Random Matrices, Robust Estimation, and Applications

(ICASSP'2015, Brisbane)

Romain COUILLET

CentraleSupélec, France

April 19, 2015

CentraleSupélec

Outline

Outline

Outline

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N is the sample covariance matrix (SCM)

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

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N is the sample covariance matrix (SCM)

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

- If $n \to \infty$, then, strong law of large numbers

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

  or equivalently, in spectral norm

$$\left\| \hat{C}_N - C_N \right\| \xrightarrow{\text{a.s.}} 0.$$

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N is the sample covariance matrix (SCM)

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

- If $n \to \infty$, then, strong law of large numbers

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

  or equivalently, in spectral norm

$$\left\| \hat{C}_N - C_N \right\| \xrightarrow{\text{a.s.}} 0.$$

Random Matrix Regime

- No longer valid if $N, n \to \infty$ with $N/n \to c \in (0, \infty)$,

$$\left\| \hat{C}_N - C_N \right\| \not\to 0.$$

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N is the sample covariance matrix (SCM)

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

- If $n \to \infty$, then, strong law of large numbers

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

  or equivalently, in spectral norm

$$\left\| \hat{C}_N - C_N \right\| \xrightarrow{\text{a.s.}} 0.$$

Random Matrix Regime

- No longer valid if $N, n \to \infty$ with $N/n \to c \in (0, \infty)$,

$$\left\| \hat{C}_N - C_N \right\| \not\to 0.$$

- For practical N, n with $N \simeq n$, leads to dramatically wrong conclusions

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N is the sample covariance matrix (SCM)

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

- If $n \to \infty$, then, strong law of large numbers

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

  or equivalently, in spectral norm

$$\left\| \hat{C}_N - C_N \right\| \xrightarrow{\text{a.s.}} 0.$$

Random Matrix Regime

- No longer valid if $N, n \to \infty$ with $N/n \to c \in (0, \infty)$,

$$\left\| \hat{C}_N - C_N \right\| \not\to 0.$$

- For practical N, n with $N \simeq n$, leads to dramatically wrong conclusions
- Even for $N = n/100$.

Setting: $x_i \in \mathbb{C}^N$ i.i.d., $x_1 \sim \mathcal{CN}(0, I_N)$

The Large Dimensional Fallacies

Setting: $x_i \in \mathbb{C}^N$ i.i.d., $x_1 \sim \mathcal{CN}(0, I_N)$

- assume $N = N(n)$ such that $N/n \to c > 1$

The Large Dimensional Fallacies

Setting: $x_i \in \mathbb{C}^N$ i.i.d., $x_1 \sim \mathcal{CN}(0, I_N)$

- assume $N = N(n)$ such that $N/n \to c > 1$
- then, joint point-wise convergence

$$\max_{1 \leq i,j \leq N} \left| \left[\hat{C}_N - I_N \right]_{ij} \right| = \max_{1 \leq i,j \leq N} \left| \frac{1}{n} X_{j,\cdot} X_{i,\cdot}^* - \boldsymbol{\delta}_{ij} \right| \xrightarrow{\text{a.s.}} 0.$$

Setting: $x_i \in \mathbb{C}^N$ i.i.d., $x_1 \sim \mathcal{CN}(0, I_N)$

- assume $N = N(n)$ such that $N/n \to c > 1$
- then, joint point-wise convergence

$$\max_{1 \le i,j \le N} \left| \left[\hat{C}_N - I_N \right]_{ij} \right| = \max_{1 \le i,j \le N} \left| \frac{1}{n} X_{j,\cdot} X_{i,\cdot}^* - \boldsymbol{\delta}_{ij} \right| \xrightarrow{\text{a.s.}} 0.$$

- however, eigenvalue mismatch

$$0 = \lambda_1(\hat{C}_N) = \ldots = \lambda_{N-n}(\hat{C}_N) \le \lambda_{N-n+1}(\hat{C}_N) \le \ldots \le \lambda_N(\hat{C}_N)$$
$$1 = \lambda_1(I_N) = \ldots = \lambda_{N-n}(I_N) = \lambda_{N-n+1}(\hat{C}_N) = \ldots = \lambda_N(I_N)$$

Setting: $x_i \in \mathbb{C}^N$ i.i.d., $x_1 \sim \mathcal{CN}(0, I_N)$

- assume $N = N(n)$ such that $N/n \to c > 1$
- then, joint point-wise convergence

$$\max_{1\leq i,j\leq N} \left| \left[\hat{C}_N - I_N \right]_{ij} \right| = \max_{1\leq i,j\leq N} \left| \frac{1}{n} X_{j,\cdot} X_{i,\cdot}^* - \boldsymbol{\delta}_{ij} \right| \xrightarrow{\text{a.s.}} 0.$$

- however, eigenvalue mismatch

$$0 = \lambda_1(\hat{C}_N) = \ldots = \lambda_{N-n}(\hat{C}_N) \leq \lambda_{N-n+1}(\hat{C}_N) \leq \ldots \leq \lambda_N(\hat{C}_N)$$
$$1 = \lambda_1(I_N) = \ldots = \lambda_{N-n}(I_N) = \lambda_{N-n+1}(\hat{C}_N) = \ldots = \lambda_N(I_N)$$

$\Rightarrow$ no convergence in spectral norm.

The Marčenko–Pastur law

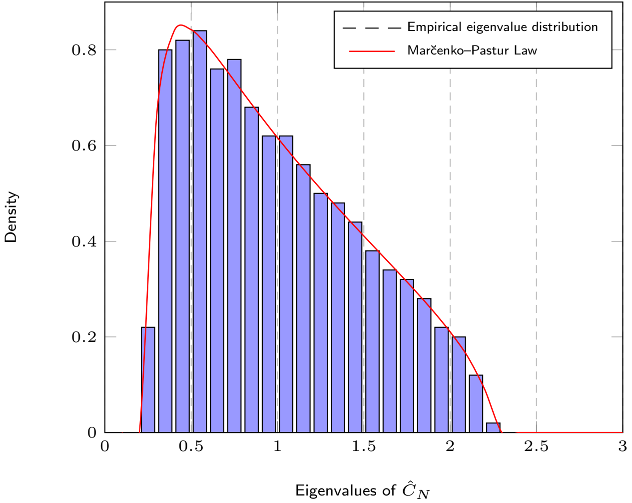


Figure: Histogram of the eigenvalues of $\hat{C}_N$ for $N = 500$, $n = 2000$, $C_N = I_N$.

Definition (Empirical Spectral Density)

Empirical spectral density (e.s.d.) μ_N of Hermitian matrix $A_N \in \mathbb{C}^{N\times N}$ is

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\delta}_{\lambda_i(A_N)}.$$

Definition (Empirical Spectral Density)

Empirical spectral density (e.s.d.) μ_N of Hermitian matrix $A_N \in \mathbb{C}^{N\times N}$ is

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\delta}_{\lambda_i(A_N)}.$$

Theorem (Marčenko–Pastur Law **[Marčenko,Pastur'67]**)

$X_N \in \mathbb{C}^{N\times n}$ with i.i.d. zero mean, unit variance entries.
As $N, n \to \infty$ with $N/n \to c \in (0,\infty)$, e.s.d. μ_N of $\frac{1}{n}X_N X_N^*$ satisfies

$$\mu_N \xrightarrow{\text{a.s.}} \mu_c$$

weakly, where

- $\mu_c(\{0\}) = \max\{0, 1 - c^{-1}\}$

Definition (Empirical Spectral Density)

Empirical spectral density (e.s.d.) μ_N of Hermitian matrix $A_N \in \mathbb{C}^{N\times N}$ is

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\delta}_{\lambda_i(A_N)}.$$

Theorem (Marčenko–Pastur Law **[Marčenko,Pastur'67]**)

$X_N \in \mathbb{C}^{N\times n}$ with i.i.d. zero mean, unit variance entries.
As $N, n \to \infty$ with $N/n \to c \in (0,\infty)$, e.s.d. μ_N of $\frac{1}{n}X_N X_N^*$ satisfies

$$\mu_N \xrightarrow{\text{a.s.}} \mu_c$$

weakly, where

- $\mu_c(\{0\}) = \max\{0, 1 - c^{-1}\}$
- *on $(0,\infty)$, μ_c has continuous density f_c supported on $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$*

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

The Marčenko–Pastur law

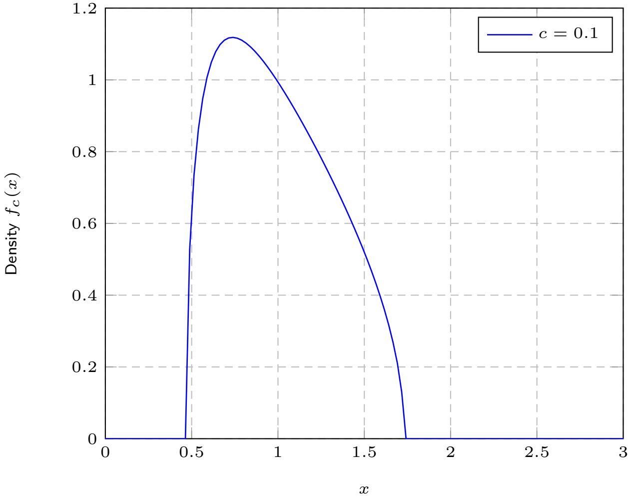


Figure: Marčenko-Pastur law for different limit ratios $c = \lim_{N\to\infty} N/n$.

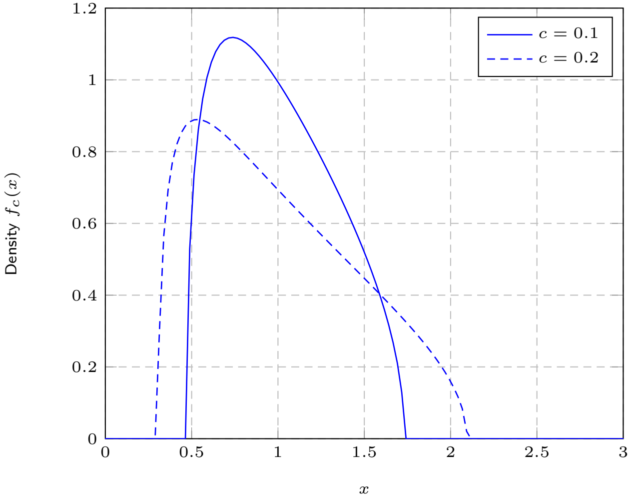


Figure: Marčenko-Pastur law for different limit ratios $c = \lim_{N \to \infty} N/n$.

The Marčenko–Pastur law

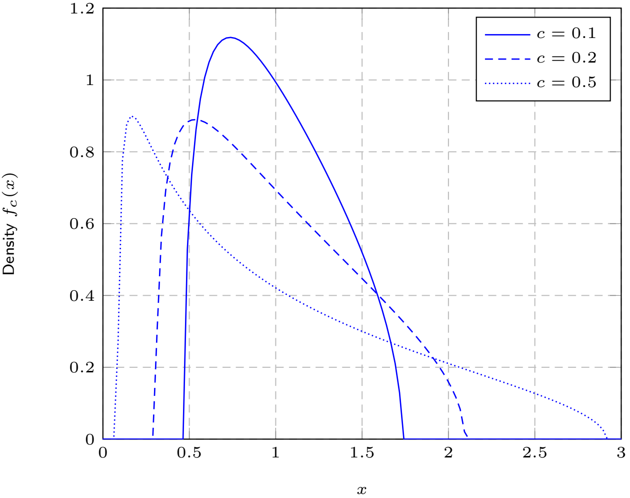


Figure: Marčenko-Pastur law for different limit ratios $c = \lim_{N\to\infty} N/n$.

In Wireless Communications.
Since 1998,

- **[Telatar'98],[Hachem et al.'07]** Mutual information $\mathcal{I}(x,y)$ of multivariate channels (MIMO, CDMA, MAC, etc.):
  $y = Hx + \sigma w$, $H \in \mathbb{C}^{N\times n}$,

$$\mathcal{I}(x,y) = \log\det\left(I_N + \sigma^{-2}HH^*\right) = \int \log(1+\sigma^{-2}t)\mu_N(dt)$$

  with $\mu_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\delta}_{\lambda_i(HH^*)}$.

In Wireless Communications.
Since 1998,

- **[Telatar'98],[Hachem et al.'07]** Mutual information $\mathcal{I}(x,y)$ of multivariate channels (MIMO, CDMA, MAC, etc.):
  $y = Hx + \sigma w$, $H \in \mathbb{C}^{N\times n}$,

$$\mathcal{I}(x,y) = \log\det\left(I_N + \sigma^{-2}HH^*\right) = \int \log(1+\sigma^{-2}t)\mu_N(dt)$$

  with $\mu_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\delta}_{\lambda_i(HH^*)}$.

- **[Shamai,Verdù'01],[Wagner et al.'12]** Multi-user MIMO rate r of broadcast channels, linear receivers:
  $y_i = H_{i,\cdot}(HH^* + \alpha I_N)^{-1}Hx + \sigma w_i$

$$r = \log\left(1 + \frac{|H_{i,\cdot}(HH^*+\alpha I_N)^{-1}H_{i,\cdot}^*|^2}{\sigma^2 + \|H_{i,\cdot}(HH^*+\alpha I_N)^{-1}H_{-i,\cdot}H_{-i,\cdot}^*\|^2}\right).$$

In Signal Processing.
Since 2005 (mostly),

- **[Cardoso et al.'08],[Bianchi et al.'11]** Hypothesis tests:
  for $i = 1, \ldots, n$, $y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ a + \sigma w_i & , \mathcal{H}_1 \end{cases}$,

$$\frac{\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n y_i y_i^*\right)}{\frac{1}{N}\operatorname{tr}\frac{1}{n}\sum_{i=1}^n y_i y_i^*} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$$

$$\frac{|y_n^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} a|}{\sqrt{y_n^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} y_n}\sqrt{a^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} a}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma.$$

In Signal Processing.
Since 2005 (mostly),

- **[Cardoso et al.'08],[Bianchi et al.'11]** Hypothesis tests:
  for $i = 1, \ldots, n$, $y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ a + \sigma w_i & , \mathcal{H}_1 \end{cases}$,

$$\frac{\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n y_i y_i^*\right)}{\frac{1}{N}\operatorname{tr}\frac{1}{n}\sum_{i=1}^n y_i y_i^*} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$$

$$\frac{|y_n^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} a|}{\sqrt{y_n^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} y_n}\sqrt{a^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} a}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma.$$

- **[Mestre'08],[Couillet et al.'11]** Subspace and energy estimation:
  for $i = 1, \ldots, n$, $y_i = \sum_{\ell=1}^L \sqrt{p_\ell}\, a(\theta_\ell) + \sigma w_i$,

$$f\left(\frac{1}{n}\sum_{i=1}^n y_i y_i^*\right) \triangleq \hat{p}_\ell \overset{?}{\to} p_\ell$$

$$g\left(\frac{1}{n}\sum_{i=1}^n y_i y_i^*\right) \triangleq \hat{\theta}_\ell \overset{?}{\to} \theta_\ell$$

In Signal Processing.
Since 2005 (mostly),

- **[Cardoso et al.'08],[Bianchi et al.'11]** Hypothesis tests:
  for $i=1,\ldots,n$, $y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ a+\sigma w_i & , \mathcal{H}_1 \end{cases}$,

$$\frac{\lambda_{\max}\left(\frac{1}{n}\sum_{i=1}^n y_i y_i^*\right)}{\frac{1}{N}\operatorname{tr}\frac{1}{n}\sum_{i=1}^n y_i y_i^*} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma$$

$$\frac{|y_n^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} a|}{\sqrt{y_n^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} y_n}\sqrt{a^*(\frac{1}{n-1}\sum_{i=1}^{n-1} y_i y_i^*)^{-1} a}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma.$$

- **[Mestre'08],[Couillet et al.'11]** Subspace and energy estimation:
  for $i=1,\ldots,n$, $y_i = \sum_{\ell=1}^L \sqrt{p_\ell} a(\theta_\ell) + \sigma w_i$,

$$f\left(\frac{1}{n}\sum_{i=1}^n y_i y_i^*\right) \triangleq \hat{p_\ell} \overset{?}{\to} p_\ell$$

$$g\left(\frac{1}{n}\sum_{i=1}^n y_i y_i^*\right) \triangleq \hat{\theta_\ell} \overset{?}{\to} \theta_\ell$$

  - former (usually) involves eigenvalues of $\frac{1}{n}\sum_{i=1}^n y_i y_i^*$
  - later (usually) involves eigenspaces of $\frac{1}{n}\sum_{i=1}^n y_i y_i^*$.

Outline

Definition (Stieltjes Transform)

For μ real probability measure of support $\operatorname{supp}(\mu)$, Stieltjes transform m_μ defined, for $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) = \int \frac{1}{t - z} \mu(dt).$$

Definition (Stieltjes Transform)

For μ real probability measure of support $\operatorname{supp}(\mu)$, Stieltjes transform m_μ defined, for $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) = \int \frac{1}{t - z} \mu(dt).$$

Property (Inverse Stieltjes Transform)

For $a < b$ continuity points of μ,

$$\mu([a, b]) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \Im[m_F(x + \imath\varepsilon)] dx$$

Definition (Stieltjes Transform)

For μ real probability measure of support $\operatorname{supp}(\mu)$, Stieltjes transform m_μ defined, for $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$, as

$$m_\mu(z) = \int \frac{1}{t-z} \mu(dt).$$

Property (Inverse Stieltjes Transform)

For $a < b$ continuity points of μ,

$$\mu([a,b]) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \Im[m_F(x + \imath\varepsilon)]dx$$

Besides, if μ has a density f at x,

$$f(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \Im[m_F(x + \imath\varepsilon)].$$

Property (Relation to e.s.d.)

If μ e.s.d. of Hermitian $A \in \mathbb{C}^{N \times N}$, (i.e., $\mu = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\delta}_{\lambda_i(A)}$)

$$m_\mu(z) = \frac{1}{N} \operatorname{tr} (A - zI_N)^{-1}$$

Property (Relation to e.s.d.)

If μ e.s.d. of Hermitian $A \in \mathbb{C}^{N \times N}$, (i.e., $\mu = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\delta}_{\lambda_i(A)}$)

$$m_\mu(z) = \frac{1}{N} \mathrm{tr}\, (A - zI_N)^{-1}$$

Proof:

$$m_\mu(z) = \int \frac{\mu(dt)}{t - z} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\lambda_i(A) - z} = \frac{1}{N} \mathrm{tr}\, (\mathrm{diag}\{\lambda_i(A)\} - zI_N)^{-1}$$

$$= \frac{1}{N} \mathrm{tr}\, (A - zI_N)^{-1}.$$

The Stieltjes transform

Property (Stieltjes transform and moments)

For compactly supported μ,

$$m_\mu(z) = -\sum_{k=0}^{\infty} M_{\mu,k} z^{-k-1}$$

with $M_{\mu,k} = \int t^k \mu(dt)$.

Property (Stieltjes transform and moments)

For compactly supported μ,

$$m_\mu(z) = -\sum_{k=0}^{\infty} M_{\mu,k} z^{-k-1}$$

with $M_{\mu,k} = \int t^k \mu(dt)$.

If μ e.s.d. of A,

$$m_\mu(z) = -\sum_{k=0}^{\infty} z^{-k-1} \frac{1}{N} \operatorname{tr} A^k.$$

Property (Stieltjes transform of Gram matrices)

For $X \in \mathbb{C}^{N \times n}$, and

- μ e.s.d. of XX^*
- $\tilde{\mu}$ e.s.d. X^*X

Then

$$m_\mu(z) = \frac{n}{N} m_{\tilde{\mu}}(z) - \frac{N-n}{N} \frac{1}{z}.$$

Property (Stieltjes transform of Gram matrices)

For $X \in \mathbb{C}^{N\times n}$, and

- μ e.s.d. of XX^*
- $\tilde{\mu}$ e.s.d. X^*X

Then

$$m_\mu(z) = \frac{n}{N} m_{\tilde{\mu}}(z) - \frac{N-n}{N}\frac{1}{z}.$$

Proof:

$$m_\mu(z) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i(XX^*) - z} = \frac{1}{N}\sum_{i=1}^{n} \frac{1}{\lambda_i(X^*X) - z} + \frac{1}{N}(N-n)\frac{1}{0-z}.$$

Side remark (for wireless communications)

Definition (Shannon Transform)

μ real probability measure with Stieltjes transform m_μ and support $\operatorname{supp}(\mu) \subset \mathbb{R}^+$, then Shannon Transform $\mathcal{V}_\mu$ is

$$\begin{aligned}\mathcal{V}_\mu(x) &= \int_0^\infty \log(1+xt)\mu(dt) \\ &= \int_x^\infty \left(\frac{1}{t} - m_\mu(-t)\right) dt.\end{aligned}$$

Side remark (for wireless communications)

Definition (Shannon Transform)

μ real probability measure with Stieltjes transform m_μ and support $\mathrm{supp}(\mu) \subset \mathbb{R}^+$, then Shannon Transform $\mathcal{V}_\mu$ is

$$\begin{aligned}\mathcal{V}_\mu(x) &= \int_0^\infty \log(1+xt)\mu(dt) \\ &= \int_x^\infty \left(\frac{1}{t} - m_\mu(-t)\right) dt.\end{aligned}$$

- Fundamental to capacity calculus in wireless communications
- Can be computed from m_μ alone, no need to know μ.

Three fundamental lemmas in all proofs.

Lemma (Resolvent Identity)

For $A, B \in \mathbb{C}^{N \times N}$ invertible,

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}.$$

Three fundamental lemmas in all proofs.

Lemma (Resolvent Identity)

For $A, B \in \mathbb{C}^{N\times N}$ invertible,

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}.$$

Corollary

For $t \in \mathbb{C}$, $x \in \mathbb{C}^N$, $A \in \mathbb{C}^{N\times N}$, with A and $A + txx^*$ invertible,

$$(A + txx^*)^{-1}x = \frac{A^{-1}x}{1 + tx^*A^{-1}x}.$$

Three fundamental lemmas in all proofs.

Lemma (Rank-one perturbation)

For $A, B \in \mathbb{C}^{N\times N}$ Hermitian nonnegative definite, e.s.d. μ of A, $t > 0$, $x \in \mathbb{C}^N$, $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$,

$$\left| \frac{1}{N} \operatorname{tr} B \left(A + txx^* - zI_N\right)^{-1} - \frac{1}{N} \operatorname{tr} B \left(A - zI_N\right)^{-1} \right| \leq \frac{1}{N} \frac{\|B\|}{\operatorname{dist}(z, \operatorname{supp}(\mu))}$$

Three fundamental lemmas in all proofs.

Lemma (Rank-one perturbation)

For $A, B \in \mathbb{C}^{N\times N}$ Hermitian nonnegative definite, e.s.d. μ of A, $t > 0$, $x \in \mathbb{C}^N$, $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$,

$$\left| \frac{1}{N} \operatorname{tr} B \left(A + txx^* - zI_N\right)^{-1} - \frac{1}{N} \operatorname{tr} B \left(A - zI_N\right)^{-1} \right| \leq \frac{1}{N} \frac{\|B\|}{\operatorname{dist}(z, \operatorname{supp}(\mu))}$$

In particular, as $N \to \infty$, if $\limsup_N \|B\| < \infty$,

$$\frac{1}{N} \operatorname{tr} B \left(A + txx^* - zI_N\right)^{-1} - \frac{1}{N} \operatorname{tr} B \left(A - zI_N\right)^{-1} \to 0.$$

Three fundamental lemmas in all proofs.

Lemma (Trace Lemma)

For

- $x \in \mathbb{C}^N$ with i.i.d. entries with zero mean, unit variance, finite eighth moment,
- $A \in \mathbb{C}^{N\times N}$ deterministic (or independent of x), $\limsup_N \|A\| = 0$ (or a.s.),

then

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

Theorem (Marčenko–Pastur Law **[Marčenko,Pastur'67]**)

$X_N \in \mathbb{C}^{N\times n}$ *with i.i.d. zero mean, unit variance entries.*
As $N, n \to \infty$ *with* $N/n \to c \in (0,\infty)$, *e.s.d.* μ_N *of* $\frac{1}{n}X_N X_N^*$ *satisfies*

$$\mu_N \xrightarrow{\text{a.s.}} \mu_c$$

weakly, where

- $\mu_c(\{0\}) = \max\{0, 1 - c^{-1}\}$

Proof of the Marčenko–Pastur law

Theorem (Marčenko–Pastur Law **[Marčenko,Pastur'67]**)

$X_N \in \mathbb{C}^{N\times n}$ *with i.i.d. zero mean, unit variance entries.*
As $N, n \to \infty$ *with* $N/n \to c \in (0,\infty)$, *e.s.d.* μ_N *of* $\frac{1}{n}X_N X_N^*$ *satisfies*

$$\mu_N \xrightarrow{\text{a.s.}} \mu_c$$

weakly, where

- $\mu_c(\{0\}) = \max\{0, 1 - c^{-1}\}$
- *on* $(0,\infty)$, μ_c *has continuous density* f_c *supported on* $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$

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

Stieltjes transform approach.

Stieltjes transform approach.

Proof

- With μ_N e.s.d. of $\frac{1}{n}X_N X_N^*$,

$$m_{\mu_N}(z) = \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}X_N X_N^* - zI_N\right)^{-1} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1}{n}X_N X_N^{\mathsf{H}} - zI_N\right)^{-1}\right]_{ii}.$$

Stieltjes transform approach.

Proof

- With μ_N e.s.d. of $\frac{1}{n}X_N X_N^*$,

$$m_{\mu_N}(z) = \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}X_N X_N^* - zI_N\right)^{-1} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1}{n}X_N X_N^{\mathsf{H}} - zI_N\right)^{-1}\right]_{ii}.$$

- Write

$$X_N = \begin{bmatrix} y^* \\ Y_{N-1} \end{bmatrix} \in \mathbb{C}^{N\times n}$$

Stieltjes transform approach.

Proof

- With μ_N e.s.d. of $\frac{1}{n}X_N X_N^*$,

$$m_{\mu_N}(z) = \frac{1}{N}\mathrm{tr}\left(\frac{1}{n}X_N X_N^* - zI_N\right)^{-1} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1}{n}X_N X_N^{\mathsf{H}} - zI_N\right)^{-1}\right]_{ii}.$$

- Write

$$X_N = \begin{bmatrix} y^* \\ Y_{N-1} \end{bmatrix} \in \mathbb{C}^{N\times n}$$

so that, for $\Im[z] > 0$,

$$\left(\frac{1}{n}X_N X_N^{\mathsf{H}} - zI_N\right)^{-1} = \begin{pmatrix} \frac{1}{n}y^*y - z & \frac{1}{n}y^*Y_{N-1} \\ \frac{1}{n}Y_{N-1}y & \frac{1}{n}Y_{N-1}Y_{N-1}^* - zI_{N-1} \end{pmatrix}^{-1}.$$

Proof (continued)

- From block matrix inverse formula

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

we have

$$\left[\left(\frac{1}{n}X_N X_N^* - zI_N\right)^{-1}\right]_{11} = \frac{1}{-z - z\frac{1}{n}y^*(\frac{1}{n}Y_{N-1}^* Y_{N-1} - zI_n)^{-1}y}.$$

Proof (continued)

- From block matrix inverse formula

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

we have

$$\left[\left(\frac{1}{n} X_N X_N^* - z I_N \right)^{-1} \right]_{11} = \frac{1}{-z - z\frac{1}{n} y^* (\frac{1}{n} Y_{N-1}^* Y_{N-1} - z I_n)^{-1} y}.$$

- By Trace Lemma, as $N, n \to \infty$

$$\left[\left(\frac{1}{n} X_N X_N^* - z I_N \right)^{-1} \right]_{11} - \frac{1}{-z - z\frac{1}{n} \operatorname{tr} (\frac{1}{n} Y_{N-1}^* Y_{N-1} - z I_n)^{-1}} \xrightarrow{\text{a.s.}} 0.$$

Proof (continued)

- By Rank-1 Perturbation Lemma ($X_N^* X_N = Y_{N-1}^* Y_{N-1} + yy^*$), as $N, n \to \infty$

$$\left[\left(\frac{1}{n} X_N X_N^* - z I_N\right)^{-1}\right]_{11} - \frac{1}{-z - z\frac{1}{n}\operatorname{tr}\left(\frac{1}{n} X_N^* X_N - z I_n\right)^{-1}} \xrightarrow{\text{a.s.}} 0.$$

Proof (continued)

- By Rank-1 Perturbation Lemma ($X_N^* X_N = Y_{N-1}^* Y_{N-1} + yy^*$), as $N, n \to \infty$

$$\left[\left(\frac{1}{n} X_N X_N^* - zI_N\right)^{-1}\right]_{11} - \frac{1}{-z - z\frac{1}{n}\operatorname{tr}\left(\frac{1}{n}X_N^* X_N - zI_n\right)^{-1}} \xrightarrow{\text{a.s.}} 0.$$

- Since $\frac{1}{n}\operatorname{tr}\left(\frac{1}{n}X_N^* X_N - zI_n\right)^{-1} = \frac{1}{n}\operatorname{tr}\left(\frac{1}{n}X_N X_N^* - zI_N\right)^{-1} - \frac{n-N}{n}\frac{1}{z}$,

$$\left[\left(\frac{1}{n} X_N X_N^* - zI_N\right)^{-1}\right]_{11} - \frac{1}{1 - \frac{N}{n} - z - z\frac{1}{n}\operatorname{tr}\left(\frac{1}{n}X_N X_N^* - zI_N\right)^{-1}} \xrightarrow{\text{a.s.}} 0.$$

Proof (continued)

- By Rank-1 Perturbation Lemma ($X_N^* X_N = Y_{N-1}^* Y_{N-1} + yy^*$), as $N, n \to \infty$

$$\left[\left(\frac{1}{n} X_N X_N^* - z I_N \right)^{-1} \right]_{11} - \frac{1}{-z - z \frac{1}{n} \operatorname{tr} \left(\frac{1}{n} X_N^* X_N - z I_n \right)^{-1}} \xrightarrow{\text{a.s.}} 0.$$

- Since $\frac{1}{n} \operatorname{tr} \left(\frac{1}{n} X_N^* X_N - z I_n \right)^{-1} = \frac{1}{n} \operatorname{tr} \left(\frac{1}{n} X_N X_N^* - z I_N \right)^{-1} - \frac{n-N}{n} \frac{1}{z}$,

$$\left[\left(\frac{1}{n} X_N X_N^* - z I_N \right)^{-1} \right]_{11} - \frac{1}{1 - \frac{N}{n} - z - z \frac{1}{n} \operatorname{tr} \left(\frac{1}{n} X_N X_N^* - z I_N \right)^{-1}} \xrightarrow{\text{a.s.}} 0.$$

- Repeating for entries $(2, 2), \ldots, (N, N)$, and averaging, we get (for $\Im[z] > 0$)

$$m_{\mu_N}(z) - \frac{1}{1 - \frac{N}{n} - z - z \frac{N}{n} m_{\mu_N}(z)} \xrightarrow{\text{a.s.}} 0.$$

Proof (continued)

- Then $m_{\mu_N}(z) \xrightarrow{\text{a.s.}} m(z)$ solution to

$$m(z) = \frac{1}{1 - c - z - czm(z)}$$

Proof (continued)

- Then $m_{\mu_N}(z) \xrightarrow{\text{a.s.}} m(z)$ solution to

$$m(z) = \frac{1}{1 - c - z - czm(z)}$$

i.e., (with branch of $\sqrt{f(z)}$ such that $m(z) \to 0$ as $|z| \to \infty$)

$$m(z) = \frac{1-c}{2cz} - \frac{1}{2c} + \frac{\sqrt{\left(z - (1+\sqrt{c})^2\right)\left(z - (1-\sqrt{c})^2\right)}}{2cz}.$$

Proof (continued)

- Then $m_{\mu_N}(z) \xrightarrow{\text{a.s.}} m(z)$ solution to

$$m(z) = \frac{1}{1 - c - z - czm(z)}$$

i.e., (with branch of $\sqrt{f(z)}$ such that $m(z) \to 0$ as $|z| \to \infty$)

$$m(z) = \frac{1-c}{2cz} - \frac{1}{2c} + \frac{\sqrt{\left(z - (1+\sqrt{c})^2\right)\left(z - (1-\sqrt{c})^2\right)}}{2cz}.$$

- Finally, by inverse Stieltjes Transform, for $x > 0$,

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \Im[m(x + \imath\varepsilon)] = \frac{\sqrt{\left((1+\sqrt{c})^2 - x\right)\left(x - (1-\sqrt{c})^2\right)}}{2\pi c x} 1_{\{x \in [(1-\sqrt{c})^2, (1+\sqrt{c})^2]\}}.$$

And for $x = 0$,

$$\lim_{\varepsilon \downarrow 0} \imath\varepsilon \Im[m(\imath\varepsilon)] = \left(1 - c^{-1}\right) 1_{\{c>1\}}.$$

Theorem (Sample Covariance Matrix Model **[Silverstein,Bai'95]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N \in \mathbb{C}^{N\times n}$, *with*

- $C_N \in \mathbb{C}^{N\times N}$ *nonnegative definite with e.s.d.* $\nu_N \to \nu$ *weakly,*
- $X_N \in \mathbb{C}^{N\times n}$ *has i.i.d. entries of zero mean and unit variance.*

As $N, n \to \infty$, $N/n \to c \in (0,\infty)$, $\tilde{\mu}_N$ *e.s.d. of* $\frac{1}{n} Y_N^* Y_N \in \mathbb{C}^{n\times n}$ *satisfies*

$$\tilde{\mu}_N \xrightarrow{\text{a.s.}} \tilde{\mu}$$

weakly, with $m_{\tilde{\mu}}(z)$, $\Im[z] > 0$, *unique solution with* $\Im[m_{\tilde{\mu}}(z)] > 0$ *of*

$$m_{\tilde{\mu}}(z) = \left(-z + c \int \frac{t}{1 + t m_{\tilde{\mu}}(z)} \nu(dt) \right)^{-1}.$$

Theorem (Sample Covariance Matrix Model **[Silverstein,Bai'95]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N \in \mathbb{C}^{N\times n}$, *with*

- $C_N \in \mathbb{C}^{N\times N}$ *nonnegative definite with e.s.d.* $\nu_N \to \nu$ *weakly,*
- $X_N \in \mathbb{C}^{N\times n}$ *has i.i.d. entries of zero mean and unit variance.*

As $N, n \to \infty$, $N/n \to c \in (0,\infty)$, $\tilde{\mu}_N$ *e.s.d. of* $\frac{1}{n} Y_N^* Y_N \in \mathbb{C}^{n\times n}$ *satisfies*

$$\tilde{\mu}_N \xrightarrow{\text{a.s.}} \tilde{\mu}$$

weakly, with $m_{\tilde{\mu}}(z)$, $\Im[z] > 0$, *unique solution with* $\Im[m_{\tilde{\mu}}(z)] > 0$ *of*

$$m_{\tilde{\mu}}(z) = \left(-z + c \int \frac{t}{1 + t m_{\tilde{\mu}}(z)} \nu(dt) \right)^{-1}.$$

Moreover, $\tilde{\mu}$ *is continuous on* $\mathbb{R}^+$ *and real analytic wherever positive.*

Theorem (Sample Covariance Matrix Model [Silverstein,Bai'95])

Let $Y_N = C_N^{\frac{1}{2}} X_N \in \mathbb{C}^{N\times n}$, *with*

- $C_N \in \mathbb{C}^{N\times N}$ *nonnegative definite with e.s.d.* $\nu_N \to \nu$ *weakly,*
- $X_N \in \mathbb{C}^{N\times n}$ *has i.i.d. entries of zero mean and unit variance.*

As $N, n \to \infty$, $N/n \to c \in (0,\infty)$, $\tilde{\mu}_N$ *e.s.d. of* $\frac{1}{n} Y_N^* Y_N \in \mathbb{C}^{n\times n}$ *satisfies*

$$\tilde{\mu}_N \xrightarrow{\text{a.s.}} \tilde{\mu}$$

weakly, with $m_{\tilde{\mu}}(z)$, $\Im[z] > 0$, *unique solution with* $\Im[m_{\tilde{\mu}}(z)] > 0$ *of*

$$m_{\tilde{\mu}}(z) = \left(-z + c \int \frac{t}{1 + t m_{\tilde{\mu}}(z)} \nu(dt) \right)^{-1}.$$

Moreover, $\tilde{\mu}$ *is continuous on* $\mathbb{R}^+$ *and real analytic wherever positive.*

Immediate corollary: For μ_N e.s.d. of $\frac{1}{n} Y_N Y_N^* = \frac{1}{n} \sum_{i=1}^n C_N^{\frac{1}{2}} x_i x_i^* C_N^{\frac{1}{2}}$,

$$\mu_N \xrightarrow{\text{a.s.}} \mu$$

weakly, with $\tilde{\mu} = c\mu + (1-c)\boldsymbol{\delta}_0$.

Side note on other models.
Similar results for multiple matrix models:

Side note on other models.
Similar results for multiple matrix models:

- **Doubly-correlated (or separable variance profile)**: $Y_N = C_N^{\frac{1}{2}} X_N T_N^{\frac{1}{2}}$ (2 fixed point equations)
  Applications in Section "Robust Estimation and Random Matrices"

Side note on other models.
Similar results for multiple matrix models:

- **Doubly-correlated (or separable variance profile)**: $Y_N = C_N^{\frac{1}{2}} X_N T_N^{\frac{1}{2}}$ (2 fixed point equations)
  Applications in Section "Robust Estimation and Random Matrices"
- **Information-plus-noise**: $Y_N = A_N + X_N$, A_N deterministic
- **Variance profile**: $Y_N = P_N \odot X_N$ (entry-wise product)
- **Per-column covariance**: $Y_N = [y_1, \ldots, y_n]$, $y_i = C_{N,i}^{\frac{1}{2}} x_i$
- etc.

Retrieving μ (or $\tilde{\mu}$) from $m_{\tilde{\mu}}$.

- Since μ differentiable (unless maybe in zero), recall that

$$f(x) \triangleq \mu'(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \Im[m_\mu(x + \imath\varepsilon)].$$

Retrieving μ (or $\tilde{\mu}$) from $m_{\tilde{\mu}}$.

- Since μ differentiable (unless maybe in zero), recall that

$$f(x) \triangleq \mu'(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \Im[m_\mu(x + \imath\varepsilon)].$$

- Thus, to plot f,
  - span $z = x + \imath\varepsilon$ for all $x \in \mathbb{R}$ and ε small (say, $\varepsilon = 10^{-3}$)
  - solve $m_{\tilde{\mu}}(x + \imath\varepsilon)$ by fixed-point algorithm (provably convergent) for each x
  - plot $(x, \Im[m_{\tilde{\mu}}(x + \imath\varepsilon)])$ for each x.

Sample Covariance Matrices

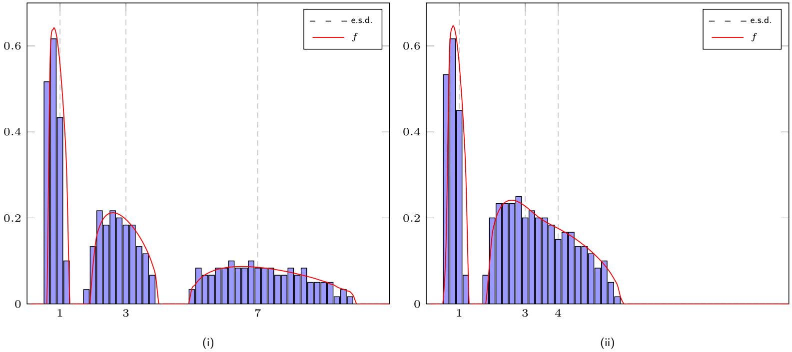


Figure: Histogram of the eigenvalues of $\frac{1}{n}Y_N Y_N^*$, $n = 3000$, $N = 300$, with C_N diagonal with evenly weighted masses in (i) 1, 3, 7, (ii) 1, 3, 4.

Property [Spectral gaps]

Let $Y_N = C_N^{\frac{1}{2}} X_N$ as above. Assume $[a^\mu, b^\mu] \subset \mathbb{R} \setminus \mathrm{supp}(\mu)$, then

$$[a^\nu, b^\nu] \subset \mathbb{R} \setminus \mathrm{supp}(\nu), \text{ where } a^\nu = -1/m_\mu(a^\mu), \; b^\nu = -1/m_\mu(b^\mu).$$

Property [Spectral gaps]

Let $Y_N = C_N^{\frac{1}{2}} X_N$ as above. Assume $[a^\mu, b^\mu] \subset \mathbb{R} \setminus \mathrm{supp}(\mu)$, then

$$[a^\nu, b^\nu] \subset \mathbb{R} \setminus \mathrm{supp}(\nu), \text{ where } a^\nu = -1/m_\mu(a^\mu),\ b^\nu = -1/m_\mu(b^\mu).$$

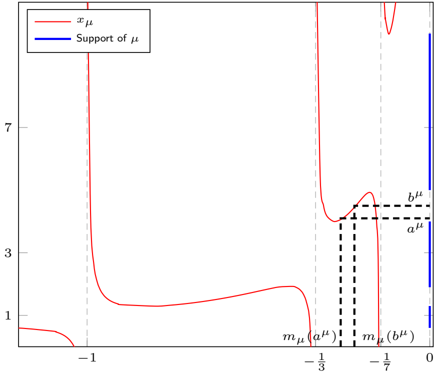


Figure: Function $x_\mu(m)$, extended inverse of $m_\mu(x)$ on real axis, for $\nu = \frac{1}{3}\boldsymbol{\delta}_1 + \frac{1}{3}\boldsymbol{\delta}_3 + \frac{1}{3}\boldsymbol{\delta}_7$.

Theorem (Exact Separation [Bai,Silverstein'99])

Let $Y_N = C_N^{\frac{1}{2}} X_N$ *as above with additionally*

- X_N *has entries of bounded fourth order moment*
- $\max_{1 \leq i \leq n} \operatorname{dist}(\lambda_i(C_N), \operatorname{supp}(\nu)) \to 0$
- $\limsup_N \|C_N\| < \infty$.

Then, letting $[a^\mu, b^\mu] \subset \mathbb{R}^+ \setminus \operatorname{supp}(\mu)$ *with corresponding* $[a^\nu, b^\nu] \subset \mathbb{R}^+ \setminus \operatorname{supp}(\nu)$,

$$\#\left\{\lambda_i\left(\frac{1}{n} Y_N Y_N^*\right) < a^\mu\right\} = \#\left\{\lambda_i(C_N) < a^\nu\right\}$$

for all large n *a.s., except for zero eigenvalues.*

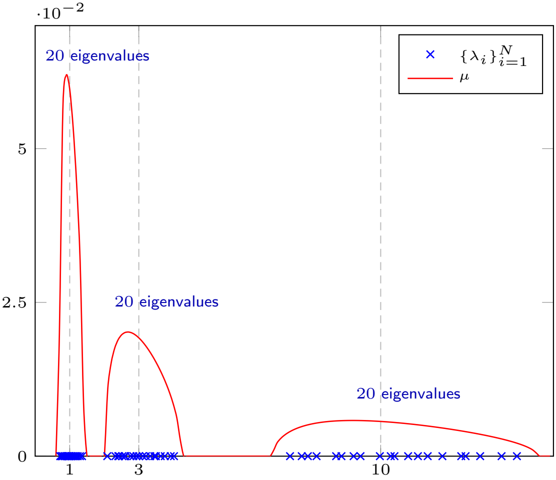


Figure: Eigenvalues of $\frac{1}{n}\mathbf{Y}_N\mathbf{Y}_N^{\mathsf{H}}$ versus μ, $\nu = \frac{1}{3}\boldsymbol{\delta}_1 + \frac{1}{3}\boldsymbol{\delta}_3 + \frac{1}{3}\boldsymbol{\delta}_{10}$, $N = 60$, $n = 600$.

Estimation of C_N.

- For f complex analytic, estimate of $\int f(t)\nu(dt)$?

Estimation of C_N.

- For f complex analytic, estimate of $\int f(t)\nu(dt)$?
- By Cauchy's integral, for $\mathcal{C}_\nu$ enclosing $\mathrm{supp}(\nu)$,

$$\int f(t)\nu(dt) = -\int \left(\frac{1}{2\pi\imath} \oint_{\mathcal{C}_\nu} \frac{f(z)}{t-z} dz \right) \nu(dt) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}_\nu} f(z) m_\nu(z) dz.$$

Estimation of C_N.

- For f complex analytic, estimate of $\int f(t)\nu(dt)$?
- By Cauchy's integral, for $\mathcal{C}_\nu$ enclosing $\mathrm{supp}(\nu)$,

$$\int f(t)\nu(dt) = -\int \left(\frac{1}{2\pi\imath} \oint_{\mathcal{C}_\nu} \frac{f(z)}{t-z} dz \right) \nu(dt) = -\frac{1}{2\pi\imath} \oint_{\mathcal{C}_\nu} f(z) m_\nu(z) dz.$$

- Reminder: by **[Silverstein,Bai'95]**,

$$m_{\tilde{\mu}}(\omega) = \left(-\omega + c \int \frac{t}{1 + t m_{\tilde{\mu}}(\omega)} \nu(dt) \right)^{-1}$$

or equivalently

$$m_\nu \left(-\frac{1}{m_{\tilde{\mu}}(\omega)} \right) = -\omega m_\mu(\omega) m_{\tilde{\mu}}(\omega).$$

Estimation of C_N.

- Together, with $z = -1/m_{\tilde{\mu}}(\omega)$,

$$\int f(t)\nu(dt) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mu} f\left(-\frac{1}{m_{\tilde{\mu}}(\omega)}\right)\omega m_\mu(\omega)\frac{m'_{\tilde{\mu}}(\omega)}{m_{\tilde{\mu}}(\omega)}d\omega$$

with $-1/m_{\tilde{\mu}}(\mathcal{C}_\mu) = \mathcal{C}_\nu$ (thus must be well defined!).

Estimation of C_N.

- Together, with $z = -1/m_{\tilde{\mu}}(\omega)$,

$$\int f(t)\nu(dt) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mu} f\left(-\frac{1}{m_{\tilde{\mu}}(\omega)}\right)\omega m_\mu(\omega)\frac{m'_{\tilde{\mu}}(\omega)}{m_{\tilde{\mu}}(\omega)}d\omega$$

  with $-1/m_{\tilde{\mu}}(\mathcal{C}_\mu) = \mathcal{C}_\nu$ (thus must be well defined!).

- By uniform convergence over $\mathcal{C}_\mu$

$$\sup_{z\in\mathcal{C}_\mu} |m_\mu(z) - m_{\mu_N}(z)| \xrightarrow{\text{a.s.}} 0$$

  ($\mu_N = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\delta}_{\lambda_i(\frac{1}{n}Y_N Y_N^*)}$), we get

$$-\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mu} f\left(-\frac{1}{m_{\tilde{\mu}_N}(\omega)}\right)\omega m_{\mu_N}(\omega)\frac{m'_{\tilde{\mu}_N}(\omega)}{m_{\tilde{\mu}_N}(\omega)}d\omega \xrightarrow{\text{a.s.}} \int f(t)\nu(dt).$$

Estimation of C_N.

- Together, with $z = -1/m_{\tilde{\mu}}(\omega)$,

$$\int f(t)\nu(dt) = -\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mu} f\left(-\frac{1}{m_{\tilde{\mu}}(\omega)}\right)\omega m_\mu(\omega)\frac{m'_{\tilde{\mu}}(\omega)}{m_{\tilde{\mu}}(\omega)}d\omega$$

  with $-1/m_{\tilde{\mu}}(\mathcal{C}_\mu) = \mathcal{C}_\nu$ (thus must be well defined!).

- By uniform convergence over $\mathcal{C}_\mu$

$$\sup_{z\in\mathcal{C}_\mu} |m_\mu(z) - m_{\mu_N}(z)| \xrightarrow{\text{a.s.}} 0$$

  ($\mu_N = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\delta}_{\lambda_i(\frac{1}{n}Y_N Y_N^*)}$), we get

$$-\frac{1}{2\pi\imath}\oint_{\mathcal{C}_\mu} f\left(-\frac{1}{m_{\tilde{\mu}_N}(\omega)}\right)\omega m_{\mu_N}(\omega)\frac{m'_{\tilde{\mu}_N}(\omega)}{m_{\tilde{\mu}_N}(\omega)}d\omega \xrightarrow{\text{a.s.}} \int f(t)\nu(dt).$$

- Since $m_{\tilde{\mu}_N}(\omega) = \frac{1}{N}\sum_{i=1}^N(\lambda_i - z)^{-1}$, computation possible via
  - numerical integrals
  - residue calculus

Estimation of C_N for atomic ν (i.e., $f(t) = t$).

Theorem (Eigen-Inference **[Mestre'08]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$ with exact separation condition. Assume

$$\nu_N = \sum_{k=1}^{K} \frac{N_k}{N} \boldsymbol{\delta}_{t_k}, \text{ with } \sum_{k=1}^{K} N_k = N, \; N_k/N \to c_k$$

and there is a cluster in μ that maps $\{t_j\}$ only.

Estimation of C_N for atomic ν (i.e., $f(t) = t$).

Theorem (Eigen-Inference **[Mestre'08]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$ with exact separation condition. Assume

$$\nu_N = \sum_{k=1}^{K} \frac{N_k}{N} \boldsymbol{\delta}_{t_k}, \text{ with } \sum_{k=1}^{K} N_k = N, \ N_k/N \to c_k$$

and there is a cluster in μ that maps $\{t_j\}$ only. Let

$$\hat{t}_j = \frac{n}{N_j} \sum_{m \in \mathcal{N}_j} (\lambda_m - \kappa_m)$$

- $\mathcal{N}_j = \{\sum_{i=1}^{j-1} N_i + 1, \ldots, \sum_{i=1}^{j} N_i\}$
- $\lambda_1 \geq \ldots \geq \lambda_N$ *eigenvalues of* $\frac{1}{n} Y_N Y_N^*$
- $\kappa_1 \geq \ldots \geq \kappa_N$ *eigenvalues of diag*$(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^*$.

Then, as $N, n \to \infty$, $N/n \to c \in (0, \infty)$,

$$\hat{t}_j \xrightarrow{\text{a.s.}} t_j.$$

Outline

Reminder. According to **[Bai,Sil'98]**, asymptotically no eigenvalue of $\frac{1}{n}Y_N Y_N^*$ outside $\mathrm{supp}(\mu)$ if

1. $E[|X_{ij}|^4] < \infty$
2. $\max_{1\leq i\leq n} \mathrm{dist}(\lambda_i(C_N), \mathrm{supp}(\nu)) \to 0$
3. $\limsup_N \|C_N\| < \infty$.

Reminder. According to **[Bai,Sil'98]**, asymptotically no eigenvalue of $\frac{1}{n}Y_N Y_N^*$ outside $\mathrm{supp}(\mu)$ if

1. $E[|X_{ij}|^4] < \infty$
2. $\max_{1\leq i\leq n} \mathrm{dist}(\lambda_i(C_N), \mathrm{supp}(\nu)) \to 0$
3. $\limsup_N \|C_N\| < \infty$.

Breaking the rules. If we break

- **Rule 1**: Infinitely many eigenvalues may wander away from $\mathrm{supp}(\mu)$.

Spiked Models

Reminder. According to **[Bai,Sil'98]**, asymptotically no eigenvalue of $\frac{1}{n}Y_NY_N^*$ outside $\operatorname{supp}(\mu)$ if

1. $E[|X_{ij}|^4] < \infty$
2. $\max_{1\leq i\leq n} \operatorname{dist}(\lambda_i(C_N), \operatorname{supp}(\nu)) \to 0$
3. $\limsup_N \|C_N\| < \infty$.

Breaking the rules. If we break

- **Rule 1**: Infinitely many eigenvalues may wander away from $\operatorname{supp}(\mu)$.
- **Rule 2**: C_N may create isolated eigenvalues in $\frac{1}{n}Y_NY_N^*$, called spikes.

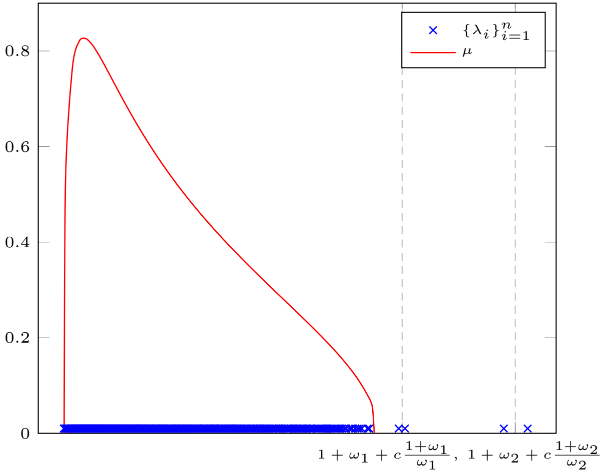


Figure: Eigenvalues of $\frac{1}{n}Y_N Y_N^*$, $\nu_N = \frac{N-4}{N}\boldsymbol{\delta}_1 + \frac{2}{N}\boldsymbol{\delta}_2 + \frac{2}{N}\boldsymbol{\delta}_3$ (hence $\nu = \boldsymbol{\delta}_1$), $N = 500$, $n = 1500$.

Theorem (Eigenvalues [Baik,Silverstein'06])

Let $Y_N = C_N^{\frac{1}{2}} X_N$*, with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $\nu_N = \frac{N - \sum_{m=1}^M k_m}{N} \boldsymbol{\delta}_1 + \sum_{m=1}^M \frac{k_m}{N} \boldsymbol{\delta}_{1+\omega_m}$*, with* $\omega_1 > \ldots > \omega_M$*.*

Theorem (Eigenvalues **[Baik,Silverstein'06]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $\nu_N = \frac{N - \sum_{m=1}^M k_m}{N} \boldsymbol{\delta}_1 + \sum_{m=1}^M \frac{k_m}{N} \boldsymbol{\delta}_{1+\omega_m}$, *with* $\omega_1 > \ldots > \omega_M$.

Then, as $N, n \to \infty$, $N/n \to c \in (0, \infty)$, *with* $\lambda_1 \geq \ldots \geq \lambda_N$, $\lambda_i = \lambda_i(\frac{1}{n} Y_N Y_N^*)$,

- *if* $\omega_m > \sqrt{c}$, *for* $i = 1, \ldots, k_m$,

$$\lambda_{k_1+\ldots+k_{m-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_m + c\frac{1+\omega_m}{\omega_m} > (1+\sqrt{c})^2$$

Theorem (Eigenvalues [Baik,Silverstein'06])

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $\nu_N = \frac{N - \sum_{m=1}^{M} k_m}{N} \boldsymbol{\delta}_1 + \sum_{m=1}^{M} \frac{k_m}{N} \boldsymbol{\delta}_{1+\omega_m}$, *with* $\omega_1 > \ldots > \omega_M$.

Then, as $N, n \to \infty$, $N/n \to c \in (0, \infty)$, *with* $\lambda_1 \geq \ldots \geq \lambda_N$, $\lambda_i = \lambda_i(\frac{1}{n} Y_N Y_N^*)$,

- *if* $\omega_m > \sqrt{c}$, *for* $i = 1, \ldots, k_m$,

$$\lambda_{k_1+\ldots+k_{m-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_m + c\frac{1+\omega_m}{\omega_m} > (1+\sqrt{c})^2$$

- *if* $\omega_m \in (0, \sqrt{c}]$, $\lambda_{k_1+\ldots+k_{m-1}+1} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$

Theorem (Eigenvalues [Baik,Silverstein'06])

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $\nu_N = \frac{N - \sum_{m=1}^{M} k_m}{N} \boldsymbol{\delta}_1 + \sum_{m=1}^{M} \frac{k_m}{N} \boldsymbol{\delta}_{1+\omega_m}$, *with* $\omega_1 > \ldots > \omega_M$.

Then, as $N, n \to \infty$, $N/n \to c \in (0, \infty)$, *with* $\lambda_1 \geq \ldots \geq \lambda_N$, $\lambda_i = \lambda_i(\frac{1}{n} Y_N Y_N^*)$,

- *if* $\omega_m > \sqrt{c}$, *for* $i = 1, \ldots, k_m$,

$$\lambda_{k_1+\ldots+k_{m-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_m + c\frac{1+\omega_m}{\omega_m} > (1+\sqrt{c})^2$$

- *if* $\omega_m \in (0, \sqrt{c}]$, $\lambda_{k_1+\ldots+k_{m-1}+1} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$
- *if* $\omega_{k_j} \in [-\sqrt{c}, 0)$, $\lambda_{\min(N,n)} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$

Theorem (Eigenvalues [Baik,Silverstein'06])

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $\nu_N = \frac{N-\sum_{m=1}^M k_m}{N}\boldsymbol{\delta}_1 + \sum_{m=1}^M \frac{k_m}{N}\boldsymbol{\delta}_{1+\omega_m}$, *with* $\omega_1 > \ldots > \omega_M$.

Then, as $N,n\to\infty$, $N/n \to c \in (0,\infty)$, *with* $\lambda_1 \geq \ldots \geq \lambda_N$, $\lambda_i = \lambda_i(\frac{1}{n}Y_N Y_N^*)$,

- *if* $\omega_m > \sqrt{c}$, *for* $i = 1,\ldots,k_m$,

$$\lambda_{k_1+\ldots+k_{m-1}+i} \xrightarrow{\text{a.s.}} 1+\omega_m + c\frac{1+\omega_m}{\omega_m} > (1+\sqrt{c})^2$$

- *if* $\omega_m \in (0,\sqrt{c}]$, $\lambda_{k_1+\ldots+k_{m-1}+1} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$
- *if* $\omega_{k_j} \in [-\sqrt{c},0)$, $\lambda_{\min(N,n)} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$
- *other eigenvalues discriminated over* c:
  - *if* $\omega_m < -\sqrt{c}$, $c<1$, *for* $i=1,\ldots,k_m$,

$$\lambda_{N-k_M-\ldots-k_m+i} \xrightarrow{\text{a.s.}} 1+\omega_m + c\frac{1+\omega_m}{\omega_m}$$

  - *if* $\omega_{k_j} < -\sqrt{c}$, $c>1$, $\lambda_{n-k_M-\ldots-k_m+1} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$

Proof

- **Two ingredients**: Algebraic calculus + trace lemma

Proof

- **Two ingredients**: Algebraic calculus + trace lemma
- **Find eigenvalues away from eigenvalues of $\frac{1}{n}X_N X_N^*$:**

$$\begin{aligned} 0 &= \det\left(\frac{1}{n}Y_N Y_N^* - \lambda I_N\right) \\ &= \det(C_N)\det\left(\frac{1}{n}X_N X_N^* - \lambda C_N^{-1}\right) \\ &= \det\left(\frac{1}{n}X_N X_N^* - \lambda I_N + \lambda(I_N - C_N^{-1})\right) \\ &= \det\left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)\det\left(I_N + \lambda(I_N - C_N^{-1})\left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)^{-1}\right). \end{aligned}$$

Proof

- **Two ingredients**: Algebraic calculus + trace lemma
- **Find eigenvalues away from eigenvalues of** $\frac{1}{n}X_NX_N^*$:

$$\begin{aligned}
0 &= \det\left(\frac{1}{n}Y_NY_N^* - \lambda I_N\right) \\
&= \det(C_N)\det\left(\frac{1}{n}X_NX_N^* - \lambda C_N^{-1}\right) \\
&= \det\left(\frac{1}{n}X_NX_N^* - \lambda I_N + \lambda(I_N - C_N^{-1})\right) \\
&= \det\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)\det\left(I_N + \lambda(I_N - C_N^{-1})\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)^{-1}\right).
\end{aligned}$$

- **Use low rank property**:

$$I_N - C_N^{-1} = I_N - (I_N + U\Omega U^*)^{-1} = U(I_K + \Omega^{-1})^{-1}U^*,\ \Omega \in \mathbb{C}^{K\times K},\ K = \sum k_m.$$

Hence

$$0 = \det\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)\det\left(I_N + U(I_K + \Omega^{-1})^{-1}U^*\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)^{-1}\right)$$

Proof

- **Two ingredients**: Algebraic calculus + trace lemma
- **Find eigenvalues away from eigenvalues of $\frac{1}{n}X_NX_N^*$**:

$$\begin{aligned}
0 &= \det\left(\frac{1}{n}Y_NY_N^* - \lambda I_N\right) \\
&= \det(C_N)\det\left(\frac{1}{n}X_NX_N^* - \lambda C_N^{-1}\right) \\
&= \det\left(\frac{1}{n}X_NX_N^* - \lambda I_N + \lambda(I_N - C_N^{-1})\right) \\
&= \det\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)\det\left(I_N + \lambda(I_N - C_N^{-1})\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)^{-1}\right).
\end{aligned}$$

- **Use low rank property**:

$$I_N - C_N^{-1} = I_N - (I_N + U\Omega U^*)^{-1} = U(I_K + \Omega^{-1})^{-1}U^*,\ \Omega \in \mathbb{C}^{K\times K},\ K = \sum k_m.$$

Hence

$$0 = \det\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)\det\left(I_N + U(I_K + \Omega^{-1})^{-1}U^*\left(\frac{1}{n}X_NX_N^* - \lambda I_N\right)^{-1}\right)$$

Proof (2)

- **Sylverster's identity** $(\det(I + AB) = \det(I + BA))$,

$$0 = \det\left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)\det\left(I_K + (I_K + \Omega^{-1})^{-1} U^* \left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)^{-1} U\right)$$

Proof (2)

- **Sylverster's identity** $(\det(I + AB) = \det(I + BA))$,

$$0 = \det\left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)\det\left(I_K + (I_K + \Omega^{-1})^{-1}U^*\left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)^{-1}U\right)$$

- **No eigenvalue outside the support [Bai,Sil'98]**: $\det(\frac{1}{n}X_N X_N^* - \lambda I_N)$ has no zero beyond $(1 + \sqrt{c})^2$ for all large n a.s.

Proof (2)

- **Sylverster's identity** $(\det(I+AB)=\det(I+BA))$,

$$0=\det\left(\frac{1}{n}X_NX_N^*-\lambda I_N\right)\det\left(I_K+(I_K+\Omega^{-1})^{-1}U^*\left(\frac{1}{n}X_NX_N^*-\lambda I_N\right)^{-1}U\right)$$

- **No eigenvalue outside the support [Bai,Sil'98]**: $\det(\frac{1}{n}X_NX_N^*-\lambda I_N)$ has no zero beyond $(1+\sqrt{c})^2$ for all large n a.s.
- **Extension of Trace Lemma**: for each $z\in\mathbb{C}\setminus\operatorname{supp}(\mu)$,

$$U^*\left(\frac{1}{n}X_NX_N^*-zI_N\right)^{-1}U\xrightarrow{\text{a.s.}}m_\mu(z)I_K.$$

Proof (2)

- **Sylverster's identity** ($\det(I + AB) = \det(I + BA)$),

$$0 = \det\left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)\det\left(I_K + (I_K + \Omega^{-1})^{-1}U^*\left(\frac{1}{n}X_N X_N^* - \lambda I_N\right)^{-1}U\right)$$

- **No eigenvalue outside the support [Bai,Sil'98]**: $\det(\frac{1}{n}X_N X_N^* - \lambda I_N)$ has no zero beyond $(1+\sqrt{c})^2$ for all large n a.s.
- **Extension of Trace Lemma**: for each $z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$,

$$U^*\left(\frac{1}{n}X_N X_N^* - zI_N\right)^{-1}U \xrightarrow{\text{a.s.}} m_\mu(z) I_K.$$

- As a result, for all large n a.s.,

$$\begin{aligned} 0 &= \det\left(I_K + (I_K + \Omega^{-1})^{-1}U^*(\frac{1}{n}X_N X_N^* - \lambda I_N)^{-1}U\right) \\ &\simeq \prod_{m=1}^{M}\left(1 + \frac{1}{1+\omega_m^{-1}}m_\mu(\lambda)\right)^{k_m} \end{aligned}$$

Proof (3)

- **Limiting solutions:** zeros (with multiplicity) of

$$1 + \frac{\omega_m}{1+\omega_m} m_\mu(\lambda) = 0.$$

Proof (3)

- **Limiting solutions:** zeros (with multiplicity) of

$$1 + \frac{\omega_m}{1+\omega_m} m_\mu(\lambda) = 0.$$

- Using Marčenko–Pastur law properties $(m_\mu(z) = (1 - c - z - czm_\mu(z))^{-1})$,

$$\lambda \in \left\{ 1 + \omega_m + c\frac{1+\omega_m}{\omega_m} \right\}_{m=1}^{M}.$$

Theorem (Eigenvectors **[Paul'07]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $C_N = I_N + \sum_{m=1}^{M} \omega_m U_m U_m^*$, $U_m \in \mathbb{C}^{N \times k_m}$ *eigenbasis*, $\omega_1 > \ldots > \omega_M > 0$.

Theorem (Eigenvectors **[Paul'07]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $C_N = I_N + \sum_{m=1}^{M} \omega_m U_m U_m^*$, $U_m \in \mathbb{C}^{N \times k_m}$ *eigenbasis,* $\omega_1 > \ldots > \omega_M > 0$.

Then, as $N, n \to \infty$, $N/n \to c \in (0, \infty)$, *with* $\hat{u}_i$ *eigenvectors of* $\lambda_i(\frac{1}{n} Y_N Y_N^*)$,

- *if* $\omega_m > \sqrt{c}$, *with* $\hat{U}_m = [\hat{u}_{k_1+\ldots+k_{m-1}+1}, \ldots, \hat{u}_{k_1+\ldots+k_m}]$,

$$\hat{U}_m^* U_m U_m^* \hat{U}_m \xrightarrow{\text{a.s.}} \frac{1 - c\omega_m^{-2}}{1 + c\omega^{-1}} I_{k_m}$$

Theorem (Eigenvectors **[Paul'07]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $C_N = I_N + \sum_{m=1}^{M} \omega_m U_m U_m^*$, $U_m \in \mathbb{C}^{N \times k_m}$ *eigenbasis,* $\omega_1 > \ldots > \omega_M > 0$.

Then, as $N, n \to \infty$, $N/n \to c \in (0, \infty)$, *with* $\hat{u}_i$ *eigenvectors of* $\lambda_i(\frac{1}{n} Y_N Y_N^*)$,

- *if* $\omega_m > \sqrt{c}$, *with* $\hat{U}_m = [\hat{u}_{k_1+\ldots+k_{m-1}+1}, \ldots, \hat{u}_{k_1+\ldots+k_m}]$,

$$\hat{U}_m^* U_m U_m^* \hat{U}_m \xrightarrow{\text{a.s.}} \frac{1 - c\omega_m^{-2}}{1 + c\omega^{-1}} I_{k_m}$$

- *if* $\omega_m < \sqrt{c}$, $\hat{U}_m^* U_m U_m^* \hat{U}_m \xrightarrow{\text{a.s.}} 0$.

Theorem (Eigenvectors **[Paul'07]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. zero mean, unit variance, finite fourth order moment entries*
- $C_N = I_N + \sum_{m=1}^{M} \omega_m U_m U_m^*$, $U_m \in \mathbb{C}^{N \times k_m}$ *eigenbasis,* $\omega_1 > \ldots > \omega_M > 0$.

Then, as $N, n \to \infty$, $N/n \to c \in (0, \infty)$, *with* $\hat{u}_i$ *eigenvectors of* $\lambda_i(\frac{1}{n} Y_N Y_N^*)$,

- *if* $\omega_m > \sqrt{c}$, *with* $\hat{U}_m = [\hat{u}_{k_1+\ldots+k_{m-1}+1}, \ldots, \hat{u}_{k_1+\ldots+k_m}]$,

$$\hat{U}_m^* U_m U_m^* \hat{U}_m \xrightarrow{\text{a.s.}} \frac{1 - c\omega_m^{-2}}{1 + c\omega^{-1}} I_{k_m}$$

- *if* $\omega_m < \sqrt{c}$, $\hat{U}_m^* U_m U_m^* \hat{U}_m \xrightarrow{\text{a.s.}} 0$.

Proof: *Based on Cauchy integral + similar ingredients as eigenvalue proof*

$$a^* U_m U_m^* b = \frac{1}{2\pi\imath} \oint_{\mathcal{C}_m} a^* \left(\frac{1}{n} Y_N Y_N^* - z I_N \right)^{-1} b dz$$

for $\mathcal{C}_m$ *contour circling around* $\lambda_{k_1+\ldots+k_{m-1}+1}, \ldots, \lambda_{k_1+\ldots+k_m}$ *only.*

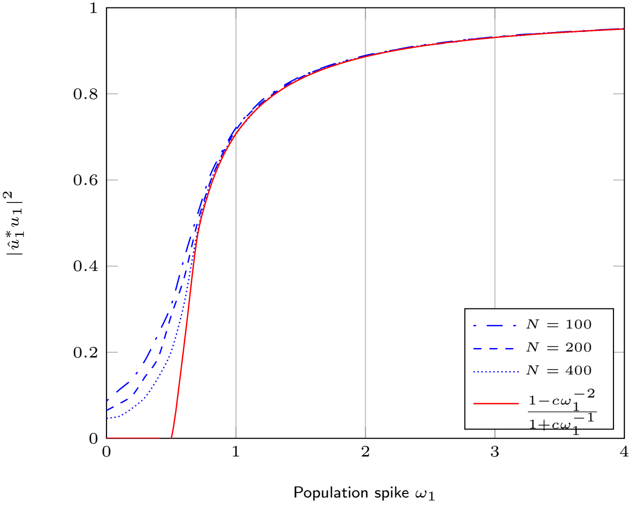


Figure: Simulated versus limiting $|\hat{u}_1^* u_1|^2$ for $Y_N = C_N^{\frac{1}{2}} X_N$, $C_N = I_N + \omega_1 u_1 u_1^*$, $N/n = 1/3$, varying ω_1.

Outline

Theorem (Fluctuations of Linear Statistics **[Bai,Silverstein'04]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, X_N *with i.i.d. complex Gaussian zero mean, unit variance entries. Assume* $N(\nu_N - \nu) \xrightarrow{\mathcal{L}} 0$ *and let* $\Delta_{\mu,N} = N(\mu_N - \mu)$*. Then, as* $N/n \to \infty$*,* $N/n = c + o(N^{-1})$*, for* $f_1, \ldots, f_k$ *analytic,*

$$\left(\int f_1(x) \Delta_{\mu,N}(dx), \ldots, \int f_k(x) \Delta_{\mu,N}(dx) \right) \xrightarrow{\mathcal{L}} (X_{f_1}, \ldots, X_{f_k})$$

Gaussian vector with zero mean and covariance matrix

$$\mathrm{Cov}(X_f, X_g) = -\frac{1}{4\pi^2} \oint \oint \frac{f(z_1) g(z_2)}{(m_{\tilde{\mu}}(z_1) - m_{\tilde{\mu}}(z_2))^2} m'_{\tilde{\mu}}(z_1) m'_{\tilde{\mu}}(z_2) dz_1 dz_2.$$

Theorem (Fluctuations of Linear Statistics **[Bai,Silverstein'04]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, X_N *with i.i.d. complex Gaussian zero mean, unit variance entries.*
Assume $N(\nu_N - \nu) \xrightarrow{\mathcal{L}} 0$ *and let* $\Delta_{\mu,N} = N(\mu_N - \mu)$. *Then, as* $N/n \to \infty$, $N/n = c + o(N^{-1})$, *for* $f_1, \ldots, f_k$ *analytic,*

$$\left(\int f_1(x) \Delta_{\mu,N}(dx), \ldots, \int f_k(x) \Delta_{\mu,N}(dx) \right) \xrightarrow{\mathcal{L}} (X_{f_1}, \ldots, X_{f_k})$$

Gaussian vector with zero mean and covariance matrix

$$\mathrm{Cov}(X_f, X_g) = -\frac{1}{4\pi^2} \oint \oint \frac{f(z_1) g(z_2)}{(m_{\tilde{\mu}}(z_1) - m_{\tilde{\mu}}(z_2))^2} m'_{\tilde{\mu}}(z_1) m'_{\tilde{\mu}}(z_2) dz_1 dz_2.$$

Several generalizations:

- if $N(\nu_N - \nu) \not\to 0$, valid with $\Delta_{\mu,N} = N(\mu_N - \bar{\mu}_N)$, $\bar{\mu}_N$ deterministic equivalent.

Theorem (Fluctuations of Linear Statistics [Bai,Silverstein'04])

Let $Y_N = C_N^{\frac{1}{2}} X_N$, X_N *with i.i.d. complex Gaussian zero mean, unit variance entries.*
Assume $N(\nu_N - \nu) \xrightarrow{\mathcal{L}} 0$ *and let* $\Delta_{\mu,N} = N(\mu_N - \mu)$. *Then, as* $N/n \to \infty$, $N/n = c + o(N^{-1})$, *for* $f_1, \ldots, f_k$ *analytic,*

$$\left(\int f_1(x)\Delta_{\mu,N}(dx), \ldots, \int f_k(x)\Delta_{\mu,N}(dx) \right) \xrightarrow{\mathcal{L}} (X_{f_1}, \ldots, X_{f_k})$$

Gaussian vector with zero mean and covariance matrix

$$\mathrm{Cov}(X_f, X_g) = -\frac{1}{4\pi^2} \oint \oint \frac{f(z_1)g(z_2)}{(m_{\tilde{\mu}}(z_1) - m_{\tilde{\mu}}(z_2))^2} m'_{\tilde{\mu}}(z_1) m'_{\tilde{\mu}}(z_2) dz_1 dz_2.$$

Several generalizations:

- if $N(\nu_N - \nu) \not\to 0$, valid with $\Delta_{\mu,N} = N(\mu_N - \bar{\mu}_N)$, $\bar{\mu}_N$ deterministic equivalent.
- non-Gaussian X_N: additional bias and variance terms proportional to kurtosis.

Theorem (Fluctuations of Linear Statistics **[Bai,Silverstein'04]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, X_N *with i.i.d. complex Gaussian zero mean, unit variance entries. Assume* $N(\nu_N - \nu) \xrightarrow{\mathcal{L}} 0$ *and let* $\Delta_{\mu,N} = N(\mu_N - \mu)$*. Then, as* $N/n \to \infty$, $N/n = c + o(N^{-1})$, *for* $f_1, \ldots, f_k$ *analytic,*

$$\left(\int f_1(x) \Delta_{\mu,N}(dx), \ldots, \int f_k(x) \Delta_{\mu,N}(dx) \right) \xrightarrow{\mathcal{L}} (X_{f_1}, \ldots, X_{f_k})$$

Gaussian vector with zero mean and covariance matrix

$$\mathrm{Cov}(X_f, X_g) = -\frac{1}{4\pi^2} \oint \oint \frac{f(z_1) g(z_2)}{(m_{\tilde{\mu}}(z_1) - m_{\tilde{\mu}}(z_2))^2} m'_{\tilde{\mu}}(z_1) m'_{\tilde{\mu}}(z_2) dz_1 dz_2.$$

Several generalizations:

- if $N(\nu_N - \nu) \not\to 0$, valid with $\Delta_{\mu,N} = N(\mu_N - \bar{\mu}_N)$, $\bar{\mu}_N$ deterministic equivalent.
- non-Gaussian X_N: additional bias and variance terms proportional to kurtosis.
- non-analytic f: result holds for C^3 functions, different result for not C^1 functions.

Theorem (Fluctuations of Linear Statistics **[Bai,Silverstein'04]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, X_N with i.i.d. complex Gaussian zero mean, unit variance entries. Assume $N(\nu_N - \nu) \xrightarrow{\mathcal{L}} 0$ and let $\Delta_{\mu,N} = N(\mu_N - \mu)$. Then, as $N/n \to \infty$, $N/n = c + o(N^{-1})$, for $f_1, \ldots, f_k$ analytic,

$$\left(\int f_1(x) \Delta_{\mu,N}(dx), \ldots, \int f_k(x) \Delta_{\mu,N}(dx) \right) \xrightarrow{\mathcal{L}} (X_{f_1}, \ldots, X_{f_k})$$

Gaussian vector with zero mean and covariance matrix

$$\mathrm{Cov}(X_f, X_g) = -\frac{1}{4\pi^2} \oint \oint \frac{f(z_1) g(z_2)}{(m_{\tilde{\mu}}(z_1) - m_{\tilde{\mu}}(z_2))^2} m'_{\tilde{\mu}}(z_1) m'_{\tilde{\mu}}(z_2) dz_1 dz_2.$$

Several generalizations:

- if $N(\nu_N - \nu) \not\to 0$, valid with $\Delta_{\mu,N} = N(\mu_N - \bar{\mu}_N)$, $\bar{\mu}_N$ deterministic equivalent.
- non-Gaussian X_N: additional bias and variance terms proportional to kurtosis.
- non-analytic f: result holds for C^3 functions, different result for not C^1 functions.
- CLT also exist for bilinear forms: e.g., $a^*(\frac{1}{n} Y_N Y_N^* - z I_N)^{-1} b$.

Theorem (Phase Transition **[Baik,BenArous,Péché'05]**)

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d.complex Gaussian zero mean, unit variance entries,*
- $\nu_N = \frac{N - \sum_{m=1}^{M} k_m}{N} \boldsymbol{\delta}_1 + \sum_{m=1}^{M} \frac{k_m}{N} \boldsymbol{\delta}_{1+\omega_m}$, $\omega_1 > \ldots > \omega_M > 0$ *(*$M \geq 0$*).*

Theorem (Phase Transition [Baik,BenArous,Péché'05])

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d.complex Gaussian zero mean, unit variance entries,*
- $\nu_N = \frac{N - \sum_{m=1}^{M} k_m}{N} \boldsymbol{\delta}_1 + \sum_{m=1}^{M} \frac{k_m}{N} \boldsymbol{\delta}_{1+\omega_m}$, $\omega_1 > \ldots > \omega_M > 0$ *(*$M \geq 0$*).*

Then, as $N, n \to \infty$, $N/n \to c < 1$,

- *If* $\omega_1 < \sqrt{c}$ *(or* $M = 0$*),*

$$N^{\frac{2}{3}} \frac{\lambda_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{\frac{1}{2}}} \xrightarrow{\mathcal{L}} T_2, \text{ (complex Tracy–Widom law)}$$

Tracy–Widom Theorem

Theorem (Phase Transition [Baik,BenArous,Péché'05])

Let $Y_N = C_N^{\frac{1}{2}} X_N$, *with*

- X_N *with i.i.d. complex Gaussian zero mean, unit variance entries,*
- $\nu_N = \frac{N - \sum_{m=1}^{M} k_m}{N} \boldsymbol{\delta}_1 + \sum_{m=1}^{M} \frac{k_m}{N} \boldsymbol{\delta}_{1+\omega_m}$, $\omega_1 > \ldots > \omega_M > 0$ *(*$M \geq 0$*).*

Then, as $N, n \to \infty$, $N/n \to c < 1$,

- *If* $\omega_1 < \sqrt{c}$ *(or* $M = 0$*),*

$$N^{\frac{2}{3}} \frac{\lambda_1 - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{\frac{1}{2}}} \xrightarrow{\mathcal{L}} T_2, \text{ (complex Tracy–Widom law)}$$

- *If* $\omega_1 > \sqrt{c}$,

$$\left(\frac{(1+\omega_1)^2}{c} - \frac{(1+\omega_1)^2}{\omega_1^2} \right)^{\frac{1}{2}} N^{\frac{1}{2}} \left[\lambda_1 - \left(1 + \omega_1 + c \frac{1+\omega_1}{\omega_1} \right) \right] \xrightarrow{\mathcal{L}} G_k$$

with G_k *law of largest eigenvalue of the* $k \times k$ *GUE matrix. In particular,* $G_1(x)$ *real Gaussian distribution function.*

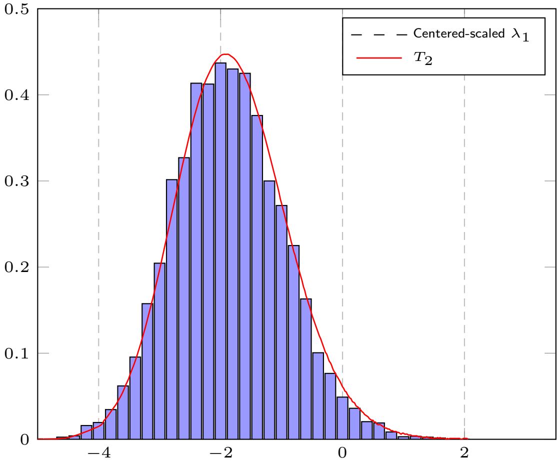


Figure: Distribution of $N^{\frac{2}{3}} c^{-\frac{1}{2}} (1+\sqrt{c})^{-\frac{4}{3}} \left[\lambda_1(\frac{1}{n} X_N X_N^*) - (1+\sqrt{c})^2\right]$ versus Tracy–Widom (T_2), $N = 500$, $n = 1500$.

Outline

Source Detection and Enumeration

Context. Observations $y_1, \ldots, y_n \in \mathbb{C}^N$, independent with

$$y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ h s_i + \sigma w_i & , \mathcal{H}_1 \end{cases}$$

with $\sigma > 0$ unknown, $s_i \in \mathbb{C}$ random, $h \in \mathbb{C}^N$ unknown, $w_i \sim \mathcal{CN}(0, I_N)$.

Context. Observations $y_1, \ldots, y_n \in \mathbb{C}^N$, independent with

$$y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ h s_i + \sigma w_i & , \mathcal{H}_1 \end{cases}$$

with $\sigma > 0$ unknown, $s_i \in \mathbb{C}$ random, $h \in \mathbb{C}^N$ unknown, $w_i \sim \mathcal{CN}(0, I_N)$.

GLRT procedure. [Bianchi,Najim,Debbah,Maida'10]

- Hypothesis test for $Y_N = [y_1, \ldots, y_n]$

$$\frac{\sup_{h,s,\sigma} p(Y_N | h, s, \sigma)}{\sup_\sigma p(Y_N | \sigma)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \Gamma.$$

Context. Observations $y_1, \ldots, y_n \in \mathbb{C}^N$, independent with

$$y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ h s_i + \sigma w_i & , \mathcal{H}_1 \end{cases}$$

with $\sigma > 0$ unknown, $s_i \in \mathbb{C}$ random, $h \in \mathbb{C}^N$ unknown, $w_i \sim \mathcal{CN}(0, I_N)$.

GLRT procedure. [Bianchi,Najim,Debbah,Maida'10]

- Hypothesis test for $Y_N = [y_1, \ldots, y_n]$

$$\frac{\sup_{h,s,\sigma} p(Y_N | h, s, \sigma)}{\sup_\sigma p(Y_N | \sigma)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \Gamma.$$

- Provably equivalent to hypothesis test (with $\lambda_i = \lambda_i(\frac{1}{n} Y_N Y_N^*)$)

$$T(Y_N) \triangleq \frac{\lambda_1}{\frac{1}{N}\sum_{i=1}^N \lambda_i} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma$$

  for some γ continuous growing with Γ.

Context. Observations $y_1, \ldots, y_n \in \mathbb{C}^N$, independent with

$$y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ h s_i + \sigma w_i & , \mathcal{H}_1 \end{cases}$$

with $\sigma > 0$ unknown, $s_i \in \mathbb{C}$ random, $h \in \mathbb{C}^N$ unknown, $w_i \sim \mathcal{CN}(0, I_N)$.

GLRT procedure. [Bianchi,Najim,Debbah,Maida'10]

- Hypothesis test for $Y_N = [y_1, \ldots, y_n]$

$$\frac{\sup_{h,s,\sigma} p(Y_N | h, s, \sigma)}{\sup_\sigma p(Y_N | \sigma)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \Gamma.$$

- Provably equivalent to hypothesis test (with $\lambda_i = \lambda_i(\frac{1}{n} Y_N Y_N^*)$)

$$T(Y_N) \triangleq \frac{\lambda_1}{\frac{1}{N}\sum_{i=1}^N \lambda_i} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma$$

  for some γ continuous growing with Γ.
- As $N, n \to \infty$, $N/n \to c$, $\frac{1}{N}\sum_{i=1}^N \lambda_i \xrightarrow{\text{a.s.}} 1$, and

$$P\left(N^{\frac{2}{3}} (1 + \sqrt{c})^{-\frac{4}{3}} c^{-\frac{1}{2}} \left(T(Y_N) - (1 + \sqrt{c})^2 \right) < \gamma \mid \mathcal{H}_0 \right) \to T_2(\gamma)$$

Context. Observations $y_1,\ldots,y_n \in \mathbb{C}^N$, independent with

$$y_i = \begin{cases} \sigma w_i & , \mathcal{H}_0 \\ h s_i + \sigma w_i & , \mathcal{H}_1 \end{cases}$$

with $\sigma > 0$ unknown, $s_i \in \mathbb{C}$ random, $h \in \mathbb{C}^N$ unknown, $w_i \sim \mathcal{CN}(0, I_N)$.

GLRT procedure. [Bianchi,Najim,Debbah,Maida'10]

- Hypothesis test for $Y_N = [y_1,\ldots,y_n]$

$$\frac{\sup_{h,s,\sigma} p(Y_N|h,s,\sigma)}{\sup_\sigma p(Y_N|\sigma)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \Gamma.$$

- Provably equivalent to hypothesis test (with $\lambda_i = \lambda_i(\frac{1}{n} Y_N Y_N^*)$)

$$T(Y_N) \triangleq \frac{\lambda_1}{\frac{1}{N}\sum_{i=1}^N \lambda_i} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma$$

  for some γ continuous growing with Γ.
- As $N, n \to \infty$, $N/n \to c$, $\frac{1}{N}\sum_{i=1}^N \lambda_i \xrightarrow{\text{a.s.}} 1$, and

$$P\left(N^{\frac{2}{3}}(1+\sqrt{c})^{-\frac{4}{3}} c^{-\frac{1}{2}} \left(T(Y_N) - (1+\sqrt{c})^2\right) < \gamma \mid \mathcal{H}_0\right) \to T_2(\gamma)$$

- Setting false alarm rate to η implies $\gamma \geq T_2^{-1}(\eta)$, i.e., test

$$T(Y_N) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} (1+\sqrt{c})^2 + N^{-\frac{2}{3}} T_2^{-1}(\eta)(1+\sqrt{c})^{\frac{4}{3}} c^{\frac{1}{2}}.$$

Power Estimation

Context. Estimation of $p_1, \ldots, p_k$ from $y_1, \ldots, y_n$ independent with

$$y_i = \sum_{k=1}^{K} \sqrt{p_k} H_k s_{k,i} + \sigma w_i$$

$H_k \in \mathbb{C}^{N \times N_k}$ i.i.d. zero mean, variance $1/N$, $s_{k,i}$ i.i.d. zero mean, unit variance.

Context. Estimation of $p_1, \ldots, p_k$ from $y_1, \ldots, y_n$ independent with

$$y_i = \sum_{k=1}^{K} \sqrt{p_k} H_k s_{k,i} + \sigma w_i$$

$H_k \in \mathbb{C}^{N \times N_k}$ i.i.d. zero mean, variance $1/N$, $s_{k,i}$ i.i.d. zero mean, unit variance.

- **Spike regime**: N_k fixed, use $\lambda_i \to 1 + p_m + c\frac{1+p_m}{p_m}$, $(H_m^* H_{m'} \xrightarrow{\text{a.s.}} \boldsymbol{\delta}_{m-m'} I_{N_k})$

$$p_m \xleftarrow{\text{a.s.}} \hat{p}_m = \frac{1}{N_m} \sum_{i \in \mathcal{N}_m} \frac{\lambda_i - (c+1)}{2} + \sqrt{(c+1-\lambda_i)^2 - 4c}.$$

Power Estimation

Context. Estimation of $p_1, \ldots, p_k$ from $y_1, \ldots, y_n$ independent with

$$y_i = \sum_{k=1}^{K} \sqrt{p_k} H_k s_{k,i} + \sigma w_i$$

$H_k \in \mathbb{C}^{N \times N_k}$ i.i.d. zero mean, variance $1/N$, $s_{k,i}$ i.i.d. zero mean, unit variance.

- **Spike regime**: N_k fixed, use $\lambda_i \to 1 + p_m + c\frac{1+p_m}{p_m}$, $(H_m^* H_{m'} \xrightarrow{\text{a.s.}} \boldsymbol{\delta}_{m-m'} I_{N_k})$

$$p_m \xleftarrow{\text{a.s.}} \hat{p}_m = \frac{1}{N_m} \sum_{i \in \mathcal{N}_m} \frac{\lambda_i - (c+1)}{2} + \sqrt{(c+1-\lambda_i)^2 - 4c}.$$

- **Large N_k regime**: $N_k/N \to c_k > 0$, use contour integration method

$$p_m \xleftarrow{\text{a.s.}} \hat{p}_m = \begin{cases} \frac{Nn}{N_m(n-N)} \sum_{i \in \mathcal{N}_m} (\eta_i - \kappa_i) & , n \neq N \\ \frac{N}{N_m(N - \sum_i N_i)} \sum_{i \in \mathcal{N}_m} \left(\sum_{j=1}^{N} \frac{\eta_i}{(\lambda_j - \eta_i)^2} \right)^{-1} & , n = N \end{cases}$$

(under cluster m separability condition) with

- $\eta_1 \geq \ldots \geq \eta_N$ eigenvalues of $\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^*$
- $\kappa_1 \geq \ldots \geq \kappa_N$ eigenvalues of $\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^*$

(details in **[Couillet,Silverstein,Bai,Debbah'11]**).

Power Estimation

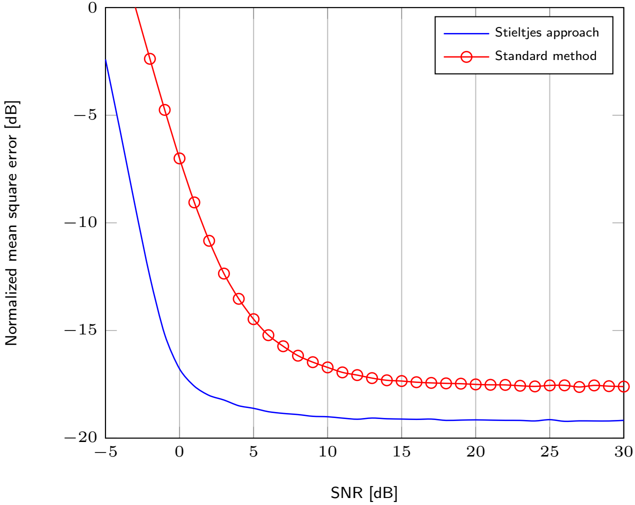


Figure: Estimate NMSE for p_1 (**large N_k regime**), three sources, $p_1 = 1$, $p_2 = 1/4$, $p_3 = 1/16$, $N_1 = N_2 = N_3 = 4$,$N = 24$, $n = 128$. Comparison between standard statistics (assumes $n \gg N \gg N_k$) and Stieltjes transform approach.

Power Estimation

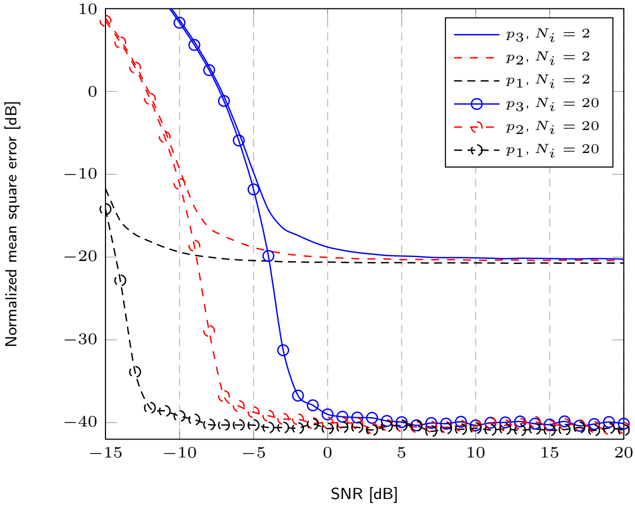


Figure: Estimate NMSE (**large N_k regime**), $p_1 = 10$, $p_2 = 3$, $p_3 = 1$, $N_1 = N_2 = N_3$,$\sum N_k/N = N/n = 1/10$.

Context. Estimate $\theta_1, \ldots, \theta_M$ from $y_1, \ldots, y_n \in \mathbb{C}^N$ independent,

$$y_i = \sum_{m=1}^{M} \sqrt{p_m} a(\theta_m) s_{m,i} + \sigma w_i$$

with $a(\theta) \in \mathbb{C}^N$ steering vector, e.g., ULA case $[a(\theta)]_k = \frac{1}{\sqrt{N}} \exp(2\pi\imath k \sin(\theta) d)$.

Context. Estimate $\theta_1, \ldots, \theta_M$ from $y_1, \ldots, y_n \in \mathbb{C}^N$ independent,

$$y_i = \sum_{m=1}^{M} \sqrt{p_m} a(\theta_m) s_{m,i} + \sigma w_i$$

with $a(\theta) \in \mathbb{C}^N$ steering vector, e.g., ULA case $[a(\theta)]_k = \frac{1}{\sqrt{N}} \exp(2\pi \imath k \sin(\theta) d)$.

MUSIC. Uses $a(\theta)^* U_W = 0$, $U_W = [u_{M+1}, \ldots, u_N]$ "noise subspace" of $E[y_1 y_1^*]$.

Context. Estimate $\theta_1, \ldots, \theta_M$ from $y_1, \ldots, y_n \in \mathbb{C}^N$ independent,

$$y_i = \sum_{m=1}^{M} \sqrt{p_m} a(\theta_m) s_{m,i} + \sigma w_i$$

with $a(\theta) \in \mathbb{C}^N$ steering vector, e.g., ULA case $[a(\theta)]_k = \frac{1}{\sqrt{N}} \exp(2\pi\imath k \sin(\theta) d)$.

MUSIC. Uses $a(\theta)^* U_W = 0$, $U_W = [u_{M+1}, \ldots, u_N]$ "noise subspace" of $E[y_1 y_1^*]$.

- **Classical MUSIC**: by $\hat{U}_W \hat{U}_W^* \xrightarrow{\text{a.s.}} U_W U_W^*$ ($\hat{U}_W = [\hat{u}_{M+1}, \ldots, \hat{u}_N]$) as $n \to \infty$,

$$\hat{\theta}_1, \ldots, \hat{\theta}_M = \text{ deep minima of } \left\{ 1 - \sum_{k=1}^{M} a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta) \right\}.$$

Context. Estimate $\theta_1, \ldots, \theta_M$ from $y_1, \ldots, y_n \in \mathbb{C}^N$ independent,

$$y_i = \sum_{m=1}^{M} \sqrt{p_m} a(\theta_m) s_{m,i} + \sigma w_i$$

with $a(\theta) \in \mathbb{C}^N$ steering vector, e.g., ULA case $[a(\theta)]_k = \frac{1}{\sqrt{N}} \exp(2\pi \imath k \sin(\theta) d)$.

MUSIC. Uses $a(\theta)^* U_W = 0$, $U_W = [u_{M+1}, \ldots, u_N]$ "noise subspace" of $E[y_1 y_1^*]$.

- **Classical MUSIC**: by $\hat{U}_W \hat{U}_W^* \xrightarrow{\text{a.s.}} U_W U_W^*$ ($\hat{U}_W = [\hat{u}_{M+1}, \ldots, \hat{u}_N]$) as $n \to \infty$,

$$\hat{\theta}_1, \ldots, \hat{\theta}_M = \text{ deep minima of } \left\{ 1 - \sum_{k=1}^{M} a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta) \right\}.$$

- **Spiked (G)-MUSIC**: by $a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta) \xrightarrow{\text{a.s.}} \frac{1 - c\hat{p}_k^{-2}}{1 + c\hat{p}_k^{-1}} a(\theta)^* u_k u_k^* a(\theta)$,

$$\hat{\theta}_1, \ldots, \hat{\theta}_M = \text{ deep minima of } \left\{ 1 - \sum_{k=M+1}^{N} \frac{1 + c\hat{p}_k^{-1}}{1 - c\hat{p}_k^{-2}} a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta) \right\}.$$

(details in **[Mestre,Lagunas'08],[Hachem et al.'13]**).

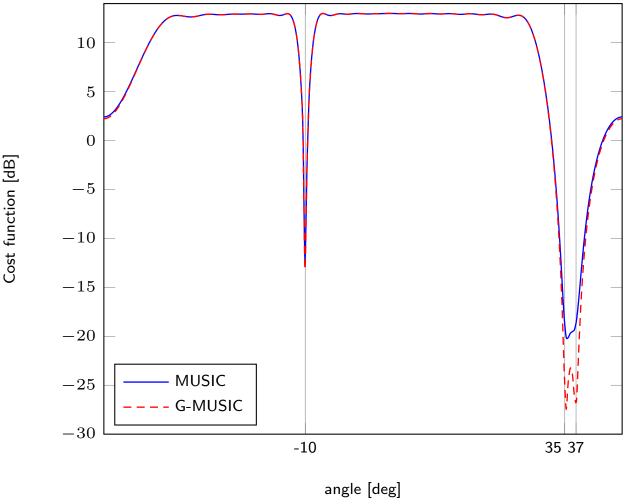


Figure: MUSIC versus G-MUSIC, $M = 3$ sources, $N = 20$, $n = 150$, $\sigma^2 = 0.1$. Angles $\theta_1 = 10^\circ$, $\theta_2 = 35^\circ$, $\theta_3 = 37^\circ$.

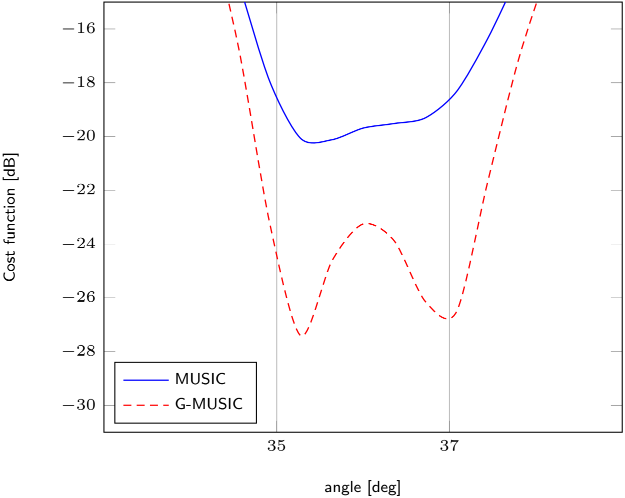


Figure: MUSIC versus G-MUSIC, $M = 3$ sources, $N = 20$, $n = 150$, $\sigma^2 = 0.1$. Angles $\theta_1 = 10^\circ$, $\theta_2 = 35^\circ$, $\theta_3 = 37^\circ$.

Other SP applications

- **Covariance Matrix Estimation**. Inconsistent in large N, n regime, but
  - **Linear shrinkage**: optimize ρ in estimate $\hat{C}_N = (1-\rho)\frac{1}{n}Y_N Y_N^* + \rho I_N$
  - **Non-linear shrinkage**: optimize $\hat{C}_N = U f(\Lambda) U^*$, with $\frac{1}{n}Y_N Y_N^* = U \Lambda U^*$.

- **Covariance Matrix Estimation**. Inconsistent in large N, n regime, but
  - **Linear shrinkage**: optimize ρ in estimate $\hat{C}_N = (1-\rho)\frac{1}{n}Y_N Y_N^* + \rho I_N$
  - **Non-linear shrinkage**: optimize $\hat{C}_N = U f(\Lambda) U^*$, with $\frac{1}{n}Y_N Y_N^* = U\Lambda U^*$.

- **Improved (sparse) PCA**. PCA on $\frac{1}{n}Y_{\mathcal{I}}Y_{\mathcal{I}}^*$, $\mathcal{I} \subset \{1,\ldots,N\}$, $Y_{\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|\times N}$ such that

$$\left|(\hat{u}_{1,\mathcal{I}}^e)^* u_1\right|^2 \text{ maximum}$$

with $\hat{u}_{1,\mathcal{I}}^e \in \mathbb{C}^N$ PCA vector of $\frac{1}{n}Y_{\mathcal{I}}Y_{\mathcal{I}}^*$ extended with zeros.

- **Toepitz covariance matrices**: Toeplitzification $\mathcal{T}(\frac{1}{n}Y_N Y_N^*)$ consistent in large N, n regime,

$$\left\| \mathcal{T}\left(\frac{1}{n}Y_N Y_N^*\right) - C_N \right\| \xrightarrow{\text{a.s.}} 0$$

(with $[\mathcal{T}(X)]_{ij} = \frac{1}{N}\sum_k [X]_{k,k+(i-j)}$). Many results beyond RMT regime **[Bickel,Levina'08],[Wu,Pouramadi'09],[Vinogradova,Couillet,Hachem'14]**.

- **Toepitz covariance matrices**: Toeplitzification $\mathcal{T}(\frac{1}{n}Y_N Y_N^*)$ consistent in large N, n regime,

$$\left\| \mathcal{T}\left(\frac{1}{n}Y_N Y_N^*\right) - C_N \right\| \xrightarrow{\text{a.s.}} 0$$

(with $[\mathcal{T}(X)]_{ij} = \frac{1}{N}\sum_k [X]_{k,k+(i-j)}$). Many results beyond RMT regime **[Bickel,Levina'08]**,**[Wu,Pouramadi'09]**,**[Vinogradova,Couillet,Hachem'14]**.

- **Kernel matrices**: used in BSS, spectral clustering, etc.

$$K_{i,j} = k(x_i, x_j) = \begin{cases} x_i^* x_j & \text{, sample covariance matrix} \\ |x_i^* x_j|^2 & \text{, kurtosis-based BSS} \\ \exp(-\|x_i - x_j\|^2) & \text{, Gaussian kernel.} \end{cases}$$

Few results **[ElKaroui'10]**.

Outline

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

- **[Huber'64]** If $x_1 \sim (1-\varepsilon)\mathcal{N}(0, C_N) + \varepsilon G$, G unknown, robust estimator ($n > N$)

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^{n} \max\left\{\ell_1, \frac{\ell_2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i}\right\} x_i x_i^* \text{ for some } \ell_1, \ell_2 > 0.$$

Context

Baseline scenario: $x_1,\ldots,x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1]=0$, $E[x_1x_1^*]=C_N$:

- If $x_1 \sim \mathcal{N}(0,C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

- **[Huber'64]** If $x_1 \sim (1-\varepsilon)\mathcal{N}(0,C_N)+\varepsilon G$, G unknown, robust estimator ($n>N$)

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^n \max\left\{\ell_1, \frac{\ell_2}{\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i}\right\} x_i x_i^* \text{ for some } \ell_1,\ell_2>0.$$

- **[Maronna'76]** If x_1 elliptical (and $n>N$), ML estimator for C_N given by

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^n u\left(\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i\right) x_i x_i^* \text{ for some non-increasing } u.$$

Context

Baseline scenario: $x_1, \ldots, x_n \in \mathbb{C}^N$ (or $\mathbb{R}^N$) i.i.d. with $E[x_1] = 0$, $E[x_1 x_1^*] = C_N$:

- If $x_1 \sim \mathcal{N}(0, C_N)$, ML estimator for C_N given by sample covariance matrix (SCM)

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

- **[Huber'64]** If $x_1 \sim (1-\varepsilon)\mathcal{N}(0, C_N) + \varepsilon G$, G unknown, robust estimator ($n > N$)

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^{n} \max \left\{ \ell_1, \frac{\ell_2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i} \right\} x_i x_i^* \text{ for some } \ell_1, \ell_2 > 0.$$

- **[Maronna'76]** If x_1 elliptical (and $n > N$), ML estimator for C_N given by

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i \right) x_i x_i^* \text{ for some non-increasing } u.$$

- **[Pascal'13; Chen'11]** If $N > n$, x_1 elliptical or with outliers, shrinkage extensions

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

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

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

We study such $\hat{C}_N$ in the regime

$$N, n \to \infty, \; N/n \to c \in (0, \infty).$$

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

We study such $\hat{C}_N$ in the regime

$$N, n \to \infty, \;\; N/n \to c \in (0, \infty).$$

- Math interest:
  - limiting eigenvalue distribution of $\hat{C}_N$
  - limiting values and fluctuations of functionals $f(\hat{C}_N)$

Context

Results only known for N fixed and $n \to \infty$:

- not appropriate in settings of interest today (BigData, array processing, MIMO)

We study such $\hat{C}_N$ in the regime

$$N, n \to \infty, \; N/n \to c \in (0, \infty).$$

- Math interest:
  - limiting eigenvalue distribution of $\hat{C}_N$
  - limiting values and fluctuations of functionals $f(\hat{C}_N)$
- Application interest:
  - comparison between SCM and robust estimators
  - performance of robust/non-robust estimation methods
  - improvement thereof (by proper parametrization)

Outline of Theoretical Content

- First order convergence:

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

for some tractable random matrices $\hat{S}_N$.

- First order convergence:

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

for some tractable random matrices $\hat{S}_N$.

- Second order results:

$$N^{1-\varepsilon}\left(a^*\hat{C}_N^k b - a^*\hat{S}_N^k b \right) \xrightarrow{\text{a.s.}} 0$$

allowing transfer of CLT results.

Outline of Theoretical Content

- First order convergence:

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

  for some tractable random matrices $\hat{S}_N$.

- Second order results:

$$N^{1-\varepsilon} \left(a^* \hat{C}_N^k b - a^* \hat{S}_N^k b \right) \xrightarrow{\text{a.s.}} 0$$

  allowing transfer of CLT results.

- Applications:
  - improved robust covariance matrix estimation
  - improved robust tests / estimators
  - specific examples in statistics at large, array processing, statistical finance, etc.

Outline

Definition (Maronna's Estimator)

For $x_1, \ldots, x_n \in \mathbb{C}^N$ with $n > N$, $\hat{C}_N$ is the solution (upon existence and uniqueness) of

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

Definition (Maronna's Estimator)

For $x_1, \ldots, x_n \in \mathbb{C}^N$ with $n > N$, $\hat{C}_N$ is the solution (upon existence and uniqueness) of

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

where $u : [0, \infty) \to (0, \infty)$ is

- non-increasing
- such that $\phi(x) \triangleq xu(x)$ increasing of supremum ϕ_∞ with

$$1 < \phi_\infty < c^{-1}, \; c \in (0, 1).$$

Model Description

Definition (Maronna's Estimator)

For $x_1, \ldots, x_n \in \mathbb{C}^N$ with $n > N$, $\hat{C}_N$ is the solution (upon existence and uniqueness) of

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

where $u : [0, \infty) \to (0, \infty)$ is

- non-increasing
- such that $\phi(x) \triangleq x u(x)$ increasing of supremum ϕ_∞ with

$$1 < \phi_\infty < c^{-1}, \; c \in (0, 1).$$

Remark (Correlation Invariance)

For some $C_N \succ 0$, calling $\tilde{x}_i \triangleq C_N^{-\frac{1}{2}} x_i$, $\tilde{C}_N \triangleq C_N^{-\frac{1}{2}} \hat{C}_N C_N^{-\frac{1}{2}}$,

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

If $E[x_i x_i^*] = C_N$, sufficient to assume $E[x_i x_i^*] = I_N$.

Assumption ("Elliptical" Data)

$x_1, \ldots, x_n$ independent,

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

- $w_i \in \mathbb{C}^N$ isotropic, $\|w_i\|^2 = N$
- $C_N \succ 0$, $\limsup_N \|C_N\| < \infty$
- $\tau_i > 0$ deterministic (or random independent of w_i)

Assumption ("Elliptical" Data)

$x_1, \ldots, x_n$ independent,

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

- $w_i \in \mathbb{C}^N$ isotropic, $\|w_i\|^2 = N$
- $C_N \succ 0$, $\limsup_N \|C_N\| < \infty$
- $\tau_i > 0$ deterministic (or random independent of w_i)
- for $\tilde{\nu}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\tau_i}$ and some $m > 0$,

$$\tilde{\nu}_n([0, m)) < 1 - \phi_\infty^{-1} \text{ for all large } n \text{ (a.s.)}$$

- $\int \tau \tilde{\nu}_n(d\tau) \to 1$ (a.s.).

Assumption ("Elliptical" Data)

$x_1, \ldots, x_n$ independent,

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

- $w_i \in \mathbb{C}^N$ isotropic, $\|w_i\|^2 = N$
- $C_N \succ 0$, $\limsup_N \|C_N\| < \infty$
- $\tau_i > 0$ deterministic (or random independent of w_i)
- for $\tilde{\nu}_n = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{\tau_i}$ and some $m > 0$,

$$\tilde{\nu}_n([0, m)) < 1 - \phi_\infty^{-1} \text{ for all large } n \text{ (a.s.)}$$

- $\int \tau \tilde{\nu}_n(d\tau) \to 1$ (a.s.).

Fact (Existence and Uniqueness)

By **[Kent&Tyler'91]**, for each $n > N$, $\hat{C}_N$ is a.s. well-defined.

Assumption (Tail Control)

For each $a > b > 0$,

$$\frac{\limsup_n \tilde{\nu}_n([t, \infty))}{\phi(at) - \phi(bt)} \to 0$$

as $t \to \infty$.

Assumption (Tail Control)

For each $a > b > 0$,

$$\frac{\limsup_n \tilde{\nu}_n([t,\infty))}{\phi(at) - \phi(bt)} \to 0$$

as $t \to \infty$.

Example: If $u(x) = \frac{\alpha+1}{\alpha+x}$, τ_i i.i.d., sufficient to have $E[\tau_1^{1+\varepsilon}] < \infty$.

Assumption (Random Matrix Regime)

As $n \to \infty$,

$$c_N \triangleq \frac{N}{n} \to c \in (0,1).$$

Definition (v and ψ)

Letting $g(x) = x(1 - c\phi(x))^{-1}$ (on $\mathbb{R}_+$),

$$v(x) \triangleq (u \circ g^{-1})(x) \quad \text{non-increasing}$$
$$\psi(x) \triangleq xv(x) \quad \text{increasing and bounded by } \psi_\infty.$$

Definition (v and ψ)

Letting $g(x) = x(1 - c\phi(x))^{-1}$ (on $\mathbb{R}_+$),

$$v(x) \triangleq (u \circ g^{-1})(x) \quad \text{non-increasing}$$
$$\psi(x) \triangleq xv(x) \quad \text{increasing and bounded by } \psi_\infty.$$

Lemma (Rewriting $\hat{C}_N$)

It holds (with $C_N = I_N$) that

$$\hat{C}_N \triangleq \frac{1}{n}\sum_{i=1}^{n} \tau_i v\left(\tau_i d_i\right) w_i w_i^*$$

with $(d_1, \ldots, d_n) \in \mathbb{R}_+^n$ a.s. unique solution to

$$d_i = \frac{1}{N} w_i^* \hat{C}_{(i)}^{-1} w_i = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j \neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i, \; i = 1, \ldots, n.$$

Remark (Quadratic Form close to Trace)

Random matrix insight: $(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*)^{-1}$ "almost independent" of w_i, so

Remark (Quadratic Form close to Trace)

Random matrix insight: $(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*)^{-1}$ "almost independent" of w_i, so

$$d_i = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j\neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i \simeq \frac{1}{N} \mathrm{tr} \left(\frac{1}{n} \sum_{j\neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} \simeq \gamma_N$$

for some deterministic sequence $(\gamma_N)_{N=1}^{\infty}$, irrespective of i.

Remark (Quadratic Form close to Trace)

Random matrix insight: $(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*)^{-1}$ "almost independent" of w_i, so

$$d_i = \frac{1}{N} w_i^* \left(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*\right)^{-1} w_i \simeq \frac{1}{N}\mathrm{tr}\left(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j d_j) w_j w_j^*\right)^{-1} \simeq \gamma_N$$

for some deterministic sequence $(\gamma_N)_{N=1}^{\infty}$, irrespective of i.

Lemma (Key Lemma)

Letting $e_i \triangleq \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$ with γ_N unique solution to

$$1 = \int \frac{\psi(\tau\gamma_N)}{1 + c\psi(\tau\gamma_N)} \tilde{\nu}_n(d\tau),$$

we have

$$\max_{1\leq i\leq n} |e_i - 1| \xrightarrow{\text{a.s.}} 0.$$

(Proof in a few slides.)

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_N) x_i x_i^*.$$

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_N) x_i x_i^*.$$

$$\left[\textit{equivalently,} \quad \hat{S}_N = \gamma_N^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi(\tau_i \gamma_N) C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} \right]$$

Corollaries

- **Spectral measure**: $\mu_N^{\hat{C}_N} - \mu_N^{\hat{S}_N} \xrightarrow{\mathcal{L}} 0$ a.s. ($\mu_N^X \triangleq \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{\lambda_i(X)}$)

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_N) x_i x_i^*.$$

$$\left[\textit{equivalently,} \quad \hat{S}_N = \gamma_N^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi(\tau_i \gamma_N) C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} \right]$$

Corollaries

- **Spectral measure**: $\mu_N^{\hat{C}_N} - \mu_N^{\hat{S}_N} \xrightarrow{\mathcal{L}} 0$ a.s. ($\mu_N^X \triangleq \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\delta}_{\lambda_i(X)}$)
- **Local convergence**: $\max_{1 \leq i \leq N} |\lambda_i(\hat{C}_N) - \lambda_i(\hat{S}_N)| \xrightarrow{\text{a.s.}} 0.$

Theorem (Large dimensional behavior [C,Pascal,Silverstein'13])

With the notations and assumptions above,

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

with

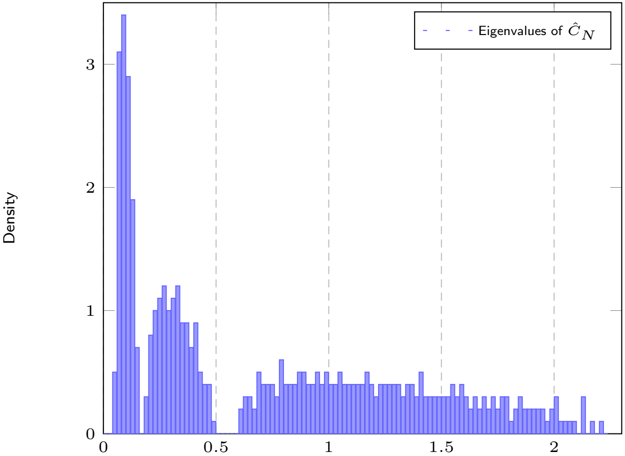
$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma_N) x_i x_i^*.$$

$$\left[\textit{equivalently,} \quad \hat{S}_N = \gamma_N^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi(\tau_i \gamma_N) C_N^{\frac{1}{2}} w_i w_i^* C_N^{\frac{1}{2}} \right]$$

Corollaries

- **Spectral measure**: $\mu_N^{\hat{C}_N} - \mu_N^{\hat{S}_N} \xrightarrow{\mathcal{L}} 0$ a.s. ($\mu_N^X \triangleq \frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{\lambda_i(X)}$)
- **Local convergence**: $\max_{1 \leq i \leq N} |\lambda_i(\hat{C}_N) - \lambda_i(\hat{S}_N)| \xrightarrow{\text{a.s.}} 0.$
- **Norm boundedness**: $\limsup_N \|\hat{C}_N\| < \infty$

$\longrightarrow$ Bounded spectrum (unlike SCM!)



(i)

Figure: $n = 2500$, $N = 500$, $C_N = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

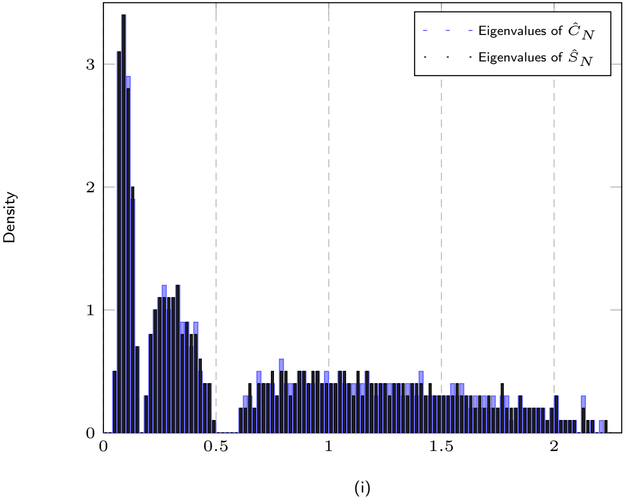


Figure: $n = 2500$, $N = 500$, $C_N = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

Large dimensional behavior

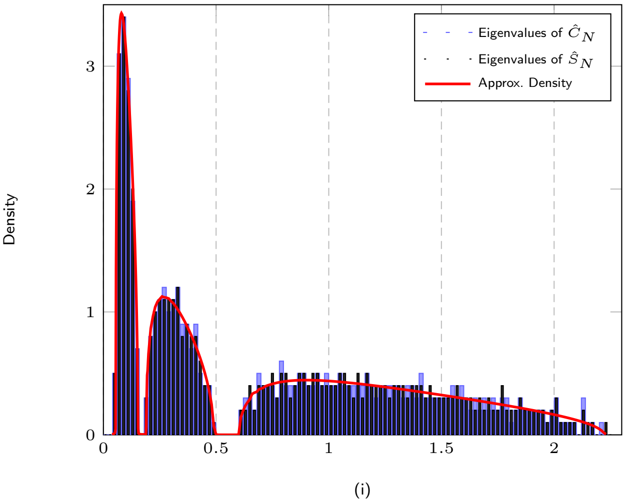


Figure: $n = 2500$, $N = 500$, $C_N = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_i \sim \Gamma(.5, 2)$ i.i.d.

Proof of the Key Lemma: $\max_i |e_i - 1| \xrightarrow{\text{a.s.}} 0$, $e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$

Property (Quadratic form and γ_N)

$$\max_{1 \le i \le n} \left| \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j \neq i} \tau_j v(\tau_j \gamma_N) w_j w_j^* \right)^{-1} w_i - \gamma_N \right| \xrightarrow{\text{a.s.}} 0.$$

Property (Quadratic form and γ_N)

$$\max_{1\leq i\leq n}\left|\frac{1}{N}w_i^*\left(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j\gamma_N)w_jw_j^*\right)^{-1}w_i - \gamma_N\right| \xrightarrow{\text{a.s.}} 0.$$

Proof of the Property

- Uniformity easy (moments of all orders for $[w_i]_j$).
- By a "quadratic form similar to trace" approach, we get

$$\max_{1\leq i\leq n}\left|\frac{1}{N}w_i^*\left(\frac{1}{n}\sum_{j\neq i}\tau_j v(\tau_j\gamma_N)w_jw_j^*\right)^{-1}w_i - m(0)\right| \xrightarrow{\text{a.s.}} 0$$

  with $m(0)$ unique positive solution to **[MarPas'67;SilBai'95]**

$$m(0) = \int \frac{\tau v(\tau\gamma_N)}{1 + c\tau v(\tau\gamma_N)m(0)}\tilde{\nu}_n(d\tau).$$

- γ_N precisely solves this equation, thus $m(0) = \gamma_N$.

Proof of the Key Lemma: $\max_i |e_i - 1| \xrightarrow{\text{a.s.}} 0$, $e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$

Substitution Trick (case $\tau_i \in [a, b] \subset (0, \infty)$)

Up to relabelling $e_1 \leq \ldots \leq e_n$, use

$$
\begin{aligned}
v(\tau_n \gamma_N) e_n = v(\tau_n d_n) &= v\left(\tau_n \frac{1}{N} w_n^* \left(\frac{1}{n} \sum_{i<n} \tau_i \underbrace{v(\tau_i d_i)}_{= v(\tau_i \gamma_N) e_i} w_i w_i^* \right)^{-1} w_n \right) \\
&\leq v\left(\tau_n e_n^{-1} \frac{1}{N} w_n^* \left(\frac{1}{n} \sum_{i<n} \tau_i v(\tau_i \gamma_N) w_i w_i^* \right)^{-1} w_n \right) \\
&\leq v\left(\tau_n e_n^{-1} (\gamma_N - \varepsilon_n) \right) \text{ a.s., } \varepsilon_n \to 0 \text{ (slow).}
\end{aligned}
$$

Substitution Trick (case $\tau_i \in [a,b] \subset (0,\infty)$)

Up to relabelling $e_1 \leq \ldots \leq e_n$, use

$$v(\tau_n\gamma_N)e_n = v(\tau_n d_n) = v\left(\tau_n \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i \underbrace{v(\tau_i d_i)}_{=v(\tau_i\gamma_N)e_i} w_i w_i^*\right)^{-1} w_n\right)$$

$$\leq v\left(\tau_n e_n^{-1} \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i v(\tau_i\gamma_N) w_i w_i^*\right)^{-1} w_n\right)$$

$$\leq v\left(\tau_n e_n^{-1}(\gamma_N - \varepsilon_n)\right) \text{ a.s., } \varepsilon_n \to 0 \text{ (slow).}$$

Use properties of ψ to get

$$\psi\left(\tau_n\gamma_N\right) \leq \psi\left(\tau_n e_n^{-1}\gamma_N\right)\left(1 - \varepsilon_n\gamma_N^{-1}\right)^{-1}$$

Substitution Trick (case $\tau_i \in [a,b] \subset (0,\infty)$)

Up to relabelling $e_1 \leq \ldots \leq e_n$, use

$$v(\tau_n \gamma_N) e_n = v(\tau_n d_n) = v\left(\tau_n \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i \underbrace{v(\tau_i d_i)}_{=v(\tau_i\gamma_N)e_i} w_i w_i^*\right)^{-1} w_n\right)$$
$$\leq v\left(\tau_n e_n^{-1} \frac{1}{N} w_n^* \left(\frac{1}{n}\sum_{i<n} \tau_i v(\tau_i \gamma_N) w_i w_i^*\right)^{-1} w_n\right)$$
$$\leq v\left(\tau_n e_n^{-1}(\gamma_N - \varepsilon_n)\right) \text{ a.s., } \varepsilon_n \to 0 \text{ (slow).}$$

Use properties of ψ to get

$$\psi\left(\tau_n \gamma_N\right) \leq \psi\left(\tau_n e_n^{-1} \gamma_N\right)\left(1 - \varepsilon_n \gamma_N^{-1}\right)^{-1}$$

Conclusion: If $e_n > 1 + \ell$ i.o., as $\tau_n \in [a,b]$, on subsequence $\begin{cases} \tau_n \to \tau_0 > 0 \\ \gamma_N \to \gamma_0 > 0 \end{cases}$,

$$\psi(\tau_0 \gamma_0) \leq \psi\left(\frac{\tau_0 \gamma_0}{1+\ell}\right), \text{ a contradiction.}$$

General τ_i case

- Control of

$$\Delta_M = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j \neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i - \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{\substack{j \neq i \\ \tau_j \leq M}} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i.$$

- Rationale: Large M bring small Δ_M but (possibly) large τ_n
  $\longrightarrow$ Relative control between tail of $\tilde{\nu}_n$ and flattening of ψ.

Proof of the Key Lemma: $\max_i |e_i - 1| \xrightarrow{\text{a.s.}} 0$, $e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma_N)}$

General τ_i case

- Control of

$$\Delta_M = \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{j \neq i} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i - \frac{1}{N} w_i^* \left(\frac{1}{n} \sum_{\substack{j \neq i \\ \tau_j \leq M}} \tau_j v(\tau_j d_j) w_j w_j^* \right)^{-1} w_i.$$

- Rationale: Large M bring small Δ_M but (possibly) large τ_n
$\longrightarrow$ Relative control between tail of $\tilde{\nu}_n$ and flattening of ψ.

This concludes the proof.

Assumption (Signal Model)

$x_1, \ldots, x_n$ independent,

$$x_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i$$

- $w_i \in \mathbb{C}^N$, τ_i as previously, (for simplicity) $\tilde{\nu}_n \to \tilde{\nu}$
- $s_{li} \in \mathbb{C}$ i.i.d., mean 0, variance 1
- $p_1 \geq \ldots \geq p_L \geq 0$
- $a_1, \ldots, a_L \in \mathbb{C}^N$ deterministic with $\sum_{l=1}^{L} p_l a_l a_l^* \to \mathrm{diag}(p_i)_{i=1}^{L}$.

Spiked Model Extension

Assumption (Signal Model)

$x_1, \ldots, x_n$ independent,

$$x_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i$$

- $w_i \in \mathbb{C}^N$, τ_i as previously, (for simplicity) $\tilde{\nu}_n \to \tilde{\nu}$
- $s_{li} \in \mathbb{C}$ i.i.d., mean 0, variance 1
- $p_1 \geq \ldots \geq p_L \geq 0$
- $a_1, \ldots, a_L \in \mathbb{C}^N$ deterministic with $\sum_{l=1}^{L} p_l a_l a_l^* \to \text{diag}(p_i)_{i=1}^{L}$.

Theorem (Extension of pure-noise model [C'2014])

As $n \to \infty$, under previous assumptions,

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

where

$$\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma) x_i x_i^*.$$

(same result but different model, $\gamma = \lim_N \gamma_N$)

$\longrightarrow \hat{S}_N$ follows a spiked random matrix model.

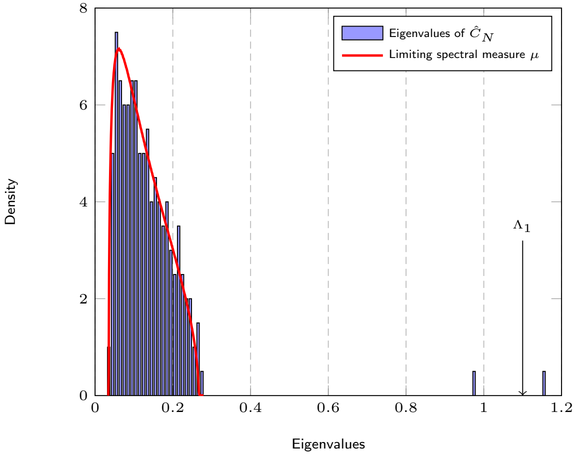


Figure: Eigenvalues of $\hat{C}_N$, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Spiked Model Extension

$\longrightarrow$ But eigenvalues allowed to wander away from limiting support.

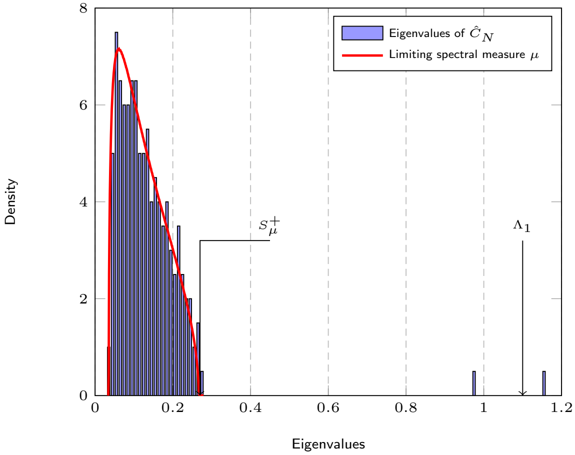


Figure: Eigenvalues of $\hat{C}_N$, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Spiked Model Extension

$\longrightarrow$ Noise eigenvalues are bounded by some S^+.

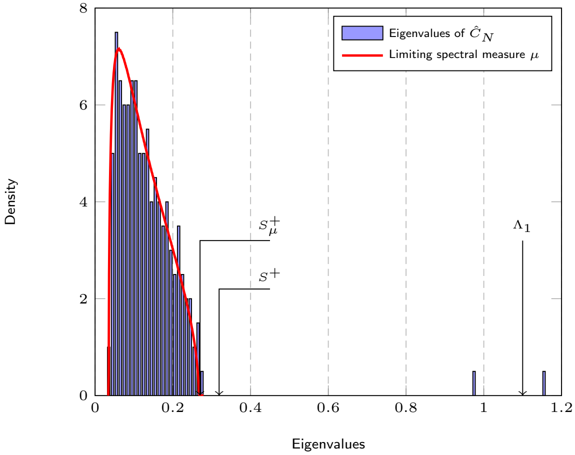


Figure: Eigenvalues of $\hat{C}_N$, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Spiked Model Extension

$\longrightarrow$ To be compared versus SCM $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$

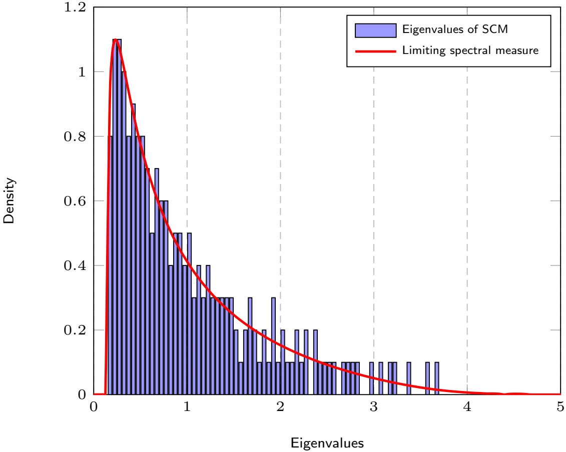


Figure: Eigenvalues of $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Sudent-t impulsions.

Spiked Model Extension

Some important remarks:

- If $p_1 = \ldots = p_L = 0$, noise-only model and

$$\limsup_N \|\hat{C}_N\| = \limsup_N \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\psi(\tau_i \gamma)}{\gamma} w_i w_i^* \right\| \leq S^+ \triangleq \frac{\phi_\infty (1+\sqrt{c})^2}{(1-c\phi_\infty)\gamma}.$$

Spiked Model Extension

Some important remarks:

- If $p_1 = \ldots = p_L = 0$, noise-only model and

$$\limsup_N \|\hat{C}_N\| = \limsup_N \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\psi(\tau_i \gamma)}{\gamma} w_i w_i^* \right\| \leq S^+ \triangleq \frac{\phi_\infty (1+\sqrt{c})^2}{(1-c\phi_\infty)\gamma}.$$

- If $p_1 \geq \ldots \geq p_L > 0$, informative spikes if $\det(\hat{S}_N - xI_N)$ has solutions beyond S^+ (and not S_μ^+!), i.e., if

$$p_l > p_- \triangleq \lim_{x \downarrow S^+} -c \left(\int \frac{\delta(x) v(t\gamma)}{1+\delta(x) t v(t\gamma)} \tilde{\nu}(dt) \right)^{-1}$$

with $\delta(x)$, $x > S_\mu^+$, unique solution to

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

Theorem (Spiked estimation, known $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$ *(*$\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N$*),*

Extreme eigenvalues. *For each* j *with* $p_j > p_-$*,*

$$\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+ \text{ a.s., where } -c\left(\int \frac{\delta(\Lambda_j)v(\tau\gamma)}{1+\delta(\Lambda_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} = p_j.$$

Theorem (Spiked estimation, known $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$ *(*$\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N$*),*

Extreme eigenvalues. *For each* j *with* $p_j > p_-$*,*

$$\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+ \text{ a.s., where } -c\left(\int \frac{\delta(\Lambda_j)v(\tau\gamma)}{1+\delta(\Lambda_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} = p_j.$$

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

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

Theorem (Spiked estimation, known $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$ ($\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N$),

Extreme eigenvalues. *For each* j *with* $p_j > p_-$,

$$\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+ \text{ a.s., where } -c\left(\int \frac{\delta(\Lambda_j)v(\tau\gamma)}{1+\delta(\Lambda_j)\tau v(\tau\gamma)}\tilde{\nu}(d\tau)\right)^{-1} = p_j.$$

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

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

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

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

where

$$w_k \triangleq \int \frac{v(t\gamma)\tilde{\nu}(dt)}{\left(1+\delta(\hat{\lambda}_k)tv(t\gamma)\right)^2}\left[\int \frac{v(t\gamma)\tilde{\nu}(dt)}{1+\delta(\hat{\lambda}_k)tv(t\gamma)}\left(1-\frac{1}{c}\int \frac{\delta(\hat{\lambda}_k)^2 t^2 v(t\gamma)^2 \tilde{\nu}(dt)}{\left(1+\delta(\hat{\lambda}_k)tv(t\gamma)\right)^2}\right)\right]^{-1}$$

Theorem (Spiked estimation, unknown $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$,

Empirical estimates.

$$\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0, \ \hat{\gamma}_n \triangleq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i$$

$$\max_{\tau_j < M} |\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0, \ \hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}_n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i.$$

Theorem (Spiked estimation, unknown $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$,

Empirical estimates.

$$\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0,\ \hat{\gamma}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i$$

$$\max_{\tau_j < M} |\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0,\ \hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}_n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i.$$

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

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

Theorem (Spiked estimation, unknown $\tilde{\nu}$ [C'2014])

With the SVD $AA^* = \sum_{l=1}^{L} q_l u_l u_l^*$ *and* $\hat{C}_N = \sum_{i=1}^{N} \hat{\lambda}_i \hat{u}_i \hat{u}_i^*$,

Empirical estimates.

$$\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0,\ \hat{\gamma}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i$$

$$\max_{\tau_j < M} |\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0,\ \hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}_n} \frac{1}{N} x_i^* \hat{C}_{(i)}^{-1} x_i.$$

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

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

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

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

for the corresponding $\hat{w}_k = f(\{\hat{\tau}_i\}, \hat{\delta}(\hat{\lambda}_k))$.

$\longrightarrow$ Application to angle estimation with

$$a_l = a(\theta_l), \ \theta_l \in [0, 2\pi)$$

$\longrightarrow$ Application to angle estimation with

$$a_l = a(\theta_l),\ \theta_l \in [0, 2\pi)$$

Corollary (Robust G-MUSIC)

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

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

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

Then, for each j *with* $p_j > p_-$,

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

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

where

$$\hat{\theta}_j \triangleq \mathrm{argmin}_{\theta \in V(\theta_j)} \left\{ \hat{\eta}_{\mathrm{RG}}(\theta) \right\}$$

$$\hat{\theta}_j^{\mathrm{emp}} \triangleq \mathrm{argmin}_{\theta \in V(\theta_j)} \left\{ \hat{\eta}_{\mathrm{RG}}^{\mathrm{emp}}(\theta) \right\}.$$

Spiked Model Extension

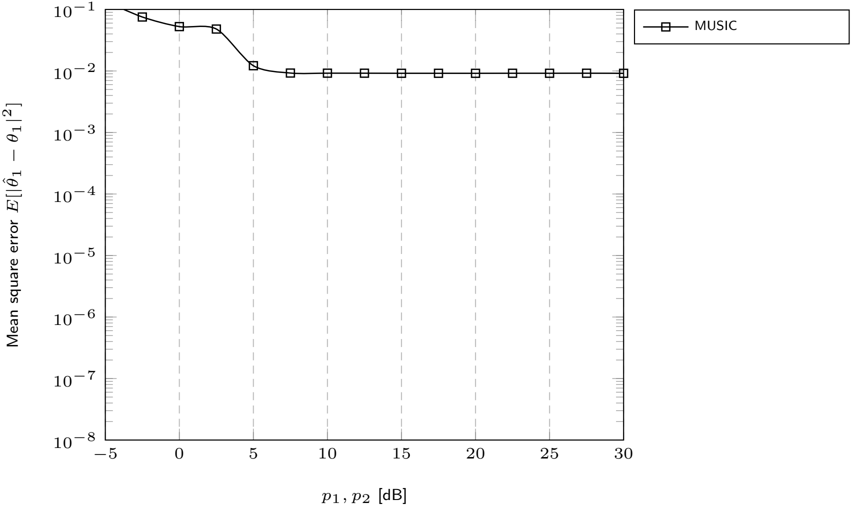


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

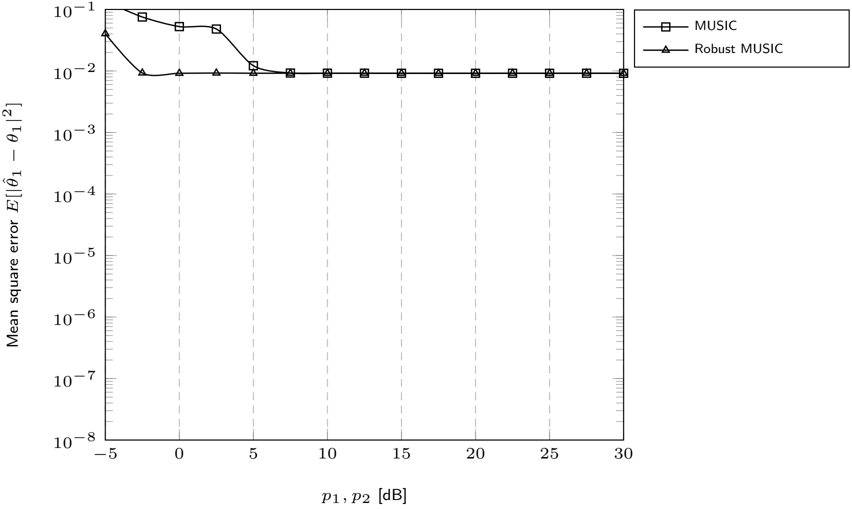


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

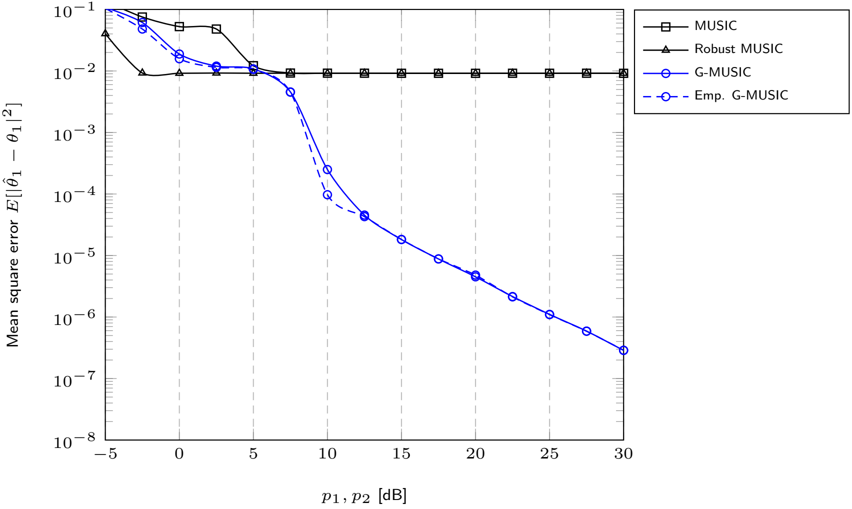


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

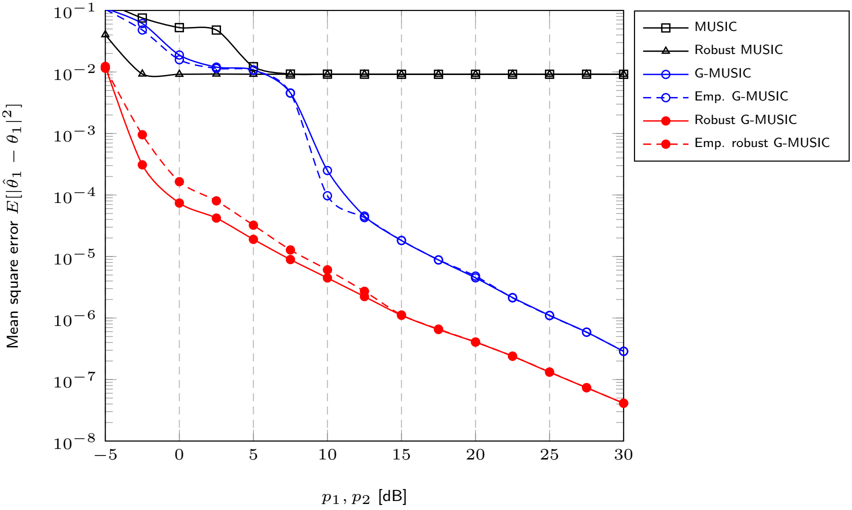


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, Student-t impulsions, $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $p_1 = p_2$.

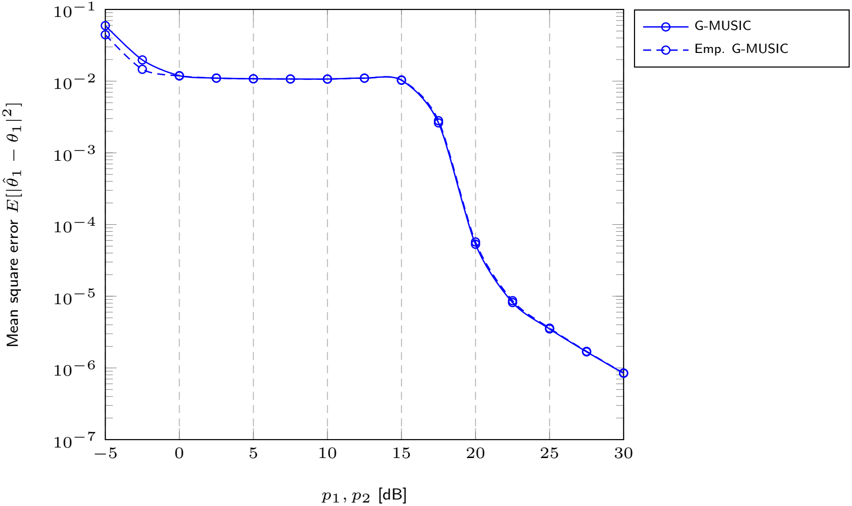


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, sample outlier scenario $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

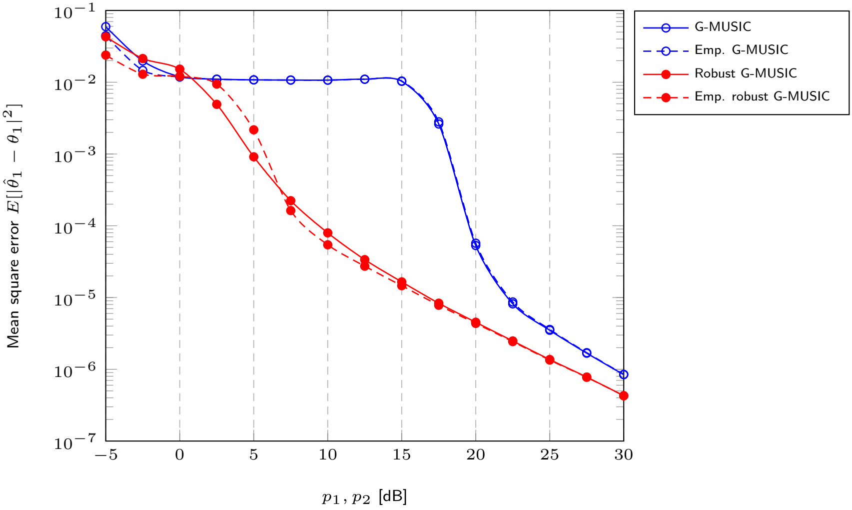


Figure: MSE for estimate of $\theta_1 = 10^\circ$, $N = 20$, $n = 100$, $L = 2$ sources at 10° and 12°, sample outlier scenario $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Outlier Data

Original setting of Huber

Assumption (Outlying Data)

Observation set

$$X = \left[x_1, \ldots, x_{(1-\varepsilon_n)n}, a_1, \ldots, a_{\varepsilon_n n}\right]$$

where $x_i \sim \mathcal{CN}(0, C_N)$ and $a_1, \ldots, a_{\varepsilon_n n} \in \mathbb{C}^N$ deterministic with

$$\limsup_n \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} \frac{1}{N} a_i^* C_N^{-1} a_i < \infty$$

(or only a.s. if a_i random).

Theorem (Outlier Rejection [Morales-Jimenez,C,McKay'14])

As $n \to \infty$,

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

where

$$\hat{S}_N \triangleq v\left(\gamma_N\right) \frac{1}{n} \sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} v\left(\alpha_{i,n}\right) a_i a_i^*$$

with γ_N and $\alpha_{1,n}, \ldots, \alpha_{\varepsilon_n n,n}$ unique positive solutions to

$$\gamma_N = \frac{1}{N} \mathrm{tr} C_N \left(\frac{(1-\varepsilon)v(\gamma_N)}{1+cv(\gamma_N)\gamma_N} C_N + \frac{1}{n} \sum_{i=1}^{\varepsilon_n n} v\left(\alpha_{i,n}\right) a_i a_i^* \right)^{-1}$$

$$\alpha_{i,n} = \frac{1}{N} a_i^* \left(\frac{(1-\varepsilon)v(\gamma_N)}{1+cv(\gamma_N)\gamma_N} C_N + \frac{1}{n} \sum_{j \neq i}^{\varepsilon_n n} v\left(\alpha_{j,n}\right) a_j a_j^* \right)^{-1} a_i, \;\; i = 1, \ldots, \varepsilon_n n.$$

Outlier Data

- For $\varepsilon_n n = 1$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{n-1} x_i x_i^* + \left(v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}a_1^* C_N^{-1} a_1\right) + o(1)\right) a_1 a_1^*$$

Outlier rejection relies on $\frac{1}{N}a_1^* C_N^{-1} a_1 \lessgtr 1$.

Outlier Data

- For $\varepsilon_n n = 1$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{n-1} x_i x_i^* + \left(v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}a_1^* C_N^{-1} a_1\right) + o(1)\right) a_1 a_1^*$$

Outlier rejection relies on $\frac{1}{N}a_1^* C_N^{-1} a_1 \lessgtr 1$.

- For $a_i \sim \mathcal{CN}(0, D_N)$, $\varepsilon_n \to \varepsilon \geq 0$,

$$\hat{S}_N = v(\gamma_n)\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + v(\alpha_n)\frac{1}{n}\sum_{i=1}^{\varepsilon_n n} a_i a_i^*$$

$$\gamma_n = \frac{1}{N}\mathrm{tr}\, C_N \left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n} C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n} D_N\right)^{-1}$$

$$\alpha_n = \frac{1}{N}\mathrm{tr}\, D_N \left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n} C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n} D_N\right)^{-1}.$$

Outlier Data

- For $\varepsilon_n n = 1$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{n-1} x_i x_i^* + \left(v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}a_1^* C_N^{-1} a_1\right) + o(1)\right) a_1 a_1^*$$

Outlier rejection relies on $\frac{1}{N} a_1^* C_N^{-1} a_1 \lessgtr 1$.

- For $a_i \sim \mathcal{CN}(0, D_N)$, $\varepsilon_n \to \varepsilon \geq 0$,

$$\hat{S}_N = v\left(\gamma_n\right)\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + v\left(\alpha_n\right)\frac{1}{n}\sum_{i=1}^{\varepsilon_n n} a_i a_i^*$$

$$\gamma_n = \frac{1}{N}\operatorname{tr} C_N \left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n}D_N\right)^{-1}$$

$$\alpha_n = \frac{1}{N}\operatorname{tr} D_N \left(\frac{(1-\varepsilon)v(\gamma_n)}{1+cv(\gamma_n)\gamma_n}C_N + \frac{\varepsilon v(\alpha_n)}{1+cv(\alpha_n)\alpha_n}D_N\right)^{-1}.$$

For $\varepsilon_n \to 0$,

$$\hat{S}_N = v\left(\frac{\phi^{-1}(1)}{1-c}\right)\frac{1}{n}\sum_{i=1}^{(1-\varepsilon_n)n} x_i x_i^* + \frac{1}{n}\sum_{i=1}^{\varepsilon_n n} v\left(\frac{\phi^{-1}(1)}{1-c}\frac{1}{N}\operatorname{tr} D_N C_N^{-1}\right) a_i a_i^*$$

Outlier rejection relies on $\frac{1}{N}\operatorname{tr} D_N C_N^{-1} \lessgtr 1$.

Outlier Data

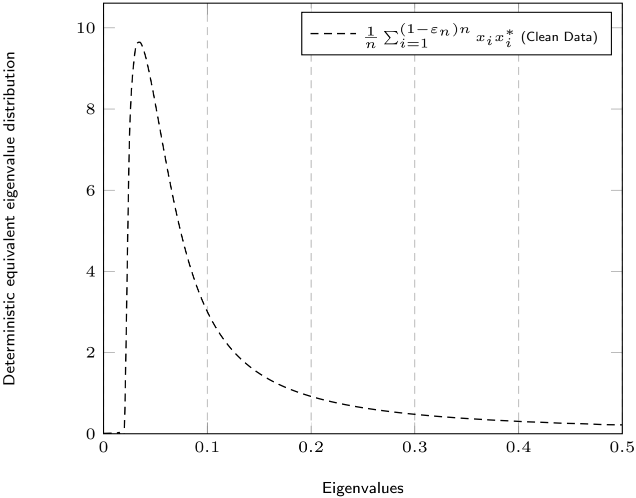


Figure: Limiting eigenvalue distributions. $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$.

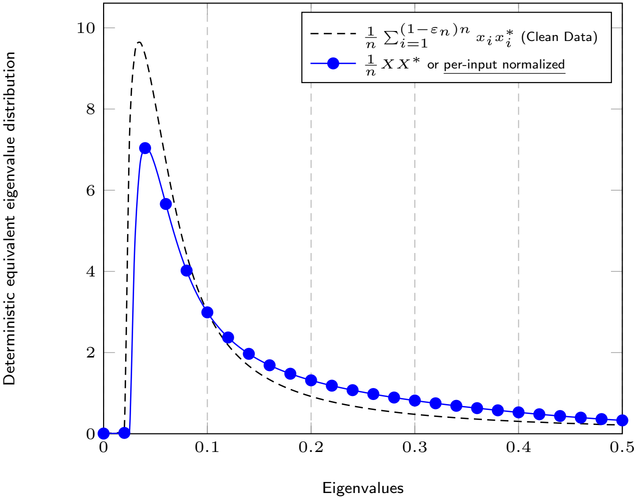


Figure: Limiting eigenvalue distributions. $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$.

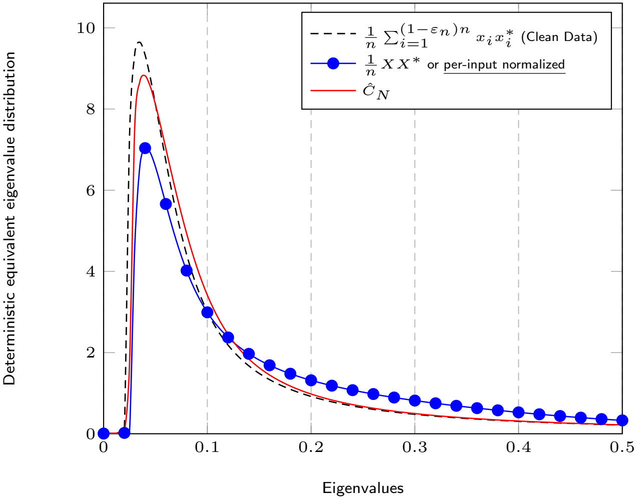


Figure: Limiting eigenvalue distributions. $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$.

Outline

Context

- Generalize robust estimators to $N > n$

Shrinkage Estimators

Context

- Generalize robust estimators to $N > n$
- Exploit Ledoit–Wolf type shrinkage ($+\rho I_N$ regularization)

Context

- Generalize robust estimators to $N > n$
- Exploit Ledoit–Wolf type shrinkage ($+\rho I_N$ regularization)
- Optimize degree of freedom ρ.

Context

- Generalize robust estimators to $N > n$
- Exploit Ledoit–Wolf type shrinkage ($+\rho I_N$ regularization)
- Optimize degree of freedom ρ.

Assumption (Pure-noise model)

Independent $x_1, \ldots, x_n \in \mathbb{C}^N$,

$$x_i = \sqrt{\tau}_i z_i$$

with

- $\tau_i > 0$ arbitrary
- $z_i \sim \mathcal{CN}(0, C_N)$, $\limsup_N \|C_N\| < \infty$
- $\nu_n \triangleq \frac{1}{N} \sum_{i=1}^N \boldsymbol{\delta}_{\lambda_i(C_N)} \to \nu$.

Two estimators in the literature

Definition (Abramovich–Pascal estimate)

For $\rho \in (\max\{0, 1-n/N\}, 1]$, unique solution $\hat{C}_N(\rho)$ to

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

Property: $\frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}(\rho) = 1$.

Two estimators in the literature

Definition (Abramovich–Pascal estimate)

For $\rho \in (\max\{0, 1 - n/N\}, 1]$, unique solution $\hat{C}_N(\rho)$ to

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

Property: $\frac{1}{N} \operatorname{tr} \hat{C}_N^{-1}(\rho) = 1$.

Definition (Chen estimate)

For $\rho \in (0, 1]$, unique solution $\check{C}_N(\rho)$ to

$$\check{C}_N(\rho) = \frac{\check{B}_N(\rho)}{\frac{1}{N} \operatorname{tr} \check{B}_N(\rho)}$$

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

Property: $\frac{1}{N} \operatorname{tr} \check{C}_N(\rho) = 1$.

Theorem (Abramovich–Pascal estimator [C,McKay'14])

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

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

with

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

and $\hat{\gamma}(\rho)$ *unique positive solution to*

$$1 = \int \frac{t}{\rho \hat{\gamma}(\rho) + (1 - \rho)t} \nu(dt).$$

Theorem (Chen estimator [C,McKay'14])

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

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

where

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

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 $\check{\gamma}(\rho)$ *unique positive solution to*

$$1 = \int \frac{t}{\rho\check{\gamma}(\rho) + \frac{1-\rho}{(1-\rho)c + F(\check{\gamma}(\rho);\rho)} t} \nu(dt).$$

Corollary (Model Equivalence)

For $\rho \in (0,1]$, there exists a unique $(\hat{\rho},\check{\rho})$ such that

$$\frac{\hat{S}_N(\hat{\rho})}{\lim_N \frac{1}{N}\operatorname{tr}\hat{S}_N(\hat{\rho})} = \check{S}_N(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} z_i z_i^* + \rho I_N.$$

Besides, $\rho \mapsto \hat{\rho}$ and $\rho \mapsto \check{\rho}$ are continuously increasing and onto.

Consequence: both estimators equivalent in limit to Ledoit–Wolf on z_i (not x_i).

Uniform convergence allows for optimization over ρ.

Proposition (Optimal Frobenius-norm Shrinkage)

For each ρ, define

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

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

$$D^\star = c\frac{M_{\nu,2} - 1}{c + M_{\nu,2} - 1} \quad (M_{\nu,2},\ \text{order-2 moment})$$

$$\rho^\star = \frac{c}{c + M_{\nu,2} - 1}$$

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,

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

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

Proposition (Optimal Frobenius-norm shrinkage estimate)

Let $\hat{\rho}_N, \check{\rho}_N$ be solutions to

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

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

Then

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

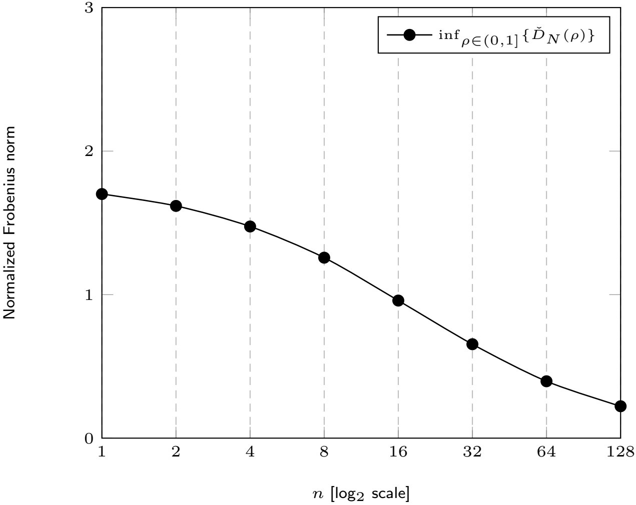


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* C_N{}^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

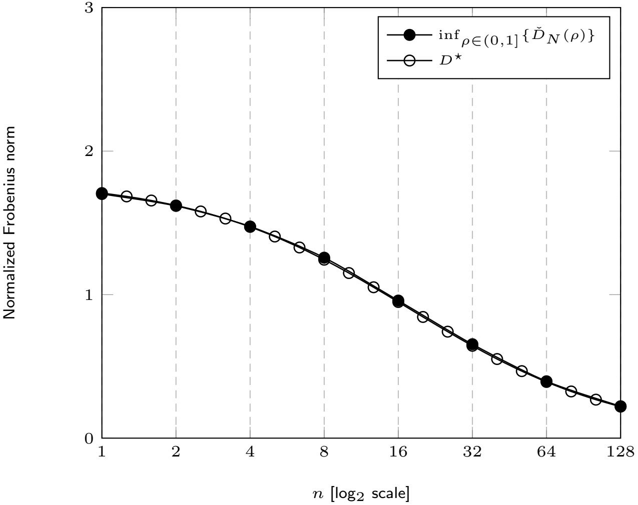


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* C_N{}^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

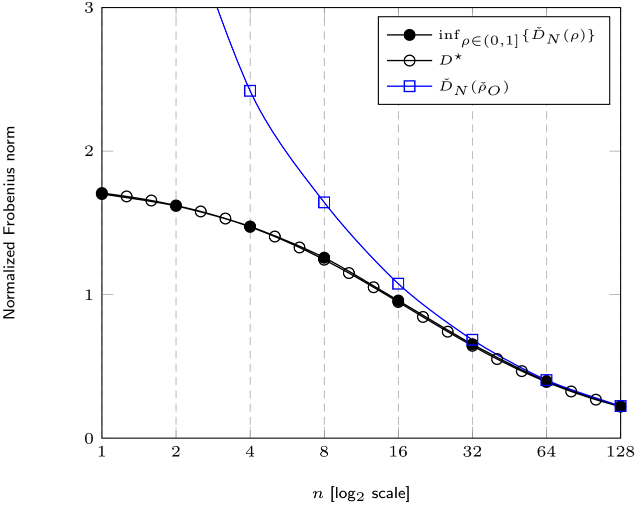


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* C_N{}^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

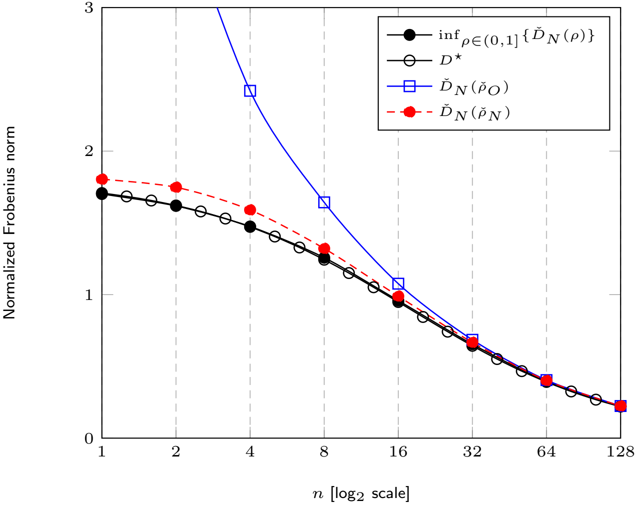


Figure: Optimal shrinkage, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$; $\check{\rho}_O$ clairvoyant estimator of (Chen et al., 2011) assuming $\hat{C}_N(\rho) \simeq (1-\rho)\frac{1}{n}\sum_i \frac{x_i x_i^*}{\frac{1}{N} x_i^* {C_N}^{-1} x_i} + \rho I_N$.

Optimal asymptotic shrinkage

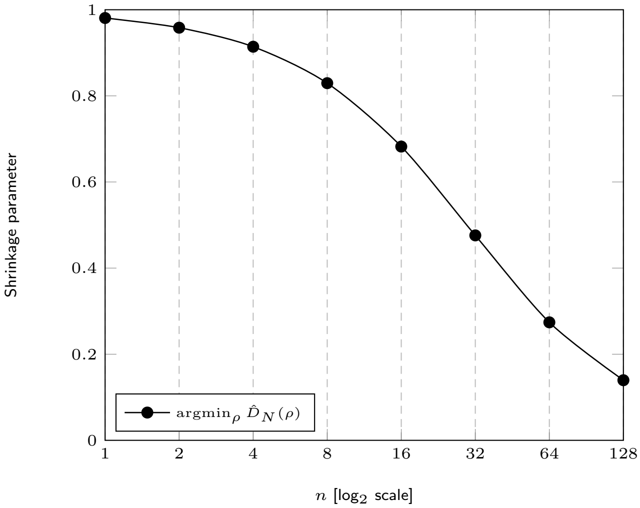


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

Optimal asymptotic shrinkage

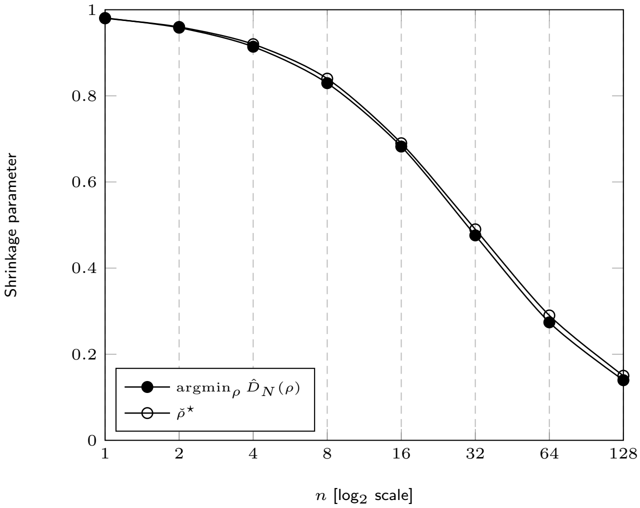


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

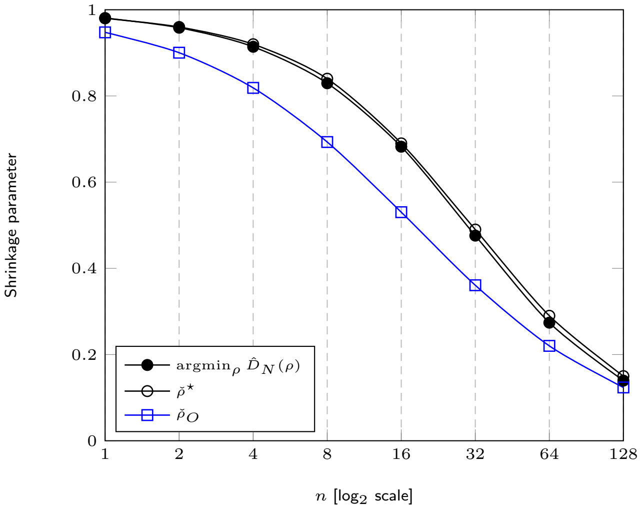


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

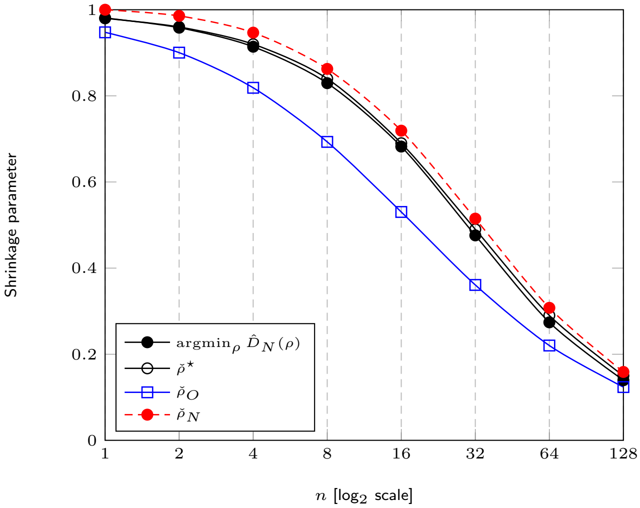


Figure: Shrinkage parameter ρ, $N = 32$, $[C_N]_{ij} = .7^{|i-j|}$.

Outline

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0,\sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0,\sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \ \ ?$$

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ $(\|a\| = \|b\| = 1)$

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0,\sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0,\sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \;\; ?$$

Conjectures

- From simulations, it seems that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| = O(N^{-\frac{1}{2}})$. Weak result.

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0, \sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \quad ?$$

Conjectures

- From simulations, it seems that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| = O(N^{-\frac{1}{2}})$. Weak result.
- Because of self-averaging, we hope: $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b = o(N^{-\frac{1}{2}})$

Fluctuations of $\hat{C}_N$

Context (about $\sup_\rho \|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| \xrightarrow{\text{a.s.}} 0$)

Implies: propagation to $\hat{S}_N(\rho)$ of first order results on $\hat{C}_N(\rho)$

- Linear statistics $f(\hat{C}_N(\rho)) - f(\hat{S}_N(\rho)) \xrightarrow{\text{a.s.}} 0$
- Anisotropic results $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b \xrightarrow{\text{a.s.}} 0$ ($\|a\| = \|b\| = 1$)

Does not imply: propagation to $\hat{S}_N(\rho)$ of second-order results on $\hat{C}_N(\rho)$

- If $N^\alpha f(\hat{S}_N(\rho)) \to \mathcal{N}(0, \sigma^2)$, what about $N^\alpha f(\hat{C}_N(\rho))$?
- If $N^\alpha a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, what about

$$N^\alpha a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \ ?$$

Conjectures

- From simulations, it seems that $\|\hat{C}_N(\rho) - \hat{S}_N(\rho)\| = O(N^{-\frac{1}{2}})$. Weak result.
- Because of self-averaging, we hope: $a^*\hat{C}_N(\rho)b - a^*\hat{S}_N(\rho)b = o(N^{-\frac{1}{2}})$
- Since $\sqrt{N}a^*(\hat{S}_N(\rho) - E[\hat{S}_N(\rho)])b \to \mathcal{N}(0, \sigma^2)$, this would imply

$$\sqrt{N}a^*(\hat{C}_N(\rho) - E[\hat{C}_N(\rho)])b \to \mathcal{N}(0, \sigma^2).$$

Theorem (Fluctuation of bilinear forms [C,Kammoun,Pascal'14])

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

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

Theorem (Fluctuation of bilinear forms [C,Kammoun,Pascal'14])
Let $a, b \in \mathbb{C}^N$ with $\|a\| = \|b\| = 1$. Then, as $n \to \infty$, $N/n \to c \in (0, \infty)$, for all $\varepsilon > 0$, $k \in \mathbb{Z}$,

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

(with $\varepsilon < \frac{1}{2}$, desired result)

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

- We prove easily (classical proof but with speed)

$$\max_{1 \leq i \leq n} N^{\frac{1}{2}-\varepsilon} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$$

Not good enough.

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

- We prove easily (classical proof but with speed)

$$\max_{1\leq i\leq n} N^{\frac{1}{2}-\varepsilon} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$$

Not good enough.

- **IDEA 1**: Exploit self-averaging

$$a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$
$$= \frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right]$$

Proof idea

- First write (with $d_i = \frac{1}{N} z_i^* \hat{C}_{(i)}^{-1} z_i$)

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

- We prove easily (classical proof but with speed)

$$\max_{1 \leq i \leq n} N^{\frac{1}{2}-\varepsilon} |d_i - \gamma_N| \xrightarrow{\text{a.s.}} 0$$

Not good enough.

- **IDEA 1**: Exploit self-averaging

$$a^* \hat{C}_N^{-1} \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right] z_i z_i^* \right) \hat{S}_N^{-1} b$$

$$= \frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{i=1}^n a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{d_i} \right]$$

- But too hard. Since d_i implicit.

Proof idea

- **IDEA 2**: Introduce intermediate quantity

$$\tilde{d}_i(\rho) = \frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i = \frac{1}{N} z_i^* \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{j \neq i}^{n} \frac{z_j z_j^*}{\gamma_N} + \rho I_N \right)^{-1} z_i$$

and write

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = \frac{1-\rho}{1-(1-\rho)c_N} \underbrace{\frac{1}{n} \sum_{i=1}^{n} a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right]}_{\text{Term (A)}}$$

$$+ \frac{1-\rho}{1-(1-\rho)c_N} a^* \hat{C}_N^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \underbrace{\left[\frac{1}{\tilde{d}_i} - \frac{1}{d_i} \right]}_{\text{Term (B)}} z_i z_i^* \right) \hat{S}_N^{-1} b.$$

Proof idea

- **IDEA 2**: Introduce intermediate quantity

$$\tilde{d}_i(\rho) = \frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i = \frac{1}{N} z_i^* \left(\frac{1-\rho}{1-(1-\rho)c_N} \frac{1}{n} \sum_{j \neq i}^{n} \frac{z_j z_j^*}{\gamma_N} + \rho I_N \right)^{-1} z_i$$

and write

$$a^* \hat{C}_N^{-1} b - a^* \hat{S}_N^{-1} b = \frac{1-\rho}{1-(1-\rho)c_N} \underbrace{\frac{1}{n} \sum_{i=1}^{n} a^* \hat{C}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left[\frac{1}{\gamma_N} - \frac{1}{\tilde{d}_i} \right]}_{\text{Term (A)}}$$

$$+ \frac{1-\rho}{1-(1-\rho)c_N} a^* \hat{C}_N^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \underbrace{\left[\frac{1}{\tilde{d}_i} - \frac{1}{d_i} \right]}_{\text{Term (B)}} z_i z_i^* \right) \hat{S}_N^{-1} b.$$

- Key lemma for both Terms (A)-(B):

Lemma (Key Lemma, Self-averaging)

$$E\left[\left| \frac{1}{n} \sum_{i=1}^{n} a^* \hat{S}_N^{-1} z_i z_i^* \hat{S}_N^{-1} b \left(\frac{1}{N} z_i^* \hat{S}_{(i)}^{-1} z_i - \gamma_N \right) \right|^{2p} \right] = O\left(N^{-2p}\right)$$

Context (Hypothesis Test)

We observe $x_1, \ldots, x_n$, $x_i = \sqrt{\tau}_i w_i$, $\|w_i\|^2 = N$ isotropic, and receive

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

with $\alpha > 0$ unknown, $p \in \mathbb{C}^N$ known.

Context (Hypothesis Test)

We observe $x_1, \ldots, x_n$, $x_i = \sqrt{\tau}_i w_i$, $\|w_i\|^2 = N$ isotropic, and receive

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

with $\alpha > 0$ unknown, $p \in \mathbb{C}^N$ known.

Definition (GLRT Detector)

$$T_N(\rho) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{\gamma}{\sqrt{N}}$$

with

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

Theorem (Asymptotic detector performance [C,Kammoun,Pascal'14])

Under $\mathcal{H}_0$, as $n \to \infty$ with $N/n \to c \in (0,\infty)$,

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

where

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

with $Q_N(\underline{\rho}) \triangleq (I_N + (1-\underline{\rho}) m(-\underline{\rho}) C_N)^{-1}$ *and* $\underline{\rho} = \rho(\rho + \frac{1}{\gamma_N(\rho)} \frac{1-\rho}{1-(1-\rho)c})^{-1}$*.*

Proposition (Empirical performance optimum)

Let

$$\hat{\sigma}_N^2(\rho) \triangleq \frac{1}{2} \frac{1 - \rho \frac{p^* \hat{C}_N^{-2}(\rho) p}{p^* \hat{C}_N^{-1}(\rho) p}}{(1 - c_N + c_N \rho)(1 - \rho)}.$$

Then,

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

Proposition (Empirical performance optimum)

Let

$$\hat{\sigma}_N^2(\rho) \triangleq \frac{1}{2} \frac{1 - \rho \frac{p^* \hat{C}_N^{-2}(\rho) p}{p^* \hat{C}_N^{-1}(\rho) p}}{(1 - c_N + c_N \rho)(1 - \rho)}.$$

Then,

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

Besides, let

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

Then, for every $\gamma > 0$,

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

Application to GLRT detection

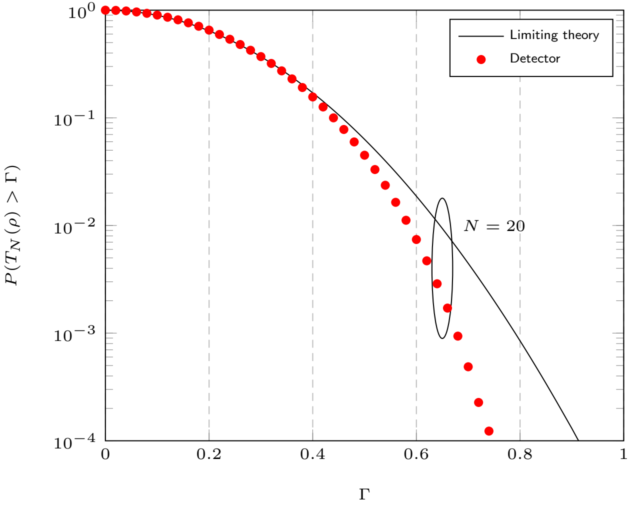


Figure: False alarm rate $P(T_N(\hat{\rho}_N^*) > \Gamma)$, $N = 20$ or $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $[C_N]_{ij} = .7^{|i-j|}$, $N/n = 1/2$.

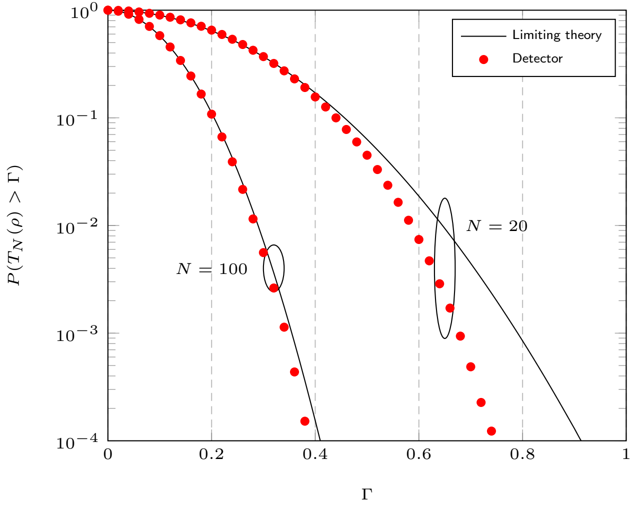


Figure: False alarm rate $P(T_N(\hat{\rho}_N^*) > \Gamma)$, $N = 20$ or $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $[C_N]_{ij} = .7^{|i-j|}$, $N/n = 1/2$.

Outline

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^* \beta\right), \; y_i \in \mathbb{R}, \; x_i \in \mathbb{R}^N.$$

Perspectives

Takeaway messages

- Robust estimators of scatter: $\left\{ \begin{array}{l} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{array} \right.$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^* \beta\right), \; y_i \in \mathbb{R}, \; x_i \in \mathbb{R}^N.$$

More general framework: BigData RMT

- Kernel random matrices (application to spectral clustering, kernel PCA)

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N,n \\ \text{become tractable when } N,n\to\infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \operatorname{argmin}_{\beta\in\mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^*\beta\right),\ y_i \in \mathbb{R},\ x_i \in \mathbb{R}^N.$$

More general framework: BigData RMT

- Kernel random matrices (application to spectral clustering, kernel PCA)
- (A)symmetric graphical models (adjacency, Laplacian matrices), neural nets

Perspectives

Takeaway messages

- Robust estimators of scatter: $\begin{cases} \text{difficult to study for each } N, n \\ \text{become tractable when } N, n \to \infty. \end{cases}$
- Approximations useful up to second-order fluctuations.
- In practice, allows for (optimized) robust improvement of classical statistics.

Still on robust statistics

- Further properties of robust estimators of scatter (fluctuations of linear statistics).
- Accurate characterization of gain versus SCM with deterministic outliers.
- Extension to robust regression: first results in **[El Karoui'13]**

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^N} \sum_{i=1}^{n} f\left(y_i - x_i^* \beta\right), \; y_i \in \mathbb{R}, \; x_i \in \mathbb{R}^N.$$

More general framework: BigData RMT

- Kernel random matrices (application to spectral clustering, kernel PCA)
- (A)symmetric graphical models (adjacency, Laplacian matrices), neural nets
- Sparsity considerations (sparse PCA).

Outline

Bibliographical references I

- P. Bianchi, J. Najim, M. Maida, M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, 2011.
- L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative spectrum sensing using random matrix theory," International Symposium on Wireless Pervasive Computing, Santorini, Greece, 2008.
- Y. Chen, A. Wiesel, A. O. Hero. "Robust shrinkage estimation of high-dimensional covariance matrices." IEEE Transactions on Signal Processing, vol. 59, no. 9, pp. 4097-4107, 2011.
- R. Couillet, "Robust spiked random matrices and a robust G-MUSIC estimator", (submitted to) Elsevier Journal of Multivariate Analysis, 2014, arXiv Preprint 1404.7685.
- R. Couillet, A. Kammoun, F. Pascal, "Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals", (submitted to) Elsevier Journal of Multivariate Analysis, 2014, arXiv Preprint 1410.0817.
- R. Couillet, M. McKay, "Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators", Elsevier Journal of Multivariate Analysis, vol. 131, pp. 99-120, 2014.
- R. Couillet, F. Pascal, J. W. Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples", (to appear in) Elsevier Journal of Multivariate Analysis, 2013.

Bibliographical references II

- 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.
- N. El Karoui, "The spectrum of kernel random matrices." The Annals of Statistics, vol. 38, no. 1, pp. 1-50, 2010.
- N. El Karoui, "Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results." arXiv preprint arXiv:1311.2445, 2013.
- W. Hachem, P. Loubaton, J. Najim, "Deterministic equivalents for certain functionals of large random matrices," Annals of Applied Probability, vol. 17, no. 3, pp. 875-930, 2007.
- W. Hachem, P. Loubaton, X. Mestre, J. Najim, P. Vallet, "A subspace estimator for fixed rank perturbations of large random matrices," Journal of Multivariate Analysis, vol. 114, pp. 427-447, 2013.
- P. J. Huber, "Robust estimation of a location parameter," The Annals of Mathematical Statistics, vol. 35, no. 1, pp. 73-101, 1964.
- J. T. Kent, D. E. Tyler. "Redescending M-estimates of multivariate location and scatter." The Annals of Statistics, pp. 2102-2119, 1991.
- 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.
- R. A. Maronna, "Robust M-estimators of multivariate location and scatter", Annals of Statistics, vol. 4, pp. 51-67, 1976.

Bibliographical references III

- X. Mestre, “Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates,” IEEE Transactions on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.
- X. Mestre, M. A. Lagunas, “Modified subspace algorithms for DoA estimation with large arrays.” IEEE Transactions on Signal Processing, vol. 56, no. 2, pp. 598-614, 2008.
- D. Morales-Jimenez, R. Couillet, M. McKay, “Large dimensional analysis of Maronnas M-estimator with outliers”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’15), Brisbane, Australia, 2015.
- F. Pascal, Y. Chitour, Y. Quek, “Generalized robust shrinkage estimator – Application to STAP data”, 2013.
- 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.
- S. Shamai and S. Verdú, “The impact of frequency-flat fading on the spectral efficiency of CDMA,” IEEE Trans. on Information Theory, vol. 47, no. 4, pp. 1302-1327, 2001.
- I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” European Transactions on Telecommunications, vol. 10, no. 6, pp. 585-595, 1999.
- S. Wagner, R. Couillet, M. Debbah, D. T. M. Slock, “Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels under Limited Feedback,” IEEE Transactions on Information Theory, vol. 58, no. 7, pp. 4509-4537, 2012.

The End

Thank you.