Future Random Matrix Tools for Large Dimensional Signal Processing

EUSIPCO 2014, Lisbon, Portugal.

Abla KAMMOUN[1] and Romain COUILLET[2]

[1]King's Abdullah University of Technology and Science, Saudi Arabia
[2]SUPELEC, France

September 1st, 2014

High-dimensional data

- Consider n observations $\mathbf{x}_1, \cdots, \mathbf{x}_n$of size N, independent and identically distributed with zero-mean and covariance $\mathbf{C}_N$, i.e, $\mathbb{E}\left[\mathbf{x}_1\mathbf{x}_1^{\mathrm{H}}\right] = C_N$,
- Let $\mathbf{X}_N = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$. The sample covariance estimate $\hat{S}_N$ of $\mathbf{C}_N$ is given by: $\hat{S}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$,
- From the law of large numbers, as $n \to +\infty$,

$$\hat{S}_N \xrightarrow{\text{a.s.}} \mathbf{C}_N.$$

$\longrightarrow$ Convergence in the operator norm

High-dimensional data

- Consider n observations $\mathbf{x}_1, \cdots, \mathbf{x}_n$ of size N, independent and identically distributed with zero-mean and covariance $\mathbf{C}_N$, i.e, $\mathbb{E}\left[\mathbf{x}_1\mathbf{x}_1^{\mathrm{H}}\right] = C_N$,
- Let $\mathbf{X}_N = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$. The sample covariance estimate $\hat{S}_N$ of $\mathbf{C}_N$ is given by:
  $\hat{S}_N = \frac{1}{n}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$,
- From the law of large numbers, as $n \to +\infty$,

$$\hat{S}_N \xrightarrow{\text{a.s.}} \mathbf{C}_N.$$

$\longrightarrow$ Convergence in the operator norm

- In practice, it might be difficult to afford $n \to +\infty$,
  - if $n \gg N$, $\hat{S}_N$ can be sufficiently accurate,
  - if $N/n = \mathcal{O}(1)$, we model this scenario by the following assumption: $N \to +\infty$ and $n \to +\infty$ with $\frac{N}{n} \to c$,
  - Under this assumption, we have pointwise convergence to each element of C_N, i.e,

$$\left(\hat{S}_N\right)_{i,j} \xrightarrow{\text{a.s.}} (C_N)_{i,j}$$

  but $\|S_N - C_N\|$ does not converge to zero.

  $\longrightarrow$ The convergence in the operator norm does not hold.

Illustration

Consider $C_N = I_N$, the spectrum of $\hat{S}_N$ is different from that of C_N

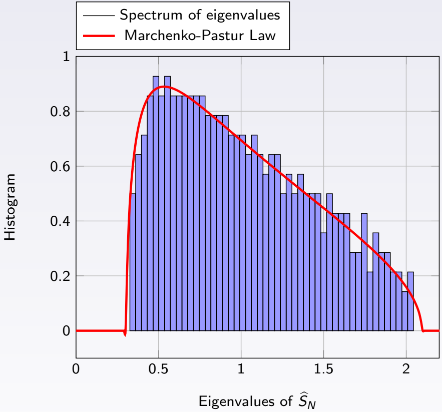


Figure: Spectrum of eigenvalues when $N = 400$ and $n = 2000$

$\longrightarrow$ The asymptotic spectrum can be characterized by the Marchenko-Pastur Law.

Reasons of interest for signal processing

- Scale similarity in array processing applications: large antenna arrays vs limited number of observations,
- Need for detection and estimation based on large dimensional random inputs: subspace methods in array processing.
- The assumption "number of obervations $\gg$ dimension of observation" is no longer 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 $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{N \times n}$ the n samples, taken from

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

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 $\mathbf{X}\mathbf{X}^{\mathsf{H}} = \hat{\mathbf{U}}_S \hat{\mathbf{\Lambda}}_S \hat{\mathbf{U}}_S^{\mathsf{H}} + \hat{\mathbf{U}}_W \hat{\mathbf{\Lambda}}_W \hat{\mathbf{U}}_W^{\mathsf{H}}$.

Reasons of interest for signal processing

- Scale similarity in array processing applications: large antenna arrays vs limited number of observations,
- Need for detection and estimation based on large dimensional random inputs: subspace methods in array processing.
- The assumption "number of obervations $\gg$ dimension of observation" is no longer 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 $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{N \times n}$ the n samples, taken from

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

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 $\mathbf{X}\mathbf{X}^{\mathsf{H}} = \hat{\mathbf{U}}_S\hat{\mathbf{\Lambda}}_S\hat{\mathbf{U}}_S^{\mathsf{H}} + \hat{\mathbf{U}}_W\hat{\mathbf{\Lambda}}_W\hat{\mathbf{U}}_W^{\mathsf{H}}$.
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

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

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

Outline

Outline

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 is to the Cauchy transform as the characteristic functin is to the Fourier transform.

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 is to the Cauchy transform as the characteristic functin is to the Fourier transform.

Equivalence $F \leftrightarrow m_F$

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

Stieltjes transform of a Hermitian matrix

- Let $\mathbf{X}$ be a $N \times N$ random matrix. Denote by dF^X the empirical measure of its eigenvalues $\lambda_1, \cdots, \lambda_N$, i.e, $dF^X = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$. The Stieltjes transform of $\mathbf{X}$ denoted by $m_{\mathbf{X}} = m_F$ is the stieltjes transform of its empirical measure:

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

- The Stieltjes transform of a random matrix is the trace of the resolvent matrix $\mathbf{Q}(z) = (\mathbf{X} - z\mathbf{I}_N)^{-1}$. The resolvent matrix plays a key role in the derivation of many of the results of random matrix theory.
- For compactly supported F, $m_F(z)$ is linked to the moments $M_k = \mathbb{E}\frac{1}{N}\operatorname{tr}\mathbf{X}^k$,

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

- m_F is defined in general on $\mathbb{C}_+$ but exists everywhere outside the support of F.

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

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

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

The Marčenko-Pastur density

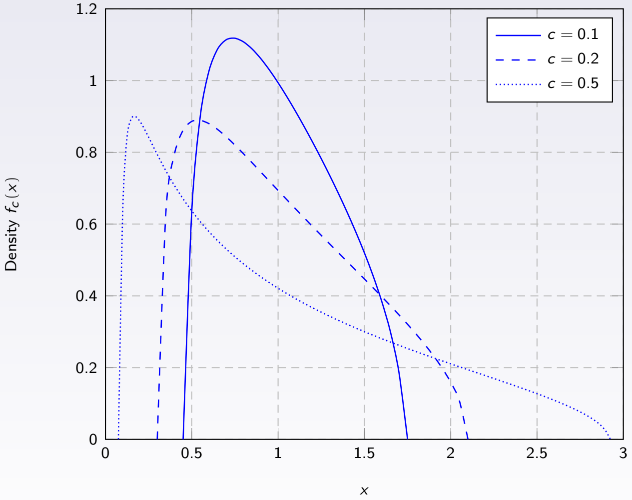


Figure: Marčenko-Pastur law for different limit ratios $c = \lim_{N\to\infty} N/n$.

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

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

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,

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

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.

Asymptotic results involving 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 $\mathbf{Y}_N = \frac{1}{\sqrt{n}}\mathbf{X}_N\mathbf{C}_N^{\frac{1}{2}}$*, where* $\mathbf{X}_N \in \mathbb{C}^{n\times N}$ *has i.i.d entries of mean* 0 *and variance* 1*. Consider the regime* $n, N \to +\infty$ *with* $\frac{N}{n} \to c$*. Let* $\hat{\underline{m}}_N$ *be the Stieltjes transform associated to* $\mathbf{X}_N\mathbf{X}_N^*$*. Then,* $\hat{\underline{m}}_N - \underline{m}_N \to 0$ *almost surely for all* $z \in \mathbb{C}\backslash\mathbb{R}_+$*, where* $\underline{m}_N(z)$ *is the unique solution in the set* $\{z \in \mathbb{C}_+, \underline{m}_N(z) \in \mathbb{C}_+\}$ *to:*

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

Asymptotic results involving 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 $\mathbf{Y}_N = \frac{1}{\sqrt{n}}\mathbf{X}_N\mathbf{C}_N^{\frac{1}{2}}$, where $\mathbf{X}_N \in \mathbb{C}^{n\times N}$ has i.i.d entries of mean 0 and variance 1. Consider the regime $n, N \to +\infty$ with $\frac{N}{n} \to c$. Let $\hat{\underline{m}}_N$ be the Stieltjes transform associated to $\mathbf{X}_N\mathbf{X}_N^*$. Then, $\hat{\underline{m}}_N - \underline{m}_N \to 0$ almost surely for all $z \in \mathbb{C}\backslash\mathbb{R}_+$, where $\underline{m}_N(z)$ is the unique solution in the set $\{z \in \mathbb{C}_+, \underline{m}_N(z) \in \mathbb{C}_+\}$ to:

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

- in general, no explicit expression for $\underline{F}_N$, the distribution whose Stietljes transform is $\underline{m}_N(z)$.
- The theorem above characterizes also the Stieltjes transform of $\mathbf{B}_N = \mathbf{X}_N^{\mathrm{H}}\mathbf{X}_N$ denoted by m_N,

$$m_N = c\underline{m}_N + (c-1)\frac{1}{z}$$

This gives access to the spectrum of the sample covariance matrix model of $\mathbf{x}$, when $\mathbf{y}_i = \mathbf{C}_N^{\frac{1}{2}}\mathbf{x}_i$, $\mathbf{x}_i$ i.i.d., $\mathbf{C}_N = E[\mathbf{y}\mathbf{y}^{\mathrm{H}}]$.

Getting F' from m_F

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

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

where m_F is (up to now) only defined on $\mathbb{C}^+$.

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 $F^C(x) = \frac{1}{3}\mathbf{1}_{x \leqslant 1} + \frac{1}{3}\mathbf{1}_{x \leqslant 3} + \frac{1}{3}\mathbf{1}_{x \leqslant K}$ and let $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{Z}_N^{\mathrm{H}}\mathbf{Z}_N\mathbf{C}_N^{\frac{1}{2}}$ with $F^{B_N} \to F$, then

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

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

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

Spectrum of the sample covariance matrix

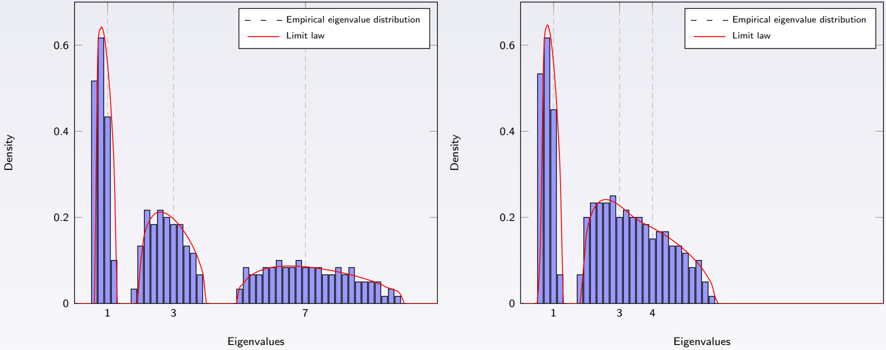


Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{Z}_N^{\mathsf{H}}\mathbf{Z}_N\mathbf{C}_N^{\frac{1}{2}}$, $N = 3000$, $n = 300$, with $\mathbf{C}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

Outline

Support of a distribution

The support of a density f is the closure of the set $\{x, f(x) \neq 0\}$.
For instance the support of the marčenko-Pastur law is $\left[(1-\sqrt{c})^2, (1+\sqrt{c})^2\right]$.

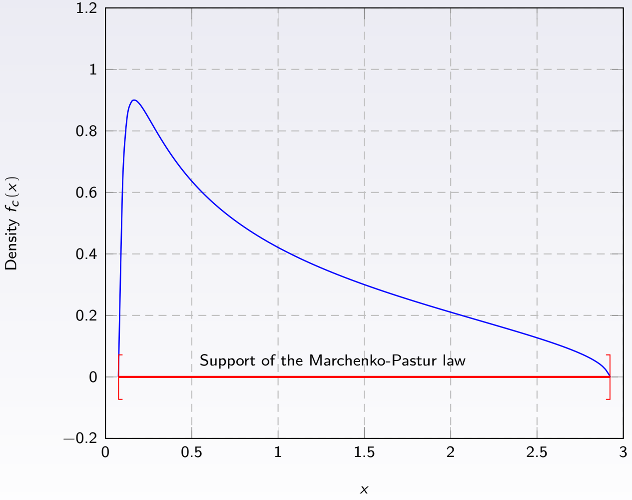


Figure: Marčenko-Pastur law for different limit ratios $c = 0.5$.

Extreme eigenvalues

- Limiting spectral results are insufficient to infer about the location of extreme eigenvalues.
- Example: Consider $dF_N(x) = \frac{1}{N}\sum_{k=1}^{N} \delta_{a_k}$. Then, $dF_N^0 = \frac{N-1}{N} dF_N + \frac{1}{N}\delta_{A_N}(x)$ and dF_N with $A_N \geqslant a_N$ satisfy:

$$dF_N - dF_N^0 \Rightarrow 0.$$

- However, the supports of F_N and F_{N_0} differ by the mass A_N.

Question: How is the behaviour of the extreme eigenvalues of random covariance matrices?

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, unit variance and infinite fourth order. Let $\mathbf{C}_N \in \mathbb{C}^{N \times N}$ be nonrandom and bounded in norm. Let $\underline{m}_N$ be the unique solution in $\mathbb{C}_+$ of

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

Let F_N be the distribution associated to the Stieltjes transform $m_N(z)$. Consider $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}\mathbf{C}_N^{\frac{1}{2}}$. We know that $F^{\mathbf{B}_N} - F_N$ converge weakly to zero. Choose $N_0 \in \mathbb{N}$ and $[a, b]$, $a > 0$, outside the support of F_N for all $N \geqslant N_0$. Denote $\mathcal{L}_N$ the set of eigenvalues of $\mathbf{B}_N$. Then,

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

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*: $\mathbf{C}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{C}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{C}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ 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*: $\mathbf{C}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{C}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{C}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ 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*: $\mathbf{C}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{C}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{C}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ 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

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

  and the generally correlation model where each column of $\mathbf{X}$ has correlation $\mathbf{R}_i$.

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

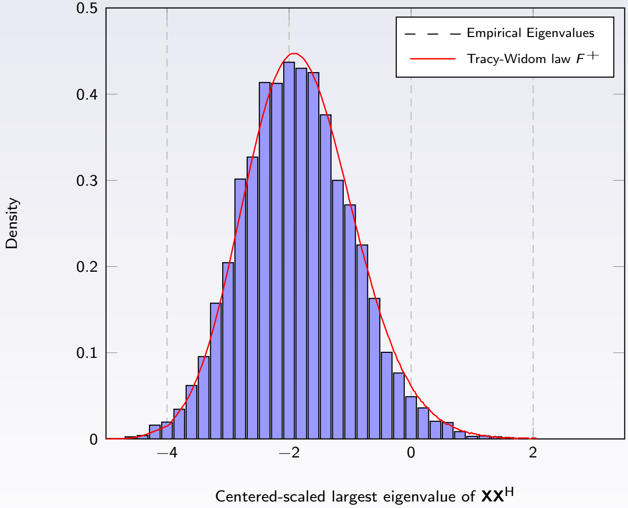


Figure: Distribution of $N^{\frac{2}{3}} c^{-\frac{1}{2}} (1+\sqrt{c})^{-\frac{4}{3}} \left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the distribution of X^+ (distributed as Tracy-Widom law) for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{XX}^{\mathsf{H}}$. Empirical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

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

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

Spiked models

- We consider n independent observations $\mathbf{x}_1, \cdots, \mathbf{x}_n$ of size N,
- The correlation structure is in general "white + low rank",

$$\mathbb{E}\left[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}\right] = \mathbf{I} + \mathbf{P}$$

where $\mathbf{P}$ is of low rank,
- Objective: to infer the eigenvalues and/or the eigenvectors of $\mathbf{P}$

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 $\mathbf{B}_N = \frac{1}{n}(\mathbf{I}+\mathbf{P})^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}(\mathbf{I}+\mathbf{P})^{\frac{1}{2}}$*, where* $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *has i.i.d., zero mean and unit variance entries, and* $\mathbf{P}_N \in \mathbb{R}^{N\times N}$ *with eigenvalues given by:*

$$\mathrm{eig}(\mathbf{P}) = \mathit{diag}(\omega_1,\ldots,\omega_K,\underbrace{0,\ldots,\ldots,0}_{N-K})$$

with $\omega_1 > \ldots > \omega_K > -1$*,* $c = \lim_N N/n$*. Let* $\lambda_1,\cdots,\lambda_N$ *be the eigenvalues of* $\mathbf{B}_N$*. We then have*

- *if* $\omega_j > \sqrt{c}$*,* $\lambda_j \xrightarrow{\text{a.s.}} 1+\omega_j + c\frac{1+\omega_j}{\omega_j}$ *(i.e. beyond the Marčenko–Pastur bulk!)*
- *if* $\omega_j \in (0,\sqrt{c}]$*,* $\lambda_j \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$ *(i.e. right-edge of the Marčenko–Pastur bulk!)*
- *if* $\omega_j \in [-\sqrt{c},0)$*,* $\lambda_j \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$ *(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 $\mathbf{B}_N = \frac{1}{n}(\mathbf{I}+\mathbf{P})^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}(\mathbf{I}+\mathbf{P})^{\frac{1}{2}}$*, where* $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *has i.i.d., zero mean and unit variance entries, and* $\mathbf{P}_N \in \mathbb{R}^{N\times N}$ *with eigenvalues given by:*

$$\mathrm{eig}(\mathbf{P}) = \mathit{diag}(\omega_1,\ldots,\omega_K,\underbrace{0,\ldots,\ldots,0}_{N-K})$$

with $\omega_1 > \ldots > \omega_K > -1$, $c = \lim_N N/n$*. Let* $\lambda_1,\cdots,\lambda_N$ *be the eigenvalues of* $\mathbf{B}_N$*. We then have*

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

Illustration of spiked models

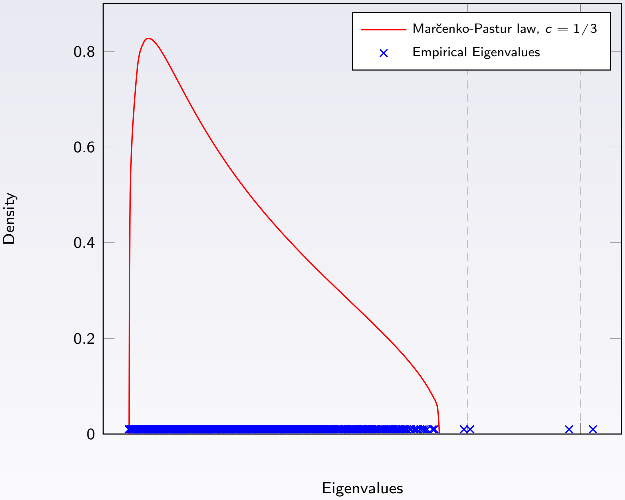


Figure: Eigenvalues of $\mathbf{B}_N = \frac{1}{n}(\mathbf{P}+\mathbf{I})^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N{}^{\mathsf{H}}(\mathbf{P}+\mathbf{I})^{\frac{1}{2}}$, where $\omega_1 = \omega_2 = 1$ and $\omega_3 = \omega_4 = 2$ Dimensions: $N = 500$, $n = 1500$.

Illustration of spiked models

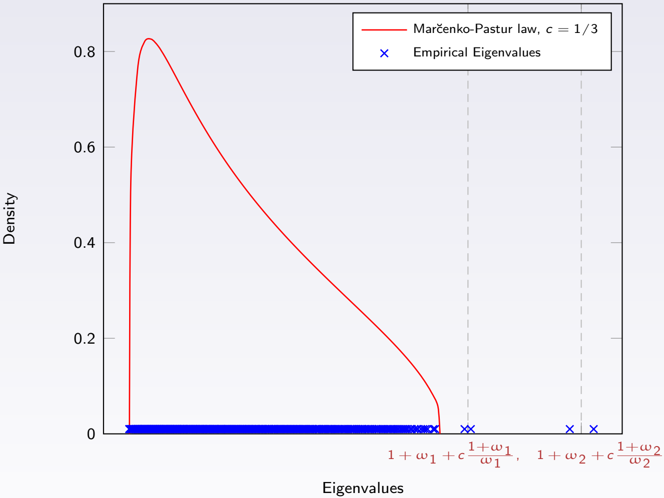


Figure: Eigenvalues of $\mathbf{B}_N = \frac{1}{n}(\mathbf{P}+\mathbf{I})^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N{}^{\mathsf{H}}(\mathbf{P}+\mathbf{I})^{\frac{1}{2}}$, where $\omega_1 = \omega_2 = 1$ and $\omega_3 = \omega_4 = 2$ Dimensions: $N = 500$, $n = 1500$.

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

Sketch of the proof

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

Sketch of the proof

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

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

Sketch of the proof

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

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

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

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

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

Sketch of the proof

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

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

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

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

  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).

Sketch of the proof

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

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

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

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

  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 x_k satisfy $x_k m(x_k) + (1+\omega_k)\omega_k^{-1} = 0$.
- replacing $m(x)$, this is finally:

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

Comments on the result

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

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.

Outline

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

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

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

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

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

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

  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

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

- 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^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.

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

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

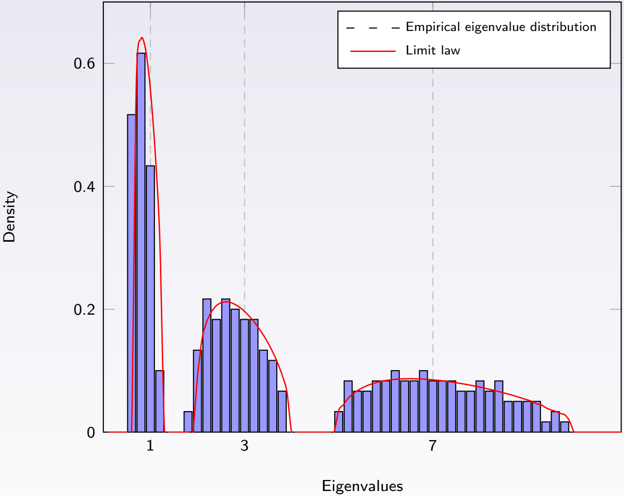


Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{C}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7.

Another way to determine the spectrum: inverse function method

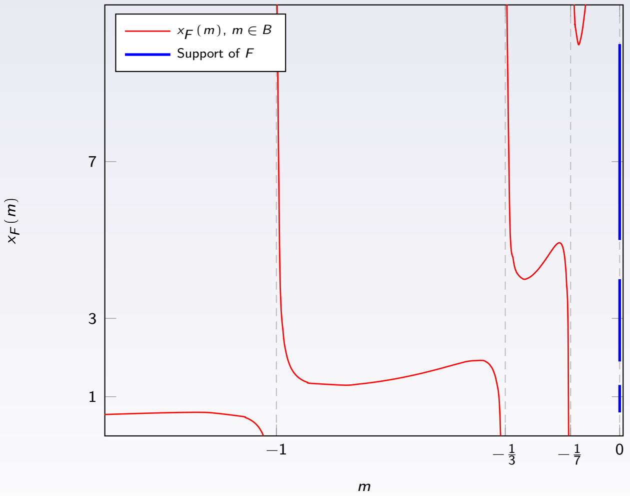


Figure: Stieltjes transform of $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{C}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7. The support of F is read on the vertical axis, 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, unit variance, and $\mathbf{C}_N$ be diagonal such that $F^{\mathbf{C}_N} \Rightarrow F^C$, as $n, N \to \infty$, $N/n \to c$, where F^C has K masses in $t_1, \ldots, t_K$ with multiplicity $n_1, \ldots, n_K$ respectively. Then the l.s.d. of $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N^{\frac{1}{2}}$ 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 $(Q = K)$ or not $(Q < K)$ 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 $(Q = K)$ or not $(Q < K)$ 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 $m_1, \ldots, m_K$ 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.

Exact eigenvalue separation

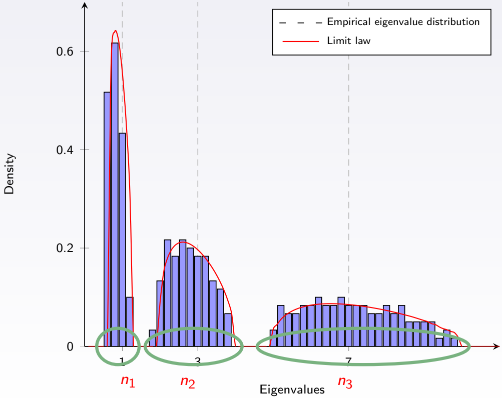
Z. D. Bai, J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.

- Recall that 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.

- Recall that 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,



Eigeninference: Introduction of the problem

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

$$\hat{\mathbf{C}}_N = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathsf{H}}$$

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

Eigeninference: Introduction of the problem

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

$$\hat{\mathbf{C}}_N = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathsf{H}}$$

  is an n-consistent estimator of $\mathbf{C}_N = E[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.
- 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,

$$\hat{\mathbf{C}}_N = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathsf{H}}$$

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

$$\hat{\mathbf{C}}_N = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathsf{H}}$$

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

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

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

$$G_1(\hat{\mathbf{C}}_N) - \alpha_n^{-1} \log \det(\mathbf{C}_N) \to 0$$

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

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

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

$$G_1(\hat{\mathbf{C}}_N) - \alpha_n^{-1}\log\det(\mathbf{C}_N) \to 0$$

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 $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}\mathbf{C}_N^{\frac{1}{2}}$, where $F^{\mathbf{C}_N}$ 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 $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N^{\frac{1}{2}}$, where $F^{\mathbf{C}_N}$ 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 $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N^{\frac{1}{2}}$, where $F^{\mathbf{C}_N}$ 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 $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N^{\frac{1}{2}}$.
- 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 $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}\mathbf{C}_N^{\frac{1}{2}}$.
- 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.

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 $\mathbf{B}_N = \frac{1}{n}\mathbf{C}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N^{\frac{1}{2}}$*, with* $\mathbf{X}_N \in \mathbb{C}^{N\times n}$*, i.i.d. with entries of zero mean, unit variance, and* $\mathbf{C}_N \in \mathbb{R}^{N\times N}$ *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*

$$\underline{m}_{\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N\mathbf{X}_N}(\mu) = 0$$

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 $\underline{\mathbf{B}}_N = \frac{1}{n}\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N\mathbf{X}_N$,

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

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 $\underline{\mathbf{B}}_N = \frac{1}{n}\mathbf{X}_N^{\mathsf{H}}\mathbf{C}_N\mathbf{X}_N$,

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

  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.

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

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

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

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

  or equivalently

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

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

Reminders of complex analysis

- 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,*

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

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,

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

Complex integration

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

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

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

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

Complex integration

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

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

- 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,

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

Complex integration

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

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

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

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

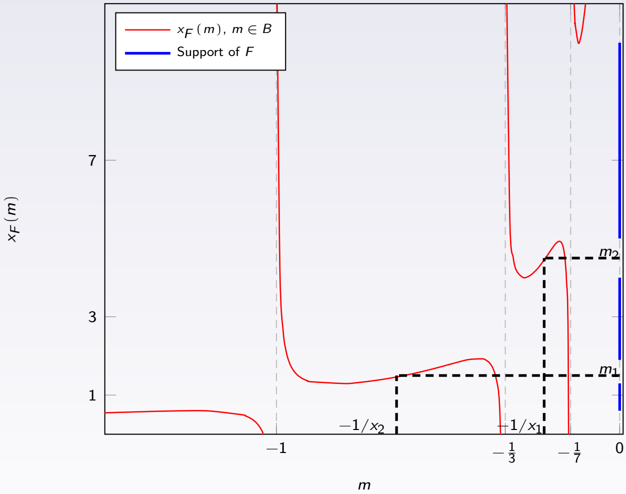
- When the system dimensions are large,

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

- Dominated convergence arguments then show

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

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 $\hat{\underline{m}}_N(z) = 0$.
- 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

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

  - 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

Outline

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 $\mathbf{W}$ 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 $F_{\mathbf{YY}^\mathsf{H}}$ 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 $t_1 = \sum_{k=1}^{N} |h_k|^2 + \sigma^2$

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 $F_{\mathbf{YY}^H}$ 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 $\mathbf{YY}^H$ is therefore asymptotically, as $N, n \to \infty$, $N/n \to c$,
  - if $\mathcal{H}_0$,

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

  - if $\mathcal{H}_1$,

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

  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,

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

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

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

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

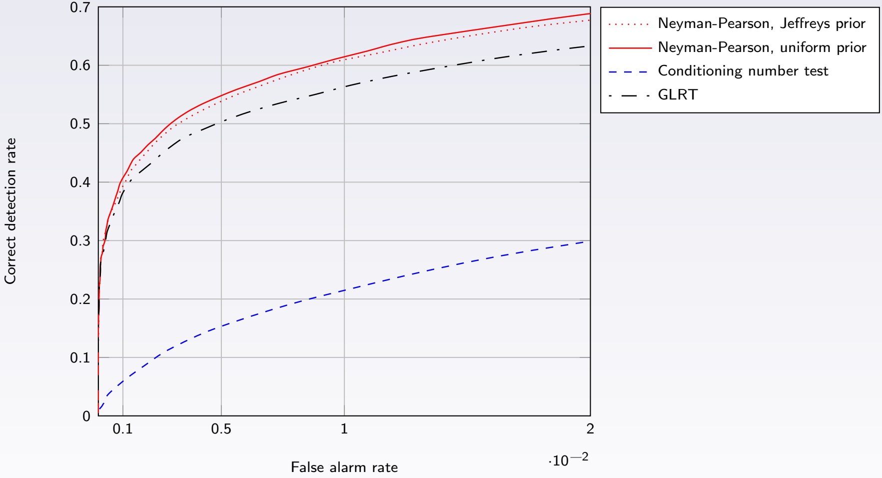


Figure: ROC curve for *a priori* unknown σ^2 of the Neyman-Pearson test, conditioning number method and GLRT, $K = 1$, $N = 4$, $M = 8$, $\mathrm{SNR} = 0 \ \mathrm{dB}$. For the Neyman-Pearson test, both uniform and Jeffreys prior, with exponent $\beta = 1$, are provided.

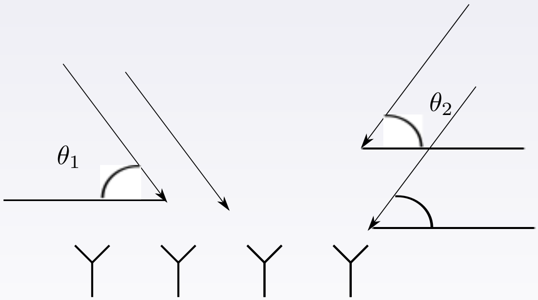
Related biography

- 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

Source localization

A uniform array of M antennas receives signal from K radio sources during n signal snapshots.
Objective: Estimate the arrival angles $\theta_1, \cdots, \theta_K$.



Source Localization using Music Algorithm

We consider the scenario of K sources and N antenna-array capturing n observations:

$$\mathbf{x}_t = \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_{k,t} + \sigma \mathbf{w}_t,\ t = 1, \cdots, n$$

- $\mathbf{A}_N = [\mathbf{a}_N(\theta_1), \cdots, \mathbf{a}_N(\theta_K)]$ with $\mathbf{a}_N(\theta) = \begin{bmatrix} 1 \\ e^{\imath\pi \sin\theta} \\ \cdots \\ e^{\imath(N-1)\pi\sin\theta} \end{bmatrix}$
- σ^2 is the noise variance and is set 1 for simplicity,
- Objective: infer $\theta_1, \cdots, \theta_K$ from the n observations
- Let $\mathbf{X}_N = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$, then,

$$\mathbf{X} = \mathbf{AS} + \mathbf{W} = [\mathbf{A} \quad \mathbf{I}_N] \begin{bmatrix} \mathbf{S} \\ \mathbf{W} \end{bmatrix}$$

- If K is finite while $n, N \to +\infty$, the model correponds to the spiked covariance model.
- MUSIC Algorithm: Let $\mathbf{\Pi}$ be the orthogonal projection matrix on the span of $\mathbf{AA}^*$ and $\mathbf{\Pi}^\perp = \mathbf{I}_N - \mathbf{\Pi}$ (orthogonal projector on the noise subspace). Angles $\theta_1, \cdots, \theta_K$ are the unique ones verifying

$$\eta(\theta) \triangleq \mathbf{a}_N(\theta)^* \mathbf{\Pi} \mathbf{a}_N(\theta) = 0$$

Traditional MUSIC algorithm

- Traditional MUSIC algorithm: Angles are estimated as local minima of:

$$\mathbf{a}_N(\theta)^* \hat{\mathbf{\Pi}} \mathbf{a}_N(\theta)$$

  where $\hat{\mathbf{\Pi}}$ is the orthogonal projection matrix on the eigenspace associated to the K largest eigenvalues of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^*$
- It is well-known that this estimator is consistent when $n \to +\infty$ with K, N fixed,
- We consider the case of K finite $\longrightarrow$ spiked covariance model
- What happens when $n, N \to +\infty$?

Asymptotic behaviour of the traditional MUSIC (1)

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

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

Asymptotic behaviour of the traditional MUSIC (1)

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

- 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 $\mathbf{Z} = [\mathbf{S}^{\mathsf{T}}\ \mathbf{W}^{\mathsf{T}}]^{\mathsf{T}}$ to recover (up to one row) the generic spiked model

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

Asymptotic behaviour of the traditional MUSIC (1)

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

- 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 $\mathbf{Z} = [\mathbf{S}^{\mathsf{T}}\ \mathbf{W}^{\mathsf{T}}]^{\mathsf{T}}$ to recover (up to one row) the generic spiked model

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

- Reminder: If x eigenvalue of $\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathsf{H}}$ 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.

Asymptotic behaviour of the traditional MUSIC (2)

→ Recall the MUSIC approach: we want to estimate

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

Asymptotic behaviour of the traditional MUSIC (2)

→ 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:

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

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

Asymptotic behaviour of the traditional MUSIC (2)

→ 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:

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

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:

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

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

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

Asymptotic behaviour of the traditional MUSIC (3)

- 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

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

Asymptotic behaviour of the traditional MUSIC (3)

- 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

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

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

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

Therefore,

$$\hat{\eta}(\theta) = \mathbf{a}(\theta)^{\mathsf{H}}\hat{\boldsymbol{\Pi}}\mathbf{a}(\theta) \xrightarrow{\text{a.s.}} \mathbf{a}(\theta)\mathbf{a}(\theta)^{\mathsf{H}} - \sum_{i=1}^{K}\frac{1 - c\omega_i^{-2}}{1 + c\omega_i^{-1}}\mathbf{a}(\theta)^{\mathsf{H}}\mathbf{u}_i\mathbf{u}_i^{\mathsf{H}}\mathbf{a}(\theta)$$

Improved G-MUSIC

Recall that:

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

→ 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

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

with

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

Improved G-MUSIC

Recall that:

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

→ 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

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

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!

Simulation results

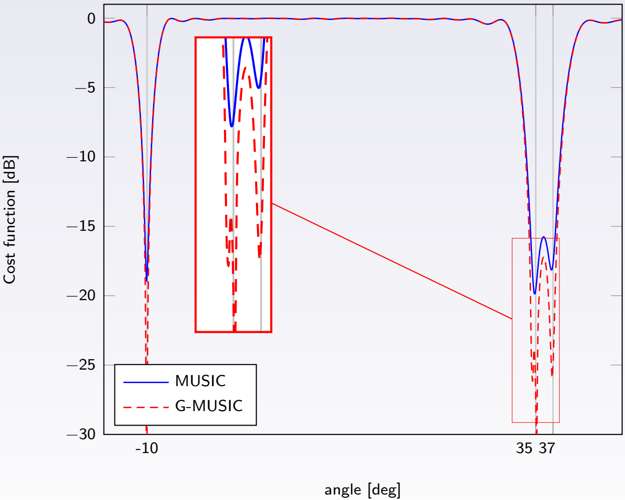


Figure: MUSIC against G-MUSIC for DoA detection of $K = 3$ signal sources, $N = 20$ sensors, $M = 150$ samples, SNR of 10 dB. Angles of arrival of $10°$, $35°$, and $37°$.

Outline of the tutorial

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

Outline of the tutorial

Outline of the tutorial

Outline of the tutorial

Outline

Outline

Covariance estimation and sample covariance matrices

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

$\longrightarrow$ 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. $x \in \mathbb{C}^N$.

Covariance estimation and sample covariance matrices

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

$\longrightarrow$ 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 $\hat{\theta} = f(\hat{S}_N)$.

Covariance estimation and sample covariance matrices

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

$\longrightarrow$ 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

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

    $\to$ 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
    $\to$ One therefore expects $\hat{\theta}$ to closely approximate θ for all finite n.

Covariance estimation and sample covariance matrices

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

$\longrightarrow$ 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

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

    $\to$ 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
    $\to$ 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,

$$\|\hat{S}_N - C_N\| \not\to 0 \text{ as } N, n \to \infty, \ N/n \to c > 0, \text{ so that in general } |\hat{\theta} - \theta| \not\to 0$$

    $\to$ 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.
    $\to$ 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 strictly increasing $u(s)$ (inconsistent with Huber's estimate)
  - 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 $\hat{S}_N$.

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

Robust Estimation and RMT

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

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

Robust Estimation and RMT

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

- $x = A_N w$, w 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 $\hat{C}_N$?

Setting and assumptions

- **Assumptions:**
  - Take $x_1, \ldots, x_n \in \mathbb{C}^N$ "elliptical-like" random vectors, i.e. $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ where
    - $\tau_1, \ldots, \tau_n \in \mathbb{R}^+$ random or deterministic with $\frac{1}{n}\sum_{i=1}^n \tau_i \xrightarrow{\text{a.s.}} 1$
    - $w_1, \ldots, w_n \in \mathbb{C}^N$ random independent with $w_i/\sqrt{N}$ uniformly distributed over the unit-sphere
    - $C_N \in \mathbb{C}^{N\times N}$ deterministic, with $C_N \succ 0$ and $\limsup_N \|C_N\| < \infty$
  - We denote $c_N \triangleq N/n$ and consider the growth regime $c_N \to c \in (0,1)$.

Setting and assumptions

- **Assumptions:**
  - Take $x_1, \ldots, x_n \in \mathbb{C}^N$ "elliptical-like" random vectors, i.e. $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ where
    - $\tau_1, \ldots, \tau_n \in \mathbb{R}^+$ random or deterministic with $\frac{1}{n}\sum_{i=1}^n \tau_i \xrightarrow{\text{a.s.}} 1$
    - $w_1, \ldots, w_n \in \mathbb{C}^N$ random independent with $w_i/\sqrt{N}$ uniformly distributed over the unit-sphere
    - $C_N \in \mathbb{C}^{N\times N}$ deterministic, with $C_N \succ 0$ and $\limsup_N \|C_N\| < \infty$
  - We denote $c_N \triangleq N/n$ and consider the growth regime $c_N \to c \in (0,1)$.
- **Maronna's estimator of scatter:** (almost sure) unique solution to

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

where u satisfies

(i) $u : [0,\infty) \to (0,\infty)$ nonnegative continuous and non-increasing

(ii) $\phi : x \mapsto xu(x)$ increasing and bounded with $\lim_{x\to\infty} \phi(x) \triangleq \phi_\infty > 1$

(iii) $\phi_\infty < c_+^{-1}$.

Setting and assumptions

- **Assumptions:**
  - Take $x_1, \ldots, x_n \in \mathbb{C}^N$ "elliptical-like" random vectors, i.e. $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ where
    - $\tau_1, \ldots, \tau_n \in \mathbb{R}^+$ random or deterministic with $\frac{1}{n}\sum_{i=1}^n \tau_i \xrightarrow{\text{a.s.}} 1$
    - $w_1, \ldots, w_n \in \mathbb{C}^N$ random independent with $w_i/\sqrt{N}$ uniformly distributed over the unit-sphere
    - $C_N \in \mathbb{C}^{N\times N}$ deterministic, with $C_N \succ 0$ and $\limsup_N \|C_N\| < \infty$
  - We denote $c_N \triangleq N/n$ and consider the growth regime $c_N \to c \in (0,1)$.
- **Maronna's estimator of scatter:** (almost sure) unique solution to

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

where u satisfies

(i) $u : [0,\infty) \to (0,\infty)$ nonnegative continuous and non-increasing
(ii) $\phi : x \mapsto xu(x)$ increasing and bounded with $\lim_{x\to\infty}\phi(x) \triangleq \phi_\infty > 1$
(iii) $\phi_\infty < c_+^{-1}$.

- **Additional technical assumption:** Let $\nu_n \triangleq \frac{1}{n}\sum_{i=1}^n \delta_{\tau_i}$. For each $a > b > 0$, a.s.

$$\limsup_{t\to\infty} \frac{\limsup_n \nu_n((t,\infty))}{\phi(at) - \phi(bt)} = 0.$$

$\to$ Controls relative speed of the tail of ν_n versus the flattening speed of $\phi(x)$ as $x \to \infty$.
Examples:
  - $\tau_i < M$ for each i. In this case, $\nu_n((t,\infty)) = 0$ a.s. for $t > M$.
  - For $u(t) = (1+\alpha)/(\alpha+t)$, $\alpha > 0$, and τ_i i.i.d., it is sufficient to have $E[\tau_1^{1+\varepsilon}] < \infty$.

Heuristic approach

- **Major issues with $\hat{C}_N$:**
  - Defined implicitly
  - Sum of non-independent rank-one matrices from vectors $\sqrt{u(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i)}x_i$ ($\hat{C}_N$ depends on all x_j's).

Heuristic approach

- **Major issues with $\hat{C}_N$:**
  - Defined implicitly
  - Sum of non-independent rank-one matrices from vectors $\sqrt{u(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i)}x_i$ ($\hat{C}_N$ depends on all x_j's).
- **But there is some hope:**
  - First remark: we can work with $C_N = I_N$ without generality restriction!
  - Denote

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

$\longrightarrow$ Then intuitively, $\hat{C}_{(j)}$ and x_j are only "weakly" dependent.

Heuristic approach

- **Major issues with $\hat{C}_N$:**
  - Defined implicitly
  - Sum of non-independent rank-one matrices from vectors $\sqrt{u(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i)}x_i$ ($\hat{C}_N$ depends on all x_j's).
- **But there is some hope:**
  - First remark: we can work with $C_N = I_N$ without generality restriction!
  - Denote

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

    $\longrightarrow$ Then intuitively, $\hat{C}_{(j)}$ and x_j are only "weakly" dependent.
  - We expect in particular (highly non-rigorous but intuitive!!):

$$\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i \simeq \tau_i\frac{1}{N}\operatorname{tr}\hat{C}_{(i)}^{-1} \simeq \tau_i\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}.$$

Heuristic approach

- **Major issues with $\hat{C}_N$:**
  - Defined implicitly
  - Sum of non-independent rank-one matrices from vectors $\sqrt{u(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i)}x_i$ ($\hat{C}_N$ depends on all x_j's).
- **But there is some hope:**
  - First remark: we can work with $C_N = I_N$ without generality restriction!
  - Denote

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

    $\longrightarrow$ Then intuitively, $\hat{C}_{(j)}$ and x_j are only "weakly" dependent.
  - We expect in particular (highly non-rigorous but intuitive!!):

$$\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i \simeq \tau_i\frac{1}{N}\operatorname{tr}\hat{C}_{(i)}^{-1} \simeq \tau_i\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}.$$

- **Our heuristic approach:**
  - Rewrite $\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i$ as $f(\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i)$ for some function f (later called g^{-1})
  - Deduce that

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

  - Use $\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i \simeq \tau_i\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}$ to get

$$\hat{C}_N \simeq \frac{1}{n}\sum_{i=1}^{n}(u\circ f)\left(\tau_i\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}\right)x_ix_i^*$$

  - Use random matrix results to find a limiting value γ for $\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}$, and conclude

$$\hat{C}_N \simeq \frac{1}{n}\sum_{i=1}^{n}(u\circ f)(\tau_i\gamma)x_ix_i^*.$$

Heuristic approach in detail: f and γ

- **Determination of f:** Recall the identity $(A + tvv^*)^{-1}v = A^{-1}/(1 + tv^*A^{-1}v)$. Then

$$\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i = \frac{\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i}{1 + c_N u(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i)\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i}$$

so that

$$\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i = \frac{\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i}{1 - c_N\phi(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i)}.$$

Now the function $g : x \mapsto x/(1 - c_N\phi(x))$ is monotonous increasing (we use the assumption $\phi_\infty < c^{-1}$!), hence, with $f = g^{-1}$,

$$\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i = g^{-1}\left(\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i\right).$$

Heuristic approach in detail: f and γ

- **Determination of γ:** From previous calculus, we expect

$$\hat{C}_N \simeq \frac{1}{n}\sum_{i=1}^n (u\circ g^{-1})\left(\tau_i \frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}\right) x_i x_i^* \simeq \frac{1}{n}\sum_{i=1}^n (u\circ g^{-1})(\tau_i\gamma)\, x_i x_i^*.$$

Hence

$$\gamma \simeq \frac{1}{N}\operatorname{tr}\hat{C}_N^{-1} \simeq \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}\sum_{i=1}^n (u\circ g^{-1})(\tau_i\gamma)\,\tau_i w_i w_i^*\right)^{-1}.$$

Since τ_i are independent of w_i and γ deterministic, this is a Bai-Silverstein model

$$\frac{1}{n}WDW^*,\ W=[w_1,\ldots,w_n],\ D=\operatorname{diag}(D_{ii})=u\circ g^{-1}(\tau_i\gamma).$$

And we have:

$$\gamma \simeq \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}WDW^*\right)^{-1} = m_{\frac{1}{n}WDW^*}(0) \simeq \left(0+\int \frac{t(u\circ g^{-1})(t\gamma)}{1+c(u\circ g^{-1})(t\gamma)m_{\frac{1}{n}WDW^*}(0)}\nu_N(dt)\right)^{-1}$$

$$= \left(\frac{1}{n}\sum_{i=1}^n \frac{\tau_i(u\circ g^{-1})(\tau_i\gamma)}{1+c\tau_i(u\circ g^{-1})(\tau_i\gamma)m_{\frac{1}{n}WDW^*}(0)}\right)^{-1}.$$

Since $\gamma \simeq m_{\frac{1}{n}WDW^*}(0)$, this defines γ as a solution of a fixed-point equation:

$$\gamma = \left(\frac{1}{n}\sum_{i=1}^n \frac{\tau_i(u\circ g^{-1})(\tau_i\gamma)}{1+c\tau_i(u\circ g^{-1})(\tau_i\gamma)\gamma}\right)^{-1}.$$

Main result

R. Couillet, F. Pascal, J. W. Silverstein, “The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples”, (submitted to) Elsevier Journal of Multivariate Analysis.

Theorem (Asymptotic Equivalence)

Under the assumptions defined earlier, we have

$$\left\| \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{a.s.}} 0, \text{ where } \hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma) x_i x_i^*$$

$v(x) = (u \circ g^{-1})(x)$, $\psi(x) = xv(x)$, $g(x) = x/(1 - c\phi(x))$ *and* $\gamma > 0$ *unique solution of*

$$1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\psi(\tau_i \gamma)}{1 + c\psi(\tau_i \gamma)}.$$

Main result

R. Couillet, F. Pascal, J. W. Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples", (submitted to) Elsevier Journal of Multivariate Analysis.

Theorem (Asymptotic Equivalence)

Under the assumptions defined earlier, we have

$$\left\| \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{a.s.}} 0, \text{ where } \hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma) x_i x_i^*$$

$v(x) = (u \circ g^{-1})(x)$, $\psi(x) = xv(x)$, $g(x) = x/(1 - c\phi(x))$ *and* $\gamma > 0$ *unique solution of*

$$1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\psi(\tau_i \gamma)}{1 + c\psi(\tau_i \gamma)}.$$

- **Remarks:**
  - Th. says: first order substitution of $\hat{C}_N$ by $\hat{S}_N$ allowed for large N, n.
  - It turns out that $v \sim u$ and $\psi \sim \phi$ in general behavior.
  - Corollaries:

$$\max_{1 \leqslant i \leqslant n} \left| \lambda_i(\hat{S}_N) - \lambda_i(\hat{C}_N) \right| \xrightarrow{\text{a.s.}} 0$$

$$\frac{1}{N} \operatorname{tr} (\hat{C}_N - zI_N)^{-1} - \frac{1}{N} \operatorname{tr} (\hat{S}_N - zI_N)^{-1} \xrightarrow{\text{a.s.}} 0$$

$\longrightarrow$ Important feature for detection and estimation.

Main result

R. Couillet, F. Pascal, J. W. Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples", (submitted to) Elsevier Journal of Multivariate Analysis.

Theorem (Asymptotic Equivalence)

Under the assumptions defined earlier, we have

$$\left\| \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{a.s.}} 0, \text{ where } \hat{S}_N \triangleq \frac{1}{n}\sum_{i=1}^{n} v(\tau_i \gamma) x_i x_i^*$$

$v(x) = (u \circ g^{-1})(x)$, $\psi(x) = xv(x)$, $g(x) = x/(1 - c\phi(x))$ *and* $\gamma > 0$ *unique solution of*

$$1 = \frac{1}{n}\sum_{i=1}^{n} \frac{\psi(\tau_i \gamma)}{1 + c\psi(\tau_i \gamma)}.$$

- **Remarks:**
  - Th. says: first order substitution of $\hat{C}_N$ by $\hat{S}_N$ allowed for large N, n.
  - It turns out that $v \sim u$ and $\psi \sim \phi$ in general behavior.
  - Corollaries:

$$\max_{1 \leqslant i \leqslant n} \left| \lambda_i(\hat{S}_N) - \lambda_i(\hat{C}_N) \right| \xrightarrow{\text{a.s.}} 0$$

$$\frac{1}{N}\operatorname{tr}(\hat{C}_N - zI_N)^{-1} - \frac{1}{N}\operatorname{tr}(\hat{S}_N - zI_N)^{-1} \xrightarrow{\text{a.s.}} 0$$

  $\longrightarrow$ Important feature for detection and estimation.
- **Proof:** So far in the tutorial, we do not have a rigorous proof!

Proof

- **Fundamental idea:** Showing that all $\frac{1}{\tau_i}\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i$ converge to the same limit γ.

Proof

- **Fundamental idea:** Showing that all $\frac{1}{\tau_i}\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i$ converge to the same limit γ.
- **Technical trick:** Denote

$$e_i \triangleq \frac{v\left(\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i\right)}{v(\tau_i\gamma)}$$

and relabel terms such that

$$e_1 \leqslant \ldots \leqslant e_n$$

We shall prove that, for each $\ell > 0$,

$$e_1 > 1 - \ell \text{ i.o. and } e_n < 1 + \ell \text{ i.o.}$$

Proof

- **Fundamental idea:** Showing that all $\frac{1}{\tau_i}\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i$ converge to the same limit γ.
- **Technical trick:** Denote

$$e_i \triangleq \frac{v\left(\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i\right)}{v(\tau_i\gamma)}$$

and relabel terms such that

$$e_1 \leqslant \ldots \leqslant e_n$$

We shall prove that, for each $\ell > 0$,

$$e_1 > 1-\ell \text{ i.o. and } e_n < 1+\ell \text{ i.o.}$$

- **Some basic inequalities:** Denoting $d_i \triangleq \frac{1}{\tau_i}\frac{1}{N}x_i^*\hat{C}_{(i)}^{-1}x_i = \frac{1}{N}w_i^*\hat{C}_{(i)}^{-1}w_i$, we have

$$e_j = \frac{v\left(\tau_j\frac{1}{N}w_j^*\left(\frac{1}{n}\sum_{i\neq j}\tau_i v(\tau_i d_i)w_iw_i^*\right)^{-1}w_j\right)}{v(\tau_j\gamma)} = \frac{v\left(\tau_j\frac{1}{N}w_j^*\left(\frac{1}{n}\sum_{i\neq j}\tau_i v(\tau_i\gamma)e_iw_iw_i^*\right)^{-1}w_j\right)}{v(\tau_j\gamma)}$$

$$\leqslant \frac{v\left(\tau_j\frac{1}{N}w_j^*\left(\frac{1}{n}\sum_{i\neq j}\tau_i v(\tau_i\gamma)e_nw_iw_i^*\right)^{-1}w_j\right)}{v(\tau_j\gamma)} = \frac{v\left(\frac{\tau_j}{e_n}\frac{1}{N}w_j^*\left(\frac{1}{n}\sum_{i\neq j}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_j\right)}{v(\tau_j\gamma)}$$

Proof

- **Specialization to e_n:**

$$e_n \leqslant \frac{v\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_i w_i^*\right)^{-1}w_n\right)}{v(\tau_n\gamma)}$$

or equivalently, recalling $\psi(x) = xv(x)$,

$$\frac{\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_i w_i^*\right)^{-1}w_n}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_i w_i^*\right)^{-1}w_n\right)}{\psi(\tau_n\gamma)}.$$

Proof

- **Specialization to e_n:**

$$e_n \leqslant \frac{v\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n\right)}{v(\tau_n\gamma)}$$

or equivalently, recalling $\psi(x) = xv(x)$,

$$\frac{\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n\right)}{\psi(\tau_n\gamma)}.$$

- **Random Matrix results:**
  - By trace lemma, we should have

$$\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n \simeq \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1} \simeq \gamma$$

(by definition of γ as in previous slides). . .

Proof

- **Specialization to e_n:**

$$e_n \leqslant \frac{v\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n\right)}{v(\tau_n\gamma)}$$

or equivalently, recalling $\psi(x) = xv(x)$,

$$\frac{\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n\right)}{\psi(\tau_n\gamma)}.$$

- **Random Matrix results:**
  - By trace lemma, we should have

$$\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n \simeq \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1} \simeq \gamma$$

  (by definition of γ as in previous slides)...
  - **DANGER**: by relabeling, w_n no longer independent of $w_1, \ldots, w_{n-1}$!
  $\Rightarrow$ Broken trace lemma!

Proof

- **Specialization to e_n:**

$$e_n \leqslant \frac{v\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n\right)}{v(\tau_n\gamma)}$$

or equivalently, recalling $\psi(x) = xv(x)$,

$$\frac{\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n\right)}{\psi(\tau_n\gamma)}.$$

- **Random Matrix results:**
  - By trace lemma, we should have

$$\frac{1}{N}w_n^*\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_n \simeq \frac{1}{N}\mathrm{tr}\left(\frac{1}{n}\sum_{i\neq n}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1} \simeq \gamma$$

    (by definition of γ as in previous slides). . .
  - **DANGER**: by relabeling, w_n no longer independent of $w_1,\ldots,w_{n-1}$!
    $\Rightarrow$ Broken trace lemma!
  - **Solution**: uniform convergence result.
    By (higher order) moment bounds, Markov inequality, and Borel Cantelli, for all large n a.s.

$$\max_{1\leqslant j\leqslant n}\left|\frac{1}{N}w_j^*\left(\frac{1}{n}\sum_{i\neq j}\tau_i v(\tau_i\gamma)w_iw_i^*\right)^{-1}w_j - \gamma\right| < \varepsilon.$$

Proof

- **Back to original problem**: For all large n a.s., we then have (using growth of ψ)

$$\frac{\gamma - \varepsilon}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}(\gamma + \varepsilon)\right)}{\psi(\tau_n \gamma)}.$$

Proof

- **Back to original problem**: For all large n a.s., we then have (using growth of ψ)

$$\frac{\gamma - \varepsilon}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}(\gamma + \varepsilon)\right)}{\psi(\tau_n \gamma)}.$$

- **Proof by contradiction**: Assume $e_n > 1 + \ell$ i.o., then on a subsequence $e_n > 1 + \ell$ always and

$$\frac{\gamma - \varepsilon}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{1+\ell}(\gamma + \varepsilon)\right)}{\psi(\tau_n \gamma)}.$$

Proof

- **Back to original problem**: For all large n a.s., we then have (using growth of ψ)

$$\frac{\gamma - \varepsilon}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}(\gamma + \varepsilon)\right)}{\psi(\tau_n \gamma)}.$$

- **Proof by contradiction**: Assume $e_n > 1 + \ell$ i.o., then on a subsequence $e_n > 1 + \ell$ always and

$$\frac{\gamma - \varepsilon}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{1+\ell}(\gamma + \varepsilon)\right)}{\psi(\tau_n \gamma)}.$$

  - **Bounded support for τ_i:** If $0 < \tau_- < \tau_i < \tau_+ < \infty$ for all i, n, then on a subsequence where $\tau_n \to \tau_0$,

$$\underbrace{\frac{\gamma - \varepsilon}{\gamma}}_{\to 1 \text{ as } \varepsilon \to 0} \leqslant \underbrace{\frac{\psi\left(\frac{\tau_0}{1+\ell}(\gamma + \varepsilon)\right)}{\psi(\tau_0 \gamma)}}_{\to \frac{\psi\left(\frac{\tau_0}{1+\ell}\gamma\right)}{\psi(\tau_0 \gamma)} < 1 \text{ as } \varepsilon \to 0}$$

CONTRADICTION!

Proof

- **Back to original problem**: For all large n a.s., we then have (using growth of ψ)

$$\frac{\gamma - \varepsilon}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{e_n}(\gamma+\varepsilon)\right)}{\psi(\tau_n \gamma)}.$$

- **Proof by contradiction**: Assume $e_n > 1+\ell$ i.o., then on a subsequence $e_n > 1+\ell$ always and

$$\frac{\gamma - \varepsilon}{\gamma} \leqslant \frac{\psi\left(\frac{\tau_n}{1+\ell}(\gamma+\varepsilon)\right)}{\psi(\tau_n \gamma)}.$$

  - **Bounded support for τ_i:** If $0 < \tau_- < \tau_i < \tau_+ < \infty$ for all i, n, then on a subsequence where $\tau_n \to \tau_0$,

$$\underbrace{\frac{\gamma - \varepsilon}{\gamma}}_{\to 1 \text{ as } \varepsilon \to 0} \leqslant \underbrace{\frac{\psi\left(\frac{\tau_0}{1+\ell}(\gamma+\varepsilon)\right)}{\psi(\tau_0 \gamma)}}_{\to \frac{\psi\left(\frac{\tau_0}{1+\ell}\gamma\right)}{\psi(\tau_0\gamma)} < 1 \text{ as } \varepsilon \to 0}$$

CONTRADICTION!

  - **Unbounded support for τ_i:** Importance of relative growth of τ_n versus convergence of ψ to ψ_∞.
Proof consists in dividing $\{\tau_i\}$ in two groups: few large ones versus all others.
Sufficient condition:

$$\limsup_{t\to\infty} \frac{\limsup_n \nu_n((t,\infty))}{\phi(at) - \phi(bt)} = 0.$$

Simulations

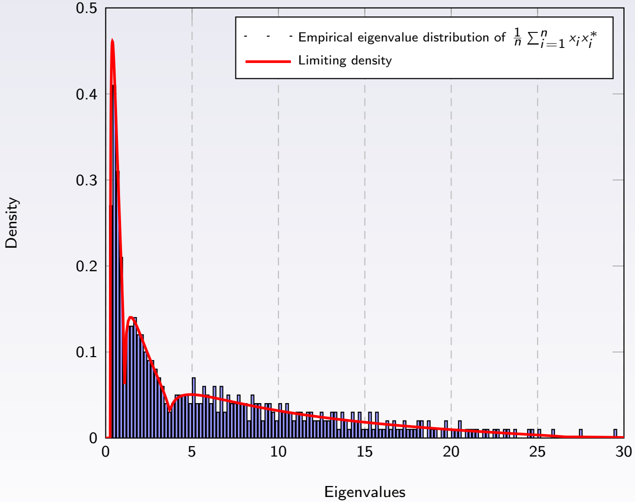


Figure: Histogram of the eigenvalues of $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$ for $n = 2500$, $N = 500$, $C_N = \mathrm{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

Simulations

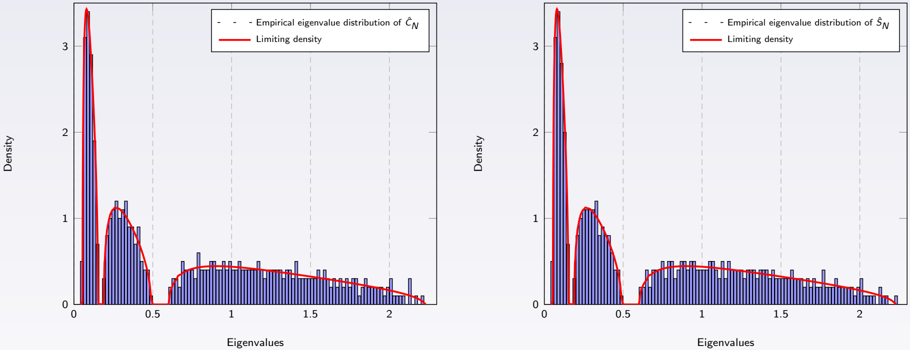


Figure: Histogram of the eigenvalues of $\hat{C}_N$ (left) and $\hat{S}_N$ (right) for $n = 2500$, $N = 500$, $C_N = \mathrm{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

Simulations

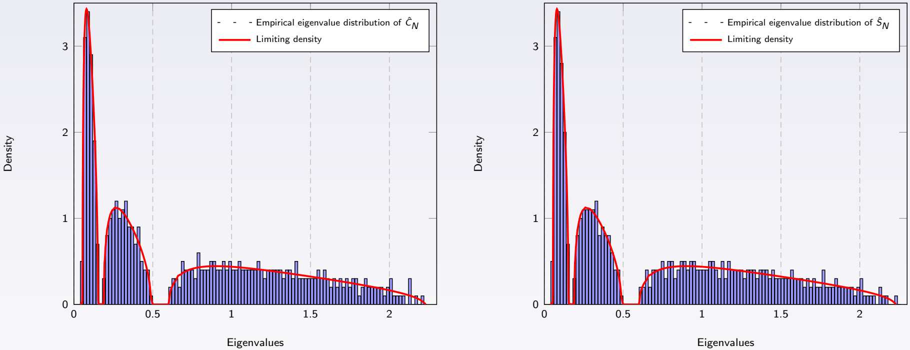


Figure: Histogram of the eigenvalues of $\hat{C}_N$ (left) and $\hat{S}_N$ (right) for $n = 2500$, $N = 500$, $C_N = \mathrm{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

- **Remark/Corollary:** Spectrum of $\hat{C}_N$ a.s. bounded uniformly on n.

Hint on potential applications

- **Spectrum boundedness:** for impulsive noise scenarios,
  - SCM spectrum grows unbounded
  - robust scatter estimator spectrum remains bounded

$\Rightarrow$ Robust estimators improve spectrum separability (important for many statistical inference techniques seen previously)

Hint on potential applications

- **Spectrum boundedness:** for impulsive noise scenarios,
  - SCM spectrum grows unbounded
  - robust scatter estimator spectrum remains bounded

  ⇒ Robust estimators improve spectrum separability (important for many statistical inference techniques seen previously)
- **Spiked model generalization:** we may expect a generalization to spiked models
  - spikes are swallowed by the bulk in SCM context
  - we expect spikes to re-emerge in robust scatter context

  ⇒ We shall see that we get **even better** than this. . .

Hint on potential applications

- **Spectrum boundedness:** for impulsive noise scenarios,
  - SCM spectrum grows unbounded
  - robust scatter estimator spectrum remains bounded

  $\Rightarrow$ Robust estimators improve spectrum separability (important for many statistical inference techniques seen previously)
- **Spiked model generalization:** we may expect a generalization to spiked models
  - spikes are swallowed by the bulk in SCM context
  - we expect spikes to re-emerge in robust scatter context

  $\Rightarrow$ We shall see that we get **even better** than this...
- **Application scenarios:**
  - Radar detection in impulsive noise (non-Gaussian noise, possibly clutter)
  - Financial data analytics
  - Any application where Gaussianity is too strong an assumption...

Outline

System Setting

- **Signal model:**

$$y_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i = A_i \bar{w}_i$$

$$A_i \triangleq \begin{bmatrix} \sqrt{p_1} a_1 & \ldots & \sqrt{p_L} a_L & \sqrt{\tau_i} I_N \end{bmatrix}, \quad \bar{w}_i \triangleq [s_{1i}, \ldots, s_{Li}, w_i]^{\mathsf{T}}.$$

with $y_1, \ldots, y_n \in \mathbb{C}^N$ satisfying:

1. $\tau_1, \ldots, \tau_n > 0$ random such that $\nu_n \triangleq \frac{1}{n} \sum_{i=1}^{n} \delta_{\tau_i} \to \nu$ weakly and $\int t\nu(dt) = 1$;
2. $w_1, \ldots, w_n \in \mathbb{C}^N$ random independent unitarily invariant $\sqrt{N}$-norm;
3. $L \in \mathbb{N}$, $p_1 \geqslant \ldots \geqslant p_L \geqslant 0$ deterministic;
4. $a_1, \ldots, a_L \in \mathbb{C}^N$ deterministic or random with $A^* A \xrightarrow{\text{a.s.}} \operatorname{diag}(p_1, \ldots, p_L)$ as $N \to \infty$, with $A \triangleq [\sqrt{p_1} a_1, \ldots, \sqrt{p_L} a_L] \in \mathbb{C}^{N \times L}$.
5. $s_{1,1}, \ldots, s_{Ln} \in \mathbb{C}$ independent with zero mean, unit variance.

System Setting

- **Signal model:**

$$y_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i = A_i \bar{w}_i$$

$$A_i \triangleq \left[\sqrt{p_1} a_1 \quad \ldots \quad \sqrt{p_L} a_L \quad \sqrt{\tau_i} I_N \right], \quad \bar{w}_i \triangleq [s_{1i}, \ldots, s_{Li}, w_i]^{\mathsf{T}}.$$

with $y_1, \ldots, y_n \in \mathbb{C}^N$ satisfying:

1. $\tau_1, \ldots, \tau_n > 0$ random such that $\nu_n \triangleq \frac{1}{n} \sum_{i=1}^{n} \delta_{\tau_i} \to \nu$ weakly and $\int t\nu(dt) = 1$;
2. $w_1, \ldots, w_n \in \mathbb{C}^N$ random independent unitarily invariant $\sqrt{N}$-norm;
3. $L \in \mathbb{N}$, $p_1 \geqslant \ldots \geqslant p_L \geqslant 0$ deterministic;
4. $a_1, \ldots, a_L \in \mathbb{C}^N$ deterministic or random with $A^* A \xrightarrow{\text{a.s.}} \operatorname{diag}(p_1, \ldots, p_L)$ as $N \to \infty$, with $A \triangleq [\sqrt{p_1} a_1, \ldots, \sqrt{p_L} a_L] \in \mathbb{C}^{N \times L}$.
5. $s_{1,1}, \ldots, s_{Ln} \in \mathbb{C}$ independent with zero mean, unit variance.

- **Relation to previous model:** If $L = 0$, $y_i = \sqrt{\tau_i} w_i$.
  $\Rightarrow$ Elliptical model with covariance a low-rank (L) perturbation of I_N.
  $\Rightarrow$ We expect a spiked version of previous results.

System Setting

- **Signal model:**

$$y_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i = A_i \bar{w}_i$$

$$A_i \triangleq \begin{bmatrix} \sqrt{p_1} a_1 & \ldots & \sqrt{p_L} a_L & \sqrt{\tau_i} I_N \end{bmatrix}, \quad \bar{w}_i \triangleq [s_{1i}, \ldots, s_{Li}, w_i]^{\mathsf{T}}.$$

with $y_1, \ldots, y_n \in \mathbb{C}^N$ satisfying:

1. $\tau_1, \ldots, \tau_n > 0$ random such that $\nu_n \triangleq \frac{1}{n} \sum_{i=1}^{n} \delta_{\tau_i} \to \nu$ weakly and $\int t \nu(dt) = 1$;
2. $w_1, \ldots, w_n \in \mathbb{C}^N$ random independent unitarily invariant $\sqrt{N}$-norm;
3. $L \in \mathbb{N}$, $p_1 \geqslant \ldots \geqslant p_L \geqslant 0$ deterministic;
4. $a_1, \ldots, a_L \in \mathbb{C}^N$ deterministic or random with $A^* A \xrightarrow{\text{a.s.}} \operatorname{diag}(p_1, \ldots, p_L)$ as $N \to \infty$, with $A \triangleq [\sqrt{p_1} a_1, \ldots, \sqrt{p_L} a_L] \in \mathbb{C}^{N \times L}$.
5. $s_{1,1}, \ldots, s_{Ln} \in \mathbb{C}$ independent with zero mean, unit variance.

- **Relation to previous model:** If $L = 0$, $y_i = \sqrt{\tau_i} w_i$.
  $\Rightarrow$ Elliptical model with covariance a low-rank (L) perturbation of I_N.
  $\Rightarrow$ We expect a spiked version of previous results.
- **Application contexts:**
  - *wireless communications:* signals s_{li} from L transmitters, N-antenna receiver; a_l random i.i.d. channels ($a_l^* a_{l'} \to \delta_{l-l'}$, e.g. $a_l \sim \mathcal{CN}(0, I_N/N)$);
  - *array processing:* L sources emit signals s_{li} at steering angle $a_l = a(\theta_l)$. For ULA,

$$[a(\theta)]_j = N^{-\frac{1}{2}} \exp(2\pi\imath dj \sin(\theta)).$$

Some intuition

- **Signal detection/estimation in impulsive environments:** Two scenarios
  - heavy-tailed noise (elliptical, Gaussian mixtures, etc.)
  - Gaussian noise with spurious impulsions

Some intuition

- **Signal detection/estimation in impulsive environments:** Two scenarios
  - heavy-tailed noise (elliptical, Gaussian mixtures, etc.)
  - Gaussian noise with spurious impulsions
- **Problems expected with SCM:** Respectively,
  - unbounded limiting spectrum, no source separation!
    $\Rightarrow$ Invalidates G-MUSIC
  - isolated eigenvalues due to spikes in *time direction*
    $\Rightarrow$ False alarms induced by noise impulses!

Some intuition

- **Signal detection/estimation in impulsive environments:** Two scenarios
  - heavy-tailed noise (elliptical, Gaussian mixtures, etc.)
  - Gaussian noise with spurious impulsions
- **Problems expected with SCM:** Respectively,
  - unbounded limiting spectrum, no source separation!
    $\Rightarrow$ Invalidates G-MUSIC
  - isolated eigenvalues due to spikes in *time direction*
    $\Rightarrow$ False alarms induced by noise impulses!
- **Our results:** In a spiked model with noise impulsions,
  - whatever noise impulsion type, spectrum of $\hat{C}_N$ remains bounded
  - isolated largest eigenvalues may appear, two classes:
    - isolated eigenvalues due to noise impulses CANNOT exceed a threshold!
    - all isolated eigenvalues beyond this threshold are due to signal

    $\Rightarrow$ Detection criterion: everything above threshold is signal.

Theoretical results

Theorem (Extension to spiked robust model)

Under the same assumptions as in previous section,

$$\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{S}_N \triangleq \frac{1}{n}\sum_{i=1}^{n} v(\tau_i\gamma)A_i\bar{w}_i\bar{w}_i^* A_i^*$$

with γ the unique solution to

$$1 = \int \frac{\psi(t\gamma)}{1 + c\psi(t\gamma)}\nu(dt)$$

and we recall

$$A_i \triangleq \begin{bmatrix} \sqrt{p_1}a_1 & \dots & \sqrt{p_L}a_L & \sqrt{\tau_i}I_N \end{bmatrix}$$
$$\bar{w}_i = [s_{1i}, \dots, s_{Li}, w_i]^{\mathsf{T}}.$$

Theoretical results

Theorem (Extension to spiked robust model)

Under the same assumptions as in previous section,

$$\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{S}_N \triangleq \frac{1}{n}\sum_{i=1}^{n} v(\tau_i\gamma) A_i \bar{w}_i \bar{w}_i^* A_i^*$$

with γ *the unique solution to*

$$1 = \int \frac{\psi(t\gamma)}{1 + c\psi(t\gamma)} \nu(dt)$$

and we recall

$$A_i \triangleq \begin{bmatrix} \sqrt{p_1}a_1 & \dots & \sqrt{p_L}a_L & \sqrt{\tau_i}I_N \end{bmatrix}$$
$$\bar{w}_i = [s_{1i}, \dots, s_{Li}, w_i]^{\mathsf{T}}.$$

- **Remark:** For $L = 0$, $A_i = [0, \dots, 0, I_N]$.
  $\Rightarrow$ Recover previous result $A_i\bar{w}_i$ becomes w_i.

Localization of eigenvalues

Theorem (Eigenvalue localization)

Denote

- u_k *eigenvector of k-th largest eigenvalue of* $AA^* = \sum_{i=1}^{L} p_i a_i a_i^*$
- $\hat{u}_k$ *eigenvector of k-th largest eigenvalue of* $\hat{C}_N$

Also define $\delta(x)$ *unique positive solution to*

$$\delta(x) = c\left(-x + \int \frac{tv_c(t\gamma)}{1+\delta(x)tv_c(t\gamma)}\nu(dt)\right)^{-1}.$$

Further denote

$$p_- \triangleq \lim_{x \downarrow S^+} -c\left(\int \frac{\delta(x)v_c(t\gamma)}{1+\delta(x)tv_c(t\gamma)}\nu(dt)\right)^{-1}, \quad S^+ \triangleq \frac{\phi_\infty(1+\sqrt{c})^2}{\gamma(1-c\phi_\infty)}.$$

Localization of eigenvalues

Theorem (Eigenvalue localization)

Denote

- u_k *eigenvector of* k*-th largest eigenvalue of* $AA^* = \sum_{i=1}^{L} p_i a_i a_i^*$
- $\hat{u}_k$ *eigenvector of* k*-th largest eigenvalue of* $\hat{C}_N$

Also define $\delta(x)$ *unique positive solution to*

$$\delta(x) = c\left(-x + \int \frac{t v_c(t\gamma)}{1+\delta(x) t v_c(t\gamma)} \nu(dt)\right)^{-1}.$$

Further denote

$$p_- \triangleq \lim_{x \downarrow S^+} -c\left(\int \frac{\delta(x) v_c(t\gamma)}{1+\delta(x) t v_c(t\gamma)} \nu(dt)\right)^{-1}, \quad S^+ \triangleq \frac{\phi_\infty(1+\sqrt{c})^2}{\gamma(1-c\phi_\infty)}.$$

Then, if $p_j > p_-$, $\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+$, *otherwise* $\limsup_n \hat{\lambda}_j \leqslant S^+$ *a.s., with* Λ_j *unique positive solution to*

$$-c\left(\delta(\Lambda_j) \int \frac{v_c(\tau\gamma)}{1+\delta(\Lambda_j)\tau v_c(\tau\gamma)} \nu(d\tau)\right)^{-1} = p_j.$$

Simulation

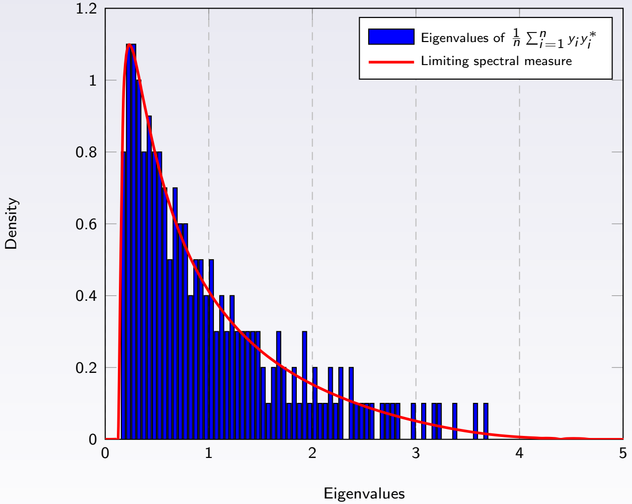


Figure: Histogram of the eigenvalues of $\frac{1}{n}\sum_i y_i y_i^*$ against the limiting spectral measure, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Sudent-t impulsions.

Simulation

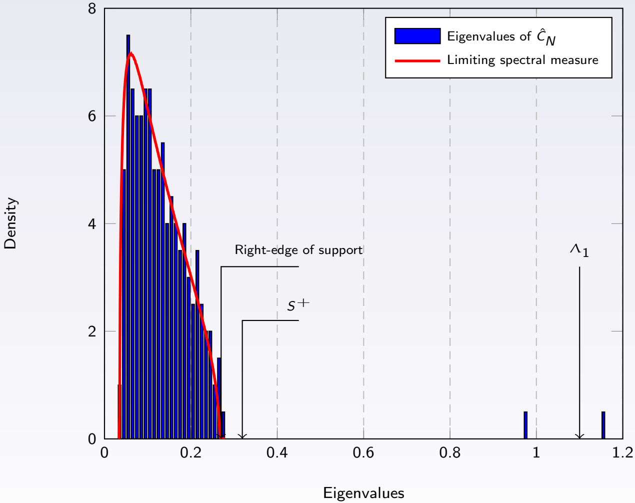


Figure: Histogram of the eigenvalues of $\hat{C}_N$ against the limiting spectral measure, for $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Comments

- **SCM vs robust:** Spikes invisible in SCM in impulsive noise, reborn in robust estimate of scatter.

Comments

- **SCM vs robust:** Spikes invisible in SCM in impulsive noise, reborn in robust estimate of scatter.
- **Largest eigenvalues:**
  - $\lambda_i(\hat{C}_N) > S^+ \Rightarrow$ Presence of a source!
  - $\lambda_i(\hat{C}_N) \in (\sup(\text{Support}), S^+) \Rightarrow$ May be due to a source or to a noise impulse.
  - $\lambda_i(\hat{C}_N) < \sup(\text{Support}) \Rightarrow$ As usual, nothing can be said.

$\Rightarrow$ Induces a natural source detection algorithm.

Eigenvalue and eigenvector projection estimates

- **Two scenarios:**
  - known $\nu = \lim_n \frac{1}{n} \sum_{i=1}^{n} \delta_{\tau_i}$
  - unknown ν

Eigenvalue and eigenvector projection estimates

- **Two scenarios:**
  - known $\nu = \lim_n \frac{1}{n} \sum_{i=1}^{n} \delta_{\tau_i}$
  - unknown ν

Theorem (Estimation under known ν)

1. Power estimation. *For each $p_j > p_-$,*

$$-c\left(\delta(\hat{\lambda}_j)\int \frac{v_c(\tau\gamma)}{1+\delta(\hat{\lambda}_j)\tau v_c(\tau\gamma)}\nu(d\tau)\right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

2. Bilinear form estimation. *For each $a,b \in \mathbb{C}^N$ with $\|a\| = \|b\| = 1$, and $p_j > p_-$*

$$\sum_{k,p_k=p_j} a^* u_k u_k^* b - \sum_{k,p_k=p_j} w_k a^* \hat{u}_k \hat{u}_k^* b \xrightarrow{\text{a.s.}} 0$$

where

$$w_k = \frac{\displaystyle\int \frac{v_c(t\gamma)}{\left(1+\delta(\hat{\lambda}_k)tv_c(t\gamma)\right)^2}\nu(dt)}{\displaystyle\int \frac{v_c(t\gamma)}{1+\delta(\hat{\lambda}_k)tv_c(t\gamma)}\nu(dt)\left(1-\frac{1}{c}\int \frac{\delta(\hat{\lambda}_k)^2 t^2 v_c(t\gamma)^2}{\left(1+\delta(\hat{\lambda}_k)tv_c(t\gamma)\right)^2}\nu(dt)\right)}.$$

Eigenvalue and eigenvector projection estimates

Theorem (Estimation under unknown ν)

1. Purely empirical power estimation. *For each* $p_j > p_-$,

$$-\left(\hat{\delta}(\hat{\lambda}_j)\frac{1}{N}\sum_{i=1}^{n}\frac{v(\hat{\tau}_i\hat{\gamma}_n)}{1+\hat{\delta}(\hat{\lambda}_j)\hat{\tau}_i v(\hat{\tau}_i\hat{\gamma}_n)}\right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

2. Purely empirical bilinear form estimation. *For each* $a,b\in\mathbb{C}^N$ *with* $\|a\|=\|b\|=1$, *and each* $p_j > p_-$,

$$\sum_{k,p_k=p_j} a^* u_k u_k^* b - \sum_{k,p_k=p_j} \hat{w}_k a^* \hat{u}_k \hat{u}_k^* b \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{w}_k = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{v(\hat{\tau}_i\hat{\gamma})}{\left(1+\hat{\delta}(\hat{\lambda}_k)\hat{\tau}_i v(\hat{\tau}_i\hat{\gamma})\right)^2}}{\frac{1}{n}\sum_{i=1}^{n}\frac{v(\hat{\tau}_i\hat{\gamma})}{1+\hat{\delta}(\hat{\lambda}_k)\hat{\tau}_i v(\hat{\tau}_i\hat{\gamma})}\left(1-\frac{1}{N}\sum_{i=1}^{n}\frac{\hat{\delta}(\hat{\lambda}_k)^2\hat{\tau}_i^2 v(\hat{\tau}_i\hat{\gamma})^2}{\left(1+\hat{\delta}(\hat{\lambda}_k)\hat{\tau}_i v(\hat{\tau}_i\hat{\gamma})\right)^2}\right)}$$

$$\hat{\gamma} \triangleq \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N}y_i^*\hat{C}_{(i)}^{-1}y_i, \quad \hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}}\frac{1}{N}y_i^*\hat{C}_{(i)}^{-1}y_i, \quad \hat{\delta}(x) \text{ as } \delta(x) \text{ but for } (\tau_i,\gamma)\to(\hat{\tau}_i,\hat{\gamma}).$$

Application to G-MUSIC

- Assume the model $a_i = a(\theta_i)$ with

$$a(\theta) = N^{-\frac{1}{2}} [\exp(2\pi \imath d j \sin(\theta))]_{j=0}^{N-1}.$$

Application to G-MUSIC

- Assume the model $a_i = a(\theta_i)$ with

$$a(\theta) = N^{-\frac{1}{2}} [\exp(2\pi \imath dj \sin(\theta))]_{j=0}^{N-1}.$$

Corollary (Robust G-MUSIC)

Define $\hat{\eta}_{\rm RG}(\theta)$ *and* $\hat{\eta}_{\rm RG}^{\rm emp}(\theta)$ *as*

$$\hat{\eta}_{\rm RG}(\theta) = 1 - \sum_{k=1}^{|\{j, p_j > p_-\}|} w_k a(\theta)^* \hat{u}_k \hat{u}_k a(\theta)$$

$$\hat{\eta}_{\rm RG}^{\rm emp}(\theta) = 1 - \sum_{k=1}^{|\{j, p_j > p_-\}|} \hat{w}_k a(\theta)^* \hat{u}_k \hat{u}_k a(\theta).$$

Then, for each $p_j > p_-$,

$$\hat{\theta}_j \xrightarrow{\rm a.s.} \theta_j$$

$$\hat{\theta}_j^{\rm emp} \xrightarrow{\rm a.s.} \theta_j$$

where

$$\hat{\theta}_j \triangleq \operatorname{argmin}_{\theta \in \mathcal{R}_j^\kappa} \{\hat{\eta}_{\rm RG}(\theta)\}$$

$$\hat{\theta}_j^{\rm emp} \triangleq \operatorname{argmin}_{\theta \in \mathcal{R}_j^\kappa} \{\hat{\eta}_{\rm RG}^{\rm emp}(\theta)\}.$$

Simulations: Single-shot in elliptical noise

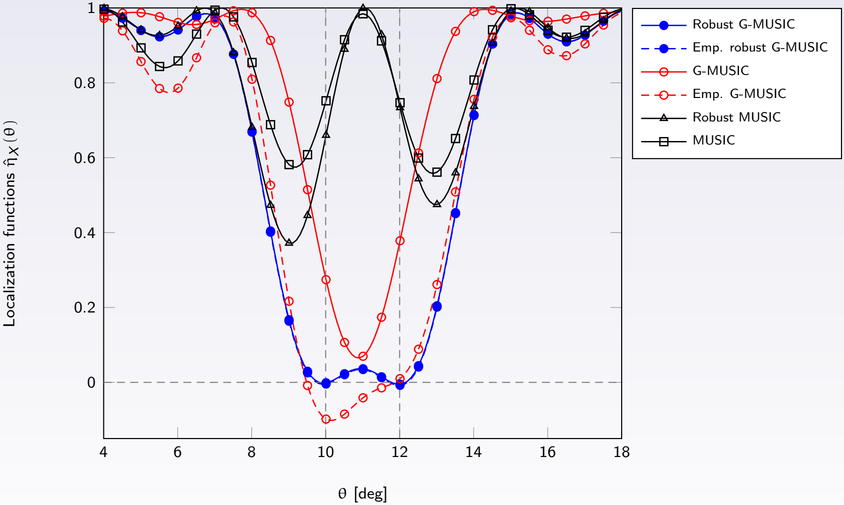


Figure: Random realization of the localization functions for the various MUSIC estimators, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 100$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$. Powers $p_1 = p_2 = 10^{0.5} = 5$ dB.

Simulations: Elliptical noise

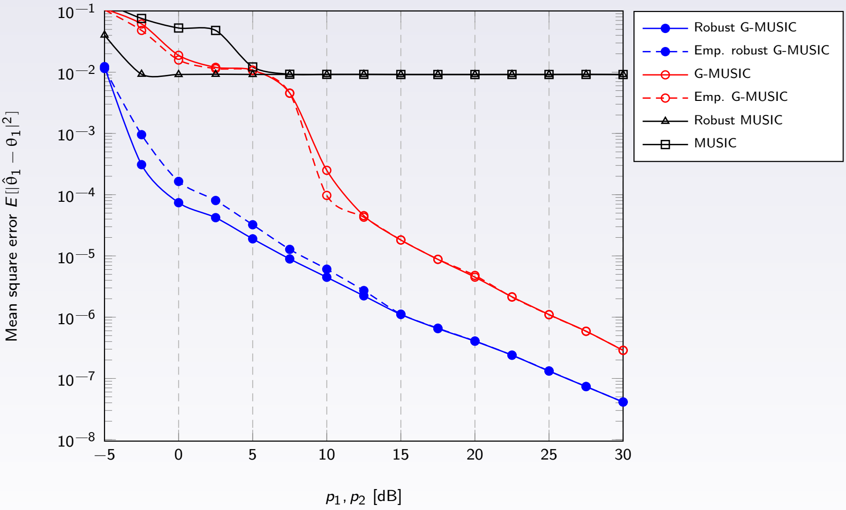


Figure: Means square error performance of the estimation of $\theta_1 = 10^\circ$, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 10$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Simulations: Spurious impulses

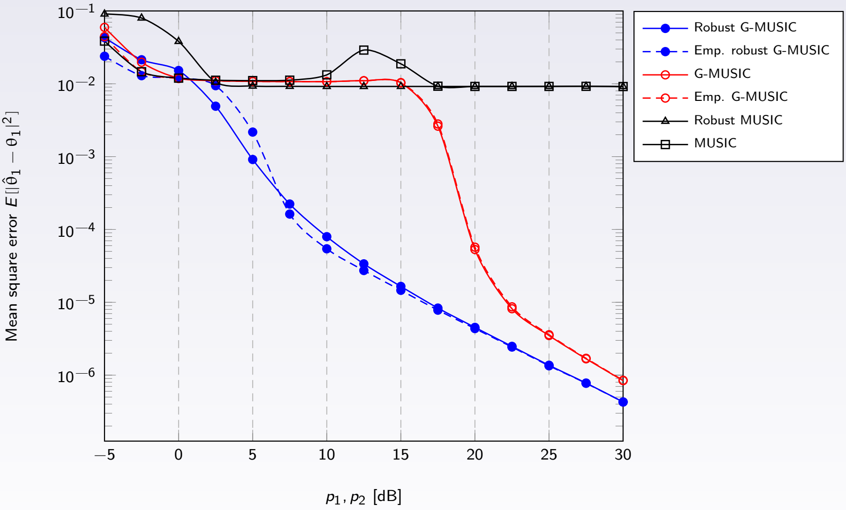


Figure: Means square error performance of the estimation of $\theta_1 = 10^\circ$, with $N = 20$, $n = 100$, two sources at 10° and 12°, sample outlier scenario $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Outline

Context

Ledoit and Wolf, 2004. A well-conditioned estimator for large-dimensional covariance matrices.
Pascal, Chitour, Quek, 2013. Generalized robust shrinkage estimator – Application to STAP data.
Chen, Wiesel, Hero, 2011. Robust shrinkage estimation of high-dimensional covariance matrices.

- **Shrinkage covariance estimation:** For $N > n$ or $N \simeq n$, shrinkage estimator

$$(1-\rho)\frac{1}{n}\sum_{i=1}^{n} x_i x_i^* + \rho I_N, \text{ for some } \rho \in [0,1].$$

  - allows for invertibility, better conditioning
  - ρ may be chosen to minimize an expected error metric

Context

Ledoit and Wolf, 2004. A well-conditioned estimator for large-dimensional covariance matrices.
Pascal, Chitour, Quek, 2013. Generalized robust shrinkage estimator – Application to STAP data.
Chen, Wiesel, Hero, 2011. Robust shrinkage estimation of high-dimensional covariance matrices.

- **Shrinkage covariance estimation:** For $N > n$ or $N \simeq n$, shrinkage estimator

$$(1-\rho)\frac{1}{n}\sum_{i=1}^{n} x_i x_i^* + \rho I_N, \text{ for some } \rho \in [0,1].$$

  - allows for invertibility, better conditioning
  - ρ may be chosen to minimize an expected error metric
- **Limitation of Maronna's estimator:**
  - Maronna and Tyler estimators limited to $N < n$, otherwise do not exist
  - introducing shrinkage in robust estimator cannot do much harm anyhow...

Context

Ledoit and Wolf, 2004. A well-conditioned estimator for large-dimensional covariance matrices.
Pascal, Chitour, Quek, 2013. Generalized robust shrinkage estimator – Application to STAP data.
Chen, Wiesel, Hero, 2011. Robust shrinkage estimation of high-dimensional covariance matrices.

- **Shrinkage covariance estimation:** For $N > n$ or $N \simeq n$, shrinkage estimator

$$(1-\rho)\frac{1}{n}\sum_{i=1}^{n} x_i x_i^* + \rho I_N, \text{ for some } \rho \in [0,1].$$

  - allows for invertibility, better conditioning
  - ρ may be chosen to minimize an expected error metric
- **Limitation of Maronna's estimator:**
  - Maronna and Tyler estimators limited to $N < n$, otherwise do not exist
  - introducing shrinkage in robust estimator cannot do much harm anyhow...
- **Introducing the robust-shrinkage estimator:** The literature proposes two such estimators

$$\hat{C}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{x_i x_i^*}{\frac{1}{N} x_i^* \hat{C}_N^{-1}(\rho) x_i} + \rho I_N, \quad \rho \in \left(\max\{0, \frac{N-n}{N}\}, 1\right] \qquad \text{(Pascal)}$$

$$\check{C}_N(\rho) = \frac{\check{B}_N(\rho)}{\frac{1}{N}\operatorname{tr}\check{B}_N(\rho)}, \quad \check{B}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{x_i x_i^*}{\frac{1}{N} x_i^* \hat{C}_N^{-1}(\rho) x_i} + \rho I_N, \quad \rho \in (0,1] \qquad \text{(Chen)}$$

Main theoretical result

- **Which estimator is better?**

  Having asked to authors of both papers, their estimator was much better than the other, but the arguments we received were quite vague...

Main theoretical result

- **Which estimator is better?**

  Having asked to authors of both papers, their estimator was much better than the other, but the arguments we received were quite vague...

- **Our result:** In the random matrix regime, both estimators tend to be one and the same!

Main theoretical result

- **Which estimator is better?**

  Having asked to authors of both papers, their estimator was much better than the other, but the arguments we received were quite vague...

- **Our result:** In the random matrix regime, both estimators tend to be one and the same!
- **Assumptions:** As before, “elliptical-like” model

$$x_i = \tau_i C_N^{\frac{1}{2}} w_i$$

$\longrightarrow$ This time, C_N cannot be taken I_N (due to $+\rho I_N$)!
$\longrightarrow$ Maronna-based shrinkage is possible but more involved...

Pascal's estimator

Theorem (Pascal's estimator)

For $\varepsilon \in (0, \min\{1, c^{-1}\})$*, define* $\hat{\mathcal{R}}_\varepsilon = [\varepsilon + \max\{0, 1 - c^{-1}\}, 1]$*. Then, as* $N, n \to \infty$*,* $N/n \to c \in (0, \infty)$*,*

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} \left\| \hat{C}_N(\rho) - \hat{S}_N(\rho) \right\| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{C}_N(\rho) = (1 - \rho)\frac{1}{n}\sum_{i=1}^{n} \frac{x_i x_i^*}{\frac{1}{N} x_i^* \hat{C}_N(\rho)^{-1} x_i} + \rho I_N$$

$$\hat{S}_N(\rho) = \frac{1}{\hat{\gamma}(\rho)} \frac{1 - \rho}{1 - (1 - \rho)c} \frac{1}{n} \sum_{i=1}^{n} C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} + \rho I_N$$

and $\hat{\gamma}(\rho)$ *is the unique positive solution to the equation in* $\hat{\gamma}$

$$1 = \frac{1}{N}\sum_{i=1}^{N} \frac{\lambda_i(C_N)}{\hat{\gamma}\rho + (1 - \rho)\lambda_i(C_N)}.$$

Moreover, $\rho \mapsto \hat{\gamma}(\rho)$ *is continuous on* $(0, 1]$*.*

Chen's estimator

Theorem (Chen's estimator)

For $\varepsilon \in (0,1)$, *define* $\check{\mathcal{R}}_\varepsilon = [\varepsilon, 1]$. *Then, as* $N, n \to \infty$, $N/n \to c \in (0, \infty)$,

$$\sup_{\rho \in \check{\mathcal{R}}_\varepsilon} \left\| \check{C}_N(\rho) - \check{S}_N(\rho) \right\| \xrightarrow{\text{a.s.}} 0$$

where

$$\check{C}_N(\rho) = \frac{\check{B}_N(\rho)}{\frac{1}{N} tr \check{B}_N(\rho)}, \quad \check{B}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N} x_i^* \check{C}_N(\rho)^{-1} x_i} + \rho I_N$$

$$\check{S}_N(\rho) = \frac{1-\rho}{1-\rho+T_\rho}\frac{1}{n}\sum_{i=1}^n C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} + \frac{T_\rho}{1-\rho+T_\rho} I_N$$

in which $T_\rho = \rho\check{\gamma}(\rho)F(\check{\gamma}(\rho);\rho)$ *with, for all* $x > 0$,

$$F(x;\rho) = \frac{1}{2}(\rho - c(1-\rho)) + \sqrt{\frac{1}{4}(\rho - c(1-\rho))^2 + (1-\rho)\frac{1}{x}}$$

and $\check{\gamma}(\rho)$ *is the unique positive solution to the equation in* $\check{\gamma}$

$$1 = \frac{1}{N}\sum_{i=1}^N \frac{\lambda_i(C_N)}{\check{\gamma}\rho + \frac{1-\rho}{(1-\rho)c + F(\check{\gamma};\rho)}\lambda_i(C_N)}.$$

Moreover, $\rho \mapsto \check{\gamma}(\rho)$ *is continuous on* $(0,1]$.

Asymptotic Model Equivalence

Theorem (Model Equivalence)

For each $\rho \in (0,1]$, *there exist unique* $\hat{\rho} \in (\max\{0, 1-c^{-1}\}, 1]$ *and* $\check{\rho} \in (0,1]$ *such that*

$$\frac{\hat{S}_N(\hat{\rho})}{\frac{1}{\hat{\gamma}(\hat{\rho})}\frac{1-\hat{\rho}}{1-(1-\hat{\rho})c} + \hat{\rho}} = \check{S}_N(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} + \rho I_N.$$

Besides, $(0,1] \to (\max\{0, 1-c^{-1}\}, 1]$, $\rho \mapsto \hat{\rho}$ *and* $(0,1] \to (0,1]$, $\rho \mapsto \check{\rho}$ *are increasing and onto.*

Asymptotic Model Equivalence

Theorem (Model Equivalence)

For each $\rho \in (0,1]$, *there exist unique* $\hat{\rho} \in (\max\{0, 1-c^{-1}\}, 1]$ *and* $\check{\rho} \in (0,1]$ *such that*

$$\frac{\hat{S}_N(\hat{\rho})}{\frac{1}{\hat{\gamma}(\hat{\rho})}\frac{1-\hat{\rho}}{1-(1-\hat{\rho})c} + \hat{\rho}} = \check{S}_N(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} + \rho I_N.$$

Besides, $(0,1] \to (\max\{0,1-c^{-1}\},1]$, $\rho \mapsto \hat{\rho}$ *and* $(0,1] \to (0,1]$, $\rho \mapsto \check{\rho}$ *are increasing and onto.*

- Up to normalization, both estimators behave the same!
- Both estimators behave the same as an impulsion-free Ledoit-Wolf estimator

Asymptotic Model Equivalence

Theorem (Model Equivalence)

For each $\rho \in (0,1]$, *there exist unique* $\hat{\rho} \in (\max\{0, 1-c^{-1}\}, 1]$ *and* $\check{\rho} \in (0,1]$ *such that*

$$\frac{\hat{S}_N(\hat{\rho})}{\frac{1}{\hat{\gamma}(\hat{\rho})}\frac{1-\hat{\rho}}{1-(1-\hat{\rho})c} + \hat{\rho}} = \check{S}_N(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} + \rho I_N.$$

Besides, $(0,1] \to (\max\{0, 1-c^{-1}\}, 1]$, $\rho \mapsto \hat{\rho}$ *and* $(0,1] \to (0,1]$, $\rho \mapsto \check{\rho}$ *are increasing and onto.*

- Up to normalization, both estimators behave the same!
- Both estimators behave the same as an impulsion-free Ledoit-Wolf estimator
- **About uniformity:** Uniformity over ρ in the theorems is essential to find optimal values of ρ.

Optimal Shrinkage parameter

- Chen sought for a Frobenius norm minimizing ρ but got stuck by implicit nature of $\check{C}_N(\rho)$

Optimal Shrinkage parameter

- Chen sought for a Frobenius norm minimizing ρ but got stuck by implicit nature of $\check{C}_N(\rho)$
- Our results allow for a simplification of the problem for large N, n!
- Model equivalence says only one problem needs be solved.

Optimal Shrinkage parameter

- Chen sought for a Frobenius norm minimizing ρ but got stuck by implicit nature of $\check{C}_N(\rho)$
- Our results allow for a simplification of the problem for large N, n!
- Model equivalence says only one problem needs be solved.

Theorem (Optimal Shrinkage)

For each $\rho \in (0,1]$, *define*

$$\hat{D}_N(\rho) = \frac{1}{N} tr\left(\left(\frac{\hat{C}_N(\rho)}{\frac{1}{N} tr\, \hat{C}_N(\rho)} - C_N\right)^2\right), \quad \check{D}_N(\rho) = \frac{1}{N} tr\left(\left(\check{C}_N(\rho) - C_N\right)^2\right).$$

Denote $D^\star = c\frac{M_2-1}{c+M_2-1}$, $\rho^\star = \frac{c}{c+M_2-1}$, $M_2 = \lim_N \frac{1}{N}\sum_{i=1}^N \lambda_i^2(C_N)$ *and* $\hat{\rho}^\star$, $\check{\rho}^\star$ *unique solutions to*

$$\frac{\hat{\rho}^\star}{\frac{1}{\hat{\gamma}(\hat{\rho}^\star)}\frac{1-\hat{\rho}^\star}{1-(1-\hat{\rho}^\star)c} + \hat{\rho}^\star} = \frac{T_{\check{\rho}^\star}}{1-\check{\rho}^\star + T_{\check{\rho}^\star}} = \rho^\star.$$

Then, letting ε *small enough,*

$$\inf_{\rho \in \hat{\mathcal{R}}_\varepsilon} \hat{D}_N(\rho) \xrightarrow{\text{a.s.}} D^\star, \qquad \inf_{\rho \in \check{\mathcal{R}}_\varepsilon} \check{D}_N(\rho) \xrightarrow{\text{a.s.}} D^\star$$

$$\hat{D}_N(\hat{\rho}^\star) \xrightarrow{\text{a.s.}} D^\star, \qquad \check{D}_N(\check{\rho}^\star) \xrightarrow{\text{a.s.}} D^\star.$$

Estimating $\hat{\rho}^\star$ and $\check{\rho}^\star$

- Theorem only useful if $\hat{\rho}^\star$ and $\check{\rho}^\star$ can be estimated!

Estimating $\hat{\rho}^\star$ and $\check{\rho}^\star$

- Theorem only useful if $\hat{\rho}^\star$ and $\check{\rho}^\star$ can be estimated!
- Careful control of the proofs provide many ways to estimate these.
- Proposition below provides one example.

Estimating $\hat{\rho}^\star$ and $\check{\rho}^\star$

- Theorem only useful if $\hat{\rho}^\star$ and $\check{\rho}^\star$ can be estimated!
- Careful control of the proofs provide many ways to estimate these.
- Proposition below provides one example.

Optimal Shrinkage Estimate

Let $\hat{\rho}_N \in (\max\{0, 1 - c^{-1}\}, 1]$ and $\check{\rho}_N \in (0, 1]$ be solutions (not necessarily unique) to

$$\frac{\hat{\rho}_N}{\frac{1}{N}\operatorname{tr}\hat{C}_N(\hat{\rho}_N)} = \frac{c_N}{\frac{1}{N}\operatorname{tr}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N}\|x_i\|^2}\right)^2\right] - 1}$$

$$\frac{\check{\rho}_N \frac{1}{n}\sum_{i=1}^n \frac{x_i^* \check{C}_N(\check{\rho}_N)^{-1} x_i}{\|x_i\|^2}}{1 - \check{\rho}_N + \check{\rho}_N \frac{1}{n}\sum_{i=1}^n \frac{x_i^* \check{C}_N(\check{\rho}_N)^{-1} x_i}{\|x_i\|^2}} = \frac{c_N}{\frac{1}{N}\operatorname{tr}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{x_i x_i^*}{\frac{1}{N}\|x_i\|^2}\right)^2\right] - 1}$$

defined arbitrarily when no such solutions exist. Then

$$\hat{\rho}_N \xrightarrow{\text{a.s.}} \hat{\rho}^\star, \quad \check{\rho}_N \xrightarrow{\text{a.s.}} \check{\rho}^\star$$

$$\hat{D}_N(\hat{\rho}_N) \xrightarrow{\text{a.s.}} D^\star, \quad \check{D}_N(\check{\rho}_N) \xrightarrow{\text{a.s.}} D^\star.$$

Simulations

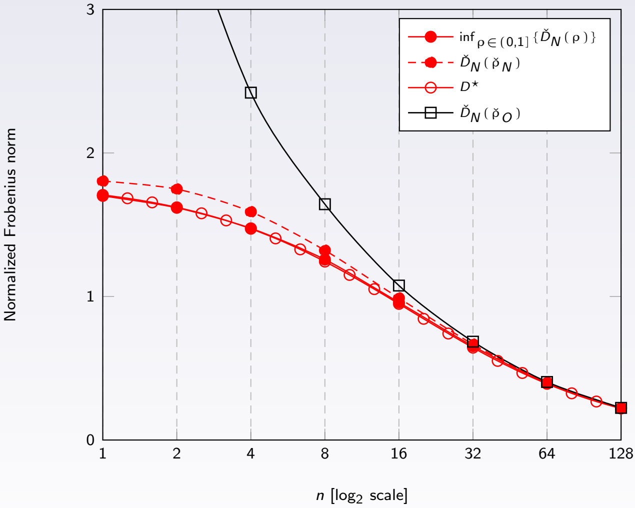


Figure: Performance of optimal shrinkage averaged over 10 000 Monte Carlo simulations, for $N = 32$, various values of n, $[C_N]_{ij} = r^{|i-j|}$ with $r = 0.7$; $\check{\rho}_N$ as above; $\check{\rho}_O$ the clairvoyant estimator proposed in (Chen'11).

Simulations

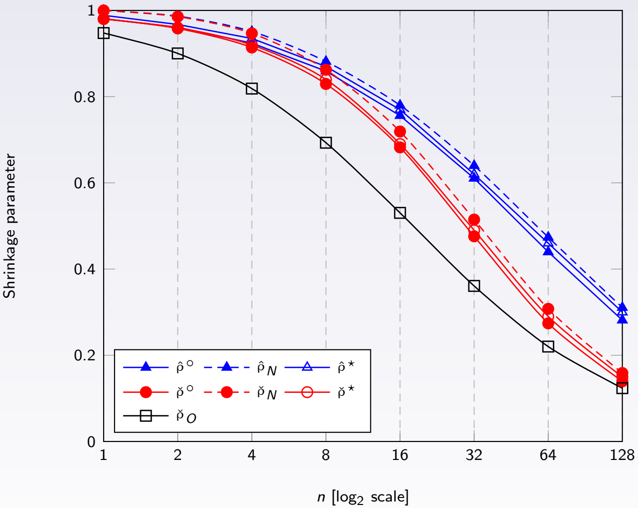


Figure: Shrinkage parameter ρ averaged over 10 000 Monte Carlo simulations, for $N = 32$, various values of n, $[C_N]_{ij} = r^{|i-j|}$ with $r = 0.7$; $\hat{\rho}_N$ and $\check{\rho}_N$ as above; $\check{\rho}_O$ the clairvoyant estimator proposed in (Chen'11); $\hat{\rho}^\circ = \operatorname{argmin}_{\{\rho \in (\max\{0, 1 - c_N^{-1}\}, 1]\}} \{\hat{D}_N(\rho)\}$ and $\check{\rho}^\circ = \operatorname{argmin}_{\{\rho \in (0,1]\}} \{\check{D}_N(\rho)\}$.

Outline

Context

- **Hypothesis testing problem:** Two sets of data
  - Initial pure-noise data: $x_1, \ldots, x_n$, $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ as before.
  - New incoming data y given by:

$$y = \begin{cases} x & , \mathcal{H}_0 \\ \alpha p + x & , \mathcal{H}_1 \end{cases}$$

with $x = \sqrt{\tau} C_N^{\frac{1}{2}} w$, $p \in \mathbb{C}^N$ deterministic known, α unknown.

Context

- **Hypothesis testing problem:** Two sets of data
  - Initial pure-noise data: $x_1, \ldots, x_n$, $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ as before.
  - New incoming data y given by:

$$y = \begin{cases} x & , \mathcal{H}_0 \\ \alpha p + x & , \mathcal{H}_1 \end{cases}$$

with $x = \sqrt{\tau} C_N^{\frac{1}{2}} w$, $p \in \mathbb{C}^N$ deterministic known, α unknown.

- **GLRT detection test:**

$$T_N(\rho) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \Gamma$$

for some detection threshold Γ where

$$T_N(\rho) \triangleq \frac{|y^* \hat{C}_N^{-1}(\rho) p|}{\sqrt{y^* \hat{C}_N^{-1}(\rho) y} \sqrt{p^* \hat{C}_N^{-1}(\rho) p}}.$$

and $\hat{C}_N(\rho)$ defined in previous section.

Context

- **Hypothesis testing problem:** Two sets of data
  - Initial pure-noise data: $x_1, \ldots, x_n$, $x_i = \sqrt{\tau_i} C_N^{\frac{1}{2}} w_i$ as before.
  - New incoming data y given by:

$$y = \begin{cases} x & , \mathcal{H}_0 \\ \alpha p + x & , \mathcal{H}_1 \end{cases}$$

  with $x = \sqrt{\tau} C_N^{\frac{1}{2}} w$, $p \in \mathbb{C}^N$ deterministic known, α unknown.
- **GLRT detection test:**

$$T_N(\rho) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \Gamma$$

for some detection threshold Γ where

$$T_N(\rho) \triangleq \frac{|y^* \hat{C}_N^{-1}(\rho) p|}{\sqrt{y^* \hat{C}_N^{-1}(\rho) y} \sqrt{p^* \hat{C}_N^{-1}(\rho) p}}.$$

and $\hat{C}_N(\rho)$ defined in previous section.

$\longrightarrow$ In fact, originally found to be $\hat{C}_N(0)$ but
  - only valid for $N < n$
  - introducing ρ may bring improved for arbitrary N/n ratios.

Objectives and main results

- **Initial observations:**
  - As $N, n \to \infty$, $N/n \to c > 0$, under $\mathcal{H}_0$,

$$T_N(\rho) \xrightarrow{\text{a.s.}} 0.$$

$\Rightarrow$ Trivial result of little interest!

Objectives and main results

- **Initial observations:**
  - As $N, n \to \infty$, $N/n \to c > 0$, under $\mathcal{H}_0$,

$$T_N(\rho) \xrightarrow{\text{a.s.}} 0.$$

    $\Rightarrow$ Trivial result of little interest!
  - Natural question: for finite N, n and given Γ, find ρ such that

$$P\left(T_N(\rho) > \Gamma\right) = \min$$

Objectives and main results

- **Initial observations:**
  - As $N, n \to \infty$, $N/n \to c > 0$, under $\mathcal{H}_0$,

$$T_N(\rho) \xrightarrow{\text{a.s.}} 0.$$

  $\Rightarrow$ Trivial result of little interest!
  - Natural question: for finite N, n and given Γ, find ρ such that

$$P\left(T_N(\rho) > \Gamma\right) = \min$$

  - Turns out the correct non-trivial object is, for $\gamma > 0$ fixed

$$P\left(\sqrt{N}T_N(\rho) > \gamma\right) = \min$$

Objectives and main results

- **Initial observations:**
  - As $N, n \to \infty$, $N/n \to c > 0$, under $\mathcal{H}_0$,

$$T_N(\rho) \xrightarrow{\text{a.s.}} 0.$$

    $\Rightarrow$ Trivial result of little interest!
  - Natural question: for finite N, n and given Γ, find ρ such that

$$P\left(T_N(\rho) > \Gamma\right) = \min$$

  - Turns out the correct non-trivial object is, for $\gamma > 0$ fixed

$$P\left(\sqrt{N} T_N(\rho) > \gamma\right) = \min$$

- **Objectives:**
  - for each ρ, develop central limit theorem to evaluate

$$\lim_{\substack{N,n\to\infty \\ N/n\to c}} P\left(\sqrt{N} T_N(\rho) > \gamma\right)$$

  - determine limiting minimizing ρ
  - empirically estimate minimizing ρ

What do we need?

CLT over $\hat{C}_N$ statistics

- We know that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$
  $\longrightarrow$ Key result so far!
- What about $\|\sqrt{N}(\hat{C}_N(\rho) - \hat{S}_N(\rho))\|$?

What do we need?

CLT over $\hat{C}_N$ statistics

- We know that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$
  $\longrightarrow$ Key result so far!
- What about $\|\sqrt{N}(\hat{C}_N(\rho) - \hat{S}_N(\rho))\|$?
  $\longrightarrow$ Does not converge to zero!!!

What do we need?

CLT over $\hat{C}_N$ statistics

- We know that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$
  $\longrightarrow$ Key result so far!
- What about $\|\sqrt{N}(\hat{C}_N(\rho) - \hat{S}_N(\rho))\|$?
  $\longrightarrow$ Does not converge to zero!!!
- But there is hope. . . :

$$\sqrt{N}(a^* \hat{C}_N^{-1}(\rho) b - a^* \hat{S}_N^{-1}(\rho) b) \xrightarrow{\text{a.s.}} 0$$

$\Rightarrow$ This is our main result!

What do we need?

CLT over $\hat{C}_N$ statistics

- We know that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$
  $\longrightarrow$ Key result so far!
- What about $\|\sqrt{N}(\hat{C}_N(\rho) - \hat{S}_N(\rho))\|$?
  $\longrightarrow$ Does not converge to zero!!!
- But there is hope. . . :

$$\sqrt{N}(a^* \hat{C}_N^{-1}(\rho) b - a^* \hat{S}_N^{-1}(\rho) b) \xrightarrow{\text{a.s.}} 0$$

  $\Rightarrow$ This is our main result!
- This requires much more delicate treatment, not discussed in this tutorial.

Main results

Theorem (Fluctuation of bilinear forms)

Let $a, b \in \mathbb{C}^N$ *with* $\|a\| = \|b\| = 1$. *Then, as* $N, n \to \infty$ *with* $N/n \to c > 0$, *for any* $\varepsilon > 0$ *and every* $k \in \mathbb{Z}$,

$$\sup_{\rho \in \mathcal{R}_\kappa} N^{1-\varepsilon} \left| a^* \hat{C}_N^k(\rho) b - a^* \hat{S}_N^k(\rho) b \right| \xrightarrow{\text{a.s.}} 0$$

where $\mathcal{R}_\kappa = [\kappa + \max\{0, 1 - 1/c\}, 1]$.

False alarm performance

Theorem (Asymptotic detector performance)

As $N, n \to \infty$ with $N/n \to c \in (0, \infty)$,

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}} \right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\hat{\rho})} \right) \right| \to 0$$

where $\rho \mapsto \hat{\rho}$ is the aforementioned mapping and

$$\sigma_N^2(\hat{\rho}) \triangleq \frac{1}{2} \frac{p^* C_N Q_N^2(\hat{\rho}) p}{p^* Q_N(\hat{\rho}) p \cdot \frac{1}{N} \operatorname{tr} C_N Q_N(\hat{\rho}) \cdot \left(1 - c(1-\rho)^2 m(-\hat{\rho})^2 \frac{1}{N} \operatorname{tr} C_N^2 Q_N^2(\hat{\rho})\right)}$$

with $Q_N(\hat{\rho}) \triangleq (I_N + (1-\hat{\rho}) m(-\hat{\rho}) C_N)^{-1}$.

False alarm performance

Theorem (Asymptotic detector performance)

As $N, n \to \infty$ with $N/n \to c \in (0, \infty)$,

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| P\left(T_N(\rho) > \frac{\gamma}{\sqrt{N}} \right) - \exp\left(-\frac{\gamma^2}{2\sigma_N^2(\hat{\rho})} \right) \right| \to 0$$

where $\rho \mapsto \hat{\rho}$ is the aforementioned mapping and

$$\sigma_N^2(\hat{\rho}) \triangleq \frac{1}{2} \frac{p^* C_N Q_N^2(\hat{\rho}) p}{p^* Q_N(\hat{\rho}) p \cdot \frac{1}{N} tr\, C_N Q_N(\hat{\rho}) \cdot \left(1 - c(1-\rho)^2 m(-\hat{\rho})^2 \frac{1}{N} tr\, C_N^2 Q_N^2(\hat{\rho})\right)}$$

with $Q_N(\hat{\rho}) \triangleq (I_N + (1-\hat{\rho}) m(-\hat{\rho}) C_N)^{-1}$.

- Limiting Rayleigh distribution
  $\Rightarrow$ Weak convergence to Rayleigh variable $R_N(\hat{\rho})$
- **Remark:** σ_N and $\hat{\rho}$ not a function of γ
  $\Rightarrow$ There exists a uniformly optimal ρ!

Simulation

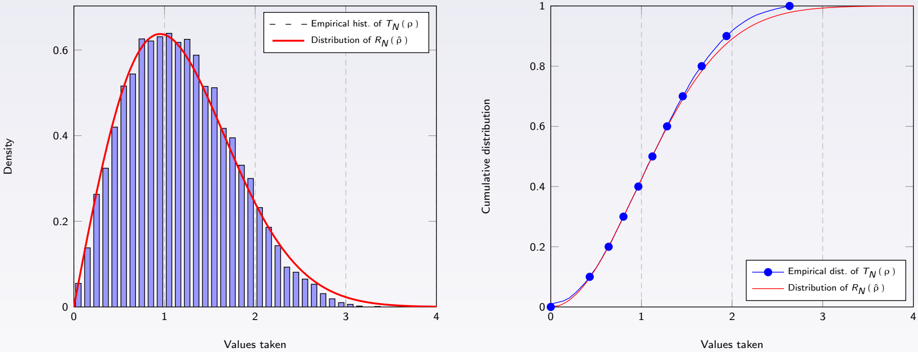


Figure: Histogram distribution function of the $\sqrt{N}T_N(\rho)$ versus $R_N(\hat{\rho})$, $N = 20$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^{\mathsf{T}}$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$, $\rho = 0.2$.

Simulation

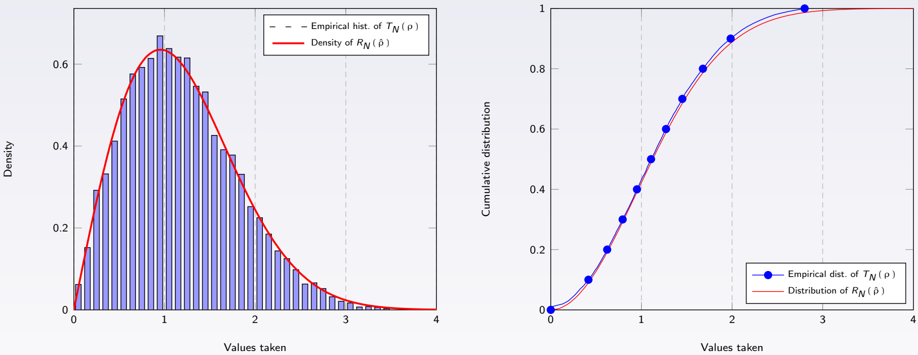


Figure: Histogram distribution function of the $\sqrt{N}T_N(\rho)$ versus $R_N(\hat{\rho})$, $N = 100$, $p = N^{-\frac{1}{2}}[1,\ldots,1]^{\mathsf{T}}$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$, $\rho = 0.2$.

Empirical estimation of optimal ρ

- Optimal ρ can be found by line search... but C_N unknown!
- We shall successively:
  - empirical estimate $\sigma_N(\hat{\rho})$
  - minimize the estimate
  - prove by uniformity asymptotic optimality of estimate

Empirical estimation of optimal ρ

- Optimal ρ can be found by line search... but C_N unknown!
- We shall successively:
  - empirical estimate $\sigma_N(\hat{\rho})$
  - minimize the estimate
  - prove by uniformity asymptotic optimality of estimate

Theorem (Empirical performance estimation)

For $\rho \in (\max\{0, 1 - c_N^{-1}\}, 1)$, *let*

$$\hat{\sigma}_N^2(\hat{\rho}) \triangleq \frac{1}{2} \frac{1 - \hat{\rho} \cdot \frac{p^* \hat{C}_N^{-2}(\rho) p}{p^* \hat{C}_N^{-1}(\rho) p} \cdot \frac{1}{N} tr\, \hat{C}_N(\rho)}{\left(1 - c + c\hat{\rho} \frac{1}{N} tr\, \hat{C}_N^{-1}(\rho) \cdot \frac{1}{N} tr\, \hat{C}_N(\rho)\right)\left(1 - \hat{\rho} \frac{1}{N} tr\, \hat{C}_N^{-1}(\rho) \cdot \frac{1}{N} tr\, \hat{C}_N(\rho)\right)}.$$

Also let $\hat{\sigma}_N^2(1) \triangleq \lim_{\hat{\rho} \uparrow 1} \hat{\sigma}_N^2(\hat{\rho})$. *Then*

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| \sigma_N^2(\hat{\rho}) - \hat{\sigma}_N^2(\hat{\rho}) \right| \xrightarrow{\text{a.s.}} 0.$$

Final result

Theorem (Optimality of empirical estimator)

Define

$$\hat{\rho}_N^* = \operatorname{argmin}_{\{\rho \in \mathcal{R}'_\kappa\}} \left\{ \hat{\sigma}_N^2(\hat{\rho}) \right\}.$$

Then, for every $\gamma > 0$,

$$P\left(\sqrt{N} T_N(\hat{\rho}_N^*) > \gamma\right) - \inf_{\rho \in \mathcal{R}_\kappa} \left\{ P\left(\sqrt{N} T_N(\rho) > \gamma\right) \right\} \to 0.$$

Simulations

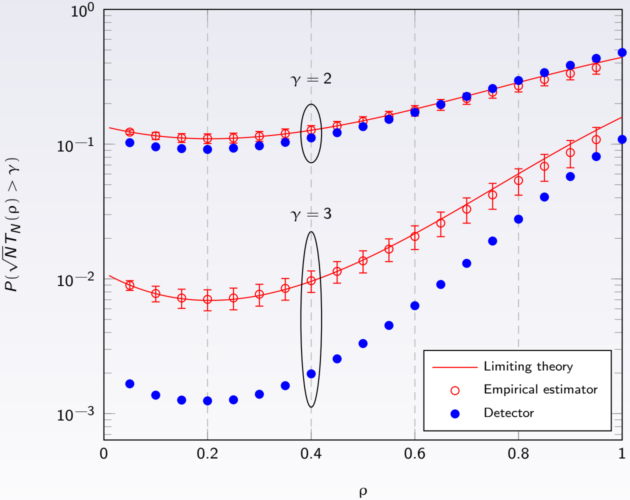


Figure: False alarm rate $P(\sqrt{N} T_N(\rho) > \gamma)$, $N = 20$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^{\mathsf{T}}$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$.

Simulations

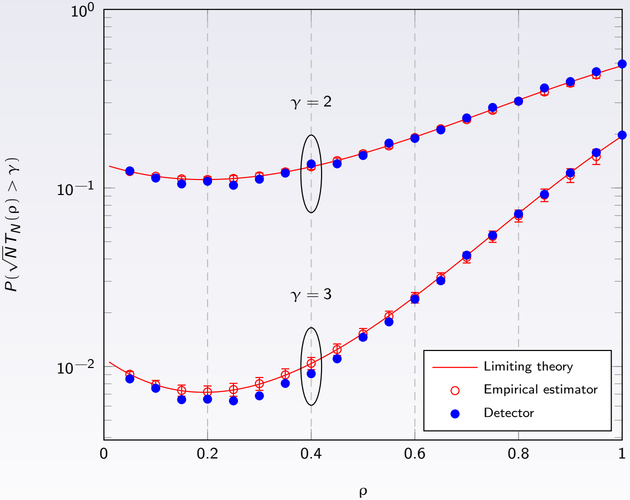


Figure: False alarm rate $P(\sqrt{N}T_N(\rho) > \gamma)$, $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$.

Simulations

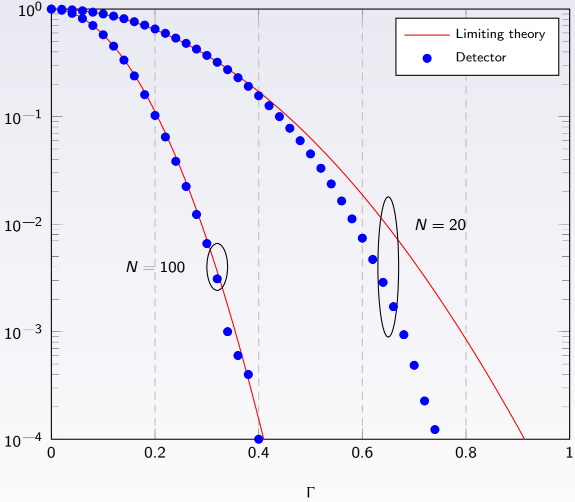


Figure: False alarm rate $P(T_N(\rho) > \Gamma)$ for $N = 20$ and $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^{\mathsf{T}}$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$.

Outline

Outline

Motivation: Spectral Clustering

N. El Karoui. The spectrum of kernel random matrices. The Annals of Statistics, 38(1):150, 2010.

- **Objective:** Clustering data $x_1, \ldots, x_n \in \mathbb{C}^N$ in k similarity classes
  - classical machine learning problem $\Rightarrow$ brought here to big data!

Motivation: Spectral Clustering

N. El Karoui. The spectrum of kernel random matrices. The Annals of Statistics, 38(1):150, 2010.

- **Objective:** Clustering data $x_1, \ldots, x_n \in \mathbb{C}^N$ in k similarity classes
  - classical machine learning problem $\Rightarrow$ brought here to big data!
  - assumes similarity function, e.g. Gaussian kernel

$$f(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

Motivation: Spectral Clustering

N. El Karoui. The spectrum of kernel random matrices. The Annals of Statistics, 38(1):150, 2010.

- **Objective:** Clustering data $x_1, \ldots, x_n \in \mathbb{C}^N$ in k similarity classes
  - classical machine learning problem $\Rightarrow$ brought here to big data!
  - assumes similarity function, e.g. Gaussian kernel

$$f(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

  - naturally brings kernel matrix:

$$W = [W_{ij}]_{1 \leqslant i,j \leqslant n} = [f(x_i, x_j)]_{1 \leqslant i,j \leqslant n}.$$

Motivation: Spectral Clustering

N. El Karoui. The spectrum of kernel random matrices. The Annals of Statistics, 38(1):150, 2010.

- **Objective:** Clustering data $x_1, \ldots, x_n \in \mathbb{C}^N$ in k similarity classes
  - classical machine learning problem $\Rightarrow$ brought here to big data!
  - assumes similarity function, e.g. Gaussian kernel

$$f(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

  - naturally brings kernel matrix:

$$W = [W_{ij}]_{1 \leqslant i,j \leqslant n} = [f(x_i, x_j)]_{1 \leqslant i,j \leqslant n}.$$

- Letting $x_1, \ldots, x_n$ random, leads naturally to studying kernel random matrices.

Motivation: Spectral Clustering

N. El Karoui. The spectrum of kernel random matrices. The Annals of Statistics, 38(1):150, 2010.

- **Objective:** Clustering data $x_1, \ldots, x_n \in \mathbb{C}^N$ in k similarity classes
  - classical machine learning problem $\Rightarrow$ brought here to big data!
  - assumes similarity function, e.g. Gaussian kernel

$$f(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

  - naturally brings kernel matrix:

$$W = [W_{ij}]_{1 \leqslant i,j \leqslant n} = [f(x_i, x_j)]_{1 \leqslant i,j \leqslant n}.$$

- Letting $x_1, \ldots, x_n$ random, leads naturally to studying kernel random matrices.
- Little is known on such random matrices, but for x_i i.i.d. zero mean and covariance I_N:

$$\left\| W - \left(\alpha 1 1^\mathsf{T} - \beta \frac{1}{n} W W^* \right) \right\| \xrightarrow{\text{a.s.}} 0$$

for some α, β depending on f and its derivatives.

Motivation: Spectral Clustering

N. El Karoui. The spectrum of kernel random matrices. The Annals of Statistics, 38(1):150, 2010.

- **Objective:** Clustering data $x_1, \ldots, x_n \in \mathbb{C}^N$ in k similarity classes
  - classical machine learning problem $\Rightarrow$ brought here to big data!
  - assumes similarity function, e.g. Gaussian kernel

$$f(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

  - naturally brings kernel matrix:

$$W = [W_{ij}]_{1 \leqslant i,j \leqslant n} = [f(x_i, x_j)]_{1 \leqslant i,j \leqslant n}.$$

- Letting $x_1, \ldots, x_n$ random, leads naturally to studying kernel random matrices.
- Little is known on such random matrices, but for x_i i.i.d. zero mean and covariance I_N:

$$\left\| W - \left(\alpha 1 1^{\mathsf{T}} - \beta \frac{1}{n} W W^* \right) \right\| \xrightarrow{\text{a.s.}} 0$$

for some α, β depending on f and its derivatives.
$\Rightarrow$ Basically, W gets equivalent to a rank-one matrix.

Motivation: Spectral Clustering

- Clustering $x_1, \ldots, x_n$ in k often written as:

$$\text{(RatioCut)} \quad \min_{\substack{\mathcal{S}_1,\ldots,\mathcal{S}_k \\ \mathcal{S}_1 \cup \ldots \cup \mathcal{S}_k = \mathcal{S} \\ \forall i \neq j,\ \mathcal{S}_i \cap \mathcal{S}_j = \emptyset}} \sum_{i=1}^{k} \sum_{j \in \mathcal{S}_i, \bar{j} \in \mathcal{S}_i^c} \frac{f(x_j, x_{\bar{j}})}{|\mathcal{S}_i|}.$$

$\longrightarrow$ But difficult to solve, NP hard!

Motivation: Spectral Clustering

- Clustering $x_1, \ldots, x_n$ in k often written as:

$$\text{(RatioCut)} \quad \min_{\substack{\mathcal{S}_1,\ldots,\mathcal{S}_k \\ \mathcal{S}_1\cup\ldots\cup\mathcal{S}_k=\mathcal{S} \\ \forall i\neq j,\ \mathcal{S}_i\cap\mathcal{S}_j=\emptyset}} \sum_{i=1}^{k} \sum_{j\in\mathcal{S}_i, \bar{j}\in\mathcal{S}_i^c} \frac{f(x_j, x_{\bar{j}})}{|\mathcal{S}_i|}.$$

$\longrightarrow$ But difficult to solve, NP hard!

- Can be equivalently rewritten

$$\text{(RatioCut)} \quad \min_{M\in\mathcal{M},\ M^\mathsf{T}M=I_k} \operatorname{tr}\left(M^\mathsf{T}LM\right)$$

where $\mathcal{M} = \{M = [m_{ij}]_{1\leqslant i\leqslant n, 1\leqslant j\leqslant k},\ m_{ij} = |\mathcal{S}_j|^{-\frac{1}{2}}\,\delta_{x_i\in\mathcal{S}_j}\}$ and

$$L = [L_{ij}]_{1\leqslant i,j\leqslant n} = [-W + \operatorname{diag}(W\cdot 1)]_{1\leqslant i,j\leqslant n} = \left[-f(x_i,x_j) + \delta_{i,j}\sum_{l=1}^{n} f(x_i,x_l)\right]_{1\leqslant i,j\leqslant n}.$$

Motivation: Spectral Clustering

- Clustering $x_1, \ldots, x_n$ in k often written as:

$$\text{(RatioCut)} \quad \min_{\substack{\mathcal{S}_1,\ldots,\mathcal{S}_k \\ \mathcal{S}_1\cup\ldots\cup\mathcal{S}_k=\mathcal{S} \\ \forall i\neq j,\ \mathcal{S}_i\cap\mathcal{S}_j=\emptyset}} \sum_{i=1}^{k} \sum_{j\in\mathcal{S}_i,\bar{j}\in\mathcal{S}_i^c} \frac{f(x_j,x_{\bar{j}})}{|\mathcal{S}_i|}.$$

$\longrightarrow$ But difficult to solve, NP hard!

- Can be equivalently rewritten

$$\text{(RatioCut)} \quad \min_{M\in\mathcal{M},\ M^{\mathsf{T}}M=I_k} \operatorname{tr}\left(M^{\mathsf{T}}LM\right)$$

where $\mathcal{M} = \{M = [m_{ij}]_{1\leqslant i\leqslant n,1\leqslant j\leqslant k},\ m_{ij} = |\mathcal{S}_j|^{-\frac{1}{2}}\,\delta_{x_i\in\mathcal{S}_j}\}$ and

$$L = [L_{ij}]_{1\leqslant i,j\leqslant n} = [-W + \operatorname{diag}(W\cdot 1)]_{1\leqslant i,j\leqslant n} = \left[-f(x_i,x_j) + \delta_{i,j}\sum_{l=1}^{n} f(x_i,x_l)\right]_{1\leqslant i,j\leqslant n}.$$

- Relaxing M to unitary leads to a simple eigenvalue/eigenvector problem:
$\Rightarrow$ **Spectral clustering**.

Objectives

- Generalization to k distributions for $x_1, \ldots, x_n$ should lead to asymptotically rank-k W matrices.
- If established, specific choices of known "good" kernel better understood.
- Eventually, find optimal choices for kernels.

Outline

Echo-state neural networks

- **Neural network:**
  - Input neuron signal $s_t \in \mathbb{R}$ (could be multivariate)
  - Output neuron signal $y_t \in \mathbb{R}$ (could be multivariate)
  - N neurons with
    - state $x_t \in \mathbb{R}^N$ at time t
    - connectivity matrix $W \in \mathbb{R}^{N \times N}$
    - connectivity vector to input $w_I \in \mathbb{R}^N$
    - connectivity vector to output $w_O \in \mathbb{R}^N$
  - State evolution $x_0 = 0$ (say) and

$$x_{t+1} = S\left(Wx_t + w_I s_t\right)$$

    with S entry-wise sigmoid function.
  - Output observation

$$y_t = w_O^T x_t.$$

Echo-state neural networks

- **Neural network:**
  - Input neuron signal $s_t \in \mathbb{R}$ (could be multivariate)
  - Output neuron signal $y_t \in \mathbb{R}$ (could be multivariate)
  - N neurons with
    - state $x_t \in \mathbb{R}^N$ at time t
    - connectivity matrix $W \in \mathbb{R}^{N \times N}$
    - connectivity vector to input $w_I \in \mathbb{R}^N$
    - connectivity vector to output $w_O \in \mathbb{R}^N$
  - State evolution $x_0 = 0$ (say) and

$$x_{t+1} = S\left(Wx_t + w_I s_t\right)$$

    with S entry-wise sigmoid function.
  - Output observation

$$y_t = w_O^T x_t.$$

- **Classical neural networks:**
  - Learning phase: input-output data (s_t, y_t) used to learn W, w_O, w_I (via e.g. LS)
  - Interpolation phase: W, w_O, w_I fixed, we observe output y_t from new data s_t.

$\Rightarrow$ Poses overlearning problems, difficult to set up, demands lots of learning data.

Echo-state neural networks

- **Neural network:**
  - Input neuron signal $s_t \in \mathbb{R}$ (could be multivariate)
  - Output neuron signal $y_t \in \mathbb{R}$ (could be multivariate)
  - N neurons with
    - state $x_t \in \mathbb{R}^N$ at time t
    - connectivity matrix $W \in \mathbb{R}^{N \times N}$
    - connectivity vector to input $w_I \in \mathbb{R}^N$
    - connectivity vector to output $w_O \in \mathbb{R}^N$
  - State evolution $x_0 = 0$ (say) and

$$x_{t+1} = S\left(Wx_t + w_I s_t\right)$$

    with S entry-wise sigmoid function.
  - Output observation

$$y_t = w_O^T x_t.$$

- **Classical neural networks:**
  - Learning phase: input-output data (s_t, y_t) used to learn W, w_O, w_I (via e.g. LS)
  - Interpolation phase: W, w_O, w_I fixed, we observe output y_t from new data s_t.

$\Rightarrow$ Poses overlearning problems, difficult to set up, demands lots of learning data.

- **Echo-state neural networks:** To solve the problems of neural networks
  - W and w_I set to be a random matrix, no longer learned
  - only w_O is learned

$\Rightarrow$ Reduces amount of data to learn, shows striking performances in some scenarios.

ESN and random matrices

- W, w_I being random, performance study involves random matrices.
  $\Rightarrow$ Stability, chaos regime, etc. involve extreme eigenvalues of W

ESN and random matrices

- W, w_I being random, performance study involves random matrices.
  $\Rightarrow$ Stability, chaos regime, etc. involve extreme eigenvalues of W
- main difficulty is non-linearity caused by S

ESN and random matrices

- W, w_I being random, performance study involves random matrices.
  $\Rightarrow$ Stability, chaos regime, etc. involve extreme eigenvalues of W
- main difficulty is non-linearity caused by S
- **Performance measures:**
  - MSE for training data
  - MSE for interpolated data

  $\Rightarrow$ Optimization to be performed on regression method!, e.g.

$$w_O = (X_{\rm train}X_{\rm train}^T + \gamma I_N)^{-1}X_{\rm train}y_{\rm train}$$

with $X_{\rm train} = [x_1, \ldots, x_T]$, $y_{\rm train} = [y_1, \ldots, y_T]^{\sf T}$, T train period.

ESN and random matrices

- W, w_I being random, performance study involves random matrices.
  $\Rightarrow$ Stability, chaos regime, etc. involve extreme eigenvalues of W
- main difficulty is non-linearity caused by S
- **Performance measures:**
  - MSE for training data
  - MSE for interpolated data

  $\Rightarrow$ Optimization to be performed on regression method!, e.g.

$$w_O = (X_{\rm train} X_{\rm train}^T + \gamma I_N)^{-1} X_{\rm train} y_{\rm train}$$

  with $X_{\rm train} = [x_1, \ldots, x_T]$, $y_{\rm train} = [y_1, \ldots, y_T]^{\sf T}$, T train period.
- **In first approximation:** $S = {\rm Id}$.
  $\Rightarrow$ MSE performance with stationary inputs leads to study

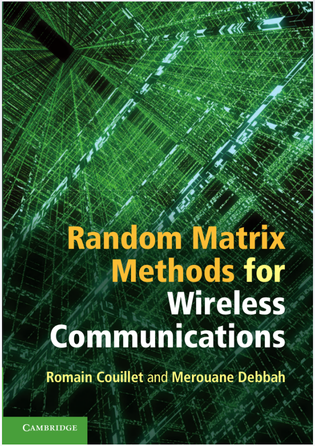
$$\sum_{j=1}^{\infty} W^j w_I w_I^{\sf T} (W^T)^j$$

  $\Rightarrow$ New random matrix model, can be analyzed with usual tools though.

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



Our webpages:

- http://couillet.romain.perso.sfr.fr
- http://sri-uq.kaust.edu.sa/Pages/KammounAbla.aspx

Spraed it!

To download this presentation (PDF format):

- Log in to your Spraed account (www.spraed.net)
- Scan this QR code.