In the rich scattering environment, there are many path exist from one transmitter antenna to one receiver antenna. Based on the central limit theorem, the summed received signal is Gaussian distributed. So it is natural to choose each entry H_{ij} as an i.i.d. Gaussian complex random variable with independent real and imaginary part, each with variance $1/2$. The Gaussian distribution with independent imaginary and real part results from rich scattering environment. This is usual *Rayleigh fading model* in wireless communication. So H_{ij} has uniform phase and Rayleigh amplitude. The i.i.d. entries assumption holds when the transmitter and receiver antennas were enough physically separated.

Assume perfect knowledge of channel side information (CSI), the channel is consist of input $\mathbf{x}$ and output $(\mathbf{y}, \mathbf{H})$. Define the ergodic (mean) capacity as

$$\begin{aligned} C &= \max_{p(\mathbf{x})} I\left(\mathbf{x}; \mathbf{y}, \mathbf{H}\right) \\ &= \max_{p(\mathbf{x})} I\left(\mathbf{x}; \mathbf{H}\right) + I\left(\mathbf{x}; \mathbf{y}|\mathbf{H}\right) \quad (\mathbf{x}, H \text{ independent}, I\left(\mathbf{x}; \mathbf{H}\right) = 0) \\ &= \max_{p(\mathbf{x})} \mathcal{E}_{\mathbf{H}}\left[I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H)\right]. \end{aligned} \tag{15}$$

In the fading situation, because the transmitter has no knowledge of the channel state, so the most robust and conservative strategy for the transmitter is allocating power equally to each antenna and makes their outputs independent. Fortunately, this strategy will achieve the capacity in a long run if the channel is ergodic.

Theorem 1: The capacity of multiple antenna Gaussian channel with fading is $\mathcal{E}[\log\det[I_r + (P/t)HH^\dagger]]$. The capacity is achieved when $\mathbf{x}$ is circularly symmetric complex Gaussian with zero-mean and covariance $(P/t)I_t$.

Sketch of the proof: If $\mathbf{x}$ has a specific covariance Q, then the choice of $\mathbf{x}$ to maximize $I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H)$ is the circularly symmetric complex Gaussian random vector with covariance Q. Thus we need to maximize

$$\Psi(Q) = \mathcal{E}[\log\det(I_r + \mathbf{H}Q\mathbf{H}^\dagger)] \tag{16}$$

Since Q is covariance matrix, so it's non-negative definite, so we can diagonalize it using a unitary matrix U. The U is absorbed by $\mathbf{H}$ and will not affect its distribution, so we can further constraint the choice of Q as a diagonal matrix. Like the random coding method in the proof of channel coding theorem, we select a Q and then permute it in all possible ways. Denote the permuted matrix as $Q^{\Pi} = \Pi Q \Pi^{\dagger}$. The average of all this permuted matrices

$$\tilde{Q} = \frac{1}{t!} \sum_{\Pi} Q^{\Pi} \tag{17}$$

satisfies $\Psi(\tilde{Q}) \geq \Psi(Q)$ and power constraint $tr(\tilde{Q}) = tr(Q)$. The first claim is based on the concavity of $\log\det$ [3] and the second is that $tr(AB) = tr(BA)$ if both AB and BA is possible. Obviously $\tilde{Q}$ is multiple of identity matrix. So we choose the optimal matrix Q to be the largest possible, i.e., $(P/t)I_t$.

Although through Theorem 1, we know when will the capacity achieve, but the evaluation of the capacity is still a problem when r and t get large. So Theorem 2 gives the close form of the capacity.

Theorem 2: The capacity of the channel with t transmitters and r receivers under power constraint P equals

$$\int_0^{\infty} \log(1 + P\lambda/t) \sum_{k=0}^{m-1} \frac{k!}{(k+n-m)!} [L_k^{n-m}(\lambda)]^2 \lambda^{n-m} e^{-\lambda} d\lambda \tag{18}$$

where $m = \min(r,t)$ and $n = \min(r,t)$, and L_j^i are the associated Laguerre polynomials.

Sketch of the proof: From the capacity formula, note that $\det(I_r + (P/t)\mathbf{H}\mathbf{H}^{\dagger}) = \det(I_t + (P/t)\mathbf{H}^{\dagger}\mathbf{H})$ and define

$$\mathbf{W} = \begin{cases} \mathbf{H}\mathbf{H}^{\dagger} & r < t \\ \mathbf{H}^{\dagger}\mathbf{H} & r \geq t. \end{cases} \tag{19}$$

$n = \max(r,t)$ and $m = \min(r,t)$. Then $\mathbf{W}$ is $m \times m$ non-negative matrix. So the capacity can be expressed in terms of its eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_m$:

$$\mathcal{E}\left[\sum_{i=1}^{m} \log(1 + (P/t)\lambda_i)\right]. \tag{20}$$

Actually the distribution law of $\mathbf{W}$ is known as *Wishart distribution with parameters* m, n. Its eigenvalues density function is

$$p_{\lambda}(\lambda_1, \lambda_2, \cdots, \lambda_m) = K(m,n) \prod_i \lambda_i^{n-m} e^{-\lambda_i} \prod_{i<j} (\lambda_i - \lambda_j)^2. \tag{21}$$

Recall the last term in the product is the determinant of Vandermonde matrix

$$D = \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_m \\ \vdots & & \vdots \\ \lambda_1^{m-1} & \cdots & \lambda_m^{m-1} \end{bmatrix}. \tag{22}$$

The distribution can be written as $p_{\lambda}(\lambda_1, \lambda_2, \cdots, \lambda_m) = K(m,n) \det^2 D \prod_i \lambda_i^{n-m} e^{-\lambda_i}$. In the space of real functions in which defines a specific inner product, applying Gram-Schmidt orthogonal transform to the sequence $1, \lambda, \cdots, \lambda^{m-1}$, the D changes into

$$\tilde{D} = \begin{bmatrix} \varphi_1(\lambda_1) & \cdots & \varphi_1(\lambda_m) \\ \vdots & & \vdots \\ \varphi_m(\lambda_1) & \cdots & \varphi_m(\lambda_m) \end{bmatrix}. \tag{23}$$

Thus $\int_0^\infty \varphi_i(\lambda)\varphi_j(\lambda)\lambda^{n-m}e^{-\lambda}d\lambda = \delta_{ij}$. By expand the determinant of $\tilde{D}$, the distribution is

$$p_\lambda(\lambda_1, \lambda_2, \cdots, \lambda_m) = C(m,n)\sum_{\alpha,\beta}(-1)^{per(\alpha)+per(\beta)}\prod_i \varphi_{\alpha_i}(\lambda_i)\varphi_{\beta_i}(\lambda_i)\lambda_i^{n-m}e^{-\lambda_i}. \tag{24}$$

Integrating over $\lambda_2, \cdots, \lambda_m$ using the property of this specific inner product, we get marginal distribution $p(\lambda_1)$. Then the expectation (20) can be calculated, which will give the last result of Theorem 2. The capacity calculated by Theorem 2 for $1 \leq r, t \leq 20$ and $P = 20dB$ shown in Figure 3.

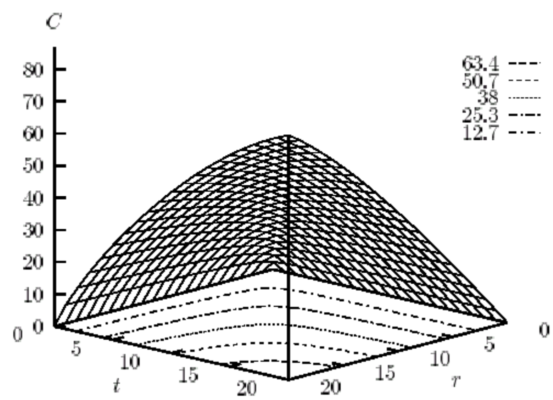

Figure 3. Capacity of MIMO fading channel

It is interesting to observe some extreme scenarios.

- Only transmitter diversity: $r$fixed$, t \to \infty$, $C \to r\log(1+P)$.
- Only receiver diversity: $t = 1, r \to \infty$, $C \to \log(1+Pr)$
- Both transmitter/receiver diversity: $r = t = n$, $C \sim r\int_0^4 \log(1+P\nu)\frac{1}{\pi}\sqrt{\frac{1}{\nu}-\frac{1}{4}}d\nu$

The last situation is most interesting because the capacity grows linearly as the number of antennas. Compare with the independent parallel Gaussian channel $r = t = n$ and $H = I_n$, of which $C = n\log(1+P/n) \to P$ when $n \to \infty$, the capacity gain is obviously. The fading seems bless the capacity. We can think it this way, although sometimes fading makes some path corrupted, but because the use of multiple antennas, there are always possible very good paths due to the constructive combining of signals, we can achieve very high data rate on this path. So in the meaning of average, the MIMO's capacity is larger than the unfaded channel.

V. NON-ERGODIC CHANNELS

In the case of channel is not ergodic, the transfer function H is random, but will be fixed once it's realized. Under this situation, the maximum mutual information is in general not equal to the channel capacity because it is not always achievable. Another measure of channel capacity that is frequently used is *outage capacity*, which is defined through the tradeoff between the outage probability and supportable rate. The capacity is treated as an random variable associated with a outage probability q. Simply,

$$Pr\{C \leq C_{\text{outage}}\} = q. \tag{25}$$

q is strictly defined as

$$q = \inf_{Q:\, Q \geq 0, tr(Q) \leq P} Pr\{\log \det(I_r + \mathbf{H} Q \mathbf{H}^{\dagger}) \leq C_{outage}\}. \tag{26}$$

VI. Further Reading

In [1], the author supposed the coherent receiver condition. For the non-coherent receiver condition, a comprehensive investigation can be found in [4].

References

[1] I. Telatar, "Capacity of multi-antenna gaussian channels," *Bell Labs Technical Memorandum*, June 1995.

[2] R. G. Gallager, *Information theory and reliable communication.* New York: John Wiley & Sons, 1968.

[3] R. Horn and R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[4] B. Hochwald and T. Marzetta, "Capacity of mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Transactions on Information Theory*, Jan. 1999.