Random Matrices in Wireless Communications
Course 1: Introduction to random matrix theory and the Stieltjes transform

Romain Couillet
ST-Ericsson, Supélec, FRANCE
romain.couillet@supelec.fr

Supélec

Outline

Outline

High-dimensional data

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}$$

  is a "good" estimator of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

  What about the global behaviour? What about the eigenvalue distribution?

Assume $\mathbf{R} = \mathbf{I}_N$ and draw the eigenvalues of $\mathbf{R}_n$ for n, N large.

High-dimensional data

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^H]$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H$$

  is a "good" estimator of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

What about the global behaviour? What about the eigenvalue distribution?

Assume $\mathbf{R} = \mathbf{I}_N$ and draw the eigenvalues of $\mathbf{R}_n$ for n, N large.

High-dimensional data

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$$

  is a "good" estimator of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

What about the global behaviour? What about the eigenvalue distribution?

Assume $\mathbf{R} = \mathbf{I}_N$ and draw the eigenvalues of $\mathbf{R}_n$ for n, N large.

High-dimensional data

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$$

  is a “good” estimator of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

What about the global behaviour? What about the eigenvalue distribution?

Assume $\mathbf{R} = \mathbf{I}_N$ and draw the eigenvalues of $\mathbf{R}_n$ for n, N large.

High-dimensional data

Let $\mathbf{x}_1, \mathbf{x}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}$$

  is a “good” estimator of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

  What about the global behaviour? What about the eigenvalue distribution?

Assume $\mathbf{R} = \mathbf{I}_N$ and draw the eigenvalues of $\mathbf{R}_n$ for n, N large.

Empirical and limit spectra of Wishart matrices

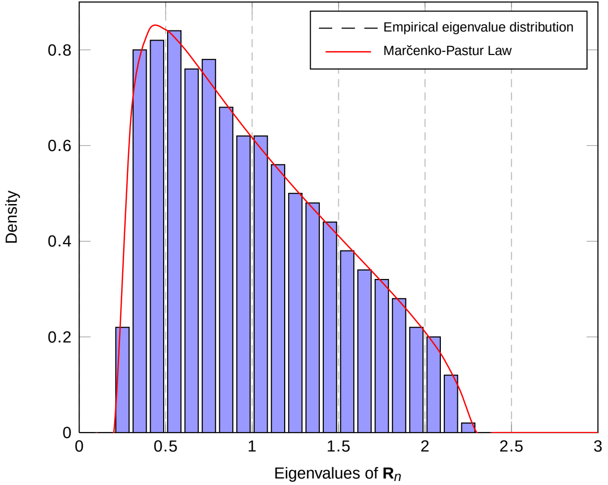

Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

Definition

Definition

Let Ω be some probability space, and let $\omega \in \Omega$. A random matrix $\mathbf{X} = \mathbf{X}(\omega)$ is a random variable whose value lies in some matrix space.

Note:

- the probability space Ω is often neglected; it is e.g. the propagation environment for MIMO channel matrices.
- for asymptotic considerations, $\omega \in \Omega$ will be the realization of an infinite *sequence* $\mathbf{X}_1(\omega), X_2(\omega), \ldots$ of size $1, 2, \ldots$ random matrices.

In practice, we are mostly interested into Hermitian matrices and especially in the distribution of their eigenvalues.

Definition

The distribution function F_N of the eigenvalues of the $N \times N$ random Hermitian matrix $\mathbf{X}_N = \mathbf{X}_N(\omega)$ is called the empirical spectrum distribution (e.s.d.) of $\mathbf{X}_N$. If F_N has a limit F when $N \to \infty$, this limit is called the limit spectral distribution of $\mathbf{X}_N$.

Definition

Definition

Let Ω be some probability space, and let $\omega \in \Omega$. A random matrix $\mathbf{X} = \mathbf{X}(\omega)$ is a random variable whose value lies in some matrix space.

Note:

- the probability space Ω is often neglected; it is e.g. the propagation environment for MIMO channel matrices.
- for asymptotic considerations, $\omega \in \Omega$ will be the realization of an infinite *sequence* $\mathbf{X}_1(\omega), X_2(\omega), \ldots$ of size $1, 2, \ldots$ random matrices.

In practice, we are mostly interested into Hermitian matrices and especially in the distribution of their eigenvalues.

Definition

The distribution function F_N of the eigenvalues of the $N \times N$ random Hermitian matrix $\mathbf{X}_N = \mathbf{X}_N(\omega)$ is called the empirical spectrum distribution (e.s.d.) of $\mathbf{X}_N$. If F_N has a limit F when $N \to \infty$, this limit is called the limit spectral distribution of $\mathbf{X}_N$.

Finite size and asymptotic considerations

The field of random matrices is often segmented into

- *Finite-size random matrices*:
  - of interest are: joint entry distributions, ordered eigenvalue distributions, e.s.d., expectation of functionals
  - particularly suitable to small size matrices
  - however, much problems arise for models more involved than i.i.d. Gaussian
- *Limiting results*:
  - of interest are: limit spectral distributions (l.s.d.), functionals of l.s.d., central limit theorems etc.
  - suitable to large matrices, but often good approximation to smaller matrices
  - much easier to work with than finite size, more flexible (i.i.d., Kronecker, variance profile models, structured matrices)
  - possesses a variety of powerful tools: Stieltjes transform, free probability

Remark: This course will mainly focus on limiting results and almost no finite size considerations.

Finite size and asymptotic considerations

The field of random matrices is often segmented into

- *Finite-size random matrices*:
  - of interest are: joint entry distributions, ordered eigenvalue distributions, e.s.d., expectation of functionals
  - particularly suitable to small size matrices
  - however, much problems arise for models more involved than i.i.d. Gaussian
- *Limiting results*:
  - of interest are: limit spectral distributions (l.s.d.), functionals of l.s.d., central limit theorems etc.
  - suitable to large matrices, but often good approximation to smaller matrices
  - much easier to work with than finite size, more flexible (i.i.d., Kronecker, variance profile models, structured matrices)
  - possesses a variety of powerful tools: Stieltjes transform, free probability

Remark: This course will mainly focus on limiting results and almost no finite size considerations.

Why is this useful to wireless communications?

- increasing number of parameters: multi-user systems, multiple concurrent cells, multiple antennas
- matrices with random entries are the basis for MIMO channels, CDMA codes
- it is no longer possible to treat large dimensional problems with classical probability approaches
- random matrices answer a widening panel of problems: system performance, detection, estimation...

Example

MIMO channel capacity Call $\mathbf{H} \in \mathbb{C}^{n \times N}$ the realization of a MIMO channel matrix whose entries and distributed according to some random process. We have the per-antenna mutual information

$$C(\sigma^2) = \frac{1}{N} \log\det \left[\mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{H}\mathbf{H}^{\mathsf{H}} \right]$$

Note that, with $\mathbf{h}_i$ the i^{th} column of $\mathbf{H}$, $\mathbf{H}\mathbf{H}^{\mathsf{H}} = \sum_{i=1}^{N} \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}}$. If $\mathbf{H}$ has i.i.d. entries, then, as both $n, N \to \infty$, $n/N \to c$,

$$C(\sigma^2) \to \int \log \left[1 + \frac{t}{\sigma^2} \right] dF_c(t)$$

with F_c the Marčenko-Pastur law with parameter c.

Why is this useful to wireless communications?

- increasing number of parameters: multi-user systems, multiple concurrent cells, multiple antennas
- matrices with random entries are the basis for MIMO channels, CDMA codes
- it is no longer possible to treat large dimensional problems with classical probability approaches
- random matrices answer a widening panel of problems: system performance, detection, estimation...

Example

MIMO channel capacity Call $\mathbf{H} \in \mathbb{C}^{n \times N}$ the realization of a MIMO channel matrix whose entries and distributed according to some random process. We have the per-antenna mutual information

$$C(\sigma^2) = \frac{1}{N} \log\det \left[\mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{H}\mathbf{H}^{\mathrm{H}} \right]$$

Note that, with $\mathbf{h}_i$ the i^{th} column of $\mathbf{H}$, $\mathbf{H}\mathbf{H}^{\mathrm{H}} = \sum_{i=1}^{N} \mathbf{h}_i \mathbf{h}_i^{\mathrm{H}}$. If $\mathbf{H}$ has i.i.d. entries, then, as both $n, N \to \infty$, $n/N \to c$,

$$C(\sigma^2) \to \int \log \left[1 + \frac{t}{\sigma^2} \right] dF_c(t)$$

with F_c the Marčenko-Pastur law with parameter c.

Outline

Wishart matrices

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

- First random matrix considerations date back to Wishart (1928) who studies the joint distribution of *Gaussian sample covariance matrices* $\mathbf{R}_n = \mathbf{X}\mathbf{X}^{\mathsf{H}} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$, $\mathbf{x}_i \in \mathbb{C}^N \sim \mathcal{N}(0, \mathbf{R})$,

$$P_{\mathbf{R}_n}(\mathbf{B}) = \frac{\pi^{N(N-1)/2}}{\det \mathbf{R}^n \prod_{i=1}^{N}(n-i)!} e^{-\operatorname{tr}(\mathbf{R}^{-1}\mathbf{B})} \det \mathbf{B}^{n-N}$$

- Subsequent work provide expressions of the joint and marginal eigenvalue distributions,

$$P_{(\lambda_i)}(\lambda_1, \ldots, \lambda_N) = \frac{\det(\{e^{-r_j^{-1}\lambda_i}\}_N)}{\Delta(\mathbf{R}^{-1})} \Delta(\mathbf{L}) \prod_{j=1}^{N} \frac{\lambda_j^{n-N}}{j!(n-j)!}$$

with $r_1 \geq \ldots \geq r_N$ the eigenvalues of $\mathbf{R}$ and $\mathbf{L} = \operatorname{diag}(\lambda_1 \geq \ldots \geq \lambda_N)$ and

$$p_\lambda(\lambda) = \frac{1}{M} \sum_{k=0}^{N-1} \frac{k!}{(k+n-N)!} [L_k^{n-N}]^2 \lambda^{n-N} e^{-\lambda}$$

where L_n^k are the Laguerre polynomials defined as

$$L_n^k(\lambda) = \frac{e^\lambda}{k!\lambda^n} \frac{d^k}{d\lambda^k}(e^{-\lambda}\lambda^{n+k})$$

Semi-circle law, Full circle law...

- First asymptotic approach is due to Wigner for nuclear physics purposes

E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," The annals of mathematics, vol. 62, pp. 546-564, 1955.

If $\mathbf{X}_N \in \mathbb{C}^{N \times N}$ is Hermitian with i.i.d. entries of mean 0, variance $1/N$, then $F^{\mathbf{X}_N} \xrightarrow{\text{a.s.}} F$ where F has density f the semi-circle law

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

- If $\mathbf{X}_N \in \mathbb{C}^{N \times N}$ has with i.i.d. 0 mean, variance $1/N$ entries, then asymptotically its complex eigenvalues distribute uniformly on the complex unit circle.

Semi-circle law

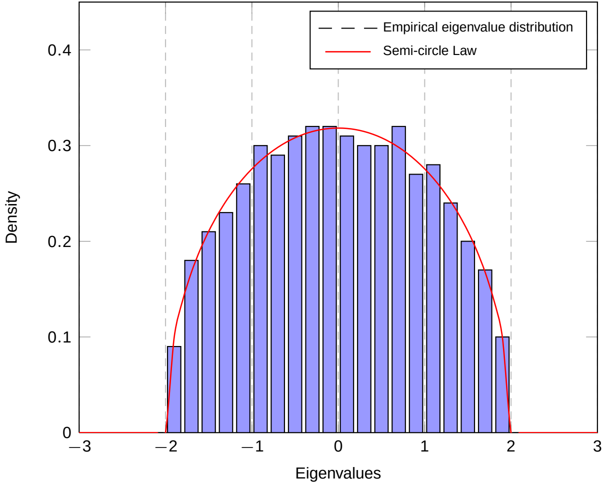

Figure: Histogram of the eigenvalues of Wigner matrices and the semi-circle law, for $N = 500$

Circular law

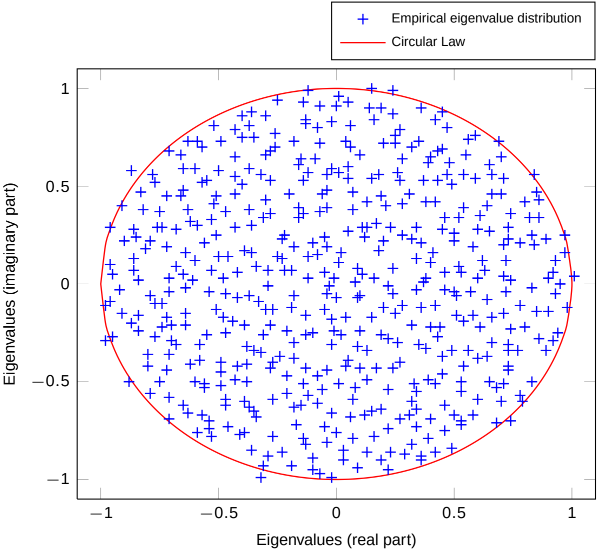

Figure: Eigenvalues of $\mathbf{X}_N$ with i.i.d. standard Gaussian entries, for $N = 500$.

More involved matrix models

- much study has surrounded the Marčenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have an expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

Tools for random matrix theory

To study these models, a consistent powerful mathematical framework is required.

More involved matrix models

- much study has surrounded the Marčenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have an expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

Tools for random matrix theory

To study these models, a consistent powerful mathematical framework is required.

More involved matrix models

- much study has surrounded the Marčenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have an expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

Tools for random matrix theory

To study these models, a consistent powerful mathematical framework is required.

Outline

Eigenvalue distribution and moments

- Moments of eigenvalue distributions,
  - The e.s.d. of an $N \times N$ Hermitian matrix $\mathbf{X}_N(\omega)$ has successive *empirical* moments $\hat{M}_k$, $k = 1, 2, \ldots$,

$$\hat{M}_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$$

  - if F_N denotes the e.s.d. of $\mathbf{X}_N(\omega)$, M_k is

$$\hat{M}_k = \int \lambda^k dF(\lambda)$$

- In classical probability theory, if A and B are independent, the moments of $A + B$ are functions of the moments of A and those of B. In particular, for A, B independent,

$$c_k(A + B) = c_k(A) + c_k(B)$$

  with $c_k(X)$ the cumulants of X (polynomial functions of the moments m_k of X).
- The cumulants c_n are connected to the moments m_n through formulas invoking partitions,

$$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{|V|}$$

- If $\mathbf{A}$, $\mathbf{B}$ are Hermitian matrices, we feel that, if they have independent entries, there should exist a relationship between the *eigenvalue distribution moments* $M_k(\mathbf{A} + \mathbf{B}) = \mathrm{E}_\omega[\hat{M}_k(\mathbf{A}(\omega) + \mathbf{B}(\omega))]$

D. V. Voiculescu, K. J. Dykema, A. Nica, "Free random variables," American Mathematical Society, 1992.

Eigenvalue distribution and moments

- Moments of eigenvalue distributions,
  - The e.s.d. of an $N \times N$ Hermitian matrix $\mathbf{X}_N(\omega)$ has successive *empirical* moments $\hat{M}_k$, $k = 1, 2, \ldots$,

$$\hat{M}_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$$

  - if F_N denotes the e.s.d. of $\mathbf{X}_N(\omega)$, M_k is

$$\hat{M}_k = \int \lambda^k dF(\lambda)$$

- In classical probability theory, if A and B are independent, the moments of $A+B$ are functions of the moments of A and those of B. In particular, for A, B independent,

$$c_k(A+B) = c_k(A) + c_k(B)$$

  with $c_k(X)$ the cumulants of X (polynomial functions of the moments m_k of X).
- The cumulants c_n are connected to the moments m_n through formulas invoking partitions,

$$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{|V|}$$

- If $\mathbf{A}$, $\mathbf{B}$ are Hermitian matrices, we feel that, if they have independent entries, there should exist a relationship between the *eigenvalue distribution moments* $M_k(\mathbf{A}+\mathbf{B}) = \mathrm{E}_\omega[\hat{M}_k(\mathbf{A}(\omega)+\mathbf{B}(\omega))]$

D. V. Voiculescu, K. J. Dykema, A. Nica, "Free random variables," American Mathematical Society, 1992.

Eigenvalue distribution and moments

- Moments of eigenvalue distributions,
  - The e.s.d. of an $N \times N$ Hermitian matrix $\mathbf{X}_N(\omega)$ has successive *empirical* moments $\hat{M}_k$, $k = 1, 2, \ldots$,

$$\hat{M}_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$$

  - if F_N denotes the e.s.d. of $\mathbf{X}_N(\omega)$, M_k is

$$\hat{M}_k = \int \lambda^k dF(\lambda)$$

- In classical probability theory, if A and B are independent, the moments of $A + B$ are functions of the moments of A and those of B. In particular, for A, B independent,

$$c_k(A + B) = c_k(A) + c_k(B)$$

  with $c_k(X)$ the cumulants of X (polynomial functions of the moments m_k of X).
- The cumulants c_n are connected to the moments m_n through formulas invoking partitions,

$$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{|V|}$$

- If $\mathbf{A}$, $\mathbf{B}$ are Hermitian matrices, we feel that, if they have independent entries, there should exist a relationship between the *eigenvalue distribution moments* $M_k(\mathbf{A} + \mathbf{B}) = \mathrm{E}_\omega[\hat{M}_k(\mathbf{A}(\omega) + \mathbf{B}(\omega))]$

D. V. Voiculescu, K. J. Dykema, A. Nica, “Free random variables,” American Mathematical Society, 1992.

Eigenvalue distribution and moments

- Moments of eigenvalue distributions,
  - The e.s.d. of an $N \times N$ Hermitian matrix $\mathbf{X}_N(\omega)$ has successive *empirical* moments $\hat{M}_k$, $k = 1, 2, \ldots$,

$$\hat{M}_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$$

  - if F_N denotes the e.s.d. of $\mathbf{X}_N(\omega)$, M_k is

$$\hat{M}_k = \int \lambda^k dF(\lambda)$$

- In classical probability theory, if A and B are independent, the moments of $A + B$ are functions of the moments of A and those of B. In particular, for A, B independent,

$$c_k(A + B) = c_k(A) + c_k(B)$$

  with $c_k(X)$ the cumulants of X (polynomial functions of the moments m_k of X).
- The cumulants c_n are connected to the moments m_n through formulas invoking partitions,

$$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{|V|}$$

- If $\mathbf{A}$, $\mathbf{B}$ are Hermitian matrices, we feel that, if they have independent entries, there should exist a relationship between the *eigenvalue distribution moments* $M_k(\mathbf{A} + \mathbf{B}) = \mathrm{E}_\omega[\hat{M}_k(\mathbf{A}(\omega) + \mathbf{B}(\omega))]$

D. V. Voiculescu, K. J. Dykema, A. Nica, “Free random variables,” American Mathematical Society, 1992.

Eigenvalue distribution and moments

- Moments of eigenvalue distributions,
  - The e.s.d. of an $N \times N$ Hermitian matrix $\mathbf{X}_N(\omega)$ has successive *empirical* moments $\hat{M}_k$, $k = 1, 2, \ldots$,

$$\hat{M}_k = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$$

  - if F_N denotes the e.s.d. of $\mathbf{X}_N(\omega)$, M_k is

$$\hat{M}_k = \int \lambda^k dF(\lambda)$$

- In classical probability theory, if A and B are independent, the moments of $A + B$ are functions of the moments of A and those of B. In particular, for A, B independent,

$$c_k(A + B) = c_k(A) + c_k(B)$$

  with $c_k(X)$ the cumulants of X (polynomial functions of the moments m_k of X).
- The cumulants c_n are connected to the moments m_n through formulas invoking partitions,

$$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{|V|}$$

- If **A**, **B** are Hermitian matrices, we feel that, if they have independent entries, there should exist a relationship between the *eigenvalue distribution moments* $M_k(\mathbf{A} + \mathbf{B}) = \mathrm{E}_\omega[\hat{M}_k(\mathbf{A}(\omega) + \mathbf{B}(\omega))]$

D. V. Voiculescu, K. J. Dykema, A. Nica, “Free random variables,” American Mathematical Society, 1992.

Free probability

Free probability applies to asymptotically large random matrices. We assume here all matrices have infinite size

- To connect the moments of $\mathbf{A} + \mathbf{B}$ to those of $\mathbf{A}$ and $\mathbf{B}$, independence is not enough. One needs for $\mathbf{A} = \mathbf{A}(\omega)$ and $\mathbf{B}(\omega)$ to be realizations of free sub-algebras of random matrices. Roughly speaking, $\mathbf{A}$ and $\mathbf{B}$ need to be independent and to have "disconnected eigen-directions".
  - two Gaussian matrices are free
  - a Gaussian matrix and any deterministic matrix are free
  - unitary (Haar distributed) matrices are free
  - a Haar matrix and a Gaussian matrix are free etc.
- Similarly as in classical probability, we define free cumulants C_k,

$$C_1 = M_1$$
$$C_2 = M_2 - M_1^2$$
$$C_3 = M_3 - 3M_1M_2 + 2M_1^2$$

R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," Mem. A.M.S., vol. 627, 1998.

- A combinatorial description of the relation moments-cumulants invokes non-crossing partitions,

$$M_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{V \in \pi} C_{|V|}$$

Free probability

Free probability applies to asymptotically large random matrices. We assume here all matrices have infinite size

- To connect the moments of $\mathbf{A}+\mathbf{B}$ to those of $\mathbf{A}$ and $\mathbf{B}$, independence is not enough. One needs for $\mathbf{A}=\mathbf{A}(\omega)$ and $\mathbf{B}(\omega)$ to be realizations of free sub-algebras of random matrices. Roughly speaking, $\mathbf{A}$ and $\mathbf{B}$ need to be independent and to have "disconnected eigen-directions".
  - two Gaussian matrices are free
  - a Gaussian matrix and any deterministic matrix are free
  - unitary (Haar distributed) matrices are free
  - a Haar matrix and a Gaussian matrix are free etc.
- Similarly as in classical probability, we define free cumulants C_k,

$$
\begin{aligned}
C_1 &= M_1 \\
C_2 &= M_2 - M_1^2 \\
C_3 &= M_3 - 3M_1M_2 + 2M_1^2
\end{aligned}
$$

R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," Mem. A.M.S., vol. 627, 1998.

- A combinatorial description of the relation moments-cumulants invokes non-crossing partitions,

$$M_n = \sum_{\pi \in \mathcal{N}C(n)} \prod_{V \in \pi} C_{|V|}$$

Non-crossing partitions

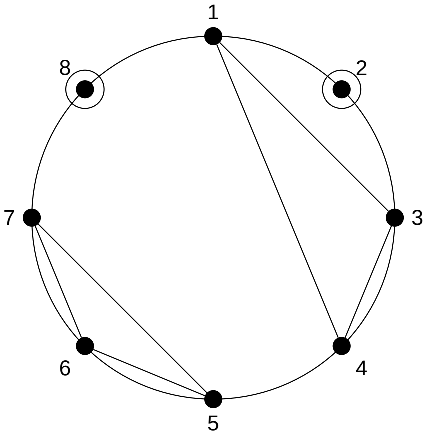

Figure: Non-crossing partition $\pi = \{\{1,3,4\},\{2\},\{5,6,7\},\{8\}\}$ of $NC(8)$.

Moments of sums and products of random matrices

- Combinatorial calculus of all moments

Theorem

For free random matrices **A** *and* **B***, we have the relationship,*

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1\\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A})C_{|V_2|}(\mathbf{B})$$

in conjunction with free moment-cumulant formula, gives all moments of sum and product.

- Denote $m_F(z)$ the moment-generating function of the l.s.d. F of a random Hermitian matrix $\mathbf{X}$, also called *Stieltjes transform*,

$$m_F(z) = -\sum_{k=0}^{\infty} M_k z^{-k-1}$$

Theorem

If F is a compactly supported distribution function, then m_F above exists for all $z \in \mathbb{C}^*$ and gives access to F through an inverse Stieltjes-transform formula (see Section 23).

- In the absence of support compactness, it is impossible to retrieve the distribution function from moments. This is in particular the case of Vandermonde matrices.

Moments of sums and products of random matrices

- Combinatorial calculus of all moments

Theorem

For free random matrices **A** *and* **B**, *we have the relationship,*

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1 \\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A})C_{|V_2|}(\mathbf{B})$$

in conjunction with free moment-cumulant formula, gives all moments of sum and product.

- Denote $m_F(z)$ the moment-generating function of the l.s.d. F of a random Hermitian matrix $\mathbf{X}$, also called *Stieltjes transform*,

$$m_F(z) = -\sum_{k=0}^{\infty} M_k z^{-k-1}$$

Theorem

If F is a compactly supported distribution function, then m_F above exists for all $z \in \mathbb{C}^*$ and gives access to F through an inverse Stieltjes-transform formula (see Section 23).

- In the absence of support compactness, it is impossible to retrieve the distribution function from moments. This is in particular the case of Vandermonde matrices.

Moments of sums and products of random matrices

- Combinatorial calculus of all moments

Theorem

For free random matrices **A** *and* **B***, we have the relationship,*

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1 \\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A})C_{|V_2|}(\mathbf{B})$$

in conjunction with free moment-cumulant formula, gives all moments of sum and product.

- Denote $m_F(z)$ the moment-generating function of the l.s.d. F of a random Hermitian matrix $\mathbf{X}$, also called *Stieltjes transform*,

$$m_F(z) = -\sum_{k=0}^{\infty} M_k z^{-k-1}$$

Theorem

If F is a compactly supported distribution function, then m_F above exists for all $z \in \mathbb{C}^*$ and gives access to F through an inverse Stieltjes-transform formula (see Section 23).

- In the absence of support compactness, it is impossible to retrieve the distribution function from moments. This is in particular the case of Vandermonde matrices.

Free convolution

- In classical probability theory, for independent A, B,

$$f_{A+B}(x) = f_A(x) * f_B(x) \triangleq \int f_A(t) f_B(x-t) dt$$

- In free probability, for free **A**, **B**, we use the notations

$$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}},\ \mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxslash \mu_{\mathbf{B}}$$

Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.

Theorem

Convolution of the information-plus-noise model Let $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ have i.i.d. Gaussian entries of mean 0 and variance 1, $\mathbf{R}_N \in \mathbb{C}^{N \times n}$, such that $\mu_{\frac{1}{n}\mathbf{R}_N\mathbf{R}_N^{\mathsf{H}}} \Rightarrow \mu_\Gamma$, as $n/N \to c$. Then the e.s.d. of

$$\mathbf{B}_N = \frac{1}{n}(\mathbf{R}_N + \sigma\mathbf{X}_N)(\mathbf{R}_N + \sigma\mathbf{X}_N)^{\mathsf{H}}$$

converges weakly and almost surely to μ_B such that

$$\mu_B = ((\mu_\Gamma \boxslash \mu_c) \boxplus \delta_{\sigma^2}) \boxtimes \mu_c$$

with μ_c the Marčenko-Pastur law.

Free convolution

- In classical probability theory, for independent A, B,

$$f_{A+B}(x) = f_A(x) * f_B(x) \triangleq \int f_A(t) f_B(x-t)dt$$

- In free probability, for free $\mathbf{A}$, $\mathbf{B}$, we use the notations

$$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}},\ \mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxslash \mu_{\mathbf{B}}$$

Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.

Theorem

Convolution of the information-plus-noise model Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *have i.i.d. Gaussian entries of mean* 0 *and variance* 1*,* $\mathbf{R}_N \in \mathbb{C}^{N\times n}$*, such that* $\mu_{\frac{1}{n}\mathbf{R}_N\mathbf{R}_N^{\mathsf{H}}} \Rightarrow \mu_\Gamma$*, as* $n/N \to c$*. Then the e.s.d. of*

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)\left(\mathbf{R}_N + \sigma\mathbf{X}_N\right)^{\mathsf{H}}$$

converges weakly and almost surely to μ_B *such that*

$$\mu_B = \left(\left(\mu_\Gamma \boxslash \mu_c\right) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

with μ_c *the Marčenko-Pastur law.*

Similarities between classical and free probability

	Classical Probability	Free probability
Moments	$m_k = \int x^k dF(x)$	$M_k = \int x^k dF(x)$
Cumulants	$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{\lvert V \rvert}$	$M_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{V \in \pi} C_{\lvert V \rvert}$
Independence	classical independence	freeness
Additive convolution	$f_{A+B} = f_A * f_B$	$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$
Multiplicative convolution	f_{AB}	$\mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$
Sum Rule	$c_k(A+B) = c_k(A) + c_k(B)$	$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$
Central Limit	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \to \mathcal{N}(0,1)$	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \Rightarrow$ semi-circle law

Bibliography on Free Probability related work

- D. Voiculescu, "Addition of certain non-commuting random variables," Journal of functional analysis, vol. 66, no. 3, pp. 323-346, 1986.
- R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," Mem. A.M.S., vol. 627, 1998.
- R. Seroul, D. O'Shea, "Programming for Mathematicians," Springer, 2000.
- H. Bercovici, V. Pata, "The law of large numbers for free identically distributed random variables," The Annals of Probability, pp. 453-465, 1996.
- A. Nica, R. Speicher, "On the multiplication of free N-tuples of noncommutative random variables," American Journal of Mathematics, pp. 799-837, 1996.
- Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.
- N. R. Rao, A. Edelman, "The polynomial method for random matrices," Foundations of Computational Mathematics, vol. 8, no. 6, pp. 649-702, 2008.
- Ø. Ryan, M. Debbah, "Asymptotic Behavior of Random Vandermonde Matrices With Entries on the Unit Circle," IEEE Trans. on Information Theory, vol. 55, no. 7, pp. 3115-3147, 2009.

Outline

The Stieltjes transform

Definition

Let F be a probability distribution function. The Stieltjes transform m_F of F is the function defined, for $z \in \mathbb{C}^+$, as

$$m_F(z) = \int \frac{1}{\lambda - z} dF(\lambda)$$

For $a < b$ real, denoting $z = x + iy$, we have the inverse formula

$$F([a,b]) = \lim_{y \to 0} \frac{1}{\pi} \int_a^b \Im[m_F(x+iy)]$$

Remark on the Stieltjes transform

- If F is the e.s.d. of a Hermitian matrix $\mathbf{X}_N \in \mathbb{C}^{N\times N}$, we might denote $m_{\mathbf{X}} \triangleq m_F$, and

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N} \operatorname{tr} (\mathbf{X}_N - z\mathbf{I}_N)^{-1}$$

- We already saw that, for compactly supported F,

$$m_F(z) = -\sum_{k=0}^{\infty} M_k z^{-k-1}$$

The Stieltjes transform is doubly more powerful than the moment approach!

- conveys more information than any K-finite sequence $M_1, \ldots, M_K$.
- is not handicapped by the support compactness constraint.

- however, Stieltjes transform methods, while stronger, are more painful to work with.

Remark on the Stieltjes transform

- If F is the e.s.d. of a Hermitian matrix $\mathbf{X}_N \in \mathbb{C}^{N\times N}$, we might denote $m_{\mathbf{X}} \triangleq m_F$, and

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N} \operatorname{tr} (\mathbf{X}_N - z\mathbf{I}_N)^{-1}$$

- We already saw that, for compactly supported F,

$$m_F(z) = -\sum_{k=0}^{\infty} M_k z^{-k-1}$$

The Stieltjes transform is doubly more powerful than the moment approach!

- conveys more information than any K-finite sequence $M_1, \ldots, M_K$.
- is not handicapped by the support compactness constraint.

- however, Stieltjes transform methods, while stronger, are more painful to work with.

Properties of the Stieltjes transform

- m_F defined in general on $\mathbb{C}^+$ but exists everywhere outside the support of F.
- if $\mathbf{X} \in \mathbb{C}^{N \times n}$, the spectral distribution of $\mathbf{X}\mathbf{X}^{\mathrm{H}}$ and $\mathbf{X}^{\mathrm{H}}\mathbf{X}$ only differ by a mass of $|N - n|$ zeros. Say $N \geq n$,

$$m_{\mathbf{X}\mathbf{X}^{\mathrm{H}}}(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z} = \frac{1}{N}\sum_{i=1}^{n}\frac{1}{\lambda_i - z} + \frac{1}{N}(N-n)\frac{-1}{z}$$

hence

$$m_{\mathbf{X}\mathbf{X}^{\mathrm{H}}}(z) = \frac{n}{N} m_{\mathbf{X}^{\mathrm{H}}\mathbf{X}} - \frac{N-n}{N}\frac{1}{z}$$

Asymptotic results using the Stieltjes transform

J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 175-192, 1995.

Theorem

Let $\underline{\mathbf{B}}_N = \mathbf{X}_N \mathbf{T}_N \mathbf{X}_N^{\mathrm{H}} \in \mathbb{C}^{N \times N}$, *where* $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ *has i.i.d. entries of mean* 0 *and variance* $1/N$, $F^{\mathbf{T}_N} \Rightarrow F^T$ *and* $n/N \to c$. *Then,* $F^{\underline{\mathbf{B}}_N}$ *converges weakly and almost surely to* $\underline{F}$ *with Stieltjes transform*

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1}$$

whose solution is unique in the set $\{z \in \mathbb{C}^+, m_{\underline{F}}(z) \in \mathbb{C}^+\}$.

The proof of a more general theorem will be given in Part 2 of this course.

- in general, no explicit expression for $\underline{F}$.
- the theorem above characterizes also the Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathrm{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with asymptotic distribution F,

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

  This gives access to the spectrum of the sample covariance matrix model, when $\mathbf{X}_N = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, with i.i.d. columns $\mathbf{T}_N = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.

Getting F' from m_F

- Remember that, for $a < b$ real,

$$F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x+iy)]$$

  where m_F is (up to now) only defined on $\mathbb{C}^+$.
  (we will show in Part 3 that it can be somehow extended to $\mathbb{C}^*$)

- to plot the density F',
  - *first approach:* span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.
  - *refined approach:* see Part 3.

Example (Sample covariance matrix)

For N multiple of 3, let $F^{T_N}(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N\mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1}$$

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Getting F' from m_F

- Remember that, for $a < b$ real,

$$F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x + iy)]$$

  where m_F is (up to now) only defined on $\mathbb{C}^+$.
  (we will show in Part 3 that it can be somehow extended to $\mathbb{C}^*$)
- to plot the density F',
  - *first approach:* span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.
  - *refined approach:* see Part 3.

Example (Sample covariance matrix)

For N multiple of 3, let $F^{T_N}(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^H \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1}$$

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Getting F' from m_F

- Remember that, for $a < b$ real,

$$F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x + iy)]$$

  where m_F is (up to now) only defined on $\mathbb{C}^+$.
  (we will show in Part 3 that it can be somehow extended to $\mathbb{C}^*$)
- to plot the density F',
  - *first approach:* span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.
  - *refined approach:* see Part 3.

Example (Sample covariance matrix)

For N multiple of 3, let $F^{T_N}(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N^H\mathbf{X}_N\mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z\right)^{-1}$$

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Getting F' from m_F

- Remember that, for $a < b$ real,

$$F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x+iy)]$$

  where m_F is (up to now) only defined on $\mathbb{C}^+$.
  (we will show in Part 3 that it can be somehow extended to $\mathbb{C}^*$)
- to plot the density F',
  - *first approach:* span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.
  - *refined approach:* see Part 3.

Example (Sample covariance matrix)

For N multiple of 3, let $F'^{\mathbf{T}_N}(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathrm{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1}$$

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Spectrum of the sample covariance matrix

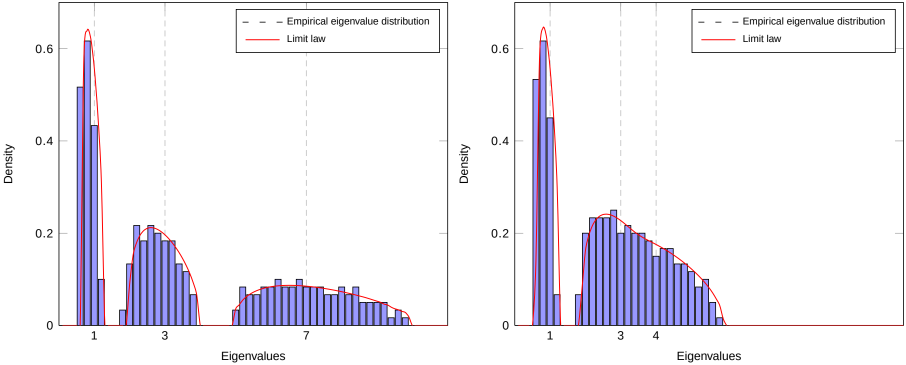

Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^H \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$, $N = 3000$, $n = 300$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

Other notorious result

V. L. Girko, "Theory of Random Determinants," Kluwer, Dordrecht, 1990.

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ with x_{ij} i.i.d. of zero mean and variance σ_{ij}^2/N where the σ_{ij}'s are uniformly bounded. Assume the distribution of σ_{ij} tends to $p_\sigma(x,y)$ as $n,N\to\infty$, $n/N\to c$. Then, almost surely, the e.s.d. of $\mathbf{B}_N = \mathbf{X}_N\mathbf{X}_N^H$ converges weakly to F with Stieltjes transform

$$m_F(z) = \int_0^1 u(x,z)dx$$

and $u(x,z)$ satisfies

$$u(x,z) = \left[-z + \int_0^c \frac{p_\sigma(x,y)dy}{1+\int_0^1 u(x',z)p_\sigma(x',y)dx'}\right]^{-1}$$

Other transforms: the R-transform

- All classically used transforms can be expressed as a function of the Stieltjes transform
- Some transforms are more handy to treat specific problems.

Definition

Let F be a distribution function m_F its Stieltjes transform. Then the R-transform of F is defined as

$$m_F(R_F(z) + z^{-1}) = -z$$

or equivalently

$$m_F(z) = \frac{1}{R_F(-m_F(z)) - z}$$

The main property of the R-transform is that, for $\mathbf{A}$, $\mathbf{B}$ free random matrices,

$$R_{\mathbf{A}+\mathbf{B}} = R_{\mathbf{A}} + R_{\mathbf{B}}$$

Other transforms: the R-transform

- All classically used transforms can be expressed as a function of the Stieltjes transform
- Some transforms are more handy to treat specific problems.

Definition

Let F be a distribution function m_F its Stieltjes transform. Then the R-transform of F is defined as

$$m_F(R_F(z) + z^{-1}) = -z$$

or equivalently

$$m_F(z) = \frac{1}{R_F(-m_F(z)) - z}$$

The main property of the R-transform is that, for **A**, **B** free random matrices,

$$R_{\mathbf{A}+\mathbf{B}} = R_{\mathbf{A}} + R_{\mathbf{B}}$$

Other transforms: the S-transform

Definition

Let F be a distribution function m_F its Stieltjes transform. Then the S-transform of F is defined as

$$m_F\left(\frac{z+1}{zS_F(z)}\right) = -zS_F(z)$$

The S-transform is the product equivalent of the R-transform, i.e. for $\mathbf{A}$, $\mathbf{B}$ free random matrices,

$$S_{\mathbf{AB}} = S_{\mathbf{A}} \cdot S_{\mathbf{B}}$$

Remark: the R- and S-transforms are convenient to use when dealing with unitary matrices.
Example of use is worked out in Part 2.

Other transforms: Shannon and η-transforms

A. M. Tulino, S. Verdù, "Random matrix theory and wireless communications," Now Publishers Inc., 2004.

Definition

Let F be a probability distribution, m_F its Stieltjes transform, then the Shannon-transform $\mathcal{V}_F$ of F is defined as

$$\mathcal{V}_F(x) \triangleq \int_0^\infty \log(1+x\lambda)dF(\lambda) = \int_x^\infty \left(\frac{1}{t} - m_F(-t)\right) dt$$

Note that this last relation is fundamental to wireless communication purposes!

Definition

Let F be a probability distribution, m_F its Stieltjes transform, then the η-transform η_F of F is defined as

$$\eta_F(x) \triangleq \int_0^\infty \frac{1}{1+x\lambda} dF(\lambda) = \frac{1}{x} m_F\left(-\frac{1}{x}\right)$$

The η-transform is only a convenient way to use the Stieltjes transform on the negative real-line.

Other transforms: Shannon and η-transforms

A. M. Tulino, S. Verdù, “Random matrix theory and wireless communications,” Now Publishers Inc., 2004.

Definition

Let F be a probability distribution, m_F its Stieltjes transform, then the Shannon-transform $\mathcal{V}_F$ of F is defined as

$$\mathcal{V}_F(x) \triangleq \int_0^\infty \log(1+x\lambda)dF(\lambda) = \int_x^\infty \left(\frac{1}{t} - m_F(-t)\right) dt$$

Note that this last relation is fundamental to wireless communication purposes!

Definition

Let F be a probability distribution, m_F its Stieltjes transform, then the η-transform η_F of F is defined as

$$\eta_F(x) \triangleq \int_0^\infty \frac{1}{1+x\lambda} dF(\lambda) = \frac{1}{x} m_F\left(-\frac{1}{x}\right)$$

The η-transform is only a convenient way to use the Stieltjes transform on the negative real-line.

Outline

The Marčenko-Pastur law

The theorem to be proven is the following

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have i.i.d. zero mean variance $1/n$ entries with finite eighth order moments. As $n, N \to \infty$ with $\frac{N}{n} \to c \in (0,\infty)$, the e.s.d. of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ converges almost surely to a nonrandom distribution function F_c with density f_c given by

$$f_c(x) = (1-c^{-1})^+\delta(x) + \frac{1}{2\pi cx}\sqrt{(x-a)^+(b-x)^+}$$

where $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$ and $\delta(x) = I_{\{0\}}(x)$.

The Marčenko-Pastur density

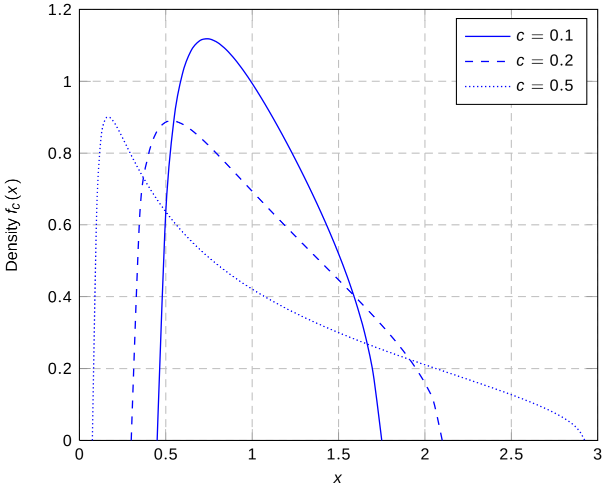

Figure: Marčenko-Pastur law for different limit ratios $c = \lim_{N\to\infty} N/n$.

Diagonal entries of the resolvent

Since we want an expression of m_F, we start by identifying the diagonal entries of the resolvent $(\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}} - z\mathbf{I}_N)^{-1}$ of $\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}$. Denote

$$\mathbf{X}_N = \begin{bmatrix} \mathbf{y}^{\mathrm{H}} \\ \mathbf{Y} \end{bmatrix}$$

Now, for $z \in \mathbb{C}^+$, we have

$$\left(\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}} - z\mathbf{I}_N\right)^{-1} = \begin{bmatrix} \mathbf{y}^{\mathrm{H}}\mathbf{y} - z & \mathbf{y}^{\mathrm{H}}\mathbf{Y}^{\mathrm{H}} \\ \mathbf{Y}\mathbf{y} & \mathbf{Y}\mathbf{Y}^{\mathrm{H}} - z\mathbf{I}_{N-1} \end{bmatrix}^{-1}$$

Consider the first diagonal element of $(\mathbf{R}_N - z\mathbf{I}_N)^{-1}$. From the matrix inversion lemma,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}$$

which here gives

$$\left[\left(\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}} - z\mathbf{I}_N\right)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathrm{H}}(\mathbf{Y}^{\mathrm{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

Diagonal entries of the resolvent

Since we want an expression of m_F, we start by identifying the diagonal entries of the resolvent $(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N)^{-1}$ of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$. Denote

$$\mathbf{X}_N = \begin{bmatrix} \mathbf{y}^{\mathsf{H}} \\ \mathbf{Y} \end{bmatrix}$$

Now, for $z \in \mathbb{C}^+$, we have

$$\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} = \begin{bmatrix} \mathbf{y}^{\mathsf{H}}\mathbf{y} - z & \mathbf{y}^{\mathsf{H}}\mathbf{Y}^{\mathsf{H}} \\ \mathbf{Y}\mathbf{y} & \mathbf{Y}\mathbf{Y}^{\mathsf{H}} - z\mathbf{I}_{N-1} \end{bmatrix}^{-1}$$

Consider the first diagonal element of $(\mathbf{R}_N - z\mathbf{I}_N)^{-1}$. From the matrix inversion lemma,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}$$

which here gives

$$\left[\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

Trace Lemma

D. N. C. Tse, O. Zeitouni, “Linear multiuser receivers in random environments,” IEEE Trans. on Information Theory, vol. 46, no. 1, pp. 171-188, 2000.

To go further, we need the following result,

Theorem

Let $\{\mathbf{A}_N\} \in \mathbb{C}^{N\times N}$*. Let* $\{\mathbf{x}_N\} \in \mathbb{C}^N$*, be a random vector of i.i.d. entries with zero mean, variance* $1/N$ *and finite* 8^{th} *order moment, independent of* $\mathbf{A}_N$*. Then*

$$\sqrt{N}\left[\mathbf{x}_N^{\mathrm{H}}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\operatorname{tr}\mathbf{A}_N\right] \to \mathcal{CN}(0,1)$$

As a corollary, we have

$$\mathbf{x}_N^{\mathrm{H}}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\operatorname{tr}\mathbf{A}_N \to 0$$

almost surely.

For large N, we therefore have approximately

$$\left[\left(\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}} - z\mathbf{I}_N\right)^{-1}\right]_{11} \simeq \frac{1}{-z - z\frac{1}{N}\operatorname{tr}(\mathbf{Y}^{\mathrm{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}}$$

Rank-1 perturbation lemma

J. W. Silverstein, Z. D. Bai, “On the empirical distribution of eigenvalues of a class of large dimensional random matrices,” Journal of Multivariate Analysis, vol. 54, no. 2, pp. 175-192, 1995.

It is somewhat intuitive that adding *a single column* to $\mathbf{Y}$ won't affect the trace in the limit.

Theorem

Let $z \in \mathbb{C}^+$, $\mathbf{A}$ *and* $\mathbf{B}$ $N \times N$ *with* $\mathbf{B}$ *Hermitian, and* $\mathbf{v} \in \mathbb{C}^N$*. Then*

$$\left| \frac{1}{N} \operatorname{tr} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{v}\mathbf{v}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} \right) \mathbf{A} \right| \leq \frac{1}{N} \frac{\|\mathbf{A}\|}{\Im[z]}$$

with $\|\mathbf{A}\|$ *the spectral norm of* $\mathbf{A}$.

Therefore, for large N, we have approximately,

$$\begin{aligned}
\left[\left(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N \right)^{-1} \right]_{11} &\simeq \frac{1}{-z - z\frac{1}{N} \operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}} \\
&\simeq \frac{1}{-z - z\frac{1}{N} \operatorname{tr}(\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N - z\mathbf{I}_n)^{-1}} \\
&= \frac{1}{-z - z\frac{n}{N} m_{\underline{F}}(z)}
\end{aligned}$$

in which we recognize the Stieltjes transform $m_{\underline{F}}$ of the l.s.d. of $\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N$.

End of the proof

We have again the relation

$$\frac{n}{N} m_{\underline{F}}(z) = m_F(z) + \frac{N-n}{N}\frac{1}{z}$$

hence

$$\left[\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}\right]_{11} \simeq \frac{1}{\frac{n}{N} - 1 - z - zm_F(z)}$$

Note that the choice $(1,1)$ is irrelevant here, so the expression is valid for all pair (i,i). Summing over the N terms and averaging, we finally have

$$m_F(z) = \frac{1}{N}\operatorname{tr}\left(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \simeq \frac{1}{c - 1 - z - zm_F(z)}$$

which solve a polynomial of second order. Finally

$$m_F(z) = \frac{c-1}{2z} - \frac{1}{2} + \frac{\sqrt{(c-1-z)^2 - 4z}}{2z}$$

from the inverse Stieltjes transform formula, we then verify that m_F is the Stieltjes transform of the Marčenko-Pastur law.

Outline

Models studied with analytic tools

- *Stieltjes transform:* models involving i.i.d. matrices
  - sample covariance matrix models, $\mathbf{XTX}^{\mathrm{H}}$ and $\mathbf{T}^{\frac{1}{2}}\mathbf{X}^{\mathrm{H}}\mathbf{XT}^{\frac{1}{2}}$
  - doubly correlated models, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}$. With $\mathbf{X}$ Gaussian, Kronecker model.
  - doubly correlated models with external matrix, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} + \mathbf{A}$.
  - variance profile, $\mathbf{XX}^{\mathrm{H}}$, where $\mathbf{X}$ has i.i.d. entries with mean 0, variance $\sigma_{i,j}^2$.
  - Ricean channels, $\mathbf{XX}^{\mathrm{H}} + \mathbf{A}$, where $\mathbf{X}$ has a variance profile.
  - sum of doubly correlated i.i.d. matrices, $\sum_{k=1}^{K}\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathrm{H}}\mathbf{R}_k^{\frac{1}{2}}$.
  - information-plus-noise models $(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathrm{H}}$
  - frequency-selective doubly-correlated channels $(\sum_{k=1}^{K}\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})(\sum_{k=1}^{K}\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})$
  - sum of frequency-selective doubly-correlated channels $\sum_{k=1}^{K}\mathbf{R}_k^{\frac{1}{2}}\mathbf{H}_k\mathbf{T}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{R}_k^{\frac{1}{2}}$, where $\mathbf{H}_k = \sum_{l=1}^{L}\mathbf{R}'^{\frac{1}{2}}_{kl}\mathbf{X}_{kl}\mathbf{T}'_{kl}\mathbf{X}_{kl}^{\mathrm{H}}\mathbf{R}'^{\frac{1}{2}}_{kl}$.
- *R- and S-transforms:* models involving a column subset $\mathbf{W}$ of unitary matrices
  - doubly correlated Haar matrix $\mathbf{R}^{\frac{1}{2}}\mathbf{WTW}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}$
  - sum of simply correlated Haar matrices $\sum_{k=1}^{K}\mathbf{W}_k\mathbf{T}_k\mathbf{W}_k^{\mathrm{H}}$

In most cases, $\mathbf{T}$ and $\mathbf{R}$ can be taken random, but independent of $\mathbf{X}$. More involved random matrices, such as Vandermonde matrices, were not yet studied.

Models studied with analytic tools

- *Stieltjes transform:* models involving i.i.d. matrices
  - sample covariance matrix models, $\mathbf{XTX}^{\mathrm{H}}$ and $\mathbf{T}^{\frac{1}{2}}\mathbf{X}^{\mathrm{H}}\mathbf{XT}^{\frac{1}{2}}$
  - doubly correlated models, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}$. With $\mathbf{X}$ Gaussian, Kronecker model.
  - doubly correlated models with external matrix, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} + \mathbf{A}$.
  - variance profile, $\mathbf{XX}^{\mathrm{H}}$, where $\mathbf{X}$ has i.i.d. entries with mean 0, variance $\sigma_{i,j}^2$.
  - Ricean channels, $\mathbf{XX}^{\mathrm{H}} + \mathbf{A}$, where $\mathbf{X}$ has a variance profile.
  - sum of doubly correlated i.i.d. matrices, $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathrm{H}}\mathbf{R}_k^{\frac{1}{2}}$.
  - information-plus-noise models $(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathrm{H}}$
  - frequency-selective doubly-correlated channels $(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})$
  - sum of frequency-selective doubly-correlated channels $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{H}_k\mathbf{T}_k\mathbf{H}_k^{\mathrm{H}}\mathbf{R}_k^{\frac{1}{2}}$, where $\mathbf{H}_k = \sum_{l=1}^{L} \mathbf{R'}_{kl}^{\frac{1}{2}}\mathbf{X}_{kl}\mathbf{T'}_{kl}\mathbf{X}_{kl}^{\mathrm{H}}\mathbf{R'}_{kl}^{\frac{1}{2}}$.
- *R- and S-transforms:* models involving a column subset $\mathbf{W}$ of unitary matrices
  - doubly correlated Haar matrix $\mathbf{R}^{\frac{1}{2}}\mathbf{WTW}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}$
  - sum of simply correlated Haar matrices $\sum_{k=1}^{K} \mathbf{W}_k\mathbf{T}_k\mathbf{W}_k^{\mathrm{H}}$

In most cases, $\mathbf{T}$ and $\mathbf{R}$ can be taken random, but independent of $\mathbf{X}$. More involved random matrices, such as Vandermonde matrices, were not yet studied.

Models studied with analytic tools

- *Stieltjes transform:* models involving i.i.d. matrices
  - sample covariance matrix models, $\mathbf{XTX}^{\mathsf{H}}$ and $\mathbf{T}^{\frac{1}{2}}\mathbf{X}^{\mathsf{H}}\mathbf{XT}^{\frac{1}{2}}$
  - doubly correlated models, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$. With $\mathbf{X}$ Gaussian, Kronecker model.
  - doubly correlated models with external matrix, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} + \mathbf{A}$.
  - variance profile, $\mathbf{XX}^{\mathsf{H}}$, where $\mathbf{X}$ has i.i.d. entries with mean 0, variance $\sigma^2_{i,j}$.
  - Ricean channels, $\mathbf{XX}^{\mathsf{H}} + \mathbf{A}$, where $\mathbf{X}$ has a variance profile.
  - sum of doubly correlated i.i.d. matrices, $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$.
  - information-plus-noise models $(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}}$
  - frequency-selective doubly-correlated channels $(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})$
  - sum of frequency-selective doubly-correlated channels $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{H}_k\mathbf{T}_k\mathbf{H}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$, where $\mathbf{H}_k = \sum_{l=1}^{L} \mathbf{R}'^{\frac{1}{2}}_{kl}\mathbf{X}_{kl}\mathbf{T}'_{kl}\mathbf{X}_{kl}^{\mathsf{H}}\mathbf{R}'^{\frac{1}{2}}_{kl}$.
- *R- and S-transforms:* models involving a column subset $\mathbf{W}$ of unitary matrices
  - doubly correlated Haar matrix $\mathbf{R}^{\frac{1}{2}}\mathbf{WTW}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$
  - sum of simply correlated Haar matrices $\sum_{k=1}^{K} \mathbf{W}_k\mathbf{T}_k\mathbf{W}_k^{\mathsf{H}}$

In most cases, $\mathbf{T}$ and $\mathbf{R}$ can be taken random, but independent of $\mathbf{X}$. More involved random matrices, such as Vandermonde matrices, were not yet studied.

Models studied with moments/free probability

- asymptotic results
  - most of the above models with Gaussian $\mathbf{X}$.
  - products $\mathbf{V}_1\mathbf{V}_1^H\mathbf{T}_1\mathbf{V}_2\mathbf{V}_2^H\mathbf{T}_2\ldots$ of Vandermonde and deterministic matrices
  - *conjecture:* any probability space of matrices invariant to row or column permutations.
- marginal studies, not yet fully explored
  - rectangular free convolution: singular values of rectangular matrices
  - finite size models. Instead of almost sure convergence of $m_{\mathbf{X}_N}$ as $N \to \infty$, we can study finite size behaviour of $\mathrm{E}[m_{\mathbf{X}_N}]$.

Models studied with moments/free probability

- asymptotic results
  - most of the above models with Gaussian $\mathbf{X}$.
  - products $\mathbf{V}_1\mathbf{V}_1^H\mathbf{T}_1\mathbf{V}_2\mathbf{V}_2^H\mathbf{T}_2\ldots$ of Vandermonde and deterministic matrices
  - *conjecture:* any probability space of matrices invariant to row or column permutations.
- marginal studies, not yet fully explored
  - rectangular free convolution: singular values of rectangular matrices
  - finite size models. Instead of almost sure convergence of $m_{\mathbf{X}_N}$ as $N \to \infty$, we can study finite size behaviour of $\mathrm{E}[m_{\mathbf{X}_N}]$.

Open problems, to be explored

- Stieltjes transform methods for more structured matrices: e.g. Vandermonde matrices
- clean framework for band matrix models
- finite dimensional methods for Ricean matrices
- other ?

Related bibliography

- R. B. Dozier, J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices," Journal of Multivariate Analysis, vol. 98, no. 4, pp. 678-694, 2007.
- J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 175-192, 1995.
- J. W. Silverstein, S. Choi "Analysis of the limiting spectral distribution of large dimensional random matrices" Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.
- F. Benaych-Georges, "Rectangular random matrices, related free entropy and free Fisher's information," Arxiv preprint math/0512081, 2005.
- Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.
- V. L. Girko, "Theory of Random Determinants," Kluwer, Dordrecht, 1990.
- R. Couillet, M. Debbah, J. W. Silverstein, "A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels," *submitted to IEEE Trans. on Information Theory*.
- V. L. Girko, "Theory of Random Determinants," Kluwer, Dordrecht, 1990.
- W. Hachem, Ph. Loubaton, J. Najim, "Deterministic Equivalents for Certain Functionals of Large Random Matrices", Annals of Applied Probability, vol. 17, no. 3, 2007.
- M. J. M. Peacock, I. B. Collings, M. L. Honig, "Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions," IEEE Trans. on Information Theory, vol. 54, no. 5, pp. 2123, 2008.
- D. Petz, J. Réffy, "On Asymptotics of large Haar distributed unitary matrices," Periodica Math. Hungar., vol. 49, pp. 103-117, 2004.
- Ø. Ryan, A. Masucci, S. Yang, M. Debbah, "Finite dimensional statistical inference," *submitted to IEEE Trans. on Information Theory, Dec. 2009.*

Technical Bibliography

- W. Rudin, “Real and complex analysis,” New York, 1966.
- P. Billingsley, “Probability and measure,” Wiley New York, 2008.
- P. Billingsley, “Convergence of probability measures,” Wiley New York, 1968.