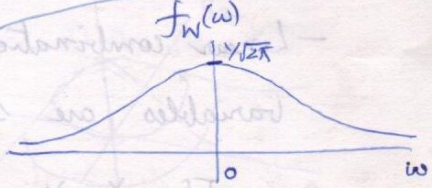
(1) 30/7/12 EE6003 Multi-antenna Digital Communication

Review of Gaussian random variables and vectors:

Scalar real Gaussian random variables

Standard Gaussian random variable W takes values over the real line $\mathbb{R}$ and has pdf

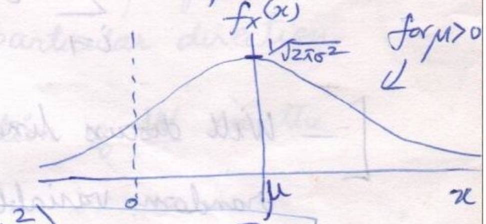
$$\boxed{f_W(\omega) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\omega^2}{2}\right), \quad \omega \in \mathbb{R}.}$$



Mean of $W = E[W] = 0$.

Variance of $W = E[W^2] (\triangleq \sigma^2) = 1$

General Gaussian random variable $X = \sigma W + \mu$



Mean of $X = E[X] = \mu$.

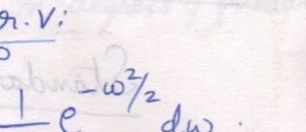
Variance of $X = E[(X-\mu)^2] = \sigma^2$

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.}$$

Notation: $X \sim N(\mu, \sigma^2)$ (Standard Gaussian $W \sim N(0,1)$).

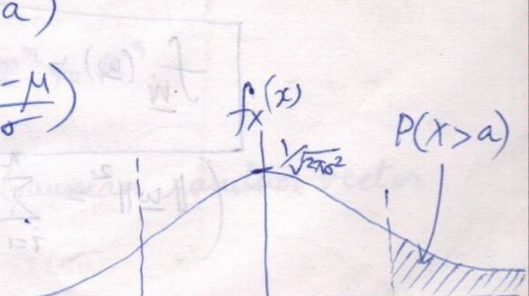
Tail probability of the standard Gaussian r.v:

$$Q(a) = P(W > a) = \int_a^\infty f_W(\omega)\, d\omega = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-\omega^2/2}\, d\omega.$$



Tail probability of a general Gaussian r.v:

$$P(X > a) = P(\sigma W + \mu > a)$$
$$= P\left(W > \frac{a-\mu}{\sigma}\right)$$
$$= Q\left(\frac{a-\mu}{\sigma}\right).$$



Useful bounds on $Q(a)$:

$$\frac{1}{\sqrt{2\pi}\,a}\left(1-\frac{1}{a^2}\right)e^{-a^2/2} < Q(a) < e^{-a^2/2} \quad \text{for } a>1$$

$\Rightarrow$ Tail probability decays exponentially fast (as a increases).

Linear transformations of Gaussian random variables:

- Linear combinations of independent Gaussian random variables are still Gaussian.

If $X_1, X_2, \dots, X_n$ are independent and $X_i \sim N(\mu_i, \sigma_i^2)$, then

$$\sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i\mu_i, \sum_{i=1}^{n} c_i^2\sigma_i^2\right).$$

[— Will discuss linear combinations of jointly Gaussian random variables after discussing jointly Gaussian random vectors]

- ## Real Gaussian random vectors:

Standard Gaussian random vector $\underline{W}$:

* Collection of n i.i.d. standard Gaussian r.v.s $W_1, W_2, \dots, W_n$.

$$\underline{W} = (W_1\ W_2\ \cdots\ W_n)^T$$

* Takes values in $\mathbb{R}^n$.

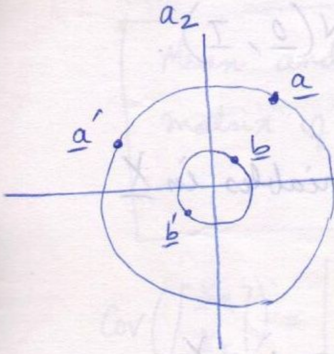
$$\boxed{f_{\underline{W}}(\underline{w}) = \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{\|\underline{w}\|^2}{2}\right), \quad \underline{w}\in\mathbb{R}^n}$$

$\left(\|\underline{w}\|^2 = \sum_{i=1}^{n} w_i^2;\right.$ Euclidean distance from origin $\underline{0}$ to $\left.\underline{w} = (w_1\ w_2\ \cdots\ w_n)^T\right)$

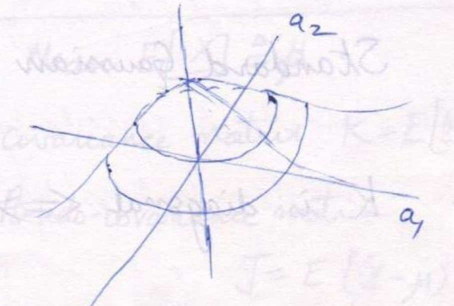
Note that the pdf $f_{\underline{W}}(\underline{w})$ depends only on the <u>magnitude</u> of the argument (i.e. only $\|\underline{w}\|$).

* If $T_{n\times n}$ is an orthonormal transformation; i.e, $T^T T = T T^T = I_n$, then $T\underline{W}$ is also standard Gaussian. ($T_{n\times n}$ preserves the magnitude of a vector. $\left(\|T\underline{w}\|^2 = \underline{w}^T T^T T \underline{w} = \underline{w}^T \underline{w} = \|\underline{w}\|^2 \right)$
Isotropic property.

(- Rotations and reflections are some such transformations)



$n=2$
$f(\underline{a}) = f(\underline{a}')$
$f(\underline{b}) = f(\underline{b}')$.



$\underline{W}$ does not prefer any particular direction.

Each element of $T\underline{W}$ is a projection of $\underline{W}$ in the direction defined by a row of T.
$\Rightarrow$ Projections of the standard Gaussian random vector in orthogonal directions are independent.

* $\|\underline{W}\|^2 = \sum_{i=1}^{n} W_i^2 \sim \chi_n^2$ chi-squared random variable with n degrees of freedom.

for $n=2$, $f_{\|\underline{W}\|^2}(a) = \frac{1}{2} e^{-a/2}$, $a \geqslant 0$

(exponential distribution for $n=2$ with parameter $\frac{1}{2}$).

L2, 31/7/12

<u>General Gaussian random vector</u> $\underline{X} = A\underline{W} + \underline{\mu}$

$\underline{\mu} \in \mathbb{R}^n$

A : linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$.

* Any linear transformation of a Gaussian random vector is also Gaussian.

* If A is invertible, then

$$\boxed{f_{\underline{X}}(\underline{x}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(AA^T)}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T (AA^T)^{-1} (\underline{x}-\underline{\mu})\right), \quad \underline{x} \in \mathbb{R}^n}$$

* Mean vector $E[\underline{X}] = \underline{\mu}$.

Covariance matrix $= E\left[(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})^T\right] = AA^T \triangleq K$.

$\underline{X} \sim N(\underline{\mu}, K)$. [Note that $\underline{X}$ is completely characterized by $\underline{\mu}$ and K].

Standard Gaussian random vector $\underline{W} \sim N(\underline{0}, I)$.

* K is diagonal $\Longleftrightarrow$ component random variables in $\underline{X}$ are independent.

$K = I \Longleftrightarrow$ component random variables in $\underline{X}$ are independent and have the same unit variance.
(also called white Gaussian random vector).

* If A is not invertible,

$A\underline{W}$ maps $\underline{W}$ into a subspace of dimension less than n

Density of $A\underline{W}$ equals 0 outside this subspace and impulsive inside. (or)

Some components of $A\underline{W}$ can be expressed as linear combinations of the others.

$\Rightarrow$ We can think of $\underline{X}$ to have some components $\tilde{\underline{X}}$ that are linearly independent combinations of $\underline{W}$ and other components are linear combinations of elements in $\tilde{\underline{X}}$.
$\tilde{\underline{X}}$ will have an invertible covariance matrix.
We will take K to be invertible in general for simplicity. Otherwise, we can handle using the above separation of $\underline{X}$ into $\tilde{\underline{X}}$ and other components.

Complex Gaussian random vectors:

$$\underline{X} = \underline{X}_R + j\,\underline{X}_I$$

where $\underline{X}_R$, $\underline{X}_I$ are real random vectors.

→ $\underline{X}$ is complex Gaussian if $\begin{bmatrix} \underline{X}_R \\ \underline{X}_I \end{bmatrix}$ is a real Gaussian random vector.

→ Distribution of $\underline{X}$ is specified completely by

Mean and covariance matrix of $\begin{bmatrix} X_R \\ X_I \end{bmatrix}$	(or) equivalently	Mean $E[\underline{X}] \triangleq \mu$. Covariance matrix $K = E[(\underline{X}-\mu)(\underline{X}-\mu)^H]$ Pseudo-covariance matrix $J = E[(\underline{X}-\mu)(\underline{X}-\mu)^T]$

$$\mathrm{Cov}\left(\begin{bmatrix} X_R \\ X_I \end{bmatrix}\right) = \begin{bmatrix} C_R & C_{RI} \\ C_{IR} & C_I \end{bmatrix}$$

where $C_{IR} = C_{RI}^T$.

$$K = (C_R + C_I) + j\,(C_{IR} - C_{RI})$$

$$J = (C_R - C_I) + j\,(C_{IR} + C_{RI}).$$

→ $\underline{X}$ is circularly symmetric if $e^{j\theta}\underline{X}$ has the same distribution as $\underline{X}$ for any θ.

$$\Rightarrow (1)\ E[\underline{X}] = E[e^{j\theta}\underline{X}] \Rightarrow E[\underline{X}] = \underline{0}\ \text{(zero-mean)}$$

$$(2)\ E[\underline{X}\,\underline{X}^T] = E[e^{j\theta}\underline{X}\,e^{j\theta}\underline{X}^T] = e^{j2\theta}E[\underline{X}\,\underline{X}^T]$$

$$\Rightarrow E[\underline{X}\,\underline{X}^T] = 0$$

$$\Rightarrow J = 0 \text{ (zero pseudo-covariance).}$$

$$\Rightarrow C_R = C_I,\quad C_{IR} = -C_{RI}.$$

If $\underline{X}$ is circularly symmetric, K fully specifies the second-order statistics. Further, if $\underline{X}$ is also Gaussian, K specifies the entire statistics.

$$\boxed{\text{Denoted } CN(\underline{0}, K)}$$

Special cases :

(1) $\underline{n=1}$ Complex Gaussian random variable

$$X = X_R + j X_I \rightarrow \text{5 parameters in general}$$

$E[X_R], E[X_I], E[X_R^2], E[X_I^2], E[X_R X_I]$.

X_R, X_I i.i.d zero-mean Gaussian $\Longleftrightarrow$ circularly symmetric X.

$\sim CN(0, \sigma^2)$.

$\sigma^2 = E[X_R^2] = E[X_I^2]$.

Only one parameter σ^2.

Standard circularly symmetric complex Gaussian $CN(0,1)$.

(2) Standard circularly symmetric complex Gaussian random vector $\underline{W} \sim CN(\underline{0}, I_n)$

$$f_{\underline{W}}(\underline{w}) = \frac{1}{\pi^n} \exp\left(-\|\underline{w}\|^2\right), \quad \underline{w} \in \mathbb{C}^n.$$

$U\underline{W}$ has the same distribution as $\underline{W}$ if

U is unitary, i.e., $U^H U = I$. (Isotropic property)

$\|\underline{W}\|^2 \sim \chi^2_{2n}$ (chi-squared r.v. with $2n$ degrees of freedom)

$\rightarrow$ $\underline{X} = A\underline{W}$ is also circularly symmetric complex Gaussian with covariance matrix $K = AA^H$.

$\underline{X} \sim CN(\underline{0}, K)$.

Also, any circular symmetric complex Gaussian r. vectors can be written as a linearly transformed version of a standard circularly symmetric complex Gaussian r. vector.

If A is invertible, we have

$$f_{\underline{X}}(\underline{x}) = \frac{1}{\pi^n \det(K)} \exp\left(-\underline{x}^H K^{-1} \underline{x}\right), \quad \underline{x} \in \mathbb{C}^n.$$

L3
8/12

Detection in Gaussian noise:

Scalar case: $Y = U + W$ (real AWGN channel)

U is equally likely to be u_A or u_B

$W \sim N(0, \frac{N_0}{2})$.

Detection problem: Observe Y and decide on U.

$\hat{U} = g(Y)$.

$g(\cdot)$ specifies detection rule.

Choose $g(\cdot)$ to minimize probability of error.

$P_e = Pr(\hat{U} \neq U)$

Minimize $Pr(\hat{U} \neq U / Y = y)$ for each y.

(MAP rule)

$\Rightarrow \hat{U} = u_A$ if $\boxed{Pr(U = u_A / Y = y) \geq Pr(U = u_B / Y = y)}$

$\Rightarrow \hat{U} = u_A$ if $f_Y(y/U = u_A)\, P(U = u_A) \geq f_Y(y/U = u_B)\, P(U = u_B)$

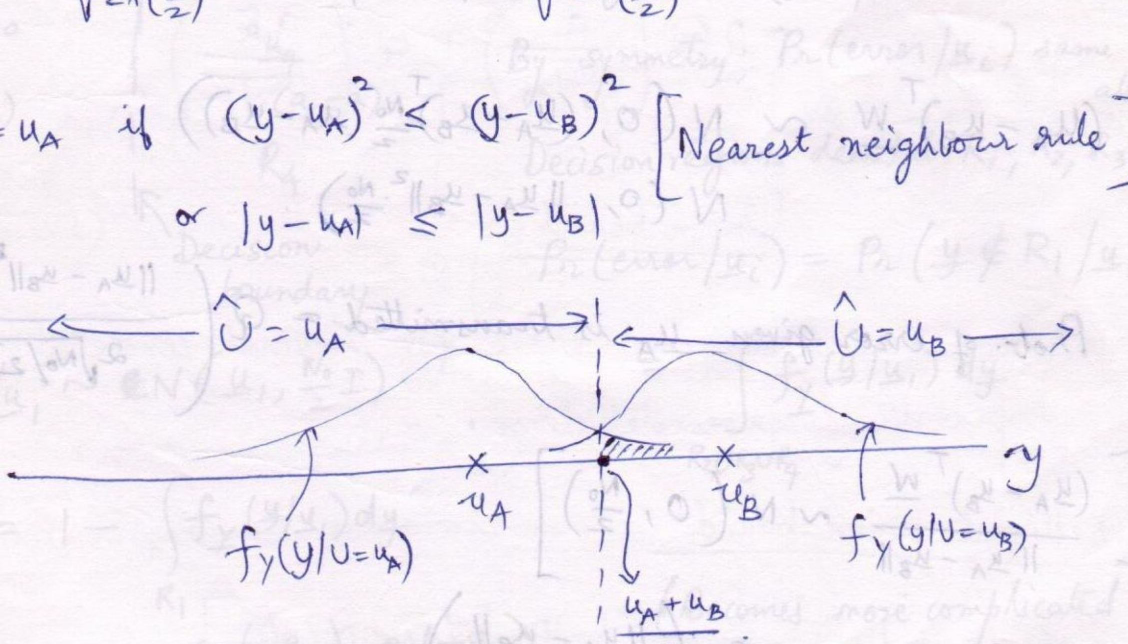
$\Rightarrow \hat{U} = u_A$ if $\boxed{f_Y(y/U = u_A) \geq f_Y(y/U = u_B)}$ (since $P(U = u_A) = P(U = u_B)$)

(ML rule)

$$\frac{1}{\sqrt{2\pi(\frac{N_0}{2})}} e^{-\frac{(y-u_A)^2}{2(N_0/2)}} \geq \frac{1}{\sqrt{2\pi(\frac{N_0}{2})}} e^{-\frac{(y-u_B)^2}{2(\frac{N_0}{2})}}$$

$\hat{U} = u_A$ if $(y - u_A)^2 \leq (y - u_B)^2$ [Nearest neighbour rule]

or $|y - u_A| \leq |y - u_B|$



— Easily extends to M-ary signaling $(u_1, u_2, \cdots, u_M)$ instead of binary signaling (u_A, u_B). [Nearest neighbour detection]

$\boxed{\text{Vector case}}$: $\quad \underline{Y}_{n\times 1} = \underline{U}_{n\times 1} + \underline{W}_{n\times 1}$

$\underline{W} \sim N\left(\underline{0}, \frac{N_0}{2} I\right)$ white Gaussian noise

$\underline{U}$ is either $\underline{u}_A$ or $\underline{u}_B$ with equal probability

<u>ML rule</u>: $\hat{\underline{U}} = \underline{u}_A$ if $f_{\underline{Y}}(\underline{y} | \underline{U} = \underline{u}_A) \geq f_{\underline{Y}}(\underline{y} | \underline{U} = \underline{u}_B)$

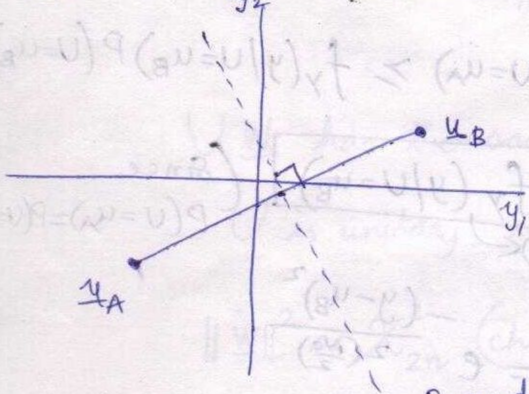
(or) $\|\underline{y} - \underline{u}_A\|^2 \leq \|\underline{y} - \underline{u}_B\|^2$

$$\left[\text{Since } f_{\underline{Y}}(\underline{y} | \underline{U} = \underline{u}_A) = \frac{1}{\left(2\pi \frac{N_0}{2}\right)^{n/2}} \cdot \exp\left(-\left(\frac{\|\underline{y} - \underline{u}_A\|^2}{N_0}\right)\right)\right]$$

Nearest neighbour rule.

— Projection of $\underline{Y}$ onto $\underline{u}_A - \underline{u}_B$ sufficient

<u>e.g.</u> 2D case



— Projection of noise along $\underline{u}_A - \underline{u}_B$ determines performance

— Noise orthogonal to $\underline{u}_A - \underline{u}_B$ is independent of noise along $\underline{u}_A - \underline{u}_B$ and irrelevant.

↳ Boundary: Hyperplane perpendicular to $\underline{u}_B - \underline{u}_A$.

$$(\underline{u}_A - \underline{u}_B)^T \underline{W} \sim N\left(0, (\underline{u}_A - \underline{u}_B)^T \frac{N_0}{2} (\underline{u}_A - \underline{u}_B)\right)$$

$$N\left(0, \|\underline{u}_A - \underline{u}_B\|^2 \cdot \frac{N_0}{2}\right)$$

Prob. of error given $\underline{u}_A$ is transmitted $= Q\left(\dfrac{\|\underline{u}_A - \underline{u}_B\|}{2\sqrt{N_0/2}}\right)$

$$\left[\frac{(\underline{u}_A - \underline{u}_B)^T \underline{W}}{\|\underline{u}_A - \underline{u}_B\|} \sim N\left(0, \frac{N_0}{2}\right)\right]$$

Prob. of error $= Q\left(\dfrac{\|\underline{u}_A - \underline{u}_B\|}{2\sqrt{N_0/2}}\right)$

— <u>Minimum distance/nearest neighbour rule</u> for M-ary case as well.

Complex vector case:

$$\underset{n\times 1}{\underline{Y}} = \underset{n\times 1}{\underline{U}} + \underset{n\times 1}{\underline{W}}$$

$\underline{W} \sim CN(\underline{0}, N_0 I_{n\times n})$ White Gaussian noise.

$\underline{U}$ is either $\underline{u}_A$ or $\underline{u}_B$ with equal probability

ML rule: $\hat{\underline{U}} = \underline{u}_A$ if $\|\underline{y} - \underline{u}_A\|^2 \le \|\underline{y} - \underline{u}_B\|^2$.

- Projection of $\underline{y}$ onto $\underline{u}_A - \underline{u}_B$ sufficient.

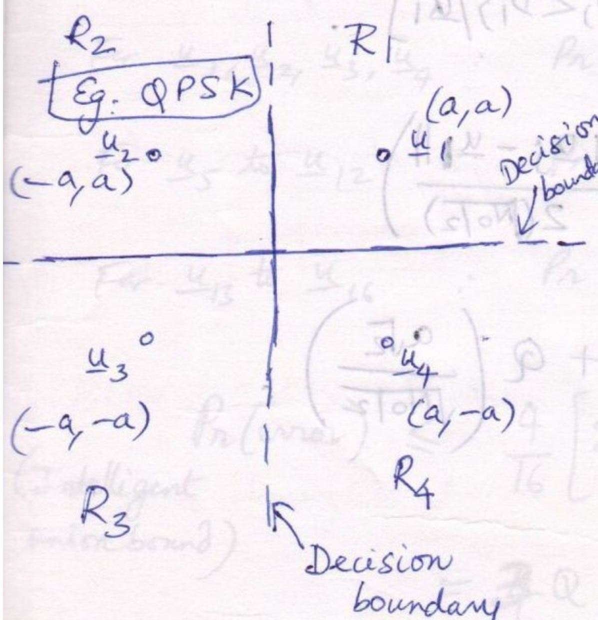
$$\frac{(\underline{u}_A - \underline{u}_B)^H \underline{Y}}{\|\underline{u}_A - \underline{u}_B\|} \rightarrow \text{a complex scalar r. variable.}$$

- Only noise along this vector determines performance.
- Can think of this as real vector case with dimension $2n$.

Probability of error and bounds (M-ary case):

→ $\underline{U}$ can take values $\underline{u}_1, \underline{u}_2, \cdots, \underline{u}_M$.

→ Decision rule: $\hat{\underline{U}} = \arg\min_{\underline{u}_i} \|\underline{y} - \underline{u}_i\|^2 = \arg\min_{\underline{u}_i} D_i^2$



$$Pr(\text{error}) = \sum_{i=1}^{M} Pr(\text{error}/\underline{u}_i)\, Pr(\underline{u}_i)$$

Assume $Pr(\underline{u}_i) = \frac{1}{4}$.

By symmetry, $Pr(\text{error}/\underline{u}_i)$ same for all i.

Decision regions denoted R_1, R_2, R_3, R_4

$$Pr(\text{error}/\underline{u}_i) = Pr(\underline{y} \notin R_1/\underline{u}_1)$$

$\underline{Y}/\underline{u}_1 \sim N(\underline{u}_1, \frac{N_0}{2} I)$

$$= \int_{R_2\cup R_3\cup R_4} f_{\underline{Y}}(\underline{y}/\underline{u}_1)\, d\underline{y}$$

$$= 1 - \int_{R_1} f_{\underline{Y}}(\underline{y}/\underline{u}_1)\, d\underline{y}$$

$$= 1 - Q\left(\frac{-a}{N_0/2}\right) \cdot Q\left(\frac{-a}{N_0/2}\right)$$

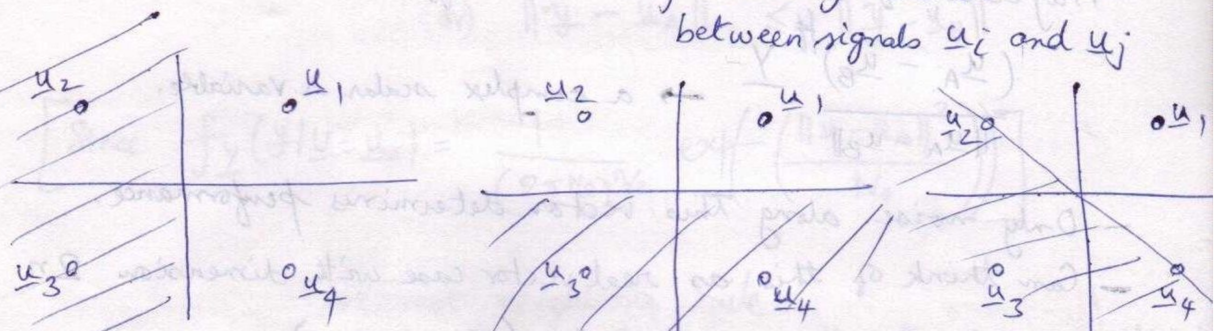
$$= 1 - \left(1 - Q\left(\frac{a}{N_0/2}\right)\right)^2$$

(Becomes more complicated to evaluate for large M and more complicated decision boundaries)

Union bound

$$P_r(\text{error}/\underline{u}_i) = P_r\left[\bigcup_{j\neq i}\{D_j < D_i\} / \underline{u}_i\right]$$

$$\leq \sum_{j\neq i} \underbrace{P_r\left[\{D_j < D_i\}/\underline{u}_i\right]}_{\substack{\text{Pairwise error probability}\\ \text{for binary detection}\\ \text{between signals } \underline{u}_i \text{ and } \underline{u}_j}}$$



$P_r(\{D_2 < D_1\}/\underline{u}_1)$ $\quad$ $P_r\{D_4 < D_1\}/\underline{u}_1)$ $\quad$ $P_r(\{D_3 < D_1\}/\underline{u}_1)$

$$P_r(\text{error}/\underline{u}_i) \leq \sum_{j=2}^{4} P_r\left[\{D_j < D_1\}/\underline{u}_1\right].$$

$$= \sum_{j=2}^{4} Q\left(\frac{\|\underline{u}_j - \underline{u}_1\|}{2(\sqrt{N_0/2})}\right)$$

$$= 2Q\left(\frac{a}{\sqrt{N_0/2}}\right) + Q\left(\frac{a\sqrt{2}}{\sqrt{N_0/2}}\right).$$

Union bound.

→ Union bound can be loose for several constellations, especially large constellations.

(4) 3/8/12

Intelligent union bound

$$P_r(\text{error}/\underline{u}_i) \leq \sum_{j\in N_{ML}(i)} P_r\left[\{D_j < D_i\}/\underline{u}_i\right]$$

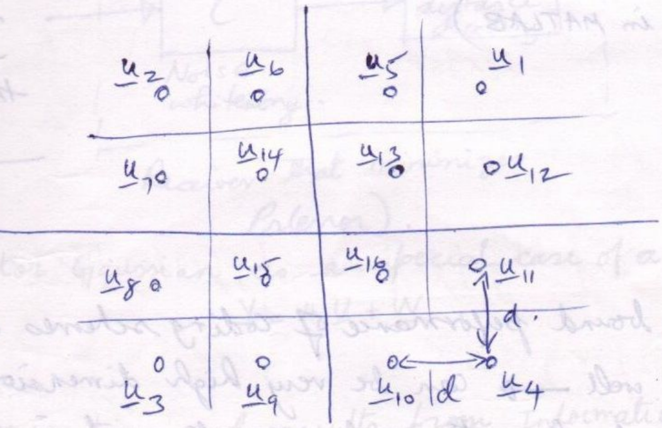
Set of neighbours of i that characterize the decision region R_i

– In the above QPSK example, only $\underline{u}_2$ and $\underline{u}_4$ define R_1.

$$\Rightarrow Pr(\text{error}/\underline{u}_1) \leq 2Q\left(\frac{a}{\sqrt{N_0/2}}\right).$$

Note: Any pairwise error probability $P[\{D_j < D_i\}/\underline{u}_i]$ is a lower bound for $Pr(\text{error}/\underline{u}_i)$.

Eg: 16-QAM



For $\underline{u}_1, \underline{u}_2, \underline{u}_3, \underline{u}_4$: $Pr(\text{error}/\underline{u}_i) \leq 2Q\left(\frac{d}{2\sigma}\right)$

For $\underline{u}_5$ to $\underline{u}_{12}$: $Pr(\text{error}/\underline{u}_i) \leq 3Q\left(\frac{d}{2\sigma}\right)$

For $\underline{u}_{13}$ to $\underline{u}_{16}$: $Pr(\text{error}/\underline{u}_i) \leq 4Q\left(\frac{d}{2\sigma}\right)$

$$Pr(\text{error}) \leq \frac{4}{16}\left[2Q\left(\frac{d}{2\sigma}\right)\right] + \frac{8}{16}\left[3Q\left(\frac{d}{2\sigma}\right)\right] + \frac{4}{16}\left[4Q\left(\frac{d}{2\sigma}\right)\right]$$

(Intelligent union bound)

$$= 3Q\left(\frac{d}{2\sigma}\right)$$

(Simple union bound will be looser and involve many more terms – 15 terms for each $\underline{u}_i$).

Nearest neighbours approximation

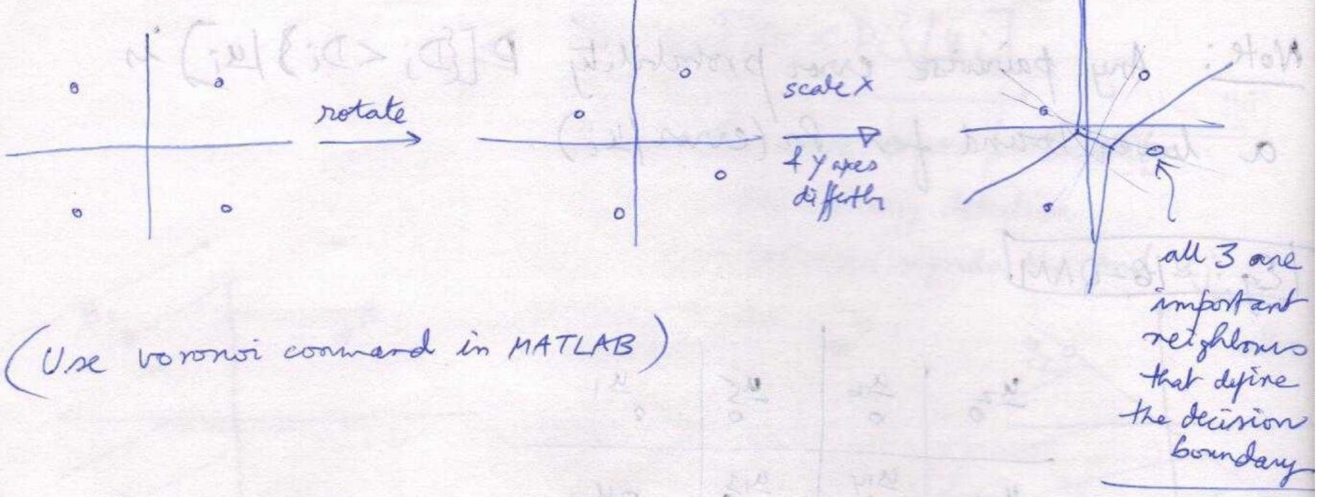
$$Pr(\text{error}/\underline{u}_i) \approx \underbrace{N_{d_{min}}(i)}_{\text{No. of nearest neighbours of } \underline{u}_i} Q\left(\frac{d_{min}}{2\sigma}\right)$$

d_{min} → distance at which nearest neighbours are.

– $Q(x)$ decays rapidly with x for large x. ⇒ terms with smaller x dominate.

- Can easily use these bounds in higher dimensions as well.

- Another example:



(Use voronoi command in MATLAB)

- Could use this to bound performance of coding schemes over AWGN channels as well → can be very high dimension depending on block length of the code, exact analysis very difficult.

- Will use this bound to analyze and design space-time codes.

Summary thus far:

→ Gaussian random vectors (real and complex)

→ Vector Gaussian channel

$$\underline{Y} = \underline{U} + \underline{W}$$

For $\underline{U} \in \{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_M\}$, detection rule that minimizes Pr(error). In the equally likely input case, minimum distance decoding rule. Calculation of Pr(error) and upper bounds / lower bounds / approximation.

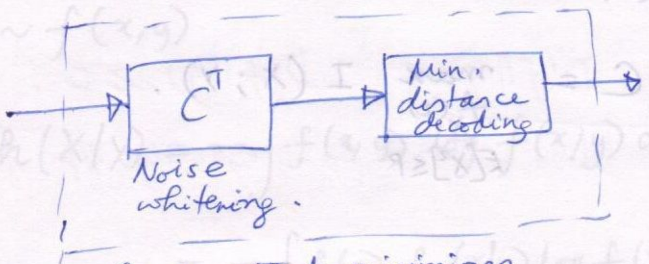
For a given transmission scheme, we discussed Pr(error). What about rate? Will talk a bit about capacity of vector Gaussian channel next.

→ Detection rule with correlated noise

$$\underline{Y} = \underline{U} + \underset{\hookrightarrow \sim N(\underline{0}, K)}{\underline{W}}$$

$$\max_i f_{\underline{Y}}(\underline{y}/\underline{u}_i) \equiv \min_i (\underline{y} - \underline{u}_i)^T K^{-1} (\underline{y} - \underline{u}_i)$$

$$K^{-1} = CC^T \Rightarrow \min_i \| C^T(\underline{y} - \underline{u}_i)\|^2$$



Receiver that minimizes Pr(error).

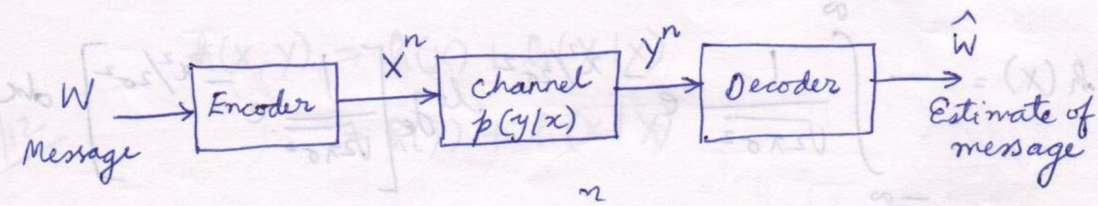
→ Vector Gaussian as a special case of a MIMO channel (diagonal H)

$$\underline{Y} = H\,\underline{U} + \underline{W}.$$

15 6/8/12

<u>Some definitions and results from Information Theory:</u>

A communication system model.



Memoryless channel : $p(y^n/x^n) = \prod_{i=1}^{n} p(y_i/x_i)$.

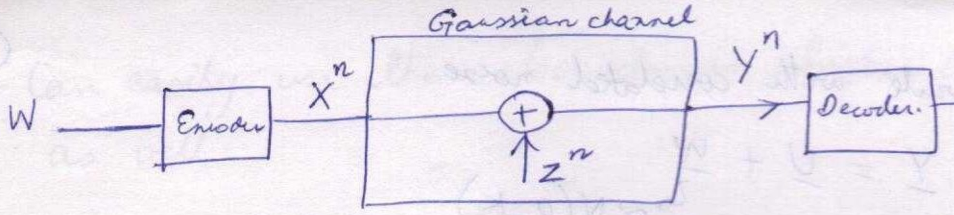
Discrete Memoryless channel (DMC): X, Y take values from discrete alphabets $\mathcal{X}, \mathcal{Y}$.

→ Channel used n times to send message W.
<u>As $n \to \infty$</u>, what is the <u>highest rate</u> C in bits/channel use at which information can be sent with <u>arbitrarily low probability of error</u>?

Shannon's result :

$$C = \max_{p(x)} I(X; Y).$$

where $I(X;Y)$ is the <u>mutual information</u> between X and Y.



$$Z_i \sim N(0, N) \text{ i.i.d.}$$

→ If input is unconstrained, capacity is infinite
(for the above channel with continuous input/output)

→ Usually, there is a transmit power constraint (average).

$$E[X^2] \leq P.$$

In this case, $C = \max\limits_{\substack{f_X(x) \\ E[X^2] \leq P}} I(X;Y).$

<u>Differential entropy</u>:

$X \sim f_X(x)$; $\quad h(X) = E[-\log f_X(X)] = -\int_S f_X(x) \log f_X(x)\, dx$.

(Depends only on the pdf) $\qquad S = \{x : f_X(x) > 0\}$.

<u>Eg</u>: $X \sim \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$ $\quad$ [i.e. $N(0, \sigma^2)$]

$$h(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \log_e\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\right] dx$$

$$= -\log_e \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\, dx$$

$$+ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \left[\frac{+x^2}{2\sigma^2}\right] dx$$

$$= \log_e \sqrt{2\pi\sigma^2} + \frac{1}{2} = \frac{1}{2}\log_e 2\pi\sigma^2 + \frac{1}{2}$$

$$= \frac{1}{2}\log_e(2\pi e \sigma^2) \text{ nats}$$

$$= \frac{1}{2}\log_2(2\pi e \sigma^2) \text{ bits}$$

Joint and conditional differential entropy:

$$X_1, X_2, \ldots, X_n \sim f(x_1, x_2, \ldots, x_n)$$

Joint entropy $[\, h(\underline{X}) \,]$

$$h(X_1, X_2, \ldots, X_n) = -E\left[\log f(X_1, X_2, \ldots, X_n)\right]$$

$$= -\int f(x_1 x_2, \ldots, x_n) \log f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n$$

$$X, Y \sim f(x,y)$$

Conditional entropy

$$h(X/Y) = -\int f(x,y) \log f(x/y)\, dx\, dy$$

$$= -\int f(y) \underbrace{f(x/y) \log f(x/y)\, dx}_{h(X|Y=y)}\, dy$$

Similarly $h(X \mid Y, Z) = -\int f(x,y,z) \log f(x/y,z)\, dx\, dy\, dz$.

Chain rule: $h(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} h(X_i \mid X_{i-1}, \ldots, X_1)$

$$h(X,Y) = h(X) + h(Y|X)$$

$$= h(Y) + h(X|Y)$$

(b) 7/8/12

Relative entropy:

$D(f\|g)$ between 2 densities $f(\cdot)$ and $g(\cdot)$ is

$$D(f\|g) = \int f(x) \log \frac{f(x)}{g(x)}\, dx$$

[Finite only if the support set of f is contained in the support set of g.]

Mutual information between 2 r.v's X, Y with joint pdf $f(x,y)$ is

$$I(X;Y) = D\left(f(x,y) \,\|\, f(x) f(y) \right)$$

$$= \int f(x,y) \log \frac{f(x,y)}{f(x) f(y)}\, dx\, dy$$

8

$$= \int f(x) f(y|x) \log \frac{f(x) f(y|x)}{f(x) f(y)} \, dx \, dy$$

$$= \int f(x,y) \log f(y|x) \, dx \, dy - \int f(x,y) \log f(y) \, dx \, dy$$

$$= h(Y) - h(Y|X)$$

$$= h(X) - h(X|Y).$$

Important property of $D(f\|g)$: $D(f\|g) \geq 0$.

with equality iff $f = g$.

$\Rightarrow$ (1) $I(X;Y) \geq 0$

(2) $h(X|Y) \leq h(X)$.

Entropy of a Gaussian random vector $\underset{n\times 1}{\underline{X}} \sim N(\mu, K)$

$$h(\underline{X}) = \frac{1}{2} \log_2 \left[(2\pi e)^n |K| \right] \text{ bits.}$$

Proof: $f_{\underline{X}}(\underline{x}) = \dfrac{1}{(\sqrt{2\pi})^n |K|^{1/2}} e^{-\frac{1}{2}(\underline{x}-\mu)^T K^{-1} (\underline{x}-\mu)}$

$$h(\underline{X}) = -\int f(\underline{x}) \ln \left[\frac{1}{(\sqrt{2\pi})^n |K|^{1/2}} e^{-\frac{1}{2}(\underline{x}-\mu)^T K^{-1}(\underline{x}-\mu)} \right] d\underline{x}$$

$$= \ln \left[(\sqrt{2\pi})^n |K|^{1/2} \right] + \frac{1}{2} E\left[(\underline{X}-\mu)^T K^{-1} (\underline{X}-\mu) \right]$$

$$= \frac{1}{2} \ln \left((2\pi)^n |K| \right) + \frac{1}{2} E\left[(\underline{X}-\mu)^T K^{-1} (\underline{X}-\mu) \right]$$

$$E\left[(\underline{X}-\mu)^T K^{-1} (\underline{X}-\mu) \right] = \mathrm{Tr}\left[E\left[(\underline{X}-\mu)^T K^{-1} (\underline{X}-\mu) \right] \right]$$

$$= E\left[\mathrm{Tr}\left((\underline{X}-\mu)^T K^{-1} (\underline{X}-\mu) \right) \right]$$

$$= E\left[\mathrm{Tr}\left((\underline{X}-\mu)(\underline{X}-\mu)^T K^{-1} \right) \right]$$

$$= \mathrm{Tr}\left(E\left[(\underline{X}-\mu)(\underline{X}-\mu)^T K^{-1} \right] \right) = \mathrm{Tr}(I_n) = n$$

$$h(\underline{X}) = \frac{1}{2}\log_e\left((2\pi)^n |K|\right) + \frac{1}{2} n \log_e e$$

$$= \frac{1}{2}\log_e\left((2\pi e)^n |K|\right) \text{ nats}$$

$$= \frac{1}{2}\log_2\left((2\pi e)^n |K|\right) \text{ bits.}$$

Thm:

Consider $\underline{X} \in \mathbb{R}^n$ which are zero-mean with covariance matrix $K = E[\underline{X}\,\underline{X}^T]$. Then,

$$h(\underline{X}) \leq \frac{1}{2}\log_2\left((2\pi e)^n |K|\right) \text{ bits.}$$

with equality iff $\underline{X} \sim N(\underline{0}, K)$.

Proof:

Let $g(\underline{x})$ be any density satisfying $E[\underline{X}] = \underline{0}$, $E[\underline{X}\,\underline{X}^T] = K$.

Let $\phi_K(\underline{x})$ be the pdf of a $N(\underline{0}, K)$ vector.

$$D(g \,||\, \phi_K) \geq 0.$$

$$\Rightarrow \quad 0 \leq \int g(\underline{x}) \log \frac{g(\underline{x})}{\phi_K(\underline{x})} d\underline{x}$$

$$= -h(g) - \int g(\underline{x}) \log \phi_K(\underline{x})\, d\underline{x}$$

Since $\phi_K(\underline{x}) = \frac{1}{(\sqrt{2\pi})^n |K|^{1/2}} e^{-\frac{1}{2}(\underline{x}^T K^{-1} \underline{x})}$.

$$\log \phi_K(\underline{x}) = -\log (\sqrt{2\pi})^n |K|^{1/2} - \frac{1}{2}\underline{x}^T K^{-1} \underline{x}.$$

$$E_g\left[\log \phi_K(\underline{X})\right] = -\log (\sqrt{2\pi})^n |K|^{1/2} - \frac{1}{2}E_g\left[\underline{X}^T K^{-1} \underline{X}\right]$$

$$= -\log (\sqrt{2\pi})^n |K|^{1/2} - \frac{1}{2}E_g\left[Tr\left(\underline{X}\,\underline{X}^T K^{-1}\right)\right]$$

$$= -\log (\sqrt{2\pi})^n |K|^{1/2} - \frac{n}{2}$$

$$= E_{\phi_K}\left[\log \phi_K(\underline{X})\right].$$

$$\Rightarrow h(\underline{Y}) \leq \frac{1}{2}\log\left((2\pi e)^n |K|\right) \text{ bits.}$$

L7
8/8/12

Capacity of the vector Gaussian channel:

$$\underline{Y} = \underline{X} + \underbrace{\underline{W}_{n\times 1}}_{\sim N(\underline{0}, K) \text{ and independent of } \underline{X}}$$

$$C = \max_{\substack{f_{\underline{X}}(x) \\ E[\underline{X}^T\underline{X}] \leq P}} I(\underline{X};\underline{Y})$$

$E[\underline{X}^T\underline{X}] \leq P$ → average power constraint.

$\Big[E[\underline{X}^T\underline{X}] \leq P$ is also equivalently $Tr(E[\underline{X}\,\underline{X}^T]) \leq P$
$Tr(K_X) \leq P \Big]$.

$$\begin{aligned} I(\underline{X};\underline{Y}) &= h(\underline{Y}) - h(\underline{Y}/\underline{X}) \\ &= h(\underline{Y}) - h(\underline{X}+\underline{W}/\underline{X}) \\ &= h(\underline{Y}) - h(\underline{W}/\underline{X}) \\ &= h(\underline{Y}) - h(\underline{W}) \\ &= h(\underline{Y}) - \frac{1}{2}\log\left[(2\pi e)^n |K|\right] \end{aligned}$$

Covariance of $\underline{Y}$, $K_Y = E[\underline{Y}\,\underline{Y}^H] = K_X + K$
(actually $E[(\underline{Y}-\mu_Y)(\underline{Y}-\mu_Y)^H]$)

For a given K_X

$$I(\underline{X};\underline{Y}) \leq \frac{1}{2}\log\left((2\pi e)^n |K_X + K|\right) - \frac{1}{2}\log\left((2\pi e)^n |K|\right)$$

We can further optimize over all K_X subject to $tr(K_X) \leq P$.

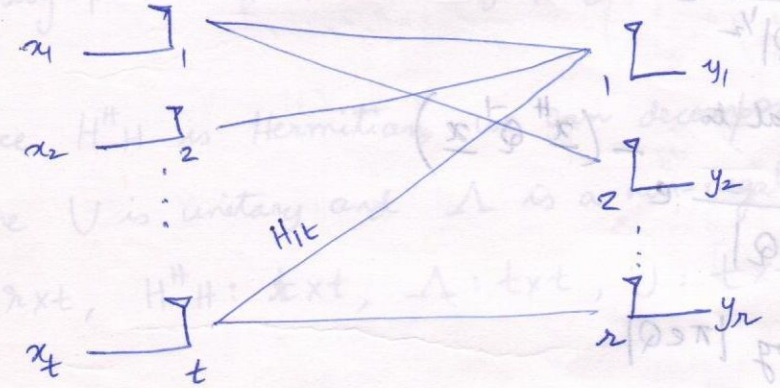
$$C = \max_{\substack{K_X \\ tr(K_X)\leq P}} \frac{1}{2}\log\frac{|K_X + K|}{|K|}.$$

This can be solved. We will do it for the MIMO case (and the vector Gaussian case will be a special case of this).

Emre Telatar, "Capacity of Multi-antenna Gaussian Channels," European Transactions on Telecommunications, 1999.

- Formulas for capacity
- Methods to evaluate these formulas
- Potential gains of multi-antenna systems over single-antenna systems
- Deterministic & fading multi-antenna channels.

Model



$$\underline{y} = \underbrace{H}_{r \times t \text{ matrix}} \underline{x} + \underline{n} \quad \underline{n} \sim CN(\underline{0}, I_r)$$

* Flat fading model between each tx-rx antenna pair

* Power constraint: $E[\underline{x}^H \underline{x}] \le P$ i.e. $tr(E[\underline{x}\,\underline{x}^H]) \le P$.

Three cases considered:

(1) H is deterministic

(2) H is a random matrix and independent for each use of the channel.

(3) H is a random matrix but fixed ~~for~~ all channel uses once chosen.

channel known at receiver, statistics known at transmitter.

Section 2: (self-study) discusses complex Gaussian random vectors in detail.

Important results:

(1) $\underline{X}_{n\times 1}$ Complex Gaussian. $\quad \hat{\underline{X}}_{2n\times 1} = \begin{bmatrix} Re(\underline{X}) \\ Im(\underline{X}) \end{bmatrix}$

Circularly symmetric $\underline{X}$ with cov. Q $\longrightarrow$ Cov $\hat{\underline{X}}$ is $\frac{1}{2}\begin{bmatrix} Re(Q) & -Im(Q) \\ Im(Q) & Re(Q) \end{bmatrix}$

(2) $\underline{X} \sim CN(\underline{0}, Q)$ has pdf

$$\frac{1}{|\pi \hat{Q}|^{1/2}} e^{-(\hat{\underline{x}}^H \hat{Q}^{-1} \hat{\underline{x}})}$$

which is equal to

$$\frac{1}{|\pi Q|} e^{-(\underline{x}^H Q^{-1} \underline{x})}$$

$$h(\underline{X}) = \log |\pi e Q|$$

(3) $\underline{X}$: zero-mean, with $E[\underline{X}\underline{X}^H] = Q$.

$$h(\underline{X}) \leq \log |\pi e Q|$$

with equality iff $\underline{X} \sim CN(\underline{0}, Q)$

(4) If $\underline{X} \sim CN$, then so is $\underline{Y}_{m\times 1} = A\underline{X}_{n\times 1}$ for any $A \in \mathbb{C}^{m\times n}$.

$$K_Y = A K_X A^H.$$

(5) If $\underline{X}$ & $\underline{Y}$ are independent circularly symmetric complex Gaussians, then $\underline{Z} = \underline{X} + \underline{Y}$ is circularly symmetric complex Gaussian.

Now, let us consider the case

$$\underline{Y} = H\underline{X} + \underbrace{\underline{N}}_{CN(\underline{0}, I_n)}$$

where H is deterministic.

Following the same steps as in the vector Gaussian case,

$$C = \max_{\operatorname{tr}(Q) \le P} \log |HQH^H + I_r| \triangleq \psi(Q, H)$$

$$C = \max_{\substack{Q \\ \operatorname{tr}(Q) \le P}} \psi(Q, H),$$

18
10/8/12

→ Since $\det(I + AB) = \det(I + BA)$

$$\log |I_r + HQH^H| = \log |I_t + QH^H H|$$

→ Since $H^H H$ is Hermitian, we can decompose $H^H H = U^H \Lambda U$

where U is unitary and Λ is a non-negative diagonal matrix.

$(H : r \times t,\ H^H H : t \times t,\ \Lambda : t \times t,\ U : t \times t)$

$$\psi(Q,H) = \log |I_t + QU^H \Lambda U| = \log |I_t + \underbrace{QU^H \Lambda^{1/2}}_{A} \underbrace{\Lambda^{1/2} U}_{B}|$$

$$= \log |I_t + \underbrace{\Lambda^{1/2} U}_{B} \underbrace{QU^H \Lambda^{1/2}}_{A}|$$

→ Observe $\tilde{Q} = UQU^H$ is non-negative definite, i.e., $\tilde{Q} \geq 0$ if and only if $Q \geq 0$. Also, $\operatorname{tr}(\tilde{Q}) = \operatorname{tr}(Q)$.

$$(\because \operatorname{tr}(UQU^H) = \operatorname{tr}(QU^H U) = \operatorname{tr}(Q)).$$

$$\Rightarrow \max_{\substack{Q \geq 0 \\ \operatorname{tr}(Q) \le P}} \psi(Q,H) = \max_{\substack{\tilde{Q} \geq 0 \\ \operatorname{tr}(\tilde{Q}) \le P}} \psi(\tilde{Q}, H)$$

$$= \log |I_t + \Lambda^{1/2} \tilde{Q} \Lambda^{1/2}|$$

Maximization of $\psi(Q,H)$ can be done over $\tilde{Q}$.

→ For any $A \geq 0$, $\det(A) \le \prod_i A_{ii}$, with equality when A is diagonal.

$$\tilde{Q} \geq 0 \Rightarrow \Lambda^{1/2} \tilde{Q} \Lambda^{1/2} + I_t \geq 0.$$

$$\Rightarrow \det(I_t + \Lambda^{1/2} \tilde{Q} \Lambda^{1/2}) \le \prod_i (i^{th} \text{ diagonal entry of } I_t + \Lambda^{1/2} \tilde{Q} \Lambda^{1/2})$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_t \end{bmatrix}$$

$$\Rightarrow \left| I_t + \Lambda^{1/2} \tilde{Q} \Lambda^{1/2} \right| \leq \prod_i \left(1 + \tilde{Q}_{ii} \lambda_i\right)$$

with equality when $\tilde{Q}$ is diagonal.

$\rightarrow$ $\tilde{Q}$ is diagonal. What should be the diagonal values?

Solve

$$\max_{\substack{\tilde{Q}_{ii} \geq 0 \\ \sum_i \tilde{Q}_{ii} \leq P}} \prod_i \left(1 + \tilde{Q}_{ii} \lambda_i\right)$$

Solution: $\tilde{Q}_{ii} = \left(\mu - \dfrac{1}{\lambda_i}\right)^{+}$ for $i = 1, 2, \ldots, t$.

where $\sum_i \tilde{Q}_{ii} = P$

(Waterfilling solution).

$$\rightarrow \quad C = \log\left(\prod_i \left(1 + \tilde{Q}_{ii} \lambda_i\right)\right) \qquad \tilde{Q}_{ii} \lambda_i = (\mu \lambda_i - 1)^{+}$$

$$= \sum_i \log\left(1 + \tilde{Q}_{ii} \lambda_i\right)$$

$$= \sum_{\substack{i \\ (\mu\lambda_i - 1 \geq 0)}} \log(\mu \lambda_i)$$

$$\underline{X} \sim \mathcal{CN}(\underline{0}, Q)$$

$$\underline{X} \longrightarrow \boxed{H} \longrightarrow \underline{Y}$$

$$\tilde{Q} = U Q U^H$$

$$\Rightarrow Q = U^H \tilde{Q} U$$

$$\underbrace{\underset{t\times 1}{\underline{\tilde{X}}} \sim \mathcal{CN}(\underline{0}, \tilde{Q}) \longrightarrow \boxed{U^H}}_{\text{Optimal input}} \xrightarrow{\underset{t\times 1}{\underline{X}} \sim \mathcal{CN}(\underline{0}, Q)} \boxed{H} \longrightarrow \underline{Y}$$

$$\underline{Y} = H\underline{X} + \underline{N}$$
$$= H\,U^H\underline{\tilde{X}} + \underline{N}$$

Alternative approach:

$H_{r\times t}$ can be written as $H = UDV^H$

where $U_{r\times r}$, $V_{t\times t}$ are unitary, $D \geq 0$ diagonal.

- Entries of D are the non-negative square roots of the eigenvalues of HH^H,
- Columns of U are the eigenvectors of HH^H
- Columns of V are the eigenvectors of $H^H H$.

$$\underline{X} \to \boxed{V^H} \xrightarrow{\underline{\tilde{X}}} \boxed{H} \xrightarrow{\underline{\tilde{Y}}} \boxed{U^H} \to \underline{Y}$$

$$\underline{Y} = H\underline{X} + \underline{N}$$

$$\Rightarrow \underline{Y} = UDV^H\underline{X} + \underbrace{\underline{N}}_{\sim CN(\underline{0}, I)}$$

$$\Rightarrow U^H\underline{Y} = D\underline{V^H X} + \underbrace{U^H\underline{N}}_{\tilde{N} \sim CN(\underline{0}, I)}$$

U, V are invertible and $\underline{\tilde{N}}$ has the same distribution as $\underline{N}$

$\Rightarrow$ $\underline{Y} = H\underline{X} + N$

equivalent to

$$\underline{\tilde{Y}} = D\underline{\tilde{X}} + \underline{\tilde{N}} \quad \text{where } \underline{\tilde{Y}} = U^H\underline{Y},\ \underline{\tilde{X}} = V^H\underline{X}.$$

Since H is a $r\times t$ matrix, its rank is at most $\min\{r, t\}$.

Only $\min\{r, t\}$ of the diag. entries of D are non-zero.

For these channels,

$$\tilde{y}_i = \lambda_i^{1/2}\,\tilde{x}_i + n_i \qquad i = 1, 2, \ldots, \min(r, t)$$

[The noise in channels where λ_i is zero does not have any effect.

Only the channels with $\lambda_i > 0$ need to be considered.]

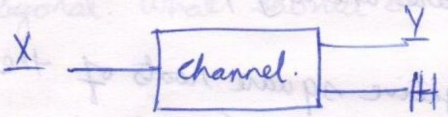
$\underline{\tilde{X}} \sim CN(\underline{0}, \tilde{Q})$ where $\tilde{Q}$ is determined by waterfilling

Multi-antenna Gaussian channel with Rayleigh Fading :

$$\underline{Y} = \mathbb{H}\,\underline{X} + \underline{N}.$$

each entry is independent and
$$\mathbb{H} \sim CN(0,1)$$

Assume H is known at the receiver.



$$I(\underline{X};\underline{Y},H) = I(\underline{X};\mathbb{H}) + I(\underline{X};\underline{Y}/\mathbb{H})$$
$$= 0 + I(\underline{X};Y/\mathbb{H})$$
$$= E_{\mathbb{H}}\left[I(\underline{X};\underline{Y}/\mathbb{H}=H)\right]$$

L9
12/8/12

Now
$$C = \max_{\substack{f_{\underline{X}}(x)\\ tr(K_X)\le P}} E_{\mathbb{H}}\left[I(\underline{X};\underline{Y}/\mathbb{H}=H)\right]$$

$$\le \max_{\substack{\underline{X}\sim CN(0,Q)\\ tr(Q)\le P\\ Q\ge 0}} E_{\mathbb{H}}\left[\log\det(I_r + \mathbb{H}Q\mathbb{H}^H)\right]$$

(Since $I(\underline{X};\underline{Y}/\mathbb{H}=H) \le \log\det(I_r + HQH^H)$ for any $\underline{X}$ with $K_X = Q$), $E_{\mathbb{H}}\left[I(\underline{X};\underline{Y}|\mathbb{H}=H)\right] \le E_{\mathbb{H}}\left[\log\det(I_r + \mathbb{H}Q\mathbb{H}^H)\right]$

$$C = \max_{\substack{tr(Q)\le P\\ Q\ge 0}} \underbrace{E_{\mathbb{H}}\left[\log\det(I_r + \mathbb{H}Q\mathbb{H}^H)\right]}_{\triangleq \psi(Q)}$$

$\rightarrow$ $Q \ge 0 \Rightarrow Q = UDU^H$ where U is unitary, $D \ge 0$ is diagonal.

$$\psi(Q) = E_{\mathbb{H}}\left[\log\det(I_r + \mathbb{H}UDU^H\mathbb{H}^H)\right]$$

$$= E_{\mathbb{H}}\left[\log\det\left(I_r + (\mathbb{H}U)D(\mathbb{H}U)^H\right)\right]$$

(Lemma 5: $\mathbb{H}_{r\times t}$ i.i.d. entries each $CN(0,1)$.

Then for any unitary $U_{r\times r}$, $V_{t\times t}$,

$U\mathbb{H}V^H$ has the same distribution as $\mathbb{H}$.

Pf:

Suffices to show $\mathbb{G} = U\mathbb{H}$ has the same distribution as $\mathbb{H}$

(Then, apply the result to $\mathbb{G}^H$).

$$\mathbb{G} = \begin{bmatrix} \underline{G}_1 & \underline{G}_2 & \cdots & \underline{G}_t \end{bmatrix} = U\begin{bmatrix} \underline{H}_1 & \underline{H}_2 & \cdots & \underline{H}_t \end{bmatrix}$$

$\Rightarrow \quad \underline{G}_i = U\underline{H}_i \sim CN(\underline{0}, I)$ same as $\underline{H}_i$.

Also, columns of $\mathbb{G}$ are independent since columns of $\mathbb{H}$ are independent ■)

→ Since the distribution of $\mathbb{H}U$ is the same as $\mathbb{H}$,

$$\Psi(Q) = \Psi(D) = E_{\mathbb{H}}\left[\log\det\left(I_r + \mathbb{H}D\mathbb{H}^H\right)\right]$$

Therefore, we can restrict to $\underline{\text{diagonal } Q \geq 0}$.

→ Let Π be any permutation matrix.

Define $Q^{\Pi} = \Pi Q \Pi^H$. $\quad$ Q^{Π} has diagonal entries that are a permutation of diagonal entries of Q.

$$\Psi(Q^{\Pi}) = E_{\mathbb{H}}\left[\log\det\left(I_r + \mathbb{H}\Pi Q \Pi^H \mathbb{H}^H\right)\right]$$

Since $\mathbb{H}\Pi$ has the same distribution as $\mathbb{H}$,

$$\Psi(Q^{\Pi}) = \Psi(Q).$$

→ (1) For any H, $\quad Q \longrightarrow I_r + HQH^H$ is linear and preserves positive definiteness.

(2) Since log det is concave on the set of positive definite matrices,

$$Q \to \psi(Q, \mathbb{H}) = \log\det(I_r + \mathbb{H} Q \mathbb{H}^H) \text{ is concave.}$$

(1) & (2) $\Rightarrow$ $Q \to \psi(Q)$ is <u>concave</u>.

$\longrightarrow$ $\displaystyle \tilde{Q} = \frac{1}{t!}\sum_{\Pi} Q^{\Pi}$ satisfies $\psi(\tilde{Q}) \geq \psi(Q)$

and $tr(\tilde{Q}) = tr(Q)$.

$\Rightarrow$ $\tilde{Q} = \alpha I$ is the optimal Q.

and $\alpha = \dfrac{P}{t}$.

$$\boxed{C = E_{\mathbb{H}}\left[\log\det\left(I_r + \frac{P}{t}\mathbb{H}\mathbb{H}^H\right)\right].}$$

<u>Remarks</u>:

(1)
$$\mathbb{H} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1t} \\ \vdots & & & \\ \vdots & & & \\ H_{r1} & H_{r2} & \cdots & H_{rt} \end{bmatrix}_{r\times t}$$

$$\mathbb{H}\mathbb{H}^H = \begin{bmatrix} \sum_{i=1}^{t}|H_{1i}|^2 & & & \\ & \sum_{i=1}^{t}|H_{2i}|^2 & & \\ & ((i,j)^{th}\text{ element}) & \ddots & \\ & & & \sum_{i=1}^{t}|H_{ri}|^2 \end{bmatrix}_{r\times r}$$

$$\sum_{k=1}^{t} H_{ik} H_{kj}^{*}$$

For a fixed r, as $t \to \infty$

$$\frac{1}{t}\mathbb{H}\cdot\mathbb{H}^H \longrightarrow I_r \quad \underline{\text{almost surely}} \text{ (by law of large numbers)}.$$

$$\Rightarrow C \longrightarrow r \log(1+P)$$

14/8/12

(2) For $r=1$, $\mathbb{H} = [H_{11} \ H_{12} \ \cdots \ H_{1t}]$

$$\mathbb{H}\mathbb{H}^H = \sum_{i=1}^{t} |H_{1i}|^2$$

For $t=1$, $\mathbb{H} = \begin{bmatrix} H_{11} \\ H_{21} \\ \vdots \\ H_{r1} \end{bmatrix}$ $\quad \left|I_r + \frac{P}{t}\mathbb{H}\mathbb{H}^H\right| = \left|1 + \frac{P}{t}\mathbb{H}^H\mathbb{H}\right|$

$$\mathbb{H}^H\mathbb{H} = \sum_{i=1}^{r} |H_{i1}|^2.$$

$$C = E\Bigg[\underbrace{\log \left(1 + \frac{P}{t}\sum_{i=1}^{t}|H_{1i}|^2\right)}_{(r=1)\text{ case}} \text{ (or) } \underbrace{E\log\left(1+P\sum_{i=1}^{r}|H_{i1}|^2\right)}_{(t=1)\text{ case}}\Bigg]$$

(3) Evaluation of the capacity

$$\det\left(I_r + \frac{P}{t}\mathbb{H}\mathbb{H}^H\right) = \det\left(I_t + \frac{P}{t}\mathbb{H}^H\mathbb{H}\right)$$

(Use this when $r<t$) (Use this when $r \geq t$)

$$\text{Let } W = \begin{cases} \mathbb{H}\mathbb{H}^H & r<t \\ \mathbb{H}^H\mathbb{H} & r \geq t \end{cases}$$

Let $n = \max(r,t)$

$m = \min(r,t)$

W : $m \times n$ random non-negative definite matrix

follows the <u>Wishart distribution with parameters m, n</u>

$\lambda_1, \lambda_2, \ldots \lambda_m$: Eigenvalues of W (random variables)

$$C = E_{\underline{\lambda}}\left[\sum_{i=1}^{m} \log\left(1 + \frac{P}{t}\lambda_i\right)\right]$$

Joint density of the unordered eigenvalues:

$$f_{\underline{\lambda}}(\lambda_1, \lambda_2, \ldots, \lambda_m) = \frac{1}{m!\, K_{m,n}} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2$$

Joint density of ordered eigenvalues:

$$f_{\underline{\lambda},\text{ordered}}(\lambda_1, \ldots, \lambda_m) = \frac{1}{K_{m,n}} e^{-\sum_i \lambda_i} \prod_i \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2$$

for $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_m \geq 0$

($K_{m,n}$ is a normalizing factor)

Density of one of the unordered eigenvalues:

$$f_{\lambda_1}(\lambda_1) = \frac{1}{m} \sum_{i=1}^{m} \varphi_i(\lambda_1)^2 \lambda_1^{n-m} e^{-\lambda_1} \qquad \text{— (1)}$$

where $\varphi_{k+1}(\lambda) = \left[\dfrac{k!}{(k+n-m)!}\right]^{1/2} L_k^{n-m}(\lambda) \quad k = 0, \ldots, m-1.$

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

$$C = E\left[\sum_{i=1}^{m} \log\left(1 + \frac{P}{t}\lambda_i\right)\right] = m\, E\left[\log\left(1 + \frac{P}{t}\lambda_1\right)\right]$$

where λ_1 is distributed as in (1).

$$\Rightarrow \boxed{C = \int_0^{\infty} \log\left(1 + \frac{P}{t}\lambda\right) \sum_{k=0}^{m-1} \frac{k!}{(k+n-m)!} \left[L_k^{n-m}(\lambda)\right]^2 \lambda^{n-m} e^{-\lambda}\, d\lambda.}$$

where $m = \min\{r, t\}$, $n = \max\{r, t\}$

L_j^i are the associated Laguerre polynomials.

Special cases

(1) $\boxed{t=1}$

Note $L_0^{n-m}(\lambda) = 1$.

$$\Rightarrow C = \frac{1}{\Gamma(r)} \int_0^\infty \log(1+P\lambda)\, \lambda^{r-1} e^{-\lambda} d\lambda$$

(See Figure 2)

As $r \to \infty$, $C \to \log(1+Pr)$.

(2) $\boxed{r=1}$

$$C = \frac{1}{\Gamma(t)} \int_0^\infty \log\left(1+\frac{P}{t}\lambda\right) \lambda^{t-1} e^{-\lambda} d\lambda$$

(See Figure 3)

As $t \to \infty$, $C \to \log(1+P)$

(3) $\boxed{r=t}$

$$C = \int_0^\infty \log\left(1+\frac{P\lambda}{r}\right) \sum_{k=0}^{r-1} L_k(\lambda)^2 e^{-\lambda} d\lambda.$$

where $L_k = L_k^0$ is the Laguerre polynomial of order k.

Using results from the theory of random matrices, we get

$$C \sim r \int_0^4 \log(1+P\nu) \frac{1}{\pi} \sqrt{\frac{1}{\nu} - \frac{1}{4}}\, d\nu$$

which is linear in r. (See Fig. 4).

(4) $C(r,t,P)$: Capacity of channel with r receivers, t transmitters, P total transmit power.

$$\boxed{C(a,b,Pb) = C(b,a,Pa)}$$

e.g. $C(1,t,Pt) = C(t,1,P)$

$t \times 1$ with power Pt $(\equiv)$ $1 \times t$ with power P.

(5) Above, capacity is also ergodic capacity : when $\mathbb{H}$ is ergodic (not only when $\mathbb{H}$ is i.i.d. memoryless)

Non-ergodic channels:

→ $\mathbb{H}$ is chosen randomly at the beginning and held fixed for all uses of the channel.

$\mathbb{H}$ i.i.d. entries $\mathcal{CN}(0,1)$.

→ Capacity $= 0$.

For any rate $R>0$, there is a non-zero probability that the realized $\mathbb{H}$ is incapable of supporting it.

$\Rightarrow$ outage.

$$P_{out}(R,P) = \inf_{\substack{Q:Q\geq 0 \\ tr(Q)\leq P}} Pr\left(\Psi(Q,\mathbb{H}) < R\right).$$

For a given rate R and power P, we can find $P_{out}(R,P)$ 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 $\mathbb{H}$ whose probability is less than $P_{out}(R,P)$.

$$\Psi(Q,\mathbb{H}) = \log\det\left(I_r + \mathbb{H}Q\mathbb{H}^H\right).$$

→ $\boxed{t=1}$ case $\Rightarrow$ Clearly $Q_{1\times 1} = P$.

$$\mathbb{H} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_{r\times 1} \qquad \log\det(I_r + \mathbb{H}Q\mathbb{H}^H) = \log(1 + P\mathbb{H}^H\mathbb{H})$$

$$\text{Outage probability} = Pr\left(\log(1 + P\mathbb{H}^H\mathbb{H}) < R\right)$$

$\mathbb{H}^H\mathbb{H} \sim \chi^2$ with $2r$ degrees of freedom and mean r.

$$P_{out}(R,P) = \frac{\gamma\left(r, \frac{(e^R-1)}{P}\right)}{\Gamma(r)}$$

where $\gamma(a,x)$ is the incomplete gamma function

$$= \int_0^x u^{a-1} e^{-u}\, du .$$

Let $\psi(P,\epsilon)$ be the rate R at which $P(\psi(P,\mathbb{H}) \le R) = \epsilon$.

i.e. $\psi(P,\epsilon)$ is the $\underline{\epsilon\text{-outage capacity}}$.

→ $\mathbb{H}U$ has the same distribution as $\mathbb{H}$ for unitary U.

$$\Rightarrow \psi(UQU^H, \mathbb{H}) = \psi(Q, \mathbb{H})$$

$\Rightarrow$ Q can be chosen to be diagonal without affecting the outage prob.

$$P_{out}(R,P) = \inf_{\substack{Q:\,Q \ge 0 \\ Q \text{ diag} \\ \sum_i Q_{ii} \le P}} P_r(\psi(Q,\mathbb{H}) < R)$$

→ Only a <u>conjecture</u> for optimal Q.

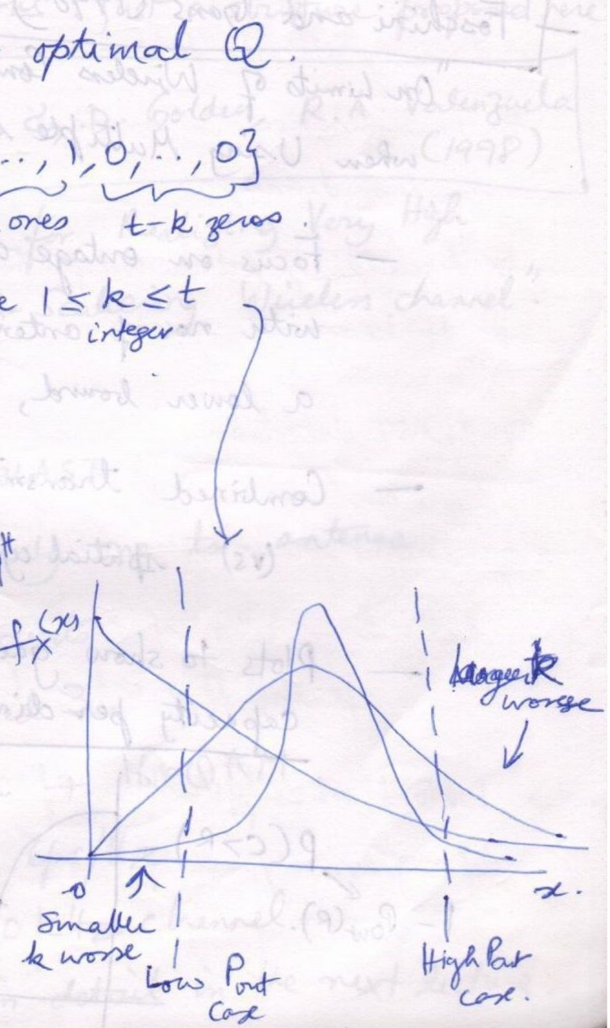
$$Q = \frac{P}{k}\,\text{diag}\{\underbrace{1,\dots,1}_{k \text{ ones}}, \underbrace{0,\dots,0}_{t-k \text{ zeros}}\}$$

for some $1 \le k \le t$ integer

→ Example: $\boxed{r=1}$ case.

See Fig. 6 & 7.

$X \triangleq \mathbb{H}\mathbb{H}^H$



$\boxed{\text{So far}}$: Telatar (1999)

→ deterministic H

— SVD + waterfilling

→ Random H (fast fading)

— Ergodic capacity

— $C \propto \min(r, t)$

→ Random H (slow fading) - non-ergodic

— Outage probability

— Outage capacity

— Some evaluation/plots for $r=1$ case.

$\boxed{\text{Today}}$: Summary of some other parallel work (also from Bell labs)

— $\boxed{\text{Foschini and Gans (1998)}}$

"On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas".

— Focus on outage capacity: How does it increase with no. of antennas? Linear. (analysis based on a lower bound, numerical evaluation).

— Combined transmit-receive diversity (vs) spatial cycling, tx. diversity, rx. diversity

— Plots to show gains in <u>outage capacity</u> and <u>capacity per dimension</u>