Theory of Large Dimensional Random Matrices for Engineers (Part I)

Antonia M. Tulino
Università degli Studi di Napoli, "Federico II"

The 9th International Symposium on Spread Spectrum Techniques and Applications,
Manaus, Amazon, Brazil,
August 28-31, 2006

Outline

- A brief historical tour of the main results in random matrix theory.

Outline

- A brief historical tour of the main results in random matrix theory.
- Overview some of the main transforms.

Outline

- A brief historical tour of the main results in random matrix theory.
- Overview some of the main transforms.
- Fundamental limits of wireless communications: basic channels.

Outline

- A brief historical tour of the main results in random matrix theory.
- Overview some of the main transforms.
- Fundamental limits of wireless communications: basic channels.
- Performance measures of engineering interest (Signal Processing / Information Theory).

Outline

- A brief historical tour of the main results in random matrix theory.
- Overview some of the main transforms.
- Fundamental limits of wireless communications: basic channels.
- Performance measures of engineering interest (Signal Processing / Information Theory).
- Stieltjes transform, and its role in understanding eigenvalues of random matrices (Part II)

Outline

- A brief historical tour of the main results in random matrix theory.
- Overview some of the main transforms.
- Fundamental limits of wireless communications: basic channels.
- Performance measures of engineering interest (Signal Processing / Information Theory).
- Stieltjes transform, and its role in understanding eigenvalues of random matrices (Part II)
- Limit theorems of three classes of random matrices (Part II)

Outline

- A brief historical tour of the main results in random matrix theory.
- Overview some of the main transforms.
- Fundamental limits of wireless communications: basic channels.
- Performance measures of engineering interest (Signal Processing / Information Theory).
- Stieltjes transform, and its role in understanding eigenvalues of random matrices (Part II)
- Limit theorems of three classes of random matrices (Part II)
- Proof of one of the theorems (Part II)

Introduction

- Today random matrices find applications in fields as diverse as the Riemann hypothesis, stochastic differential equations, statistical physics, chaotic systems, numerical linear algebra, neural networks, etc.

- Random matrices are also finding an increasing number of applications in the context of information theory and signal processing.

Random Matrices & Information Theory

- The applications in information theory include, among others:
  - ✓ Wireless communications channels
  - ✓ Learning and neural networks
  - ✓ Capacity of ad hoc networks
  - ✓ Speed of convergence of iterative algorithms for multiuser detection
  - ✓ Direction of arrival estimation in sensor arrays

- Earliest applications to wireless communication : works of Foschini and Telatar, in the mid-90s, on characterizing the capacity of multi-antenna channels.

A. M. Tulino and S. Verdú
"Random Matrices and Wireless Communications,"
Foundations and Trends in Communications and Information Theory,
vol. 1, no. 1, June 2004.

Wireless Channels

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

- $\mathbf{x}$ = K-dimensional complex-valued input vector,
- $\mathbf{y}$ = N-dimensional complex-valued output vector,
- $\mathbf{n}$ = N-dimensional additive Gaussian noise
- $\mathbf{H}$ = $N \times K$ random channel matrix known to the receiver

This model applies to a variety of communication problems by simply reinterpreting K, N, and $\mathbf{H}$

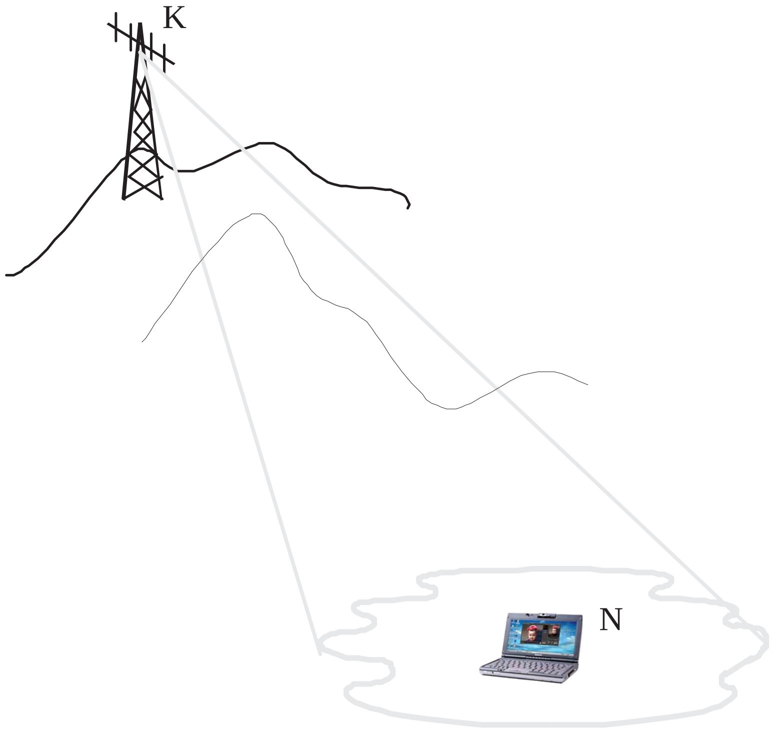
⋇ Fading

⋇ Wideband

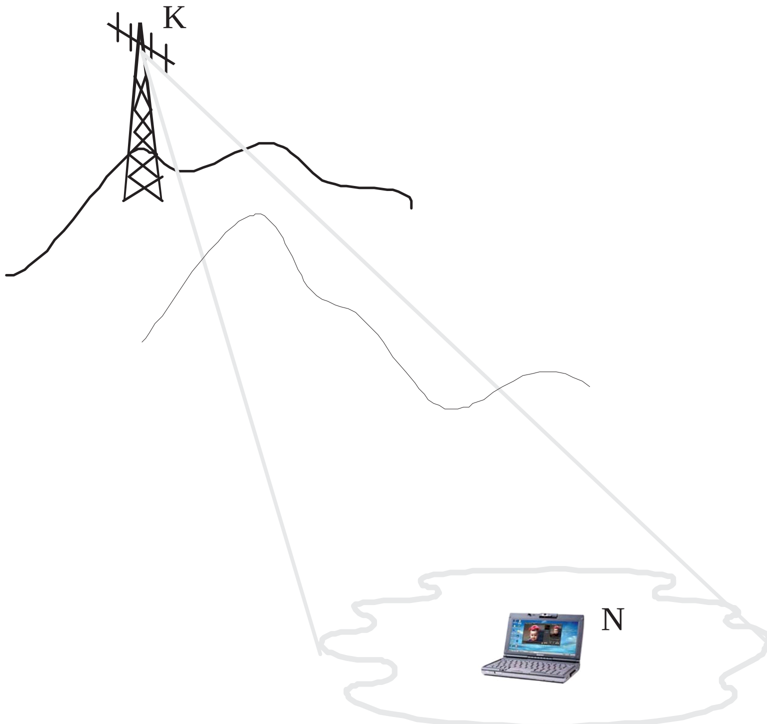
⋇ Multiuser

⋇ Multiantenna

Multi-Antenna channels



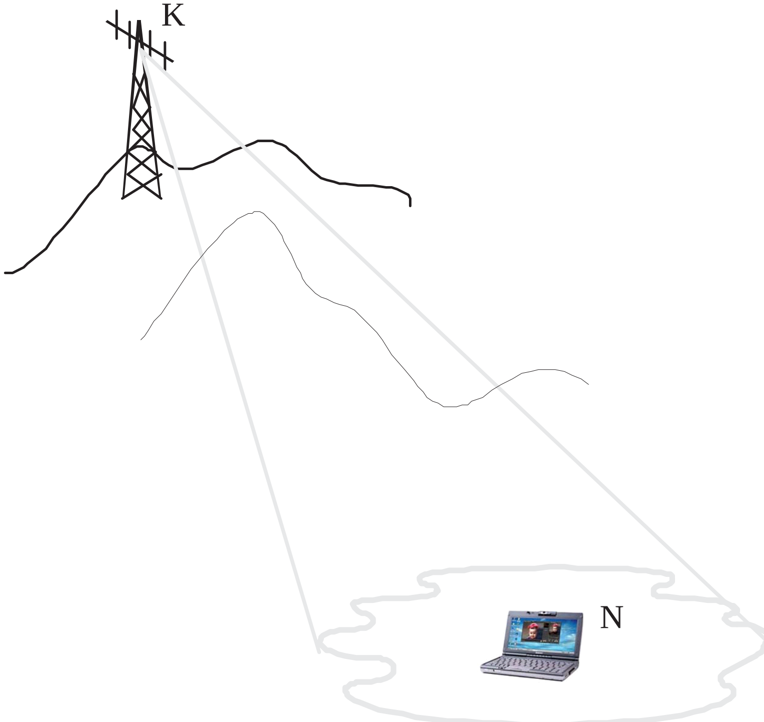
Multi-Antenna channels

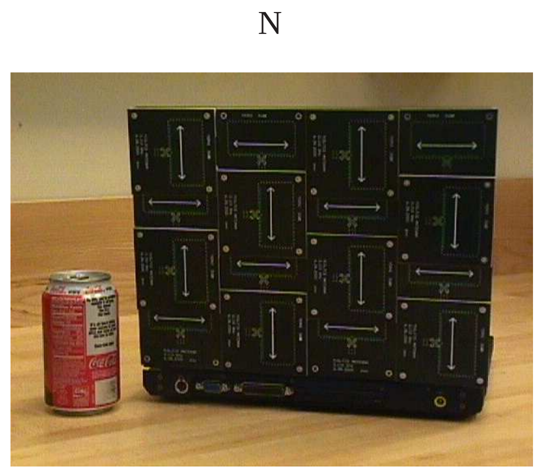


N

Prototype picture courtesy of
Bell Labs (Lucent Technologies)

Multi-Antenna channels





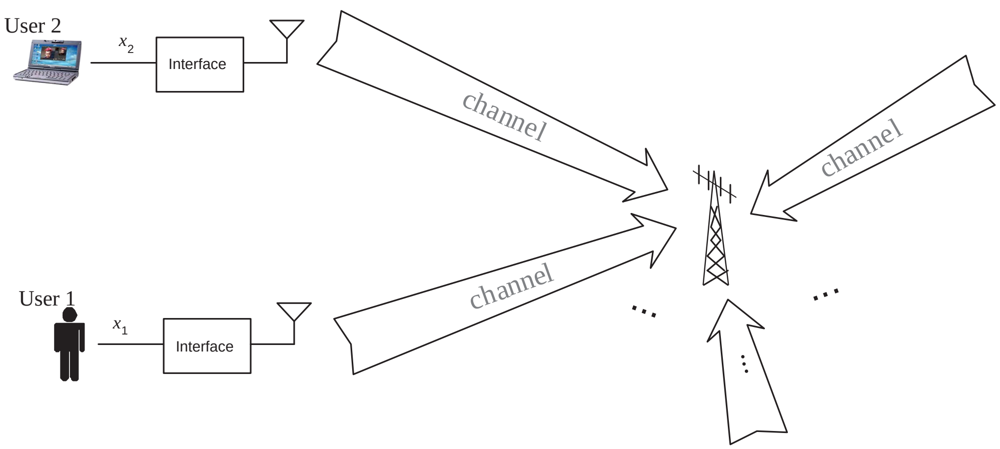
Prototype picture courtesy of Ball Labs (Lucent Technologies)

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

- K and N number of transmit and receive antennas
- $\mathbf{H}$ = propagation matrix: $N \times K$ complex matrix whose entries represent the gains between each transmit and each receive antenna.

CDMA (Code-Division Multiple Access) Channel

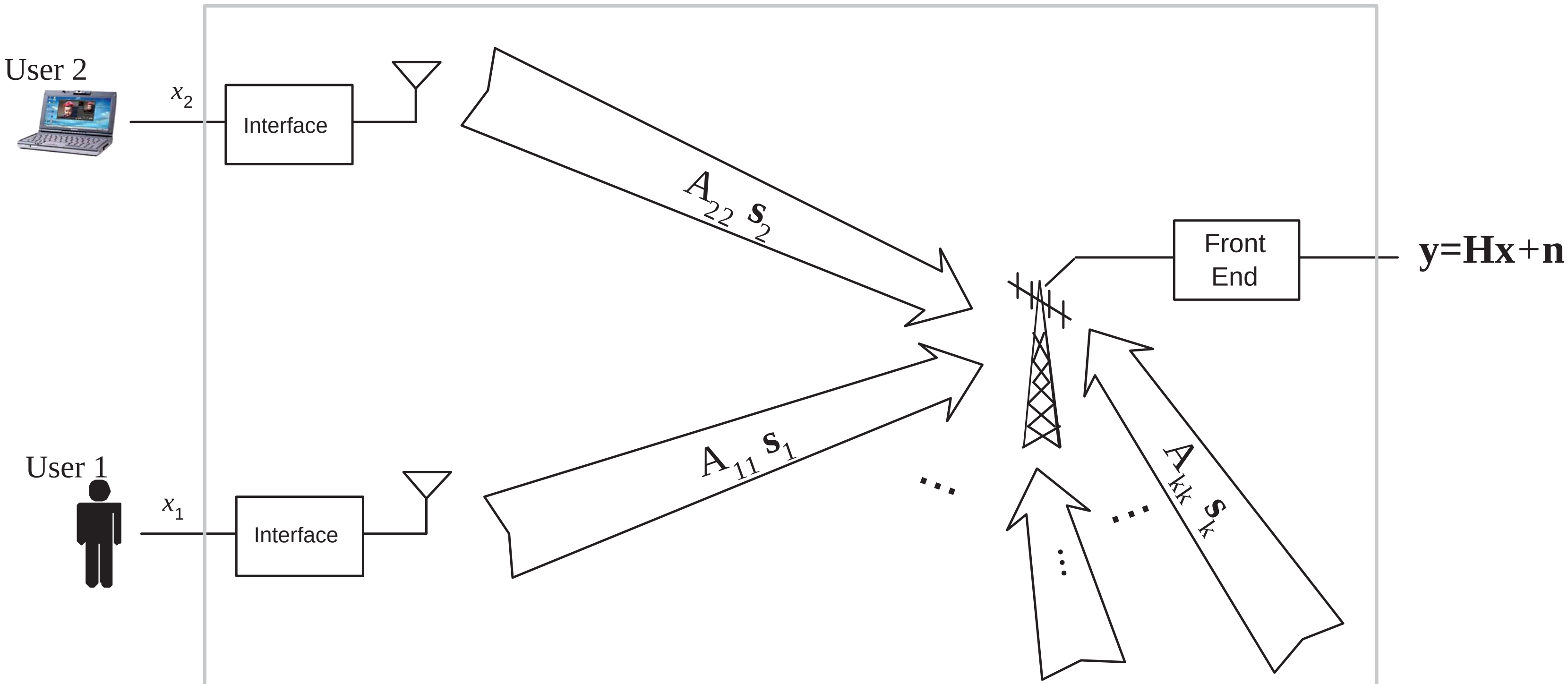
Signal space with N dimensions.
N= "spreading gain" = proportional to Bandwidth
Each user assigned a "signature vector" known at the receiver



- DS-CDMA (Direct sequence CDMA) used in many current cellular systems (IS-95, cdma2000, UMTS).

- MC-CDMA (Multi-Carrier CDMA) being considered for 4G (Fourth Generation) wireless.

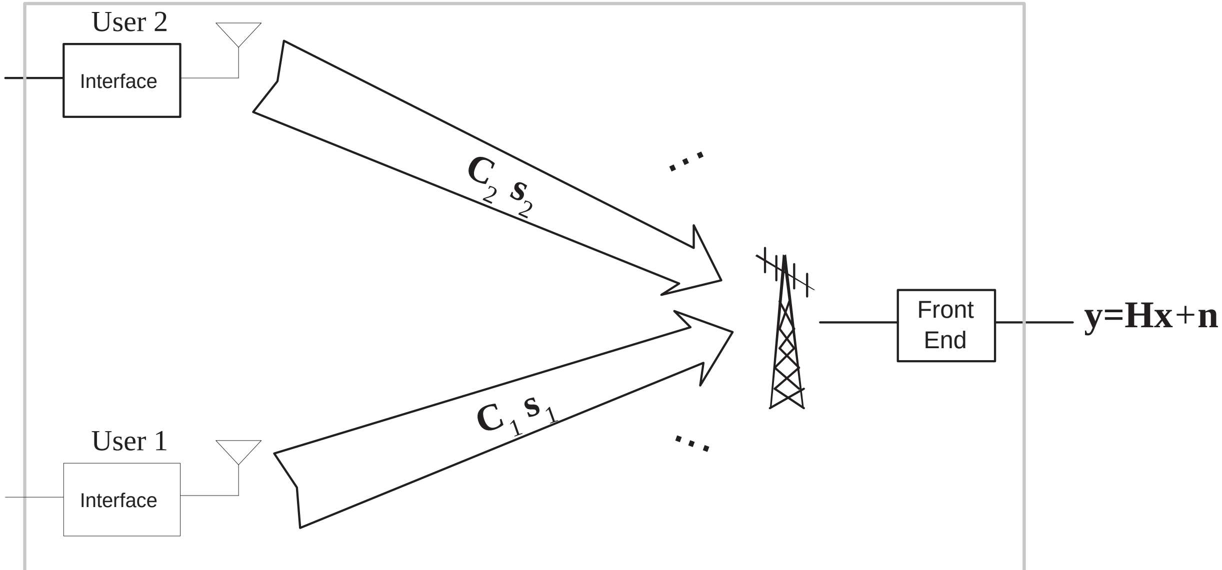
DS-CDMA Flat-faded Channel



$$\mathbf{y} = \underbrace{\mathbf{H}}_{\mathbf{SA}}\mathbf{x} + \mathbf{n} = \mathbf{SAx} + \mathbf{n}$$

- K= number of users; N= processing gain.
- $\mathbf{S} = [\mathbf{s}_1 \quad | \quad \ldots \quad | \quad \mathbf{s}_K]$ with $\mathbf{s}_k$ the signature vector of the k^{th} user.
- $\mathbf{A}$ is a $K \times K$ diagonal matrix containing the independent complex fading coefficients for each user.

Multi-Carrier CDMA (MC-CDMA)



$$\mathbf{y} = \underbrace{\mathbf{H}}_{\mathbf{G}\circ\mathbf{S}}\mathbf{x} + \mathbf{n} = \mathbf{G}\circ\mathbf{S}\mathbf{x} + \mathbf{n}$$

- K and N represent the number of users and of subcarriers.
- $\mathbf{H}$ incorporates both the spreading and the frequency-selective fading i.e.

$$h_{nk} = g_{nk}s_{nk} \qquad\qquad n = 1,\ldots,N \quad k = 1,\ldots,K$$

- $\mathbf{S}=[\mathbf{s}_1 \quad | \quad \ldots \quad | \quad \mathbf{s}_K]$ with $\mathbf{s}_k$ the signature vector of the k^{th} user.
- $\mathbf{G}=[\mathbf{g}_1 \quad | \quad \ldots \quad | \quad \mathbf{g}_K]$ is an $N \times K$ matrix whose columns are independent N-dimensional random vectors.

Role of Singular Values
in
Wireless Communication

Empirical (Asymptotic) Spectral Distribution

Definition: The ESD (Empirical Spectral Distribution) of an $N \times N$ Hermitian random matrix $\mathbf{A}$, $\mathsf{F}_{\mathbf{A}}^{N}(\cdot)$,

$$\mathsf{F}_{\mathbf{A}}^{N}(x) = \frac{1}{N}\sum_{i=1}^{N} 1\{\lambda_i(\mathbf{A}) \leq x\}$$

where $\lambda_1(\mathbf{A}), \ldots, \lambda_N(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$.

If, as $N \to \infty$, $\mathsf{F}_{\mathbf{A}}^{N}(\cdot)$ converges almost surely (a.s), the corresponding limit (asymptotic ESD) is simply denoted by $\mathsf{F}_{\mathbf{A}}(\cdot)$.

$\bar{\mathsf{F}}_{\mathbf{A}}^{N}(\cdot)$ denotes the expected ESD.

Role of Singular Values: Mutual Information

$$\mathcal{I}(\textsf{SNR}) = \frac{1}{N}\log\det\left(\mathbf{I} + \textsf{SNR}\,\mathbf{H}\mathbf{H}^\dagger\right) = \frac{1}{N}\sum_{i=1}^{N}\log\left(1 + \textsf{SNR}\,\lambda_i(\mathbf{H}\mathbf{H}^\dagger)\right)$$

$$= \int_0^\infty \log\left(1 + \textsf{SNR}\,x\right) d\mathsf{F}^N_{\mathbf{H}\mathbf{H}^\dagger}(x)$$

with $\mathsf{F}^N_{\mathbf{H}\mathbf{H}^\dagger}(x)$ the ESD of $\mathbf{H}\mathbf{H}^\dagger$ and with

$$\textsf{SNR} = \frac{N E[\|\mathbf{x}\|^2]}{K E[\|\mathbf{n}\|^2]}$$

the signal-to-noise ratio, a key performance measure.

Role of Singular Values: Ergodic Mutual Information

In an ergodic time-varying channel,

$$\begin{aligned}\mathbb{E}[\mathcal{I}(\textsf{SNR})] &= \frac{1}{N}\mathbb{E}\left[\log\det\left(\mathbf{I} + \textsf{SNR}\,\mathbf{H}\mathbf{H}^\dagger\right)\right] \\ &= \int_0^\infty \log\left(1 + \textsf{SNR}\,x\right) d\bar{\mathsf{F}}^N_{\mathbf{H}\mathbf{H}^\dagger}(x)\end{aligned}$$

where $\bar{\mathsf{F}}^N_{\mathbf{H}\mathbf{H}^\dagger}(\cdot)$ denotes the expected ESD.

High-SNR Power Offset

For $\text{SNR} \to \infty$, a regime of interest in short-range applications, the mutual information behaves as

$$I(\text{SNR}) = S_\infty \left(\log \text{SNR} + \mathcal{L}_\infty\right) + o(1)$$

where the key measures are the high-SNR slope

$$S_\infty = \lim_{\text{SNR}\to\infty} \frac{I(\text{SNR})}{\log \text{SNR}}$$

which for most channels gives $S_\infty = \min\left\{\frac{K}{N}, 1\right\}$, and the *power offset*

$$\mathcal{L}_\infty = \lim_{\text{SNR}\to\infty} \log \text{SNR} - \frac{I(\text{SNR})}{S_\infty}$$

which essentially boils down to $\log\det(\mathbf{H}\mathbf{H}^\dagger)$ or $\log\det(\mathbf{H}^\dagger\mathbf{H})$ depending on whether $K > N$ or $K < N$.

Role of Singular Values: MMSE

The minimum mean-square error (MMSE) incurred in the estimation of the input $\mathbf{x}$ based on the noisy observation at the channel output $\mathbf{y}$ for an i.i.d. Gaussian input:

$$\text{MMSE} = \frac{1}{K} E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2] = \frac{1}{K} \sum_{k=1}^{K} E[|x_k - \hat{x}_k|^2] = \frac{1}{K} \sum_{k=1}^{K} \text{MMSE}_k$$

where $\hat{\mathbf{x}}$ is the estimate of $\mathbf{x}$. For an i.i.d Gaussian input,

$$\text{MMSE} = \frac{1}{K} \text{tr}\left(\left(\mathbf{I} + \text{SNR}\, \mathbf{H}^\dagger \mathbf{H} \right)^{-1} \right)$$

$$= \frac{1}{K} \sum_{i=1}^{K} \frac{1}{1 + \text{SNR}\, \lambda_i(\mathbf{H}^\dagger \mathbf{H})}$$

$$= \int_0^\infty \frac{1}{1 + \text{SNR}\, x} d\text{F}^K_{\mathbf{H}^\dagger \mathbf{H}}(x)$$

$$= \frac{N}{K} \int_0^\infty \frac{1}{1 + \text{SNR}\, x} d\text{F}^N_{\mathbf{H}\mathbf{H}^\dagger}(x) - \frac{N-K}{K}$$

In the Beginning ...

The Birth of (Nonasymptotic) Random Matrix Theory: (Wishart, 1928)

J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20 A, pp. 32–52, 1928.

Probability density function of the Wishart matrix:

$$\mathbf{H}\mathbf{H}^{\dagger} = \mathbf{h}_1\mathbf{h}_1^{\dagger} + \ldots + \mathbf{h}_n\mathbf{h}_n^{\dagger}$$

where $\mathbf{h}_i$ are iid zero-mean Gaussian vectors.

Wishart Matrices

Definition 1. *The* $m \times m$ *random matrix* $\mathbf{A} = \mathbf{H}\mathbf{H}^\dagger$ *is a (central)* real/complex Wishart *matrix with* n *degrees of freedom and covariance matrix* $\mathbf{\Sigma}$*, (*$\mathbf{A} \sim \mathcal{W}_m(n, \mathbf{\Sigma})$*), if the columns of the* $m \times n$ *matrix* $\mathbf{H}$ *are zero-mean independent real/complex Gaussian vectors with covariance matrix* $\mathbf{\Sigma}$*.*[1]

The p.d.f. of a complex Wishart matrix $\mathbf{A} \sim \mathcal{W}_m(n, \mathbf{\Sigma})$ for $n \geq m$ is

$$f_{\mathbf{A}}(\mathbf{B}) = \frac{\pi^{-m(m-1)/2}}{\det \mathbf{\Sigma}^n \prod_{i=1}^m (n-i)!} \exp\left[-\mathrm{tr}\left\{\mathbf{\Sigma}^{-1}\mathbf{B}\right\}\right] \det \mathbf{B}^{n-m}. \qquad (1)$$

[1] If the entries of $\mathbf{H}$ have nonzero mean, $\mathbf{H}\mathbf{H}^\dagger$ is a non-central Wishart matrix.

Singular Values^2: Fisher-Hsu-Girshick-Roy

The joint p.d.f. of the ordered strictly positive eigenvalues of the Wishart matrix $\mathbf{HH}^\dagger$:

- R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *The Annals of Eugenics*, vol. 9, pp. 238–249, 1939.
- M. A. Girshick, "On the sampling theory of roots of determinantal equations," *The Annals of Math. Statistics*, vol. 10, pp. 203–204, 1939.
- P. L. Hsu, "On the distribution of roots of certain determinantal equations," *The Annals of Eugenics*, vol. 9, pp. 250–258, 1939.
- S. N. Roy, "p-statistics or some generalizations in the analysis of variance appropriate to multivariate problems," *Sankhya*, vol. 4, pp. 381–396, 1939.

Singular Values^2: Fisher-Hsu-Girshick-Roy

- Joint distribution of ordered nonzero eigenvalues (Fisher in 1939, Hsu in 1939, Girshick in 1939, Roy in 1939):

$$\gamma_{t,r} \exp\left(-\sum_{i=1}^{t} \lambda_i\right) \prod_{i=1}^{t} \lambda_i^{r-t} \prod_{j=i+1}^{t} (\lambda_i - \lambda_j)^2$$

where t and r are the minimum and maximum of the dimensions of $\mathbf{H}$.

- The marginal p.d.f. of the unordered eigenvalues is

$$\sum_{k=0}^{t-1} \frac{k!}{(k+r-t)!} \left[L_k^{r-t}(\lambda)\right]^2 \lambda^{r-t} e^{-\lambda}$$

where the Laguerre polynomials are $L_k^n(\lambda) = \frac{1}{k!} e^{\lambda} \lambda^{-n} \frac{d^k}{d\lambda^k}\left(e^{-\lambda}\lambda^{n+k}\right)$.

Singular Values^2: Fisher-Hsu-Girshick-Roy

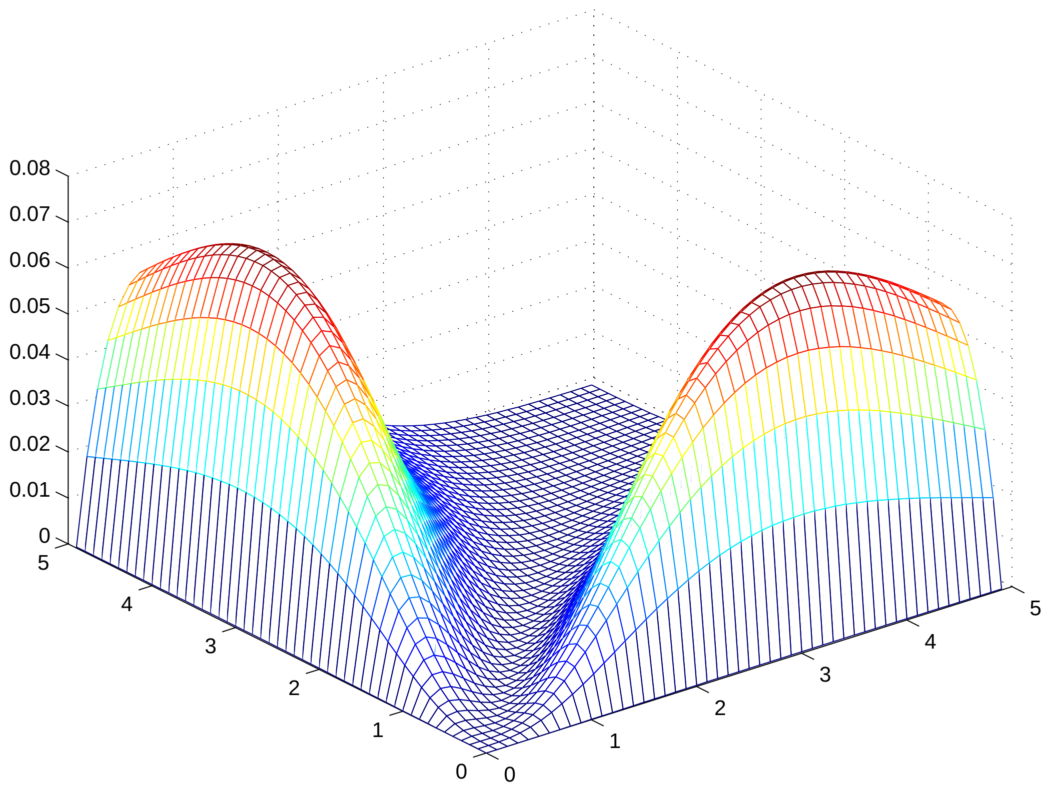


Figure 1: Joint p.d.f. of the unordered positive eigenvalues of the Wishart matrix $\mathbf{HH}^\dagger$ with $n = 3$ and $m = 2$.

Wishart Matrices: Eigenvectors

Theorem 1. *The matrix of eigenvectors of Wishart matrices is uniformly distributed on the manifold of unitary matrices (* Haar measure*)*

Unitarily invariant RMs

- Definition: An $N \times N$ self-adjoint random matrix $\mathbf{A}$ is called *unitarily invariant* if the p.d.f. of $\mathbf{A}$ is equal to that of $\mathbf{V}\mathbf{A}\mathbf{V}^\dagger$ for any unitary matrix $\mathbf{V}$.

- Property: If $\mathbf{A}$ is unitarily invariant, it admits the following eigenvalue decomposition:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger.$$

  with $\mathbf{U}$ and $\mathbf{\Lambda}$ independent.

- Example

  ⋇ A Wishart matrix is unitarily invariant.

  ⋇ $\mathbf{A} = \frac{1}{2}(\mathbf{H} + \mathbf{H}^\dagger)$ with $\mathbf{H}$ a $N \times N$ Gaussian matrix with i.i.d entries, is unitarily invariant.

  ⋇ $\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}$ with $\mathbf{U}$ Haar matrix and $\mathbf{B}$ independent on $\mathbf{U}$, is unitarily invariant.

Bi-Unitarily invariant RMs

- Definition: An $N \times N$ random matrix $\mathbf{A}$ is called *bi-unitarily invariant* if its p.d.f. equals that of $\mathbf{UAV}^\dagger$ for any unitary matrices $\mathbf{U}$ and $\mathbf{V}$.

- Property: If $\mathbf{A}$ is a bi-unitarily invariant RM, it has a polar decomposition $\mathbf{A} = \mathbf{UH}$ with

  ⋇ $\mathbf{U}$ $N \times N$ Haar RM.
  ⋇ $\mathbf{H}$ $N \times N$ unitarily invariant positive-definite RM.
  ⋇ $\mathbf{U}$ and $\mathbf{H}$ independent.

Example:

⋇ A complex Gaussian randon matrix with i.i.d. entries is bi-unitarily invariant.

⋇ An $N \times K$ matrix $\mathbf{Q}$ uniformly distributed over the Stiefel manifold of complex $N \times K$ matrices such that $\mathbf{QQ}^\dagger = \mathbf{I}$.

The Birth of Asymptotic Random Matrix Theory

E. Wigner, “Characteristic vectors of bordered matrices with infinite dimensions,” *The Annals of Mathematics*, vol. 62, pp. 546–564, 1955.

$$\mathbf{W} = \frac{1}{\sqrt{N}} \begin{bmatrix} 0 & +1 & +1 & -1 & -1 & +1 \\ +1 & 0 & -1 & -1 & +1 & +1 \\ +1 & -1 & 0 & +1 & +1 & -1 \\ -1 & -1 & +1 & 0 & +1 & +1 \\ -1 & +1 & +1 & +1 & 0 & -1 \\ +1 & +1 & -1 & +1 & -1 & 0 \end{bmatrix}$$

As the matrix dimension $N \to \infty$, the histogram of the eigenvalues converges to the *semicircle law*:

$$f(x) = \frac{1}{2\pi}\sqrt{4 - x^2}, \quad -2 < x < 2$$

Motivation: bypass the Schrödinger equation and explain the statistics of experimentally measured atomic energy levels in terms of the limiting spectrum of those random matrices.

Wigner Matrices: The Semicircle Law

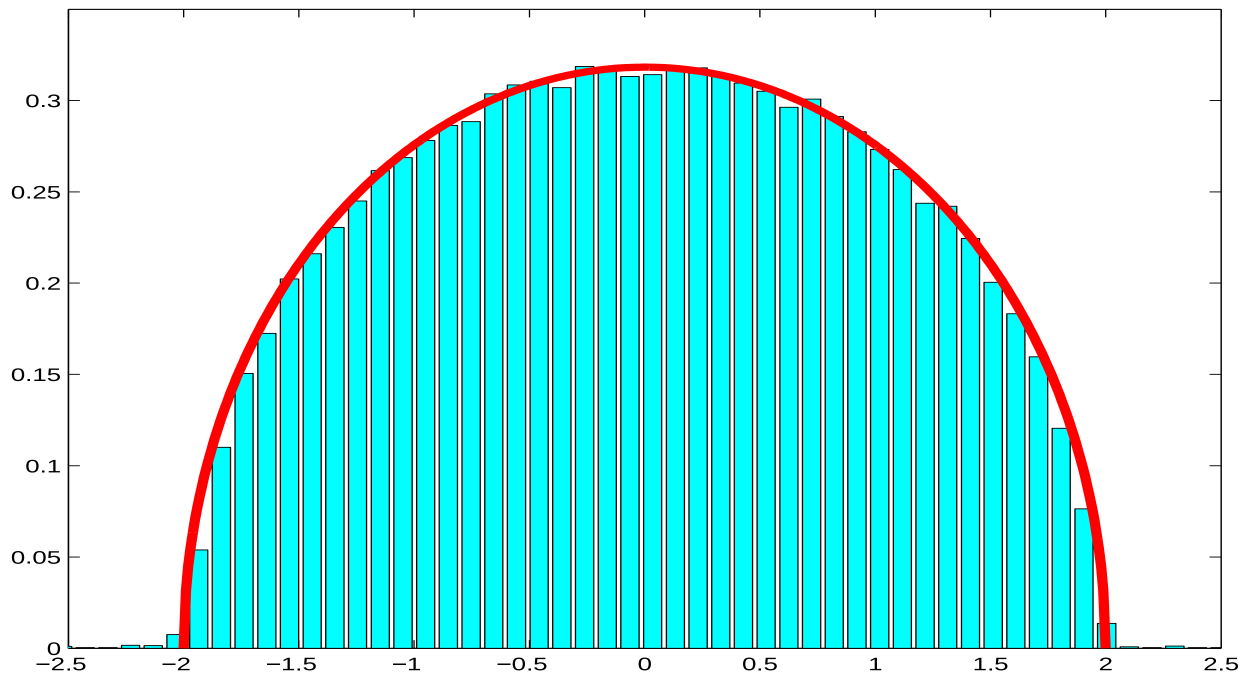
E. Wigner, "On the distribution of roots of certain symmetric matrices," *The Annals of Mathematics*, vol. 67, pp. 325–327, 1958.

If the upper-triangular entries are independent zero-mean random variables with variance $\frac{1}{N}$ (standard Wigner matrix) such that, for some constant κ, and sufficiently large N

$$\max_{1 \leq i \leq j \leq N} \mathbb{E}\left[|\mathsf{W}_{i,j}|^4\right] \leq \frac{\kappa}{N^2} \tag{2}$$

Then, the empirical distribution of $\mathbf{W}$ converges almost surely to the semicircle law

The Semicircle Law



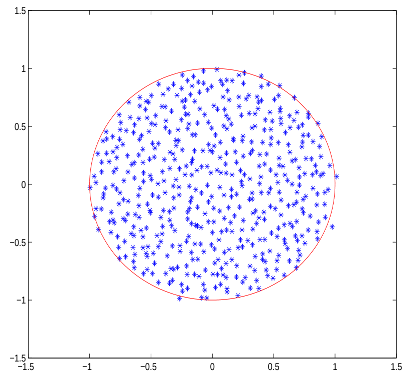
The semicircle law density function compared with the histogram of the average of 100 empirical density functions for a Wigner matrix of size $N = 10$.

Square matrix of iid coefficients

Girko (1984), *full-circle law* for the unsymmetrized matrix

$$\mathbf{H} = \frac{1}{\sqrt{N}} \begin{bmatrix} +1 & +1 & +1 & -1 & -1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 \\ +1 & -1 & -1 & -1 & +1 & +1 \\ -1 & -1 & +1 & -1 & -1 & -1 \\ -1 & -1 & +1 & +1 & +1 & +1 \end{bmatrix}$$

As $N \to \infty$, the eigenvalues of $\mathbf{H}$ are uniformly distributed on the unit disk.



The full-circle law and the eigenvalues of a realization of a 500×500 matrix

Full Circle Law

V. L. Girko, "Circular law," *Theory Prob. Appl.*, vol. 29, pp. 694–706, 1984.
Z. D. Bai, "The circle law," *The Annals of Probability*, pp. 494–529, 1997.

Theorem 2. *Let $\mathbf{H}$ be an $N \times N$ complex random matrix whose entries are independent random variables with identical mean and variance and finite kth moments for $k \geq 4$. Assume that the joint distributions of the real and imaginary parts of the entries have uniformly bounded densities. Then, the asymptotic spectrum of $\mathbf{H}$ converges almost surely to the circular law, namely the uniform distribution over the unit disk on the complex plane $\{\zeta \in \mathbb{C} : |\zeta| \leq 1\}$ whose density is given by*

$$\mathsf{f}_c(\zeta) = \frac{1}{\pi} \qquad |\zeta| \leq 1 \tag{3}$$

(also holds for real matrices replacing the assumption on the joint distribution of real and imaginary parts with the one-dimensional distribution of the real-valued entries.)

Elliptic Law (Sommers-Crisanti-Sompolinsky-Stein, 1988)

H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, Spectrum of large random asymmetric matrices, *Physical Review Letters*, vol. 60, pp. 1895- 1899, 1988.

If the off-diagonal entries are Gaussian and pairwise correlated with correlation coefficient ρ, the eigenvalues are asymptotically uniformly distributed on an ellipse in the complex plane whose axes coincide with the real and imaginary axes and have radii $1 + \rho$ and $1 - \rho$.

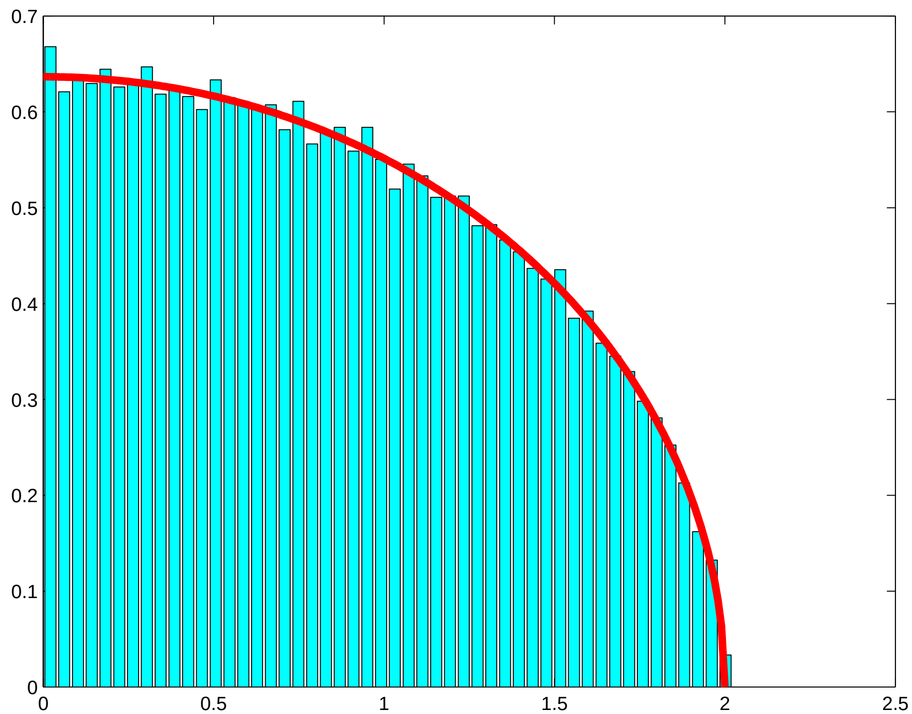
What About the Singular Values?

Asymptotic Distribution of Singular Values: Quarter circle law

Consider an $N \times N$ matrix $\mathbf{H}$ whose entries are independent zero-mean complex (or real) random variables with variance $\frac{1}{N}$, the asymptotic distribution of the singular values converges to

$$q(x) = \frac{1}{\pi}\sqrt{4 - x^2}, \quad 0 \leq x \leq 2 \tag{4}$$

Asymptotic Distribution of Singular Values: Quarter circle law



The quarter circle law compared a histogram of the average of 100 empirical singular value density functions of a matrix of size 100×100.

Minimum Singular Value of Gaussian Matrix

- A. Edelman, *Eigenvalues and condition number of random matrices*. PhD thesis, Dept. Mathematics, MIT, Cambridge, MA, 1989.
- J. Shen, "On the singular values of Gaussian random matrices," *Linear Algebra and its Applications*, vol. 326, no. 1-3, pp. 1–14, 2001.

Theorem 3. *The minimum singular value of an* $N \times N$ *standard complex Gaussian matrix* $\mathbf{H}$ *satisfies*

$$\lim_{N \to \infty} P[N\sigma_{\min} \geq x] = e^{-x-x^2/2}. \tag{5}$$

Marčenko-Pastur Law

- V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, pp. 457–483, 1967.

Rediscovering/Strenghtening the Marčenko-Pastur Law

- V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, pp. 457–483, 1967.

- U. Grenander and J. W. Silverstein, "Spectral analysis of networks with random topologies," *SIAM J. of Applied Mathematics*, vol. 32, pp. 449–519, 1977.

- K. W. Wachter, "The strong limits of random matrix spectra for sample matrices of independent elements," *The Annals of Probability*, vol. 6, no. 1, pp. 1–18, 1978.

- J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. of Multivariate Analysis*, vol. 54, pp. 175–192, 1995.

- Y. L. Cun, I. Kanter, and S. A. Solla, "Eigenvalues of covariance matrices: Application to neural-network learning," *Physical Review Letters*, vol. 66, pp. 2396–2399, 1991.

Marčenko-Pastur Law

V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, pp. 457–483, 1967.

If $N \times K$-matrix $\mathbf{H}$ has zero-mean i.i.d. entries with variance $\frac{1}{N}$, the asymptotic ESD of $\mathbf{H}\mathbf{H}^\dagger$ found in (Marčenko-Pastur, 1967) is

$$\tilde{\mathsf{f}}_\beta(x) = [1-\beta]^+ \, \delta(x) + \frac{\sqrt{[x-a]^+[b-x]^+}}{2\pi x}$$

where

$$[z]^+ = \max\{0, z\},$$

and

$$a = \left(1 - \sqrt{\beta}\right)^2 \qquad b = \left(1 + \sqrt{\beta}\right)^2.$$

$$\frac{K}{N} \to \beta$$

Marčenko-Pastur Law

V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, pp. 457–483, 1967.

If $N \times K$-matrix $\mathbf{H}$ has zero-mean i.i.d. entries with variance $\frac{1}{N}$, the asymptotic ESD of $\mathbf{H}\mathbf{H}^\dagger$ found in (Marčenko-Pastur, 1967) is

$$\tilde{\mathsf{f}}_\beta(x) = [1-\beta]^+ \, \delta(x) + \frac{\sqrt{[x-a]^+[b-x]^+}}{2\pi x}$$

(Bai 1999) The results also holds if only a unit second-moment condition is placed on the entries of $\mathbf{H}$ and

$$\frac{1}{K}\sum \mathbb{E}\left[|\mathsf{H}_{i,j}|^2 \, 1\left\{|\mathsf{H}_{i,j}| \geq \delta\right\}\right] \to 0$$

for any $\delta > 0$ (Lindeberg-type condition on the whole matrix).

Nonzero-Mean Matrices

Lemma: (Yin 1986, Bai 1999): For any $N \times K$ matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\sup_{x \geq 0} |\mathsf{F}^N_{\mathbf{A}\mathbf{A}^\dagger}(x) - \mathsf{F}^N_{\mathbf{B}\mathbf{B}^\dagger}(x)| \leq \frac{\mathrm{rank}(\mathbf{A} - \mathbf{B})}{N}.$$

Lemma: (Yin 1986, Bai 1999): For any $N \times N$ Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\sup_{x \geq 0} |\mathsf{F}^N_{\mathbf{A}}(x) - \mathsf{F}^N_{\mathbf{B}}(x)| \leq \frac{\mathrm{rank}(\mathbf{A} - \mathbf{B})}{N}.$$

Using these Lemmas, all results illustrated so far can be extended to matrices whose mean has rank r where $r > 1$ but such that

$$\lim_{N \to \infty} \frac{r}{N} = 0.$$

Generalizations needed!

- Correlated Entries

$$\mathbf{H} = \sqrt{\mathbf{\Phi}_{\mathsf{R}}}\mathbf{S}\sqrt{\mathbf{\Phi}_{\mathsf{T}}}$$

S: $N \times K$ matrix whose entries are independent complex random variables (arbitrarily distributed)

$\mathbf{\Phi}_{\mathsf{R}}$: $N \times N$ either deterministic or random matrix (whose asymptotic spectrum of converges a. s. to a compactly supported measure).

$\mathbf{\Phi}_{\mathsf{T}}$: $K \times K$ either deterministic or random matrix matrix whose asymptotic spectrum converges a. s. to a compactly supported measure.

- Non-identically Distributed Entries

$\mathbf{H}$ be an $N \times K$ complex random matrix with independent entries (arbitrarily distributed) with identical means.

$$\mathrm{Var}[\mathsf{H}_{i,j}] = \frac{\mathsf{G}_{i,j}}{N}$$

with $\mathsf{G}_{i,j}$ uniformly bounded.

Special case : Doubly Regular Channels

Transforms

1. Stieltjes transform
2. η transform
3. Shannon transform
4. R-transform
5. S-transform

The Stieltjes Transform

The Stieltjes transform (also called the Cauchy transform) of an arbitrary random variable X is defined as

$$S_X(z) = \mathbb{E}\left[\frac{1}{X - z}\right]$$

whose inversion formula was obtained in :
T. J. Stieltjes, “Recherches sur les fractions continues,” *Annales de la Faculte des Sciences de Toulouse*, vol. 8 (9), no. A (J), pp. 1–47 (1–122), 1894 (1895).

$$\mathsf{f}_X(\lambda) = \lim_{\omega \to 0^+} \frac{1}{\pi} \mathsf{Im}\left[\mathcal{S}_X(\lambda + \jmath\omega)\right]$$

The η-Transform [Tulino-Verdú 2004]

The η-transform of a nonnegative random variable X is given by

$$\eta_X(\gamma) = \mathbb{E}\left[\frac{1}{1+\gamma X}\right]$$

where γ is a nonnegative real number, and thus, $0 < \eta_X(\gamma) \leq 1$.

Note:

$$\eta_X(\gamma) = \sum_{k=0}^{\infty}(-\gamma)^k \mathbb{E}[X^k],$$

η-Transform of a Random Matrix

Given a $K \times K$ Hermitian matrix $\mathbf{A} = \mathbf{H}^\dagger\mathbf{H}$,

- the η-transform of its expected ESD is

$$\eta_{\bar{\mathsf{F}}^N_{\mathbf{A}}}(\gamma) = \frac{1}{K}\sum_{i=1}^{K}\mathbb{E}\left[\frac{1}{1+\gamma\lambda_i(\mathbf{H}^\dagger\mathbf{H})}\right] = \frac{1}{N}\mathbb{E}\left[\mathrm{tr}\left\{\left(\mathbf{I}+\gamma\mathbf{H}^\dagger\mathbf{H}\right)^{-1}\right\}\right]$$

- the η-transform of its asymptotic ESD is

$$\eta_{\mathbf{A}}(\gamma) = \int_0^\infty \frac{1}{1+\gamma x} d\mathsf{F}_{\mathbf{A}}(x) = \lim_{K\to\infty}\frac{1}{K}\mathrm{tr}\{(\mathbf{I}+\gamma\mathbf{H}^\dagger\mathbf{H})^{-1}\}$$

$\eta(\gamma)$ = generating function for the expected (asymptotic) moments of $\mathbf{A}$.

$\eta(\mathsf{SNR}) =$ Minimum Mean Square Error

The Shannon Transform [Tulino-Verdú 2004]

The Shannon transform of a nonnegative random variable X is defined as

$$\mathcal{V}_X(\gamma) = \mathbb{E}[\log(1 + \gamma X)]$$

where $\gamma > 0$.

- The Shannon transform gives the capacity of various noisy coherent communication channels.

Shannon Transform of a Random Matrix

Given a $N \times N$ Hermitian matrix $\mathbf{A} = \mathbf{H}\mathbf{H}^\dagger$,

- the Shannon transform of its expected ESD is

$$\mathcal{V}_{\bar{\mathsf{F}}_{\mathbf{A}}^{N}}(\gamma) = \frac{1}{N}\mathbb{E}\left[\log\det\left(\mathbf{I} + \gamma\mathbf{A}\right)\right]$$

- the Shannon transform of its asymptotic ESD is

$$\mathcal{V}_{\mathbf{A}}(\gamma) = \lim_{N\to\infty}\frac{1}{N}\log\det\left(\mathbf{I} + \gamma\mathbf{A}\right)$$

$$\mathcal{I}(\mathsf{SNR}\,, \mathbf{H}\mathbf{H}^\dagger) = \mathcal{V}(\mathsf{SNR}\,)$$

Stieltjes, Shannon and η

$$\frac{\gamma}{\log e}\frac{d}{d\gamma}\mathcal{V}_X(\gamma) = 1 - \frac{1}{\gamma}\mathcal{S}_X\left(-\frac{1}{\gamma}\right) = 1 - \eta_X(\gamma)$$

Stieltjes, Shannon and η

$$\frac{\gamma}{\log e}\frac{d}{d\gamma}\mathcal{V}_X(\gamma) = 1 - \frac{1}{\gamma}\mathcal{S}_X\left(-\frac{1}{\gamma}\right) = 1 - \eta_X(\gamma)$$

$$\Updownarrow$$

$$\text{SNR}\,\frac{d}{d\text{SNR}}\mathcal{I}(\text{SNR}) = \frac{K}{N}(1 - \text{MMSE})$$

S-transform

D. Voiculescu, "Multiplication of certain non-commuting random variables," *J. Operator Theory*, vol. 18, pp. 223–235, 1987.

$$\Sigma_X(x) = -\frac{x+1}{x}\eta_X^{-1}(1+x) \tag{6}$$

which maps $(-1, 0)$ onto the positive real line.

S-transform: Key Theorem

♣O. Ryan, On the limit distributions of random matrices with independent or free entries, *Com. in Mathematical Physics*, vol. 193, pp. 595-626, 1998.

♠F. Hiai and D. Petz, Asymptotic freeness almost everywhere for random matrices, *Acta Sci. Math. Szeged*, vol. 66, pp. 801-826, 2000.

Let $\mathbf{A}$ and $\mathbf{B}$ be independent random matrices, if either:

♠ $\mathbf{B}$ is unitarily or bi-unitarily invariant,

♣ or both $\mathbf{A}$ and $\mathbf{B}$ have i.i.d entries

then S-transform of the spectrum of $\mathbf{AB}$ is :

$$\Sigma_{\mathbf{AB}}(x) = \Sigma_{\mathbf{A}}(x)\Sigma_{\mathbf{B}}(x)$$

and

$$\eta_{\mathbf{AB}}(\gamma) = \eta_{\mathbf{A}}\left(\frac{\gamma}{\Sigma_{\mathbf{B}}(\eta_{\mathbf{AB}}(\gamma) - 1)}\right)$$

S-transform: Example

Let

$$\mathbf{H} = \mathbf{C}\mathbf{Q}$$

where:

- ⋇ $K \leq N$
- ⋇ $\mathbf{Q}$ is an $N \times K$ matrix independent of $\mathbf{C}$ and uniformly distributed over the Stiefel manifold of complex $N \times K$ matrices such that $\mathbf{Q}\mathbf{Q}^\dagger = \mathbf{I}$.

Since $\mathbf{Q}$ is bi-unitarily invariant then

$$\eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\text{SNR}) = \eta_{\mathbf{C}\mathbf{C}^\dagger}\left(\text{SNR}\,\frac{\beta - 1 + \eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}}{\eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\text{SNR})}\right)$$

and

$$\mathcal{V}_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(\gamma) = \int_0^{\text{SNR}} \frac{1}{x}\left(1 - \eta_{\mathbf{C}\mathbf{Q}\mathbf{Q}^\dagger\mathbf{C}^\dagger}(x)\right) dx$$

Downlink MC-CDMA with Orthogonal Sequences and equal-power

$$\mathbf{y} = \mathbf{CQAx} + \mathbf{n},$$

where:

- $\mathbf{Q}$ = the orthogonal spreading sequences
- $\mathbf{A}$ = the $K \times K$ diagonal matrix of transmitted amplitudes with $\mathbf{A} = \mathbf{I}$
- C = the $N \times N$ matrix of fading coefficients.

$$\frac{1}{K}\sum_{k=1}^{K} \mathsf{MMSE}_k \stackrel{a.s.}{\to} \eta_{\mathbf{Q}^\dagger\mathbf{C}^\dagger\mathbf{CQ}}(\mathsf{SNR}) = 1 - \frac{1}{\beta}(1 - \eta_{\mathbf{CQQ}^\dagger\mathbf{C}^\dagger}(\mathsf{SNR}))$$

An alternative characterization of the Shannon-transform (inspired by the optimality by successive cancellation with MMSE) is

$$\mathcal{V}_{\mathbf{CQQ}^\dagger\mathbf{C}^\dagger}(\gamma) = \beta\,\mathbb{E}\left[\log\left(1 + \daleth(\mathsf{Y},\gamma)\right)\right]$$

with

$$\frac{\daleth(y,\gamma)}{1+\daleth(y,\gamma)} = \mathbb{E}\left[\frac{\gamma|\mathsf{C}|^2}{\beta\, y\, \gamma|\mathsf{C}|^2 + 1 + (1-\beta\, y)\daleth(y,\gamma)}\right]$$

where Y is a random variable uniform on $[0,1]$.

R-transform

D. Voiculescu, “Addition of certain non-commuting random variables,” *J. Funct. Analysis*, vol. 66, pp. 323–346, 1986.

$$\mathsf{R}_X(z) = \mathcal{S}_X^{-1}(-z) - \frac{1}{z}. \tag{7}$$

R-transform and η-transform

The R-transform (restricted to the negative real axis) of a non-negative random variable X is given by

$$\mathsf{R}_X(\varphi) = \frac{\eta_X(\gamma) - 1}{\varphi}$$

with γ and φ satisfying $\varphi = -\gamma\,\eta_X(\gamma)$

R-transform: Key Theorem

♣O. Ryan, On the limit distributions of random matrices with independent or free entries, *Com. in Mathematical Physics*, vol. 193, pp. 595-626, 1998.

♠F. Hiai and D. Petz, Asymptotic freeness almost everywhere for random matrices, *Acta Sci. Math. Szeged*, vol. 66, pp. 801-826, 2000.

Let $\mathbf{A}$ and $\mathbf{B}$ be independent random matrices, if either:

♠ $\mathbf{B}$ is unitarily or bi-unitarily invariant,

♣ or both $\mathbf{A}$ and $\mathbf{B}$ have i.i.d entries

then the *R-transform* of the spectrum of the sum is $\mathrm{R}_{\mathbf{A}+\mathbf{B}} = \mathrm{R}_{\mathbf{A}} + \mathrm{R}_{\mathbf{B}}$ and

$$\eta_{\mathbf{A}+\mathbf{B}}(\gamma) = \eta_{\mathbf{A}}(\gamma_a) + \eta_{\mathbf{B}}(\gamma_b) - 1$$

with γ_a, γ_b and γ satisfying the following pair of equations:

$$\gamma_a\, \eta_{\mathbf{A}}(\gamma_a) = \gamma\, \eta_{\mathbf{A}+\mathbf{B}}(\gamma) = \gamma_b\, \eta_{\mathbf{B}}(\gamma_b)$$

Random Quadratic Forms

Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *The Annals of Probability*, vol. 26, pp. 316–345, 1998.

Theorem 4. *Let the components of the N-dimensional vector $\mathbf{x}$ be zero-mean and independent with variance $\frac{1}{N}$. For any $N \times N$ nonnegative definite random matrix $\mathbf{B}$ independent of $\mathbf{x}$ whose spectrum converges almost surely,*

$$\lim_{N\to\infty} \mathbf{x}^{\dagger} \left(\mathbf{I} + \gamma\mathbf{B}\right)^{-1} \mathbf{x} = \eta_{\mathbf{B}}(\gamma) \quad \textit{a.s.} \tag{8}$$

$$\lim_{N\to\infty} \mathbf{x}^{\dagger} \left(\mathbf{B} - z\mathbf{I}\right)^{-1} \mathbf{x} = \mathcal{S}_{\mathbf{B}}(z) \quad \textit{a.s.} \tag{9}$$

Rationale

Stieltjes: Description of asymptotic distribution of singular values + tool for proving results (Marčenko-Pastur (1967))

η**:** Description of asymptotic distribution of singular values + signal processing insight

Shannon: Description of asymptotic distribution of singular values + information theory insight

Non-asymptotic Shannon Transform

Example: For $N \times K$-matrix $\mathbf{H}$ having zero-mean i.i.d. Gaussian entries:

$$\mathcal{V}(\mathsf{SNR}) = \sum_{k=0}^{t-1} \sum_{\ell_1=0}^{k} \sum_{\ell_2=0}^{k} \binom{k}{\ell_1} \frac{(k+r-t)!(-1)^{\ell_1+\ell_2} I_{\ell_1+\ell_2+r-t}(\mathsf{SNR})}{(k-\ell_2)!(r-t+\ell_1)!(r-t+\ell_2)!\ell_2!}$$

$$I_0(\mathsf{SNR}) = -e^{\frac{1}{\mathsf{SNR}}} E_i(-\tfrac{1}{\mathsf{SNR}})$$

$$I_n(\mathsf{SNR}) = n I_{n-1}(\mathsf{SNR}) + (-\mathsf{SNR})^{-n} \left(I_0(\mathsf{SNR}) + \sum_{k=1}^{n} (k-1)! \, (-\mathsf{SNR})^{k} \right)$$

For the η-Transform

$$\eta(\mathsf{SNR}) = 1 - \frac{\mathsf{SNR}}{\beta} \frac{d}{d\mathsf{SNR}} \mathcal{V}(\mathsf{SNR})$$

Asymptotics

- $K \longrightarrow \infty$
- $N \longrightarrow \infty$
- $\frac{K}{N} \longrightarrow \beta$

Shannon and η-Transform of Marčenko-Pastur Law

Example: The Shannon transform of the Marčenko-Pastur law is

$$\mathcal{V}(\text{SNR}) = \log\left(1 + \text{SNR} - \frac{1}{4}\mathcal{F}(\text{SNR}, \beta)\right) + \frac{1}{\beta}\log\left(1 + \text{SNR}\,\beta - \frac{1}{4}\mathcal{F}(\text{SNR}, \beta)\right) - \frac{\log e}{4\,\beta\,\text{SNR}}\mathcal{F}(\text{SNR}, \beta)$$

where

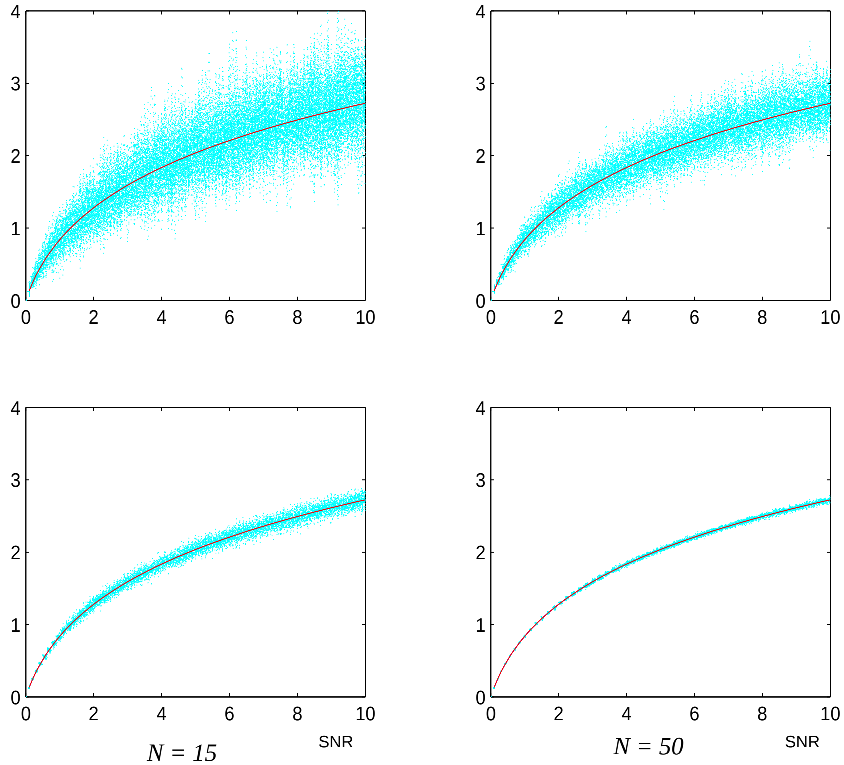
$$\mathcal{F}(x,z) = \left(\sqrt{x(1+\sqrt{z})^2+1} - \sqrt{x(1-\sqrt{z})^2+1}\right)^2$$

while its η-transform is

$$\eta_{\mathbf{HH}^\dagger}(\text{SNR}) = \left(1 - \frac{\mathcal{F}(\text{SNR}, \beta)}{4\,\text{SNR}}\right)$$

Asymptotics

$$\text{Shannon Capacity} = \mathcal{V}_{F^N_{\mathbf{HH}^\dagger}}(\text{SNR}) = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + \text{SNR}\,\lambda_i(\mathbf{HH}^\dagger)\right)$$



$\beta = 1$ for sizes: $N = 3, 5, 15, 50$

Asymptotics

Distribution Insensitivity: The asymptotic eigenvalue distribution does not depend on the distribution with which the independent matrix coefficients are generated.

"Ergodicity": The eigenvalue histogram of one matrix realization converges almost surely to the asymptotic eigenvalue distribution.

Speed of Convergence: $8 = \infty$.

Marčenko-Pastur Law: Applications

- Unfaded Equal-Power DS-CDMA
- Canonical model (i.i.d. Rayleigh fading MIMO channels)
- Multi-Carrier CDMA channels whose sequences have i.i.d. entries

More General Models

- Correlated Entries

$$\mathbf{H} = \sqrt{\mathbf{\Phi}_{\mathsf{R}}}\mathbf{S}\sqrt{\mathbf{\Phi}_{\mathsf{T}}}$$

S: $N \times K$ matrix whose entries are independent complex random variables (arbitrarily distributed) with identical means and variance $\frac{1}{N}$.
$\mathbf{\Phi}_{\mathsf{R}}$**:** $N \times N$ random matrix whose asymptotic spectrum of converges a. s. to a compactly supported measure.
$\mathbf{\Phi}_{\mathsf{T}}$**:** $K \times K$ random matrix whose asymptotic spectrum converges a. s. to a compactly supported measure.

- Non-identically Distributed Entries

H be an $N \times K$ complex random matrix with independent entries (arbitrarily distributed) with identical means.

$$\mathrm{Var}[\mathsf{H}_{i,j}] = \frac{\mathsf{G}_{i,j}}{N}$$

with $\mathsf{G}_{i,j}$ uniformly bounded.

Special case : Doubly Regular Channels

Doubly Regular Matrices [Tulino-Lozano-Verdu,2005]

Definition: An $N \times K$ matrix $\mathbf{P}$ is *asymptotically mean row-regular* if

$$\lim_{K \to \infty} \frac{1}{K} \sum_{j=1}^{K} \mathsf{P}_{i,j}$$

is independent of i as $\frac{K}{N} \to \beta$.

Definition: $\mathbf{P}$ is *asymptotically mean column-regular* if its transpose is asymptotically mean row-regular.

Definition: $\mathbf{P}$ is *asymptotically mean doubly-regular* if it is both asymptotically mean row-regular and asymptotically mean column-regular.

- If the limits $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{P}_{i,j} = \lim_{K \to \infty} \frac{1}{K} \sum_{j=1}^{K} \mathsf{P}_{i,j} = 1$ then $\mathbf{P}$ is *standard asymptotically mean doubly-regular*.

Regular Matrices: Example

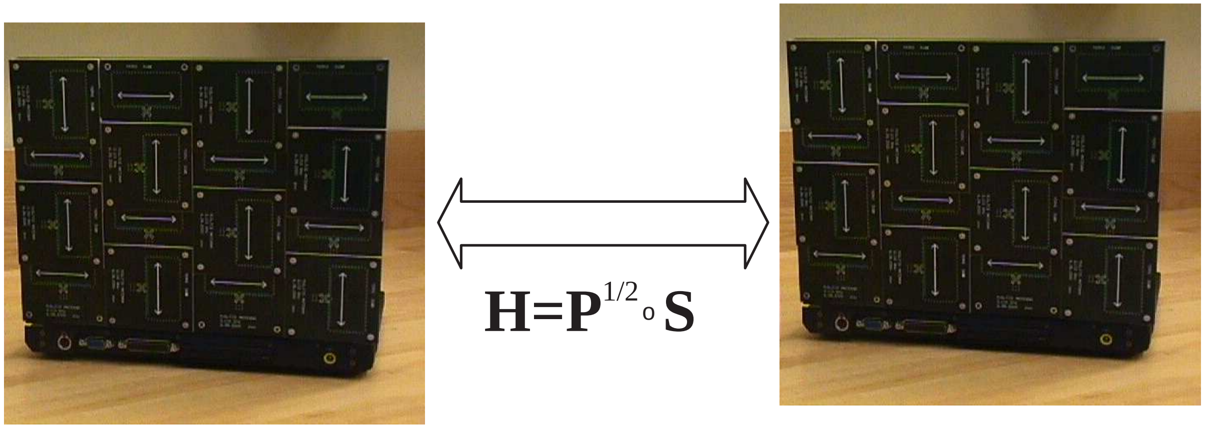
- An $N \times K$ rectangular Toeplitz matrix

$$\mathsf{P}_{i,j} = \varphi(i-j)$$

  with $K \geq N$ is an asymptotically mean row-regular matrix.

- If either the function φ is periodic or $N = K$, then the Toeplitz matrix is asymptotically mean doubly-regular.

Double Regularity: Engineering Insight



- where $\mathbf{S}$ has i.i.d. entries with variance $\frac{1}{N}$ and thus $\mathrm{Var}\left[\mathsf{H}_{i,j}\right] = \frac{\mathsf{P}_{i,j}}{N}$
- gain between copolar antennas (σ) different from gain between crosspolar antennas (χ) and thus when antennas with two orthogonal polarizations are used

$$\mathbf{P} = \begin{bmatrix} \sigma & \chi & \sigma & \chi & \cdots \\ \chi & \sigma & \chi & \sigma & \cdots \\ \sigma & \chi & \sigma & \chi & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

which is mean doubly regular.

Asymptotic ESD of a Doubly-Regular Matrix [Tulino-Lozano-Verdu, 2005]

Theorem: Define an $N \times K$ complex random matrix $\mathbf{H}$ whose entries

- are independent (arbitrarily distributed) satisfying the Lindeberg condition and with identical means.

- have variances
$$\mathrm{Var}\left[\mathsf{H}_{i,j}\right] = \frac{\mathsf{P}_{i,j}}{N}$$
with $\mathbf{P}$ an $N \times K$ deterministic standard asymptotically doubly-regular matrix whose entries are uniformly bounded for any N.

The ESD of $\mathbf{H}^{\dagger}\mathbf{H}$ converges a.s. to the Marčenko-Pastur law.

This result extends to matrices $\mathbf{H} = \mathbf{H}_0 + \bar{\mathbf{H}}$ whose mean has rank $r > 1$ such that

$$\lim_{N\to\infty} \frac{r}{N} = 0.$$

One-Side Correlated Entries

Let $\mathbf{H} = \mathbf{S}\sqrt{\mathbf{\Phi}}$ (or $\mathbf{H} = \sqrt{\mathbf{\Phi}}\mathbf{S}$) with:

$\mathbf{S}$**:** $N \times K$ matrix whose entries are independent (arbitrarily distributed) with identical mean and variance $\frac{1}{N}$.

$\mathbf{\Phi}$**:** $K \times K$ (or $N \times N$) deterministic correlation matrix whose asymptotic ESD converges to a compactly supported measure.

Then,

$$\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) = \beta\mathcal{V}_{\mathbf{\Phi}}(\eta_{\mathbf{HH}^\dagger}\gamma) + \log\frac{1}{\eta_{\mathbf{HH}^\dagger}} + (\eta_{\mathbf{HH}^\dagger} - 1)\log e$$

with $\eta_{\mathbf{HH}^\dagger}(\gamma)$ satisfying

$$\beta = \frac{1 - \eta_{\mathbf{HH}^\dagger}}{1 - \eta_{\mathbf{\Phi}}(\gamma\eta_{\mathbf{HH}^\dagger})}.$$

One-Side Correlated Entries: Applications

- Multi-Antenna Channels with correlation either only at the transmitter or at the receiver.

- DS-CDMA with Frequency-Flat Fading; in this case
  - ⋇ $\mathbf{\Phi} = \mathbf{A}\mathbf{A}^{\dagger}$ with $\mathbf{A}$ the $K \times K$ diagonal matrix of complex fading coefficients

Correlated Entries

Let

$$\mathbf{H} = \mathbf{CSA}$$

$\mathbf{S}$**:** $N \times K$ complex random matrix whose entries are i.i.d with variance $\frac{1}{N}$.

$\mathbf{\Phi}_\mathsf{R} = \mathbf{CC}^\dagger$**:** $N \times N$ either determinist or random matrix such that its ESD converges a.s. to a compactly supported measure.

$\mathbf{\Phi}_\mathsf{T} = \mathbf{AA}^\dagger$**:** $K \times K$ either determinist or random matrix such that its ESD of converges a.s. to a compactly supported measure.

Definition: Let Λ_R and Λ_T be independent random variables with distributions given by the asymptotic ESD of $\mathbf{\Phi}_\mathsf{R}$ and $\mathbf{\Phi}_\mathsf{T}$.

Correlated Entries: Applications

- Multi-Antenna Channels with correlation at the transmitter and receiver (Separable correlation model); in this case:
  - ⋇ $\mathbf{\Phi}_{\mathsf{R}} =$ the receive correlation matrix respectively,
  - ⋇ $\mathbf{\Phi}_{\mathsf{T}} =$ the transmit correlation matrix.

- Downlink MC-CDMA with frequency-selective fading and i.i.d sequences; in this case:
  - ⋇ $\mathbf{C} =$ the $N \times N$ diagonal matrix containing fading coefficient for each subcarrier,
  - ⋇ $\mathbf{A} =$ the K $\times$ K deterministic diagonal matrix containing the amplitudes of the users.

Correlated Entries: Applications

- Downlink DS-CDMA with Frequency-Selective Fading; in this case:

  ⋇ $\mathbf{C} =$ the $N \times N$ Toeplitz matrix defined as:

$$(\mathbf{C})_{i,j} = \frac{1}{W_c} c\left(\frac{i-j}{W_c}\right)$$

  with $c(\cdot)$ the impulse response of the channel,

  ⋇ $\mathbf{A} =$ K $\times$ K deterministic diagonal matrix containing the amplitudes of the users.

Correlated Entries: Shannon and η-transform [Tulino-Lozano-Verdú, 2003]

- The η-transform is:

$$\eta_{\mathbf{HH}^\dagger}(\gamma) = \eta_{\mathbf{\Phi}_{\mathsf{R}}}(\beta\,\gamma_r(\gamma)).$$

- The Shannon transform is:

$$\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) = \mathcal{V}_{\mathbf{\Phi}_{\mathsf{R}}}(\beta\gamma_r) + \beta\mathcal{V}_{\mathbf{\Phi}_{\mathsf{T}}}(\gamma_t) - \beta\frac{\gamma_r\gamma_t}{\gamma}\log e$$

where

$$\frac{\gamma_r\gamma_t}{\gamma} = 1 - \eta_{\mathbf{\Phi}_{\mathsf{T}}}(\gamma_t) \qquad\qquad \beta\frac{\gamma_r\gamma_t}{\gamma} = 1 - \eta_{\mathbf{\Phi}_{\mathsf{R}}}(\beta\gamma_r)$$

Correlated Entries: Shannon and η-transform [Tulino-Lozano-Verdú, 2003]

- The η-transform is:

$$\eta_{\mathbf{HH}^\dagger}(\gamma) = \mathbb{E}\left[\frac{1}{1+\beta\Lambda_{\mathsf{R}}\gamma_r(\gamma)}\right].$$

- The Shannon transform is:

$$\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) = \mathbb{E}\left[\log_2(1+\beta\Lambda_{\mathsf{R}}\gamma_r)\right] + \beta\mathbb{E}\left[\log_2(1+\Lambda_{\mathsf{T}}\gamma_t)\right] - \beta\frac{\gamma_r\gamma_t}{\gamma}\log_2 e$$

where

$$\frac{\gamma_r\gamma_t}{\gamma} = \gamma_t\mathbb{E}\left[\frac{\Lambda_{\mathsf{T}}}{1+\Lambda_{\mathsf{T}}\gamma_t}\right] \qquad \beta\frac{\gamma_r\gamma_r}{\gamma} = \beta\gamma_r\mathbb{E}\left[\frac{\Lambda_{\mathsf{R}}}{1+\beta\Lambda_{\mathsf{R}}\gamma_r}\right] \qquad (10)$$

Arbitrary Numbers of Dimensions: Shannon Transform of Correlated channels

- The η-transform is:

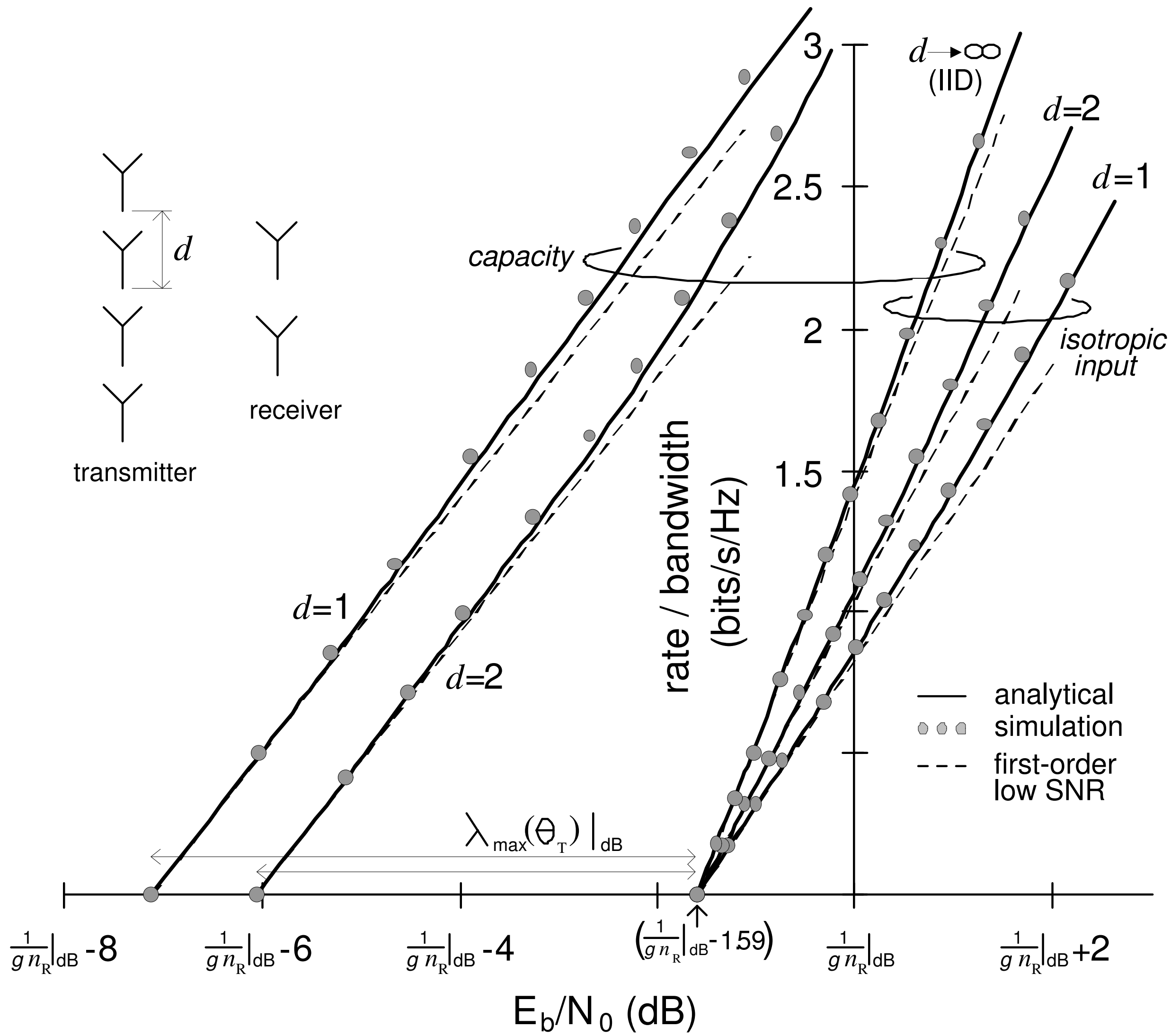
$$\eta_{\mathbf{H}\mathbf{H}^\dagger}(\gamma) \approx \frac{1}{n_{\mathrm{R}}} \sum_{i=1}^{n_{\mathrm{R}}} \frac{1}{1+\beta\,\lambda_i(\mathbf{\Phi}_{\mathsf{R}})\,\gamma_r}.$$

- The Shannon transform is:

$$\mathcal{V}_{\mathbf{H}\mathbf{H}^\dagger}(\gamma) \approx \sum_{i=1}^{n_{\mathrm{R}}} \log_2\left(1+\beta\,\lambda_i(\mathbf{\Phi}_{\mathsf{R}})\,\gamma_r\right) + \beta \sum_{j=1}^{n_{\mathrm{T}}} \log_2\left(1+\lambda_j(\mathbf{\Phi}_{\mathsf{T}})\,\gamma_t\right) - \beta \frac{\gamma_t\,\gamma_r}{\gamma} \log_2 e$$

$$\frac{\gamma_r}{\gamma} = \frac{1}{n_{\mathrm{T}}} \sum_{j=1}^{n_{\mathrm{T}}} \frac{\lambda_j(\mathbf{\Phi}_{\mathsf{T}})}{1+\lambda_j(\mathbf{\Phi}_{\mathsf{T}})\gamma_t} \qquad \frac{\gamma_t}{\gamma} = \frac{1}{n_{\mathrm{R}}} \sum_{i=1}^{n_{\mathrm{R}}} \frac{\lambda_i(\mathbf{\Phi}_{\mathsf{R}})}{1+\beta\,\lambda_i(\mathbf{\Phi}_{\mathsf{R}})\,\gamma_r}.$$

Example: Mutual Information of a Multi-Antenna Channel



The transmit correlation matrix: $(\mathbf{\Phi}_{\mathsf{T}}{}^{\dagger})_{i,j} \approx e^{-0.05\, d^2 (i-j)^2}$ with d antenna spacing (wavelengths).

Correlated Entries (Hanly-Tse, 2001)

- $\mathbf{S}$ be a $N \times K$ matrix with i.i.d entries
- $\mathbf{A}_\ell = \mathrm{diag}\{\mathsf{A}_{1,\ell}, \ldots, \mathsf{A}_{K,\ell}\}$ where $\{\mathsf{A}_{k,\ell}\}$ are i.i.d. random variables
- $\bar{\mathbf{S}}$ be a $NL \times K$ matrix with i.i.d entries
- $\mathbf{P}$ a $K \times K$ diagonal matrix whose k-th diagonal entry $(\mathbf{P})_{k,k} = \sum_{\ell=1}^{L} \mathsf{A}^2_{k,\ell}$.

The distribution of the singular values of the matrix

$$\mathbf{H} = \begin{bmatrix} \mathbf{S}\mathbf{A}_1 \\ \cdots \\ \mathbf{S}\mathbf{A}_L \end{bmatrix} \tag{11}$$

is the same as the distribution of the singular values of the matrix

$$\bar{\mathbf{S}}\sqrt{\mathbf{P}}$$

Applications: DS-CDMA with Flat Fading and Antenna Diversity: $\{\mathsf{A}_{k,\ell}\}$ are the i.i.d. fading coefficients of the kth user at the ℓth antenna and $\mathbf{S}$ is the signature matrix.

Engineering interpretation: the effective spreading gain $=$ the CDMA spreading gain $\times$ the number of receive antennas

Non-identically Distributed Entries

Let $\mathbf{H}$ be an $N \times K$ complex random matrix:

- Entries are independent (arbitrarily distributed) satisfying the Lindeberg condition and with identical means,

- $$\mathrm{Var}[\mathsf{H}_{i,j}] = \frac{\mathsf{P}_{i,j}}{N}$$

  where $\mathbf{P}$ is an $N \times K$ deterministic matrix whose entries are uniformly bounded.

Arbitrary Numbers of Dimensions: Shannon Transform for IND Channels

$$\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma) \approx \beta \sum_{j=1}^{n_\mathrm{T}} \log_2 (1 + \gamma \, \Gamma_j) + \sum_{i=1}^{n_\mathrm{R}} \log_2 \left(1 + \frac{\gamma\beta}{n_\mathrm{T}} \sum_{j=1}^{n_\mathrm{T}} (\mathbf{P})_{i,j} \Upsilon_j \right) - \frac{\gamma\beta}{n_\mathrm{T}} \sum_{j=1}^{n_\mathrm{T}} \Gamma_j \Upsilon_j$$

where

$$\Gamma_j = \frac{1}{n_\mathrm{R}} \sum_{i=1}^{n_\mathrm{R}} \frac{(\mathbf{P})_{i,j}}{1 + \beta \frac{1}{n_\mathrm{T}} \sum_{j=1}^{n_\mathrm{T}} (\mathbf{P})_{i,j} \Upsilon_j}$$

$$\Upsilon_j = \frac{\gamma}{1 + \gamma \, \Gamma_j}$$

- SNR Γ_j = SINR exhibited by x_j at the output of a linear MMSE receiver,
- $\Upsilon_j /$SNR= the corresponding MSE.

Non-identically Distributed Entries: Special cases

- $\mathbf{P}$ is asymptotic doubly regular. In which case:

$\mathcal{V}_{\mathbf{HH}^\dagger}(\gamma)$ and $\eta_{\mathbf{HH}^\dagger}(\gamma)$ $\quad\equiv\quad$ Shannon and η of the Marčenko-Pastur Law.

- $\mathbf{P}$ is the outer product of the nonnegative N-vector λ_{R} and K-vector λ_{T}. In this case:

$$\mathbf{G} = \lambda_{\mathsf{R}} \lambda_{\mathsf{T}}^{\dagger} \quad \Rightarrow \quad \mathbf{H} = \sqrt{\mathrm{diag}(\lambda_{\mathsf{R}})}\mathbf{S}\sqrt{\mathrm{diag}(\lambda_{\mathsf{T}})}$$

Non-identically Distributed Entries: Applications

- MC-CDMA frequency-selective fading and i.i.d sequences (Uplink and Downlink).

- Uplink DS-CDMA with Frequency-Selective Fading:

L. Li, A. M. Tulino, and S. Verdú, Design of reduced-rank MMSE multiuser detectors using random matrix methods, *IEEE Trans. on Information Theory*, vol. 50, June 2004.

J. Evans and D. Tse, Large system performance of linear multiuser receivers in multipath fading channels, *IEEE Trans. on Information Theory*, vol. 46, Sep. 2000.

J. M. Chaufray, W. Hachem, and P. Loubaton, Asymptotic analysis of optimum and sub-optimum CDMA MMSE receivers, *Proc. IEEE Int. Symp. on Information Theory (ISIT02)*, p. 189, July 2002.

Non-identically Distributed Entries: Applications

- Multi-Antenna Channels with
  - ⋇ Polarization Diversity:

$$\mathbf{H} = \sqrt{\mathbf{P}} \circ \mathbf{H}_w$$

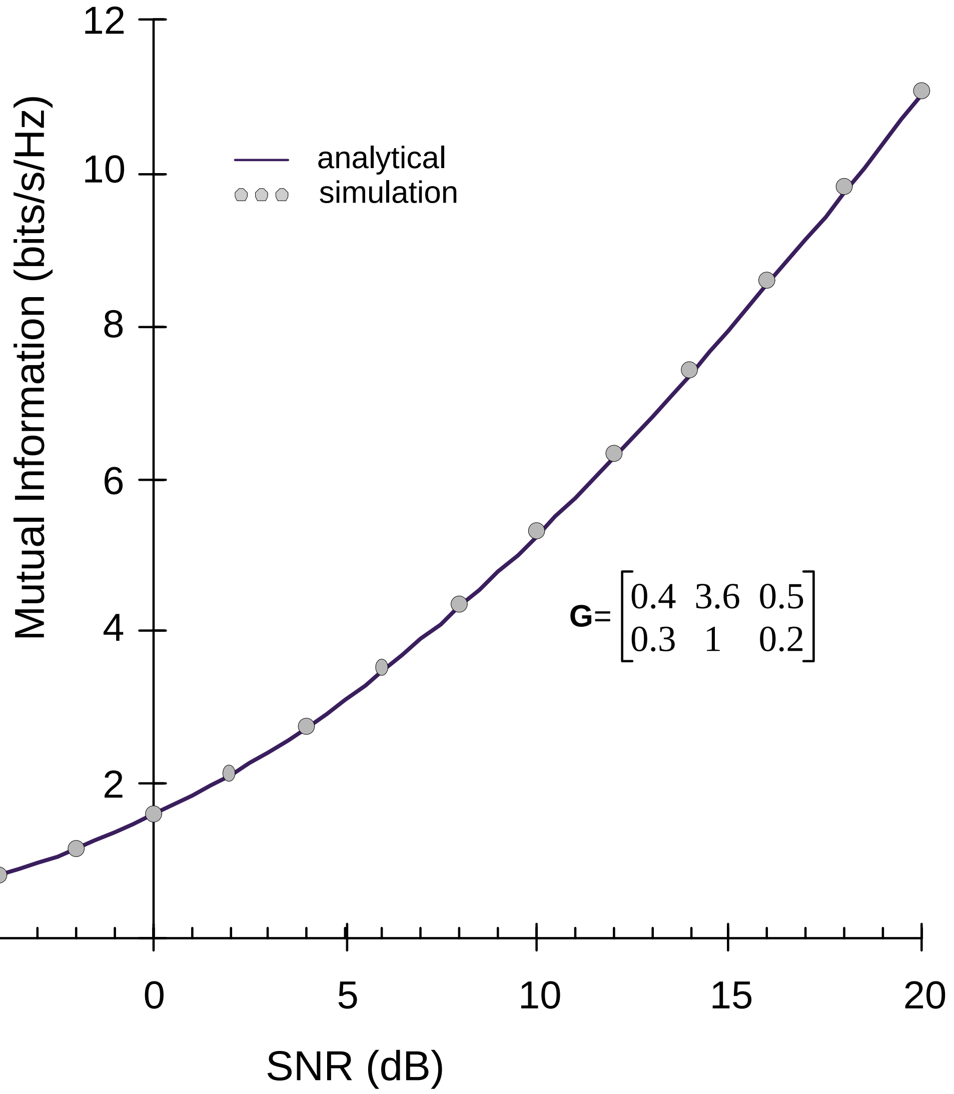
where $\mathbf{H}_w$ is zero-mean i.i.d. Gaussian and $\mathbf{P}$ is a deterministic matrix with nonnegative entries.
$(\mathbf{P})_{\mathbf{i},\mathbf{j}}$ is the power gain between the jth transmit and ith receive antennas, determined by their relative polarizations.

  - ⋇ *Non-separable Correlations*

$$\mathbf{H} = \mathbf{U}\mathbf{H}_w\mathbf{U}^\dagger$$

where $\mathbf{U}_{\mathrm{R}}$ and $\mathbf{U}_{\mathrm{T}}$ are unitary while the entries of $\tilde{\mathbf{H}}$ are independent zero-mean Gaussian. A more restrictive case is when $\mathbf{U}_{\mathrm{R}}$ and $\mathbf{U}_{\mathrm{T}}$ are Fourier matrices.
This model is advocated and experimentally supported in W. Weichselberger et all, A stochastic mimo channel model with joint correlation of both link ends, *IEEE Trans. on Wireless Com.*, vol. 5, no. 1, pp. 90–100, 2006.

Example: Mutual Information of a Multi-Antenna Channel



Ergodic Regime

- $\{\mathbf{H}_i\}$ varies ergodically over the duration of a codeword.

- The quantity of interest is then the mutual information averaged over the fading, $\mathbb{E}\left[\mathcal{I}(\text{SNR}\,, \mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger)\right]$, with

$$\mathcal{I}(\text{SNR}\,, \mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger) = \frac{1}{N}\log\det\left(\mathbf{I} + \text{SNR}\,\mathbf{H}\mathbf{\Phi}\mathbf{H}^\dagger\right)$$

Non-ergodic Conditions

- Often, however, $\mathbf{H}$ is held approximately constant during the span of a codeword

- Outage capacity (cumulative distribution of mutual information),

$$\mathbb{P}_{out}(R) = \mathbb{P}[\log\det(\mathbf{I} + \text{SNR}\,\mathbf{H}\mathbf{H}^\dagger) < R]$$

- The normalized mutual information converges a.s. to its expectation as $K, N \to \infty$ (hardening / self-averaging)

$$\frac{1}{N}\log\det(\mathbf{I} + \text{SNR}\,\mathbf{H}\mathbf{H}^\dagger) \overset{a.s.}{\to} \mathcal{V}_{\mathbf{H}\mathbf{H}^\dagger}(\text{SNR}) = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}[\log\det(\mathbf{I} + \text{SNR}\,\mathbf{H}\mathbf{H}^\dagger)]$$

  However, non-normalized mutual information

$$I(\text{SNR}, \mathbf{H}\mathbf{H}^\dagger) = \log\det(\mathbf{I} + \text{SNR}\,\mathbf{H}\mathbf{H}^\dagger)$$

  still suffers random fluctuations that, while small relative to the mean, are vital to the outage capacity.

CLT for Linear Spectral Statistics

Z. D. Bai and J. W. Silverstein, CLT of linear spectral statistics of large dimensional sample covariance matrices, Annals of Probability, vol. 32, no. 1A, pp. 553605, 2004.

IID Channel

As $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the random variable

$$\Delta_N = \log\det(\mathbf{I} + \text{SNR}\,\mathbf{H}\mathbf{H}^\dagger) - N\mathcal{V}_{\mathbf{H}\mathbf{H}^\dagger}(\text{SNR})$$

is asymptotically zero-mean Gaussian with variance

$$\mathbb{E}\left[\Delta^2\right] = -\log\left(1 - \frac{\left(1 - \eta_{\mathbf{H}\mathbf{H}^\dagger}(\text{SNR})\right)^2}{\beta}\right)$$

IID Channel

- For fixed numbers of antennas, mean and variance of the mutual information of the IID channel given by [Smith & Shafi '02] and [Wang & Giannakis '04]. Approximate normality observed numerically.

- Arguments supporting the asymptotic normality of the cumulative distribution of mutual information given:

  ⋇ in [Hochwald et al. '04], for SNR $\to 0$ or SNR $\to \infty$.

  ⋇ in [Moustakas et al. '03] using the replica method from statistical physics (not yet fully rigorized).

  ⋇ in [Kamath et al. '02], asymptotic normality proved rigorously for any SNR using Bai & Silverstein's CLT.

One-Side Correlated Wireless Channel ($\mathbb{H} = \mathbb{S}\sqrt{\mathbf{\Phi}_{\mathsf{T}}}$) [Tulino-Verdu,2004]

Theorem: As $K, N \to \infty$ with $\frac{K}{N} \to \beta$, the random variable

$$\Delta_N = \log\det(\mathbf{I} + \mathsf{SNR}\,\mathbf{S}\mathbf{\Phi}_{\mathsf{T}}\mathbf{S}^{\dagger}) - N\mathcal{V}_{\mathbf{S}\mathbf{\Phi}_{\mathsf{T}}\mathbf{S}^{\dagger}}(\mathsf{SNR})$$

is asymptotically zero-mean Gaussian with variance

$$\mathbb{E}[\Delta^2] = -\log\left(1 - \beta\,\mathbb{E}\left[\left(\frac{\mathsf{T}\,\mathsf{SNR}\,\eta_{\mathbf{S}\mathbf{\Phi}_{\mathsf{T}}\mathbf{S}^{\dagger}}(\mathsf{SNR})}{1 + \mathsf{T}\,\mathsf{SNR}\,\eta_{\mathbf{S}\mathbf{\Phi}_{\mathsf{T}}\mathbf{S}^{\dagger}}(\mathsf{SNR})}\right)^2\right]\right)$$

with expectation over the nonnegative random variable T whose distribution equals the asymptotic ESD of $\mathbf{\Phi}_{\mathsf{T}}$.

Examples

In the examples that follow, transmit antennas correlated with

$$(\mathbf{\Phi}_{\mathsf{T}})_{i,j} = e^{-0.2(i-j)^2}$$

which is typical of an elevated base station in suburbia. The receive antennas are uncorrelated.

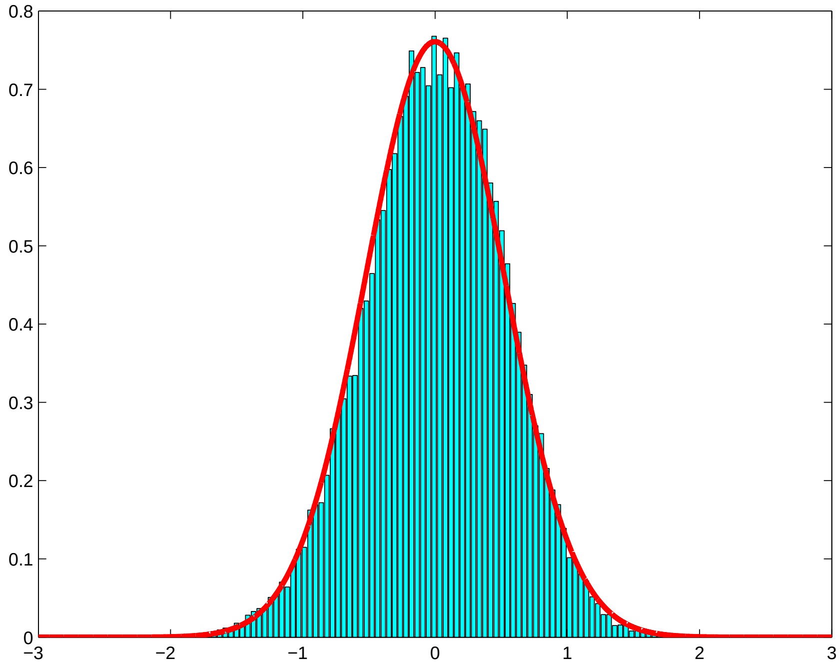
The outage capacity is computed by applying our asymptotic formulas to finite (and small) matrices,

$$\mathcal{V}_{\mathbf{S}\mathbf{\Phi}_{\mathsf{T}}\mathbf{S}^{\dagger}}(\mathsf{SNR}) \approx \frac{1}{N}\sum_{j=1}^{K} \log\left(1 + \mathsf{SNR}\lambda_j(\mathbf{\Phi}_{\mathsf{T}})\,\eta\right) - \log\eta + (\eta - 1)\log e$$

$$\eta = \frac{1}{1+\mathsf{SNR}\,\frac{1}{K}\sum_{j=1}^{K}\frac{\lambda_j(\mathbf{\Phi}_{\mathsf{T}})}{1+\mathsf{SNR}\,\lambda_j(\mathbf{\Phi}_{\mathsf{T}})\eta}}$$

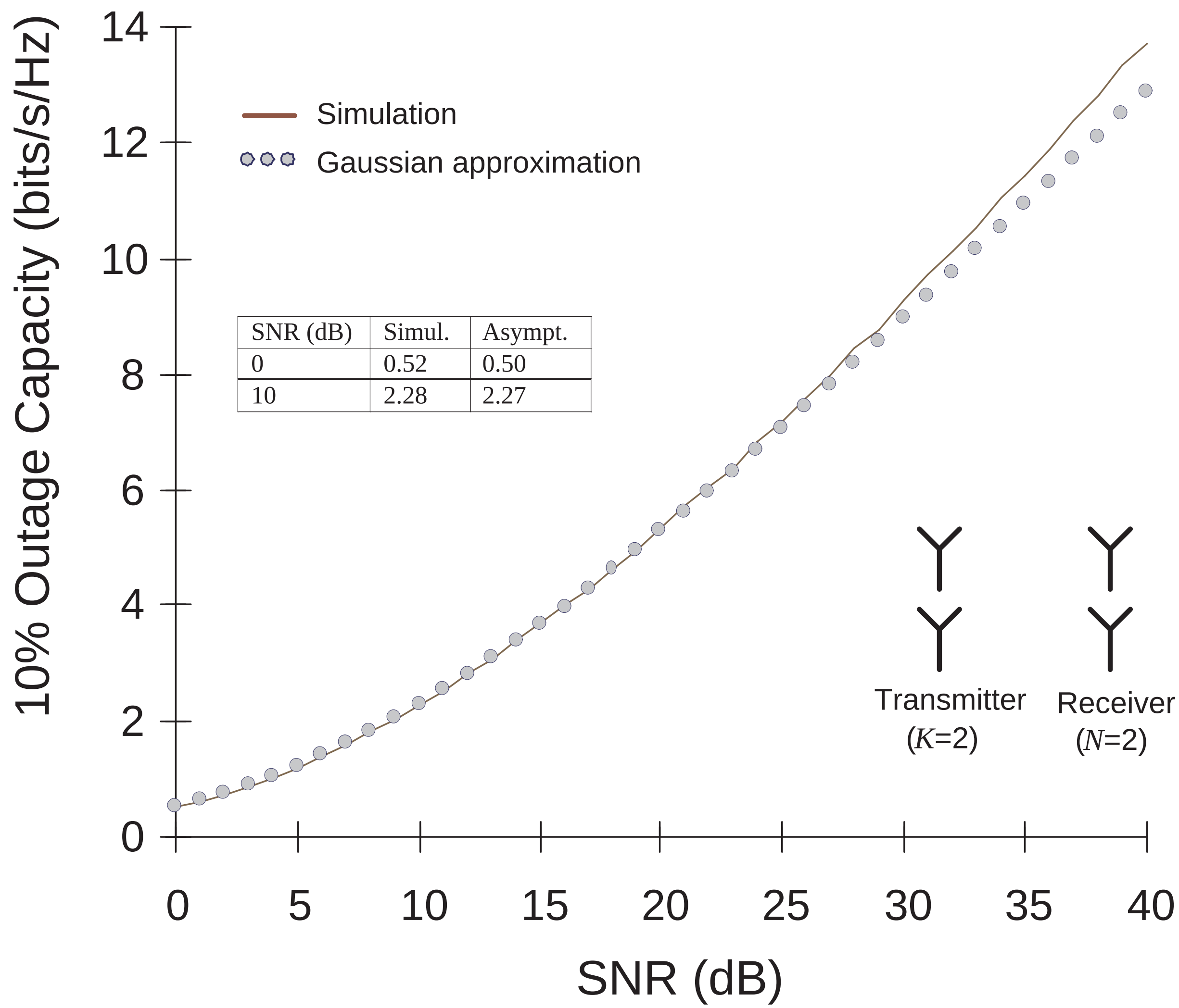
$$\mathbb{E}[\Delta^2] = -\log\left(1 - \beta\,\frac{1}{K}\sum_{j=1}^{K}\left[\left(\frac{\lambda_j(\mathbf{\Phi}_{\mathsf{T}})\mathsf{SNR}\,\eta}{1+\lambda_j(\mathbf{\Phi}_{\mathsf{T}})\mathsf{SNR}\,\eta}\right)^2\right]\right)$$

Example: Histogram



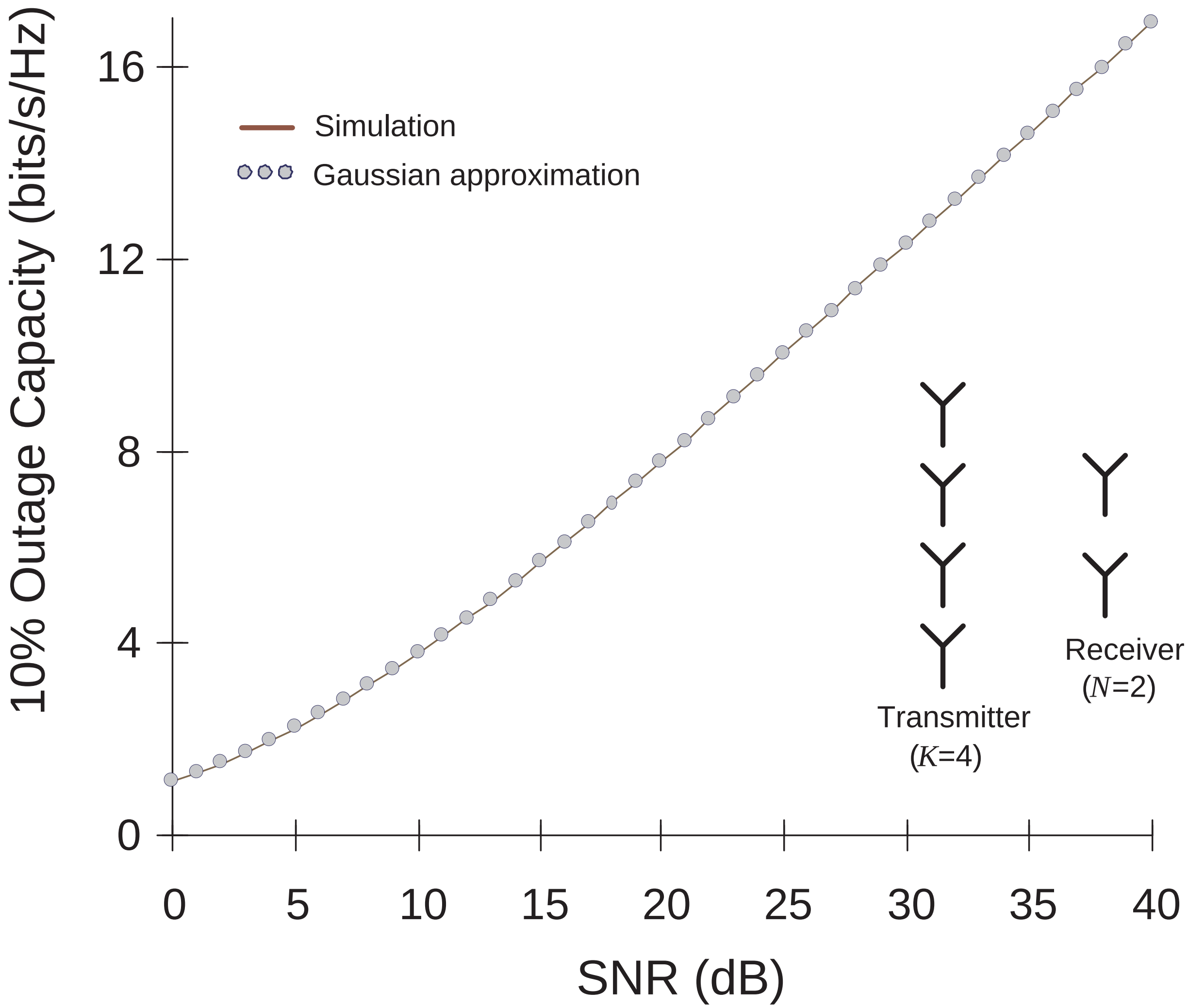
$K = 5$ transmit and $N = 10$ receive antennas, $\text{SNR} = 10$.

Example: 10%-Outage Capacity ($K = N = 2$)



SNR (dB)	Simul.	Asympt.
0	0.52	0.50
10	2.28	2.27

Example: 10%-Outage Capacity ($K = 4, N = 2$)



Summary

- Various wireless communication channels: analysis tackled with the aid of random matrix theory.
- Shannon and η-transforms, motivated by the application of random matrices to the theory of noisy communication channels.
- Shannon transforms and η-transforms for the asymptotic ESD of several classes of random matrices.
- Application of the various findings to the analysis of several wireless channel in both ergodic and non-ergodic regime.
- Succinct expressions for the asymptotic performance measures.
- Applicability of these asymptotic results to finite-size communication systems.

Reference

A. M. Tulino and S. Verdú
"Random Matrices and Wireless Communications,"
Foundations and Trends in Communications and Information Theory,
vol. 1, no. 1, June 2004.
http://dx.doi.org/10.1561/0100000001

Theory of Large Dimensional Random Matrices for Engineers (Part II)

Jack Silverstein
North Carolina State University

The 9th International Symposium on Spread Spectrum Techniques and Applications,
Manaus, Amazon, Brazil,
August 28-31, 2006

1. Introduction. Let $\mathcal{M}(\mathbb{R})$ denote the collection of all sub-probability distribution functions on $\mathbb{R}$. We say for $\{F_N\} \subset \mathcal{M}(\mathbb{R})$, F_N converges vaguely to $F \in \mathcal{M}(\mathbb{R})$ (written $F_N \xrightarrow{v} F$) if for all $[a,b]$, a,b continuity points of F, $\lim_{N\to\infty} F_N\{[a,b]\} = F\{[a,b]\}$. We write $F_N \xrightarrow{D} F$, when F_N, F are probability distribution functions (equivalent to $\lim_{N\to\infty} F_N(a) = F(a)$ for all continuity points a of F).

For $F \in \mathcal{M}(\mathbb{R})$,

$$\mathcal{S}_F(z) \equiv \int \frac{1}{x-z} dF(x), \quad z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} : \Im z > 0\}$$

is defined as the Stieltjes transform of F.

Properties:

1. $\mathcal{S}_F$ is an analytic function on $\mathbb{C}^+$.

2. $\Im \mathcal{S}_F(z) > 0$.

3. $|\mathcal{S}_F(z)| \leq \frac{1}{\Im z}$.

4. For continuity points $a < b$ of F

$$F\{[a,b]\} = \frac{1}{\pi} \lim_{\eta \to 0^+} \int_a^b \Im \mathcal{S}_F(\xi + i\eta) d\xi.$$

5. If, for $x_0 \in \mathbb{R}$, $\Im \mathcal{S}_F(x_0) \equiv \lim_{z \in \mathbb{C}^+ \to x_0} \Im \mathcal{S}_F(z)$ exists, then F is differentiable at x_0 with value $(\frac{1}{\pi}) \Im \mathcal{S}_F(x_0)$ (Silverstein and Choi (1995)).

Let $S \subset \mathbb{C}^+$ be countable with a cluster point in $\mathbb{C}^+$. Using 4., the fact that $F_N \xrightarrow{v} F$ is equivalent to

$$\int f_N(x) dF_N(x) \to \int f(x) dF(x)$$

for all continuous f vanishing at $\pm\infty$, and the fact that an analytic function defined on $\mathbb{C}^+$ is uniquely determined by the values it takes on S, we have

$$F_N \xrightarrow{v} F \quad \Longleftrightarrow \quad \mathcal{S}_{F_N}(z) \to \mathcal{S}_F(z) \quad \text{for all } z \in S.$$

The fundamental connection to random matrices:

For any Hermitian $N \times N$ matrix $\mathbf{A}$, we let $F^{\mathbf{A}}$ denote the *empirical distribution function*, or *empirical spectral distribution* (ESD), of its eigenvalues:

$$F^{\mathbf{A}}(x) = \frac{1}{N}(\text{number of eigenvalues of } \mathbf{A} \leq x).$$

Then

$$\mathcal{S}_{F^{\mathbf{A}}}(z) = \frac{1}{N}\mathsf{tr}\,(\mathbf{A} - z\mathbf{I})^{-1}.$$

So, if we have a sequence $\{\mathbf{A}_N\}$ of Hermitian random matrices, to show, with probability one, $F^{\mathbf{A}_N} \xrightarrow{v} F$ for some $F \in \mathcal{M}(\mathbb{R})$, it is equivalent to show for any $z \in \mathbb{C}^+$

$$\frac{1}{N}\mathsf{tr}\,(\mathbf{A}_N - z\mathbf{I})^{-1} \to \mathcal{S}_F(z) \quad a.s.$$

For the remainder of the lecture $\mathcal{S}_{\mathbf{A}}$ will denote $\mathcal{S}_{F^{\mathbf{A}}}$.

The main goal of this part of the tutorial is to present results on the limiting ESD of three classes of random matrices. The results are expressed in terms of limit theorems, involving convergence of the Stieltjes transforms of the ESD's. An outline of the proof of the first result will be given. The proof will clearly indicate the importance of the Stieltjes transform to limiting spectral behavior. Essential properties needed in the proof will be emphasized in order to better understand where randomness comes in and where basic properties of matrices are used.

For each of the theorems, it is assumed that the sequence of random matrices are defined on a common probability space. They all assume:

For $N = 1, 2, \ldots$ $\mathbf{X} = \mathbf{X}_N = (X_{ij}^N)$, $N \times K$, $X_{ij}^N \in \mathbb{C}$, i.d. for all N, i, j, independent across i, j for each N, $\mathsf{E}|X_{1\,1}^1 - \mathsf{E}X_{1\,1}^1|^2 = 1$, and $K = K(N)$ with $K/N \to \beta > 0$ as $N \to \infty$.

Let $\mathbf{S} = \mathbf{S}_N = (1/\sqrt{N})\mathbf{X}_N$.

THEOREM 1.1 (Marčenko and Pastur (1967), Silverstein and Bai (1995)). *Let* $\mathbf{T}$ *be a* $K \times K$ *real diagonal random matrix whose ESD converges almost surely in distribution, as* $N \to \infty$ *to a nonrandom limit. Let* T *denote a random variable with this limiting distribution. Let* $\mathbf{W}_0$ *be an* $N \times N$ *Hermitian random matrix with ESD converging, almost surely, vaguely to a nonrandom distribution* $\mathcal{W}_0$ *with Stieltjes transform denoted by* $\mathcal{S}_0$*. Assume* $\mathbf{S}$, $\mathbf{T}$, *and* $\mathbf{W}_0$ *to be independent, Then the ESD of*

$$\mathbf{W} = \mathbf{W}_0 + \mathbf{S}\mathbf{T}\mathbf{S}^\dagger$$

converges vaguely, as $N \to \infty$*, almost surely to a nonrandom distribution whose Stieltjes transform,* $\mathcal{S}(\cdot)$*, satisfies for* $z \in \mathbb{C}^+$

$$\mathcal{S}(z) = \mathcal{S}_0\left(z - \beta\,\mathbb{E}\left[\frac{\mathsf{T}}{1 + \mathsf{T}\mathcal{S}(z)}\right]\right). \tag{1.1}$$

It is the only solution to (1.1) in $\mathbb{C}^+$*.*

THEOREM 1.2 (Silverstein, in preparation). *Define* $\mathbf{H} = \mathbf{CSA}$, *where* $\mathbf{C}$ *is* $N \times N$ *and* $\mathbf{A}$ *is* $K \times K$, *both random. Assume that the ESD's of* $\mathbf{D} = \mathbf{CC}^\dagger$ *and* $\mathbf{T} = \mathbf{AA}^\dagger$ *converge almost surely in distribution to nonrandom limits, and let* D *and* T *denote random variables distributed, respectively, according to those limits. Assume* $\mathbf{C}$, $\mathbf{A}$ *and* $\mathbf{S}$ *to be independent. Then the ESD of* $\mathbf{HH}^\dagger$ *converges in distribution, as* $N \to \infty$, *almost surely to a nonrandom limit whose Stieltjes transform,* $\mathcal{S}(\cdot)$, *is given for* $z \in \mathbb{C}^+$ by

$$\mathcal{S}(z) = \mathbb{E}\left[\frac{1}{\beta\,\mathsf{D}\,\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathcal{F}(z)\mathsf{T}}\right] - z}\right],$$

where $\mathcal{F}(z)$ satisfies

$$\mathcal{F}(z) = \mathbb{E}\left[\frac{\mathsf{D}}{\beta\,\mathsf{D}\,\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathcal{F}(z)\,\mathsf{T}}\right] - z}\right]. \tag{1.2}$$

$\mathcal{F}(z)$ *is the only solution to (1.2) in* $\mathbb{C}^+$.

THEOREM 1.3 (Dozier and Silverstein). *Let* $\mathbf{H}_0$ *be* $N \times K$, *random, independent of* $\mathbf{S}$, *such that the ESD of* $\mathbf{H}_0\mathbf{H}_0^\dagger$ *converges almost surely in distribution to a nonrandom limit, and let* M *denote a random variable with this limiting distribution. Let* $\mathsf{K} > 0$ *be nonrandom. Define*

$$\mathbf{H} = \mathbf{S} + \sqrt{\mathsf{K}}\mathbf{H}_0.$$

Then the ESD of $\mathbf{H}\mathbf{H}^\dagger$ converges in distribution, as $N \to \infty$, almost surely to a nonrandom limit whose Stieltjes transform $\mathcal{S}$ satisfies for each $z \in \mathbb{C}^+$

$$\mathcal{S}(z) = \mathbb{E}\left[\frac{1}{\frac{\mathsf{KM}}{1+\mathcal{S}(z)} - z(1 + \mathcal{S}(z)) + (\beta - 1)}\right]. \tag{1.3}$$

$\mathcal{S}(z)$ is the only solution to (1.3) with both $\mathcal{S}(z)$ and $z\mathcal{S}(z)$ in $\mathbb{C}^+$.

Remark: In Theorem 1.1 if $\mathbf{W}_0 = 0$ for all N large, then $\mathcal{S}_0(z) = -1/z$ and we find that $\mathcal{S} = \mathcal{S}(z)$ has an inverse

$$z = -\frac{1}{\mathcal{S}} + \beta\,\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]. \tag{1.4}$$

All of the analytic behavior of the limiting distribution can be extracted from this equation (Silverstein and Choi).

Explicit solutions can be derived in a few cases. Consider the *Mařcenko-Pastur* distribution, where $\mathbf{T} = \mathbf{I}$, that is, the matrix is simply $\mathbf{S}\mathbf{S}^\dagger$. Then $\mathcal{S} = \mathcal{S}(z)$ solves

$$z = -\frac{1}{\mathcal{S}} + \beta\frac{1}{1+\mathcal{S}},$$

resulting in the quadratic equation

$$z\mathcal{S}^2 + \mathcal{S}(z+1-\beta) + 1 = 0$$

with solution

$$
\begin{aligned}
\mathcal{S} = \frac{-(z+1-\beta) \pm \sqrt{(z+1-\beta)^2 - 4z}}{2z} \\
= \frac{-(z+1-\beta) \pm \sqrt{z^2 - 2z(1+\beta) + (1-\beta)^2}}{2z} \\
= \frac{-(z+1-\beta) \pm \sqrt{(z-(1-\sqrt{\beta})^2)(z-(1+\sqrt{\beta})^2)}}{2z}
\end{aligned}
$$

We see the imaginary part of $\mathcal{S}$ goes to zero when z approaches the real line and lies outside the interval $[(1-\sqrt{\beta})^2, (1+\sqrt{\beta})^2]$, so we conclude from property 5. that for all $x \neq 0$ the limiting distribution has a density f given by

$$
f(x) = \begin{cases} \frac{\sqrt{(x-(1-\sqrt{\beta})^2)((1+\sqrt{\beta})^2-x)}}{2\pi x} & x \in ((1-\sqrt{\beta})^2, (1+\sqrt{\beta})^2) \\ 0 & \text{otherwise.} \end{cases}
$$

Considering the value of β (the limit of columns to rows) we can conclude that the limiting distribution has no mass at zero when $\beta \geq 1$, and has mass $1 - \beta$ at zero when $\beta < 1$.

2. Why these theorems are true. We begin with three facts which account for most of why the limiting results are true, and the appearance of the limiting equations for the Stieltjes transforms.

LEMMA 2.1 *For* $N \times N$ $\mathbf{A}$, $q \in \mathbb{C}^N$, *and* $t \in \mathbb{C}$ *with* $\mathbf{A}$ *and* $\mathbf{A} + tqq^\dagger$ *invertible, we have*

$$q^\dagger(\mathbf{A} + tqq^\dagger)^{-1} = \frac{1}{1 + tq^\dagger\mathbf{A}^{-1}q} q^\dagger\mathbf{A}^{-1}$$

(since $q^\dagger\mathbf{A}^{-1}(\mathbf{A} + tqq^\dagger) = (1 + tq^\dagger\mathbf{A}^{-1}q)q^\dagger$).

LEMMA 2.2 *For* $N \times N$ $\mathbf{A}$ *and* $\mathbf{B}$, *with* $\mathbf{B}$ *Hermitian*, $z \in \mathbb{C}^+$, $t \in \mathbb{R}$, *and* $q \in \mathbb{C}^N$, *we have*

$$|\mathsf{tr}\,[(\mathbf{B} - z\mathbf{I})^{-1} - (\mathbf{B} + tqq^\dagger - z\mathbf{I})^{-1}]\mathbf{A}| = \left| t \frac{q^\dagger(\mathbf{B} - z\mathbf{I})^{-1}\mathbf{A}((\mathbf{B} - z\mathbf{I})^{-1}q}{1 + tq^\dagger(\mathbf{B} - z\mathbf{I})^{-1}q} \right| \le \frac{\|\mathbf{A}\|}{\Im z}.$$

Proof. The identity follows from Lemma 2.1. We have

$$\left| t\frac{q^\dagger(\mathbf{B}-zI)^{-1}\mathbf{A}((\mathbf{B}-z\mathbf{I})^{-1}q}{1+tq^\dagger(\mathbf{B}-z\mathbf{I})^{-1}q} \right| \leq \|\mathbf{A}\|\,|t|\frac{\|(\mathbf{B}-z\mathbf{I})^{-1}q\|^2}{|1+tq^\dagger(\mathbf{B}-z\mathbf{I})^{-1}q|}.$$

Write $\mathbf{B} = \sum_i \lambda_i e_i e_i^*$, its spectral decomposition. Then

$$\|(\mathbf{B}-zI)^{-1}q\|^2 = \sum_i \frac{|e_i^\dagger q|^2}{|\lambda_i - z|^2}$$

and

$$|1+tq^\dagger(\mathbf{B}-zI)^{-1}q| \geq |t|\Im(q^\dagger(\mathbf{B}-z\mathbf{I})^{-1}q) = |t|\Im z\sum_i \frac{|e_i^\dagger q|^2}{|\lambda_i - z|^2}.$$

LEMMA 2.3. *For* $X = (X_1, \ldots, X_N)^T$ *i.i.d. standardized entries,* $\mathbf{C}$ $N \times N$*, we have for any* $p \geq 2$

$$\mathbb{E}|X^\dagger \mathbf{C} X - \mathsf{tr}\,\mathbf{C}|^p \leq K_p\Big(\big(\mathbb{E}|X_1|^4 \mathsf{tr}\,\mathbf{C}\mathbf{C}^\dagger\big)^{p/2} + \mathbb{E}|X_1|^{2p} \mathsf{tr}\,(\mathbf{C}\mathbf{C}^\dagger)^{p/2}\Big)$$

where the constant K_p *does not depend on* N, $\mathbf{C}$, *nor on the distribution of* X_1. (Proof given in Bai and Silverstein (1998).)

Thus we have

$$\mathbb{E}\left|\frac{X^\dagger \mathbf{C} X - \mathsf{tr}\,\mathbf{C}}{N}\right|^p \leq \frac{K_0}{N^{p/2}},$$

the constant K_0 depending on a bound on the $2p$-th moment of X_1 and on the norm of $\mathbf{C}$. Roughly speaking, for large N, a scaled quadratic form involving a vector consisting of i.i.d. standardized random variables is close to the scaled trace of the matrix. As will be seen below, this is the only place where randomness comes in.

The first step needed to prove each of the theorems is truncation and centralization of the elements of $\mathbf{X}$, that is, showing that it is sufficient to prove each result under the assumption the elements have mean zero, variance 1, and are bounded, for each N, by a rate growing slower than N ($\log N$ is sufficient). These steps will be omitted. Although not needed for Theorem 1.1, additional truncation of the eigenvalues of $\mathbf{D}$ and $\mathbf{T}$ in Theorem 1.2 and $\mathbf{H}\mathbf{H}^\dagger$ in Theorem 1.3, all at a rate slower than N is also required (again, $\ln N$ is sufficient). We are at this stage able to go through algebraic manipulations, keeping in mind the above three lemmas, and intuitively derive the equation in Theorem 1.1.

Before continuing, two more basic properties of matrices are included here.

LEMMA 2.4 *Let* $z_1, z_2 \in \mathbb{C}^+$ *with* $\max(\Im z_1, \Im z_2) \geq v > 0$, $\mathbf{A}$ *and* $\mathbf{B}$ $N \times N$ *with* $\mathbf{A}$ *Hermitian, and* $q \in \mathbb{C}^N$*. Then*

$$|\mathsf{tr}\,\mathbf{B}((\mathbf{A} - z_1\mathbf{I})^{-1} - (\mathbf{A} - z_2\mathbf{I})^{-1})| \leq |z_2 - z_1| N \|\mathbf{B}\| \frac{1}{v^2}, \text{ and}$$

$$|q^\dagger \mathbf{B}(\mathbf{A} - z_1\mathbf{I})^{-1} q - q^\dagger \mathbf{B}(\mathbf{A} - z_2\mathbf{I})^{-1} q| \leq |z_2 - z_1| \, \|q\|^2 \|\mathbf{B}\| \frac{1}{v^2}.$$

We now outline the proof of Theorem 1.1. Write $\mathbf{T} = \mathrm{diag}(t_1, \ldots, t_K)$. Let q_i denote the i^{th} column of $\mathbf{S}$. Then

$$\mathbf{STS}^\dagger = \sum_{i=1}^{K} t_i q_i q_i^*.$$

Let $\mathbf{W}_{(i)} = \mathbf{W} - t_i q_i q_i^\dagger$. For any $z \in \mathbb{C}^+$ and $x \in \mathbb{C}$ we write

$$\mathbf{W} - z\mathbf{I} = \mathbf{W}_0 - (z - x)\mathbf{I} + (1/N)\mathbf{STS}^\dagger - x\mathbf{I}.$$

Taking inverses we have

$$(\mathbf{W}_0 - (z - x)\mathbf{I})^{-1}$$
$$= (\mathbf{W} - z\mathbf{I})^{-1} + (\mathbf{W}_0 - (z - x)\mathbf{I})^{-1}((1/N)\mathbf{STS}^\dagger - x\mathbf{I})(\mathbf{W} - zI)^{-1}.$$

Dividing by N, taking traces and using Lemma 2.1 we find

$$\mathcal{S}_{\mathbf{W}_0}(z-x)-\mathcal{S}_{\mathbf{W}}(z) = (1/N)\mathsf{tr}\,(\mathbf{W}_0-(z-x)\mathbf{I})^{-1}\left(\sum_{i=1}^{K} t_i q_i q_i^\dagger - x\mathbf{I}\right)(\mathbf{W}-zI)^{-1}$$

$$= (1/N)\sum_{i=1}^{n} \frac{t_i q_i^\dagger(\mathbf{W}_{(i)} - zI)^{-1}(\mathbf{W}_0 - (z-x)\mathbf{I})^{-1} q_i}{1 + t_i q_i^\dagger(\mathbf{W}_{(i)} - zI)^{-1} q_i}$$

$$- x(1/N)\mathsf{tr}\,(\mathbf{W} - z\mathbf{I})^{-1}(\mathbf{W}_0 - (z-x)\mathbf{I})^{-1}.$$

Notice when x and q_i are independent, Lemmas 2.2, 2.3 give us

$$q_i^\dagger(\mathbf{W}_{(i)}-z\mathbf{I})^{-1}(\mathbf{W}_0-(z-x)\mathbf{I})^{-1}q_i \approx (1/N)\mathsf{tr}\,(\mathbf{W}-z\mathbf{I})^{-1}(\mathbf{W}_0-(z-x)\mathbf{I})^{-1}.$$

Letting

$$x = x_N = (1/N) \sum_{i=1}^{K} \frac{t_i}{1 + t_i \mathcal{S}_{\mathbf{W}}(z)}$$

we have

$$\mathcal{S}_{\mathbf{W}_0}(z - x_N) - \mathcal{S}_{\mathbf{W}}(z) = (1/N) \sum_{i=1}^{K} \frac{t_i}{1 + t_i \mathcal{S}_{\mathbf{W}}(z)} d_i$$

where

$$d_i = \frac{1 + t_i \mathcal{S}_{\mathbf{W}}(z)}{1 + t_i q_i^\dagger (\mathbf{W}_{(i)} - zI)^{-1} q_i} q_i^\dagger (\mathbf{W}_{(i)} - z\mathbf{I})^{-1} (\mathbf{W}_0 - (z - x_N)\mathbf{I})^{-1} q_i$$
$$- (1/N) \mathsf{tr}\, (\mathbf{W} - z\mathbf{I})^{-1} (\mathbf{W}_0 - (z - x_N)\mathbf{I})^{-1}.$$

In order to use Lemma 2.3, for each i, x_N is replaced by

$$x_{(i)} = (1/N) \sum_{j=1}^{K} \frac{t_j}{1 + t_j \mathcal{S}_{\mathbf{W}_{(i)}}(z)}.$$

Using Lemma 2.3 ($p = 6$ is sufficient) and the fact that all matrix inverses encountered are bounded in spectral norm by $1/\Im z$ we have from standard arguments using Boole's and Markov's inequalities, and the Borel-Cantelli lemma, almost surely

$$\text{(2.1)} \qquad \max_{i\le K} \max[|\,\|q_i\|^2 - 1|, |q_i^\dagger(\mathbf{W}_{(i)} - zI)^{-1}q_i - \mathcal{S}_{\mathbf{W}_{(i)}}(z)|,$$

$$|q_i^\dagger(\mathbf{W}_{(i)} - zI)^{-1}(\mathbf{W}_0 - (z - x_{(i)})\mathbf{I})^{-1}q_i - (1/N)\mathsf{tr}\,(\mathbf{W}_{(i)} - z\mathbf{I})^{-1}(\mathbf{W}_0 - (z - x_{(i)})\mathbf{I})^{-1}|]$$

$$\to 0 \quad \text{as } N \to \infty.$$

This and Lemma 2.2 imply almost surely

$$\text{(2.2)} \quad \max_{i\le K} \max[|\mathcal{S}_{\mathbf{W}}(z) - \mathcal{S}_{\mathbf{W}_{(i)}}(z)|, |\mathcal{S}_{\mathbf{W}}(z) - q_i^\dagger(\mathbf{W}_{(i)} - zI)^{-1}q_i|] \to 0,$$

and subsequently, almost surely

$$\max_{i\leq K} \max \left[\left| \frac{1 + t_i \mathcal{S}_{\mathbf{W}}(z)}{1 + t_i q_i^\dagger (\mathbf{W}_{(i)} - zI)^{-1} q_i} - 1 \right|, |x - x_{(i)}| \right] \to 0. \tag{2.3}$$

Therefore, from Lemmas 2.2, 2.4, and (2.1) -(2.3), we get $\max_{i\leq K} d_i \to 0$ almost surely, giving us

$$\mathcal{S}_{\mathbf{W}_0}(z - x_N) - \mathcal{S}_{\mathbf{W}}(z) \to 0,$$

almost surely.

On any realization for which the above holds and $F^{\mathbf{W}_0} \xrightarrow{v} \mathcal{W}_0$, consider any subsequence which $\mathcal{S}_{\mathbf{W}}(z)$ converges to, say, $\mathcal{S}$, then, on this subsequence

$$x_N = (K/N)\frac{1}{K}\sum_{i=1}^{K}\frac{t_i}{1+t_i\mathcal{S}_{\mathbf{W}}(z)} \to \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]$$

Therefore, in the limit we have

$$\mathcal{S} = \mathcal{S}_0\left(z - \beta\,\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right),$$

which is (1.1). Uniqueness gives us, for this realization, $\mathcal{S}_{\mathbf{W}}(z) \to \mathcal{S}$ as $N \to \infty$. This event occurs with probability one.

3. Proof of uniqueness of (1.1). For $\mathcal{S} \in \mathbb{C}^+$ satisfying (1.1) with $z \in \mathbb{C}^+$ we have

$$\mathcal{S} = \int \frac{1}{\tau - \left(z - \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right)} d\mathcal{W}_0(\tau)$$

$$= \int \frac{1}{\tau - \Re\left(z - \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right) - i\left(\Im z + \beta\mathbb{E}\left[\frac{\mathsf{T}^2\Im\mathcal{S}}{|1+\mathsf{T}\mathcal{S}|^2}\right]\right)} d\mathcal{W}_0(\tau)$$

Therefore

(3.1)
$$\Im\mathcal{S} = \left(\Im z + \beta\mathbb{E}\left[\frac{\mathsf{T}^2\Im\mathcal{S}}{|1+\mathsf{T}\mathcal{S}|^2}\right]\right) \int \frac{1}{\left|\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right|^2} d\mathcal{W}_0(\tau)$$

Suppose $\boldsymbol{\mathcal{S}} \in \mathbb{C}^+$ also satisfies (1.1). Then

(3.2)

$$\mathcal{S} - \boldsymbol{\mathcal{S}} = \beta \int \frac{\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\boldsymbol{\mathcal{S}}} - \frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]}{\left(\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right)\left(\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\boldsymbol{\mathcal{S}}}\right]\right)} d\mathcal{W}_0(\tau)$$

$$= (\mathcal{S} - \boldsymbol{\mathcal{S}})\beta\mathbb{E}\left[\frac{\mathsf{T}^2}{(1+\mathsf{T}\mathcal{S})(1+\mathsf{T}\boldsymbol{\mathcal{S}})}\right]$$

$$\times \int \frac{1}{\left(\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right)\left(\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\boldsymbol{\mathcal{S}}}\right]\right)} d\mathcal{W}_0(\tau).$$

Using Cauchy-Schwarz and (3.1) we have

$$\Bigg|\beta\mathbb{E}\left[\frac{\mathsf{T}^2}{(1+\mathsf{T}\mathcal{S})(1+\mathsf{T}\boldsymbol{\mathcal{S}})}\right]$$

$$\times \int \frac{1}{\left(\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right)\left(\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\boldsymbol{\mathcal{S}}}\right]\right)} d\mathcal{W}_0(\tau)\Bigg|$$

$$\leq \left(\beta\mathbb{E}\left[\frac{\mathsf{T}^2}{|1+\mathsf{T}\mathcal{S}|^2}\right]\int\frac{1}{\left|\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\mathcal{S}}\right]\right|^2}d\mathcal{W}_0(\tau)\right)^{1/2}$$

$$\times\left(\beta\mathbb{E}\left[\frac{\mathsf{T}^2}{|1+\mathsf{T}\boldsymbol{\mathcal{S}}|^2}\right]\int\frac{1}{\left|\tau - z + \beta\mathbb{E}\left[\frac{\mathsf{T}}{1+\mathsf{T}\boldsymbol{\mathcal{S}}}\right]\right|^2}d\mathcal{W}_0(\tau)\right)^{1/2}$$

$$= \left(\beta\mathbb{E}\left[\frac{\mathsf{T}^2}{|1+\mathsf{T}\mathcal{S}|^2}\right]\frac{\Im\mathcal{S}}{\left(\Im z + \beta\mathbb{E}\left[\frac{\mathsf{T}^2\Im\mathcal{S}}{|1+\mathsf{T}\mathcal{S}|^2}\right]\right)}\right)^{1/2}$$

$$\times\left(\beta\mathbb{E}\left[\frac{\mathsf{T}^2}{|1+\mathsf{T}\boldsymbol{\mathcal{S}}|^2}\right]\frac{\Im\boldsymbol{\mathcal{S}}}{\left(\Im z + \beta\mathbb{E}\left[\frac{\mathsf{T}^2\Im\boldsymbol{\mathcal{S}}}{|1+\mathsf{T}\boldsymbol{\mathcal{S}}|^2}\right]\right)}\right)^{1/2} < 1.$$

Therefore, from (3.2) we must have $\mathcal{S} = \boldsymbol{\mathcal{S}}$.