Crash Course on Random Matrix Theory

Part I: Basic notions and applications to wireless communications

Morning Session: Basic notions of RMT

Romain Couillet and Mérouane Debbah

SUPELEC

26-01-2012

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 $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.

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 $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.
From the law of large numbers, as $n \to \infty$,

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

with $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{N \times n}$.

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 $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.

From the law of large numbers, as $n \to \infty$,

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

with $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{N \times n}$.

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),

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

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 $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.
From the law of large numbers, as $n \to \infty$,

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

with $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{N \times n}$.
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?

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 $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.
From the law of large numbers, as $n \to \infty$,

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

with $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{N \times n}$.
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.

Empirical and limit spectra of Wishart matrices

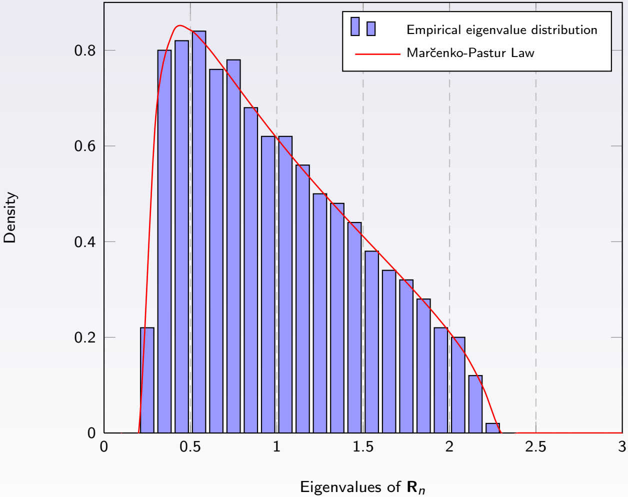

Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

Definitions

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.

Definitions

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 $\mathbf{X}_N$.

Outline

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

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

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

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.

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

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

Marčenko-Pastur law, Semi-circle law, Full circle law...

V. A. Marčenko, L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices", Math USSR-Sbornik, vol. 1, no. 4, pp. 457-483, 1967.

- If $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ has i.i.d. entries of mean 0, variance $1/n$, then (almost surely) $F^{\mathbf{X}_N \mathbf{X}_N^{\mathrm{H}}} \Rightarrow F_c$ as $N, n \to \infty$, $N/n \to c$, with F_c the Marčenko-Pastur law with density

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

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 (almost surely) $F^{\mathbf{X}_N} \Rightarrow F$ 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.

Marčenko-Pastur law

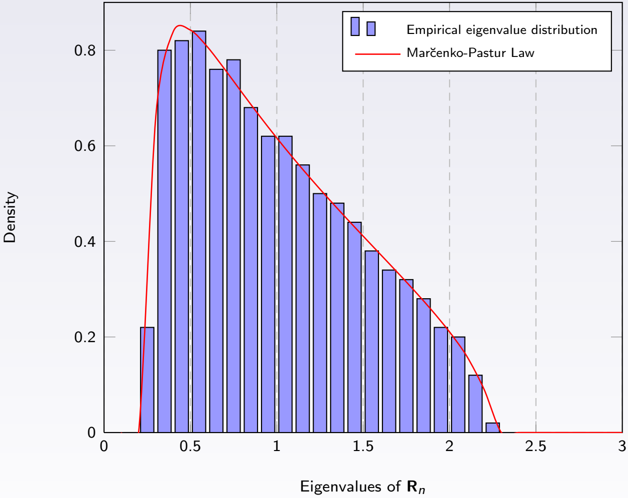

Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

Semi-circle law

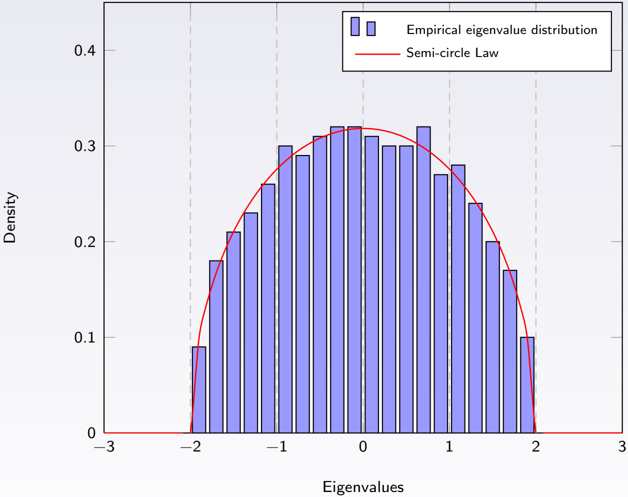

Figure: Histogram of the eigenvalues of Wigner matrices and the semi-circle law, for $N = 500$

Wigner's proof

- Proof based on the limiting moment of the eigenvalue distribution.

Wigner's proof

- Proof based on the limiting moment of the eigenvalue distribution.
- For $\mathbf{X} \in \mathbb{C}^{N \times N}$ Hermitian with $X_{ij} \sim \mathcal{CN}(0, 1/N)$, the limiting density f of the eigenvalues

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr}\left(\mathbf{X}^{2k+1}\right) = 0$$

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr}\left(\mathbf{X}^{2k}\right) = \frac{1}{k+1} C_k^{2k}$$

known as the Catalan numbers.

Wigner's proof

- Proof based on the limiting moment of the eigenvalue distribution.
- For $\mathbf{X} \in \mathbb{C}^{N \times N}$ Hermitian with $X_{ij} \sim \mathcal{CN}(0, 1/N)$, the limiting density f of the eigenvalues

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr}\left(\mathbf{X}^{2k+1}\right) = 0$$

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr}\left(\mathbf{X}^{2k}\right) = \frac{1}{k+1} C_k^{2k}$$

known as the Catalan numbers.

- These are exactly the moments of a semi-cicle distribution!

$$\alpha_{2k} = \frac{1}{\pi} \int_{-2}^{2} x^{2k} \sqrt{4 - x^2} dx = -\frac{1}{2\pi} \int_{-2}^{2} \frac{-x}{\sqrt{4-x^2}} x^{2k-1} (4 - x^2) dx$$

$$= \frac{1}{2\pi} \int_{-2}^{2} \sqrt{4 - x^2} (x^{2k-1} (4 - x^2))' dx = 4(2k-1)\alpha_{2k-2} - (2k+1)\alpha_{2k}.$$

which gives the recursive relation

$$\alpha_{2k} = \frac{2(2k-1)}{k+1} \alpha_{2k-2}, \quad \text{defining the Catalan numbers.}$$

Wigner's proof

- Proof based on the limiting moment of the eigenvalue distribution.
- For $\mathbf{X} \in \mathbb{C}^{N\times N}$ Hermitian with $X_{ij} \sim \mathcal{CN}(0,1/N)$, the limiting density f of the eigenvalues

$$\lim_{N\to\infty} \frac{1}{N}\operatorname{tr}\left(\mathbf{X}^{2k+1}\right) = 0$$
$$\lim_{N\to\infty} \frac{1}{N}\operatorname{tr}\left(\mathbf{X}^{2k}\right) = \frac{1}{k+1} C_k^{2k}$$

known as the Catalan numbers.

- These are exactly the moments of a semi-cicle distribution!

$$\alpha_{2k} = \frac{1}{\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx = -\frac{1}{2\pi}\int_{-2}^{2} \frac{-x}{\sqrt{4-x^2}} x^{2k-1}(4-x^2)dx$$
$$= \frac{1}{2\pi}\int_{-2}^{2} \sqrt{4-x^2}(x^{2k-1}(4-x^2))'dx = 4(2k-1)\alpha_{2k-2} - (2k+1)\alpha_{2k}.$$

which gives the recursive relation

$$\alpha_{2k} = \frac{2(2k-1)}{k+1}\alpha_{2k-2}, \quad \text{defining the Catalan numbers.}$$

Proof impractical for more involved models

Difficult in general to move from moments to distributions / to compute the moments directly.

Circular law

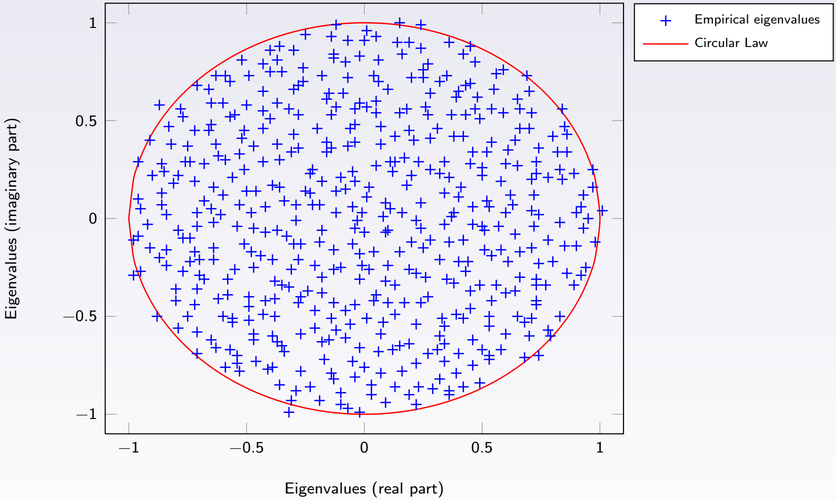

Figure: Eigenvalues of $\mathbf{X}_N$ with i.i.d. standard Gaussian entries, for $N = 500$.

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.

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.

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.

Related bibliography

- E. Wigner, “Characteristic vectors of bordered matrices with infinite dimensions,” The Annals of Mathematics, vol. 62, no. 3, pp. 548-564, 1955.
- V. A. Marčenko and L. A. Pastur, “Distributions of eigenvalues for some sets of random matrices,” Math USSR-Sbornik, vol. 1, no. 4, pp. 457-483, 1967.
- J. Wishart, “The generalized product moment distribution in samples from a normal multivariate population”, Biometrika, vol. 20A, pp. 32-52, 1928.
- T. W. Anderson, “The non-central Wishart distribution and certain problems of multivariate statistics,” The Annals of Mathematical Statistics, vol. 17, no. 4, pp. 409-431, 1946.
- H. Chandra, “Differential operators on a semi-simple Lie algebra,” American Journal of Mathematics, vol. 79, pp. 87-120, 1957.
- T. Ratnarajah and R. Vaillancourt, “Complex singular Wishart matrices and applications,” Computers and Mathematics with Applications, vol. 50, no. 3-4, pp. 399-411, 2005.

Tools for RMT

Various approaches used to deal with random matrices.

- **Asymptotic spectrum:**

Tools for RMT

Various approaches used to deal with random matrices.

- **Asymptotic spectrum:**
  - *Method of moments*: identify eigenvalue distribution through its moments [e.g. Wigner]

Tools for RMT

Various approaches used to deal with random matrices.

- **Asymptotic spectrum:**
  - *Method of moments*: identify eigenvalue distribution through its moments [e.g. Wigner]
  - *Free probability theory*: study spectrum of random matrix operations through moments of limiting distributions [e.g. Petz, Biane, Benaych-Georges]

Tools for RMT

Various approaches used to deal with random matrices.

- **Asymptotic spectrum:**
  - *Method of moments*: identify eigenvalue distribution through its moments [e.g. Wigner]
  - *Free probability theory*: study spectrum of random matrix operations through moments of limiting distributions [e.g. Petz, Biane, Benaych-Georges]
  - *Stieltjes transform method*: study spectrum of random matrix operations through Stieltjes transform [e.g. Bai, Silverstein, Pastur]

Tools for RMT

Various approaches used to deal with random matrices.

- **Asymptotic spectrum:**
  - *Method of moments*: identify eigenvalue distribution through its moments [e.g. Wigner]
  - *Free probability theory*: study spectrum of random matrix operations through moments of limiting distributions [e.g. Petz, Biane, Benaych-Georges]
  - *Stieltjes transform method*: study spectrum of random matrix operations through Stieltjes transform [e.g. Bai, Silverstein, Pastur]
  - *Gaussian tools on resolvents*: study spectrum of Gaussian random matrix operations through Gaussian tricks on the resolvent [e.g. Pastur, Loubaton, Hachem]

Tools for RMT

Various approaches used to deal with random matrices.

- **Asymptotic spectrum:**
  - *Method of moments*: identify eigenvalue distribution through its moments [e.g. Wigner]
  - *Free probability theory*: study spectrum of random matrix operations through moments of limiting distributions [e.g. Petz, Biane, Benaych-Georges]
  - *Stieltjes transform method*: study spectrum of random matrix operations through Stieltjes transform [e.g. Bai, Silverstein, Pastur]
  - *Gaussian tools on resolvents*: study spectrum of Gaussian random matrix operations through Gaussian tricks on the resolvent [e.g. Pastur, Loubaton, Hachem]
  - *Replica method*: Non-rigorous physical tools to study deterministic equivalents [e.g. Tanaka, Moustakas, Riegler]

Tools for RMT

Various approaches used to deal with random matrices.

- **Asymptotic spectrum:**
  - *Method of moments*: identify eigenvalue distribution through its moments [e.g. Wigner]
  - *Free probability theory*: study spectrum of random matrix operations through moments of limiting distributions [e.g. Petz, Biane, Benaych-Georges]
  - *Stieltjes transform method*: study spectrum of random matrix operations through Stieltjes transform [e.g. Bai, Silverstein, Pastur]
  - *Gaussian tools on resolvents*: study spectrum of Gaussian random matrix operations through Gaussian tricks on the resolvent [e.g. Pastur, Loubaton, Hachem]
  - *Replica method*: Non-rigorous physical tools to study deterministic equivalents [e.g. Tanaka, Moustakas, Riegler]
  - *Orthogonal polynomials and Fredholm determinants*: study hole probability, e.g. extreme eigenvalue distribution through determinantal equations [e.g. Johnstone, Tracy, Widom, Guionnet]

Outline of the course

- **DAY 1 – Morning:**
  - Overview *free probability theory* and moment operations on spectra of random matrices.
  - Introduce the Stieltjes transform method and its link to wireless communication and signal processing quantities.
  - Introduce deterministic equivalents as a generalization of limiting spectral distribution analysis.

Outline of the course

- **DAY 1 – Morning:**
  - Overview *free probability theory* and moment operations on spectra of random matrices.
  - Introduce the Stieltjes transform method and its link to wireless communication and signal processing quantities.
  - Introduce deterministic equivalents as a generalization of limiting spectral distribution analysis.
- **DAY 1 – Afternoon:**
  - Apply the Stieltjes transform method and deterministic equivalents to wireless communications.
  - Study of communication performance of CDMA, MIMO, relay channels, and optimization.
  - Study of inference methods based on deterministic equivalents for cognitive radios.

Outline of the course

- **DAY 1 – Morning:**
  - Overview *free probability theory* and moment operations on spectra of random matrices.
  - Introduce the Stieltjes transform method and its link to wireless communication and signal processing quantities.
  - Introduce deterministic equivalents as a generalization of limiting spectral distribution analysis.
- **DAY 1 – Afternoon:**
  - Apply the Stieltjes transform method and deterministic equivalents to wireless communications.
  - Study of communication performance of CDMA, MIMO, relay channels, and optimization.
  - Study of inference methods based on deterministic equivalents for cognitive radios.
- **DAY 2 – Morning:**
  - Extend spectrum analysis and study individual eigenvalue behaviour.
  - Introduce eigen-inference methods to extend classical signal processing methods to random matrices.
  - Discuss the specific spiked model in which a few eigenvalues escape the spectrum support.

Outline of the course

- **DAY 1 – Morning:**
  - Overview *free probability theory* and moment operations on spectra of random matrices.
  - Introduce the Stieltjes transform method and its link to wireless communication and signal processing quantities.
  - Introduce deterministic equivalents as a generalization of limiting spectral distribution analysis.
- **DAY 1 – Afternoon:**
  - Apply the Stieltjes transform method and deterministic equivalents to wireless communications.
  - Study of communication performance of CDMA, MIMO, relay channels, and optimization.
  - Study of inference methods based on deterministic equivalents for cognitive radios.
- **DAY 2 – Morning:**
  - Extend spectrum analysis and study individual eigenvalue behaviour.
  - Introduce eigen-inference methods to extend classical signal processing methods to random matrices.
  - Discuss the specific spiked model in which a few eigenvalues escape the spectrum support.
- **DAY 2 – Afternoon:**
  - Apply eigenspectrum analysis and eigen-inference methods to signal processing.
  - Study of hypothesis tests in small/large dimensional random matrix scenarios.
  - Study of eigen-inference methods to extend subspace methods: distance/DoA estimation.

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 $\hat{M}_k$, $k = 1, 2, \ldots$,

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

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

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

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,

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

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* $M_k(\mathbf{A} + \mathbf{B}) = E_\omega[\hat{M}_k(\mathbf{A}(\omega) + \mathbf{B}(\omega))]$

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* $M_k(\mathbf{A} + \mathbf{B}) = E_\omega[\hat{M}_k(\mathbf{A}(\omega) + \mathbf{B}(\omega))]$

Free probability

An extension to non-commutative random variables exists: Free probability.

Free probability

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

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.

Free probability

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

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,

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

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

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

Non-crossing partitions

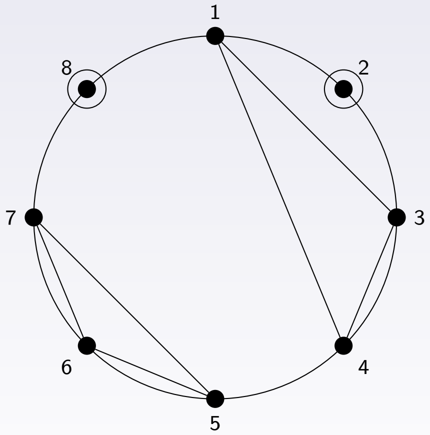

Figure: Non-crossing partition $\pi = \{\{1,3,4\},\{2\},\{5,6,7\},\{8\}\}$ of $NC(8)$.

Moments of sums and products of random matrices

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

- 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.

Moments of sums and products of random matrices

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

- 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 **X**, also called *Stieltjes transform*,

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

- 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.

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

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

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	$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{\|V\|}$	$M_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{V \in \pi} C_{\|V\|}$
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	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \rightarrow \mathcal{N}(0,1)$	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \Rightarrow$ semi-circle law

Bibliography on Free Probability related work

- F. Hiai and D. Petz, "The Semicircle Law, Free Random Variables and Entropy", Mathematical Surveys and Monographs No. 77, American Mathematical Society, 2006.
- P. Biane, "Free probability for probabilists," Quantum Probability Communications, vol. 11, pp. 55-71, 2003.
- 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.
- 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.
- F. Benaych-Georges, "Rectangular random matrices, related free entropy and free Fisher's information," Journal of Operator Theory, vol. 62, no. 2, pp. 371-419, 2009.

Outline

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

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

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

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

If F has a density f, then

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

Remark on the Stieltjes transform

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

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i - z} = \frac{1}{N}\mathrm{tr}\left(\mathrm{diag}(\{\lambda_i\}_{i=1}^{N}) - z\mathbf{I}_N\right)^{-1} = \frac{1}{N}\mathrm{tr}\left(\mathbf{X}_N - z\mathbf{I}_N\right)^{-1}$$

- For compactly supported F, $m_F(z)$ is linked to the moments $M_k = E[\frac{1}{N}\mathrm{tr}(\mathbf{X}^k)]$

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

Remark on the Stieltjes transform

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

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i - z} = \frac{1}{N}\mathrm{tr}\left(\mathrm{diag}(\{\lambda_i\}_{i=1}^{N}) - z\mathbf{I}_N\right)^{-1} = \frac{1}{N}\mathrm{tr}\left(\mathbf{X}_N - z\mathbf{I}_N\right)^{-1}$$

- For compactly supported F, $m_F(z)$ is linked to the moments $M_k = E[\frac{1}{N}\mathrm{tr}(\mathbf{X}^k)]$

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

The Stieltjes transform is 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 $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{X}$ only differ by a mass of $|N-n|$ zeros. Say $N \geqslant n$,

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

hence

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

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 $\underline{\mathbf{B}}_N = \mathbf{X}_N \mathbf{T}_N \mathbf{X}_N^{\mathsf{H}} \in \mathbb{C}^{N\times N}$, 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 later in this course.

- in general, no explicit expression for $\underline{F}$.
- the theorem above characterizes also the Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ 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 of $\mathbf{y}$, when $\mathbf{y}_i = \mathbf{T}_N^{\frac{1}{2}} \mathbf{x}_i$, $\mathbf{x}_i$ i.i.d., $\mathbf{T}_N = E[\mathbf{y}\mathbf{y}^{\mathsf{H}}]$.

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 $\mathbb{C}^*$)

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 $\Im[m_F(z)]$.

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 second part of the course.

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 second part of the course.

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 $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ 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 $K = 4$.

Spectrum of the sample covariance matrix

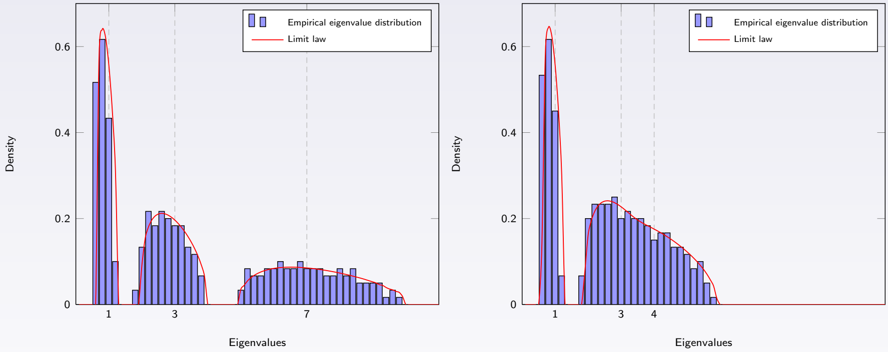

Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$, $N = 3000$, $n = 300$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

The "Shannon"-transform

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

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

This quantity is fundamental to wireless communication purposes!

Outline

The Marčenko-Pastur law

V. A. Marčenko, L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices", Math USSR-Sbornik, vol. 1, no. 4, pp. 457-483, 1967.

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 $\mathbf{X}_N \mathbf{X}_N^{\mathrm{H}}$ converges almost surely to a nonrandom distribution function F_c with density f_c given by

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

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

The Marčenko-Pastur density

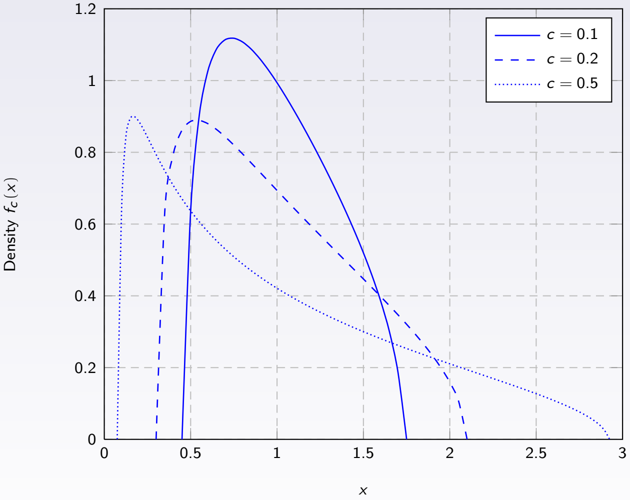

Figure: Marčenko-Pastur law for different limit ratios $c = \lim_{N\to\infty} N/n$.

Diagonal entries of the resolvent

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

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

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

Z. Bai, J. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices", Springer Series in Statistics, 2009.

To go further, we need the following result,

Theorem

Let $\{\mathbf{A}_N\} \in \mathbb{C}^{N\times N}$ *with bounded spectral norm. 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*

$$\mathbf{x}_N^{\mathsf{H}}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}tr\mathbf{A}_N \xrightarrow{\text{a.s.}} 0.$$

For large N, we therefore have approximately

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

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 $\mathbf{A}$ *and* $\mathbf{B}$ *be* $N \times N$ *with* $\mathbf{B}$ *Hermitian positive definite, and* $\mathbf{v} \in \mathbb{C}^N$*. For* $z \in \mathbb{C} \setminus \mathbb{R}^-$,

$$\left| \frac{1}{N} tr \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{v}\mathbf{v}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} \right) \mathbf{A} \right| \leqslant \frac{1}{N} \frac{\|\mathbf{A}\|}{\operatorname{dist}(z, \mathbb{R}^+)}$$

with $\|\mathbf{A}\|$ *the spectral norm of* $\mathbf{A}$*, and* $\operatorname{dist}(z, A) = \inf_{y \in A} \|y - z\|$.

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

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

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

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

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.

Related bibliography

- V. A. Marčenko, L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices", Math USSR-Sbornik, vol. 1, no. 4, pp. 457-483, 1967.
- 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.
- Z. D. Bai and J. W. Silverstein, "Spectral analysis of large dimensional random matrices, 2nd Edition" Springer Series in Statistics, 2009.
- 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.
- V. L. Girko, "Theory of Random Determinants," Kluwer, Dordrecht, 1990.
- A. M. Tulino, S. Verdù, "Random matrix theory and wireless communications," Now Publishers Inc., 2004.

Outline

Outline

Reminder and scope

- In the first part of this course,
  - we defined the Stieltjes transform:

Definition

Let F be a distribution function, and $z \in \mathbb{C}^+$. Then the Stieltjes transform $m_F(z)$ of F is defined as

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

For F the spectral distribution of an Hermitian matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$,

$$m_F(z) = \frac{1}{N} \operatorname{tr} (\mathbf{X} - z\mathbf{I}_N)^{-1}$$

- We gave limiting distribution results for some matrix models.
- We gave a sketch of the proof of the Marčenko-Pastur law.

Reminder and scope

- In the first part of this course,
  - we defined the Stieltjes transform:

Definition

Let F be a distribution function, and $z \in \mathbb{C}^+$. Then the Stieltjes transform $m_F(z)$ of F is defined as

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

For F the spectral distribution of an Hermitian matrix $\mathbf{X} \in \mathbb{C}^{N \times N}$,

$$m_F(z) = \frac{1}{N} \operatorname{tr} (\mathbf{X} - z\mathbf{I}_N)^{-1}$$

  - We gave limiting distribution results for some matrix models.
  - We gave a sketch of the proof of the Marčenko-Pastur law.
- In this second part, we will
  - extend the notion of limit distributions to deterministic equivalents
  - provide sound mathematical techniques to prove convergence/existence/uniqueness of large N results.
  - provide first wireless communication results

Limiting results against deterministic equivalents

- previously, we showed results of the type:

"let $\mathbf{X}_N$ be random, $\mathbf{T}_N$ deterministic with $F^{\mathbf{T}_N} \Rightarrow F^T$, etc. Then, when $N \to \infty$, the e.s.d. of $\mathbf{X}_N$ tends to F such that m_F is solution of a fixed-point equation,

$$m_{\mathbf{X}_N}(z) \to m_F(z)$$"

Limiting results against deterministic equivalents

- previously, we showed results of the type:

  "let $\mathbf{X}_N$ be random, $\mathbf{T}_N$ deterministic with $F^{\mathbf{T}_N} \Rightarrow F^T$, etc. Then, when $N \to \infty$, the e.s.d. of $\mathbf{X}_N$ tends to F such that m_F is solution of a fixed-point equation,

  $$m_{\mathbf{X}_N}(z) \to m_F(z) \text{ ''}$$

- this has major drawbacks
  - this assumes $\mathbf{T}_N$ *has* a limiting distribution
  - if it does, $m_{\mathbf{X}_N \mathbf{X}_N^{\mathrm{H}}}$ can at best be approximated by m_F which is a function of the limiting F^T. For finite N, $F^{\mathbf{T}_N}$ may be very different from F^T.
  - any sequence $\mathbf{T}_N$ with l.s.d. F^T engenders the same l.s.d. F.

Limiting results against deterministic equivalents

- previously, we showed results of the type:

  "let $\mathbf{X}_N$ be random, $\mathbf{T}_N$ deterministic with $F^{\mathbf{T}_N} \Rightarrow F^T$, etc. Then, when $N \to \infty$, the e.s.d. of $\mathbf{X}_N$ tends to F such that m_F is solution of a fixed-point equation,

$$m_{\mathbf{X}_N}(z) \to m_F(z) \text{ "}$$

- this has major drawbacks
  - this assumes $\mathbf{T}_N$ *has* a limiting distribution
  - if it does, $m_{\mathbf{X}_N \mathbf{X}_N^H}$ can at best be approximated by m_F which is a function of the limiting F^T. For finite N, $F^{\mathbf{T}_N}$ may be very different from F^T.
  - any sequence $\mathbf{T}_N$ with l.s.d. F^T engenders the same l.s.d. F.
- instead, we shall use results of the type

  "let $\mathbf{X}_N$ be random, $\mathbf{T}_N$ deterministic ~~with $F^{\mathbf{T}_N} \Rightarrow F^T$~~, etc. Then the e.s.d. of $\mathbf{X}_N$ ~~tends to F such that m_F is solution of a fixed-point equation~~ has Stieltjes transform $m_{\mathbf{X}_N}$ well approximated by the deterministic m_N, which is the unique solution of a fixed-point equation and such that

$$m_{\mathbf{X}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0 \text{ "}$$

In this case, m_N is a function of $\mathbf{T}_N$, for fixed N and does not require any convergence of $F^{\mathbf{T}_N}$.

Outline of the proofs

It will often be the case that the deterministic equivalent $m_N(z)$ satisfies an implicit equation. The steps are then:

1. find a suitable function f, such that the *true* Stieltjes transform $m_{\mathbf{X}_N}(z)$ satisfies, for fixed $z \in \mathbb{C}^+$,

$$m_{\mathbf{X}_N}(z) = f(m_{\mathbf{X}_N}(z)) + \varepsilon_N$$

   where $\varepsilon_N \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$.

Outline of the proofs

It will often be the case that the deterministic equivalent $m_N(z)$ satisfies an implicit equation. The steps are then:

1. find a suitable function f, such that the *true* Stieltjes transform $m_{\mathbf{X}_N}(z)$ satisfies, for fixed $z \in \mathbb{C}^+$,

$$m_{\mathbf{X}_N}(z) = f(m_{\mathbf{X}_N}(z)) + \varepsilon_N$$

   where $\varepsilon_N \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$.
   This can be done

   - using Pastur's method (see proof of Marčenko-Pastur law)
   - using guess-work (see Bai and Silverstein's proofs)

   Remark: This is as far as we went in the first part.

Outline of the proofs

It will often be the case that the deterministic equivalent $m_N(z)$ satisfies an implicit equation. The steps are then:

1. find a suitable function f, such that the *true* Stieltjes transform $m_{\mathbf{X}_N}(z)$ satisfies, for fixed $z \in \mathbb{C}^+$,
$$m_{\mathbf{X}_N}(z) = f(m_{\mathbf{X}_N}(z)) + \varepsilon_N$$
where $\varepsilon_N \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$.
This can be done
    - using Pastur's method (see proof of Marčenko-Pastur law)
    - using guess-work (see Bai and Silverstein's proofs)

   Remark: This is as far as we went in the first part.
2. For fixed N, prove the existence of a solution to
$$m = f(m)$$
This is often based on extracting a converging subsequence of $m_N, m_{2N}, \ldots$ such that m_{jN} "has the same properties as $m_{\mathbf{X}_N}(z)$ for all j".
3. For this fixed N, prove the uniqueness of the solution. This involves finding a contradiction if two solutions exist.

Outline of the proofs

It will often be the case that the deterministic equivalent $m_N(z)$ satisfies an implicit equation. The steps are then:

1. find a suitable function f, such that the *true* Stieltjes transform $m_{\mathbf{X}_N}(z)$ satisfies, for fixed $z \in \mathbb{C}^+$,
$$m_{\mathbf{X}_N}(z) = f(m_{\mathbf{X}_N}(z)) + \varepsilon_N$$
where $\varepsilon_N \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$.
This can be done
   - using Pastur's method (see proof of Marčenko-Pastur law)
   - using guess-work (see Bai and Silverstein's proofs)

   Remark: This is as far as we went in the first part.
2. For fixed N, prove the existence of a solution to
$$m = f(m)$$
This is often based on extracting a converging subsequence of $m_N, m_{2N}, \ldots$ such that m_{jN} "has the same properties as $m_{\mathbf{X}_N}(z)$ for all j".
3. For this fixed N, prove the uniqueness of the solution. This involves finding a contradiction if two solutions exist.
4. Calling $m_N(z)$ the solution, prove finally that
$$m_{\mathbf{X}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0$$

Outline

Stieltjes transform of a sum of doubly-correlated matrices

R. Couillet, M. Debbah, J. W. Silverstein, "A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels," vol. 57, no. 6, pp. 3493-3514, 2011.

Theorem

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{H}_k^{\mathsf{H}}, \text{ with } \mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}$$

$\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ with i.i.d. entries of zero mean, variance $1/n_k$, $\mathbf{R}_k$ Hermitian nonnegative definite, $\mathbf{T}_k$ diagonal nonnegative, $\limsup_N \max(\|\mathbf{R}_k\|, \|\mathbf{T}_k\|) < \infty$. Denote $c_k = N/n_k$. Then, as all N and n_k grow large, with bounded ratio c_k, for $z \in \mathbb{C} \setminus \mathbb{R}^-$,

$$m_{\mathbf{B}_N}(z) - m_N(z) \xrightarrow{\text{a.s.}} 0, \quad m_N(z) = \frac{1}{N} tr \left(-z\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_k(z) \mathbf{R}_k \right)^{-1}$$

with $e_1(z), \ldots, e_K(z)$ the unique solutions in $\{z \in \mathbb{C}^+, e_i(z) \in \mathbb{C}^+\}$ or $\{z \in \mathbb{R}^-, e_i(z) \in \mathbb{R}^+\}$ of

$$e_i(z) = \frac{1}{N} tr \mathbf{R}_i \left(-z\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_k(z) \mathbf{R}_k \right)^{-1}$$

$$\bar{e}_i(z) = \frac{1}{n_i} tr \mathbf{T}_i \left(\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{T}_i \right)^{-1}$$

Pastur's method

Pastur's method is *not* applicable here, unless all $\mathbf{R}_k$'s are diagonal.

Pastur's method

Pastur's method is *not* applicable here, unless all $\mathbf{R}_k$'s are diagonal.

Consider $K = 2$ and denote $\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}}$, with diagonal $\mathbf{R}_k$. By block-matrix inversion, we have

$$\left(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}_{11} = \left(\begin{bmatrix} \mathbf{h}_1^{\mathsf{H}} & \mathbf{h}_2^{\mathsf{H}} \\ \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 & \mathbf{U}_1^{\mathsf{H}} \\ \mathbf{h}_2 & \mathbf{U}_2^{\mathsf{H}} \end{bmatrix} - z\mathbf{I}_N\right)^{-1}_{11}$$

$$= \left[-z - z[\mathbf{h}_1^{\mathsf{H}}\mathbf{h}_2^{\mathsf{H}}]\left(\begin{bmatrix}\mathbf{U}_1^{\mathsf{H}} \\ \mathbf{U}_2^{\mathsf{H}}\end{bmatrix}[\mathbf{U}_1\mathbf{U}_2] - z\mathbf{I}_{n_1+n_2}\right)^{-1}\begin{bmatrix}\mathbf{h}_1 \\ \mathbf{h}_2\end{bmatrix}\right]^{-1}$$

with the definition $\mathbf{H}_i^{\mathsf{H}} = [\mathbf{h}_i\mathbf{U}_i^{\mathsf{H}}]$.
The inner inverse matrix is

$$\left(\begin{bmatrix}\mathbf{U}_1^{\mathsf{H}} \\ \mathbf{U}_2^{\mathsf{H}}\end{bmatrix}[\mathbf{U}_1\mathbf{U}_2] - z\mathbf{I}_{n_1+n_2}\right)^{-1} = \begin{bmatrix}\mathbf{U}_1^{\mathsf{H}}\mathbf{U}_1 - z\mathbf{I}_{n_1} & \mathbf{U}_1^{\mathsf{H}}\mathbf{U}_2 \\ \mathbf{U}_2^{\mathsf{H}}\mathbf{U}_1 & \mathbf{U}_2^{\mathsf{H}}\mathbf{U}_2 - z\mathbf{I}_{n_2}\end{bmatrix}^{-1}$$

on which we apply another block matrix inverse lemma. The upper-left $(n_1 \times n_1)$ entry equals

$$\left(-z\mathbf{U}_1^{\mathsf{H}}(\mathbf{U}_2\mathbf{U}_2^{\mathsf{H}} - z\mathbf{I}_{N-1})^{-1}\mathbf{U}_1 - z\mathbf{I}_{n_1}\right)^{-1}$$

For the second block diagonal entry, it suffices to revert all 1's in 2's and vice-versa.

Pastur's method (2)

$$\left(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}_{11} =$$

$$\left[-z - z[\mathbf{h}_1^{\mathsf{H}}\mathbf{h}_2^{\mathsf{H}}]\begin{bmatrix}\left(-z\mathbf{U}_1^{\mathsf{H}}(\mathbf{U}_2\mathbf{U}_2^{\mathsf{H}} - z\mathbf{I}_{N-1})^{-1}\mathbf{U}_1 - z\mathbf{I}_{n_1}\right)^{-1} & \star \\ \star & \left(-z\mathbf{U}_2^{\mathsf{H}}(\mathbf{U}_1\mathbf{U}_1^{\mathsf{H}} - z\mathbf{I}_{N-1})^{-1}\mathbf{U}_2 - z\mathbf{I}_{n_2}\right)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{h}_1 \\ \mathbf{h}_2\end{bmatrix}\right]^{-1}$$

The other two terms do not depend on $\mathbf{h}_1$, $\mathbf{h}_2$. We now use both results,

For $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{y} \in \mathbb{C}^N$ i.i.d. with zero mean, variance $1/N$, $\mathbf{A} \in \mathbb{C}^{N\times N}$ Hermitian with bounded spectral norm,

$$\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\operatorname{tr}\mathbf{A} \xrightarrow{\text{a.s.}} 0$$

$$\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{y} \xrightarrow{\text{a.s.}} 0$$

Since $\mathbf{R}_1$, $\mathbf{R}_2$ are diagonal, $\mathbf{h}_i = \sqrt{r_{i1}}\mathbf{T}_i^{\frac{1}{2}}\mathbf{x}_i$, with the notation $\mathbf{R}_i = \operatorname{diag}(r_{i1}, \ldots, r_{iN})$. Therefore, using the trace and rank-1 perturbation lemma,

$$\left(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}_{11} \to$$

$$\left[-z - zr_{11}\frac{1}{n_1}\operatorname{tr}\mathbf{T}_1\left(-z\mathbf{H}_1^{\mathsf{H}}(\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{H}_1 - z\mathbf{I}_{n_1}\right)^{-1} - zr_{21}\frac{1}{n_2}\operatorname{tr}\mathbf{T}_2\left(-z\mathbf{H}_2^{\mathsf{H}}(\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{H}_1 - z\mathbf{I}_{n_2}\right)^{-1}\right]^{-1}$$

Pastur's method (3)

Now, denoting $\mathbf{H}_i = [\tilde{\mathbf{h}}_i \tilde{\mathbf{U}}_i]$ (column selection instead of row),

$$\mathbf{T}_1 \left(-z\mathbf{H}_1^{\mathrm{H}} (\mathbf{H}_2\mathbf{H}_2^{\mathrm{H}} - z\mathbf{I}_N)^{-1} \mathbf{H}_1 - z\mathbf{I}_{n_1} \right)^{-1}_{11} = \tau_{11} \left[-z - z\tilde{\mathbf{h}}_1^{\mathrm{H}} \left(\tilde{\mathbf{U}}_1 \tilde{\mathbf{U}}_1^{\mathrm{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathrm{H}} - z\mathbf{I}_N \right)^{-1} \tilde{\mathbf{h}}_1 \right]^{-1}$$

$$\to \tau_{11} \left[-z - zc_1\tau_{11} \frac{1}{N} \operatorname{tr} \mathbf{R}_1 \left(\mathbf{H}_1\mathbf{H}_1^{\mathrm{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathrm{H}} - z\mathbf{I}_N \right)^{-1} \right]^{-1}$$

with τ_{ij} the j^{th} diagonal entry of $\mathbf{T}_i$. A similar result holds when changing 1's in 2's. Call now

$$f_i = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{H}_1\mathbf{H}_1^{\mathrm{H}} + \mathbf{H}_2\mathbf{H}_2^{\mathrm{H}} - z\mathbf{I}_N \right)^{-1}$$

and

$$\bar{f}_i = \frac{1}{n_i} \operatorname{tr} \mathbf{T}_i \left(\mathbf{H}_1^{\mathrm{H}} (\mathbf{H}_2\mathbf{H}_2^{\mathrm{H}} - z\mathbf{I}_N)^{-1} \mathbf{H}_1 + \mathbf{I}_{n_1} \right)^{-1}.$$

We have shown

$$f_i = \lim_{N\to\infty} \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\bar{f}_1 \mathbf{R}_1 + \bar{f}_2 \mathbf{R}_2 - z\mathbf{I}_N \right)^{-1}$$

$$\bar{f}_i = \lim_{N\to\infty} \frac{1}{n_i} \operatorname{tr} \mathbf{T}_i \left(c_i f_i \mathbf{T}_i i + \mathbf{I}_{n_i} \right)^{-1}.$$

Deterministic equivalent approach: guess work

We will use here the "guess-work" method to find the deterministic equivalent. Consider the simpler case $K = 1$.
Back to the original notations, we seek a matrix $\mathbf{D}$ such that

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{D}^{-1} \xrightarrow{\text{a.s.}} 0$$

as $N \to \infty$.

Resolvent lemma

For invertible $\mathbf{A}$, $\mathbf{B}$ matrices,

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = -\mathbf{A}^{-1}(\mathbf{A} - \mathbf{B})\mathbf{B}^{-1}$$

Taking the matrix differences,

$$\mathbf{D}^{-1} - \left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} = \mathbf{D}^{-1}\left(\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_N - \mathbf{D}\right)\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$$

It seems convenient to take $\mathbf{D} = -z\mathbf{I}_N + \bar{e}_{\mathbf{B}_N}\mathbf{R}$ with $\bar{e}_{\mathbf{B}_N}$ left to be defined (the notation $\mathbf{B}_N$ in $\bar{e}_{\mathbf{B}_N}$ reminds that we do not look *yet* for a deterministic quantity).

Deterministic equivalent approach: guess work (2)

With $\mathbf{D} = -z\mathbf{I}_N + \bar{e}_{\mathbf{B}_N}\mathbf{R}$,

$$\begin{aligned}\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} &= \mathbf{D}^{-1}(\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_N - \mathbf{D})(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\ &= \mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\left(\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\right)\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\end{aligned}$$

$$\left[\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}} = \sum_{j=1}^{n}\tau_j\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\right] = \mathbf{D}^{-1}\sum_{j=1}^{n}\tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

Deterministic equivalent approach: guess work (2)

With $\mathbf{D} = -z\mathbf{I}_N + \bar{e}_{\mathbf{B}_N}\mathbf{R}$,

$$\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \mathbf{D}^{-1}(\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_N - \mathbf{D})(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

$$= \mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\left(\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\right)\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

$$\left[\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}} = \sum_{j=1}^{n}\tau_j\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\right] = \mathbf{D}^{-1}\sum_{j=1}^{n}\tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

Taking the trace, we notice that

$$\operatorname{tr}\mathbf{D}^{-1}\sum_{j=1}^{n}\tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \sum_{j=1}^{n}\tau_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j$$

and we want to apply the trace lemma! But $\mathbf{B}_N$ is NOT independent of $\mathbf{x}_j$.

Deterministic equivalent approach: guess work (2)

With $\mathbf{D} = -z\mathbf{I}_N + \bar{e}_{\mathbf{B}_N}\mathbf{R}$,

$$\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \mathbf{D}^{-1}(\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_N - \mathbf{D})(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

$$= \mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\left(\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}}\right)\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

$$\left[\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}} = \sum_{j=1}^{n}\tau_j\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\right] = \mathbf{D}^{-1}\sum_{j=1}^{n}\tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

Taking the trace, we notice that

$$\operatorname{tr}\mathbf{D}^{-1}\sum_{j=1}^{n}\tau_j\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \sum_{j=1}^{n}\tau_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j$$

and we want to apply the trace lemma! But $\mathbf{B}_N$ is NOT independent of $\mathbf{x}_j$. We need a further step:

A matrix inversion lemma (MIL)

Let $\mathbf{A}$ be Hermitian invertible, then for any vector $\mathbf{x}$ and scalar τ such that $\mathbf{A} + \tau\mathbf{x}\mathbf{x}^{\mathrm{H}}$ is invertible

$$\mathbf{x}^{\mathrm{H}}(\mathbf{A} + \tau\mathbf{x}\mathbf{x}^{\mathrm{H}})^{-1} = \frac{\mathbf{x}^{\mathrm{H}}\mathbf{A}^{-1}}{1 + \tau\mathbf{x}\mathbf{A}^{-1}\mathbf{x}^{\mathrm{H}}}$$

Deterministic equivalent approach: guess work (2b)

Applying the MIL on $\mathbf{A} = \mathbf{B}_N - \tau_j \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j \mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}}$ and $\mathbf{A} + \tau \mathbf{x}\mathbf{x}^{\mathsf{H}} = \mathbf{B}_N$ ($\tau = \tau_j$, $\mathbf{x} = \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j$),

$$\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \sum_{j=1}^{n} \tau_j \frac{\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}}{1 + \tau_j \mathbf{x}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \bar{e}_{\mathbf{B}_N}\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

with $\mathbf{B}_{(j)} = \mathbf{B}_N - \tau_j \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j \mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}}$.
Choice of $\bar{e}_{\mathbf{B}_N}$:

$$\bar{e}_{\mathbf{B}_N} = \frac{1}{n}\sum_{j=1}^{n} \frac{\tau_j}{1 + \tau_j c \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}} \triangleq \frac{1}{n}\operatorname{tr}\mathbf{T}\left(\mathbf{I}_n + \mathbf{T}c e_{\mathbf{B}_N}\right)^{-1}$$

$$\begin{aligned}
&\frac{1}{N}\operatorname{tr}\mathbf{R}\mathbf{D}^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\
&= \frac{1}{N}\operatorname{tr}\mathbf{R}\left(\bar{e}_{\mathbf{B}_N}\mathbf{R} - z\mathbf{I}_N\right)^{-1} - e_{\mathbf{B}_N} \\
&= \frac{1}{N}\sum_{j=1}^{n}\tau_j \left[\frac{\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1 + \tau_j\mathbf{x}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j} - \frac{\frac{1}{n}\operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}}{1 + c\tau_j\frac{1}{N}\operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}}\right] \\
&\xrightarrow{\text{a.s.}} 0.
\end{aligned}$$

Deterministic equivalent approach: guess work (3)

- We now use the trace lemma and standard inequalities to show

$$\sum_N E\left[\left|\frac{1}{N}\operatorname{tr}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\left(\bar{e}_{\mathbf{B}_N}\mathbf{R} - z\mathbf{I}_N\right)^{-1}\right|^p\right] < \infty.$$

  for some integer p, imposing that $\mathbf{x}_i$ has entries of finite moments of order $2p$. The same can be done for $\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$ and we get

$$\sum_N E\left[\left|\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{R}\mathbf{D}^{-1}\right|^p\right] < \infty.$$

  Applying Markov inequality and the Borel-Cantelli lemma, we finally have the almost-sure convergence.
- to extend to more generic conditions on $\mathbf{x}_i$ (only finite variance requirement), truncation steps may be applied.

Truncation tool

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.

Truncation and centralization

Replace $\mathbf{X}_N$, $\mathbf{T}_N$ and $\mathbf{R}_N$ by $\bar{\mathbf{X}}_N$, $\bar{\mathbf{T}}_N$ and $\bar{\mathbf{R}}_N$ in the following fashion

$$\left(\bar{\mathbf{X}}_N\right)_{ij} = \left(\mathbf{X}_N\right)_{ij} \cdot I_{\{(\mathbf{X}_N)_{ij} < g_N\}}$$

for g_N that grows

- fast enough to ensure $F^{\mathbf{B}_N} - F^{\bar{\mathbf{B}}_N} \Rightarrow 0$
- slow enough to ensure $\frac{1}{N}\operatorname{tr}\left(\bar{\mathbf{B}}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\bar{\mathbf{R}}\bar{\mathbf{D}}^{-1} \xrightarrow{\text{a.s.}} 0$

Deterministic equivalent approach: existence and uniqueness

R. D. Yates, "A framework for uplink power control in cellular radio systems", IEEE Journal on Selected Areas in Communications, vol. 13, no. 7, pp. 1341-1347, 1995.

To prove existence and uniqueness, we use a constructive approach based on results on standard interference functions:

Definition

A function $\mathbf{h}(x_1, \ldots, x_K) \in \mathbb{R}^K$, $\mathbf{h} = (h_1, \ldots, h_K)$, with $x_1, \ldots, x_K \in \mathbb{R}^+$, is a standard function or a standard interference function if

1. *Positivity:* for all j, $h_j(x_1, \ldots, x_K) > 0$,
2. *Monotonicity:* if $x_1 > x_1', \ldots, x_K > x_K'$, then $h_j(x_1, \ldots, x_K) > h_j(x_1', \ldots, x_K')$, for all j,
3. *Scalability:* for all $\alpha > 1$ and j, $\alpha h_j(x_1, \ldots, x_K) > h_j(\alpha x_1, \ldots, \alpha x_K)$.

Deterministic equivalent approach: existence and uniqueness

R. D. Yates, "A framework for uplink power control in cellular radio systems", IEEE Journal on Selected Areas in Communications, vol. 13, no. 7, pp. 1341-1347, 1995.

To prove existence and uniqueness, we use a constructive approach based on results on standard interference functions:

Definition

A function $\mathbf{h}(x_1, \ldots, x_K) \in \mathbb{R}^K$, $\mathbf{h} = (h_1, \ldots, h_K)$, with $x_1, \ldots, x_K \in \mathbb{R}^+$, is a standard function or a standard interference function if

1. *Positivity:* for all j, $h_j(x_1, \ldots, x_K) > 0$,
2. *Monotonicity:* if $x_1 > x_1', \ldots, x_K > x_K'$, then $h_j(x_1, \ldots, x_K) > h_j(x_1', \ldots, x_K')$, for all j,
3. *Scalability:* for all $\alpha > 1$ and j, $\alpha h_j(x_1, \ldots, x_K) > h_j(\alpha x_1, \ldots, \alpha x_K)$.

Theorem

If $\mathbf{h}(x_1, \ldots, x_K)$ is a standard interference function and there exists $(x_1, \ldots, x_K)$ such that, for all j, $x_j \geqslant h_j(x_1, \ldots, x_K)$, then the iteration

$$x_j^{(t+1)} = h_j(x_1^{(t)}, \ldots, x_K^{(t)}), \quad x_1^{(0)}, \ldots, x_K^{(0)} > 0$$

converges to the unique solution of $x_j = h_j(x_1, \ldots, x_K)$, $j \in \{1, \ldots, K\}$.

Deterministic equivalent approach: existence and uniqueness (2)

We proceed as follows:

- we fix $z < 0$, and show that the function

$$h : e \mapsto \frac{1}{N}\operatorname{tr}\mathbf{R}\left(\bar{e}\mathbf{R} - z\mathbf{I}_N\right)^{-1}, \quad \bar{e} = \frac{1}{n}\operatorname{tr}\mathbf{T}\left(\mathbf{I}_n + ce(z)\mathbf{T}\right)^{-1}$$

  is a standard function.
- we conclude that $e^t(z) = h(e^{t-1}(z))$, $e^0(z) = 0$, converge to $e(z)$.
- standard holomorphicity arguments (Vitali's convergence theorem) then show that $e(z)$ is defined and holomorphic on $\mathbb{C} \setminus \mathbb{R}^+$.

Deterministic equivalent approach: termination of the proof

- It then suffices to show that $\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - e \xrightarrow{\text{a.s.}} 0$
  This exploits the fact that, for some ω in a probability one space,

  $$\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{R}\left(\bar{e}_{\mathbf{B}_N}\mathbf{R} - z\mathbf{I}_N\right)^{-1} = w_N(\omega) \to 0.$$

  For $z \in \mathbb{C} \setminus \mathbb{R}^+$, we have in particular

  $$e(z) - \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1} = \gamma(z)\left(e - \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1}\right) + w_N(\omega)$$

  where $\gamma(z) \to 0$ as $|z| \to \infty$. Taking z such that $\gamma(z) < 1$ for all N, n, ω, we have

  $$e(z) - \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1} = \frac{w_N}{1 - \gamma(z)} \to 0.$$

  Vitali's convergence theorem extends this result to $z \in \mathbb{C} \setminus \mathbb{R}^+$. The result is proved.
- The same argument then applies to $m_{\mathbf{B}_N}(z) - m_N(z)$.

Deterministic equivalent approach: termination of the proof

- It then suffices to show that $\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - e \xrightarrow{\text{a.s.}} 0$
  This exploits the fact that, for some ω in a probability one space,

$$\frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}\mathbf{R}\left(\bar{e}_{\mathbf{B}_N}\mathbf{R} - z\mathbf{I}_N\right)^{-1} = w_N(\omega) \to 0.$$

  For $z \in \mathbb{C} \setminus \mathbb{R}^+$, we have in particular

$$e(z) - \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1} = \gamma(z)\left(e - \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1}\right) + w_N(\omega)$$

  where $\gamma(z) \to 0$ as $|z| \to \infty$. Taking z such that $\gamma(z) < 1$ for all N, n, ω, we have

$$e(z) - \frac{1}{N}\operatorname{tr}\mathbf{R}(\mathbf{B}_N(\omega) - z\mathbf{I}_N)^{-1} = \frac{w_N}{1 - \gamma(z)} \to 0.$$

  Vitali's convergence theorem extends this result to $z \in \mathbb{C} \setminus \mathbb{R}^+$. The result is proved.
- The same argument then applies to $m_{\mathbf{B}_N}(z) - m_N(z)$.
- We have proved $F^{\mathbf{B}_N} - F_N \Rightarrow 0$, almost surely.

Result on the Shannon transform of $\mathbf{B}_N$

Remember now that

$$\int \log(1+xt)\,dF(t) = \int_{1/x}^{\infty} \left(\frac{1}{t} - m_F(-t)\right) dt.$$

Result on the Shannon transform of $\mathbf{B}_N$

Remember now that

$$\int \log(1+xt)dF(t) = \int_{1/x}^{\infty} \left(\frac{1}{t} - m_F(-t)\right) dt.$$

R. Couillet, M. Debbah, J. W. Silverstein, "A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels," IEEE Trans. on Information Theory, vol. 57, no. 6, pp. 3493-3514, 2011..

Theorem

Under the previous model for $\mathbf{B}_N$, as N, n_k grow large,

$$E\left[\frac{1}{N}\log\det(x\mathbf{B}_N + \mathbf{I}_N)\right] - \left[\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_k(-1/x)\mathbf{R}_k\right)\right.$$
$$+ \sum_{k=1}^{K} \frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-1/x)\mathbf{T}_k\right)$$
$$\left. - \frac{1}{x}\sum_{k=1}^{K} \bar{e}_k(-1/x)e_k(-1/x)\right] \xrightarrow{\text{a.s.}} 0$$

Result on the Shannon transform of $\mathbf{B}_N$

Remember now that

$$\int \log(1+xt)dF(t) = \int_{1/x}^{\infty} \left(\frac{1}{t} - m_F(-t) \right) dt.$$

R. Couillet, M. Debbah, J. W. Silverstein, "A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels," IEEE Trans. on Information Theory, vol. 57, no. 6, pp. 3493-3514, 2011..

Theorem

Under the previous model for $\mathbf{B}_N$, as N, n_k grow large,

$$E\left[\frac{1}{N}\log\det(x\mathbf{B}_N + \mathbf{I}_N)\right] - \left[\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K} \bar{e}_k(-1/x)\mathbf{R}_k\right)\right.$$

$$+ \sum_{k=1}^{K} \frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-1/x)\mathbf{T}_k^{\frac{1}{2}}\mathbf{P}_k\mathbf{T}_k^{\frac{1}{2}}\right)$$

$$\left. - \frac{1}{x}\sum_{k=1}^{K} \bar{e}_k(-1/x)e_k(-1/x)\right] \xrightarrow{\text{a.s.}} 0$$

Variance profile

W. Hachem, Ph. Loubaton, J. Najim, "Deterministic equivalents for certain functionals of large random matrices," Annals of Applied Probability, vol. 17, no. 3, pp. 875-930, 2007.

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have independent entries with $(i,j)^{th}$ entry of zero mean and variance $\frac{1}{n}\sigma_{ij}^2$. Let $\mathbf{A}_N \in \mathbb{R}^{N\times n}$ be deterministic with uniformly bounded column norm. Then

$$\frac{1}{N} tr \left((\mathbf{X}_N + \mathbf{A}_N)(\mathbf{X}_N + \mathbf{A}_N)^{\mathsf{H}} - z\mathbf{I}_N \right)^{-1} - \frac{1}{N} tr \mathbf{T}_N(z) \xrightarrow{\text{a.s.}} 0$$

where $\mathbf{T}_N(z)$ is the unique function that solves

$$\mathbf{T}_N(z) = \left(\Psi^{-1}(z) - z\mathbf{A}_N \tilde{\Psi}(z) \mathbf{A}_N^{\mathsf{T}} \right)^{-1}, \quad \tilde{\mathbf{T}}_N(z) = \left(\tilde{\Psi}^{-1}(z) - z\mathbf{A}_N^{\mathsf{T}} \Psi(z) \mathbf{A}_N \right)^{-1}$$

with $\Psi(z) = diag(\psi_i(z))$, $\tilde{\Psi}(z) = diag(\tilde{\psi}_i(z))$, with entries defined as

$$\psi_i(z) = -\left(z(1 + \frac{1}{n} tr \tilde{\mathbf{D}}_i \tilde{\mathbf{T}}(z)) \right)^{-1}, \quad \tilde{\psi}_j(z) = -\left(z(1 + \frac{1}{n} tr \mathbf{D}_j \mathbf{T}(z)) \right)^{-1}$$

and $\mathbf{D}_j = diag(\sigma_{ij}^2, 1 \leqslant i \leqslant N)$, $\tilde{\mathbf{D}}_i = diag(\sigma_{ij}^2, 1 \leqslant j \leqslant n)$

Variance profile

W. Hachem, Ph. Loubaton, J. Najim, "Deterministic equivalents for certain functionals of large random matrices," Annals of Applied Probability, vol. 17, no. 3, pp. 875-930, 2007.

Theorem

For the previous model, we also have that

$$\frac{1}{N} E \log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}(\mathbf{X}_N + \mathbf{A}_N)(\mathbf{X}_N + \mathbf{A}_N)^{\mathsf{H}}\right)$$

has deterministic equivalent

$$\frac{1}{N}\log\det\left[\frac{1}{\sigma^2}\Psi(-\sigma^2)^{-1} + \mathbf{A}_N\tilde{\Psi}(-\sigma^2)\mathbf{A}_N^{\mathsf{T}}\right]$$
$$+\frac{1}{N}\log\det\frac{1}{\sigma^2}\Psi(-\sigma^2)^{-1} - \frac{\sigma^2}{nN}\sum_{i,j}\sigma_{ij}^2\mathbf{T}_{ii}(-\sigma^2)\tilde{\mathbf{T}}_{jj}(-\sigma^2).$$

Haar random matrices

M. Debbah, W. Hachem, P. Loubaton, M. de Courville, "MMSE analysis of certain large isometric random precoded systems", IEEE Transactions on Information Theory, vol. 49, no. 5, pp. 1293-1311, 2003.

- Recent results were proposed when the matrices $\mathbf{X}_N$ are unitary and unitarily invariant (Haar matrices).

Haar random matrices

M. Debbah, W. Hachem, P. Loubaton, M. de Courville, "MMSE analysis of certain large isometric random precoded systems", IEEE Transactions on Information Theory, vol. 49, no. 5, pp. 1293-1311, 2003.

- Recent results were proposed when the matrices $\mathbf{X}_N$ are unitary and unitarily invariant (Haar matrices).
- The central result is the trace lemma

Lemma

Let $\mathbf{W} \in \mathbb{C}^{N \times n}$ be $n < N$ columns of a Haar matrix and $\mathbf{w}$ a column of $\mathbf{W}$. Let $\mathbf{B}_N \in \mathbb{C}^{N \times N}$ a random matrix, function of all columns of $\mathbf{W}$ except $\mathbf{w}$. Then, assuming that, for growing N, $c = \sup_n n/N < 1$ and $B = \sup_N \|\mathbf{B}_N\| < \infty$, we have:

$$\mathbf{w}^{\mathsf{H}} \mathbf{B}_N \mathbf{w} - \frac{1}{N-n} tr\,(\mathbf{I}_N - \mathbf{W}\mathbf{W}^{\mathsf{H}})\mathbf{B}_N \xrightarrow{\text{a.s.}} 0.$$

Haar random matrices (2)

R. Couillet, J. Hoydis, M. Debbah, "Random beamforming over quasi-static and fading channels: a deterministic equivalent approach", to appear in IEEE Trans. on Inf. Theory.

Theorem

Let $\mathbf{T}_i \in \mathbb{C}^{n_i \times n_i}$ *be nonnegative diagonal and let* $\mathbf{H}_i \in \mathbb{C}^{N \times N_i}$*. Define* $\mathbf{R}_i \triangleq \mathbf{H}_i \mathbf{H}_i^{\mathrm{H}} \in \mathbb{C}^{N \times N}$*,* $c_i = \frac{n_i}{N_i}$ *and* $\bar{c}_i = \frac{N_i}{N}$*. Denote*

$$\mathbf{B}_N = \sum_{i=1}^{K} \mathbf{H}_i \mathbf{W}_i \mathbf{T}_i \mathbf{W}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}}.$$

Then, as $N, N_1, \ldots, N_K, n_1, \ldots, n_K \to \infty$ *with ratios bounded* $\bar{c}_i$ *and* $0 \leqslant c_i \leqslant 1$ *for all* i*, almost surely*

$$F^{\mathbf{B}_N} - F_N \Rightarrow 0, \quad \text{with } m_N(z) = \frac{1}{N} tr \left(\sum_{i=1}^{K} \bar{e}_i(z) \mathbf{R}_i - z \mathbf{I}_N \right)^{-1}$$

where $(\bar{e}_1(z), \ldots, \bar{e}_K(z))$ *are the solutions (conditionally unique) of*

$$e_i(z) = \frac{1}{N} tr \mathbf{R}_i \left(\sum_{j=1}^{K} \bar{e}_j(z) \mathbf{R}_j - z \mathbf{I}_N \right)^{-1}$$

$$\bar{e}_i(z) = \frac{1}{N} tr \mathbf{T}_i \left(e_i(z) \mathbf{T}_i + [\bar{c}_i - e_i(z) \bar{e}_i(z)] \mathbf{I}_{n_i} \right)^{-1} \quad \text{(compare to i.i.d. case!)}$$

Alternative strategies

There exists alternative proof strategies for specific models.

- The Gaussian method:
  - this technique is meant for random Gaussian $\mathbf{X}$ matrices
  - based on two ingredients: a Gaussian integration by parts formula, and the Nash-Poincaré inequality.
  - *advantages:*
    - sequential method, easy to use
    - give results on convergence speed: $N(E\, m_{\mathbf{B}_N} - m_N) \to 0$
    - goes beyond the Stieltjes transform method for quadratic forms.
  - *drawbacks:*
    - somewhat painful to use, leads to much calculus, less "elegant"
    - proves convergence of Gaussian-based models of type $N(E\, m_{\mathbf{B}_N} - m_N) \to 0$ (but interpolation trick can then be used)

Alternative strategies

There exists alternative proof strategies for specific models.

- The Gaussian method:
  - this technique is meant for random Gaussian **X** matrices
  - based on two ingredients: a Gaussian integration by parts formula, and the Nash-Poincaré inequality.
  - *advantages:*
    - sequential method, easy to use
    - give results on convergence speed: $N(E\, m_{\mathbf{B}_N} - m_N) \to 0$
    - goes beyond the Stieltjes transform method for quadratic forms.
  - *drawbacks:*
    - somewhat painful to use, leads to much calculus, less "elegant"
    - proves convergence of Gaussian-based models of type $N(E\, m_{\mathbf{B}_N} - m_N) \to 0$ (but interpolation trick can then be used)
- Diagrammatic approaches: moment "drawing"-based approach that uses combinatorics from nuclear physics to infer limiting results
- Replica methods: physics-based method, non-mathematically accurate, to *conjecture* limiting results.

Outline

Fluctuations of functionals of the e.s.d.

Z. D. Bai and J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," The Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.

Theorem

$$\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}, \quad \underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$$

as usual with $\mathbf{X}_N$ *Gaussian,* $F^{\mathbf{T}_N} = diag(\{\tau_i\}) \Rightarrow H$, $|\mathbf{T}_N|$, $\tau_1 \geqslant \ldots \geqslant \tau_N$. *Denote* F *and* F_N *the l.s.d. and det. eq. of* $F^{\mathbf{B}_N}$, *and*

$$G_N \triangleq N\left[F^{\mathbf{B}_N} - F_N\right].$$

For $f_1, \ldots, f_k$ *well behaved, then*

$$\left(\int f_1(x)dG_N(x), \ldots, \int f_k(x)dG_N(x)\right) \Rightarrow (X_{f_1}, \ldots, X_{f_k})$$

of zero mean and covariance $\mathrm{Cov}(X_f, X_g)$, $(f, g) \in \{f_1, \ldots, f_k\}^2$, *such that*

$$\mathrm{Cov}(X_f, X_g) = -\frac{1}{2\pi i}\oint\oint \frac{f(z_1)g(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2}\underline{m}'(z_1)\underline{m}'(z_2)dz_1dz_2$$

for $\underline{m}(z)$ *the Stieltjes transform of the l.s.d. of* $\underline{\mathbf{B}}_N$. *The integration contours are positively defined with winding number one and enclose the support of* F.

Related bibliography

- R. B. Dozier and J. W. Silverstein, "On the Empirical Distribution of Eigenvalues of Large Dimensional Information-Plus-Noise Type Matrices," Journal of Multivariate Analysis 98(4), pp. 678-694, 2007.
- Z. D. Bai, J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," The Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.
- W. Hachem, Ph. Loubaton, and J. Najim, "A CLT for information theoretic statistics of Gram random matrices with a given variance profile," *Annals of Applied Probability*, vol. 18, no. 6, pp. 2071-2130, 2008.
- R. Couillet, M. Debbah, J. Silverstein, "A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels," IEEE Transactions on Information Theory, vol. 57, no. 6, pp. 3493-3514, 2011.
- R. Couillet, S. Wagner, M. Debbah, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels under Limited Feedback," to appear in IEEE Trans. on Information Theory.
- M. Debbah, W. Hachem, P. Loubaton, M. de Courville, "MMSE analysis of certain large isometric random precoded systems", IEEE Transactions on Information Theory, vol. 49, no. 5, pp. 1293-1311, 2003.
- R. Couillet, J. Hoydis, M. Debbah, "Random Beamforming over Quasi-Static and Fading Channels: A Deterministic Equivalent Approach", to appear in IEEE Transactions on Information Theory.
- A. Kammoun, M. Kharouf, W. Hachem, J. Najim, "A central limit theorem for the SINR at the LMMSE estimator output for large-dimensional signals," IEEE Transactions on Information Theory, vol. 55, no. 11, pp. 5048-5063, 2009.
- 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.
- A. Moustakas, S. Simon and A. Sengupta, "MIMO Capacity Through Correlated Channels in the Presence of Correlated Interferers and Noise: A (Not So) Large N Analysis," IEEE Trans. on Information Theory, vol. 49, no. 10, 2003.
- D. Guo and S. Verdú, "Multiuser Detection and Statistical Physics," Communications on Information and Network Security, Kluwer Academic Publishers, 2003.

Outline

Why iterated deterministic equivalents?

- Deterministic equivalents for very involved channel models have been established.
- Most works deal with sums of independent random matrices.

Why iterated deterministic equivalents?

- Deterministic equivalents for very involved channel models have been established.
- Most works deal with sums of independent random matrices.
- Iterated deterministic equivalents can be used for the study of more complex combinations of independents matrices.

Why iterated deterministic equivalents?

- Deterministic equivalents for very involved channel models have been established.
- Most works deal with sums of independent random matrices.
- Iterated deterministic equivalents can be used for the study of more complex combinations of independents matrices.
- Applications extend to models that cannot be treated by free probability theory.

Iterated deterministic equivalents: A simple example

$$\mathbf{B}_N = \mathbf{Y}_n \mathbf{X}_n \mathbf{X}_n^{\mathsf{H}} \mathbf{Y}_n^{\mathsf{H}}$$

- $\mathbf{X}_n \in \mathbb{C}^{n \times n}$: $[\mathbf{X}]_{i,j} \sim \mathcal{CN}\left(0, \frac{1}{n}\right)$
- $\mathbf{Y}_n \in \mathbb{C}^{n \times n}$: $[\mathbf{Y}]_{i,j} \sim \mathcal{CN}\left(0, \frac{1}{n}\right)$
- $\limsup_n \left\| \mathbf{Y}_n \mathbf{Y}_n^{\mathsf{H}} \right\| < \infty$, almost surely

Iterated deterministic equivalents: A simple example

$$\mathbf{B}_N = \mathbf{Y}_n \mathbf{X}_n \mathbf{X}_n^{\mathsf{H}} \mathbf{Y}_n^{\mathsf{H}}$$

- $\mathbf{X}_n \in \mathbb{C}^{n \times n}$: $[\mathbf{X}]_{i,j} \sim \mathcal{CN}\left(0, \frac{1}{n}\right)$
- $\mathbf{Y}_n \in \mathbb{C}^{n \times n}$: $[\mathbf{Y}]_{i,j} \sim \mathcal{CN}\left(0, \frac{1}{n}\right)$
- $\limsup_n \left\|\mathbf{Y}_n \mathbf{Y}_n^{\mathsf{H}}\right\| < \infty$, almost surely

Idea: $\mathbf{Y}_n \mathbf{X}_n$ looks like a random matrix $\mathbf{X}_n$ with random left-sided correlation $\mathbf{Y}_n$.

Iterated deterministic equivalents: A simple example

$$\mathbf{B}_N = \mathbf{Y}_n \mathbf{X}_n \mathbf{X}_n^{\mathsf{H}} \mathbf{Y}_n^{\mathsf{H}}$$

- $\mathbf{X}_n \in \mathbb{C}^{n \times n}$: $[\mathbf{X}]_{i,j} \sim \mathcal{CN}\left(0, \frac{1}{n}\right)$
- $\mathbf{Y}_n \in \mathbb{C}^{n \times n}$: $[\mathbf{Y}]_{i,j} \sim \mathcal{CN}\left(0, \frac{1}{n}\right)$
- $\limsup_n \left\| \mathbf{Y}_n \mathbf{Y}_n^{\mathsf{H}} \right\| < \infty$, almost surely

Idea: $\mathbf{Y}_n \mathbf{X}_n$ looks like a random matrix $\mathbf{X}_n$ with random left-sided correlation $\mathbf{Y}_n$.

From Bai and Silverstein result and the Fubini theorem, we have

$$\frac{1}{n} \mathrm{tr} \left(\mathbf{B}_n + x \mathbf{I}_n \right)^{-1} - \frac{1}{n} \mathrm{tr} \left(\frac{\mathbf{Y}_n \mathbf{Y}_n^{\mathsf{H}}}{1 + e_n(x)} + x \mathbf{I}_n \right)^{-1} \xrightarrow{\text{a.s.}} 0$$

where $e_n(x)$ is the unique positive solution to

$$e_n(x) = \frac{1}{n} \mathrm{tr}\, \mathbf{Y}_n \mathbf{Y}_n^{\mathsf{H}} \left(\frac{\mathbf{Y}_n \mathbf{Y}_n^{\mathsf{H}}}{1 + e_n(x)} + x \mathbf{I}_n \right)^{-1}.$$

Iterated deterministic equivalents: A simple example (2)

After straightforward calculations, we have

$$\frac{1}{n}\mathrm{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)}+x\mathbf{I}_n\right)^{-1}=(1+e_n(x))\,\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1}$$

$$e_n(x)=\sqrt{\frac{1}{x\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1}}}-1$$

Iterated deterministic equivalents: A simple example (2)

After straightforward calculations, we have

$$\frac{1}{n}\mathrm{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)}+x\mathbf{I}_n\right)^{-1}=(1+e_n(x))\,\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1}$$

$$e_n(x)=\sqrt{\frac{1}{x\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1}}}-1$$

Applying Bai and Silverstein again yields:

$$\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1}-\frac{\sqrt{1+\frac{4}{x(1+e_n(x))}}-1}{2}\xrightarrow{\text{a.s.}}0.$$

Iterated deterministic equivalents: A simple example (2)

After straightforward calculations, we have

$$\frac{1}{n}\mathrm{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)}+x\mathbf{I}_n\right)^{-1} = (1+e_n(x))\,\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1}$$

$$e_n(x) = \sqrt{\frac{1}{x\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1}}}-1$$

Applying Bai and Silverstein again yields:

$$\frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}+x\,(1+e_n(x))\,\mathbf{I}_n\right)^{-1} - \frac{\sqrt{1+\frac{4}{x(1+e_n(x))}}-1}{2} \xrightarrow{\text{a.s.}} 0.$$

Thus,

$$e_n(x) - \sqrt{\frac{2}{\sqrt{x^2+\frac{4x}{1+e_n(x)}}-x}}+1 \xrightarrow{\text{a.s.}} 0$$

Iterated deterministic equivalents: A simple example (3/3)

Putting all pieces together, one can show that:

$$(i) \qquad \frac{1}{n}\operatorname{tr}\left(\mathbf{Y}_n\mathbf{X}_n\mathbf{X}_n^{\mathsf{H}}\mathbf{Y}_n^{\mathsf{H}} + x\mathbf{I}_n\right)^{-1} - \frac{1}{n}\operatorname{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)} + x\mathbf{I}_n\right)^{-1} \xrightarrow{\text{a.s.}} 0$$

Iterated deterministic equivalents: A simple example (3/3)

Putting all pieces together, one can show that:

$$(i) \qquad \frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{X}_n\mathbf{X}_n^{\mathsf{H}}\mathbf{Y}_n^{\mathsf{H}} + x\mathbf{I}_n\right)^{-1} - \frac{1}{n}\mathrm{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)} + x\mathbf{I}_n\right)^{-1} \xrightarrow{\text{a.s.}} 0$$

$$(ii) \qquad \frac{1}{n}\mathrm{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)} + x\mathbf{I}_n\right)^{-1} - (1+\bar{e}(x))\,\frac{\sqrt{1+\frac{4}{x(1+\bar{e}(x))}}-1}{2} \xrightarrow{\text{a.s.}} 0$$

where $\bar{e}(x)$ is the unique positive solution to

$$\bar{e}(x) = \sqrt{\frac{2}{\sqrt{x^2+\frac{4x}{1+\bar{e}(x)}}-x}} - 1.$$

Iterated deterministic equivalents: A simple example (3/3)

Putting all pieces together, one can show that:

$$(i) \qquad \frac{1}{n}\mathrm{tr}\left(\mathbf{Y}_n\mathbf{X}_n\mathbf{X}_n^{\mathsf{H}}\mathbf{Y}_n^{\mathsf{H}} + x\mathbf{I}_n\right)^{-1} - \frac{1}{n}\mathrm{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)} + x\mathbf{I}_n\right)^{-1} \xrightarrow{\text{a.s.}} 0$$

$$(ii) \qquad \frac{1}{n}\mathrm{tr}\left(\frac{\mathbf{Y}_n\mathbf{Y}_n^{\mathsf{H}}}{1+e_n(x)} + x\mathbf{I}_n\right)^{-1} - (1+\bar{e}(x))\,\frac{\sqrt{1+\frac{4}{x(1+\bar{e}(x))}}-1}{2} \xrightarrow{\text{a.s.}} 0$$

where $\bar{e}(x)$ is the unique positive solution to

$$\bar{e}(x) = \sqrt{\frac{2}{\sqrt{x^2+\frac{4x}{1+\bar{e}(x)}}-x}} - 1.$$

- The same result can be obtained by arguments of free probability theory.
- We did not require the asymptotic freeness of $\mathbf{X}_n$,$\mathbf{Y}_n$ in the proof.
- The same approach can be extended to more involved channel models.

Application to other models

Multiple models can be analyzed through the iterated deterministic equivalent approach:

R. Couillet, J. Hoydis, M. Debbah, "Random beamforming over quasi-static and fading channels: a deterministic equivalent approach", (to appear in) IEEE Transactions on Information Theory, arXiv Preprint 1011.3717.

- model $\mathbf{B}_N = \mathbf{H}\mathbf{W}\mathbf{P}\mathbf{W}^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}$, with
  - $\mathbf{W} \in \mathbb{C}^{n\times p}$ Haar,
  - $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n] \in \mathbb{C}^{N\times n}$, $\mathbf{h}_i = \mathbf{R}_i^{\frac{1}{2}}\mathbf{x}_i$, $\mathbf{x}_i \sim \mathcal{CN}(0, \mathbf{I}_N)$,
  - $\mathbf{P}$ Hermitian nonnegative.

J. Hoydis, R. Couillet, M. Debbah, "Asymptotic Analysis of Double-Scattering Channels ", Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2011.

- model $\mathbf{B}_N = \mathbf{H}_1\mathbf{H}_2\mathbf{H}_2^{\mathsf{H}}\mathbf{H}_1^{\mathsf{H}}$ with $\mathbf{H}_1 = \mathbf{R}_1\mathbf{X}_1\mathbf{T}_1$, $\mathbf{H}_2 = \mathbf{R}_2\mathbf{X}_2\mathbf{T}_2$
  - $\mathbf{R}_1, \mathbf{R}_2, \mathbf{T}_1, \mathbf{T}_2$ Hermitian nonnegative
  - $\mathbf{X}_1$, $\mathbf{X}_2$ independent with i.i.d. entries

J. Hoydis, R. Couillet, M. Debbah, "Iterative deterministic equivalents for the capacity analysis of communication systems", (submitted to) IEEE Transactions on Information Theory, 2011.

- model $\mathbf{B}_N = \left(\prod_{i=1}^{N}\mathbf{H}_i\right)\left(\prod_{i=1}^{N}\mathbf{H}_i\right)^{\mathsf{H}}$, $\mathbf{H}_i$ with i.i.d. entries.

Related bibliography

- R. Couillet, J. Hoydis, M. Debbah, "Random beamforming over quasi-static and fading channels: a deterministic equivalent approach", (to appear in) IEEE Transactions on Information Theory, arXiv Preprint 1011.3717.
- J. Hoydis, R. Couillet, M. Debbah, "Asymptotic Analysis of Double-Scattering Channels ", Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2011.
- J. Hoydis, R. Couillet, M. Debbah, "Iterative deterministic equivalents for the capacity analysis of communication systems", (submitted to) IEEE Transactions on Information Theory, 2011.