Random matrices and their applications to Multi-Input Multi Output Communication Systems

Anna Scaglione

School of Electrical & Computer Engineering

$\mathbb{C}$ornell $\mathbb{U}$niversity

March 6, 2002

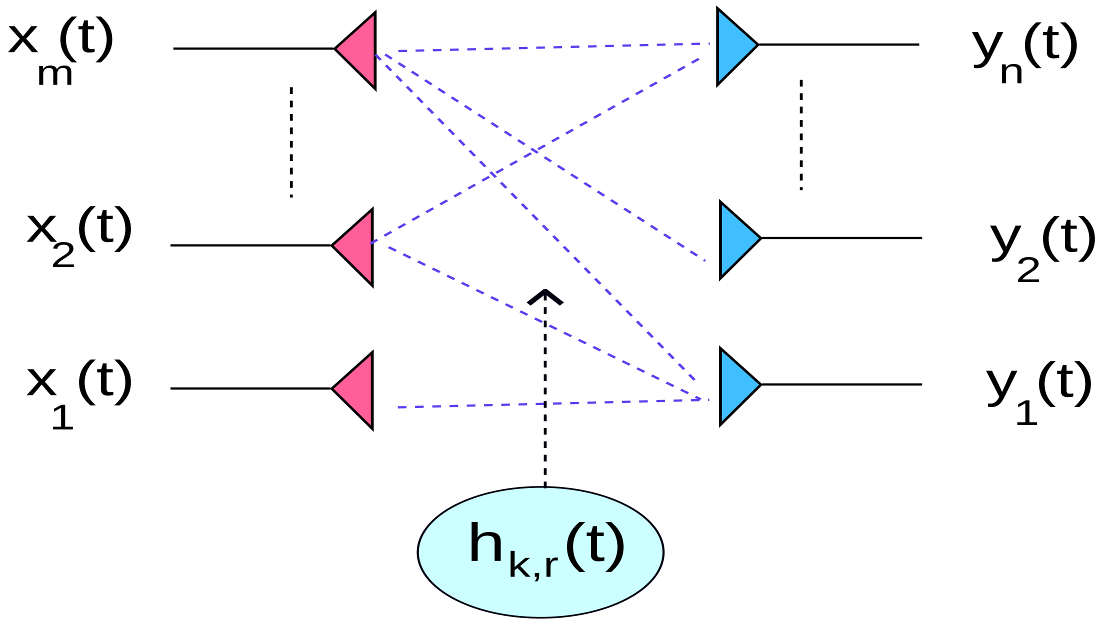
Why should we be interested in this problem

- Percentage of papers appeared in 2001 that have at least one matrix defined: 89% IT, 65% Com., 99.999999% SP
- We all use MATLAB, we all own our personal copy of "Matrix Computations" [Golub, Van Loan]. Tons of matrices are decomposed everyday by ECE students
- Random Matrices are practical, useful and have beautiful properties
- Random Matrices are good for you: they make you realize that your knowledge of Calculus is at the level of a Mickey Mouse Cartoon

Random Matrices in Communication Systems

- Multiple sources, Multiple Sensors
- System with transmit and receive diversity
  - Random Fading
  - Random Space-Time Codes
- Symbol Synchronous CDMA system
- Data Vectors with Random Covariance

MIMO Systems



$$\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))^T$$

- Multiple Access, Array Processing: $\boldsymbol{x}(t)$ from different sources
- Transmit Diversity: $\boldsymbol{x}(t) = \sum_{n=-\infty}^{+\infty} \boldsymbol{x}[n] g_T(t - nT) \qquad n \times 1$
- Space-time code $\boldsymbol{c} = (\boldsymbol{x}[n_1], \ldots, \boldsymbol{x}[n_l]) \quad n \times l$

MIMO Channel Output Model

$$\boldsymbol{y}(t) = (y_1(t), \ldots, y_n(t))^T$$

$\boldsymbol{y}[k] := \boldsymbol{y}(kT)$ signals samples, $T \approx 1/W$, $W=$ Bandwidth

- Narrowband Channel (Flat fading):

$$\boldsymbol{y}[k] = \boldsymbol{H}[k]\boldsymbol{x}[k] + \boldsymbol{n}[k] \qquad m \times 1$$

- Broadband Channel (Frequency selective):

$$\boldsymbol{y}[k] = \sum_{n=-\infty}^{\infty} \boldsymbol{H}[k-n]\boldsymbol{x}[n] + \boldsymbol{n}[k]. \qquad m \times 1$$

$\boldsymbol{H}[k]$ is $m \times n$

Symbol Synchronous CDMA system

- Multiple sources, One sensor, Multiple samples

$$y(t) = \sum_{k=1}^{K} A_k b_k s_k(t) + n(t)$$

the vector $\{\boldsymbol{y}\}_{1,p} \triangleq y(pT_c)$,

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{b} + \boldsymbol{n}$$

$\{\boldsymbol{b}\}_{1,k} \triangleq b_k$, Users symbols

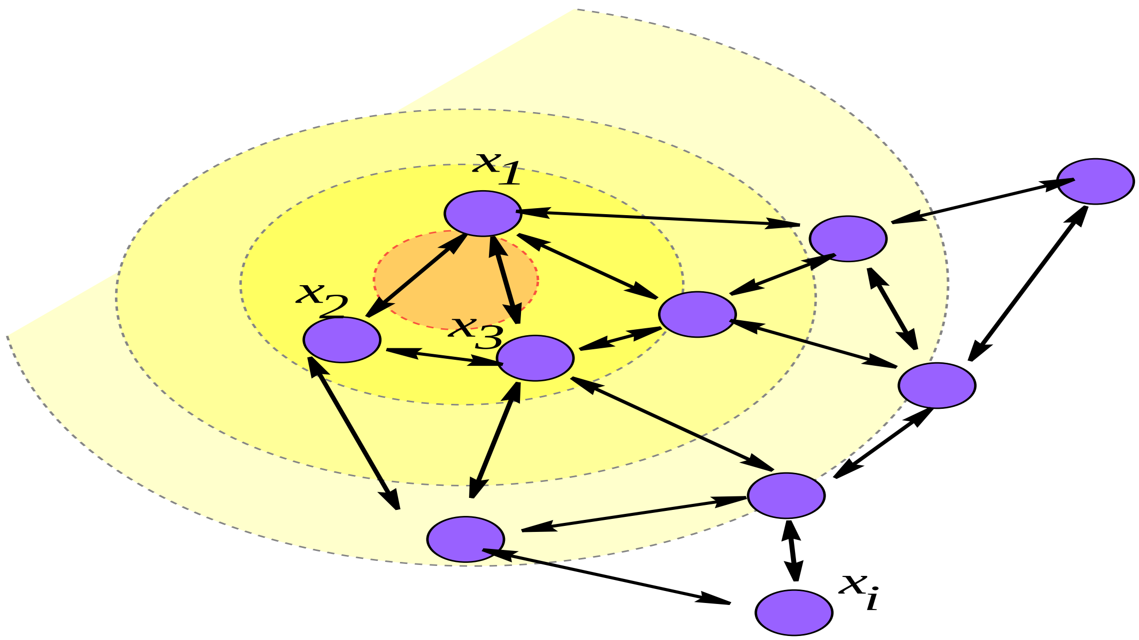
$\boldsymbol{A} \triangleq diag(A_1, \ldots, A_K)$, Users amplitudes

$\{\boldsymbol{S}\}_{k,p} \triangleq s_k(pT_c)$, Users signatures sampled at the chip rate $T_c \approx 1/W$

- This model was used in [Hanly,Tse'99] to compare the performances of Linear Multiuser Detectors

Vectors with Random Covariances

- Sensor network



- The covariance matrix of $\boldsymbol{x}[k] = (x_1[k], \ldots, x_n[k])$ depends on the relative distances between nodes $d_{i,j}$ which is random.
- We can study the rate distortion function of the data as a whole

Relevant Performance Measures

MIMO channel ⇒ $\boldsymbol{H}$ random, CDMA ⇒ $\boldsymbol{S}$ random

- MIMO channel Capacity and CDMA aggregate Capacity:

$$C = \log|\sigma^2 \boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^H| \qquad C = \log|\sigma^2 \boldsymbol{I} + \boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{S}^H|$$

- MIMO channel MMSE for LMMSE receiver

$$MMSE = Tr((\boldsymbol{I} + \boldsymbol{H}^H\boldsymbol{H}\sigma^{-2})^{-1})$$

- LMMSE User SIR

$$SIR_i = \boldsymbol{s}_i^H(\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{S}^H + \sigma^2\boldsymbol{I})^{-1}\boldsymbol{s}_i$$

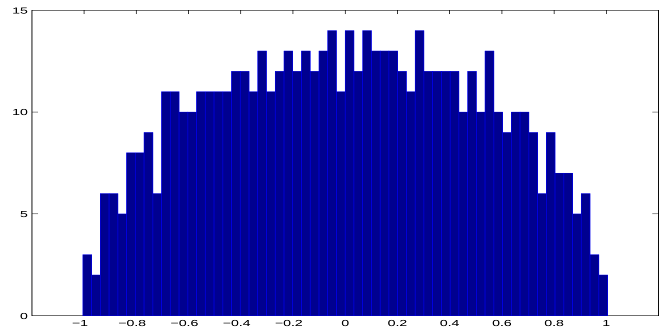
- Decorrelating receiver User SIR

$$SIR_i = \{\sigma^2(\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^H\boldsymbol{S}^H)^{-1}\}_{i,i}^{-1}$$

- Differential Entropy of Gaussian sensor data

$$H(X) = \frac{1}{2}\log(2\pi e)^n|\boldsymbol{R}_{xx}|$$

Random Eigenvalues



- Field initiated by the pioneering work by Eugene Paul Wigner
- He searched for the asymptotic empirical density of the ordered random eigenvalues of an $n \times n$ symmetric $\boldsymbol{X}(\omega)$:

$$\mu_\omega(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - \lambda_{ii}(\boldsymbol{X}(\omega)))$$

- Now we can run this 3 line worth MATLAB experiment

```
>> n=600; B=randn(n); A=(B+B')/(2*sqrt(2*n)); hist(eig(A),60)
```

and see what Wigner derived with pencil and paper (semicircle law)

Asymptotic distribution we care about

Theorem [Machenko-Pastur '67]

Let $\boldsymbol{X} = \frac{1}{n}\boldsymbol{B}^H\boldsymbol{B}$ *and* $\boldsymbol{B}$ $m \times n$ *and such that* $\alpha = m/n$*:*
(a) *The elements of* $\boldsymbol{B}_n$ *are i.i.d. random variables* $\in \mathbb{C}$ *with* $E\{[\boldsymbol{B}]_{i,j}\} = 0$, $Var\{[\boldsymbol{B}]_{i,j}\} = 1$ *and* $E\{|[\boldsymbol{B}]_{i,j}|^4\} < \infty$.
$\mu_\omega(x)$ *converges weakly as* $n \to \infty$ *to the Machenko-Pastur distribution*

$$\mu_\alpha(x) = \max(1-\alpha, 0)\delta(x) + \frac{\sqrt{(x-a)(b-x)}}{2\pi x} 1_{[a,b]}(x)$$

where $1_{[a,b]}(x) = 1$ *for* $a \leq x \leq b$ *and is zero elsewhere,* $\delta(x)$ *is a Dirac delta and:*

$$a := (\sqrt{\alpha} - 1)^2 \; , \qquad b := (\sqrt{\alpha} + 1)^2$$

- More general asymptotic results are available (used for CDMA systems).

MIMO channel- Some Asymptotic results

$\gamma := \frac{\mathcal{P}_0 \sigma_H^2}{\sigma_n^2}$ The closed form expression for the normalized Capacity is

$$C(\gamma)= \frac{1}{\log(2)} \left(\log(\gamma \ w) + \frac{1-\alpha}{\alpha} \ \log\left(\frac{1}{1-v}\right) - \frac{v}{\alpha} \right)$$

where:

$$w = \frac{1}{2}\left(1+\alpha+\gamma^{-1} + \sqrt{(1+\alpha+\gamma^{-1})^2 - 4\alpha}\right)$$

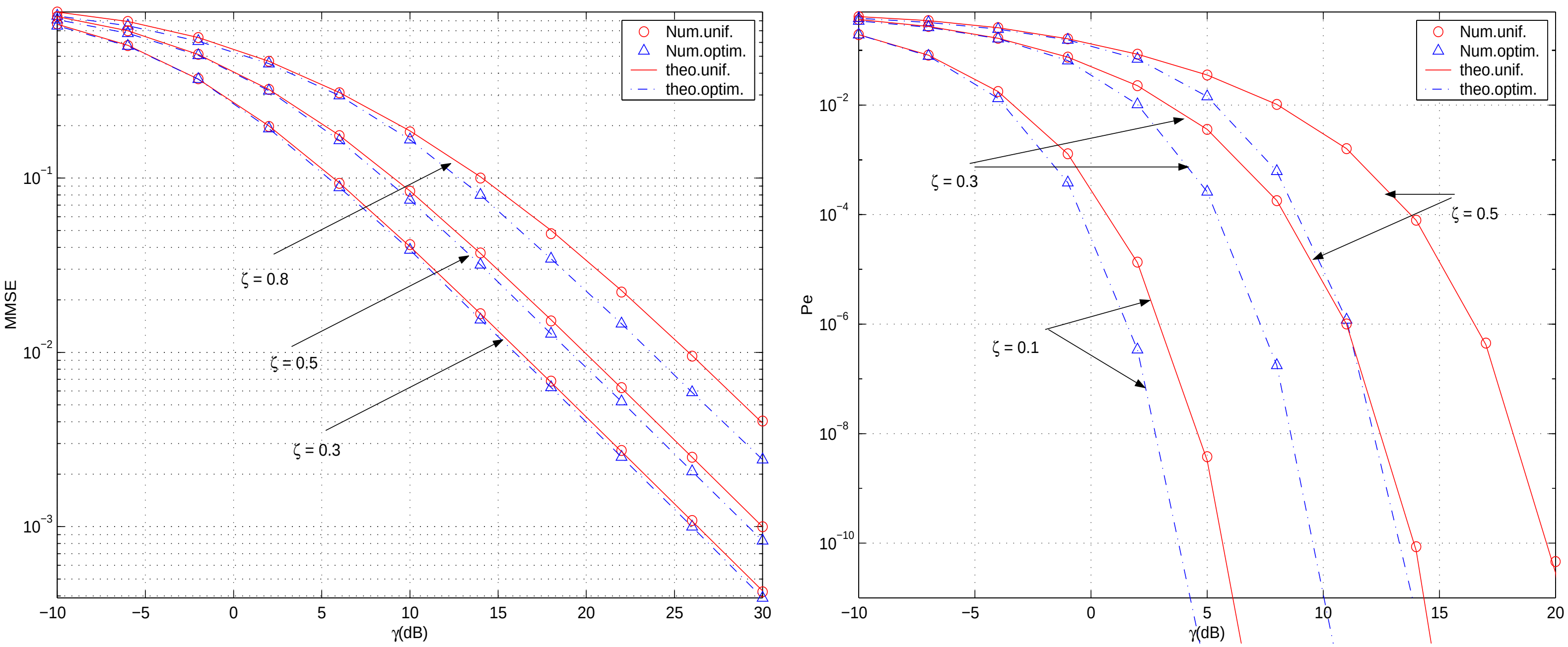
$$v = \frac{1}{2}\left(1+\alpha+\gamma^{-1} - \sqrt{(1+\alpha+\gamma^{-1})^2 - 4\alpha}\right)$$

Asymptotic results

The expressions of MSE and P_e are:

$$\overline{MSE}(\gamma) \;=\; \frac{1}{2\,\alpha\gamma}\left(-\sqrt{a\,b}\;\gamma + \sqrt{1+a\,\gamma}\sqrt{1+b\,\gamma}\;-1\right)$$

$$P_e(\gamma) \;\leq\; \exp\left(-(1+\alpha)\gamma\right)\left(\frac{1}{2\sqrt{\alpha}}\,I_1(2\sqrt{\alpha}\gamma)\frac{1+\alpha}{4\alpha}I_0(2\sqrt{\alpha}\gamma)\right)$$



Derivation of the Statistics

- Asymptotic results are ready to use (Examples later)
- Derivation of the Statistics for the finite case
  Matrix decompositions are Transformations of Random Variables!
  - First step: deriving the Jacobian of the change of variables from the original matrix to its factors
  - Verify the uniqueness: true in the case of EVD (almost true), QR or LU (lower-upper) decompositions and Cholesky decomposition
- One way of doing it: Exterior differential Calculus

Exterior Differential Calculus

- Seminal work of Élie Cartan
- Based on the concept of *exterior product* $\triangleq \wedge$, introduced by Hermann Günter Grassmann in 1844

Axioms of Grassman Exterior Algebra:

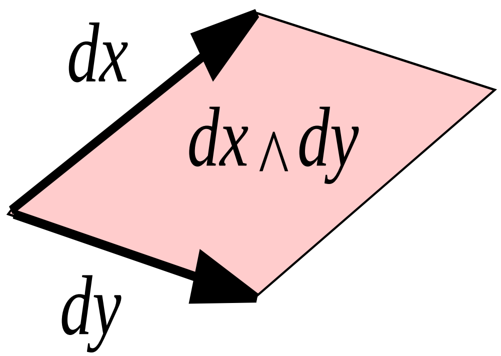
- $\alpha \wedge \alpha = 0$
- $\alpha \wedge \beta = -\beta \wedge \alpha$
- $(a\alpha) \wedge \beta = a(\alpha \wedge \beta)$.

The axioms are sufficient to establish that:

$$(\boldsymbol{A}\alpha) \wedge \beta = |\boldsymbol{A}|(\alpha \wedge \beta).$$

Amenity: Grassman at the age of 53 grew frustrated with the lack of interest in his mathematical work and turned to Sanskrit studies, writing a widely used dictionary.

What kind of product is $dxdy$? Is $dx \wedge dy$



- The product of differentials $dxdy$ behaves like $dx \wedge dy$ and we can use on it the axioms of Grassman Exterior Algebra
- To complete the description of Cartan's differential forms: Axiomatic definition of the d operator
  - $d(r\text{-form}) = (r+1)\text{-form}$
  - $d(dx) = 0$ (Poincarè Lemma)
- These rules are systematic and the results are simpler to grasp than the theory of manifolds

The Jacobian recipe

- ★ $d\boldsymbol{A}$ matrix of differentials
- ★ $(d\boldsymbol{A})$ the exterior product of the independent entries in $d\boldsymbol{A}$:
  - ▶ for an arbitrary $\boldsymbol{A}$, $(d\boldsymbol{A}) = \wedge_i \wedge_j da_{ij}$
  - ▶ if $\boldsymbol{A}$ is diagonal $(d\boldsymbol{A}) = \wedge_i da_{ii}$
  - ▶ if $\boldsymbol{A} = \boldsymbol{A}^T$ or $\boldsymbol{A}$ is lower triangular $(d\boldsymbol{A}) = \wedge_{1\le i\le j\le n} da_{ij}$
  - ▶ for $\boldsymbol{Q}$ unitary ... (not nice)
- Select the arbitrary unique matrix factorization, for Ex. $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}$
- Apply the d operator $\rightarrow d\boldsymbol{X} = d\boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}d\boldsymbol{B}$
- Evaluate $(d\boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}d\boldsymbol{B})$, $\wedge$ product of all the independent differentials

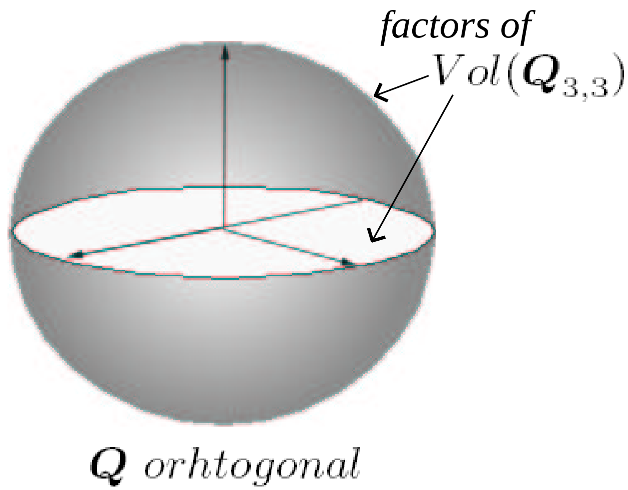
Warning: This last task requires the description of the group of matrices by mean of their independent parameters

The Stiefel Manifold

- A unitary $\boldsymbol{Q}$ is described by n^2 smooth functions that can be integrated over nice enough intervals (Stiefel Manifold)
- Clearly, the independent parameters of the Stiefel Manifold are not the real and imaginary parts of the elements of $\boldsymbol{Q}$
- n out of the n^2 parameters are redundant (in the sense that the decomposition is unique up to n parameters), hence we can assume:
  (a) The diagonal elements of $\boldsymbol{Q}$ are real
- $\boldsymbol{Q}\boldsymbol{Q}^H = \boldsymbol{I} \;\rightarrow\; \boldsymbol{Q}d\boldsymbol{Q}^H = -d\boldsymbol{Q}\boldsymbol{Q}^H$
- Under **(a)** the diagonal elements of $\boldsymbol{Q}d\boldsymbol{Q}^H$ are zero and $\boldsymbol{Q}d\boldsymbol{Q}^H$ is antisymmetric

$$(d\boldsymbol{Q}) \equiv (\boldsymbol{Q}^H d\boldsymbol{Q}) = \wedge_{i>j} \boldsymbol{q}_i^H d\boldsymbol{q}_j$$

The Stiefel Manifold



- The uniform p.d.f. in the Stiefel group of orthogonal or unitary matrices is called *Haar distribution*

The volume of $(\boldsymbol{Q}^H d\boldsymbol{Q})$ integrated over $\boldsymbol{Q}^H\boldsymbol{Q} = \boldsymbol{I}$, for $\boldsymbol{Q}$ unitary, when the diagonal elements of $\boldsymbol{Q}_{m,n}$ are constrained to be real:

$$\overline{Vol}(\boldsymbol{Q}_{m,n}) \triangleq \int_{\boldsymbol{Q}^H\boldsymbol{Q}=\boldsymbol{I}} (\boldsymbol{Q}^H d\boldsymbol{Q}) = \frac{(\pi)^{(m-1)n-n(n-1)/2}}{\prod_{i=0}^{n-1}\Gamma(m-i)}$$

The statistics of $\boldsymbol{A} = \boldsymbol{B}^H \boldsymbol{B}$

- $\boldsymbol{A} = \boldsymbol{B}^H \boldsymbol{B}$ with $p_{\boldsymbol{A}}(A)$ and $p_{\boldsymbol{B}}(B)$ the pdfs of the random matrices $\boldsymbol{A}$ and $\boldsymbol{B}$
- Trick: Use QR and Cholesky decompositions first

$$\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{R} \ , \qquad \boldsymbol{A} = \boldsymbol{R}^H \boldsymbol{R}.$$

with $(d\boldsymbol{R}) = \wedge_{i<j}(dr_{ij})$

$$(d\boldsymbol{A}) = 2^n \prod_{i=1}^{n} (|r_{ii}|^2)^{n+1-i} (d\boldsymbol{R}) \implies p_{\boldsymbol{A}}(\boldsymbol{A})(d\boldsymbol{A}) = p_{\boldsymbol{A}}(\boldsymbol{R}^H \boldsymbol{R}) \prod_{i=1}^{n} 2^n \left(|r_{ii}|^2\right)^{n+1-i} (d\boldsymbol{R})$$

$$(d\boldsymbol{B}) = \prod_{i=1}^{n} (|r_{ii}|^2)^{m+1-i} (d\boldsymbol{R})(d\boldsymbol{Q}) \implies p_{\boldsymbol{B}}(\boldsymbol{B})(d\boldsymbol{B}) = p_{\boldsymbol{B}}(\boldsymbol{Q}\boldsymbol{R}) \prod_{i=1}^{n} \left(|r_{ii}|^2\right)^{m+1-i} (d\boldsymbol{R})(d\boldsymbol{Q})$$

where $(d\boldsymbol{Q}) = (\boldsymbol{Q}^H d\boldsymbol{Q})$ is the element of volume of the Stiefel manifold

Generalized Wishart Density

$$p_{\boldsymbol{A}}(\boldsymbol{A}) = 2^{-n}|\boldsymbol{A}|^{m-n} \int p_{\boldsymbol{B}}(\boldsymbol{Q}\sqrt{\boldsymbol{A}})(\boldsymbol{Q}^H d\boldsymbol{Q})$$

- When the p.d.f. $p_{\boldsymbol{B}}(\boldsymbol{B}) = p_{\boldsymbol{B}}(\boldsymbol{B}^H\boldsymbol{B})$ then:

  - $\boldsymbol{Q}$ and $\boldsymbol{R}$ in the QR decomposition $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{R}$, are independent
  - The p.d.f. of $\boldsymbol{Q}$ has *Haar distribution*
  - The p.d.f. of $\boldsymbol{A}$ is:

$$p_{\boldsymbol{A}}(\boldsymbol{A}) = 2^{-n}|\boldsymbol{A}|^{m-n} p_{\boldsymbol{B}}(\boldsymbol{A}) Vol(\boldsymbol{Q}_{m,n})$$

The statistics of the EVD $\boldsymbol{A} = \boldsymbol{B}^H\boldsymbol{B}$

$$\begin{aligned}
(d\boldsymbol{A}) &= (d\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H + \boldsymbol{U}d\boldsymbol{\Lambda}\boldsymbol{U}^H + \boldsymbol{U}^H\boldsymbol{\Lambda}d\boldsymbol{U}) \\
(d\boldsymbol{A}) &\equiv (\boldsymbol{U}^H d\boldsymbol{A}\boldsymbol{U}) = (\boldsymbol{U}^H d\boldsymbol{U}\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{U}^H d\boldsymbol{U} + d\boldsymbol{\Lambda}) \\
&= \prod_{1\le i<k\le n}^{n} (\lambda_k - \lambda_i)^2 (d\boldsymbol{\Lambda})(\boldsymbol{U}^H d\boldsymbol{U}).
\end{aligned}$$

In the general case of $\boldsymbol{A} = \boldsymbol{B}^H\boldsymbol{B}$:

$$\begin{aligned}
p_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda}) &= 2^{-n}\prod_{1\le i<k\le n}^{n} (\lambda_k - \lambda_i)^2 \Psi(\lambda_1, \ldots, \lambda_n) \\
\Psi(\lambda_1, \ldots, \lambda_n) &\triangleq \int p_{\boldsymbol{A}}(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H)(d\boldsymbol{U}).
\end{aligned}$$

$$\{\boldsymbol{B}\}_{i,j} \sim \mathcal{N}(0, \sigma^2) \implies \Psi(\lambda_1, \ldots, \lambda_n) = \left(\prod_{i=1}^{n} \lambda_i\right)^{m-n} e^{-\frac{\sum_i \lambda_i}{\sigma^2}}$$

MIMO frequency selective channel

Let's use all this machinery!

a1. The noise is AWGN with variance $\sigma_n^2 = 1$

a2. $\{\mathbf{H}[l]\}^*_{r,t}$ are spatially uncorrelated circularly symmetric $\mathcal{N}(0,1)$ (Rayleigh fading) with $R_H[l_1, l_2, r_1, r_2, t_1, t_2] \triangleq E\{\{\mathbf{H}[l_1]\}^*_{r_1,t_1}$ $\{\mathbf{H}[l_2]\}_{r_2,t_2}\} = \delta(t_1 - t_2)\, \delta(r_1 - r_2) R_H(l_2, l_1)$

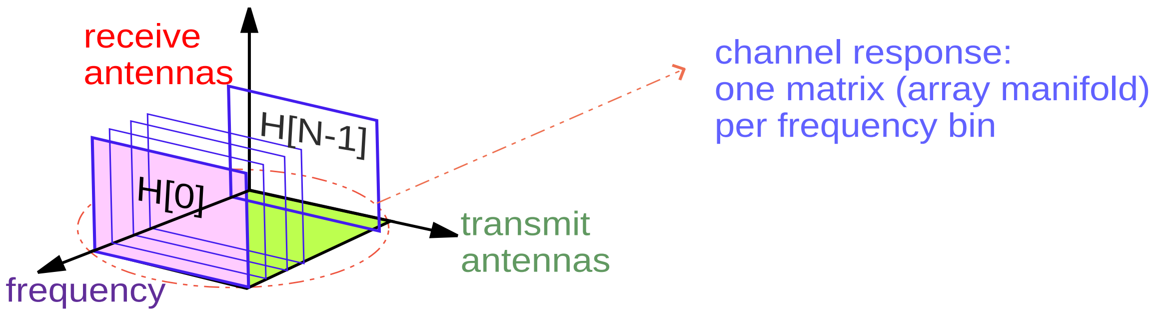
a3. $n \triangleq \min(N_T, N_R)$, $\quad m \triangleq \max(N_T, N_R)$

$$C = \log |\boldsymbol{I} + \gamma \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{H}}|$$

$$\tilde{\boldsymbol{H}} \triangleq diag(\tilde{\mathbf{H}}[\boldsymbol{d}]) \ , \quad \boldsymbol{d} \triangleq (0, \ldots, L),$$

$\tilde{\mathbf{H}}[k]$ is the MIMO transfer function at the kth frequency bin:
$\tilde{\mathbf{H}}[k] = \sum_{l=0}^{L} \mathbf{H}[l] e^{-j2\pi \frac{kl}{K}}$

Average Capacity



$$E\{C\} = \sum_{k=0}^{K-1} \sum_{l=1}^{N_T} E\left\{\log(1+\gamma\lambda_l[k])\right\}$$

Under **a1, a2**, the average Capacity for any (n, m) is:

$$E\{C\} = \sum_{k=0}^{K-1} \int_0^{\infty} \log\left(1+\gamma\sigma_H^2[k]x\right) \mu_n^{m-n}(x)dx$$

with $\alpha = m - n$ ($L_k^{\alpha}(x)$ the Laguerre polynomials):

$$\mu_n^{\alpha}(x) = \frac{1}{n}\sum_{k=0}^{n-1} \phi_k^{\alpha}(x)^2 \qquad \phi_k^{\alpha}(x) \triangleq \left[\frac{k!}{\Gamma(k+\alpha+1)} x^{\alpha} e^{-x}\right]^{1/2} L_k^{\alpha}(x)$$

Characteristic function of C

$$\Phi_C(s) = E\{e^{sC}\} = E\left\{\prod_{k=0}^{K-1} |\boldsymbol{I} + \gamma\tilde{\mathbf{H}}[k]^H\tilde{\mathbf{H}}[k]|^s\right\}$$

a3. The number of frequency bins $K = Q(L+1)$.

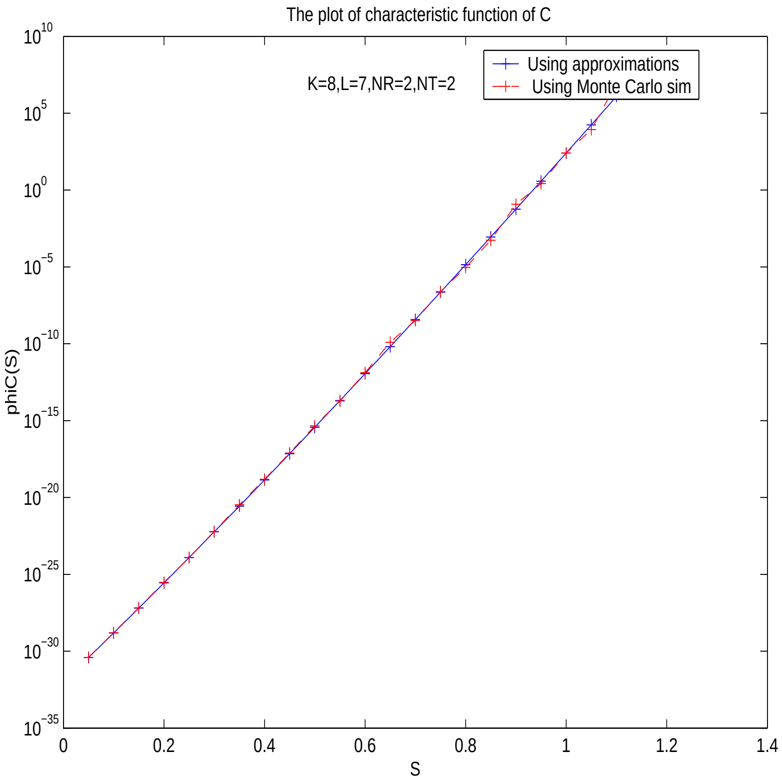
Choosing $\mathbf{p} = (0, Q, \ldots, QL)$, since $e^{-j\frac{2\pi}{Q(L+1)}lQd} = e^{-j\frac{2\pi}{(L+1)}ld}$, $\boldsymbol{W}_{L+1}$ is unitary

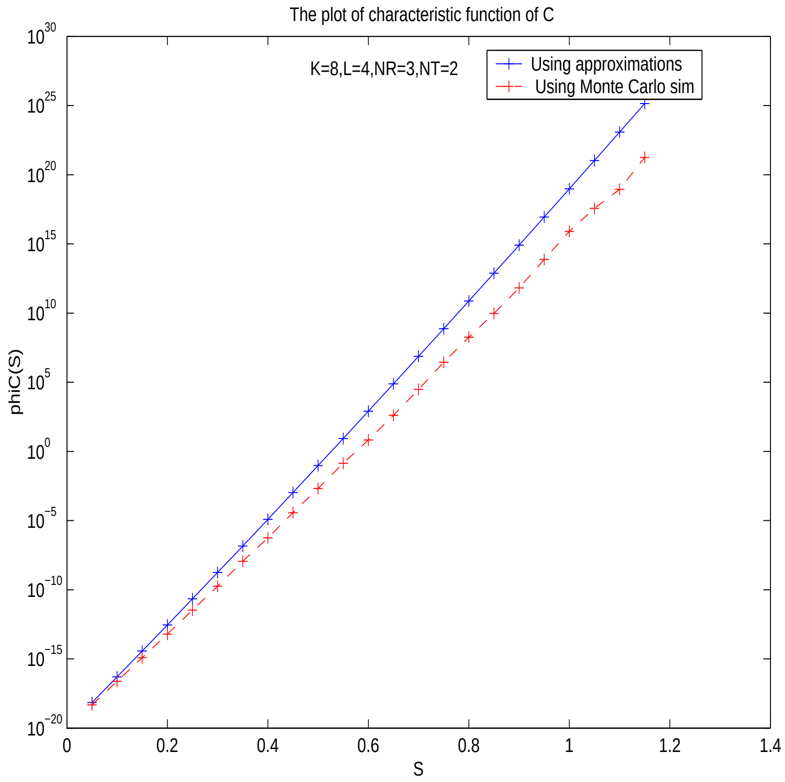
a4. $R_H(l_1, l_2) = R_H(l_2 - l_1)$

$$\Phi_C(s) \approx \gamma^{Qsn}\left(\prod_{l=0}^{L}(\sigma^2[lQ])^{Qs+m-\frac{n(n+1)}{2}}\chi_1(l)\right)\prod_{i=1}^{n}(\Gamma(i)\Gamma(m-n+Qs+i))^{L+1},$$

- Outage Capacity through Chernoff bound

Numerical versus Theory plots





Conclusion

- The study of Random Matrices has produced several beautiful results that have immediate application in Communication systems analysis