

# Random Matrices, Integrals and Space-time Systems

Babak Hassibi

California Institute of Technology

DIMACS Workshop on Algebraic Coding and Information  
Theory, Dec 15-18, 2003

# Outline

- Overview of multi-antenna systems
- Random matrices
- Rotational-invariance
- Eigendistributions
- Orthogonal polynomials
- Some important integrals
- Applications
- Open problems

# Introduction

We will be interested in multi-antenna systems of the form:

$$X = \sqrt{\frac{\rho}{M}} SH + V,$$

where  $X \in \mathbf{C}^{T \times N}$ ,  $S \in \mathbf{C}^{T \times M}$ ,  $H \in \mathbf{C}^{M \times N}$ ,  $V \in \mathbf{C}^{T \times N}$  are the receive, transmit, channel, and noise matrices, respectively. Moreover,  $M, N$  are the number of transmit/receive antennas respectively,  $T$  is the coherence interval and  $\rho$  is the SNR.

The entries of  $V$  are iid  $CN(0,1)$  and the entries of  $H$  are also  $CN(0,1)$ , but they may be correlated.

# Some Questions

We will be interested in two cases. The *coherent* case, where  $H$  is *known* to the receiver and the *non-coherent* case, where  $H$  is *unknown* to the receiver.

The following questions are natural to ask.

- What is the capacity?
- What are the capacity-achieving input distributions?
- For specific input distributions, what is the mutual information and/or cut-off rates?
- What are the (pairwise) probability of errors?

# Random Matrices

A  $m \times n$  random matrix  $A$  is simply described by the joint pdf of its entries,

$$p(A) = p(a_{ij}; i = 1, \dots, m; j = 1, \dots, n)$$

An example is the family of Gaussian random matrices, where the entries are jointly Gaussian.

# Rotational-Invariance

An important class of random matrices are (left- and right-) *rotationally-invariant* ones, with the property that their pdf is invariant to (pre- and post-) multiplication by any  $m \times m$  and  $n \times n$  unitary matrices  $\Theta$  and  $\Psi$ .

$$p(\Theta A) = p(A), \quad \Theta \Theta^* = \Theta^* \Theta = I_m$$

and

$$p(A \Psi) = p(A), \quad \Psi \Psi^* = \Psi^* \Psi = I_n$$

If a random matrix is both right- and left- rotationally-invariant we will simply call it *isotropically-random* (i.r.).

If  $G$  is a random matrix with iid Gaussian entries, then it is i.r., as are all of the matrices:

$$GG^*, G + G^*, G^p, G^{-1}, G_1 G_2, G_1 G_2^{-1}, G_1 (G_2^* G_2)^{-1} G_1^*, G_1 A G_1^*$$

# Isotropically-Random Unitary Matrices

A random unitary matrix is one for which the pdf is given by

$$p(\Theta) = f(\Theta)\delta(\Theta\Theta^* - I_m).$$

When the unitary matrix is i.r., then it is not hard to show that

$$p(\Theta) = \frac{\Gamma(m)\dots\Gamma(1)}{\pi^{m(m+1)/2}} \delta(\Theta\Theta^* - I_m).$$

Therefore an i.r. unitary matrix has a uniform distribution over the *Stiefel manifold* (space of unitary matrices). It is also called the *Haar* measure.

# A Fourier Representation

If we denote the columns of  $\Theta$  by  $\theta_k, k = 1, \dots, m$ , then

$$\delta(\Theta^* \Theta - I_m) = \prod_{k \geq l} \delta(\operatorname{Re}(\theta_k^* \theta_l) - \delta_{kl}) \delta(\operatorname{Im}(\theta_k^* \theta_l) - \delta_{kl})$$

Using the Fourier representation of the delta function

$$\delta(x) = \frac{1}{2\pi} \int d\omega e^{j\omega x}$$

It follows that we can write

$$\delta(\Theta^* \Theta - I_m) = \frac{\Gamma(m) \dots \Gamma(1)}{2^m \pi^{m(3m+1)/2}} \int_{\Omega = \Omega^*} d\Omega e^{jtr \Omega (\Theta^* \Theta - I_m)}$$



# A Few Theorems

I.r. unitary matrices come up in many applications.

**Theorem 1** *Let  $A$  be an  $m \times n$  i.r. random matrix and consider the svd  $A = U\Sigma V^*$ . Then the following two equivalent statements hold:*

- 1.  $U, \Sigma, V$  are independent random matrices and  $U$  and  $V$  are i.r. unitary.*
- 2. The pdf of  $A$  only depends on  $\Sigma$  :*

$$p(A) = f(\Sigma).$$

**Idea of Proof:**  $\Theta A \Psi^*$  and  $A$  have the same distribution for any unitary  $\Theta$  and  $\Psi$ ...

**Theorem 2** *Let  $A$  be an i.r. Hermitian matrix and consider the eigendecomposition  $A = U\Lambda U^*$ . Then the following two equivalent statements are true.*

- 1.  $U, \Lambda$  are independent random matrices and  $U$  is i.r. unitary.*
- 2. The pdf of  $A$  is independent of  $U$ :*

$$p(A) = f(\Lambda).$$

**Theorem 3** *Let  $A$  be a left rotationally-invariant random matrix and consider the QR decomposition,  $A=QR$ . Then the matrices  $Q$  and  $R$  are independent and  $Q$  is i.r. unitary.*

# Some Jacobians

The decompositions  $A = U\Lambda U^*$  and  $A = QR$  can be considered as coordinate transformations. Their corresponding Jacobians can be computed to be:

$$dA = dU d\Lambda \frac{1}{m!} \prod_{k>l} (\lambda_k - \lambda_l)^2 \delta(UU^* - I_m)$$

and

$$dA = dQ dR c \prod_{k=1}^m r_{kk}^{m-k} \delta(QQ^* - I_m)$$

for some constant  $c$ .

Note that both Jacobians are independent of  $U$  and  $Q$ .

# Eigendistributions

Thus for an i.r. Hermitian  $A$  with pdf  $p_A(\cdot)$ , we have

$$p(U, \Lambda) = p_A(\Lambda) \frac{1}{m!} \prod_{k>l} (\lambda_k - \lambda_l)^2 \delta(UU^* - I_m).$$

Integrating out the eigenvectors yields:

**Theorem 4** *Let  $A$  be an i.r. Hermitian matrix with pdf  $p_A(\cdot)$ .  
Then*

$$p(\Lambda) = \frac{\pi^{m(m-1)}}{\Gamma(m+1)\dots\Gamma(1)} p_A(\Lambda) \prod_{k>l} (\lambda_k - \lambda_l)^2$$

Note that  $\prod_{k>l} (\lambda_k - \lambda_l)^2 = \det^2 V(\Lambda)$ , a Vandermonde determinant.

# Some Examples

- Wishart matrices,  $A = G^* G$ , where  $G$  is  $m \times n, m \geq n$ .

$$p(\Lambda) = c \prod_{k=1}^n \lambda_k^{m-n} e^{-\frac{1}{2}\lambda_k} \det^2 V(\Lambda).$$

- Ratio of Wishart matrices,  $A = A_1 A_2^{-1}$  :

$$p(\Lambda) = c \prod_{k=1}^n \left( \frac{1}{1 + \lambda_k} \right)^{2n} \det^2 V(\Lambda).$$

- I.r. unitary matrix. Eigenvalues are on the unit circle and the distribution of the phases are:

$$p(\alpha_1, \dots, \alpha_m) = c \prod_{k>l} \sin^2(\alpha_k - \alpha_l).$$

# The Marginal Distribution

Note that all the previous eigendistributions were of the form:

$$p(\Lambda) = c \prod_{k=1}^m f(\lambda_k) \det^2 V(\Lambda).$$

For such pdf's the marginal can be computed using an elegant trick due to Wigner.

Define the Hankel matrix

$$F = \int d\lambda f(\lambda) \begin{bmatrix} 1 \\ \vdots \\ \lambda^{m-1} \end{bmatrix} \begin{bmatrix} 1 & \dots & \lambda^{m-1} \end{bmatrix}$$

Note that  $F \geq 0$ . Assume that  $F > 0$ . Then we can perform the Cholesky decomposition  $F=LL^*$ , with L lower triangular.

Note that  $L^{-1}FL^{-*} = I_m$  implies that the polynomials

$$\begin{bmatrix} g_0(\lambda) \\ \vdots \\ g_{m-1}(\lambda) \end{bmatrix} = L^{-1} \begin{bmatrix} 1 \\ \vdots \\ \lambda^{m-1} \end{bmatrix}$$

are orthonormal wrt to the weighting function  $f(\cdot)$ :

$$\int d\lambda f(\lambda) g_k(\lambda) g_l(\lambda) = \delta_{kl}.$$

Now the marginal distribution of one eigenvalue is given by

$$\begin{aligned} p(\lambda = \lambda_1) &= c \int d\lambda_2 \cdots d\lambda_m \prod_{k=1}^m f(\lambda_k) \det^2 V(\Lambda) \\ &= \frac{c}{\det^2 L^{-1}} \int d\lambda_2 \cdots d\lambda_m \prod_{k=1}^m f(\lambda_k) (\det L^{-1} V(\Lambda))^2 \end{aligned}$$

But

$$L^{-1}V(\Lambda) = L^{-1} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \cdots & \lambda_m^{m-1} \end{bmatrix} = \begin{bmatrix} g_0(\lambda_1) & \cdots & g_0(\lambda_m) \\ \vdots & \ddots & \vdots \\ g_{m-1}(\lambda_1) & \cdots & g_{m-1}(\lambda_m) \end{bmatrix} = V_G(\Lambda)$$

Now upon expanding out  $\det^2 V_G(\Lambda)$  and integrating over the variables  $\lambda_2 \cdots \lambda_m$  the only terms that do not vanish are those for which the indices of the orthonormal polynomials coincide.

Thus, after the smoke clears

$$p(\lambda) = c \sum_{k=0}^{m-1} f(\lambda) g_k^2(\lambda).$$

In fact, we have the following result.

**Theorem 5** *Let  $A$  be an i.r. Hermitian matrix with  $p_A(A) = \prod f(\lambda_k)$ . Then the marginal distribution of the eigenvalues of  $A$  is*

$$p(\lambda) = \frac{1}{m} \sum_{k=0}^{m-1} f(\lambda) g_k^2(\lambda).$$



# Orthogonal Polynomials

- What was just described was the connection between random matrices and orthogonal polynomials.
- For Wishart matrices, Laguerre polynomials arise. For ratios of Wishart matrices it is Jacobi polynomials, and for i.r. unitary matrices it is the complex exponential functions (orthogonal on the unit circle).
- Theorem 5 gives a Christoffel-Darboux sum and so

$$p(\lambda) = \frac{f(\lambda)}{m} \cdot \frac{a_{m-1}}{a_m} (g'_m(\lambda)g_{m-1}(\lambda) - g'_{m-1}(\lambda)g_m(\lambda))$$

- The above sum gives a uniform way to obtain the asymptotic distribution of the marginal pdf and to obtain results such as Wigner's semi-circle law.

# Remark

The attentive audience will have discerned that my choice of the Cholesky factorization of  $F$  and the resulting orthogonal polynomials was rather arbitrary.

It is possible to find the marginal distribution without resorting to orthogonal polynomials. The result is given below.

$$p(\lambda) = \frac{1}{m} f(\lambda) [1 \quad \dots \quad \lambda^{m-1}] F^{-1} \begin{bmatrix} 1 \\ \vdots \\ \lambda^{m-1} \end{bmatrix}$$

# Coherent Channels

Let us now return to the multi-antenna model

$$X = \sqrt{\frac{\rho}{M}} SH + V,$$

where we will assume that the channel  $H$  is known. We will assume that  $H = D_t G D_r$  where  $D_t, D_r$  are the correlation matrices at the transmitter and receiver and  $G$  has iid  $\text{CN}(0,1)$  entries. Note that  $D_t, D_r$  can be assumed diagonal wlog.

According to Foschini&Telatar:

1. When  $D_t = I_M, D_r = I_N$  :

$$C = E \log \det \left( I_N + \frac{\rho}{M} G^* G \right) = E \sum \log \left( 1 + \rho \lambda \left( \frac{G^* G}{M} \right) \right)$$

2. When  $D_t = I_M$  :

$$C = E \log \det(I_M + \frac{\rho}{M} GD_r^2 G^*) = E \sum \log(1 + \rho \lambda(\frac{GD_r^2 G^*}{M}))$$

3. When  $D_r = I_N$  :

$$C = \max_{tr(P)=M} E \log \det(I_N + \frac{\rho}{M} G^* D_t P D_t G) = \max_{tr(P)=M} \sum \log(1 + \rho \lambda(\frac{G^* D_t P D_t G}{M}))$$

4. In the general case:

$$C = \max_{tr(P)=M} E \sum \log(1 + \rho \lambda(\frac{D_r G^* D_t P D_t G D_r}{M}))$$

Cases 1-3 are readily dealt with using the techniques developed so far, since the matrices are rotationally-invariant.

Therefore we will do something more interesting and compute the characteristic function (not just the mean). This requires more machinery, as does Case 4, which we now develop.

# A Useful Integral Formula

Using a generalization of the technique used to prove Theorem 5, we can show the following result.

**Theorem 6** *Let functions  $f(\lambda), g_k(\lambda), h_k(\lambda), k = 0, \dots, m-1$  be given and define the matrices*

$$V_G(\Lambda) = \begin{bmatrix} g_0(\lambda_1) & \cdots & g_0(\lambda_m) \\ \vdots & \ddots & \vdots \\ g_{m-1}(\lambda_1) & \cdots & g_{m-1}(\lambda_m) \end{bmatrix}, V_H(\Lambda) = \begin{bmatrix} h_0(\lambda_1) & \cdots & h_0(\lambda_m) \\ \vdots & \ddots & \vdots \\ h_{m-1}(\lambda_1) & \cdots & h_{m-1}(\lambda_m) \end{bmatrix}$$

Then

$$\int d\Lambda \prod_{k=1}^m f(\lambda_k) \det V_G(\Lambda) \det V_H(\Lambda) = m! \det F$$

where

$$F = \int d\lambda f(\lambda) \begin{bmatrix} g_0(\lambda) \\ \vdots \\ g_{m-1}(\lambda) \end{bmatrix} [h_0(\lambda) \quad \cdots \quad h_{m-1}(\lambda)].$$

Theorem 6 was apparently first shown by Andreief in 1883.

A useful generalization has been noted in Chiani, Win and Zanella (2003).

**Theorem 7** *Let functions  $f_k(\lambda), g_k(\lambda), h_k(\lambda), k = 0, \dots, m-1$  be given. Then*

$$\int d\Lambda \prod_{k=1}^m f_k(\lambda_k) \det V_G(\Lambda) \det V_H(\Lambda) = \text{Tensor} \left( \int f_i(\lambda) g_j(\lambda) h_k(\lambda) \right)$$

*where for the tensor  $A_{ijk}$  we have defined*

$$\text{Tensor}(A) = \sum_{\mu} \text{sgn}(\mu) \sum_{\alpha} \text{sgn}(\alpha) \prod_{k=1}^m A_{\mu_k \alpha_k k}$$

*and the sums are over all possible permutations  $\mu, \alpha$  of the integers 1 to  $m$ .*

# An Exponential Integral

**Theorem 8 (Itzyskon and Zuber, 1990)** *Let  $A$  and  $B$  be  $m$ -dimensional diagonal matrices. Then*

$$\int d\Theta e^{\text{tr}\Theta A \Theta^* B} \delta(\Theta \Theta^* - I_m) = \Gamma(m+1) \cdots \Gamma(1) \frac{\det E(A, B)}{\det V(A) \det V(B)}$$

where

$$E(A, B) = \begin{bmatrix} e^{a_1 b_1} & \cdots & e^{a_1 b_m} \\ \vdots & \ddots & \vdots \\ e^{a_m b_1} & \cdots & e^{a_m b_m} \end{bmatrix}$$

**Idea of Proof:** Use induction. Start by partitioning

$$\Theta = [\Theta_1 \quad \mathcal{G}], A = \begin{bmatrix} A_1 & \\ & a_m \end{bmatrix}$$

Then rewrite  $tr(\Theta A \Theta^* B) = tr(\Theta_1 (A_1 - a_m I_{m-1}) \Theta_1^* B) + a_m tr B$  so that the desired integral becomes

$$\begin{aligned}
\int d\Theta e^{tr \Theta A \Theta^* B} \delta(\Theta \Theta^* - I_m) &= e^{a_m tr B} \int d\Theta_1 e^{tr \Theta_1 A' \Theta_1^* B} \delta(\Theta_1^* \Theta_1 - I_{m-1}) \\
&= c e^{a_m tr B} \int d\Theta_1 d\Omega e^{tr \Theta_1 A' \Theta_1^* B + j tr \Omega (\Theta_1^* \Theta_1 - I_{m-1})} \\
&= c e^{a_m tr B} \int d\Omega \frac{e^{-j tr \Omega}}{\prod_{k=1}^m \det(b_k A' + j\Omega)} \\
&= c e^{a_m tr B} \int dW \frac{e^{-j tr A' W}}{\prod_{k=1}^m \det(b_k I_{m-1} + jW)} \\
&= c e^{a_m tr B} \int dU d\Lambda \frac{e^{-j tr A' U \Lambda U^*}}{\prod_{l=1}^{m-1} \prod_{k=1}^m (b_k + j\lambda_l)} \det^2 V(\Lambda) \delta(UU^* - I_{m-1})
\end{aligned}$$



The last integral is over an  $(m-1)$ -dimensional i.r. matrix. And so if use the integral formula (at the lower dimension) to do the integral over  $U$ , we get

$$= c \frac{e^{a_m \text{tr} B}}{\det V(A')} \int d\Lambda \prod_{l=1}^{m-1} \frac{1}{\prod_{k=1}^m (b_k + j\lambda_l)} \det E(A', \Lambda) \det V(\Lambda)$$

An application of Theorem 6 now gives the result.

# Characteristic Function

Consider  $C = E \log \det(I_M + \frac{\rho}{M} GDG^*) = E \log \det(I_N + \frac{\rho}{M} DGG^*)$

The characteristic function is (assuming  $M=N$ )

$$\begin{aligned}
 E e^{j\omega \log \det(I_N + \frac{\rho}{M} DGG^*)} &= E \det(I_N + \frac{\rho}{M} DGG^*)^{j\omega} \\
 &= c \int dW \det(I_N + \frac{\rho}{M} DW)^{j\omega} e^{-trW} \\
 &= \frac{c}{\det^m(D)} \int dW \det(I_N + \frac{\rho}{M} W)^{j\omega} e^{-trWD^{-1}} \\
 &= \frac{c}{\det^m(D)} \int dU d\Lambda \prod_{k=1}^M (1 + \frac{\rho}{M} \lambda_k) e^{-trU\Lambda U^* D^{-1}} \det^2 V(\Lambda) \delta(UU^* - I_M)
 \end{aligned}$$

Successive use of Theorems 6 and 8 give the result.

# Non-coherent Channels

Let us now consider the non-coherent channel.

$$X = \sqrt{\frac{\rho}{M}} SH + V,$$

where  $H$  is unknown and has iid  $CN(0,1)$  entries.

**Theorem 9 (Hochwald and Marzetta, 1998)** *The capacity-achieving distribution is given by  $S = UD$ , where  $U$  is  $T$ -by- $M$  i.r. unitary and  $D$  is an independent diagonal.*

**Idea of Proof:** Write  $S = UDV^*$ .  $V^*$  can be absorbed in  $H$  and so is not needed. Optimal  $S$  is left rotationally-invariant.

# Mutual Information

Determining the optimal distribution on  $D$  is an open problem. However, given  $D$ , one can compute all quantities of interest. The starting point is

$$\begin{aligned}
 p(X | U, D) &= \frac{e^{-\text{tr}X^* (I_T + \frac{\rho}{M} UD^2 U^*)^{-1} X}}{\pi^{TN} \det^N (I_T + \frac{\rho}{M} UD^2 U^*)} \\
 &= \frac{e^{-\text{tr}X^* X + \text{tr}X^* U (I_M + \frac{M}{\rho} D^{-2})^{-1} U^* X}}{\pi^{TN} \det^N (I_M + \frac{\rho}{M} D^2)}
 \end{aligned}$$

The expectation over  $U$  is now readily do-able to give  $p(X|D)$ . (A little tricky since  $U$  is not square, but doable using Fourier Representation of delta functions and Theorems 6 and 8.)

# Other Problems

- Mutual information for almost any input distribution on  $D$  can be computed.
- Cut-off rates for coherent and non-coherent channels for many input distributions (Gaussian, i.r. unitary, etc.) can be computed.
- Characteristic function for coherent channel capacity in general case can be computed.
- Sum rate capacity of MIMO broadcast channel in some special cases can be computed.
- Diversity of *distributed space-time* coding in wireless networks can be determined.

# Other Work and Open Problems

- I did not touch at all upon asymptotic analysis using the Stieltjes transform.
- Open problem include determining the optimal input distribution for the non-coherent channel and finding the optimal power allocation for coherent channels when there is correlation among the transmit antennas.