Lecture Series on Math Fundamentals for MIMO Communications

Topic 1. Complex Random Vector and Circularly Symmetric Complex Gaussian Matrix

Drs. Yindi Jing and Xinwei Yu

University of Alberta

Last Modified on: January 17, 2020

Content

Part 1. Definitions and Results

1.1 Complex random variable/vector/matrix

1.2 Circularly symmetric complex Gaussian (CSCG) matrix

1.3 Isotropic distribution, decomposition of CSCG random matrix, Complex Wishart matrix and related PDFs

Part 2. Some MIMO applications

Part 3. References

Part 1. Definitions and Results

1.1 Complex random variable/vector/matrix

- Real-valued random variable
- Real-valued random vector
- Complex-valued random variable
- Complex-valued random vector
- Real-/complex-valued random matrix

Real-valued random variable

Def. A map from the sample space to the real number set $\Omega \to \mathbb{R}$.

- A **sample space** Ω;
- A **probability measure** $\mathbb{P}(\cdot)$ defined on Ω;
- An experiment on Ω;
- A **function**: each outcome of the experiment $\mapsto$ a real number.

Notation: random variable X; sample value x.

Outcomes of the experiment: In Ω. "Invisible". Does not matter.
What matters: The values **value** x assigned by the function.

Examples:
Flip a coin and head $\to 0$, Tail $\to 1$.
A random bit of 0 or 1.
Roll a die and even $\to 0$, odd $\to 1$.

How to describe a random variable?

Def. The **cumulative distribution function (CDF)** of X:

$$F_X(x) = \mathbb{P}\left(\omega \in \Omega : X(\omega) \leq x\right) = \mathbb{P}\left(X \leq x\right).$$

Def. The **probability density function (PDF)** of X:

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

Keep in mind: A random variable though the result of an experiment, can be separated to the physical experiment. The values matter, not the outcomes of the experiment.
In other words, ω is invisible, all one can see is x.

Common concepts, models, and properties of random variables.

- Discrete random variable, continuous random variable, mixed random variable
- Mean, variance, moments
- Multiple random variables: Joint CDF/PDF, marginal CDF/PDF, conditional CDF/PDF, etc.
- Multiple random variables: Independence and correlation

(Real-valued) random vector.

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix}, \text{value } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix}$$

- K dimensional.
- Each X_i: random variable;
- Describe through joint PDF:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1, X_2, \cdots, X_K}(x_1, x_2, \cdots, x_K)$$

Mean vector and Covariant Matrix

- Mean vector.

$$\mathbf{m} = \mathbb{E}(\mathbf{X}) = \begin{bmatrix} \mathbb{E}\left(X_1\right) \\ \vdots \\ \mathbb{E}\left(X_K\right) \end{bmatrix} = \begin{bmatrix} m_1 \\ \vdots \\ m_K \end{bmatrix},$$

- Covariance matrix.

$$\mathbf{\Sigma} = \mathbb{E}\{(\mathbf{X}-\mathbf{m})(\mathbf{X}-\mathbf{m})^T\} = \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \mathbb{E}\left[(X_i - m_i)(X_j - m_j)\right] & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix},$$

$\mathbf{\Sigma} \succeq 0$ (positive semidefinite matrix).

Complex-valued random variable.

$$X = X_r + jX_s.$$

Equivalent to 2-dimensional real-valued random vector:

$$\hat{X} = \begin{bmatrix} X_{\mathrm{r}} \\ X_{\mathrm{s}} \end{bmatrix}.$$

To describe $\hat{X}$, use joint PDF of $(X_r, X_s) \Longleftrightarrow$ joint PDF of the random vector $\hat{X}$.

$$f_{\hat{X}}(\hat{x}) = f_{X_r,X_s}(x_r, x_s).$$

Complex-valued random vector

Equivalent to real-valued random vector with twice dimension:

$$\mathbf{X} = \mathbf{X}_r + j\mathbf{X}_s \quad \Longleftrightarrow \quad \hat{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_\mathrm{r} \\ \mathbf{X}_\mathrm{s} \end{bmatrix}.$$

To describe $\mathbf{X}$, use the joint PDF of $(\mathbf{X}_r, \mathbf{X}_s)$ (twice the dimension of $\mathbf{X}$):

$$\begin{aligned} f_{\hat{\mathbf{X}}}(\hat{\mathbf{x}}) &= f_{\mathbf{X}_r,\mathbf{X}_s}(\mathbf{x}_r, \mathbf{x}_s) \\ &= f_{X_{r,1},\cdots,X_{r,K},X_{s,1},\cdots,X_{s,K}}(x_{r,1},\cdots,x_{r,K},x_{s,1},\cdots,x_{s,K}). \end{aligned}$$

Real-/complex-valued random matrix:

$\mathbf{X} = [x_{ij}] : M \times N$ matrix.

Each entry x_{ij} is a real-/complex-valued random variable.

Also use $\mathbf{X}$ for a sample or a realization.

$\Longleftrightarrow$ **an (MN)-dimensional real-/complex random vector.**

To make the difference between random vector and random variables, use $\mathbf{x}$ for both a random vector and its realization. Reserve $\mathbf{X}$ for both a random matrix and its realization.

The vectorization operation for $\mathbf{X} = [x_{ij}]$.

By columns:

$$\mathrm{vec}(\mathbf{X}) = \begin{bmatrix} x_{11} \\ \vdots \\ x_{M1} \\ \vdots \\ x_{1N} \\ \vdots \\ x_{MN} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{col,1} \\ \vdots \\ \mathbf{x}_{col,N} \end{bmatrix}$$

By rows:

$$\begin{aligned} \mathrm{vec}_{row}(\mathbf{X}) &= \begin{bmatrix} x_{11} & \cdots & x_{1N} & \cdots & x_{N1} & \cdots & x_{MN} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{x}_{row,1} & \cdots & \mathbf{x}_{row,M} \end{bmatrix}. \end{aligned}$$

Remarks on the vectorization operation

- **To describe the real-/complex-valued random matrix $\mathbf{X}$ is to describe the real-/complex-valued random vector $\mathrm{vec}(\mathbf{X})$.**
- For real-valued random matrix: joint PDF of entries of $\mathrm{vec}(\mathbf{X})$.
- For complex-valued random matrix: two ways.
  - Vectorize first, then separate. $\widehat{\mathrm{vec}(\mathbf{X})} = \begin{bmatrix} \mathrm{vec}(\mathbf{X})_r \\ \mathrm{vec}(\mathbf{X})_s \end{bmatrix}$.

    $\mathrm{vec}(\mathbf{X})_r$ and $\mathrm{vec}(\mathbf{X})_s$: real and imaginary parts of $\mathrm{vec}(\mathbf{X})$.
  - Separate first, then vectorize. $\mathrm{vec}(\hat{\mathbf{X}}) = \mathrm{vec}\left(\begin{bmatrix} \mathbf{X}_r \\ \mathbf{X}_s \end{bmatrix}\right)$.

  Equivalent. The difference is a permutation of indices, i.e., a linear transformation with a permutation matrix.

1.2. Circularly symmetric complex Gaussian (CSCG) matrix

- Real-valued Gaussian random variable
- Real-valued Gaussian random vector and random matrix
- Complex-valued Gaussian random vector and **CSCG random vector**
- **CSCG random matrix**

Real-valued Gaussian random variable

Gaussian/normal distribution: $X \sim \mathcal{N}(m, \sigma^2)$.
m: mean, σ^2: variance.

PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x-m)^2}{2\sigma^2}}.$$

Standard Gaussian distribution: $\mathcal{N}(0, 1)$.

Q-function: If $X \sim \mathcal{N}(0, 1)$,

$$Q(x) \triangleq \mathbb{P}\left(X > x\right) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Real-valued Gaussian random vector and random matrix

Random vector: $\mathbf{x} = \begin{bmatrix} X_1 \\ \vdots \\ X_K \end{bmatrix}$.

Def. The random vector $\mathbf{x}$ is a **Gaussian random vector** if $X_1, X_2, \cdots, X_K$ are **jointly Gaussian**.

Joint Gaussian RVs:

- X_1, X_2 are jointly Gaussian if both Gaussian and $X_1|X_2, X_2|X_1$ are Gaussian.
- Can be generalized to more Gaussian RVs, e.g., X_1, X_2, X_3 are jointly Gaussian if each two are jointly Gaussian and $(X_1, X_2)|X_3, (X_1, X_3)|X_2, (X_2, X_3)|X_1$ are jointly Gaussian.
- Independent Gaussian RVs are jointly Gaussian.
- Linear combinations of jointly Gaussian RVs are Gaussian.

- The PDF of $\mathbf{x}$ (joint PDF of $X_1, \cdots, X_K$) is:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^K \det^{\frac{1}{2}}(\boldsymbol{\Sigma})} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})},$$

  where $\mathbf{m}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix.

- Notation: $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$.

- Several special cases.
  - i.i.d.$\sim \mathcal{N}(0,1)$

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^K} e^{-\frac{\|\mathbf{x}\|_2^2}{2}} = \frac{1}{\left(\sqrt{2\pi}\right)^K} e^{-\frac{1}{2}\sum_{i=1}^K x_i^2}$$

  - Independent only

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^K \sigma_1 \cdots \sigma_K} e^{-\sum_{i=1}^K \frac{(x_i - m_i)^2}{2\sigma_i^2}}$$

  - Zero-mean
  - Other cases ...

For a random matrix $\mathbf{X}$ which is $M \times N$, conduct vectorization to have $\mathrm{vec}(\mathbf{X})$, which is an MN-random vector.

- **Work on the $M \times N$ matrix $\mathbf{X} \Longleftrightarrow$ work on the (MN)-dimensional vector $\mathrm{vec}(\mathbf{X})$.**
- Thus the **mean vector is MN-dimensional** and the **covariance matrix is $(MN) \times (MN)$**.
- $\mathbf{X}$ is a Gaussian matrix if $\mathrm{vec}(\mathbf{X})$ is a Gaussian random vector.
- **Matrix norm distribution:** if the PDF has the following format:

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{e^{-\frac{1}{2}\mathrm{tr}[\mathbf{V}^{-1}(\mathbf{X}-\mathbf{M})^T\mathbf{U}^{-1}(\mathbf{X}-\mathbf{M})]}}{(\sqrt{2\pi})^{MN}(\sqrt{\det(\mathbf{U})})^{N}(\sqrt{\det(\mathbf{V})})^{M}},$$

for some $M \times M$ matrix $\mathbf{U}$ and $N \times N$ matrix $\mathbf{V}$.

Equivalently, $\mathrm{vec}(\mathbf{X}) \sim \mathcal{N}(\mathrm{vec}(\mathbf{M}), \mathbf{V} \otimes \mathbf{U})$.

- **If columns of $\mathbf{X}$ are i.i.d. each following $\mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$**, the mean vector of $\mathrm{vec}(\mathbf{X})$ is $\tilde{\mathbf{m}} = [\mathbf{m}^t, \cdots, \mathbf{m}^t]^t$ and the covariance matrix is $\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}\{\boldsymbol{\Sigma}, \cdots, \boldsymbol{\Sigma}\} = \mathbf{I} \otimes \boldsymbol{\Sigma}$.
  - Notice that $\mathbf{m}$ is M-dimensional and $\boldsymbol{\Sigma}$ is $M \times M$.
  - PDF:

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{X}) &= f_{\mathrm{vec}(\mathbf{X})}(\mathrm{vec}(\mathbf{X})) \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^{MN} \det^{\frac{1}{2}}(\tilde{\boldsymbol{\Sigma}})} e^{-\frac{1}{2}(\mathrm{vec}(\mathbf{x})-\tilde{\mathbf{m}})^T \tilde{\boldsymbol{\Sigma}}^{-1}(\mathrm{vec}(\mathbf{x})-\tilde{\mathbf{m}})} \\
&= \prod_{n=1}^{N} \frac{1}{\left(\sqrt{2\pi}\right)^{M} \det^{\frac{1}{2}}(\boldsymbol{\Sigma})} e^{-\frac{1}{2}(\mathbf{x}_{col,n}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_{col,n}-\mathbf{m})} \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^{MN} \det^{\frac{N}{2}}(\boldsymbol{\Sigma})} e^{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{X}-\mathbf{M})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\right]},
\end{aligned}
$$

where $\mathbf{M} = [\mathbf{m}, \cdots, \mathbf{m}]$.

– For this special case,

$$\boldsymbol{\Sigma} = \frac{1}{N}\mathbb{E}\left[(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})^T\right].$$

– For zero-mean ($\mathbf{m} = \mathbf{0}$), $\boldsymbol{\Sigma} = \frac{1}{N}\mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]$ and

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{\left(\sqrt{2\pi}\right)^{MN} \det^{\frac{N}{2}}(\boldsymbol{\Sigma})} e^{-\frac{1}{2}\mathrm{tr}\left[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]}.$$

– **Many work use the simplified notation $\mathbf{X} \sim \mathcal{CN}(\mathbf{m}, \boldsymbol{\Sigma})$ for this case.**

  * Cannot use this for the general case.
  * Understand what it really means.
  * Rigorously speaking, $\boldsymbol{\Sigma}$ is not the covariance matrix of $\mathbf{X}$.
  * For the general case, the right-hand-side of the top formula is the average covariance matrix of the columns of $\mathbf{X}$.

- **If rows of $\mathbf{X}$ are i.i.d. each following $\mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$,**
  - Notice that $\mathbf{m}$ is N-dimensional (column vector) and $\boldsymbol{\Sigma}$ is $N \times N$.
  - PDF:
  $$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{\left(\sqrt{2\pi}\right)^{MN} \det^{\frac{M}{2}}(\boldsymbol{\Sigma})} e^{-\frac{1}{2}\mathrm{tr}\left[(\mathbf{X}-\mathbf{M}^T)\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mathbf{M}^T)^T\right]}.$$
  - For zero-mean,
  $$\boldsymbol{\Sigma} = \frac{1}{M}\mathbb{E}\left[\mathbf{X}^T\mathbf{X}\right]$$
  and
  $$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{\left(\sqrt{2\pi}\right)^{MN} \det^{\frac{M}{2}}(\boldsymbol{\Sigma})} e^{-\frac{1}{2}\mathrm{tr}\left[\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X}^T\right]}.$$
  - Difference to the i.i.d. column case is a $(\cdot)^T$-operation.
  - Be careful with the position of $(\cdot)^T$ and understand why.

Complex-valued Gaussian random vector and circularly symmetric complex Gaussian vector

Def. A K-dimensional complex-valued random vector $\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_s$ is a **complex Gaussian random vector** if $\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_r \\ \mathbf{x}_s \end{bmatrix}$ is a real-valued Gaussian random vector.

Def. The complex-valued Gaussian random vector $\mathbf{x}$ is **circularly symmetric** if

$$\boldsymbol{\Sigma}_{\hat{\mathbf{x}}} = \mathbb{E}\{(\hat{\mathbf{x}} - \mathbb{E}[\hat{\mathbf{x}}])(\hat{\mathbf{x}} - \mathbb{E}[\hat{\mathbf{x}}])^T\} = \frac{1}{2}\begin{bmatrix} \mathrm{Re}\{\mathbf{Q}\} & -\mathrm{Im}\{\mathbf{Q}\} \\ \mathrm{Im}\{\mathbf{Q}\} & \mathrm{Re}\{\mathbf{Q}\} \end{bmatrix}$$

for some $K \times K$ (Hermitian and) positive-semi-definite matrix $\mathbf{Q}$, i.e., $\mathbf{Q} \succeq 0$.

- Notation for $\mathbf{x}$ being CSCG: $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{Q})$, where $\mathbf{m}$ is the mean vector and $\mathbf{Q}$ is the covariance matrix, both are complex in general.
- *Notation:* $\hat{\hat{\mathbf{Q}}} = \begin{bmatrix} \mathrm{Re}\{\mathbf{Q}\} & -\mathrm{Im}\{\mathbf{Q}\} \\ \mathrm{Im}\{\mathbf{Q}\} & \mathrm{Re}\{\mathbf{Q}\} \end{bmatrix}$.
- Real part and imaginary part must be jointly Gaussian.
- Real part and imaginary part have the same covariance matrix.
- Since $\hat{\hat{\mathbf{Q}}} \succeq \mathbf{0}$ implies Hermitian, $\mathrm{Re}\{\mathbf{Q}\}$ must be symmetric and $\mathrm{Im}\{\mathbf{Q}\}$ must be anti-symmetric (skew-symmetric).

Before more details on general CSCG random vector, consider a 1-dimensional random variable: $X \sim \mathcal{CN}(m, Q)$.

- Q is a non-negative real number.
- X_r and X_s are independent and have the same variance, which equals $Q/2$.
- $X \sim \mathcal{CN}(0, 1)$ means that the real part and imaginary part of X are i.i.d.$\sim \mathcal{N}(0, 1/2)$.
- PDF of $X \sim \mathcal{CN}(m, Q)$:

$$f_X(x) = \frac{1}{\pi Q} e^{-\frac{|x-m|^2}{Q}}.$$

Back to the general **K-dimensional CSCG random vector**: $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{Q})$.

- PDF:
$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\pi^K \det(\mathbf{Q})} e^{-(\mathbf{x}-\mathbf{m})^H \mathbf{Q}^{-1} (\mathbf{x}-\mathbf{m})}.$$
Can be derived from the PDF of the real-valued case.
- Some identities related to the mappings: $\mathbf{x} \to \hat{\mathbf{x}}$ and $\mathbf{Q} \to \hat{\hat{\mathbf{Q}}}$.
  - $\mathbf{C} = \mathbf{A}\mathbf{B} \;\Leftrightarrow\; \hat{\hat{\mathbf{C}}} = \hat{\hat{\mathbf{A}}}\hat{\hat{\mathbf{B}}}$.
  - $\mathbf{C} = \mathbf{A}^{-1} \;\Leftrightarrow\; \hat{\hat{\mathbf{C}}} = \hat{\hat{\mathbf{A}}}^{-1}$.
  - $\det(\hat{\hat{\mathbf{A}}}) = \det(\mathbf{A}\mathbf{A}^H) = |\det(\mathbf{A})|^2$.
  - $\mathbf{y} = \mathbf{A}\mathbf{x} \;\Leftrightarrow\; \hat{\mathbf{y}} = \hat{\hat{\mathbf{A}}}\hat{\mathbf{x}}$.
  - $\mathrm{Re}(\mathbf{x}^H \mathbf{y}) = \hat{\mathbf{x}}^H \hat{\mathbf{y}}$.
  - $\mathbf{Q} \succeq 0 \;\Leftrightarrow\; \hat{\hat{\mathbf{Q}}} \succeq \mathbf{0}$.
  - $\mathbf{U}$ is unitary if and only if $\hat{\hat{\mathbf{U}}}$ is orthonormal.
- For zero-mean, i.e., $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$, the PDF is
$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\pi^K \det(\mathbf{Q})} e^{-\mathbf{x}^H \mathbf{Q}^{-1} \mathbf{x}}.$$

Lemma. If $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$ for any $\mathbf{A}$, then $\mathbf{y} = \mathbf{Ax} + \mathbf{b} \sim \mathcal{CN}(\mathbf{Am} + \mathbf{b}, \mathbf{A\Sigma A}^H)$: Any linear transformation (affine function) of a CSCG random vector is also a CSCG random vector.

Lemma. If $\mathbf{x}$ and $\mathbf{y}$ are independent CSCG random vectors, then $\mathbf{z} = \mathbf{x} + \mathbf{y}$ is CSCG.

Lemma. If $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{\Phi}$ is a random unitary matrix that is independent of $\mathbf{x}$, then $\mathbf{y} = \mathbf{\Phi x}$ also follows $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ and is independent of $\mathbf{\Phi}$. $\mathbf{x}$ and $\mathbf{y}$ are equivalent in distribution.

Lemma. Let $\mathbf{x}$ be a zero-mean complex-valued random vector, and $\mathbb{E}[\mathbf{xx}^H] = \mathbf{Q}$, then the entropy of $\mathbf{x}$ satisfies

$$H(\mathbf{x}) \leq \log\det(\pi e \mathbf{Q})$$

with equality if and only if $\mathbf{x}$ is CSCG, i.e., $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q})$.

Circularly symmetric complex Gaussian random matrix

- By following previous slides, a complex-valued random matrix $\mathbf{X}$ is Gaussian if $\mathrm{vec}(\mathbf{X})$ is a complex-valued Gaussian random vector.
- A complex-valued random matrix **$\mathbf{X}$ is CSCG if $\mathrm{vec}(\mathbf{X})$ is a CSCG random vector.**
- Fundamentally, **work on** $\mathrm{vec}(\mathbf{X})$ using previous definitions and results.

$$f_{\mathrm{vec}(\mathbf{X})}(\mathrm{vec}(\mathbf{X})) = \frac{1}{\pi^K \det(\mathbf{Q})} e^{-(\mathrm{vec}(\mathbf{X})-\mathbf{m})^H \mathbf{Q}^{-1} (\mathrm{vec}(\mathbf{X})-\mathbf{m})},$$

where $\mathbf{m}$ is (MN)-dimensional and $\mathbf{Q}$ is $(MN) \times (MN)$.

- If columns (or rows) of $\mathbf{X}$ $(M \times N)$ are i.i.d.$\sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, meaning that all entries of $\mathbf{X}$ are i.i.d.$\sim \mathcal{CN}(0, 1)$, $\mathbf{X}$ is a CSCG matrix and its PDF is

$$f_{\mathrm{vec}(\mathbf{X})}(\mathrm{vec}(\mathbf{X})) = \frac{1}{\pi^{MN}} e^{-\mathrm{vec}(\mathbf{X})^H \mathrm{vec}(\mathbf{X})}$$

  or equivalently, $f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{\pi^{MN}} e^{-\|\mathbf{X}\|_F^2} = \frac{1}{\pi^{MN}} e^{-\mathrm{tr}(\mathbf{X}^H \mathbf{X})}$,

  where $\|\mathbf{X}\|_F$ is the Frobenius norm of $\mathbf{X}$.

  – Many work use **the simplified notation for this case**: $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}_{M \times N}, \mathbf{I}_{M \times M})$.

  – It **implies that columns are independent and each column has the same covariance matrix $\mathbf{I}_{M \times M}$**.

  – **Not true for the general case.** May cause confusion or mistake if taken for granted.

- If columns of $\mathbf{X}$ $(M \times N)$ are independent and CSCG, i.e., $\mathbf{x}_{col,n} \sim \mathcal{CN}(\mathbf{m}_n, \mathbf{Q}_n)$, then $\mathbf{X}$ is a **CSCG matrix** and its PDF is
$$f_{\mathbf{X}}(\mathbf{X}) = \frac{e^{-\sum_{n=1}^{N}(\mathbf{x}_{col,i}-\mathbf{m}_n)^H \mathbf{Q}_n^{-1}(\mathbf{x}_{col,i}-\mathbf{m}_n)}}{\pi^{MN}\det(\mathbf{Q}_1)\cdots\det(\mathbf{Q}_N)}$$
$$= \frac{e^{-(\mathrm{vec}(\mathbf{X})-\mathbf{m}_{full})^H \mathbf{Q}_{full}^{-1}(\mathrm{vec}(\mathbf{X})-\mathbf{m}_{full})}}{\pi^{MN}\det(\mathbf{Q}_{full})},$$
where $\mathbf{m}_{full} = [\mathbf{m}_1^T, \cdots, \mathbf{m}_N^T]^T$, $\mathbf{Q}_{full} = \mathrm{diag}\{\mathbf{Q}_1, \cdots, \mathbf{Q}_N\}$.
- The case with independent row vectors can be analyzed similarly.
- **Matrix norm distribution for CSCG case:** if the PDF has the following format:
$$f_{\mathbf{X}}(\mathbf{X}) = \frac{e^{-\mathrm{tr}[\mathbf{V}^{-1}(\mathbf{X}-\mathbf{M})^H \mathbf{U}^{-1}(\mathbf{X}-\mathbf{M})]}}{(\pi)^{MN}\det^N(\mathbf{U})\det^M(\mathbf{V})},$$
for some $M \times M$ matrix $\mathbf{U}$ and $N \times N$ matrix $\mathbf{V}$.
Equivalently, $\mathrm{vec}(\mathbf{X}) \sim \mathcal{CN}(\mathrm{vec}(\mathbf{M}), \mathbf{V} \otimes \mathbf{U})$.

1.3. Isotropic distribution, decomposition of CSCG random matrix and Wishart matrix and related PDFs

- Isotropic distribution and decomposition of CSCG random matrix
- Wishart matrix and related PDFs

Isotropic distribution and decomposition of CSCG random matrix

Def. An n-dimensional random complex unit vector $\mathbf{u}$ is **isotropically distributed** if its probability density is invariant to all unitary transformations. That is,

$$f_{\mathbf{u}}(\mathbf{u}) = f_{\mathbf{u}}(\mathbf{\Phi u}) \text{ for any } \mathbf{\Phi}^H\mathbf{\Phi} = \mathbf{I}.$$

- **Uniform distribution on the set of unit vectors.**
- The PDF **depends on the magnitude (length) only**, not direction.
- Elements of $\mathbf{u}$ are dependent.

The **PDF of an isotropically distributed** unit vector $\mathbf{u}$ is

$$f(\mathbf{u}) = \frac{\Gamma(n)}{\pi^N}\delta(\mathbf{u}^H\mathbf{u} - 1).$$

The **PDF of any L-elements** of $\mathbf{u}$:

$$f(\mathbf{u}^{(L)}) = \frac{\Gamma(n)}{\pi^L\Gamma(n-L)}\delta(1-(\mathbf{u}^{(L)})^H\mathbf{u}^{(L)}), \quad \text{norm of each element} \leq 1.$$

$\mathbf{u}^{(L)}$ is a vector of any L elements of $\mathbf{u}$.

Def. An $n \times n$ unitary matrix $\mathbf{U}$ is **isotropically distributed** if its probability density is invariant when left-multiplied by any deterministic unitary matrix, that is,

$$f_{\mathbf{U}}(\mathbf{U}) = f_{\mathbf{U}}(\mathbf{\Phi U}) \text{ for any } \mathbf{\Phi}^H \mathbf{\Phi} = \mathbf{I}.$$

- **Uniform distribution on the set of unitary matrices**.
- **Same density on all "directions"**.
- PDF is also invariant when right-multiplied by unitary matrix.
- The transpose and conjugate of $\mathbf{U}$ are also isotropically distributed.
- Any column of $\mathbf{U}$ is a random complex unit vector.
- Columns of $\mathbf{U}$ are dependent.

The PDF of an isotropically distributed unitary matrix $\mathbf{U}$ is

$$f(\mathbf{U}) = \frac{\prod_{i=1}^{n} \Gamma(i)}{\pi^{n(n-1)/2}} \delta(\mathbf{U}^H \mathbf{U} - \mathbf{I}).$$

There are general results on moments of elements of $\mathbf{U}$.

Theorem: Let $\mathbf{X}$ be an $m \times n$ standard CSCG matrix, i.e., $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, where $m \geq n$. Let $\mathbf{X} = \mathbf{\Phi R}$ be the QR-decomposition normalized so that the diagonal elements of $\mathbf{R}$ are positive. Thus

- $\mathbf{\Phi}$ is an isotropically distributed unitary matrix.
- Elements of $\mathbf{R}$ are independent of each other.
- $\mathbf{R}$ is independent of $\mathbf{\Phi}$.
- The upper diagonal elements of $\mathbf{R}$ are $\mathcal{CN}(0, 1)$.
- The ith diagonal element of $\mathbf{R}$ is a half of a χ^2 random variable with $2(m - i + 1)$ degrees of freedom: $2r_{ii} \sim \chi^2_{2(m-i+1)}$.
- The case of $m < n$ can be considered similarly.

Theorem: Let $\mathbf{X}$ be an $m \times n$ standard CSCG matrix, i.e., $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Let $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ be the singular value decomposition (SVD). Thus

- $\mathbf{U}$ and $\mathbf{V}$ are isotropically distributed unitary matrices.
- $\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}$ are independent.
- See later parts on the distributions of elements of $\mathbf{\Sigma}$.

Theorem: Let $\mathbf{X}$ be an $m \times n$ standard CSCG matrix, i.e., $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, where $n \geq m$. Then $\mathbf{X}$ is unitarily similar to an $m \times n$ matrix:

$$\frac{1}{2}\begin{bmatrix} x_{2n} & & & & 0 & \cdots & 0 \\ y_{2(m-1)} & x_{2(n-1)} & & & 0 & & 0 \\ & \ddots & \ddots & & \vdots & & \vdots \\ & & y_2 & x_{2(n-(m-1))} & 0 & \cdots & 0 \end{bmatrix},$$

where x_i^2 and y_i^2 are independent and follows χ^2 distribution with i degrees of freedom.

Complex Wishart matrix and related PDFs

Def. Let $\mathbf{X}$ be an $m \times n$ ($m \geq n$) random matrix where each row is a zero-mean CSCG random vector following $\mathcal{CN}(\mathbf{0}, \mathbf{V})$ and the rows are independent. The $n \times n$ matrix

$$\mathbf{W} = \mathbf{X}^H \mathbf{X} = \sum_{i=1}^{m} \mathbf{x}_{row,i}^H \mathbf{x}_{row,i}$$

is a **(centralized) Wishart matrix**. The probability distribution of $\mathbf{W}$ is called the (centralized) Wishart distribution, denoted as $\mathcal{CW}_n(\mathbf{V}, m)$.

- PDF of complex Wishart matrix:

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{\det^{m-n}(\mathbf{W})}{\pi^{\frac{n(n-1)}{2}} \Gamma(m) \cdots \Gamma(m-n+1) \det^{m}(\mathbf{V})} e^{-\mathrm{tr}(\mathbf{V}^{-1}\mathbf{W})}.$$

- Similar to $m < n$ and column independent CSCG random matrix.

Consider the special case of $\mathbf{V} = \mathbf{I}$, i.e., $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Define

$$\mathbf{W} = \begin{cases} \mathbf{X}\mathbf{X}^H & m < n \\ \mathbf{X}^H\mathbf{X} & m \geq n \end{cases}$$

and $N = \max\{m, n\}, M = \min\{m, n\}$. The probability distribution of $\mathbf{W}$ is called Wishat distribution with parameters m and n (notice that $N \geq M$ and $\mathbf{W}$ is $M \times M$).

- PDF of $\mathbf{W}$:

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{\det^{N-M}(\mathbf{W})}{\pi^{\frac{M(M-1)}{2}}\Gamma(N) \cdots \Gamma(N-M+1)} e^{-\mathrm{tr}(\mathbf{W})}.$$

- **Joint PDF of ordered eigenvalues** $\lambda_1 \geq \cdots \geq \lambda_M \geq 0$:

$$f(\lambda_1, \cdots, \lambda_M) = C(M, N) \cdot e^{-\sum_{i=1}^{M} \lambda_i} \prod_{i=1}^{M} \lambda_i^{N-M} \prod_{i<j} (\lambda_i - \lambda_j)^2$$

for $\lambda_1 \geq \cdots \geq \lambda_M \geq 0$, where C is a constant depends on M and N only.

- **Joint PDF of unordered eigenvalues**

$$f(\lambda_1, \cdots, \lambda_M) = \frac{C(M,N)}{M!} \cdot e^{-\sum_{i=1}^{M} \lambda_i} \prod_{i=1}^{M} \lambda_i^{N-M} \prod_{i \neq j} (\lambda_i - \lambda_j)^2$$

for $\lambda_1, \cdots, \lambda_M \geq 0$.

- **Marginal PDF of an unordered eigenvalue**

$$f_\lambda = \frac{1}{M} \sum_{i=1}^{M} \frac{(i-1)!}{(i-1+N-M)!} [L_{i-1}^{N-M}(\lambda)]^2 \lambda_1^{N-M} e^{-\lambda_1}, \quad \lambda \geq 0,$$

where $L_k^{N-M}(x) = \frac{1}{k!} e^x x^{M-N} \frac{d^k}{dx^k} (e^{-x} x^{N-M+k})$ is the Laguerre polynomial of order k.

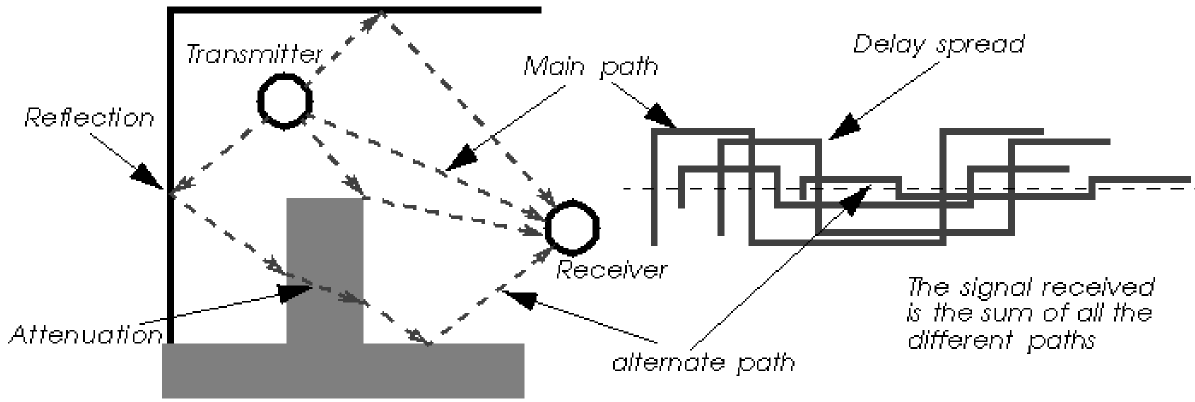
- Results on PDFs of the maximum and minimum eigenvalues.
- **Inverse Wishart distribution: Y** follows inverse Wishart distribution if its inverse follows Wishart distribution.
- **Non-centralized Wishart matrix:** when **H** has non-zero mean.

Part 2. MIMO applications

- MIMO channel model
- MIMO capacity
- Diversity analysis of distributed space-time coding with multiple antennas
- Performance analysis of massive MIMO with ZF

MIMO channel model

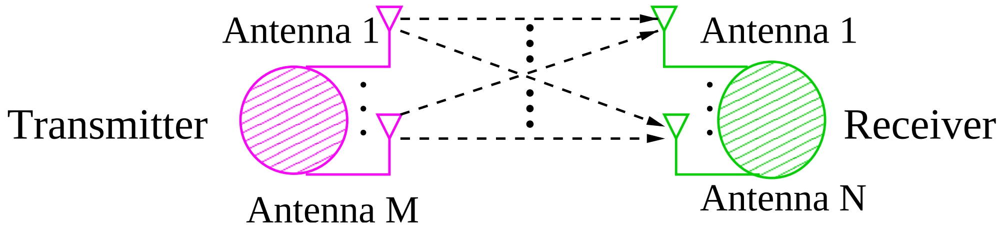
- A wireless link: Multi-path, delay spread, mobility, etc.



- Frequency-flat fading channel: The delay spread in the channel is negligible compared to symbol interval. The coherence bandwidth of the channel is much bigger than the signal bandwidth. Therefore, all frequency components of the signal will experience the same magnitude of fading.

- Frequency-selective fading channel (the counterpart).

MIMO system with M transmit antennas and N receive antennas:



- At a given time/transmission, for frequency-flat fading over the bandwidth of interest, the channel can be written as a matrix:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & \cdots & h_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \cdots & h_{NM} \end{bmatrix},$$

  where h_{nm} is the channel gain from the m-th TX antenna to the n-th RX antenna.

- Each h_{nm} is a complex value (quadrature-carrier multiplexing).
- Affected by multi-path fading, path-loss, and shadowing.

- **i.i.d. Rayleigh fading model**
  - The number of scatters is large, all scattered contributions are non-coherent and approximately equal energy, (via central limit theorem)
  - Indoor and no line-of-sight
  - Enough spacing between antennas (e.g., half-wavelength or bigger) for independent entries
  - Each channel coefficient is modeled as a circularly symmetric complex Gaussian random variable with zero mean.
  - The magnitude of each channel entry (channel gain) follows Rayleigh PDF. The magnitude-square follows exponential PDF.

– Channel variance σ_h^2 depends on the large-scale fading (e.g., distance), often normalized as 1.

$$h_{nm} \sim \mathcal{CN}(0, \sigma_h^2) \text{ and i.i.d.}$$
$$\mathrm{Re}(h_{nm}), \mathrm{Im}(h_{nm}) \sim \mathcal{N}(0, \sigma_h^2/2) \text{ independent}$$
$$f_{|h_{nm}|}(x) = \frac{2x}{\sigma_h^2} e^{-\frac{x^2}{\sigma_h^2}}, \text{ for } x \geq 0.$$
$$f_{|h_{nm}|^2}(x) = \frac{1}{\sigma_h^2} e^{-\frac{x}{\sigma_h^2}}, \text{ for } x \geq 0.$$

– $\mathbf{H}$ is CSCG: $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2 \mathbf{I})$, which is $N \times M$.

- **Correlated Rayleigh channel model**
  - $\mathbf{H}$ $(N \times M)$ is CSCG with zero-mean.
  - Correlation matrix of $\mathbf{H}$:

    $$\mathbf{R}_{\mathbf{H}} = \mathbb{E}\left\{\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H\right\},$$

    which is $MN \times MN$.
  - Another representation:

    $$\mathrm{vec}(\mathbf{H}) = \mathbf{R}_{\mathbf{H}}^{1/2}\mathrm{vec}(\tilde{\mathbf{H}}), \text{ where } \mathrm{vec}(\tilde{\mathbf{H}}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{MN}).$$

  - Transmit correlation matrix and receive correlation matrix:

    $$\mathbf{R}_{t,\mathbf{H}} = \frac{1}{N}\mathbb{E}\left\{\mathbf{H}^H\mathbf{H}\right\}, \quad \mathbf{R}_{r,\mathbf{H}} = \frac{1}{M}\mathbb{E}\left\{\mathbf{H}\mathbf{H}^H\right\}.$$

– Kronecker model: when
  1) transmit correlation is independent from the receive antennas and vice versa; and
  2) cross-correlation equals to the Kronecker product of corresponding transmit and receive correlations,

$$\mathbf{R}_{\mathbf{H}} = \mathbf{R}_{r,\mathbf{H}} \otimes \mathbf{R}_{t,\mathbf{H}}.$$

  In this case, we have $\mathbf{H} = \mathbf{R}_{r,\mathbf{H}}^{1/2}\ \tilde{\mathbf{H}}\ \mathbf{R}_{t,\mathbf{H}}^{1/2}$ where $\tilde{\mathbf{H}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

– See previous slides for the PDF.

MIMO Capacity for CSI at the receiver only

System description.

- Point-to-point multiple-antenna system with M transmit antennas and N receive antennas, with channel matrix $\mathbf{H}$.
- MIMO transceiver equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},$$

  - $\mathbf{x}$: $(M \times 1)$ contains signals sent by M transmit antennas. Its mth entry is the signal send by the mth antenna.
  - $\mathbb{E}[\mathbf{x}^H \mathbf{x}] \leq P$ and P is the maximum transmit power.
  - $\mathbf{x}$: $(N \times 1)$ contains received signals at the N receive antennas. Its nth entry is the signal received at the nth antenna.
  - $\mathbf{n}$: noise vector. Assume $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma_n})$ independent of $\mathbf{H}$ and $\mathbf{x}$.

Problem: To analyze the capacity when the receiver knows $\mathbf{H}$.

- Capacity: $C = \max_{f_{\mathbf{x}}(\mathbf{x})} I(\mathbf{x};\mathbf{y}) = \max_{f_{\mathbf{x}}(\mathbf{x})}[H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})]$.

Sketch of method and result.

- $H(\mathbf{y}|\mathbf{x}) = H(\mathbf{Hx} + \mathbf{n}|\mathbf{x}) = H(\mathbf{n}|\mathbf{x}) = H(\mathbf{n})$. From previous lemma, $H(\mathbf{n}) = \log\det(\pi e \mathbf{\Sigma_n})$.
- From the transceiver equation, $\mathbf{\Sigma_y} = \mathbf{H\Sigma_x H}^H + \mathbf{\Sigma_n}$.
- Zero-mean $\mathbf{x}$ saves power and does not hurt the capacity, so assume $\mathbb{E}[\mathbf{x}] = \mathbf{0}$. Therefore, $\mathbb{E}[\mathbf{y}] = \mathbf{0}$.

$$\mathrm{tr}\mathbf{\Sigma_x} = \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])^H(\mathbf{x} - \mathbb{E}[\mathbf{x}])] \le \mathbb{E}\,(\mathbf{x}^H\mathbf{x}).$$

- For any $\mathbf{\Sigma_x}$ where $\mathrm{tr}\mathbf{\Sigma_x} \le P$, from a previous lemma, $\max H(\mathbf{y}) = \log\det(\pi e \mathbf{\Sigma_y}) = \log\det[\pi e(\mathbf{H\Sigma_x H}^H + \mathbf{\Sigma_n})]$ achieved when $\mathbf{y}$ is CSCG, i.e., when $\mathbf{x}$ is CSCG.
- Thus,

$$\begin{aligned} C_{\mathrm{MIMO},\mathbf{H}} &= \max_{\mathrm{tr}\mathbf{\Sigma_x} \le P} \mathbb{E}\,\log\det(\mathbf{I}_N + \mathbf{\Sigma_n}^{-1}\mathbf{H\Sigma_x H}^H) \\ &= \max_{\mathbf{\Sigma_x}\text{ is diagonal, }\mathrm{tr}\mathbf{\Sigma_x} \le P} \mathbb{E}\,\log\det(\mathbf{I}_N + \mathbf{\Sigma_n}^{-1}\mathbf{H\Sigma_x H}^H), \end{aligned}$$

- For any permutation matrix $\mathbf{\Pi}$, $\mathbf{H\Pi}$ has the same distribution as $\mathbf{H}$ and $\log\det(\mathbf{X})$ is concave for $\mathbf{X} \succeq 0$. Thus

$$\begin{aligned}
&\mathbb{E}\log\det(\mathbf{I}_N + \mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{H}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{H}^H) \\
&= \frac{1}{M!}\mathbb{E}\sum_{\mathbf{\Pi}}\log\det(\mathbf{I}_N + \mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{H\Pi}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{\Pi}^H\mathbf{H}^H) \\
&\leq \mathbb{E}\log\det\left\{\mathbf{I}_N + \mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{H}\left[\frac{1}{M!}\sum_{\mathbf{\Pi}}\mathbf{\Pi}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{\Pi}^H\right]\mathbf{H}^H\right\} \\
&= \mathbb{E}\log\det\left(\mathbf{I}_N + \frac{\mathrm{tr}\mathbf{\Sigma}_{\mathbf{x}}}{M}\mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{H}\mathbf{H}^H\right)
\end{aligned}$$

with equality when $\mathbf{\Sigma}_{\mathbf{x}} = (\mathrm{tr}\mathbf{\Sigma}_{\mathbf{x}}/M)I_M$. Thus,

$$C_{\mathrm{MIMO},\mathbf{H}} = \mathbb{E}\log\det\left(\mathbf{I}_N + \frac{P}{M}\mathbf{\Sigma}_{\mathbf{n}}^{-1}\mathbf{H}\mathbf{H}^H\right)$$

with equality when $\mathbf{\Sigma}_{\mathbf{x}} = (P/M)\mathbf{I}_M$.

Special/asymptotic cases and discussions.

- If i.i.d. noises, each follows $\mathcal{CN}(0, \sigma^2)$,

$$C_{\mathrm{MIMO},\mathbf{H}} = \mathbb{E} \log\det\left(\mathbf{I}_N + \frac{P}{M\sigma^2}\mathbf{H}\mathbf{H}^H\right)$$
$$= \mathbb{E} \log\det\left(\mathbf{I}_N + \frac{\rho}{M}\mathbf{H}\mathbf{H}^H\right).$$

- When no CSI at the TX
  - No reason to transmit more energy on one antenna than another; thus, same average energy/power across antennas.
  - No reason for correlation or dependence between transmit signals of different antennas.
- When N is fixed and $M \to \infty$,

$$\frac{1}{M}\mathbf{H}\mathbf{H}^H \to \mathbf{I}_N, \quad C \to N\log(1+\rho).$$

- When M fixed and $N \to \infty$,

$$\frac{1}{N}\mathbf{H}^H\mathbf{H} \to \mathbf{I}_M, \quad C \approx M\log(1+\rho N/M).$$

- General case:

$$C_{\text{MIMO},\mathbf{H}} = \sum_{k=1}^{K} \mathbb{E}_{\lambda_k} \log\left(1 + \frac{\rho}{M}\lambda_k\right) = K\mathbb{E}_{\lambda} \log\left(1 + \frac{\rho}{M}\lambda\right),$$

where

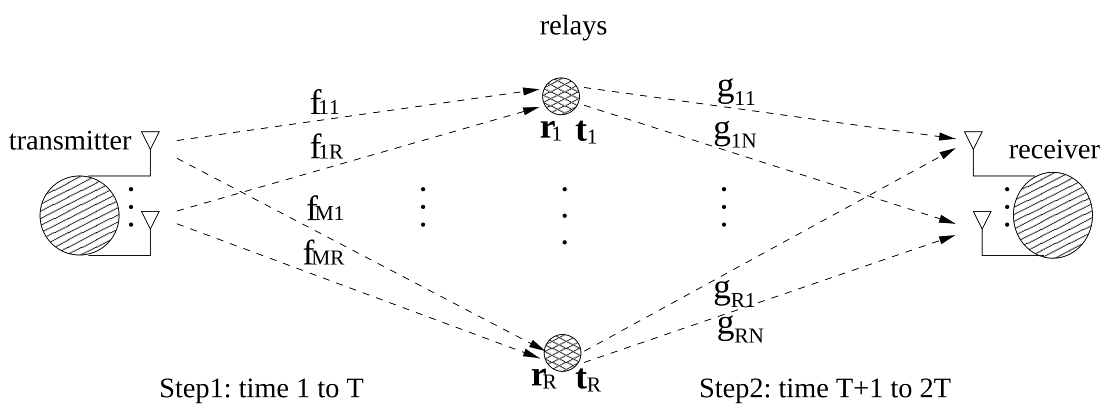
  – $\lambda_1, \cdots, \lambda_K$ are eigenvalues of $\mathbf{H}\mathbf{H}^H$,

  – λ represents one unordered eigenvalue. $K = \text{rank}(\mathbf{H})$,

  – Use eigen-distributions in previous slides for further calculations.

Performance analysis of distributed space-time coding for MIMO relay networks

System description.



- One transmitter with M transmit antennas
- One receiver with N receive antennas
- R relay each with single transmit and receive antenna
- Block-fading channels and CSI at the receiver only.
- Transmitter-relay channels: f_{mr}'s. Relay-receiver channels: g_{rn}'s.

Transceiver protocol and model.

- Two-step transmission each contains T time slots
  Step 1. Transmitter $\rightarrow$ relays. Step 2. Relays $\rightarrow$ receiver.
- Relay process: distributed space-time coding (DSTC).
  Linear transformation on its received singals then transmit:

$$\mathbf{t}_i = \sqrt{\frac{P_2}{P_1 + 1}} \mathbf{A}_i \mathbf{r}_i,$$

  - $\mathbf{r}_i$: received vector at relay in Step 1.
  - $\mathbf{t}_i$: transmit vector from relay in Step 2.
  - $\mathbf{A}_i$: a pre-determined (unitary) $T \times T$ matrix.
  - P_1 transmit power of Step 1.
  - P_2 transmit power per relay of Step 2.

- End-to-end transceiver equation:

$$\mathbf{X} = \sqrt{\beta}\mathbf{S}\mathbf{H} + \mathbf{W},$$

  - $\mathbf{X}$: $T \times N$ received signal matrix
  - $\beta \triangleq \alpha \frac{P_1 T}{M}$ where $\alpha \triangleq \frac{P_2}{P_1+1}$.
  - $\mathbf{S} \triangleq \left[\begin{array}{ccc} \mathbf{A}_1\mathbf{s} & \cdots & \mathbf{A}_R\mathbf{s} \end{array} \right]$ is the distributed space-time code depending on $\mathbf{A}_i$'s and information vector
  - $\mathbf{H} \triangleq \left[\begin{array}{ccc} (\mathbf{f}_1\mathbf{g}_1)^t & \cdots & (\mathbf{f}_R\mathbf{g}_R)^t \end{array} \right]^t$ is the $RM \times N$ equivalent channel matrix
  - $\mathbf{W} \triangleq \sqrt{\alpha}\left[\begin{array}{ccc} \sum_{i=1}^R g_{i1}\mathbf{A}_i\mathbf{v}_i & \cdots & \sum_{i=1}^R g_{iN}\mathbf{A}_i\mathbf{v}_i \end{array} \right] + \mathbf{w}$ is the equivalent noise term.
    $\mathbf{v}_i$ is noise vector at Relay i and $\mathbf{w}$ is the noise at the receiver.

Result 1. Define

$$\mathbf{R}_W = \mathbf{I} + \alpha \mathbf{G}^H \mathbf{G}.$$

Given that $\mathbf{s}_k$ is transmitted and the corresponding distributed space-time code is $\mathbf{S}_k$, the rows of $\mathbf{X}$ are independently each is CSCG distributed with the same variance $\mathbf{R}_W^t$.

The conditional PDF of $\mathbf{X}|\mathbf{S}_k$ is

$$f(\mathbf{X}|\mathbf{S}_k) = \frac{1}{\left(\pi^N \det \mathbf{R}_W\right)^{-T}} e^{-\operatorname{tr}\left(\mathbf{X}-\sqrt{\beta}\mathbf{S}_k\mathbf{H}\right)\mathbf{R}_W^{-1}\left(X-\sqrt{\beta}\mathbf{S}_k\mathbf{H}\right)^H}.$$

Sketch of Proof:

- With known CSI, since $\mathbf{v}_i$'s and $\mathbf{w}$ are independent CSCG, $\mathbf{X}$ is a linear combination of CSCG random matrices and constants, thus also a CSCG random matrix.
- Straightforward to see that $\mathbb{E}\left(\mathbf{X}|\mathbf{S}_k\right) = \sqrt{\beta}\mathbf{S}_k\mathbf{H}$.
- The rows of $\mathbf{X}$ are independent. (The columns are not.)

$$x_{tn} = \sqrt{\beta}[\mathbf{S}_k\mathbf{H}]_{tn} + \sqrt{\alpha}\sum_{i=1}^{R}\sum_{\tau=1}^{T} g_{in}a_{i,t\tau}v_{i\tau} + w_{tn}.$$

$$\mathrm{Cov}\left(x_{t_1n_1}, x_{t_2n_2}\right) = \delta_{t_1t_2}\left(\beta\begin{bmatrix} g_{1n_1} & \cdots & g_{Rn_1}\end{bmatrix}\begin{bmatrix}\bar{g}_{1n_2} \\ \vdots \\ \bar{g}_{Rn_2}\end{bmatrix} + \delta_{n_1n_2}\right).$$

By combining the results in matrix form, the covariance matrix of each row is $\mathbf{I}_N + \beta\mathbf{G}^t\bar{\mathbf{G}} = \mathbf{R}_W^t$.

- Therefore, the PDF of the ith row is

$$f([\mathbf{X}]_i|\mathbf{S}_k) = \left(\pi^N \det \mathbf{R}_W^t\right)^{-T} e^{-\mathrm{tr}\,\overline{\left[\mathbf{X}-\sqrt{\alpha}\mathbf{S}_k\mathbf{H}\right]_i}\mathbf{R}_W^{-t}\left[\mathbf{X}-\sqrt{\alpha}\mathbf{S}_k\mathbf{H}\right]_i^t}$$
$$= \left(\pi^N \det \mathbf{R}_W\right)^{-T} e^{-\mathrm{tr}\left[\mathbf{X}-\sqrt{\alpha}\mathbf{S}_k\mathbf{H}\right]_i\mathbf{R}_W^{-1}\left[\mathbf{X}-\sqrt{\alpha}\mathbf{S}_k\mathbf{H}\right]_i^H}.$$

- With independent rows, the PDF of $\mathbf{X}$ is $f(\mathbf{X}|\mathbf{S}_k) = \prod_{i=1}^{T} f([\mathbf{X}]_i|\mathbf{S}_k)$.

Result 2. The maximum-likelihood (ML) decoding is

$$\arg\min_{\mathbf{S}} \mathrm{tr}\, \left(\mathbf{X} - \sqrt{\alpha}\mathbf{S}_k\mathbf{H}\right) \mathbf{R}_W^{-1} \left(\mathbf{X} - \sqrt{\alpha}\mathbf{S}_k\mathbf{H}\right)^H .$$

With this decoding, the pairwise error probability (PEP) of mistaking $\mathbf{S}_k$ by $\mathbf{S}_l$ has the following upper bound:

$$\mathbb{P}\left(\mathbf{S}_k \to \mathbf{S}_l\right) \quad \leq \quad \mathop{\mathbb{E}}_{\mathbf{H}} e^{-\frac{\alpha}{4}\mathrm{tr}\,[(\mathbf{S}_k-\mathbf{S}_l)^*(\mathbf{S}_k-\mathbf{S}_l)\mathbf{H}\mathbf{R}_W^{-1}\mathbf{H}^H]}.$$

Sketch of Proof:

- The ML decoding rule is straightforward to obtain from the likelihood function.
- Chernoff upper bound: for any $\lambda > 0$,

$$\mathbb{P}\left(\mathbf{S}_k \rightarrow \mathbf{S}_l\right) \leq \mathbb{E}\, e^{\lambda(\ln \mathbb{P}(\mathbf{X}|\mathbf{S}_l) - \ln \mathbb{P}(\mathbf{X}|\mathbf{S}_k))}$$

$$= \mathop{\mathbb{E}}_{\mathbf{H},\mathbf{W}} e^{-\lambda \mathrm{tr}\left[\alpha(\mathbf{S}_k - \mathbf{S}_l)^H(\mathbf{S}_k - \mathbf{S}_l)\mathbf{H}\mathbf{R}_W^{-1}\mathbf{H}^H + \sqrt{\alpha}(\mathbf{S}_k - \mathbf{S}_l)\mathbf{H}\mathbf{R}_W^{-1}\mathbf{W}^H + \sqrt{\alpha}\mathbf{W}\mathbf{R}_W^{-1}(\mathbf{S}_k - \mathbf{S}_l)\mathbf{H}^H\right]}$$

$$= \mathop{\mathbb{E}}_{\mathbf{H}} \int_{\mathbf{W}} e^{-\lambda \mathrm{tr}[\cdots]} \left(\pi^N \det \mathbf{R}_W\right)^{-T} e^{-\mathrm{tr}\left(\mathbf{W}\mathbf{R}_W^{-1}\mathbf{W}^H\right)} d\mathbf{W}.$$

- The result can be proved via making sum-of-squares for the exponent and the integration over CSCG PDF is 1.

Result 3. With i.i.d. Rayleigh fading channels and fully diverse space-time code, the diversity order of DSTC is

$$d = \begin{cases} \min\{M, N\}R & \text{if } M \neq N \\ MR\left(1 - \frac{1}{M}\frac{\log\log P}{\log P}\right) & \text{if } M = N \end{cases}.$$

Sketch of Proof:

- First bound $\mathbf{R}_W$ with either of the following:

$$\mathbf{R}_W \leq (\operatorname{tr}\mathbf{R}_W)\mathbf{I} = \left(N + \frac{P_2}{P_1+1}\sum_{n=1}^{N}\sum_{i=1}^{R}|g_{in}|^2\right)\mathbf{I}_N.$$

$$\mathbf{R}_W \leq \left(1 + \frac{P_2 R}{P_1+1}\lambda_{\max}\right)\mathbf{I}_N,$$

  where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{G}^*\mathbf{G}^*/R$, whose properties can be derived from results on Wishart matrix.

- Optimal power allocation with respect to error rate bound: $P_1 = \frac{P}{2}$ and $P_2 = \frac{P}{2R}$, where P is the total transmit power.

- Use Chernoff bound in Result 2 and calculate the average over f_{mi} to get:

$$\mathbb{P}\left(\mathbf{S}_k \to \mathbf{S}_l\right) \lesssim \mathop{\mathbb{E}}_{g_{in}} \prod_{i=1}^{R} \left(1 + \frac{PT\sigma_{\min}^2}{8MNR} \frac{g_i}{1 + \frac{1}{NR}\sum_{i=1}^{R}\sum_{n=1}^{N}|g_{in}|^2}\right)^{-M}.$$

  $\sigma_{\min}^2$: the minimum singular value of $(\mathbf{S}_k - \mathbf{S}_l)^H(\mathbf{S}_k - \mathbf{S}_l)$.

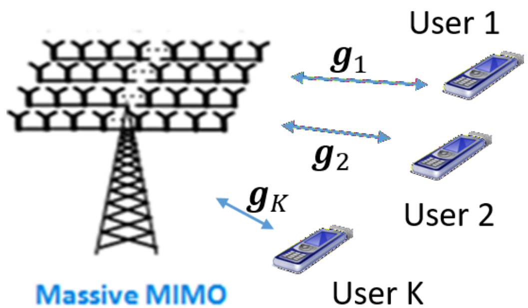
- Further conduct the calculations by splitting each integration range into $[0, x) \cup [x, \infty)$ to split the integration into 2^R ones.

- Calculate the order of each term with respect to P, and find a good/the optimal choice of x to minimize the order with respect to P.

- Refer to [4] for details.

Performance analysis of massive MIMO with ZF

Single-cell multi-user massive MIMO system model



- One base station (BS) with M antennas ($M \gg 1$)
- K single-antenna users, $M \geq K$
- Channel matrix $\mathbf{G}$ contains independent Rayleigh fading coefficients, perfect CSI at the BS.
  - Uplink: $\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}$ $(M \times K)$. The k-th column $\mathbf{g}_k$ is channel vector from BS antennas to user k.
  - Downlink: $\mathbf{G} = \mathbf{D}^{1/2}\mathbf{H}$ $(K \times M)$. The k-th row $\mathbf{g}_k$ is channel vector from User k to the BS antennas.
  - $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\mathbf{D} = \mathrm{diag}\{\beta_1, \cdots, \beta_K\}$ containing large-scale coefficients.

Models for uplink with ZF

- Users send signals to the BS.
- Transceiver equation:

$$\mathbf{y} = \sqrt{p_u}\mathbf{G}\mathbf{s} + \mathbf{w}.$$

  - $\mathbf{s} = [s_1 \cdots s_K]^T$: vector of information symbols. normalization: $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^H\right\} = \mathbf{I}$.
  - p_u: average transmit power of each user
  - $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$: noise vector
  - $\mathbf{y}$: Received vector at the BS

- Zero-forcing reception:

$$\mathbf{A} = \sqrt{\alpha}(\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H = \sqrt{\alpha}\mathbf{D}^{-1/2}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H,$$

$$\mathbf{r} = \mathbf{A}\mathbf{y} = \sqrt{\alpha p_u}\mathbf{s} + \sqrt{\alpha}\mathbf{D}^{-1/2}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{w}.$$

$\mathbf{r}$ $(K \times 1)$: the signal vector after ZF reception at the BS.

Results on uplink sum-rate

- SNR of User k:

$$\mathrm{SNR}_k = \frac{\beta_k p_u}{[(\mathbf{H}^H\mathbf{H})^{-1}]_{kk}}.$$

  Coefficient α has no effect since it appears on both signal and noise as a scaling factor.

- Need the following on CSCG and inverse Wishart distribution:

$$\mathbb{E}\{[(\mathbf{H}^H\mathbf{H})^{-1}]_{kk}\} = \frac{1}{K}\mathrm{tr}\{(\mathbf{H}^H\mathbf{H})^{-1}\} = \frac{1}{M-K}.$$

$$\mathbb{E}\left\{\frac{1}{[(\mathbf{H}^H\mathbf{H})^{-1}]_{kk}}\right\} = M+1-K.$$

  Sketch of proof:

  – From the QR-decomposition $\mathbf{H} = \mathbf{\Phi}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix}$,

$$[(\mathbf{H}^H\mathbf{H})^{-1}]_{KK} = [\mathbf{R}^{-1}\mathbf{R}^{-H}]_{KK} = \frac{1}{([\mathbf{R}]_{KK})^2}$$

  – Use the result on the distribution of $[\mathbf{R}]_{KK}$ on Page 36.

- Thus

$$\mathbb{E}\left\{\frac{1}{\mathrm{SNR}_k}\right\} = \frac{1}{\beta_k p_u(M-K)}.$$
$$\mathbb{E}\left\{\mathrm{SNR}_k\right\} = \beta_k p_u(M+1-K).$$

- Capacity approximation and bounds:

$$R_k \approx \log[1+\mathbb{E}\left(\mathrm{SNR}\right)] = \log[1+\beta_k p_u(M+1-K)].$$
$$R_k \geq \log\left[1+\frac{1}{\mathbb{E}\left(\mathrm{SNR}\right)}\right] = \log\left[1+\beta_k p_u(M-K)\right].$$
$$R_k \leq \log[1+\mathbb{E}\left(\mathrm{SNR}\right)] = \log[1+\beta_k p_u(M+1-K)]$$

Models for downlink with ZF

- BS sends signals to all users.
- $\mathbf{s} = [s_1 \cdots s_K]^T$: vector of information symbols.
  normalization: $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$.
- Zero-forcing precoding at BS:

$$\mathbf{A} = \sqrt{\alpha}\mathbf{G}^H(\mathbf{G}\mathbf{G}^H)^{-1} = \sqrt{\alpha}\mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}\mathbf{D}^{-1/2}.$$

  - BS transmitted the processed signal: $\mathbf{x} = \sqrt{p}\mathbf{A}\mathbf{s}$.
  - Average transmit power of BS: Kp. p is the power per user.
  - From $\mathbb{E}\{\mathbf{x}^H\mathbf{x}\} = Kp$, we have $\mathbb{E}\,\mathrm{tr}\{\mathbf{A}^H\mathbf{A}\} = K$. By applying the result on Page 63,

$$\Rightarrow \alpha\mathbb{E}\,\mathrm{tr}\{\mathbf{D}^{-1}(\mathbf{H}^H\mathbf{H})^{-1}\} = K \Rightarrow \alpha = \tilde{\beta}(M-K),$$

    where $\tilde{\beta} = \frac{1}{K}\sum_{k=1}^{K}\beta_k^{-1}$.

- Transceiver equation:

$$\mathbf{r} = \sqrt{p}\mathbf{G}\mathbf{A}\mathbf{s} + \mathbf{w} = \sqrt{\tilde{\beta}(M-K)p}\mathbf{s} + \mathbf{w}.$$

  - $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$: noise vector
  - $\mathbf{r}$: received signal vector at the users

- SNR of User k:

$$\mathrm{SNR}_k = \tilde{\beta}(M-K)p.$$

- Achievable rate of User k:

$$R_k = \log[1 + \tilde{\beta}(M-K)p].$$

The achievable-rate analysis method be generalized to imperfect CSI, multi-cell, and other more general cases.

References

[1] I. E. Telatar. Capacity of multi-antenna gaussian channels. *European Trans. on Telecommunications*, 10:585–595, Nov. 1999.

[2] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat-fading. *IEEE Trans. on Information Theory*, 45:139–157, Jan. 1999.

[3] A. Edelman. *Eigenvalues and Condition Numbers of Random Matrices*. Ph.D. Thesis, MIT, Department of Math. 1989.

[4] Y. Jing, B. Hassibi. Diversity analysis of distributed space-time codes in relay networks with multiple transmit/receive antennas. *EURASIP Journal on Advances in Signal Processing*. 2008.

[5] B. Clerckx and C. Oestges. *MIMO Wireless Networks: Channels, Techniques and Standards for Multi-Antenna, Multi-User and Multi-Cell Systems*. 2nd Edition, Academic Press, 2013.

[6] E. G. Larsson, H. Q. Ngo, H. Yang, and T. L. Marzetta. *Fundamentals of Massive MIMO*. Cambridge University Press. 2016.