Random Matrix Advances in Signal Processing

SPAWC 2013, Darmstadt, Germany.

Romain Couillet and Mérouane Debbah

SUPELEC, Gif sur Yvette, France

June 17th, 2013

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

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

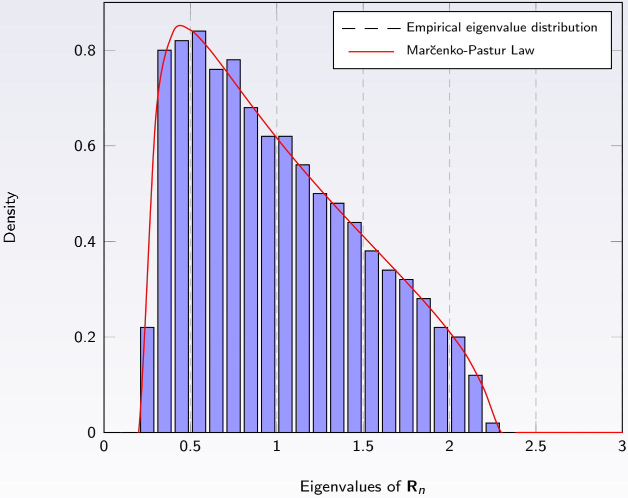


Figure : Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

Finite size against 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 against 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 tutorial will exclusively focus on limiting results.

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

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

Example

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

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

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

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

with F_c the Marčenko-Pastur law with parameter c.

Why is this useful to signal processing?

- increasing system dimensions: large antenna arrays, large datasets, (not so) large number of snapshots
- need for detection and estimation based on large dimensional random inputs: subspace methods in array processing
- the assumption "sample space $>>$ population space" is less and less valid: large arrays, systems with fast dynamics

Example

MUSIC with "few" samples (or in large arrays) Call $\mathbf{A}(\Theta) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{N \times K}$, N large, K small, the steering vectors to identify and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{C}^{N \times n}$ the n samples, taken from

$$y_t = \sum_{k=1}^{K} \mathbf{a}(\theta_k)\sqrt{p}_k s_{k,t} + \sigma w_t.$$

The MUSIC localization function reads $\gamma(\theta) = \mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{U}}_W\hat{\mathbf{U}}_W^{\mathsf{H}}\mathbf{a}(\theta)$ in the "signal vs. noise" spectral decomposition $\mathbf{YY}^{\mathsf{H}} = \hat{\mathbf{U}}_S\hat{\mathbf{\Lambda}}_S\hat{\mathbf{U}}_S^{\mathsf{H}} + \hat{\mathbf{U}}_W\hat{\mathbf{\Lambda}}_W\hat{\mathbf{U}}_W^{\mathsf{H}}$.

Why is this useful to signal processing?

- increasing system dimensions: large antenna arrays, large datasets, (not so) large number of snapshots
- need for detection and estimation based on large dimensional random inputs: subspace methods in array processing
- the assumption "sample space >> population space" is less and less valid: large arrays, systems with fast dynamics

Example

MUSIC with "few" samples (or in large arrays) Call $\mathbf{A}(\Theta) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{N\times K}$, N large, K small, the steering vectors to identify and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{C}^{N\times n}$ the n samples, taken from

$$y_t = \sum_{k=1}^{K} \mathbf{a}(\theta_k)\sqrt{p}_k s_{k,t} + \sigma w_t.$$

The MUSIC localization function reads $\gamma(\theta) = \mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{U}}_W\hat{\mathbf{U}}_W^{\mathsf{H}}\mathbf{a}(\theta)$ in the "signal vs. noise" spectral decomposition $\mathbf{Y}\mathbf{Y}^{\mathsf{H}} = \hat{\mathbf{U}}_S\hat{\mathbf{\Lambda}}_S\hat{\mathbf{U}}_S^{\mathsf{H}} + \hat{\mathbf{U}}_W\hat{\mathbf{\Lambda}}_W\hat{\mathbf{U}}_W^{\mathsf{H}}$.
Writing equivalently $\mathbf{A}(\Theta)\mathbf{P}\mathbf{A}(\Theta)^{\mathsf{H}} + \sigma^2\mathbf{I}_N = \mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^{\mathsf{H}} + \sigma^2\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}}$, as $n, N \to \infty$, $n/N \to c$, from our previous remarks

$$\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}} \not\to \hat{\mathbf{U}}_W\hat{\mathbf{U}}_W^{\mathsf{H}}$$

$\Rightarrow$ Music is NOT consistent in the large N, n regime! We need improved RMT-based solutions.

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) $F^{\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}} \Rightarrow F_c$ as $N, n \to \infty$, $N/n \to c$, with F_c the Marčenko-Pastur law with density

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

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

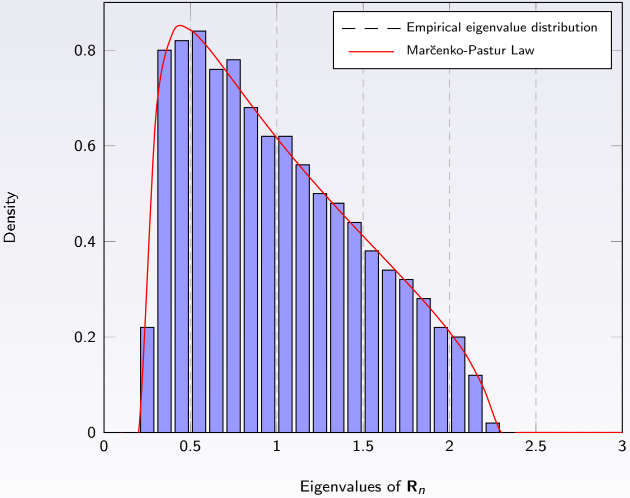


Figure : Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

Semi-circle law

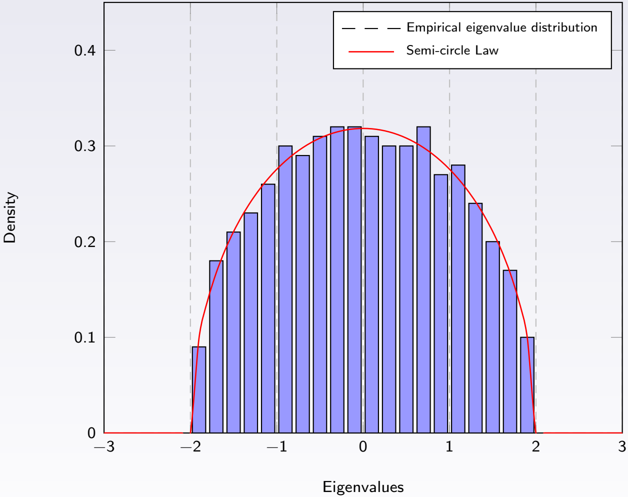


Figure : Histogram of the eigenvalues of Wigner matrices and the semi-circle law, for $N = 500$

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

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

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

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

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

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

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

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

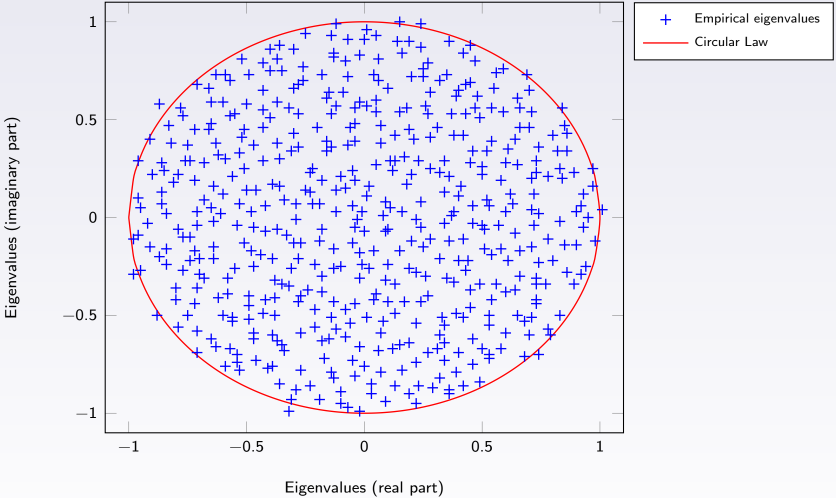


Figure : Eigenvalues of $\mathbf{X}_N$ with i.i.d. standard Gaussian entries, for $N = 500$.

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.

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 tutorial

- **Part 1: Fundamentals of Random Matrix Theory**
  - 1.1. Introduction to the Stieltjes transform method and proof of the Marčenko–Pastur law
  - 1.2. Extreme eigenvalues: no eigenvalue outside the support, exact separation, and Tracy–Widom law
  - 1.3. Extreme eigenvalues: the spiked model
  - 1.4. Spectrum analysis and G-estimation

Outline of the tutorial

- **Part 1: Fundamentals of Random Matrix Theory**
  - 1.1. Introduction to the Stieltjes transform method and proof of the Marčenko–Pastur law
  - 1.2. Extreme eigenvalues: no eigenvalue outside the support, exact separation, and Tracy–Widom law
  - 1.3. Extreme eigenvalues: the spiked model
  - 1.4. Spectrum analysis and G-estimation
- **Part 2: Source Detection**
  - 2.1. Eigenvalue-based detection

Outline of the tutorial

- **Part 1: Fundamentals of Random Matrix Theory**
  - 1.1. Introduction to the Stieltjes transform method and proof of the Marčenko–Pastur law
  - 1.2. Extreme eigenvalues: no eigenvalue outside the support, exact separation, and Tracy–Widom law
  - 1.3. Extreme eigenvalues: the spiked model
  - 1.4. Spectrum analysis and G-estimation
- **Part 2: Source Detection**
  - 2.1. Eigenvalue-based detection
- **Part 3: Statistical Inference**
  - 3.1. Generic model: source power and direction of arrival estimation (G-MUSIC)
  - 3.2. Spiked model case: spiked G-MUSIC, offline failure diagnosis in sensor networks

Outline of the tutorial

- **Part 1: Fundamentals of Random Matrix Theory**
  - 1.1. Introduction to the Stieltjes transform method and proof of the Marčenko–Pastur law
  - 1.2. Extreme eigenvalues: no eigenvalue outside the support, exact separation, and Tracy–Widom law
  - 1.3. Extreme eigenvalues: the spiked model
  - 1.4. Spectrum analysis and G-estimation
- **Part 2: Source Detection**
  - 2.1. Eigenvalue-based detection
- **Part 3: Statistical Inference**
  - 3.1. Generic model: source power and direction of arrival estimation (G-MUSIC)
  - 3.2. Spiked model case: spiked G-MUSIC, offline failure diagnosis in sensor networks
- **Part 4: Random Matrix Theory and Robust Estimation**
  - 4.1. Introduction to robust estimation
  - 4.2. Initial results and open problems

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

The Stieltjes transform

Definition

Let F be a real 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 $a < b$ continuity points of F, 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 real 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 $a < b$ continuity points of F, 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 at x, then

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

The Stieltjes transform

Definition

Let F be a real 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 $a < b$ continuity points of F, 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 at x, then

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

Equivalence $F \leftrightarrow m_F$

Similar to the Fourier transform, knowing m_F is the same as knowing F.

Stieltjes transform and Matrix Spectra

- 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

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

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

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

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

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

hence

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

Asymptotic results using the Stieltjes transform

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

Theorem

Let $\underline{\mathbf{B}}_N = \mathbf{X}_N \mathbf{T}_N \mathbf{X}_N^{\mathsf{H}} \in \mathbb{C}^{N \times N}$, *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 $\{z \in \mathbb{C}^+, m_{\underline{F}}(z) \in \mathbb{C}^+\}$.

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 $\{z \in \mathbb{C}^+, m_{\underline{F}}(z) \in \mathbb{C}^+\}$.

- 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 $\mathbb{C}^+$.

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 $\mathbb{C}^+$.
- 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 $\mathbb{C}^+$.
- 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:* spectral analysis, to come next.

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 $\mathbb{C}^+$.
- 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:* spectral analysis, to come next.

Example (Sample covariance matrix)

For N multiple of 3, let $F^{\mathbf{T}_N}(x) = \frac{1}{3}\mathbf{1}_{x \leqslant 1} + \frac{1}{3}\mathbf{1}_{x \leqslant 3} + \frac{1}{3}\mathbf{1}_{x \leqslant K}$ 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

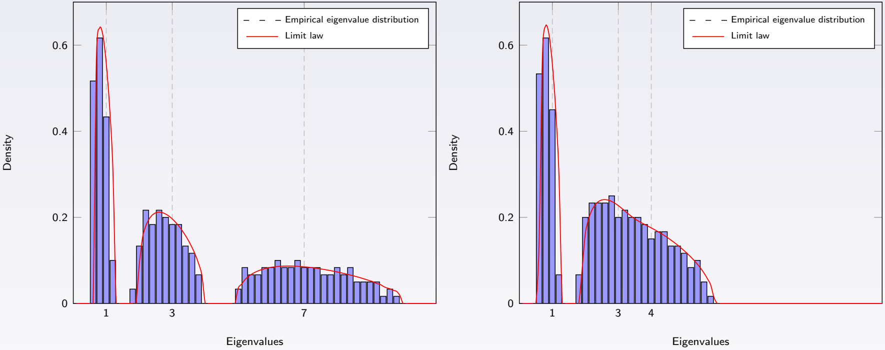


Figure : Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$, $N = 3000$, $n = 300$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

Side remark: 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!
- Note that m_F itself is of interest, not F!

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

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

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

The Marčenko-Pastur density

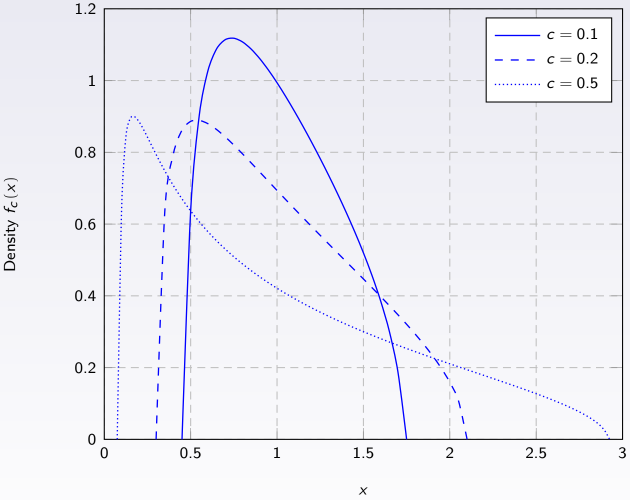


Figure : Marčenko-Pastur law for different limit ratios $c = \lim_{N \to \infty} N/n$.

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

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

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

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

which solve a polynomial of second order. Finally

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

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

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

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

A classical pitfall

- Limiting spectral results only say where the "mass" of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $F_N(x) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{x \leqslant a_k}$.

A classical pitfall

- Limiting spectral results only say where the "mass" of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $F_N(x) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{x \leqslant a_k}$.
  - $F_N^{(0)}(x) = \frac{1}{N}\delta(x) + \frac{1}{N}\sum_{k=1}^{N-1} \mathbf{1}_{x \leqslant a_k}$ also converges to F.

A classical pitfall

- Limiting spectral results only say where the "mass" of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $F_N(x) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{x \leqslant a_k}$.
  - $F_N^{(0)}(x) = \frac{1}{N}\delta(x) + \frac{1}{N}\sum_{k=1}^{N-1} \mathbf{1}_{x \leqslant a_k}$ also converges to F.
  - more generally, if F_N and $F_N^{(0)}$ are discrete and differ by $o(N)$ bounded masses, $F_N^{(0)} \Rightarrow F$.

A classical pitfall

- Limiting spectral results only say where the “mass” of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $F_N(x) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{x \leqslant a_k}$.
  - $F_N^{(0)}(x) = \frac{1}{N}\delta(x) + \frac{1}{N}\sum_{k=1}^{N-1} \mathbf{1}_{x \leqslant a_k}$ also converges to F.
  - more generally, if F_N and $F_N^{(0)}$ are discrete and differ by $o(N)$ bounded masses, $F_N^{(0)} \Rightarrow F$.
- We know that, for $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ with i.i.d. zero mean variance $1/n$,

$$F^{\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}} \Rightarrow F_c$$

with F_c is the compactly supported Marčenko-Pastur law of parameter $c = \lim_N \frac{N}{n}$.

Question: for very large N, where are the extreme eigenvalues of $\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$?

Are there eigenvalues outside the support ?

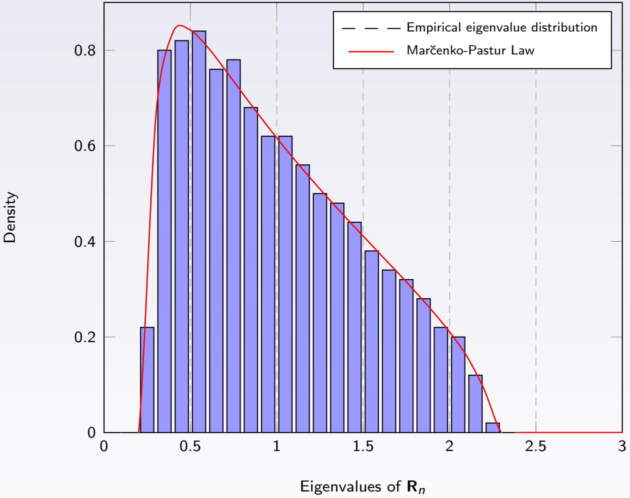


Figure : Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$

No eigenvalue outside the support of sample covariance matrices

Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," The Annals of Probability, vol. 26, no.1 pp. 316-345, 1998.

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ with i.i.d. entries with zero mean, variance $1/n$ and 4^{th} order moment of order $O(1/n^2)$. Let $\mathbf{T}_N \in \mathbb{C}^{N \times N}$ be nonrandom and bounded in norm and with $F^{\mathbf{T}_N} \Rightarrow F^T$. We know that

$$F^{\mathbf{B}_N} \Rightarrow F \text{ almost surely}, \quad \mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}.$$

Let F_N be the distribution with $m_N(z)$ solution of

$$\underline{m}_N = -\left(z - \frac{N}{n} \int \frac{\tau}{1 + \tau \underline{m}_N} dF^{\mathbf{T}_N}(\tau) \right)^{-1}, \quad \underline{m}_N(z) = \frac{N}{n} m_N(z) + \frac{N-n}{n} \frac{1}{z}.$$

Choose $N_0 \in \mathbb{N}$ and $[a, b]$, $a > 0$, outside the union of the supports of F and F_N for all $N \geqslant N_0$. Denote $\mathcal{L}_N$ the set of eigenvalues of $\mathbf{B}_N$. Then,

$$P(\mathcal{L}_N \cap [a, b] \neq \emptyset \text{ i.o.}) = 0.$$

How to read the result?

- If $\mathbf{T}_N = \mathbf{I}_N$ for all N, then this result is equivalent to

  "For $[a, b]$ outside the support of the Marčenko-Pastur law, with probability 1, $\mathbf{B}_N$ has no eigenvalue in $[a, b]$ for all large N"

How to read the result?

- If $\mathbf{T}_N = \mathbf{I}_N$ for all N, then this result is equivalent to

  "For $[a, b]$ outside the support of the Marčenko-Pastur law, with probability 1, $\mathbf{B}_N$ has no eigenvalue in $[a, b]$ for all large N"
- If $\mathbf{T}_N$ is not identity,
  - call S the support of the limiting F.
  - for some N_0, take the l.s.d. of $\mathbf{B}_N$ as if $\lim_N F^{\mathbf{T}_N} = F^{\mathbf{T}_{N_0}}$, and call its support S_{N_0}.
  - do the previous for all $N \geqslant N_0$. Call $\mathcal{A} = S \cup \bigcap_{N \geqslant N_0} S_N$.
  - take $[a, b]$ outside $\mathcal{A}$, and pick a random sequence $\mathbf{B}_1, \mathbf{B}_2, \ldots$. The result shows that, for all N large, there is no eigenvalue of $\mathbf{B}_N$ in $[a, b]$.

How to read the result?

- If $\mathbf{T}_N = \mathbf{I}_N$ for all N, then this result is equivalent to

  "For $[a,b]$ outside the support of the Marčenko-Pastur law, with probability 1, $\mathbf{B}_N$ has no eigenvalue in $[a,b]$ for all large N"
- If $\mathbf{T}_N$ is not identity,
  - call S the support of the limiting F.
  - for some N_0, take the l.s.d. of $\mathbf{B}_N$ as if $\lim_N F^{\mathbf{T}_N} = F^{\mathbf{T}_{N_0}}$, and call its support S_{N_0}.
  - do the previous for all $N \geqslant N_0$. Call $\mathcal{A} = S \cup \bigcap_{N \geqslant N_0} S_N$.
  - take $[a,b]$ outside $\mathcal{A}$, and pick a random sequence $\mathbf{B}_1, \mathbf{B}_2, \ldots$. The result shows that, for all N large, there is no eigenvalue of $\mathbf{B}_N$ in $[a,b]$.
- this is very different from taking $[a,b]$ only outside the support of F only!
- this is essential to understand spiked models, discussed later.

No eigenvalue outside the support: which models?

J. W. Silverstein, P. Debashis, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has already been shown that (for all large N) there is no eigenvalues outside the support of
  - *Marčenko-Pastur law*: $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Sample covariance matrix*: $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{T}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment.

No eigenvalue outside the support: which models?

J. W. Silverstein, P. Debashis, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has already been shown that (for all large N) there is no eigenvalues outside the support of
  - *Marčenko-Pastur law*: $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Sample covariance matrix*: $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{T}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment.

J. W. Silverstein, Z.D. Bai, Y.Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," Journal of Multivariate Analysis, vol. 26, no. 2, pp. 166-168, 1988.

- If 4^{th} order moment is infinite,

$$\limsup_N \lambda_{\max}^{\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \infty$$

No eigenvalue outside the support: which models?

J. W. Silverstein, P. Debashis, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has already been shown that (for all large N) there is no eigenvalues outside the support of
  - *Marčenko-Pastur law*: $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Sample covariance matrix*: $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{T}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment.

J. W. Silverstein, Z.D. Bai, Y.Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," Journal of Multivariate Analysis, vol. 26, no. 2, pp. 166-168, 1988.

- If 4^{th} order moment is infinite,

$$\limsup_N \lambda_{\max}^{\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \infty$$

J. Silverstein, Z. Bai, "No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices" to appear in Random Matrices: Theory and Applications.

- Only recently, information plus noise models, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment

$$(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}}$$

Sketch of Proof

- Proof entirely relies on the Stieltjes transform.
- Up to now, we know $|m_{\mathbf{B}_N}(z) - m_N(z)| \xrightarrow{\text{a.s.}} 0$ for $z \in \mathbb{C} \setminus \mathbb{R}^-$.

Sketch of Proof

- Proof entirely relies on the Stieltjes transform.
- Up to now, we know $|m_{\mathbf{B}_N}(z) - m_N(z)| \xrightarrow{\text{a.s.}} 0$ for $z \in \mathbb{C} \setminus \mathbb{R}^-$.
- This is not enough, we need in fact to show: for $z = x + i\sqrt{k}v_N$, $v_N = N^{-1/68}$, $k = 1, \ldots, 34$,

$$\max_{1 \leqslant k \leqslant 34} \sup_{x \in [a,b]} \left| m_{\mathbf{B}_N}(x + ik^{\frac{1}{2}} v_N) - m_N((x + ik^{\frac{1}{2}} v_N) \right| = o(v_N^{67}).$$

Sketch of Proof

- Proof entirely relies on the Stieltjes transform.
- Up to now, we know $|m_{\mathbf{B}_N}(z) - m_N(z)| \xrightarrow{\text{a.s.}} 0$ for $z \in \mathbb{C} \setminus \mathbb{R}^-$.
- This is not enough, we need in fact to show: for $z = x + i\sqrt{k}v_N$, $v_N = N^{-1/68}$, $k = 1, \ldots, 34$,

$$\max_{1 \leqslant k \leqslant 34} \sup_{x \in [a,b]} \left| m_{\mathbf{B}_N}(x + ik^{\frac{1}{2}} v_N) - m_N((x + ik^{\frac{1}{2}} v_N) \right| = o(v_N^{67}).$$

- Expanding the Stieltjes transforms and considering only the imaginary parts, this is

$$\max_{1 \leqslant k \leqslant 34} \sup_{x \in [a,b]} \left| \int \frac{d(F^{\mathbf{B}_N}(\lambda) - F_N(\lambda))}{(x-\lambda)^2 + kv_N^2} \right| = o(v_N^{66})$$

almost surely. Taking successive differences over the 34 values of k, we end up with

$$\sup_{x \in [a,b]} \left| \int \frac{(v_N^2)^{33} d(F^{\mathbf{B}_N}(\lambda) - F_N(\lambda))}{\prod_{k=1}^{34}((x-\lambda)^2 + kv_N^2)} \right| = o(v_N^{66})$$

Consider $a' < a$ and $b' > b$ such that $[a', b']$ is outside the support of F. We then have

$$\sup_{x \in [a,b]} \left| \int \frac{1_{\mathbb{R}^+ \setminus [a',b']}(\lambda) d(F\mathbf{B}_N(\lambda) - F_N(\lambda))}{\prod_{k=1}^{34}((x-\lambda)^2 + kv_N^2)} + \sum_{\lambda_j \in [a',b']} \frac{v_N^{68}}{\prod_{k=1}^{34}((x-\lambda_j)^2 + kv_N^2)} \right| = o(1)$$

almost surely. If, there is one eigenvalue of all $\mathbf{B}_{\phi(N)}$ in $[a, b]$, then one term of the sum is $1/34! > 0$. So the integral must away from zero. But the integral tends to 0. Contradiction.

Exact eigenvalue separation

Z. D. Bai, J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.

- The result on "no eigenvalue outside the support"
  - says where eigenvalues are not to be found
  - does not say, as we feel, that (if cluster separation) in cluster k, there are exactly n_k eigenvalues.

Exact eigenvalue separation

Z. D. Bai, J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.

- The result on "no eigenvalue outside the support"
  - says where eigenvalues are not to be found
  - does not say, as we feel, that (if cluster separation) in cluster k, there are exactly n_k eigenvalues.
- This is in fact the case,

Theorem

Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^H \mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F, $\mathbf{X}_N$ i.i.d., zero mean, variance $1/n$, finite 4^{th} moment, $F^{\mathbf{T}_N} \Rightarrow F^T$, and $\frac{N}{n} \to c$. Consider $0 < a < b$ such that $[a, b]$ is outside the support of F. Denote additionally λ_k's and τ_k's the ordered eigenvalues of $\mathbf{B}_N$ and $\mathbf{T}_N$. Then we have

1. *If $c(1 - F^T(0)) > 1$, then the smallest eigenvalue x_0 of the support of F is positive and $\lambda_N \to x_0$ almost surely, as $N \to \infty$.*
2. *If $c(1 - F^T(0)) \leqslant 1$, or $c(1 - F^T(0)) > 1$ but $[a, b]$ is not contained in $[0, x_0]$, then, almost surely, there exists N_0 such that for all $N \geqslant N_0$,*

$$\lambda_{i_N} > b, \quad \lambda_{i_N+1} < a$$

where i_N is the unique integer such that

$$\tau_{i_N} > -1/m_F(b)$$

$$\tau_{i_N+1} < -1/m_F(a).$$

Consequence of exact separation

- If eigenvalues are found outside the expected clusters, some extra "signal" must have been transmitted.
- The quantity of eigenvalues in each cluster gives an exact estimate of the multiplicity of the population!
- This is essential for eigen-inference.

Consequence of exact separation

- If eigenvalues are found outside the expected clusters, some extra “signal” must have been transmitted.
- The quantity of eigenvalues in each cluster gives an exact estimate of the multiplicity of the population!
- This is essential for eigen-inference.
- Exact separation is only known for the sample covariance matrix model so far.
- Very recently, extension to information-plus-noise model.

What's the use of all that to signal processing?

Assume N sensors wish to detect the presence of a signal. They scan successive samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then

What's the use of all that to signal processing?

Assume N sensors wish to detect the presence of a signal. They scan successive samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then

- if $\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$ has eigenvalues outside the support: with high probability, the data source is not i.i.d. white and may contain informative data.

What's the use of all that to signal processing?

Assume N sensors wish to detect the presence of a signal. They scan successive samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then

- if $\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}$ has eigenvalues outside the support: with high probability, the data source is not i.i.d. white and may contain informative data.
- if $\mathbf{R}_n$ has all eigenvalues inside the *expected* noise support, what can we say?
  - we cannot conclude so far
  - we need to further study the spectrum

Extreme eigenvalues: Deeper into the spectrum

- In order to derive statistical detection tests, we need more information on the extreme eigenvalues.

Extreme eigenvalues: Deeper into the spectrum

- In order to derive statistical detection tests, we need more information on the extreme eigenvalues.
- We will study the fluctuations of the extreme eigenvalues (second order statistics)
- However, the Stieltjes transform method is not adapted here!

Distribution of the largest eigenvalues of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$

C. A. Tracy, H. Widom, "On orthogonal and symplectic matrix ensembles," Communications in Mathematical Physics, vol. 177, no. 3, pp. 727-754, 1996.
K. Johansson, "Shape Fluctuations and Random Matrices," Comm. Math. Phys. vol. 209, pp. 437-476, 2000.

Theorem

Let $\mathbf{X} \in \mathbb{C}^{N \times n}$ have i.i.d. Gaussian entries of zero mean and variance $1/n$. Denoting λ_N^+ the largest eigenvalue of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, then

$$N^{\frac{2}{3}} \frac{\lambda_N^+ - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{\frac{1}{2}}} \Rightarrow X^+ \sim F^+$$

with $c = \lim_N N/n$ and F^+ the Tracy-Widom distribution given by

$$F^+(t) = \exp\left(- \int_t^\infty (x-t)^2 q^2(x) dx \right)$$

with q the Painlevé II function that solves the differential equation

$$q''(x) = xq(x) + 2q^3(x)$$
$$q(x) \sim_{x \to \infty} \mathrm{Ai}(x)$$

in which $\mathrm{Ai}(x)$ is the Airy function.

The law of Tracy-Widom

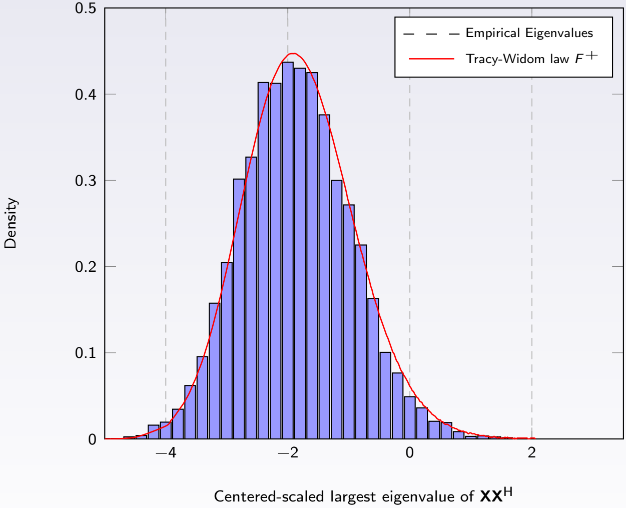


Figure : Distribution of $N^{\frac{2}{3}} c^{-\frac{1}{2}} (1+\sqrt{c})^{-\frac{4}{3}} \left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the distribution of X^+ (distributed as Tracy-Widom law) for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{XX}^{\mathsf{H}}$. Empirical distribution taken over $10,000$ Monte-Carlo simulations.

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \ldots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

with $K(x, y)$ the kernel Laguerre polynomial.

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \ldots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

  with $K(x, y)$ the kernel Laguerre polynomial.

- *Fredholm determinants*: we can write hole probability as a Fredholm determinant.

$$P\left(N^{2/3}\left(\lambda_i - (1+\sqrt{c})^2\right) \in A, i = 1, \ldots, N\right) = 1 + \sum_{k \geqslant 1} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K_N(x_i, x_j) \prod dx_i$$

$$\triangleq \det(\mathbf{I}_N - \mathcal{K}_N).$$

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \ldots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

  with $K(x,y)$ the kernel Laguerre polynomial.
- *Fredholm determinants*: we can write hole probability as a Fredholm determinant.

$$P\left(N^{2/3}\left(\lambda_i - (1+\sqrt{c})^2\right) \in A, i = 1, \ldots, N\right) = 1 + \sum_{k \geqslant 1} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K_N(x_i, x_j) \prod dx_i$$
$$\triangleq \det(\mathbf{I}_N - \mathcal{K}_N).$$

- *kernel theory*: show that K_N converges to a Airy kernel.

$$K_N(x,y) \to K_{\mathrm{Airy}}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}.$$

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \ldots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

  with $K(x,y)$ the kernel Laguerre polynomial.

- *Fredholm determinants*: we can write hole probability as a Fredholm determinant.

$$P\left(N^{2/3}\left(\lambda_i - (1+\sqrt{c})^2\right) \in A, i = 1, \ldots, N\right) = 1 + \sum_{k \geqslant 1} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K_N(x_i, x_j) \prod dx_i$$
$$\triangleq \det(\mathbf{I}_N - \mathcal{K}_N).$$

- *kernel theory*: show that K_N converges to a Airy kernel.

$$K_N(x,y) \to K_{\mathrm{Airy}}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}.$$

- *differential equation tricks*: hole probability in $[t, \infty)$ gives right-most eigenvalue distribution, which is simplified as solution of a Painelvé differential equation: the Tracy-Widom distribution.

$$F^+(t) = e^{-\int_t^\infty (x-t) q(x)^2 dx}, \quad q'' = tq + 2q^3, \; q(x) \sim_{x\to\infty} \mathrm{Ai}(x).$$

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives a hint on convergence speed
- fairly biased on the left: even fewer eigenvalues outside the support.

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives a hint on convergence speed
- fairly biased on the left: even fewer eigenvalues outside the support.
- can be shown to hold for other distributions than Gaussian under mild assumptions

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. F^T of $\mathbf{T}_N$ is a Dirac in 1.

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. F^T of $\mathbf{T}_N$ is a Dirac in 1.
- No contradiction with "no eigenvalue" theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additional clusters of positive measure in F_N.

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. F^T of $\mathbf{T}_N$ is a Dirac in 1.
- No contradiction with "no eigenvalue" theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additional clusters of positive measure in F_N.
- However, for practical purposes, the presence of "spikes" determines the presence of a signal!

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. F^T of $\mathbf{T}_N$ is a Dirac in 1.
- No contradiction with "no eigenvalue" theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additional clusters of positive measure in F_N.
- However, for practical purposes, the presence of "spikes" determines the presence of a signal!

What about the absence of spikes?

The first result

J. Baik, J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," Journal of Multivariate Analysis, vol. 97, no. 6, pp. 1382-1408, 2006.

Theorem

Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, *where* $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *has i.i.d., zero mean and variance* $1/n$ *entries, and* $\mathbf{T}_N \in \mathbb{R}^{N\times N}$ *diagonal given by*

$$\mathbf{T}_N = diag(\underbrace{1+\omega_1,\ldots,1+\omega_1}_{k_1},\ldots,\underbrace{1+\omega_M,\ldots,1+\omega_M}_{k_M},\underbrace{1,\ldots,1}_{N-\sum_{i=1}^M k_i})$$

with $\omega_1 > \ldots > \omega_M > -1$, $c = \lim_N N/n$. *We then have*

- *if* $\omega_j > \sqrt{c}$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1+\omega_j + c\frac{1+\omega_j}{\omega_j}$ *(i.e. beyond the Marčenko–Pastur bulk!)*
- *if* $\omega_{k_j} \in (0,\sqrt{c}]$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$ *(i.e. right-edge of the Marčenko–Pastur bulk!)*
- *if* $\omega_{k_j} \in [-\sqrt{c},0)$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$ *(i.e. left-edge of the Marčenko–Pastur bulk!)*

The first result

J. Baik, J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," Journal of Multivariate Analysis, vol. 97, no. 6, pp. 1382-1408, 2006.

Theorem

Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, *where* $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *has i.i.d., zero mean and variance* $1/n$ *entries, and* $\mathbf{T}_N \in \mathbb{R}^{N\times N}$ *diagonal given by*

$$\mathbf{T}_N = diag(\underbrace{1+\omega_1,\ldots,1+\omega_1}_{k_1},\ldots,\underbrace{1+\omega_M,\ldots,1+\omega_M}_{k_M},\underbrace{1,\ldots,1}_{N-\sum_{i=1}^M k_i})$$

with $\omega_1 > \ldots > \omega_M > -1$, $c = \lim_N N/n$. *We then have*

- *if* $\omega_j > \sqrt{c}$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1+\omega_j + c\frac{1+\omega_j}{\omega_j}$ *(i.e. beyond the Marčenko–Pastur bulk!)*
- *if* $\omega_{k_j} \in (0,\sqrt{c}]$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$ *(i.e. right-edge of the Marčenko–Pastur bulk!)*
- *if* $\omega_{k_j} \in [-\sqrt{c},0)$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$ *(i.e. left-edge of the Marčenko–Pastur bulk!)*
- *for the other eigenvalues, we discriminate over* c:
  - *if* $\omega_{k_j} < -\sqrt{c}$, $c<1$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1+\omega_j + c\frac{1+\omega_j}{\omega_j}$ *(i.e. beyond the Marčenko–Pastur bulk!)*
  - *if* $\omega_{k_j} < -\sqrt{c}$, $c>1$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$ *(i.e. left-edge of the Marčenko–Pastur bulk!)*

Illustration of spiked models

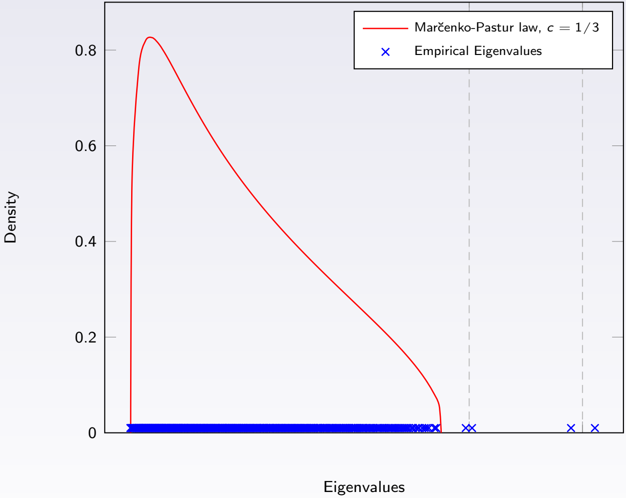


Figure : Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N} \Rightarrow 1_{[1,\infty)}$, ...Dimensions: $N = 500$, $n = 1500$.

Illustration of spiked models

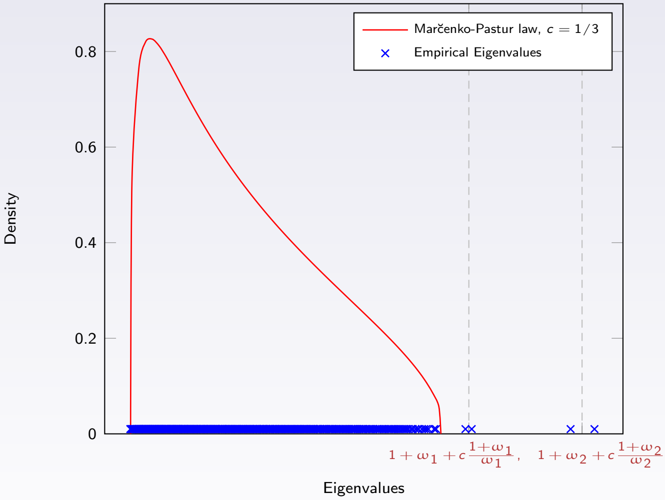


Figure : Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N} \Rightarrow 1_{[1,\infty)}$, but $\mathbf{T}_N$ is a diagonal of ones but for the first four entries set to $\{1+\omega_1, 1+\omega_1, 1+\omega_2, 1+\omega_2\}$, $\omega_1 = 1$, $\omega_2 = 2$.Dimensions: $N = 500$, $n = 1500$.

Interpretation of the result

- if c is large, or alternatively, if some “population spikes” are small, part to all of the population spikes are attracted by the support!

Interpretation of the result

- if c is large, or alternatively, if some "population spikes" are small, part to all of the population spikes are attracted by the support!
- if so, no way to decide on the existence of the spikes *from looking at the largest eigenvalues*
- in signal processing words, signals might be missed using largest eigenvalues methods.

Interpretation of the result

- if c is large, or alternatively, if some "population spikes" are small, part to all of the population spikes are attracted by the support!
- if so, no way to decide on the existence of the spikes *from looking at the largest eigenvalues*
- in signal processing words, signals might be missed using largest eigenvalues methods.
- as a consequence,
  - the more the sensors (N),
  - the larger $c = \lim N/n$,
  - the more probable we miss a spike

Interpretation of the result

- if c is large, or alternatively, if some "population spikes" are small, part to all of the population spikes are attracted by the support!
- if so, no way to decide on the existence of the spikes *from looking at the largest eigenvalues*
- in signal processing words, signals might be missed using largest eigenvalues methods.
- as a consequence,
  - the more the sensors (N),
  - the larger $c = \lim N/n$,
  - the more probable we miss a spike
  - THAT LOOKS LIKE A PARADOX.

General characterization of spiked eigenvalues

- Consider the more general model

$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P})^{\frac{1}{2}}\mathbf{X}$$

  with, for simplicity
  - $\mathbf{X}$ standard Gaussian
  - $\mathbf{P} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{C}^{N\times r}$, $\mathbf{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_r)$, $\omega_1 > \ldots > \omega_r > 0$.
- We can study the convergence properties of
  - $\lambda_1 > \ldots > \lambda_r$, the r largest eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}}$
  - $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i$, with $\hat{\mathbf{u}}_i$ the eigenvector associated to λ_i (not discussed today)

General characterization of spiked eigenvalues

- Consider the more general model
$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P})^{\frac{1}{2}}\mathbf{X}$$
with, for simplicity
  - $\mathbf{X}$ standard Gaussian
  - $\mathbf{P} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{C}^{N\times r}$, $\mathbf{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_r)$, $\omega_1 > \ldots > \omega_r > 0$.
- We can study the convergence properties of
  - $\lambda_1 > \ldots > \lambda_r$, the r largest eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}}$
  - $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i$, with $\hat{\mathbf{u}}_i$ the eigenvector associated to λ_i (not discussed today)
- Systematic study based on two ingredients:
  - random matrix tools (the Stieltjes transform method)
  - complex analysis (complex contour integration)

First order limits on eigenvalues

- We start with a study of the limiting extreme eigenvalues.

First order limits on eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\begin{aligned}\det(\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) &= \det(\mathbf{I}_N + \mathbf{P}) \det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}]) \\ &= \det(\mathbf{I}_N + \mathbf{P}) \det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1} \det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).\end{aligned}$$

First order limits on eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\det(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) = \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}])$$
$$= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).$$

- if x eigenvalue of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}$ but not of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, then for n large, $x > (1+\sqrt{c})^2$ (edge of MP law support) and

$$\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}) = \det(\mathbf{I}_r + x\boldsymbol{\Omega}\mathbf{U}^*(\mathbf{I}_N + \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}) = 0$$

with $\mathbf{P} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} \in \mathbb{C}^{N\times r}$.

First order limits on eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\det(\mathbf{\Sigma\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) = \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{XX}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}])$$
$$= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{XX}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{XX}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).$$

- if x eigenvalue of $\mathbf{\Sigma\Sigma}^{\mathsf{H}}$ but not of $\mathbf{XX}^{\mathsf{H}}$, then for n large, $x > (1+\sqrt{c})^2$ (edge of MP law support) and

$$\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{XX}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}) = \det(\mathbf{I}_r + x\mathbf{\Omega U}^*(\mathbf{I}_N + \mathbf{U\Omega U}^{\mathsf{H}})^{-1}(\mathbf{XX}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}) = 0$$

  with $\mathbf{P} = \mathbf{U\Omega U}^{\mathsf{H}}$, $\mathbf{U} \in \mathbb{C}^{N\times r}$.
- due to unitary invariance of $\mathbf{X}$,

$$\mathbf{U}^{\mathsf{H}}(\mathbf{XX}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U} \xrightarrow{\text{a.s.}} \int (t-x)^{-1} dF^{MP}(t)\mathbf{I}_r \triangleq m(x)\mathbf{I}_r$$

  with F^{MP} the MP law, and $m(x)$ the Stieltjes transform of the MP law (often known for $r = 1$ as trace lemma).

First order limits on eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\det(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) = \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}])$$
$$= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).$$

- if x eigenvalue of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}$ but not of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, then for n large, $x > (1+\sqrt{c})^2$ (edge of MP law support) and

$$\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}) = \det(\mathbf{I}_r + x\boldsymbol{\Omega}\mathbf{U}^*(\mathbf{I}_N + \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}) = 0$$

  with $\mathbf{P} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} \in \mathbb{C}^{N\times r}$.
- due to unitary invariance of $\mathbf{X}$,

$$\mathbf{U}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U} \xrightarrow{\text{a.s.}} \int (t-x)^{-1} dF^{MP}(t)\mathbf{I}_r \triangleq m(x)\mathbf{I}_r$$

  with F^{MP} the MP law, and $m(x)$ the Stieltjes transform of the MP law (often known for $r = 1$ as trace lemma).
- finally, we have that the *limiting* solutions ρ_k satisfy $\rho_k m(\rho_k) + (1+\omega_k)\omega_k^{-1} = 0$.
- replacing $m(x)$, this is finally:

$$\lambda_k \xrightarrow{\text{a.s.}} \rho_k \triangleq 1 + \omega_k + c(1+\omega_k)\omega_k^{-1}, \text{ if } \omega_k > \sqrt{c}$$

Generalization of the Tracy-Widom law

J. Baik, G. Ben Arous, S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," The Annals of Probability, vol. 33, no. 5, pp. 1643-1697, 2005.

Theorem

Let $\mathbf{X} \in \mathbb{C}^{N\times n}$ have i.i.d. Gaussian entries of zero mean and variance $1/n$ and $\mathbf{T}_N = diag(t_1, \ldots, t_N)$. Assume, for some fixed r, $t_{r+1} = \ldots = t_N = 1$ and $t_1 = \ldots = t_k$ while $t_{k+1}, \ldots, t_r$ lie in a compact subset of $(0, t_1)$. Assume further $c = \lim N/n < 1$. Denoting λ_N^+ the largest eigenvalue of $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$, we have

- *If $t_1 < 1 + \sqrt{\frac{N}{n}}$,*

$$N^{\frac{2}{3}} \frac{\lambda_N^+ - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{\frac{1}{2}}} \Rightarrow X^+ \sim F^+$$

  with F^+ the Tracy-Widom distribution.

- *If $t_1 > 1 + \sqrt{\frac{N}{n}}$,*

$$\left(t_1^2 - \frac{t_1^2 c}{(t_1-1)^2}\right)^{\frac{1}{2}} n^{\frac{1}{2}} \left[\lambda_N^+ - \left(t_1 + \frac{t_1 c}{t_1 - 1}\right)\right] \Rightarrow X_k \sim G_k$$

  for some function G_k that is the distribution of the largest eigenvalue of the $k \times k$ GUE.

$$G_k(x) = \frac{1}{Z_k} \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \prod_{1 \leqslant i < j \leqslant k} |\xi_i - \xi_j|^2 \prod_{i=1}^{k} e^{-\frac{1}{2}\xi_i^2} d\xi_1 \ldots d\xi_k$$

  In particular, $G_1(x) = \mathrm{erf}(x)$

Comments on the result

- there exists a "phase transition" when the largest population eigenvalues move from inside to outside $(0, 1 + \sqrt{c})$.

Comments on the result

- there exists a "phase transition" when the largest population eigenvalues move from inside to outside $(0, 1 + \sqrt{c})$.
- more importantly, for $t_1 < 1 + \sqrt{c}$, we still have the same Tracy-Widom,
  - no way to see the spike even when zooming in
  - in fact, simulation suggests that convergence rate to the Tracy-Widom is slower with spikes.

Presence of a spike in previous model

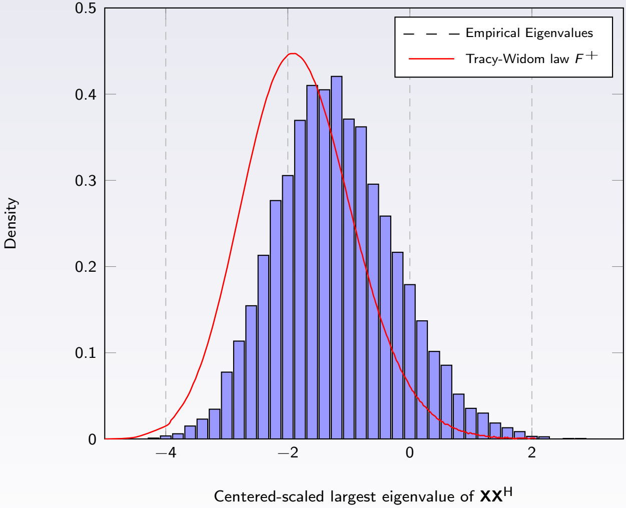


Figure : Distribution of $N^{\frac{2}{3}} c^{-\frac{1}{2}} (1+\sqrt{c})^{-\frac{4}{3}} \left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the distribution of X^+ (distributed as Tracy-Widom law) for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{T}^{\frac{1}{2}}\mathbf{XX}^\mathsf{H}\mathbf{T}^{\frac{1}{2}}$ with $\mathbf{T}$ diagonal with all entries 1 but for $T_{11} = 1.5$. Empirical distribution taken over 10,000 Monte-Carlo simulations.

Outline

Part 1: Fundamentals of Random Matrix Theory
- 1.1. The Stieltjes Transform Method
- 1.2. Extreme Eigenvalues
- 1.3. The Spiked Model
- 1.4. Spectrum Analysis and G-estimation

2. Source Detection
- 2.1. Eigenvalue-based Detection
- 2.2. Detection in unknown Noise Environment

3. Statistical Inference
- 3.1. Generic Model
  - 3.1.2. Angle-of-arrival estimation
  - 3.1.2. Angle-of-arrival estimation
- 3.2. Spiked Model
  - 3.2.1. Spiked G-MUSIC
  - 3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
- 4.1. Introduction to Robust Estimation
- 4.1. Initial Results and Open Problems

Stieltjes transform inversion for covariance matrix models

J. W. Silverstein, S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.

- We know for the model $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ that, if $F^{\mathbf{T}_N} \Rightarrow F^T$, the Stieltjes transform of the e.s.d. of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$ satisfies $m_{\underline{\mathbf{B}}_N}(z) \xrightarrow{\text{a.s.}} m_{\underline{F}}(z)$, with

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

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

Stieltjes transform inversion for covariance matrix models

J. W. Silverstein, S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.

- We know for the model $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ that, if $F^{\mathbf{T}_N} \Rightarrow F^T$, the Stieltjes transform of the e.s.d. of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$ satisfies $m_{\underline{\mathbf{B}}_N}(z) \xrightarrow{\text{a.s.}} m_{\underline{F}}(z)$, with

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

  which is unique on the set $\{z \in \mathbb{C}^+, m_{\underline{F}}(z) \in \mathbb{C}^+\}$.
- This can be inverted into

$$z_{\underline{F}}(m) = -\frac{1}{m} - c\int \frac{t}{1+tm}dF^T(t)$$

  for $m \in \mathbb{C}^+$.

Stieltjes transform inversion and spectrum characterization

Remember that we can evaluate the spectrum density by taking a complex line close to $\mathbb{R}$ and evaluating $\Im[m_{\underline{F}}(z)]$ along this line. Now we can do better.

Stieltjes transform inversion and spectrum characterization

Remember that we can evaluate the spectrum density by taking a complex line close to $\mathbb{R}$ and evaluating $\Im[m_{\underline{F}}(z)]$ along this line. Now we can do better.

It is shown that

$$\lim_{\substack{z \to x \in \mathbb{R}^* \\ z \in \mathbb{C}^+}} m_{\underline{F}}(z) = m_0(x) \quad \text{exists.}$$

We also have,

- for x_0 inside the support, the density $\underline{f}(x)$ of $\underline{F}$ in x_0 is $\frac{1}{\pi}\Im[m_0]$ with m_0 the unique solution $m \in \mathbb{C}^+$ of
$$[z_{\underline{F}}(m) =] \ x_0 = -\frac{1}{m} - c \int \frac{t}{1+tm} dF^T(t)$$
- let $m_0 \in \mathbb{R}^*$ and $x_{\underline{F}}$ the equivalent to $z_{\underline{F}}$ on the real line. Then "x_0 outside the support of $\underline{F}$" is equivalent to "$x_{\underline{F}}'(m_{\underline{F}}(x_0)) > 0$, $m_{\underline{F}}(x_0) \neq 0$, $-1/m_{\underline{F}}(x_0)$ outside the support of F^T".

Stieltjes transform inversion and spectrum characterization

Remember that we can evaluate the spectrum density by taking a complex line close to $\mathbb{R}$ and evaluating $\Im[m_{\underline{F}}(z)]$ along this line. Now we can do better.

It is shown that

$$\lim_{\substack{z \to x \in \mathbb{R}^* \\ z \in \mathbb{C}^+}} m_{\underline{F}}(z) = m_0(x) \quad \text{exists.}$$

We also have,

- for x_0 inside the support, the density $\underline{f}(x)$ of $\underline{F}$ in x_0 is $\frac{1}{\pi}\Im[m_0]$ with m_0 the unique solution $m \in \mathbb{C}^+$ of

$$[z_{\underline{F}}(m) =] \; x_0 = -\frac{1}{m} - c\int \frac{t}{1+tm} dF^T(t)$$

- let $m_0 \in \mathbb{R}^*$ and $x_{\underline{F}}$ the equivalent to $z_{\underline{F}}$ on the real line. Then "x_0 outside the support of $\underline{F}$" is equivalent to "$x_{\underline{F}}'(m_{\underline{F}}(x_0)) > 0$, $m_{\underline{F}}(x_0) \neq 0$, $-1/m_{\underline{F}}(x_0)$ outside the support of F^T".

This provides another way to determine the support!. For $m \in (-\infty, 0)$, evaluate $x_{\underline{F}}(m)$. Whenever $x_{\underline{F}}$ decreases, the image is outside the support. The rest is inside.

Another way to determine the spectrum: spectrum to analyze

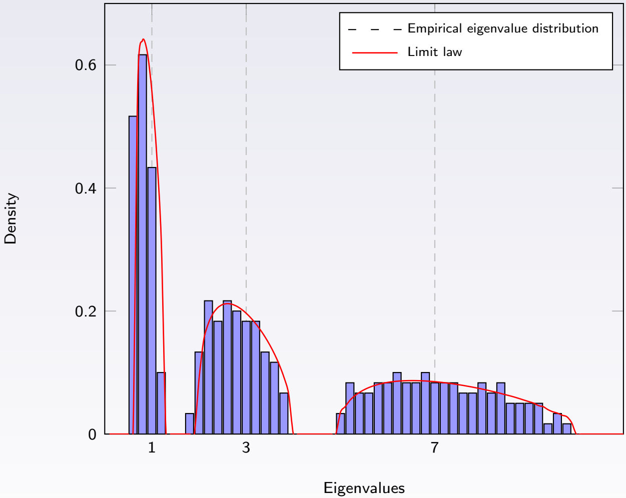


Figure : Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7.

Another way to determine the spectrum: inverse function method

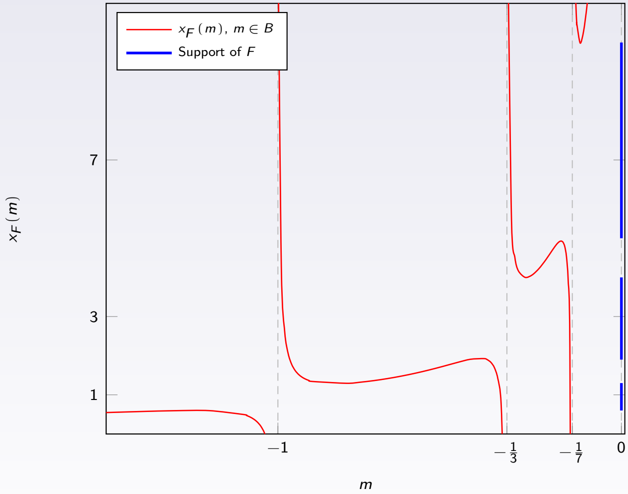


Figure : Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7. The support of F is read on the vertical axis, whenever m_F is decreasing.

Cluster boundaries in sample covariance matrix models

Xavier Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," IEEE Transactions on Information Theory, vol. 54, no. 11, Nov. 2008.

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ have i.i.d. entries of zero mean, variance $1/n$, and $\mathbf{T}_N$ be diagonal such that $F^{\mathbf{T}_N} \Rightarrow F^T$, as $n, N \to \infty$, $N/n \to c$, where F^T has K masses in $t_1, \ldots, t_K$ with multiplicity $n_1, \ldots, n_K$ respectively. Then the l.s.d. of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$ has support $\mathcal{S}$ given by

$$\mathcal{S} = [x_1^-, x_1^+] \cup [x_2^-, x_2^+] \cup \ldots \cup [x_Q^-, x_Q^+]$$

with $x_q^- = x_F(m_q^-)$, $x_q^+ = x_F(m_q^+)$, and

$$x_F(m) = -\frac{1}{m} - c\frac{1}{n} \sum_{k=1}^{K} n_k \frac{t_k}{1 + t_k m}$$

with $2Q$ the number of real-valued solutions counting multiplicities of $x_F'(m) = 0$ denoted in order $m_1^- < m_1^+ \leqslant m_2^- < m_2^+ \leqslant \ldots \leqslant m_Q^- < m_Q^+$.

Comments on spectrum characterization

Previous results allows to determine

- the spectrum boundaries
- the number Q of clusters
- as a consequence, the total separation or not of the spectrum in K clusters.

Comments on spectrum characterization

Previous results allows to determine

- the spectrum boundaries
- the number Q of clusters
- as a consequence, the total separation or not of the spectrum in K clusters.

Mestre goes further: to determine local separability of the spectrum,

- identify the K inflexion points, i.e. the K solutions $m_1, \dots, m_K$ to

$$x_F''(m) = 0$$

- check whether $x_F'(m_i) > 0$ and $x_F'(m_{i+1}) > 0$
- if so, the cluster in between corresponds to a single population eigenvalue.

Eigeninference: Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

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

is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.

Eigeninference: Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

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

  is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.

Eigeninference: Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

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

  is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.

Eigeninference: Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

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

  is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.
- If only the eigenvalues of $\mathbf{R}$ are of interest, things can be done. The process of retrieving information about eigenvalues, eigenspace projections, or functional of these is called eigen-inference.

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis,"
http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called G-estimators (Generalized estimators). Among those, we find
  - *G_1-estimator of generalized variance*. For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1}\left[\log\det(\mathbf{R}_n) + \log\frac{n(n-1)^N}{(n-N)\prod_{k=1}^N (n-k)}\right]$$

with α_n any sequence such that $\alpha_n^{-2}\log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1}\log\det(\mathbf{R}) \to 0$$

in probability.

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis,"
http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called G-estimators (Generalized estimators). Among those, we find
  - *G_1-estimator of generalized variance*. For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1}\left[\log\det(\mathbf{R}_n) + \log\frac{n(n-1)^N}{(n-N)\prod_{k=1}^N(n-k)}\right]$$

    with α_n any sequence such that $\alpha_n^{-2}\log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1}\log\det(\mathbf{R}) \to 0$$

    in probability.

- However, Girko's proofs are rarely readable, if existent.

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- It has long been thought the inverse problem of estimating $t_1, \ldots, t_K$ from the Stieltjes transform method was not possible.
- Only trials were iterative convex optimization methods.

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- It has long been thought the inverse problem of estimating $t_1, \ldots, t_K$ from the Stieltjes transform method was not possible.
- Only trials were iterative convex optimization methods.
- The problem was partially solved by Mestre in 2008!

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- It has long been thought the inverse problem of estimating $t_1, \ldots, t_K$ from the Stieltjes transform method was not possible.
- Only trials were iterative convex optimization methods.
- The problem was partially solved by Mestre in 2008!
- His technique uses elegant complex analysis tools. The description of this technique is the subject of this course.

Reminders

- Consider the sample covariance matrix model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$.
- Up to now, we saw:
  - that there is no eigenvalue outside the support with probability 1 for all large N.
  - that for all large N, when the spectrum is divided into clusters, the number of empirical eigenvalues in each cluster is exactly as we expect.

Reminders

- Consider the sample covariance matrix model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$.
- Up to now, we saw:
  - that there is no eigenvalue outside the support with probability 1 for all large N.
  - that for all large N, when the spectrum is divided into clusters, the number of empirical eigenvalues in each cluster is exactly as we expect.
- these results are of crucial importance for the following.

Inverse problem for sample covariance matrix

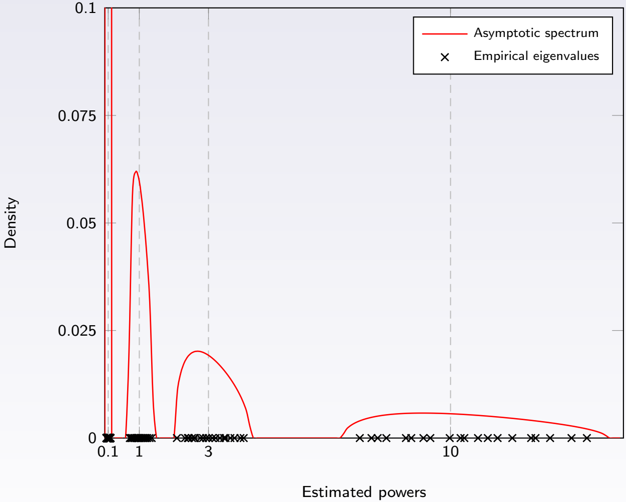


Figure : Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $N/n = 10$, $M/N = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

Eigen-inference for the sample covariance matrix model

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

Theorem

Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$, with $\mathbf{X}_N \in \mathbb{C}^{N\times n}$, i.i.d. with entries of zero mean, variance $1/n$, and $\mathbf{T}_N \in \mathbb{R}^{N\times N}$ is diagonal with K distinct entries $t_1, \ldots, t_K$ of multiplicity $N_1, \ldots, N_K$ of same order as n. Let $k \in \{1, \ldots, K\}$. Then, if the cluster associated to t_k is separated from the clusters associated to $k-1$ and $k+1$, as $N, n \to \infty$, $N/n \to c$,

$$\hat{t}_k = \frac{n}{N_k}\sum_{m\in\mathcal{N}_k}(\lambda_m - \mu_m)$$

is an N, n-consistent estimator of t_k, where $\mathcal{N}_k = \{N - \sum_{i=k}^{K} N_i + 1, \ldots, N - \sum_{i=k+1}^{K} N_i\}$, $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of $\mathbf{B}_N$ and $\mu_1, \ldots, \mu_N$ are the N solutions of

$$m_{\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N}(\mu) = 0$$

or equivalently, $\mu_1, \ldots, \mu_N$ are the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$.

Remarks on Mestre's result

Assuming cluster separation, the result consists in

- taking the empirical *ordered* λ_i's inside the cluster (note that exact separation ensures there are N_k of these!)
- getting the *ordered* eigenvalues $\mu_1, \ldots, \mu_N$ of

$$\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$$

  with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$. Keep only those of index inside $\mathcal{N}_k$.
- take the difference and scale.

How to obtain this result?

- Major trick requires tools from complex analysis

How to obtain this result?

- Major trick requires tools from complex analysis
- Silverstein's Stieltjes transform identity: for the *conjugate* model $\underline{\mathbf{B}}_N = \mathbf{X}_N^* \mathbf{T}_N \mathbf{X}_N$,

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

with $\underline{m}_N$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}$. This is the only random matrix result we need.

How to obtain this result?

- Major trick requires tools from complex analysis
- Silverstein's Stieltjes transform identity: for the *conjugate* model $\underline{\mathbf{B}}_N = \mathbf{X}_N^* \mathbf{T}_N \mathbf{X}_N$,

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

  with $\underline{m}_N$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}$. This is the only random matrix result we need.
- Before going further, we need some reminders from complex analysis.

Reminders of complex analysis

- Cauchy integration formula

Theorem

Let $U \subset \mathbb{C}$ be an open set and $f : U \to \mathbb{C}$ be holomorphic on U. Let $\gamma \subset U$ be a continuous contour (i.e. closed path). Then, for a **inside** *the surface formed by γ, we have*

$$\frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z-a} dz = f(a)$$

while for a **outside** *the surface formed by γ,*

$$\frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z-a} dz = 0.$$

Limiting spectrum of the sample covariance matrix

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

Reminder:

- If $F^{\mathbf{T}_N} \Rightarrow F^T$, then $m_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} m_F(z)$ such that

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

Limiting spectrum of the sample covariance matrix

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

Reminder:

- If $F^{\mathbf{T}_N} \Rightarrow F^T$, then $m_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} m_F(z)$ such that

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

or equivalently

$$m_{F^T}\left(-1/m_{\underline{F}}(z)\right) = -z m_{\underline{F}}(z) m_F(z)$$

with $m_{\underline{F}}(z) = c m_F(z) + (c-1)\frac{1}{z}$ and $N/n \to c$.

Reminders of complex analysis (2)

- Residue calculus

Theorem

Let γ be a contour on $\mathbb{C}$. For f holomorphic inside γ but on a discrete number of points, to compute the expression

$$\frac{1}{2\pi i}\oint_\gamma f(z)dz$$

one must

1. *determine the poles of f lying inside the surface formed by γ, i.e. those values a such that*

$$\lim_{z\to a}|f(z)|=\infty$$

2. *determine the order of each pole, i.e. the smallest k such that*

$$\lim_{z\to a}|(z-a)^k f(z)|<\infty$$

3. *compute the residues of f at the poles, i.e. evaluate the value*

$$\mathrm{Res}(f,a)\triangleq\lim_{z\to a}\frac{d^{k-1}}{dz^{k-1}}\left[(z-a)^k f(z)\right]$$

4. *the integral is then the sum of all residues.*

$$\frac{1}{2\pi i}\oint_\gamma f(z)dz=\sum_{a\in\{\text{ poles of } f\}}\mathrm{Res}(f,a)$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

- After the variable change $\omega = -1/m_{\underline{F}}(z)$,

$$t_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_F(z) \frac{m'_{\underline{F}}(z)}{m^2_{\underline{F}}(z)} dz,$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

- After the variable change $\omega = -1/m_{\underline{F}}(z)$,

$$t_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_F(z) \frac{m'_{\underline{F}}(z)}{m^2_{\underline{F}}(z)} dz,$$

- When the system dimensions are large,

$$m_F(z) \simeq m_{\mathbf{B}_N}(z) \triangleq \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\lambda_k - z}, \quad \text{with} \quad (\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{B}_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}).$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

- After the variable change $\omega = -1/m_{\underline{F}}(z)$,

$$t_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_F(z) \frac{m_{\underline{F}}'(z)}{m_{\underline{F}}^2(z)} dz,$$

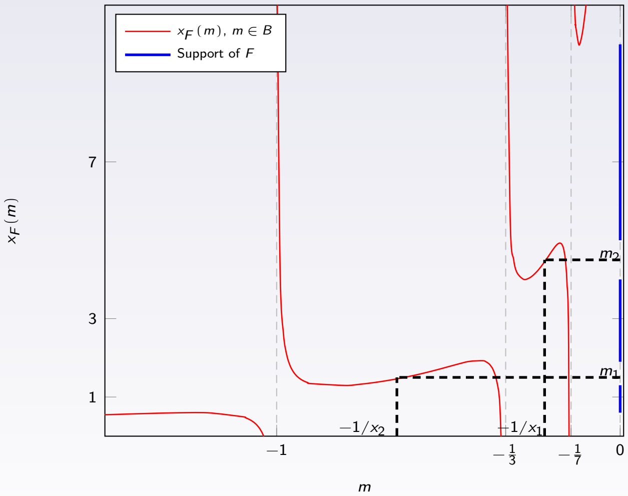
- When the system dimensions are large,

$$m_F(z) \simeq m_{\mathbf{B}_N}(z) \triangleq \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k - z}, \quad \text{with} \quad (\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{B}_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}).$$

- Dominated convergence arguments then show

$$t_k - \hat{t}_k \xrightarrow{\text{a.s.}} 0 \quad \text{with} \quad \hat{t}_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_{\mathbf{B}_N}(z) \frac{m_{\underline{\mathbf{B}}_N}'(z)}{m_{\underline{\mathbf{B}}_N}^2(z)} dz$$

Understanding the contour change



- **IF** $\mathcal{C}_{\underline{F},k}$ encloses cluster k with real points $m_1 < m_2$
- **THEN** $-1/m_1 = x_1 < t_k < x_2 = -1/m_2$ and $\mathcal{C}_k$ encloses t_k.

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.
- remember that

$$m_{\mathbf{B}_N}(w) = \frac{n}{N} m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\frac{1}{w}$$

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.
- remember that

$$m_{\mathbf{B}_N}(w) = \frac{n}{N} m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N} \frac{1}{w}$$

- residue calculus, denote $f(w) = \left(\frac{n}{N} w m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N} \right) \frac{m'_{\underline{\mathbf{B}}_N}(w)}{m_{\underline{\mathbf{B}}_N}(w)^2}$,
  - the λ_k's are poles of order 1 and

$$\lim_{z \to \lambda_k} (z - \lambda_k) f(z) = -\frac{n}{N} \lambda_k$$

  - the μ_k's are also poles of order 1 and by L'Hospital's rule

$$\lim_{z \to \mu_k} (z - \lambda_k) f(z) = \lim_{z \to \mu_k} \frac{n}{N} \frac{(z - \mu_k) z m'_{\underline{\mathbf{B}}_N}(z)}{m_{\underline{\mathbf{B}}_N}(z)} = \frac{n}{N} \mu_k$$

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.
- remember that

$$m_{\mathbf{B}_N}(w) = \frac{n}{N} m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\frac{1}{w}$$

- residue calculus, denote $f(w) = \left(\frac{n}{N} w m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\right) \frac{m'_{\underline{\mathbf{B}}_N}(w)}{m_{\underline{\mathbf{B}}_N}(w)^2}$,
  - the λ_k's are poles of order 1 and

$$\lim_{z\to\lambda_k}(z-\lambda_k)f(z) = -\frac{n}{N}\lambda_k$$

  - the μ_k's are also poles of order 1 and by L'Hospital's rule

$$\lim_{z\to\mu_k}(z-\lambda_k)f(z) = \lim_{z\to\mu_k}\frac{n}{N}\frac{(z-\mu_k)zm'_{\underline{\mathbf{B}}_N}(z)}{m_{\underline{\mathbf{B}}_N}(z)} = \frac{n}{N}\mu_k$$

  - So, finally

$$\hat{t}_k = \frac{n}{N_k}\sum_{m\in\text{contour}}(\lambda_m - \mu_m)$$

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.
- Since the μ_i are rank-1 perturbations of the λ_i, they have the interleaving property

$$\lambda_1 < \mu_2 < \lambda_2 < \ldots < \mu_N < \lambda_N$$

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.
- Since the μ_i are rank-1 perturbations of the λ_i, they have the interleaving property

$$\lambda_1 < \mu_2 < \lambda_2 < \ldots < \mu_N < \lambda_N$$

- what about μ_1? the trick is to use the fact that

$$\frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{z} dz = 0$$

which leads to

$$\frac{1}{2\pi i} \oint_{\partial \Gamma_k} \frac{m'_{\underline{F}}(w)}{m_{\underline{F}}(w)^2} dw = 0$$

the empirical version of which is

$$\#\{i : \lambda_i \in \Gamma_k\} - \#\{i : \mu_i \in \Gamma_k\}$$

Since their difference tends to 0, there are as many λ_k's as μ_k's in the contour, hence μ_1 is asymptotically in the integration contour.

Related bibliography

- C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," Communications in Mathematical Physics, vol. 177, no. 3, pp. 727-754, 1996.
- G. W. Anderson, A. Guionnet, O. Zeitouni, "An introduction to random matrices", Cambridge studies in advanced mathematics, vol. 118, 2010.
- F. Bornemann, "On the numerical evaluation of distributions in random matrix theory: A review," Markov Process. Relat. Fields, vol. 16, pp. 803-866, 2010.
- Y. Q. Yin, Z. D. Bai, P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," Probability Theory and Related Fields, vol. 78, no. 4, pp. 509-521, 1988.
- J. W. Silverstein, Z.D. Bai and Y.Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," Journal of Multivariate Analysis, vol. 26, no. 2, pp. 166-168. 1988.
- C. A. Tracy, H. Widom, "On orthogonal and symplectic matrix ensembles," Communications in Mathematical Physics, vol. 177, no. 3, pp. 727-754, 1996.
- Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," The Annals of Probability, vol. 26, no.1 pp. 316-345, 1998.
- Z. D. Bai, J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.
- J. W. Silverstein, P. Debashis, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.
- J. W. Silverstein, J. Baik, "Eigenvalues of large sample covariance matrices of spiked population models" Journal of Multivariate Analysis, vol. 97, no. 6, pp. 1382-1408, 2006.
- I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," Annals of Statistics, vol. 99, no. 2, pp. 295-327, 2001.
- K. Johansson, "Shape Fluctuations and Random Matrices," Comm. Math. Phys. vol. 209, pp. 437-476, 2000.
- J. Baik, G. Ben Arous, S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," The Annals of Probability, vol. 33, no. 5, pp. 1643-1697, 2005.

Related bibliography (2)

- J. W. Silverstein, S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.
- W. Hachem, P. Loubaton, X. Mestre, J. Najim, P. Vallet, "A Subspace Estimator for Fixed Rank Perturbations of Large Random Matrices," arxiv preprint 1106.1497, 2011.
- R. Couillet, W. Hachem, "Local failure detection and diagnosis in large sensor networks", (submitted to) IEEE Transactions on Information Theory, arXiv preprint 1107.1409.
- F. Benaych-Georges, R. Rao, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," Advances in Mathematics, vol. 227, no. 1, pp. 494-521, 2011.
- X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," IEEE Transactions on Signal Processing, vol. 56, no.11, 2008.
- X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.
- R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources", IEEE Transactions on Information Theory, vol. 57, no. 4, pp. 2420-2439, 2011.
- P. Vallet, P. Loubaton and X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case," arxiv preprint 1002.3234, 2010.

Outline

Part 1: Fundamentals of Random Matrix Theory
- 1.1. The Stieltjes Transform Method
- 1.2. Extreme Eigenvalues
- 1.3. The Spiked Model
- 1.4. Spectrum Analysis and G-estimation

2. Source Detection
- 2.1. Eigenvalue-based Detection
- 2.2. Detection in unknown Noise Environment

3. Statistical Inference
- 3.1. Generic Model
  - 3.1.2. Angle-of-arrival estimation
  - 3.1.2. Angle-of-arrival estimation
- 3.2. Spiked Model
  - 3.2.1. Spiked G-MUSIC
  - 3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
- 4.1. Introduction to Robust Estimation
- 4.1. Initial Results and Open Problems

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Problem formulation

- We want to test the hypothesis $\mathcal{H}_0$ against $\mathcal{H}_1$,

$$\mathbb{C}^{N\times n} \ni \mathbf{Y} = \begin{cases} \mathbf{h}\mathbf{x}^T + \sigma\mathbf{W} & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \sigma\mathbf{W} & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

with $\mathbf{h} \in \mathbb{C}^N$, $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{W} \in \mathbb{C}^{N\times n}$.

Problem formulation

- We want to test the hypothesis $\mathcal{H}_0$ against $\mathcal{H}_1$,

$$\mathbb{C}^{N\times n} \ni \mathbf{Y} = \begin{cases} \mathbf{h}\mathbf{x}^T + \sigma\mathbf{W} & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \sigma\mathbf{W} & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

  with $\mathbf{h} \in \mathbb{C}^N$, $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{W} \in \mathbb{C}^{N\times n}$.
- We assume no knowledge whatsoever but that **W** has i.i.d. (non-necessarily Gaussian) entries.

Exploiting the conditioning number

L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative spectrum sensing using random matrix theory," International Symposium on Wireless Pervasive Computing, pp. 334-338 , 2008.

- under either hypothesis,
  - if $\mathcal{H}_0$, for N large, we expect $F_{\mathbf{YY}^\mathrm{H}}$ close to the Marčenko-Pastur law, of support $[\sigma^2\left(1-\sqrt{c}\right)^2, \sigma^2\left(1+\sqrt{c}\right)^2]$.
  - if $\mathcal{H}_1$, if population spike more than $1+\sqrt{\frac{N}{n}}$, largest eigenvalue is further away.
- the conditioning number of $\mathbf{YY}^\mathrm{H}$ is therefore asymptotically, as $N, n \to \infty$, $N/n \to c$,
  - if $\mathcal{H}_0$,

$$\mathrm{cond}(\mathbf{Y}) \triangleq \frac{\lambda_{\max}}{\lambda_{\min}} \to \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  - if $\mathcal{H}_1$,

$$\mathrm{cond}(\mathbf{Y}) \to t_1 + \frac{ct_1}{t_1-1} > \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  with $t_1 = \sum_{k=1}^{N} |h_k|^2 + \sigma^2$

Exploiting the conditioning number

L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative spectrum sensing using random matrix theory," International Symposium on Wireless Pervasive Computing, pp. 334-338 , 2008.

- under either hypothesis,
  - if $\mathcal{H}_0$, for N large, we expect $F_{\mathbf{YY}^\mathsf{H}}$ close to the Marčenko-Pastur law, of support $[\sigma^2\left(1-\sqrt{c}\right)^2, \sigma^2\left(1+\sqrt{c}\right)^2]$.
  - if $\mathcal{H}_1$, if population spike more than $1+\sqrt{\frac{N}{n}}$, largest eigenvalue is further away.
- the conditioning number of $\mathbf{YY}^\mathsf{H}$ is therefore asymptotically, as $N, n \to \infty$, $N/n \to c$,
  - if $\mathcal{H}_0$,

$$\mathrm{cond}(\mathbf{Y}) \triangleq \frac{\lambda_{\max}}{\lambda_{\min}} \to \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  - if $\mathcal{H}_1$,

$$\mathrm{cond}(\mathbf{Y}) \to t_1 + \frac{ct_1}{t_1-1} > \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  with $t_1 = \sum_{k=1}^{N} |h_k|^2 + \sigma^2$
- the conditioning number is independent of σ. We then have the decision criterion, whether or not σ is known,

$$\text{decide} \begin{cases} \mathcal{H}_0 : & \text{if } \mathrm{cond}(\mathbf{YY}^\mathsf{H}) \leqslant \frac{\left(1-\sqrt{\frac{N}{n}}\right)^2}{\left(1+\sqrt{\frac{N}{n}}\right)^2} + \varepsilon \\ \mathcal{H}_1 : & \text{otherwise.} \end{cases}$$

for some security margin ε.

Comments on the method

- Advantages:
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known

Comments on the method

- Advantages:
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known
- Drawbacks:
  - only stands for very large N (dimension N for which asymptotic results arise function of σ!)

Comments on the method

- Advantages:
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known
- Drawbacks:
  - only stands for very large N (dimension N for which asymptotic results arise function of σ!)
  - *ad-hoc* method, does not rely on performance criterion.

Generalized likelihood ratio test

P. Bianchi, M. Debbah, M. Maida, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, 2011.

- Alternative generalized likelihood ratio test (GLRT) decision criterion, i.e.

$$C(\mathbf{Y}) = \frac{\sup_{\sigma^2,\mathbf{h}} P_{\mathbf{Y}|\mathbf{h},\sigma^2}(\mathbf{Y},\mathbf{h},\sigma^2)}{\sup_{\sigma^2} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y}|\sigma^2)}.$$

- Denote

$$T_N = \frac{\lambda_{\max}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\frac{1}{N}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

To guarantee a maximum false alarm ratio of α,

$$\text{decide} \begin{cases} \mathcal{H}_1 : & \text{if } \left(1-\frac{1}{N}\right)^{(1-N)n} T_N^{-n}\left(1-\frac{\mathbf{T}_N}{N}\right)^{(1-N)n} > \xi_N \\ \mathcal{H}_0 : & \text{otherwise.} \end{cases}$$

for some threshold ξ_N that can be explicitly given as a function of α.

Generalized likelihood ratio test

P. Bianchi, M. Debbah, M. Maida, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, 2011.

- Alternative generalized likelihood ratio test (GLRT) decision criterion, i.e.

$$C(\mathbf{Y}) = \frac{\sup_{\sigma^2,\mathbf{h}} P_{\mathbf{Y}|\mathbf{h},\sigma^2}(\mathbf{Y},\mathbf{h},\sigma^2)}{\sup_{\sigma^2} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y}|\sigma^2)}.$$

- Denote

$$T_N = \frac{\lambda_{\max}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\frac{1}{N}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

  To guarantee a maximum false alarm ratio of α,

$$\text{decide} \begin{cases} \mathcal{H}_1 : & \text{if } \left(1-\frac{1}{N}\right)^{(1-N)n} T_N^{-n} \left(1-\frac{\mathbf{T}_N}{N}\right)^{(1-N)n} > \xi_N \\ \mathcal{H}_0 : & \text{otherwise.} \end{cases}$$

  for some threshold ξ_N that can be explicitly given as a function of α.
- Optimal test with respect to GLR.
- Performs better than conditioning number test.

Performance comparison for unknown σ^2, P

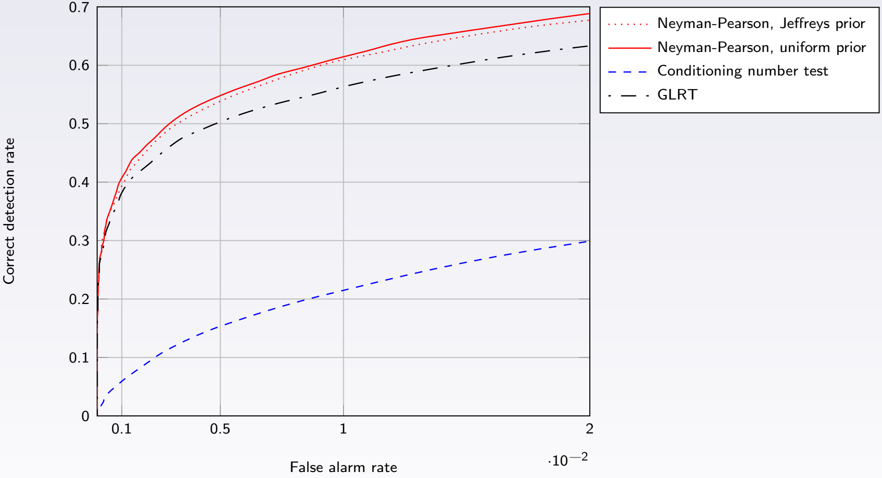


Figure : ROC curve for *a priori* unknown σ^2 of the Neyman-Pearson test, conditioning number method and GLRT, $K = 1$, $N = 4$, $M = 8$, $\text{SNR} = 0 \text{ dB}$. For the Neyman-Pearson test, both uniform and Jeffreys prior, with exponent $\beta = 1$, are provided.

Related biography

- R. Couillet, M. Debbah, “A Bayesian Framework for Collaborative Multi-Source Signal Sensing”, IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5186-5195, 2010.
- T. Ratnarajah, R. Vaillancourt, M. Alvo, “Eigenvalues and condition numbers of complex random matrices,” SIAM Journal on Matrix Analysis and Applications, vol. 26, no. 2, pp. 441-456, 2005.
- M. Matthaiou, M. R. McKay, P. J. Smith, J. A. Mossek, “On the condition number distribution of complex Wishart matrices,” IEEE Transactions on Communications, vol. 58, no. 6, pp. 1705-1717, 2010.
- C. Zhong, M. R. McKay, T. Ratnarajah, K. Wong, “Distribution of the Demmel condition number of Wishart matrices,” IEEE Trans. on Communications, vol. 59, no. 5, pp. 1309-1320, 2011.
- L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, “Cooperative spectrum sensing using random matrix theory,” International Symposium on Wireless Pervasive Computing, pp. 334-338 , 2008.
- P. Bianchi, M. Debbah, M. Maida, J. Najim, “Performance of Statistical Tests for Source Detection using Random Matrix Theory,” IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, 2011.

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Generic inference scenario

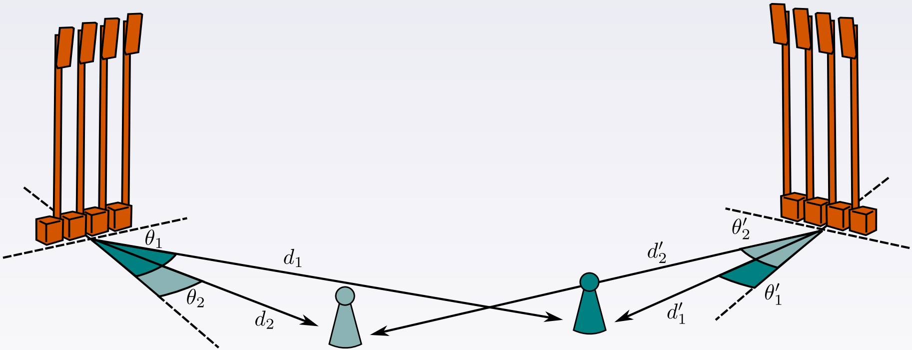


Figure : Signal sensing and angle of arrival detection

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Power estimation: problem Statement

- Consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{x}_k^{(m)} + \sigma \mathbf{w}^{(m)}$$

and wish to infer $P_1, \ldots, P_K$.

Power estimation: problem Statement

- Consider the model
$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}$$
and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).

Power estimation: problem Statement

- Consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{x}_k^{(m)} + \sigma \mathbf{w}^{(m)}$$

  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{X}_k + \sigma \mathbf{W}$$

Power estimation: problem Statement

- Consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{x}_k^{(m)} + \sigma \mathbf{w}^{(m)}$$

  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{X}_k + \sigma \mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma \mathbf{W}$$

Power estimation: problem Statement

- Consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{x}_k^{(m)} + \sigma \mathbf{w}^{(m)}$$

and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{X}_k + \sigma \mathbf{W} = \underbrace{\begin{bmatrix} \sqrt{P_1}\mathbf{H}_1 & \cdots & \sqrt{P_K}\mathbf{H}_K \end{bmatrix}}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma \mathbf{W} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}.$$

Power estimation: problem Statement

- Consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}$$

  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{X}_k + \sigma\mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix}\mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K\end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma\mathbf{W} = \left[\mathbf{H}\mathbf{P}^{\frac{1}{2}} \quad \sigma\mathbf{I}_N\right]\begin{bmatrix}\mathbf{X} \\ \mathbf{W}\end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^\mathsf{T}\ \mathbf{W}^\mathsf{T})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

Power estimation: problem Statement

- Consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}$$

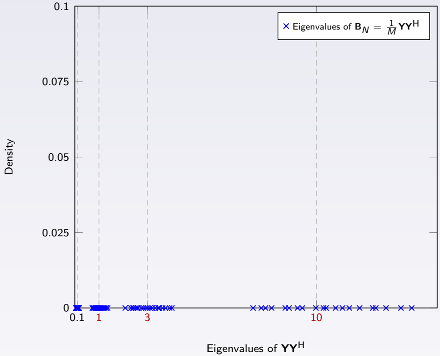
  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{X}_k + \sigma\mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix}\mathbf{X}_1\\ \vdots \\ \mathbf{X}_K\end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma\mathbf{W} = \left[\mathbf{H}\mathbf{P}^{\frac{1}{2}} \quad \sigma\mathbf{I}_N\right]\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^{\mathsf{T}} \ \mathbf{W}^{\mathsf{T}})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

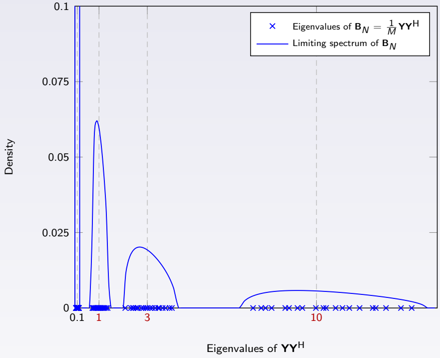
Most information about $P_1, \ldots, P_K$ is contained in the eigenvalues of $\mathbf{B}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

From small to large system analysis



Classical approach requires to assume $M >> N$ as well as $N >> n_k$ for each k!

From small to large system analysis



Assuming dimensions N, n_k, M grow large, large dimensional random matrix theory provides

- a link between:
  - **the “observation”:** the limiting spectral distribution (l.s.d.) of $\mathbf{B}_N$;
  - **the “hidden parameters”:** the powers $P_1, \ldots, P_K$, i.e. the l.s.d. of $\mathbf{P}$.
- consistent estimators of the hidden parameters.

Power estimation with random matrices

- **Reminder:** Method consists in:

Power estimation with random matrices

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.

Power estimation with random matrices

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.

Power estimation with random matrices

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.
  - **Step 3:** Using $m_{\mathbf{B}_N}$ as an approximation of m_F, residue calculus provides estimator.

Power estimation with random matrices

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.
  - **Step 3:** Using $m_{\mathbf{B}_N}$ as an approximation of m_F, residue calculus provides estimator.
- Extending $\mathbf{Y}$ with zeros, our model is a "double sample covariance matrix"

$$\underbrace{\underline{\mathbf{Y}}}_{(N+n)\times M} = \underbrace{\begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \\ 0 & 0 \end{bmatrix}}_{(N+n)\times(N+n)} \underbrace{\begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}}_{(N+n)\times M} .$$

Power estimation with random matrices

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.
  - **Step 3:** Using $m_{\mathbf{B}_N}$ as an approximation of m_F, residue calculus provides estimator.
- Extending $\mathbf{Y}$ with zeros, our model is a "double sample covariance matrix"

$$\underbrace{\mathbf{Y}}_{(N+n)\times M} = \underbrace{\begin{bmatrix} \mathbf{HP}^{\frac{1}{2}} & \sigma \mathbf{I}_N \\ 0 & 0 \end{bmatrix}}_{(N+n)\times(N+n)} \underbrace{\begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}}_{(N+n)\times M}.$$

- Limiting distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$

Theorem (Spectral analysis of $\mathbf{B}_N$)

Let $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ with eigenvalues $\lambda_1, \ldots, \lambda_N$. Denote $m_{\underline{\mathbf{B}}_N}(z) \triangleq \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\lambda_k - z}$, with $\lambda_i = 0$ for $i > N$. Then, for $M/N \to c$, $N/n_k \to c_k$, $N/n \to c_0$, for any $z \in \mathbb{C}^+$,

$$m_{\underline{\mathbf{B}}_N}(z) \xrightarrow{\text{a.s.}} m_{\underline{F}}(z)$$

with $m_{\underline{F}}(z)$ the unique solution in $\mathbb{C}^+$ of

$$\frac{1}{m_{\underline{F}}(z)} = -\sigma^2 + \frac{1}{f(z)}\left[\frac{c_0 - 1}{c_0} + m_P\left(-\frac{1}{f(z)}\right)\right], \quad \text{with} \quad f(z) = (c-1)m_{\underline{F}}(z) - czm_{\underline{F}}(z)^2.$$

Stieltjes transform method (2)

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," *to appear in* IEEE Trans. on Inf. Theory, 2010.

- estimator calculus

Theorem (Estimator of $P_1, \ldots, P_K$)

Let $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ be defined as in Theorem 19, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, $\lambda_1 < \ldots < \lambda_N$. Assume that asymptotic cluster separability condition is fulfilled for some k. Then, as N, n, $M \to \infty$,

$$\hat{P}_k - P_k \xrightarrow{\text{a.s.}} 0,$$

where

$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{i\in\mathcal{N}_k} (\eta_i - \mu_i)$$

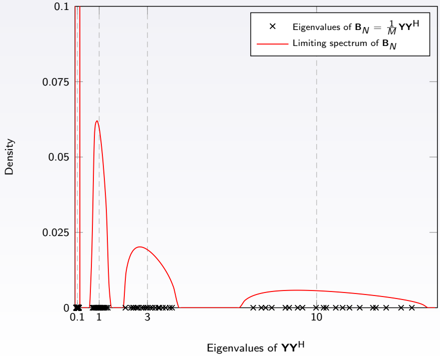
with $\mathcal{N}_k$ the set indexing the eigenvalues in cluster k of F, $\eta_1 < \ldots < \eta_N$ the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ and $\mu_1 < \ldots < \mu_N$ the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$.

Remarks

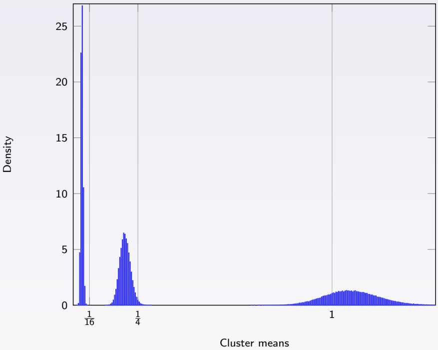
- solution is computationally simple, explicit, and the final formula compact.

Remarks

- solution is computationally simple, explicit, and the final formula compact.
- cluster separability condition is fundamental. This requires
  - for all other parameters fixed, the P_k cannot be too close top one another: source separation problem.
  - for all other parameters fixed, σ^2 must be kept low: low SNR undecidability problem.
  - for all other parameters fixed, M/N cannot be too low: sample deficiency issue (not such an issue though).
  - for all other parameters fixed, N/n cannot be too low: diversity issue.
- exact spectrum separability is an essential ingredient (known for very few models to this day).



Simulations



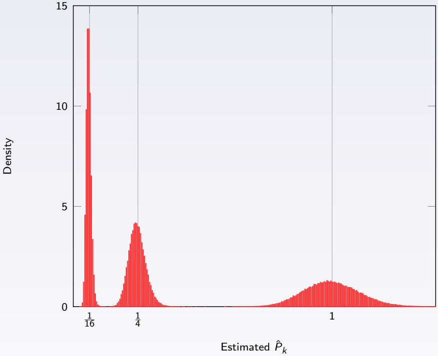


Figure : Histogram of the cluster-mean approach and of $\hat{P}_k$ for $k \in \{1,2,3\}$, $P_1 = 1/16$, $P_2 = 1/4$, $P_3 = 1$, $n_1 = n_2 = n_3 = 4$ antennas per user, $N = 24$ sensors, $M = 128$ samples and $\mathrm{SNR} = 20$ dB.

Performance comparison

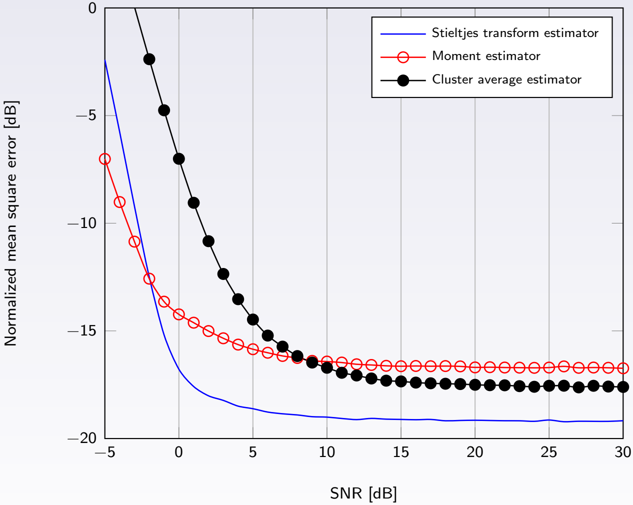


Figure : Normalized mean square error of largest estimated power $\hat{P}_3$, $P_1 = 1/16, P_2 = 1/4, P_3 = 1$, $n_1 = n_2 = n_3 = 4$,$N = 24$, $M = 128$. Comparison between classical, moment and Stieltjes transform approaches.

Outline

Part 1: Fundamentals of Random Matrix Theory
- 1.1. The Stieltjes Transform Method
- 1.2. Extreme Eigenvalues
- 1.3. The Spiked Model
- 1.4. Spectrum Analysis and G-estimation

2. Source Detection
- 2.1. Eigenvalue-based Detection
- 2.2. Detection in unknown Noise Environment

3. Statistical Inference
- 3.1. Generic Model
  - 3.1.2. Angle-of-arrival estimation
  - 3.1.2. Angle-of-arrival estimation
- 3.2. Spiked Model
  - 3.2.1. Spiked G-MUSIC
  - 3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
- 4.1. Introduction to Robust Estimation
- 4.1. Initial Results and Open Problems

Direction-of-arrival estimation: Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.

Direction-of-arrival estimation: Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.
- Received signal $\mathbf{y}^{(t)} \in \mathbb{C}^N$ at time t

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{s}(\theta_k) x_k^{(t)} + \sigma \mathbf{w}^{(t)}$$

with $E[s_k] = 0$, $E[|x_k|^2] = P_k$.

Direction-of-arrival estimation: Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.
- Received signal $\mathbf{y}^{(t)} \in \mathbb{C}^N$ at time t

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{s}(\theta_k) x_k^{(t)} + \sigma \mathbf{w}^{(t)}$$

with $E[s_k] = 0$, $E[|x_k|^2] = P_k$.

- Therefore

$$E[\mathbf{y}^{(t)} \mathbf{y}^{(y)\mathrm{H}}] \triangleq \mathbf{R} = \mathbf{S}(\Theta)\mathbf{P}\mathbf{S}(\Theta)^{\mathrm{H}} + \sigma^2 \mathbf{I}_N$$

where $\mathbf{S}(\Theta) = [\mathbf{s}(\theta_1), \ldots, \mathbf{s}(\theta_K)] \in \mathbb{C}^{N \times K}$, $\mathbf{P} = \operatorname{diag}(P_1, \ldots, P_K)$.

Direction-of-arrival estimation: Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.
- Received signal $\mathbf{y}^{(t)} \in \mathbb{C}^N$ at time t

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{s}(\theta_k) x_k^{(t)} + \sigma \mathbf{w}^{(t)}$$

with $E[s_k] = 0$, $E[|x_k|^2] = P_k$.

- Therefore

$$E[\mathbf{y}^{(t)} \mathbf{y}^{(y)\mathrm{H}}] \triangleq \mathbf{R} = \mathbf{S}(\Theta)\mathbf{P}\mathbf{S}(\Theta)^{\mathrm{H}} + \sigma^2 \mathbf{I}_N$$

where $\mathbf{S}(\Theta) = [\mathbf{s}(\theta_1), \ldots, \mathbf{s}(\theta_K)] \in \mathbb{C}^{N \times K}$, $\mathbf{P} = \mathrm{diag}(P_1, \ldots, P_K)$.

- **Objective:** Based on $\mathbf{Y} \triangleq [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, estimate $\theta_1, \ldots, \theta_K$,

MUSIC method

- Write

$$\mathbf{R} = \begin{pmatrix} \mathbf{E}_W & \mathbf{E}_S \end{pmatrix} \begin{pmatrix} \sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S \end{pmatrix} \begin{pmatrix} \mathbf{E}_W^H \\ \mathbf{E}_S^H \end{pmatrix}$$

with $\mathbf{L}_S = \mathrm{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.

MUSIC method

- Write

$$\mathbf{R} = \begin{pmatrix} \mathbf{E}_W & \mathbf{E}_S \end{pmatrix} \begin{pmatrix} \sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S \end{pmatrix} \begin{pmatrix} \mathbf{E}_W^H \\ \mathbf{E}_S^H \end{pmatrix}$$

  with $\mathbf{L}_S = \text{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.

- By definition,

$$\eta(\theta_k) \triangleq \mathbf{s}(\theta_k)^H \mathbf{E}_W \mathbf{E}_W^H \mathbf{s}(\theta_k) = 0$$

MUSIC method

- Write

$$\mathbf{R} = \begin{pmatrix}\mathbf{E}_W & \mathbf{E}_S\end{pmatrix}\begin{pmatrix}\sigma^2\mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S\end{pmatrix}\begin{pmatrix}\mathbf{E}_W^{\mathrm{H}} \\ \mathbf{E}_S^{\mathrm{H}}\end{pmatrix}$$

  with $\mathbf{L}_S = \operatorname{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.
- By definition,

$$\eta(\theta_k) \triangleq \mathbf{s}(\theta_k)^{\mathrm{H}}\mathbf{E}_W\mathbf{E}_W^{\mathrm{H}}\mathbf{s}(\theta_k) = 0$$

- MUSIC algorithm consists in finding θ such that

$$\hat{\eta}(\theta) \triangleq \mathbf{s}(\theta)^{\mathrm{H}}\hat{\mathbf{E}}_W\hat{\mathbf{E}}_W^{\mathrm{H}}\mathbf{s}(\theta).$$

  reaches a local minimum, with $\hat{\mathbf{E}}_W = [\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_{N-K}] \in \mathbb{C}^{N\times(N-K)}$ the subspace spanned by the $N-K$ smallest eigenvalues of

$$\mathbf{R}_N = \frac{1}{M}\sum_{t=1}^{M}\mathbf{y}^{(t)}\mathbf{y}^{(t)\mathrm{H}}.$$

MUSIC method

- Write
$$\mathbf{R} = \begin{pmatrix}\mathbf{E}_W & \mathbf{E}_S\end{pmatrix} \begin{pmatrix}\sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S\end{pmatrix} \begin{pmatrix}\mathbf{E}_W^{\mathsf{H}} \\ \mathbf{E}_S^{\mathsf{H}}\end{pmatrix}$$
with $\mathbf{L}_S = \mathrm{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.
- By definition,
$$\eta(\theta_k) \triangleq \mathbf{s}(\theta_k)^{\mathsf{H}} \mathbf{E}_W \mathbf{E}_W^{\mathsf{H}} \mathbf{s}(\theta_k) = 0$$
- MUSIC algorithm consists in finding θ such that
$$\hat{\eta}(\theta) \triangleq \mathbf{s}(\theta)^{\mathsf{H}} \hat{\mathbf{E}}_W \hat{\mathbf{E}}_W^{\mathsf{H}} \mathbf{s}(\theta).$$
reaches a local minimum, with $\hat{\mathbf{E}}_W = [\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_{N-K}] \in \mathbb{C}^{N \times (N-K)}$ the subspace spanned by the $N-K$ smallest eigenvalues of
$$\mathbf{R}_N = \frac{1}{M} \sum_{t=1}^{M} \mathbf{y}^{(t)} \mathbf{y}^{(t)\mathsf{H}}.$$

Only *M*-consistent!

RMT will provide an (N, M)-consistent procedure.

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.
- Starting point: Cauchy integration formula

$$\mathbf{s}(\theta_k)^{\mathsf{H}}\mathbf{E}_W\mathbf{E}_W^{\mathsf{H}}\mathbf{s}(\theta_k) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathbf{s}(\theta_k)\left(\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{s}(\theta_k)dz$$

with $\mathcal{C}$ circling around σ^2 only (only one pole in $z = \sigma^2$).

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.
- Starting point: Cauchy integration formula

$$\mathbf{s}(\theta_k)^{\mathsf{H}}\mathbf{E}_W\mathbf{E}_W^{\mathsf{H}}\mathbf{s}(\theta_k) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathbf{s}(\theta_k)\left(\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{s}(\theta_k)dz$$

with $\mathcal{C}$ circling around σ^2 only (only one pole in $z = \sigma^2$).

- We then use the result:

Lemma

For $\mathbf{a} \in \mathbb{C}^N$ *deterministic bounded, independent of* $\mathbf{R}_N$,

$$\mathbf{a}^{\mathsf{H}}\left(\mathbf{R}_N - z\mathbf{I}_N\right)^{-1}\mathbf{a} - \mathbf{a}^{\mathsf{H}}\left(\frac{1}{1 + ce_N(z)}\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{a} \xrightarrow{\text{a.s.}} 0$$

with $e_N(z)$ *solution to*

$$e = \int \frac{t}{\frac{t}{1+ce} - z} dF^{\mathbf{R}}(t).$$

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.
- Starting point: Cauchy integration formula

$$\mathbf{s}(\theta_k)^{\mathsf{H}}\mathbf{E}_W\mathbf{E}_W^{\mathsf{H}}\mathbf{s}(\theta_k) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathbf{s}(\theta_k)\left(\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{s}(\theta_k)dz$$

with $\mathcal{C}$ circling around σ^2 only (only one pole in $z = \sigma^2$).

- We then use the result:

Lemma

For $\mathbf{a} \in \mathbb{C}^N$ deterministic bounded, independent of $\mathbf{R}_N$,

$$\mathbf{a}^{\mathsf{H}}\left(\mathbf{R}_N - z\mathbf{I}_N\right)^{-1}\mathbf{a} - \mathbf{a}^{\mathsf{H}}\left(\frac{1}{1+ce_N(z)}\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{a} \xrightarrow{\text{a.s.}} 0$$

with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{t}{1+ce} - z} dF^{\mathbf{R}}(t).$$

- By change of variable, dominated convergence arguments, and residue calculus, we conclude.

G-MUSIC

X. Mestre, M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," IEEE Trans. on Signal Processing, vol. 54, no. 1, pp. 69-82, 2006.

Theorem

Under the above conditions,

$$\eta(\theta) - \bar{\eta}(\theta) \xrightarrow{\text{a.s.}} 0$$

as $N, M \to \infty$ *with* $0 < \lim N/M < \infty$, *where*

$$\bar{\eta}(\theta) = \mathbf{s}(\theta)^{\mathsf{H}} \left(\sum_{n=1}^{N} \phi(n) \hat{\mathbf{e}}_n \hat{\mathbf{e}}_n^{\mathsf{H}} \right) \mathbf{s}(\theta)$$

with $\phi(n)$ *defined as*

$$\phi(n) = \begin{cases} 1 + \sum_{k=N-K+1}^{N} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_n - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_n - \hat{\mu}_k} \right) & , \; n \leqslant N-K \\ -\sum_{k=1}^{N-K} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_n - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_n - \hat{\mu}_k} \right) & , \; n > N-K \end{cases}$$

and with $\mu_1 \leqslant \ldots \leqslant \mu_N$ *the eigenvalues of diag*$(\hat{\boldsymbol{\lambda}}) - \frac{1}{M}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^{\mathsf{T}}$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^{\mathsf{T}}$.

Simulation results

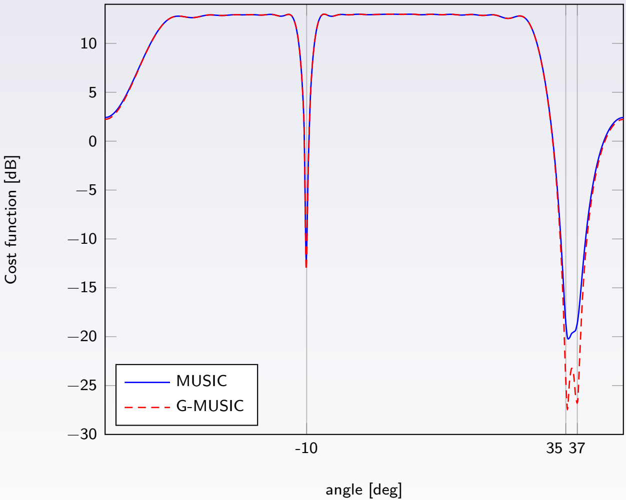


Figure : MUSIC against G-MUSIC for DoA detection of $K = 3$ signal sources, $N = 20$ sensors, $M = 150$ samples, SNR of 10 dB. Angles of arrival of 10°, 35°, and 37°.

Simulation results (2)

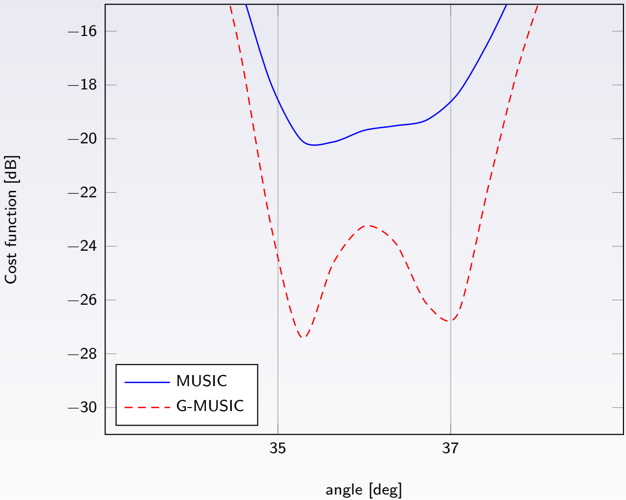


Figure : MUSIC against G-MUSIC for DoA detection of $K = 3$ signal sources, $N = 20$ sensors, $M = 150$ samples, SNR of 10 dB. Angles of arrival of 10°, 35°, and 37°.

Related bibliography

- Ø. Ryan, M. Debbah, "Free deconvolution for signal processing applications," ISIT, pp. 1846-1850, 2007.
- R. Rao, J. A. Mingo, R. Speicher, A. Edelman, "Statistical eigen-inference from large Wishart matrices," Annals of Statistics, vol. 36, no. 6, pp. 2850-2885, 2008.
- A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Finite dimensional statistical inference," vol. 57, no. 4, pp. 2457-2473, 2011.
- J. W. Silverstein, P. L. Combettes, "Large dimensional random matrix theory for signal detection and estimation in array processing," Workshop on Statistical Signal and Array Processing, pp. 276-279, 1992.
- R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," *to appear in* IEEE Trans. on Inf. Theory, 2010.
- J. Yao, R. Couillet, J. Najim, M. Debbah, "Fluctuations of an Improved Population Eigenvalue Estimator in Sample Covariance Matrix Models", (submitted to) IEEE Transactions on Information Theory.
- X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices", IEEE Transactions on Signal Processing, vol. 56, no. 11, 2008.
- X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates", IEEE Transactions on Information Theory, vol. 54, no. 11, Nov. 2008.
- A. Johnson, Y. Abramovich, X. Mestre, "MUSIC, G-MUSIC and Maximum Likelihood Performance Breakdown", IEEE Transactions on Signal Processing, vol. 56, no. 8, pp. 3944-3958, 2008.
- X. Mestre, M. A. Lagunas, "Modified Subspace Algorithms for DoA Estimation in the Small Sample Size Regime" IEEE Transactions on Signal Processing, Vol. 56, pp. 598-614, Feb. 2008.
- P. Vallet, P. Loubaton, X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case," arxiv preprint 1002.3234, 2010.
- W. Hachem, P. Loubaton, X. Mestre, P. Vallet, "A subspace estimator of finite rank perturbations of large random matrices," arXiv preprint, 2011.

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Covariance matrix against spike models

$\rightarrow$ The problems under consideration are of the type

$$\mathbf{Y} = \mathbf{A}(\Theta)\mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$ and

- $\mathbf{X}$ is random with i.i.d. entries
- $\mathbf{A}(\Theta)$ is a deterministic matrix-function of Θ (which can be recovered from spectrum information)

Covariance matrix against spike models

→ The problems under consideration are of the type

$$\mathbf{Y} = \mathbf{A}(\Theta)\mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N\times n}$ and

- $\mathbf{X}$ is random with i.i.d. entries
- $\mathbf{A}(\Theta)$ is a deterministic matrix-function of Θ (which can be recovered from spectrum information)

→ We want to retrieve Θ from the observation $\mathbf{Y}$, when both N and n are large, i.e. derivate (N, n)-consistent estimators

→ As opposed to finite N regime, two RMT approaches:

- $\mathbf{A}(\Theta)$ is a large rank matrix:
  - analysis of the link between $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $\mathbf{A}(\Theta)$
  - use of Bai–Silverstein method
  - use of statistical inference tools to retrieve Θ

Covariance matrix against spike models

→ The problems under consideration are of the type

$$\mathbf{Y} = \mathbf{A}(\Theta)\mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$ and

- $\mathbf{X}$ is random with i.i.d. entries
- $\mathbf{A}(\Theta)$ is a deterministic matrix-function of Θ (which can be recovered from spectrum information)

→ We want to retrieve Θ from the observation $\mathbf{Y}$, when both N and n are large, i.e. derivate (N, n)-consistent estimators

→ As opposed to finite N regime, two RMT approaches:

- $\mathbf{A}(\Theta)$ is a large rank matrix:
  - analysis of the link between $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $\mathbf{A}(\Theta)$
  - use of Bai–Silverstein method
  - use of statistical inference tools to retrieve Θ
  - → Improves classical n-consistent estimators

  → This is the case we already studied.

Covariance matrix against spike models

→ The problems under consideration are of the type

$$\mathbf{Y} = \mathbf{A}(\Theta)\mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$ and

- $\mathbf{X}$ is random with i.i.d. entries
- $\mathbf{A}(\Theta)$ is a deterministic matrix-function of Θ (which can be recovered from spectrum information)

→ We want to retrieve Θ from the observation $\mathbf{Y}$, when both N and n are large, i.e. derivate (N, n)-consistent estimators

→ As opposed to finite N regime, two RMT approaches:

- $\mathbf{A}(\Theta)$ is a large rank matrix:
  - analysis of the link between $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and $\mathbf{A}(\Theta)$
  - use of Bai–Silverstein method
  - use of statistical inference tools to retrieve Θ
  - → Improves classical n-consistent estimators

  → This is the case we already studied.
- $\mathbf{A}(\Theta)$ is a low-rank matrix:
  - $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ has a Marcenko-Pastur spectrum
  - use of known results for the Marcenko-Pastur law
  - use of statistical inference tools to retrieve Θ

Covariance matrix against spike models

→ The problems under consideration are of the type

$$\mathbf{Y} = \mathbf{A}(\Theta)\mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$ and

- $\mathbf{X}$ is random with i.i.d. entries
- $\mathbf{A}(\Theta)$ is a deterministic matrix-function of Θ (which can be recovered from spectrum information)

→ We want to retrieve Θ from the observation $\mathbf{Y}$, when both N and n are large, i.e. derivate (N, n)-consistent estimators

→ As opposed to finite N regime, two RMT approaches:

- $\mathbf{A}(\Theta)$ is a large rank matrix:
  - analysis of the link between $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and $\mathbf{A}(\Theta)$
  - use of Bai–Silverstein method
  - use of statistical inference tools to retrieve Θ
  - → Improves classical n-consistent estimators

  → This is the case we already studied.
- $\mathbf{A}(\Theta)$ is a low-rank matrix:
  - $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ has a Marcenko-Pastur spectrum
  - use of known results for the Marcenko-Pastur law
  - use of statistical inference tools to retrieve Θ
  - → Simpler but usually less accurate approach

Localization of small-dimensional sources (1)

$\rightarrow$ We consider the scenario of K sources and an N-antenna array capturing

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k^{(m)} + \mathbf{w}^{(m)}$$

- $s_k^{(m)}$ and $\mathbf{w}^{(m)}$ are random with zero mean and unit variance entries
- $m = 1, \ldots, n$ with N, n large assuming $N/n \rightarrow c > 0$, and K fixed

(we take $\sigma = 1$ for simplicity, which can be included in the $\mathbf{a}(\theta)$)

Localization of small-dimensional sources (1)

→ We consider the scenario of K sources and an N-antenna array capturing

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k^{(m)} + \mathbf{w}^{(m)}$$

- $s_k^{(m)}$ and $\mathbf{w}^{(m)}$ are random with zero mean and unit variance entries
- $m = 1, \ldots, n$ with N, n large assuming $N/n \to c > 0$, and K fixed

(we take $\sigma = 1$ for simplicity, which can be included in the $\mathbf{a}(\theta)$)

→ We consider a spiked random matrix approach. Denoting $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}]$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{A} & \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{W} \end{bmatrix}$$

with $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta)_K]$.

Localization of small-dimensional sources (1)

→ We consider the scenario of K sources and an N-antenna array capturing

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k^{(m)} + \mathbf{w}^{(m)}$$

- $s_k^{(m)}$ and $\mathbf{w}^{(m)}$ are random with zero mean and unit variance entries
- $m = 1, \ldots, n$ with N, n large assuming $N/n \to c > 0$, and K fixed

(we take $\sigma = 1$ for simplicity, which can be included in the $\mathbf{a}(\theta)$)

→ We consider a spiked random matrix approach. Denoting $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}]$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{A} & \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{W} \end{bmatrix}$$

with $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta)_K]$.

→ Spectral decomposition of the population covariance

$$\mathbf{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}] = \mathbf{A}\mathbf{A}^{\mathsf{H}} + \mathbf{I}_N = \mathbf{U}_S \boldsymbol{\Omega} \mathbf{U}_S^{\mathsf{H}} + \mathbf{I}_N$$

with $\mathbf{U}_S = [\mathbf{u}_1, \ldots, \mathbf{u}_K] \in \mathbb{C}^{N \times K}$ isometric, $\boldsymbol{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_K)$, $\omega_1 \geqslant \ldots \geqslant \omega_K$.

- $\mathbf{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}]$ is a small-rank perturbation of the identity matrix: spike model
- $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ is the empirical sample covariance matrix for this model

Localization of small-dimensional sources (1)

→ We consider the scenario of K sources and an N-antenna array capturing

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k^{(m)} + \mathbf{w}^{(m)}$$

- $s_k^{(m)}$ and $\mathbf{w}^{(m)}$ are random with zero mean and unit variance entries
- $m = 1, \ldots, n$ with N, n large assuming $N/n \to c > 0$, and K fixed

(we take $\sigma = 1$ for simplicity, which can be included in the $\mathbf{a}(\theta)$)

→ We consider a spiked random matrix approach. Denoting $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}]$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{A} & \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{W} \end{bmatrix}$$

with $\mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta)_K]$.

→ Spectral decomposition of the population covariance

$$\mathbf{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}] = \mathbf{A}\mathbf{A}^{\mathsf{H}} + \mathbf{I}_N = \mathbf{U}_S \boldsymbol{\Omega} \mathbf{U}_S^{\mathsf{H}} + \mathbf{I}_N$$

with $\mathbf{U}_S = [\mathbf{u}_1, \ldots, \mathbf{u}_K] \in \mathbb{C}^{N \times K}$ isometric, $\boldsymbol{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_K)$, $\omega_1 \geqslant \ldots \geqslant \omega_K$.

- $\mathbf{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}]$ is a small-rank perturbation of the identity matrix: spike model
- $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ is the empirical sample covariance matrix for this model

→ Some consequences of the model in the RMT setting (see e.g. Weyl's inequality)

- limiting weak spectrum is the Marcenko–Pastur law!
- up to K eigenvalues can leave the limiting support

Localization of small-dimensional sources (2)

$\rightarrow$ We first need to understand the spectrum of $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$

- We know that the weak spectrum is the MP law
- Up to K eigenvalues can leave the support: we identify here these eigenvalues

Localization of small-dimensional sources (2)

→ We first need to understand the spectrum of $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$

- We know that the weak spectrum is the MP law
- Up to K eigenvalues can leave the support: we identify here these eigenvalues

→ Denote $\mathbf{P} = \mathbf{A}\mathbf{A}^{\mathsf{H}} = \mathbf{U}_S \boldsymbol{\Omega} \mathbf{U}_S^{\mathsf{H}}$, $\boldsymbol{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_K)$, and $\mathbf{X} = [\mathbf{x}^{\mathsf{T}} \ \mathbf{W}^{\mathsf{T}}]^{\mathsf{T}}$ to recover (up to one row) the generic spiked model

$$\mathbf{Y} = (\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}} \mathbf{X}.$$

Localization of small-dimensional sources (2)

→ We first need to understand the spectrum of $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$

- We know that the weak spectrum is the MP law
- Up to K eigenvalues can leave the support: we identify here these eigenvalues

→ Denote $\mathbf{P} = \mathbf{A}\mathbf{A}^{\mathsf{H}} = \mathbf{U}_S\boldsymbol{\Omega}\mathbf{U}_S^{\mathsf{H}}$, $\boldsymbol{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_K)$, and $\mathbf{X} = [\mathbf{x}^{\mathsf{T}}\ \mathbf{W}^{\mathsf{T}}]^{\mathsf{T}}$ to recover (up to one row) the generic spiked model

$$\mathbf{Y} = (\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}\mathbf{X}.$$

- Reminder: If x eigenvalue of $\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ with $x > (1+\sqrt{c})^2$ (edge of MP law), for all large n,

$$x \triangleq \lambda_k \xrightarrow{\text{a.s.}} \rho_k \triangleq 1 + \omega_k + c(1+\omega_k)\omega_k^{-1}, \ \textit{if}\ \ \omega_k > \sqrt{c}$$

  for some k.

Localization of small-dimensional sources (3)

→ Recall the MUSIC approach: we want to estimate

$$\eta(\theta) = \mathbf{a}(\theta)^{\mathsf{H}} \mathbf{U}_W \mathbf{U}_W^{\mathsf{H}} \mathbf{a}(\theta) \quad (\mathbf{U}_W \in \mathbb{C}^{N \times (N-K)} \text{ such that } \mathbf{U}_W^{\mathsf{H}} \mathbf{U}_S = 0)$$

Localization of small-dimensional sources (3)

→ Recall the MUSIC approach: we want to estimate

$$\eta(\theta) = \mathbf{a}(\theta)^{\mathsf{H}}\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}}\mathbf{a}(\theta) \quad (\mathbf{U}_W \in \mathbb{C}^{N\times(N-K)} \text{ such that } \mathbf{U}_W^{\mathsf{H}}\mathbf{U}_S = 0)$$

→ Instead of this quantity, we start with the study of

$$\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{a}(\theta),\ k = 1,\ldots,K$$

with $\hat{\mathbf{u}}_1,\ldots,\hat{\mathbf{u}}_N$ the eigenvectors belonging to $\lambda_1 \geqslant \ldots \geqslant \lambda_N$.

→ To fall back on known RMT quantities, we use the Cauchy-integral:

$$\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{a}(\theta) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_i}\mathbf{a}(\theta)^{\mathsf{H}}(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{a}(\theta)dz$$

with $\mathcal{C}_i$ a contour enclosing λ_i only.

Localization of small-dimensional sources (3)

→ Recall the MUSIC approach: we want to estimate

$$\eta(\theta) = \mathbf{a}(\theta)^{\mathsf{H}}\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}}\mathbf{a}(\theta) \quad (\mathbf{U}_W \in \mathbb{C}^{N\times(N-K)} \text{ such that } \mathbf{U}_W^{\mathsf{H}}\mathbf{U}_S = 0)$$

→ Instead of this quantity, we start with the study of

$$\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{a}(\theta), \; k = 1, \ldots, K$$

with $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N$ the eigenvectors belonging to $\lambda_1 \geqslant \ldots \geqslant \lambda_N$.

→ To fall back on known RMT quantities, we use the Cauchy-integral:

$$\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{a}(\theta) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{a}(\theta)^{\mathsf{H}}(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{a}(\theta)dz$$

with $\mathcal{C}_i$ a contour enclosing λ_i only.

→ Woodbury's identity $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$ gives:

$$\mathbf{a}^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{a} = \frac{-1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{a}^{\mathsf{H}}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}(\frac{\mathbf{X}\mathbf{X}^{\mathsf{H}}}{n} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{a}dz + \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \hat{\mathbf{a}}_1^{\mathsf{H}}\widehat{\mathbf{H}}^{-1}\hat{\mathbf{a}}_2dz$$

where $\mathbf{P} = \mathbf{U}_S\boldsymbol{\Omega}\mathbf{U}_S^{\mathsf{H}}$, and

$$\begin{cases} \widehat{\mathbf{H}} & = \mathbf{I}_K + z\boldsymbol{\Omega}(\mathbf{I}_K+\boldsymbol{\Omega})^{-1}\mathbf{U}_S^{\mathsf{H}}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{U}_S \\ \hat{\mathbf{a}}_1^{\mathsf{H}} & = z\mathbf{a}(\theta)^{\mathsf{H}}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{U}_S \\ \hat{\mathbf{a}}_2 & = \boldsymbol{\Omega}(\mathbf{I}_K+\boldsymbol{\Omega})^{-1}\mathbf{U}_S^{\mathsf{H}}(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{a}(\theta). \end{cases}$$

Localization of small-dimensional sources (4)

- For large n, the first term has no pole, while the second converges to

$$T_i \triangleq \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{a}_1^{\mathsf{H}}\mathbf{H}^{-1}\mathbf{a}_2 dz, \text{ with } \begin{cases} \mathbf{H} & = \mathbf{I}_K + zm(z)\boldsymbol{\Omega}(\mathbf{I}_K + \boldsymbol{\Omega})^{-1} \\ \mathbf{a}_1^{\mathsf{H}} & = zm(z)\mathbf{a}^*(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}\mathbf{U}_S \\ \mathbf{a}_2 & = m(z)\boldsymbol{\Omega}(\mathbf{I}_K + \boldsymbol{\Omega})^{-1}\mathbf{U}_S^{\mathsf{H}}(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}\mathbf{a} \end{cases}$$

which after development is

$$T_i = \sum_{\ell=1}^{K} \frac{1}{1+\omega_\ell}\frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \frac{zm^2(z)}{\frac{1+\omega_\ell}{\omega_\ell} + zm(z)} dz.$$

Localization of small-dimensional sources (4)

- For large n, the first term has no pole, while the second converges to

$$T_i \triangleq \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{a}_1^{\mathsf{H}}\mathbf{H}^{-1}\mathbf{a}_2 dz, \text{ with } \begin{cases} \mathbf{H} & = \mathbf{I}_K + zm(z)\mathbf{\Omega}(\mathbf{I}_K+\mathbf{\Omega})^{-1} \\ \mathbf{a}_1^{\mathsf{H}} & = zm(z)\mathbf{a}^*(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{U}_S \\ \mathbf{a}_2 & = m(z)\mathbf{\Omega}(\mathbf{I}_K+\mathbf{\Omega})^{-1}\mathbf{U}_S^{\mathsf{H}}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{a} \end{cases}$$

which after development is

$$T_i = \sum_{\ell=1}^{K} \frac{1}{1+\omega_\ell}\frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \frac{zm^2(z)}{\frac{1+\omega_\ell}{\omega_\ell}+zm(z)}dz.$$

- Using residue calculus, the sole pole is in ρ_i and we find

$$\mathbf{a}^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{a}_i \xrightarrow{\text{a.s.}} \frac{1-c\omega_i^{-2}}{1+c\omega_i^{-1}}\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{H}}\mathbf{a}(\theta).$$

Localization of small-dimensional sources (5)

→ We now conclude

$$\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}}\mathbf{a}(\theta) = \mathbf{a}(\theta)^{\mathsf{H}}\mathbf{a}(\theta) - \sum_{k=1}^{K}\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{u}_k\mathbf{u}_k^{\mathsf{H}}\mathbf{a}(\theta)$$

where

$$\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{u}_k\mathbf{u}_k^{\mathsf{H}}\mathbf{a}(\theta) - \frac{1+c\omega_k^{-1}}{1-c\omega_k^{-2}}\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathsf{H}}\mathbf{a}(\theta) \xrightarrow{\text{a.s.}} 0$$

→ The ω_k are however unknown. But they can be estimated from

$$\lambda_k \xrightarrow{\text{a.s.}} \rho_k = 1 + \omega_k + c(1+\omega_k)\omega_k^{-1}$$

→ This gives finally

$$\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}}\mathbf{a}(\theta) \simeq \mathbf{a}(\theta)^{\mathsf{H}}\mathbf{a}(\theta) - \sum_{k=1}^{K}\frac{1+c\hat{\omega}_k^{-1}}{1-c\hat{\omega}_k^{-2}}\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathsf{H}}\mathbf{a}(\theta)$$

with

$$\hat{\omega}_k = \frac{\hat{\lambda}_k - (c+1)}{2} + \sqrt{(c+1-\hat{\lambda}_k)^2 - 4c)}$$

Localization of small-dimensional sources (5)

→ We now conclude

$$\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}}\mathbf{a}(\theta) = \mathbf{a}(\theta)^{\mathsf{H}}\mathbf{a}(\theta) - \sum_{k=1}^{K}\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{u}_k\mathbf{u}_k^{\mathsf{H}}\mathbf{a}(\theta)$$

where

$$\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{u}_k\mathbf{u}_k^{\mathsf{H}}\mathbf{a}(\theta) - \frac{1+c\omega_k^{-1}}{1-c\omega_k^{-2}}\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathsf{H}}\mathbf{a}(\theta) \xrightarrow{\text{a.s.}} 0$$

→ The ω_k are however unknown. But they can be estimated from

$$\lambda_k \xrightarrow{\text{a.s.}} \rho_k = 1 + \omega_k + c(1+\omega_k)\omega_k^{-1}$$

→ This gives finally

$$\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{U}_W\mathbf{U}_W^{\mathsf{H}}\mathbf{a}(\theta) \simeq \mathbf{a}(\theta)^{\mathsf{H}}\mathbf{a}(\theta) - \sum_{k=1}^{K}\frac{1+c\hat{\omega}_k^{-1}}{1-c\hat{\omega}_k^{-2}}\mathbf{a}(\theta)^{\mathsf{H}}\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathsf{H}}\mathbf{a}(\theta)$$

with

$$\hat{\omega}_k = \frac{\hat{\lambda}_k - (c+1)}{2} + \sqrt{(c+1-\hat{\lambda}_k)^2 - 4c)}$$

→ We then obtain another (N,n)-consistent MUSIC estimator, only valid for K finite!

Outline

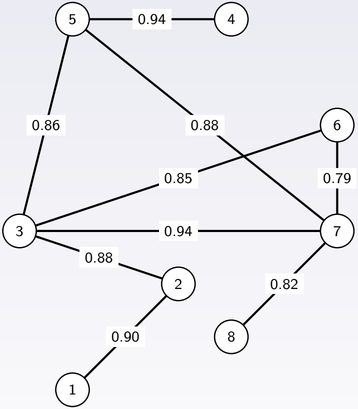
Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

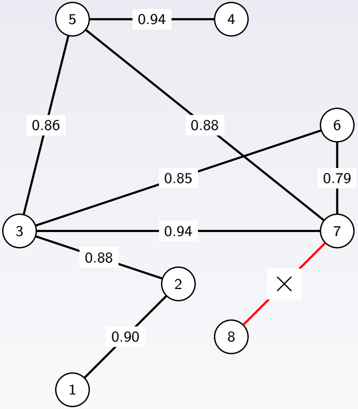
3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Problem statement



Problem statement



- Localize local failures based on observations from a sensor network.
- Focus on failures modeled as small rank perturbations of large random matrices.

Target

- Systems with failures modeled by small rank perturbations
- Observation matrix $\mathbf{\Sigma} = [\mathbf{s}_1, \ldots, \mathbf{s}_n] \in \mathbb{C}^{N \times n}$ modeled by

$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P}_k)^{\frac{1}{2}} \mathbf{X}$$

with $\mathbf{P}_k \in \mathbb{C}^{N \times N}$ of rank $r_k \ll N$, $\mathbf{X}$ with independent $\mathcal{CN}(0, 1/n)$ entries.

Target

- Systems with failures modeled by small rank perturbations
- Observation matrix $\mathbf{\Sigma} = [\mathbf{s}_1, \ldots, \mathbf{s}_n] \in \mathbb{C}^{N \times n}$ modeled by

$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P}_k)^{\frac{1}{2}} \mathbf{X}$$

  with $\mathbf{P}_k \in \mathbb{C}^{N \times N}$ of rank $r_k \ll N$, $\mathbf{X}$ with independent $\mathcal{CN}(0, 1/n)$ entries.
- Failure scenarios:
  - $(\mathcal{H}_0)$: no failure, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N$.
  - $(\mathcal{H}_k)$: $1 \leqslant k \leqslant K$, failure of type k, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$.

Target

- Systems with failures modeled by small rank perturbations
- Observation matrix $\mathbf{\Sigma} = [\mathbf{s}_1, \ldots, \mathbf{s}_n] \in \mathbb{C}^{N \times n}$ modeled by

$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P}_k)^{\frac{1}{2}} \mathbf{X}$$

  with $\mathbf{P}_k \in \mathbb{C}^{N \times N}$ of rank $r_k \ll N$, $\mathbf{X}$ with independent $\mathcal{CN}(0, 1/n)$ entries.
- Failure scenarios:
  - $(\mathcal{H}_0)$: no failure, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N$.
  - $(\mathcal{H}_k)$: $1 \leqslant k \leqslant K$, failure of type k, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$.
- Subspace approach for:
  - detecting a failure: decide between $\mathcal{H}_0$ and $\bar{\mathcal{H}}_0$
  - diagnosing a failure: upon failure detection, decide on the most probable $\mathcal{H}_k$.

Example 1

Node failure in sensor networks

- Consider the model

$$\mathbf{y} = \mathbf{H}\theta + \sigma\mathbf{w}$$

with $\mathbf{H} \in \mathbb{C}^{N\times p}$ deterministic, $\theta \sim \mathcal{CN}(0, \mathbf{I}_p)$, $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}_N)$.

  - In particular $E[\mathbf{y}] = 0$ and $E[\mathbf{y}\mathbf{y}^{\mathsf{H}}] = \mathbf{R} \triangleq \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$
  - With $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}$, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N$.

Example 1

Node failure in sensor networks

- Consider the model
$$\mathbf{y} = \mathbf{H}\theta + \sigma\mathbf{w}$$
with $\mathbf{H} \in \mathbb{C}^{N\times p}$ deterministic, $\theta \sim \mathcal{CN}(0, \mathbf{I}_p)$, $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}_N)$.
  - In particular $E[\mathbf{y}] = 0$ and $E[\mathbf{y}\mathbf{y}^{\mathsf{H}}] = \mathbf{R} \triangleq \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$
  - With $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}$, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N$.
- Upon failure of sensor k, $\mathbf{y}$ becomes
$$\mathbf{y}' = (\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^{\mathsf{H}})\mathbf{H}\theta + \sigma_k\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}}\theta' + \sigma\mathbf{w}$$
for some noise variance σ_k^2.
  - Now $E[\mathbf{y}'] = 0$ and $E[\mathbf{y}'\mathbf{y}'^{\mathsf{H}}] = (\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^{\mathsf{H}})\mathbf{H}\mathbf{H}^{\mathsf{H}}(\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^{\mathsf{H}}) + \sigma_k^2\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N$.
  - With now $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}'$,
$$E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$$
with
$$\mathbf{P}_k = -\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}} + \mathbf{R}^{-\frac{1}{2}}\mathbf{e}_k\left[(\mathbf{e}_k^{\mathsf{H}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{e}_k + \sigma_k^2)\mathbf{e}_k^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}} - \mathbf{e}_k^{\mathsf{H}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}}\right]$$
of rank-2 (image of $\mathbf{P}_k$ in $\mathrm{Span}(\mathbf{R}^{-\frac{1}{2}}\mathbf{e}_k, \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{e}_k)$)

Example 2

Sudden parameter change detection in sensor networks

- Upon sudden change of parameter θ_k,

$$\mathbf{y}' = \mathbf{H}(\mathbf{I}_p + \alpha_k \mathbf{e}_k \mathbf{e}_k^*)\theta + \mu_k \mathbf{H}\mathbf{e}_k + \sigma \mathbf{w}$$

- Then

$$E[\mathbf{y}'\mathbf{y}'^{\mathsf{H}}] = \mathbf{H}(\mathbf{I}_p + [\mu_k^2 + (1+\alpha_k)^2 - 1]\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}})\mathbf{H}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N.$$

- With $\mathbf{R} = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N$ and $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}'$,

$$E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$$

with

$$\mathbf{P}_k = [\mu_k^2 + (1+\alpha_k)^2 - 1]\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}}.$$

of rank-1.

Eigenvalue and eigenvectors statistics: Method

- Consider the model

$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P})^{\frac{1}{2}}\mathbf{X}$$

with, for simplicity

  - $\mathbf{X}$ standard Gaussian
  - $\mathbf{P} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{C}^{N \times r}$, $\mathbf{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_r)$, $\omega_1 > \ldots > \omega_r > 0$.

- Convergence properties of
  - $\lambda_1 > \ldots > \lambda_r$, the r largest eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}}$
  - $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i$, with $\hat{\mathbf{u}}_i$ the eigenvector associated to λ_i.

Eigenvalue and eigenvectors statistics: Method

- Consider the model

$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P})^{\frac{1}{2}}\mathbf{X}$$

  with, for simplicity
  - $\mathbf{X}$ standard Gaussian
  - $\mathbf{P} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{C}^{N\times r}$, $\mathbf{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_r)$, $\omega_1 > \ldots > \omega_r > 0$.
- Convergence properties of
  - $\lambda_1 > \ldots > \lambda_r$, the r largest eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}}$
  - $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i$, with $\hat{\mathbf{u}}_i$ the eigenvector associated to λ_i.
- Study based on two ingredients
  - the Stieltjes transform method
  - complex analysis

First order limits

- (Reminder) The *limiting* ρ_k are given by:

$$\lambda_k \xrightarrow{\text{a.s.}} \rho_k \triangleq 1 + \omega_k + c(1 + \omega_k)\omega_k^{-1}, \; \textit{if} \;\; \omega_k > \sqrt{c}$$

- Consider ω_i and its corresponding eigenvector $\mathbf{u}_i$, then

$$\mathbf{u}_i^{\mathsf{H}} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^{\mathsf{H}} \mathbf{u}_i \xrightarrow{\text{a.s.}} \zeta_i \triangleq \frac{1 - c\omega_i^{-2}}{1 + c\omega_i^{-1}}.$$

Fluctuations

Second order behaviour for the joint variable

$$\left(\left(\sqrt{N}(\lambda_i - \rho_i)\right)_{i=1}^r, \left(\sqrt{N}(\mathbf{u}_i^H \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \mathbf{u}_i - \zeta_i)\right)_{i=1}^r\right)$$

R. Couillet, W. Hachem, "Local failure detection and diagnosis in large sensor networks", IEEE Transactions on Information Theory, arXiv Preprint 1107.1409.

Theorem

Under the conditions above, assuming $\omega_i > \sqrt{c}$ for each $i \in \{1, \ldots, r\}$,

$$\left(\left(\sqrt{N}(\lambda_i - \rho_i)\right)_{i=1}^r, \left(\sqrt{N}(\mathbf{u}_i^H \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \mathbf{u}_i - \zeta_i)\right)_{i=1}^r\right) \Rightarrow \mathcal{N}\left(0, \begin{bmatrix} C(\rho_1) & & \\ & \ddots & \\ & & C(\rho_r) \end{bmatrix}\right)$$

where

$$C(\rho_i) \triangleq \begin{bmatrix} \frac{c^2(1+\omega_i)^2}{(c+\omega_i)^2(\omega_i^2-c)}\left(c\frac{(1+\omega_i)^2}{(c+\omega_i)^2}+1\right) & \frac{(1+\omega_i)^3c^2}{(\omega_i+c)^2\omega_i} \\ \frac{(1+\omega_i)^3c^2}{(\omega_i+c)^2\omega_i} & \frac{c(1+\omega_i)^2(\omega_i^2-c)}{\omega_i^2} \end{bmatrix}.$$

Reminder: Fluctuations at the edge of the bulk

- The previous theorem holds for $\omega_i > \sqrt{c}$, i.e. "strong perturbations"

Reminder: Fluctuations at the edge of the bulk

- The previous theorem holds for $\omega_i > \sqrt{c}$, i.e. "strong perturbations"
- For $\omega_i < \sqrt{c}$, the eigenvalue fluctuations are:

Theorem

If $0 \leqslant \omega_i < \sqrt{c}$,

$$N^{\frac{2}{3}}(1+\sqrt{c})^{-\frac{4}{3}}c^{-\frac{1}{2}}(\lambda_i - (1+\sqrt{c})^2) \Rightarrow T_2$$

where T_2 *is the complex Tracy-Widom distribution function.*

Failure detection and localization

- The proposed subspace procedure is a two-step approach:
  - **Failure detection procedure**, $\mathcal{H}_0$ vs. $\bar{\mathcal{H}}_0$: We evaluate the statistics of λ_1 against the Tracy-Widom law for a false alarm rate η,

$$\lambda_1' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta)$$

where $\lambda_1' \triangleq N^{\frac{2}{3}}(1+\sqrt{c_N})^{-\frac{4}{3}} c_N^{-\frac{1}{2}}(\lambda_1 - (1+\sqrt{c_N})^2)$.

Failure detection and localization

- The proposed subspace procedure is a two-step approach:
  - **Failure detection procedure**, $\mathcal{H}_0$ vs. $\bar{\mathcal{H}}_0$: We evaluate the statistics of λ_1 against the Tracy-Widom law for a false alarm rate η,

$$\lambda_1' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta)$$

    where $\lambda_1' \triangleq N^{\frac{2}{3}}(1+\sqrt{c_N})^{-\frac{4}{3}} c_N^{-\frac{1}{2}}(\lambda_1 - (1+\sqrt{c_N})^2)$.
  - **Failure diagnosis**, selection of $\mathcal{H}_k$: We evaluate the joint statistics of λ_i, $\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_{k,i}$ for each $k \in \{1,\ldots,K\}$, and obtain the maximum-likelihood test,

$$\hat{k} = \arg \max_{1 \leqslant k \leqslant K} \prod_{i=1}^{r} f\left(\left(\left(\sqrt{N}(\lambda_i - \rho_{k,i})\right)_{i=1}^{r}, \left(\sqrt{N}(\mathbf{u}_{k,i}^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_{k,i} - \zeta_{k,i})\right)_{i=1}^{r}\right); C(\rho_{k,i})\right)$$

    with $f(x;\mathbf{R})$ the Gaussian density with zero mean and variance $\mathbf{R}$, and indices k corresponding to hypothesis $\mathcal{H}_k$.

Results

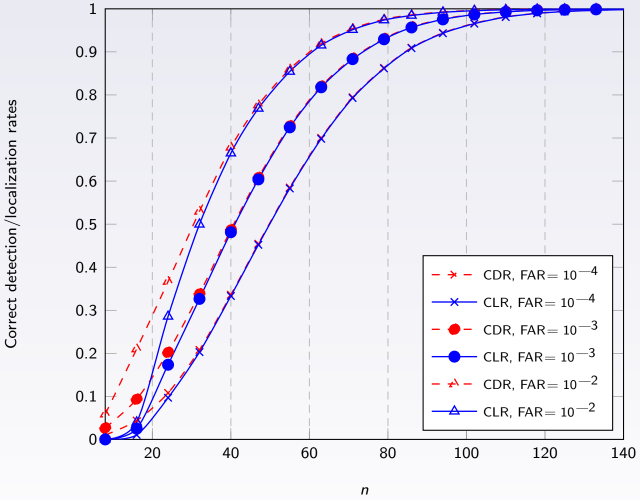


Figure : Simulation of sensor failure in an $N = 10$ node network. Correct detection (CDR) and localization (CLR) rates for different false alarm rates (FAR) and different n.

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Outline

Part 1: Fundamentals of Random Matrix Theory

1.1. The Stieltjes Transform Method

1.2. Extreme Eigenvalues

1.3. The Spiked Model

1.4. Spectrum Analysis and G-estimation

2. Source Detection

2.1. Eigenvalue-based Detection

2.2. Detection in unknown Noise Environment

3. Statistical Inference

3.1. Generic Model

3.1.2. Angle-of-arrival estimation

3.1.2. Angle-of-arrival estimation

3.2. Spiked Model

3.2.1. Spiked G-MUSIC

3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation

4.1. Introduction to Robust Estimation

4.1. Initial Results and Open Problems

Parameter estimation and sample covariance matrix

P.J. Huber, “Robust Statistics”, 1981.

→ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

Parameter estimation and sample covariance matrix

P.J. Huber, "Robust Statistics", 1981.

$\rightarrow$ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

- The main reasons are:
  - Assuming $E[x] = 0$, $E[xx^*] = C_N$, with $X = [x_1, \ldots, x_n]$, by the LLN

$$\hat{S}_N \triangleq \frac{1}{n} XX^* \xrightarrow{\text{a.s.}} C_N \text{ as } n \to \infty.$$

$\rightarrow$ Hence, if $\theta = f(C_N)$, we often use the n-consistent estimate $\hat{\theta} = f(\hat{S}_N)$.

Parameter estimation and sample covariance matrix

P.J. Huber, "Robust Statistics", 1981.

$\rightarrow$ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

- The main reasons are:
  - Assuming $E[x] = 0$, $E[xx^*] = C_N$, with $X = [x_1, \ldots, x_n]$, by the LLN

$$\hat{S}_N \triangleq \frac{1}{n} XX^* \xrightarrow{\text{a.s.}} C_N \text{ as } n \to \infty.$$

    $\rightarrow$ Hence, if $\theta = f(C_N)$, we often use the n-consistent estimate $\hat{\theta} = f(\hat{S}_N)$.
  - The SCM $\hat{S}_N$ is the ML estimate of C_N for Gaussian x
    $\rightarrow$ One therefore expects $\hat{\theta}$ to closely approximate θ for all finite n.

Parameter estimation and sample covariance matrix

P.J. Huber, "Robust Statistics", 1981.

→ Many statistical inference techniques rely on the sample covariance matrix (SCM) taken from i.i.d. observations $x_1, \ldots, x_n$ of a r.v. $x \in \mathbb{C}^N$.

- The main reasons are:
  - Assuming $E[x] = 0$, $E[xx^*] = C_N$, with $X = [x_1, \ldots, x_n]$, by the LLN
  $$\hat{S}_N \triangleq \frac{1}{n} XX^* \xrightarrow{\text{a.s.}} C_N \text{ as } n \to \infty.$$
    → Hence, if $\theta = f(C_N)$, we often use the n-consistent estimate $\hat{\theta} = f(\hat{S}_N)$.
  - The SCM $\hat{S}_N$ is the ML estimate of C_N for Gaussian x
    → One therefore expects $\hat{\theta}$ to closely approximate θ for all finite n.
- This approach however has two limitations:
  - if N, n are of the same order of magnitude,
  $$\|\hat{S}_N - C_N\| \not\to 0 \text{ as } N, n \to \infty, \; N/n \to c > 0, \text{ so that in general } |\hat{\theta} - \theta| \not\to 0$$
    → This motivated the introduction of G-estimators.
  - if x is not Gaussian, but has heavier tails, $\hat{S}_N$ is a poor estimator for C_N.
    → This motivated the introduction of robust estimators.

Reminders on robust estimation

J. T. Kent, D. E. Tyler, "Redescending M-estimates of multivariate location and scatter", 1991.
R. A. Maronna, "Robust M-estimators of multivariate location and scatter", 1976.
Y. Chitour, F. Pascal, "Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis", 2008.

→ The objectives of robust estimators:

- Replace the SCM $\hat{S}_N$ by another estimate $\hat{C}_N$ of C_N which:
  - rejects (or downscales) observations deterministically
  - or rejects observations inconsistent with the full set of observations

→ **Example**: Huber estimator, $\hat{C}_N$ defined as solution of

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} \beta_i x_i x_i^* \text{ with } \beta_i = \alpha \min\left\{1, \frac{k^2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i}\right\} \text{ for some } \alpha > 1, k^2 \text{ function of } \hat{C}_N.$$

Reminders on robust estimation

J. T. Kent, D. E. Tyler, "Redescending M-estimates of multivariate location and scatter", 1991.
R. A. Maronna, "Robust M-estimators of multivariate location and scatter", 1976.
Y. Chitour, F. Pascal, "Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis", 2008.

→ The objectives of robust estimators:

- Replace the SCM $\hat{S}_N$ by another estimate $\hat{C}_N$ of C_N which:
  - rejects (or downscales) observations deterministically
  - or rejects observations inconsistent with the full set of observations

  → **Example**: Huber estimator, $\hat{C}_N$ defined as solution of

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} \beta_i x_i x_i^* \text{ with } \beta_i = \alpha \min\left\{1, \frac{k^2}{\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i}\right\} \text{ for some } \alpha > 1, k^2 \text{ function of } \hat{C}_N.$$

- Provide scale-free estimators of C_N:

  → **Example**: Tyler's estimator: if one observes $x_i = \tau_i z_i$ for unknown scalars τ_i,

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i} x_i x_i^*$$

  - existence and uniqueness of $\hat{C}_N$ defined up to a constant.
  - few constraints on $x_1, \ldots, x_n$ ($N+1$ of them must be linearly independent)

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by the ML estimate for C_N.
  → **Example**: Maronna's estimator for elliptical x

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i\right) x_i x_i^*$$

with $u(s)$ such that

(i) $u(s)$ is continuous and non-increasing on $[0, \infty)$
(ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.

(note that Huber's estimator is compliant with Maronna's estimators)

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by the ML estimate for C_N.
  → **Example**: Maronna's estimator for elliptical x

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i\right) x_i x_i^*$$

  with $u(s)$ such that
  (i) $u(s)$ is continuous and non-increasing on $[0,\infty)$
  (ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.
  (note that Huber's estimator is compliant with Maronna's estimators)
  - existence is not too demanding
  - uniqueness imposes constraints on $N, n, u(s)$, e.g. $\phi_\infty > \frac{n}{n-N}$. Inconsistent with random matrix regime!
  - consistency result: $\hat{C}_N \to C_N$ **if** $u(s)$ meets the ML estimator for C_N.

Robust Estimation and RMT

→ So far, RMT has mostly focused on the SCM $\hat{S}_N$.

- $x = A_N y$, y having i.i.d. zero-mean unit variance entries,

Robust Estimation and RMT

→ So far, RMT has mostly focused on the SCM $\hat{S}_N$.

- $x = A_N y$, y having i.i.d. zero-mean unit variance entries,
- x satisfies concentration inequalities, e.g. elliptically distributed x.

Robust Estimation and RMT

→ So far, RMT has mostly focused on the SCM $\hat{S}_N$.

- $x = A_N y$, y having i.i.d. zero-mean unit variance entries,
- x satisfies concentration inequalities, e.g. elliptically distributed x.

Robust RMT estimation

Can we study the performance of estimators based on the $\hat{C}_N$?

- what are the spectral properties of $\hat{C}_N$?
- can we generate RMT-based estimators relying on $\hat{C}_N$?

Outline

Part 1: Fundamentals of Random Matrix Theory
1.1. The Stieltjes Transform Method
1.2. Extreme Eigenvalues
1.3. The Spiked Model
1.4. Spectrum Analysis and G-estimation

2. Source Detection
2.1. Eigenvalue-based Detection
2.2. Detection in unknown Noise Environment

3. Statistical Inference
3.1. Generic Model
3.1.2. Angle-of-arrival estimation
3.1.2. Angle-of-arrival estimation
3.2. Spiked Model
3.2.1. Spiked G-MUSIC
3.2.2. Local Failure Detection in Sensor Networks

4. Random Matrix Theory and Robust Estimation
4.1. Introduction to Robust Estimation
4.1. Initial Results and Open Problems

Some first answers

→ Recall that

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{N} u\left(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i\right) x_i x_i^*$$

for some i.i.d. $x_1, \ldots, x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

  $$F^{\alpha \hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N, n \to \infty,\ N/n \to c > 0$$

  so in particular, for $C_N = I_N$,

  $$F^{\alpha \hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

Some first answers

→ Recall that

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{N} u\left(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i\right) x_i x_i^*$$

for some i.i.d. $x_1, \ldots, x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

  $$F^{\alpha \hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N, n \to \infty,\ N/n \to c > 0$$

  so in particular, for $C_N = I_N$,

  $$F^{\alpha \hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

- Zooming in on the eigenvalues suggests also that

  $$\lambda_i(\alpha \hat{C}_N) - \lambda_i(\hat{S}_N) \Rightarrow 0$$

Some first answers

→ Recall that

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{N} u\left(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i\right)x_ix_i^*$$

for some i.i.d. $x_1,\ldots,x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

$$F^{\alpha\hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N,n\to\infty,\ \ N/n\to c>0$$

  so in particular, for $C_N = I_N$,

$$F^{\alpha\hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

- Zooming in on the eigenvalues suggests also that

$$\lambda_i(\alpha\hat{C}_N) - \lambda_i(\hat{S}_N) \Rightarrow 0$$

→ This behavior seems to be linked to a concentration result on

$$\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i,\ \ i=1,\ldots,n$$

⇒ This is what we are going to prove.

Some first answers

→ Recall that

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{N} u\left(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i\right)x_ix_i^*$$

for some i.i.d. $x_1, \ldots, x_n$ taken from a random vector x, and for some function $u(s)$.

→ For x Gaussian and $u(s)$ of Maronna or Tyler type, simulations suggest:

- Marčenko-Pastur/Bai-Silverstein distribution, i.e. up to some constant α

$$F^{\alpha\hat{C}_N} - F^{\hat{S}_N} \Rightarrow 0 \text{ as } N, n \to \infty, \; N/n \to c > 0$$

so in particular, for $C_N = I_N$,

$$F^{\alpha\hat{C}_N} \Rightarrow F^{\mathrm{MP}}$$

- Zooming in on the eigenvalues suggests also that

$$\lambda_i(\alpha\hat{C}_N) - \lambda_i(\hat{S}_N) \Rightarrow 0$$

→ This behavior seems to be linked to a concentration result on

$$\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i, \; i = 1, \ldots, n$$

⇒ This is what we are going to prove.

Then, what happens to $\hat{C}_N$ when no concentration result occurs?

⇒ So far, we have no general answer to this question!

Some first answers (2)

$\rightarrow$ Main difficulties for handling $\hat{C}_N$:

- $\hat{C}_N$ does not always exist/is not always unique.
  - sometimes, uniqueness results inconsistent with random matrix regime

Some first answers (2)

→ Main difficulties for handling $\hat{C}_N$:

- $\hat{C}_N$ does not always exist/is not always unique.
  - sometimes, uniqueness results inconsistent with random matrix regime
- Contrary to classical RMT, the column vectors $\sqrt{u(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i)}x_i$ are not independent
  - difficult to find an angle to reuse previous results
- In general, it is already difficult to show that both $\|\hat{C}_N\|$ and $\|\hat{C}_N^{-1}\|$ remain bounded as $N, n \to \infty$, $N/n \to c > 0$.

Robust model

Assumptions

- Assumptions on $u(s)$,
  (i) $u(s)$ is continuous and non-increasing on $[0, \infty)$
  (ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.
- Assumptions on $x_1, \ldots, x_n$,
  - $x_i = A_N y_i \in \mathbb{C}^N$, $y_i \in \mathbb{C}^M$ has independent entries with
    - $E[y_{i,j}] = 0$
    - $E[y_{i,j}^2] = 0$, $E[|y_{i,j}|^2] = 1$
    - $\sup_{i,j} E[|y_{i,j}|^{8+\eta}] < \infty$.
  - With $c_N = N/n$, $\bar{c}_N = M/N \geqslant 1$,

$$0 < \liminf_n c_N \leqslant \limsup_n c_N < 1, \quad \limsup_n \bar{c}_N < \infty$$

  - Denoting $C_N = A_N A_N^*$,

$$0 < \liminf_N \{\lambda_1(C_N)\} \leqslant \limsup_N \{\lambda_N(C_N)\} < \infty$$

Robust SCM estimator in the RMT regime

R. Couillet, F. Pascal, J. W. Silverstein, "Robust M-Estimation for Array Processing: A Random Matrix Approach", (submitted to) IEEE Trans. on Information Theory, 2013.

Theorem

Assume the above and consider the fixed-point equation in $Z \in \mathbb{C}^{N \times N}$,

$$Z = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} x_i^* Z^{-1} x_i\right) x_i x_i^* . \tag{1}$$

Then,

(I) *Equation (1) has a unique solution $\hat{C}_N$ for all large N a.s., defined as*

$$\hat{C}_N = \lim_{t \to \infty} Z^{(t)}$$

where

$$\begin{cases} Z^{(0)} & = I_N \\ Z^{(t+1)} & = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} x_i^* (Z^{(t)})^{-1} x_i\right) x_i x_i^*, \; t \in \mathbb{N}. \end{cases}$$

Robust SCM estimator in the RMT regime

R. Couillet, F. Pascal, J. W. Silverstein, "Robust M-Estimation for Array Processing: A Random Matrix Approach", (submitted to) IEEE Trans. on Information Theory, 2013.

Theorem

Assume the above and consider the fixed-point equation in $Z \in \mathbb{C}^{N\times N}$,

$$Z = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^* Z^{-1} x_i\right) x_i x_i^*. \tag{1}$$

Then,

(I) *Equation* (1) *has a unique solution $\hat{C}_N$ for all large N a.s., defined as*

$$\hat{C}_N = \lim_{t\to\infty} Z^{(t)}$$

where

$$\begin{cases} Z^{(0)} & = I_N \\ Z^{(t+1)} & = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}x_i^*(Z^{(t)})^{-1}x_i\right) x_i x_i^*,\, t \in \mathbb{N}. \end{cases}$$

(II) *Defining $\hat{C}_N = I_N$ when* (1) *does not have a unique solution,*

$$\left\| \phi^{-1}(1)\hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{a.s.}} 0.$$

Robust statistical inference in RMT regime

→ From Theorem 1,

- Weak convergence results on $\hat{S}_N$ propagate to $\hat{C}_N$;
- No eigenvalues and exact separation results propagate to $\hat{S}_N$;
- First order results on spiked models as well, etc.
- Irrelevant of underlying distribution of x, as opposed to the finite N regime

Robust statistical inference in RMT regime

→ From Theorem 1,

- Weak convergence results on $\hat{S}_N$ propagate to $\hat{C}_N$;
- No eigenvalues and exact separation results propagate to $\hat{S}_N$;
- First order results on spiked models as well, etc.
- Irrelevant of underlying distribution of x, as opposed to the finite N regime

→ (Almost) immediate consequence:

- RMT-based statistical estimators using $\hat{S}_N$ can be replaced by identical estimators using $\hat{C}_N$
- e.g. Mestre's DoA estimator

Robust statistical inference in RMT regime

→ From Theorem 1,

- Weak convergence results on $\hat{S}_N$ propagate to $\hat{C}_N$;
- No eigenvalues and exact separation results propagate to $\hat{S}_N$;
- First order results on spiked models as well, etc.
- Irrelevant of underlying distribution of x, as opposed to the finite N regime

→ (Almost) immediate consequence:

- RMT-based statistical estimators using $\hat{S}_N$ can be replaced by identical estimators using $\hat{C}_N$
- e.g. Mestre's DoA estimator

→ Theorem 1 however does not say anything about second order results.

- Current investigation: CLT on linear statistics for $\hat{C}_N$, for x with i.i.d. entries.
- This should provide the asymptotic performance comparison between robust-RMT estimators and traditional RMT estimators.
- So far, it seems that limiting variance depends mostly on C_N, c, $u'(\phi^{-1}(1))$, and the kurtosis of the entries of x.

Robust G-MUSIC estimator

→ Consider the model

$$x_t = \sum_{k=1}^{K} \sqrt{p_k} s(\theta_k) z_{k,t} + \sigma w_t = A_N y_t, \quad A_N \triangleq \begin{bmatrix} S(\Theta) P^{\frac{1}{2}} & \sigma I_N \end{bmatrix}, \text{ with}$$

- $S(\Theta) = [s(\theta_1), \ldots, s(\theta_K)]$ deterministic bounded norm steering vectors,
- $P = \mathrm{diag}(p_1, \ldots, p_K)$ diagonal of powers,
- $y_t = (z_{1,t}, \ldots, z_{K,t}, w_t^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{C}^{N+K}$, signals and noise vector.

Robust G-MUSIC estimator

→ Consider the model

$$x_t = \sum_{k=1}^{K} \sqrt{p_k} s(\theta_k) z_{k,t} + \sigma w_t = A_N y_t, \quad A_N \triangleq \left[S(\Theta) P^{\frac{1}{2}} \quad \sigma I_N \right], \text{ with}$$

- $S(\Theta) = [s(\theta_1), \ldots, s(\theta_K)]$ deterministic bounded norm steering vectors,
- $P = \operatorname{diag}(p_1, \ldots, p_K)$ diagonal of powers,
- $y_t = (z_{1,t}, \ldots, z_{K,t}, w_t^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{C}^{N+K}$, signals and noise vector.

→ From the above results and Mestre's G-MUSIC,

Theorem (Robust G-MUSIC)

Denote $E_W \in \mathbb{C}^{N \times (N-K)}$ the "noise subspace" of C_N, $\hat{e}_k$ the eigenvector of $\hat{C}_N$ with eigenvalue $\hat{\lambda}_k \triangleq \lambda_k(\hat{C}_N)$. Then, as $N, n \to \infty$ and K fixed,

$$\gamma(\theta) - \hat{\gamma}(\theta) \xrightarrow{\text{a.s.}} 0, \quad \gamma(\theta) = s(\theta)^* E_W E_W^* s(\theta), \quad \hat{\gamma}(\theta) = \sum_{i=1}^{N} \beta_i s(\theta)^* \hat{e}_i \hat{e}_i^* s(\theta)$$

and

$$\beta_i = \begin{cases} 1 + \sum_{k=N-K+1}^{N} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_i - \hat{\mu}_k} \right) & , \; i \leqslant N-K \\ -\sum_{k=1}^{N-K} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_i - \hat{\mu}_k} \right) & , \; i > N-K \end{cases}$$

with $\hat{\mu}_1 \leqslant \ldots \leqslant \hat{\mu}_N$ the eigenvalues of $\operatorname{diag}(\hat{\boldsymbol{\lambda}}) - \frac{1}{n}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^{\mathsf{T}}$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^{\mathsf{T}}$.

Results

→ The interest of the above robust-DoA scheme is to:

- handle noise that is "only well-approximated by Gaussian"
- handle model based on bursts of errors on individual antennas
- handle noise distributions with heavier-than-Gaussian tails in radars with distributed antennas (e.g. MIMO radars)

Results

→ The interest of the above robust-DoA scheme is to:

- handle noise that is "only well-approximated by Gaussian"
- handle model based on bursts of errors on individual antennas
- handle noise distributions with heavier-than-Gaussian tails in radars with distributed antennas (e.g. MIMO radars)

→ Some strong limitations:

- cannot handle distributions with heavier-than-Gaussian tails in classical radars
  - this would impose to choose x e.g. elliptically distributed
  - our proof technique collapses here
- cannot handle scale-free detectors/estimators, with $u(s) = 1/s$

Simulation results: The Gaussian noise reference

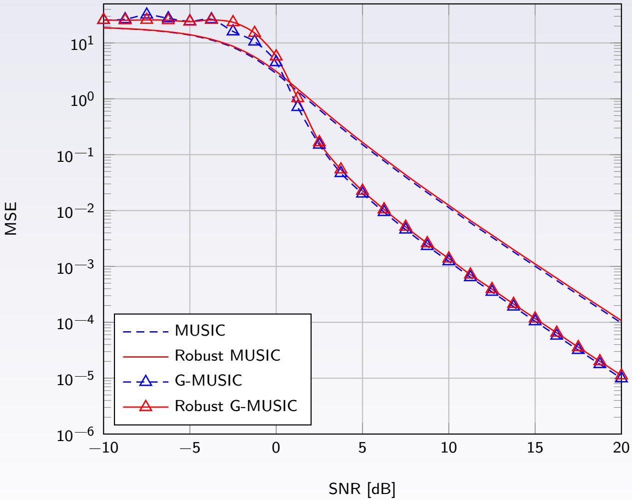


Figure : MSE performance of the various MUSIC estimators for $K = 1$, Gaussian noise, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Simulation results: Close-to-Gaussian noise with i.i.d. Student entries

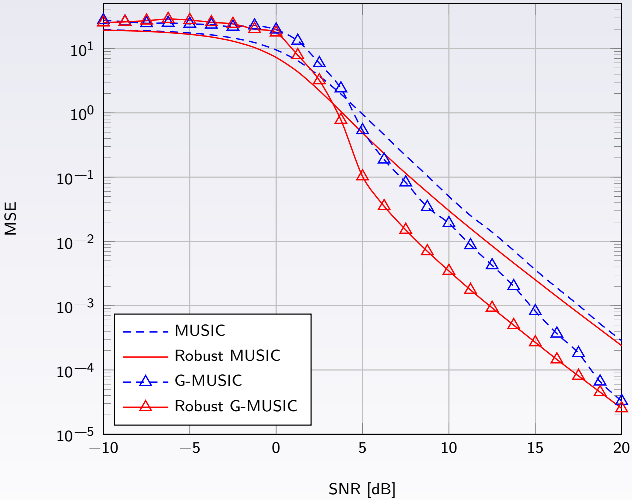


Figure : MSE performance of the various MUSIC estimators for $K = 1$, Student-t noise with $\nu = 5$, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Simulation results: Far-from-Gaussian noise with i.i.d. Student entries

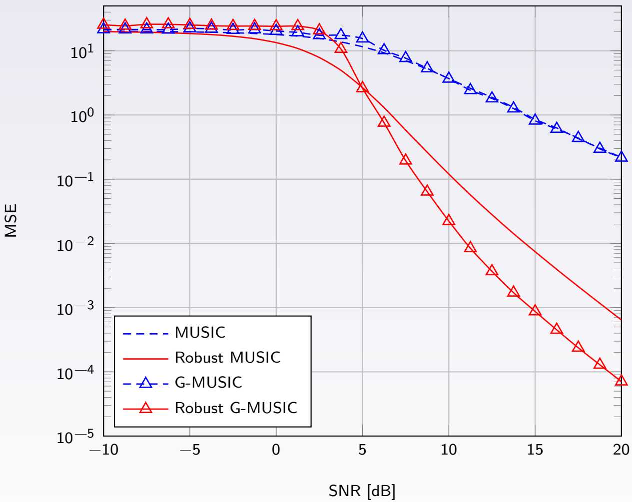


Figure : MSE performance of the various MUSIC estimators for $K = 1$, Student-t noise with $\nu = 2.5$, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Simulation results: Resolution power

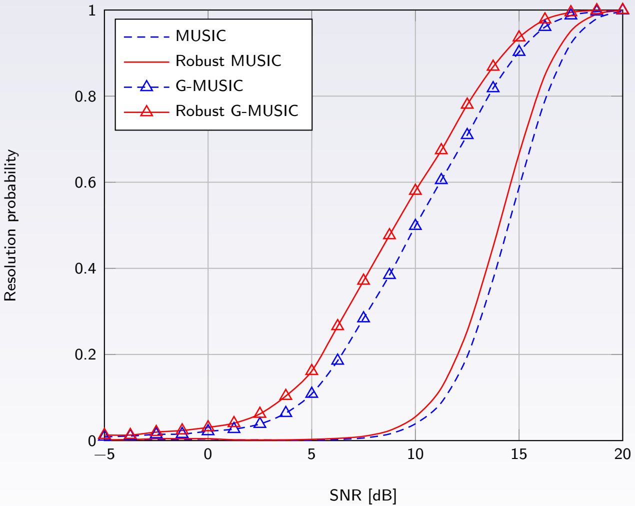


Figure : Resolution performance of the various MUSIC estimators, $\theta_1 = 10^\circ$, $\theta_2 = 15^\circ$, Student-t noise with $\nu = 5$, $N = 10$, and $n = 50$, $u(s) = (1 + \nu')/(s + \nu')$, $\nu' = 0.5$.

Sketch of proof

- Proving existence and uniqueness of a solution $\hat{C}_N$ is not simple.
- We only prove the convergence result here.

Sketch of proof

- Proving existence and uniqueness of a solution $\hat{C}_N$ is not simple.
- We only prove the convergence result here.

→ Take $(d_1, \ldots, d_n)$, $d_i = \frac{1}{N} x_i^* \hat{C}_N^{-1} x_i$ with $\hat{C}_N$ the (almost surely) unique solution:

- We assume $d_1 \leqslant \ldots \leqslant d_n$;
- We also define $D = \operatorname{diag}(u(d_1), \ldots, u(d_n))$.

→ $u(s)$ is non-increasing, so

$$XDX^* \succeq u(d_n) XX^*$$

so that

$$\frac{1}{u(d_n)} \hat{S}_N^{-1} \succeq \hat{C}_N^{-1}$$

and then

$$\frac{1}{u(d_n)} \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n \geqslant d_n$$

from which

$$\phi(d_n) \leqslant \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n$$

→ Proceeding similarly for d_1, and using ϕ non-decreasing, we conclude, for all i

$$\frac{1}{N} x_1^* \hat{S}_N^{-1} x_1 \leqslant \phi(d_1) \leqslant \phi(d_i) \leqslant \phi(d_n) \leqslant \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n$$

Sketch of proof: Convergence

→ It is then possible to show that "the $\frac{1}{N}x_i^*\hat{S}_N^{-1}x_i$ concentrate" as $n \to \infty$, so that

$$\max_{i \leqslant n} |\phi(d_i) - 1| \xrightarrow{\text{a.s.}} 0.$$

→ But $\phi_\infty > 1$ so that ϕ invertible in a neighborhood of 1, and

$$\max_{i \leqslant n} |d_i - \phi^{-1}(1)| \xrightarrow{\text{a.s.}} 0$$

so that

$$\max_{i \leqslant n} \left| u(d_i) - \frac{1}{\phi^{-1}(1)} \right| \xrightarrow{\text{a.s.}} 0$$

(note that $\phi^{-1}(1)u(\phi^{-1}(1)) = 1$)

Sketch of proof: Convergence

→ It is then possible to show that "the $\frac{1}{N} x_i^* \hat{S}_N^{-1} x_i$ concentrate" as $n \to \infty$, so that

$$\max_{i \leqslant n} |\phi(d_i) - 1| \xrightarrow{\text{a.s.}} 0.$$

→ But $\phi_\infty > 1$ so that ϕ invertible in a neighborhood of 1, and

$$\max_{i \leqslant n} |d_i - \phi^{-1}(1)| \xrightarrow{\text{a.s.}} 0$$

so that

$$\max_{i \leqslant n} \left| u(d_i) - \frac{1}{\phi^{-1}(1)} \right| \xrightarrow{\text{a.s.}} 0$$

(note that $\phi^{-1}(1) u(\phi^{-1}(1)) = 1$)

→ We then conclude with

$$\min_{i \leqslant n} \left\{ u(d_i) - \frac{1}{\phi^{-1}(1)} \right\} \frac{1}{n} XX^* \preceq \frac{1}{n} \sum_{i=1}^{n} \left(u(d_i) - \frac{1}{\phi^{-1}(1)} \right) x_i x_i^* \preceq \max_{i \leqslant n} \left\{ u(d_i) - \frac{1}{\phi^{-1}(1)} \right\} \frac{1}{n} XX^*$$

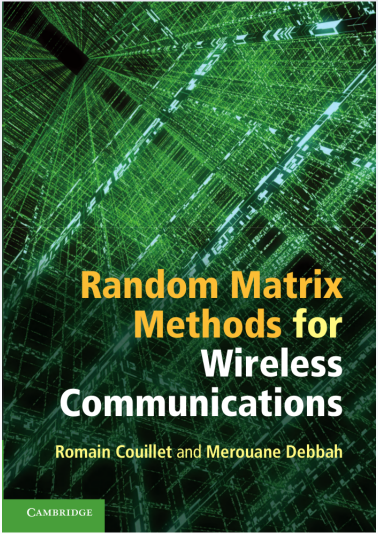
which entails, along with the a.s. boundedness of $\| \frac{1}{n} XX^* \|$,

$$\left\| \hat{C}_N - \frac{1}{\phi^{-1}(1)} \hat{S}_N \right\| \xrightarrow{\text{a.s.}} 0$$

Related biography

- J. T. Kent, D. E. Tyler, “Redescending M-estimates of multivariate location and scatter”, 1991.
- R. A. Maronna, “Robust M-estimators of multivariate location and scatter”, 1976.
- Y. Chitour, F. Pascal, “Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis”, 2008.
- N. El Karoui, “Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond”, 2009.
- R. Couillet, F. Pascal, J. W. Silverstein, “Robust M-Estimation for Array Processing: A Random Matrix Approach”, 2012.
- J. Vinogradova, R. Couillet, W. Hachem, “Statistical Inference in Large Antenna Arrays under Unknown Noise Pattern”, (submitted to) IEEE Transactions on Signal Processing, 2012.
- F. Chapon, R. Couillet, W. Hachem, X. Mestre, “On the isolated eigenvalues of large Gram random matrices with a fixed rank deformation”, (submitted to) Electronic Journal of Probability, 2012, arXiv Preprint 1207.0471.
- R. Couillet, M. Debbah, “Signal Processing in Large Systems: a New Paradigm”, IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 24-39, 2013.
- P. Loubaton, P. Vallet, “Almost sure localization of the eigenvalues in a Gaussian information plus noise model. Application to the spiked models”, Electronic Journal of Probability, 2011.
- P. Vallet, W. Hachem, P. Loubaton, X. Mestre, J. Najim, “On the consistency of the G-MUSIC DOA estimator.” IEEE Statistical Signal Processing Workshop (SSP), 2011.

To know more about all this



The end

Thank you