Random Matrices and Robust Estimation

Random Matrices and Their Application Workshop.

Romain COUILLET

The University of HongKong

January 7, 2015

Outline

Covariance estimation and sample covariance matrices

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

Covariance estimation and sample covariance matrices

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

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

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

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

Covariance estimation and sample covariance matrices

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

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

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

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

Covariance estimation and sample covariance matrices

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

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

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

    $\to$ Hence, if $\theta = f(C_N)$, we often use the n-consistent estimate $\hat{\theta} = f(\hat{S}_N)$.
  - The SCM $\hat{S}_N$ is the ML estimate of C_N for Gaussian x
    $\to$ One therefore expects $\hat{\theta}$ to closely approximate θ for all finite n.
- This approach however has two limitations:
  - if N, n are of the same order of magnitude,

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

    $\to$ This motivated the introduction of G-estimators.
  - if x is not Gaussian, but has heavier tails, $\hat{S}_N$ is a poor estimator for C_N.
    $\to$ This motivated the introduction of robust estimators.

Reminders on robust estimation

→ The objectives of robust estimators:

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

→ **Example**: Huber estimator (Huber'67), $\hat{C}_N$ defined as solution of

$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} \alpha \min\left\{1, \frac{k^2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i}\right\} x_i x_i^* \text{ for some } \alpha > 1, k^2 > 0.$$

Reminders on robust estimation

$\rightarrow$ The objectives of robust estimators:

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

  $\rightarrow$ **Example**: Huber estimator (Huber'67), $\hat{C}_N$ defined as solution of

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^{n} \alpha \min \left\{ 1, \frac{k^2}{\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i} \right\} x_i x_i^* \text{ for some } \alpha > 1, k^2 > 0.$$

- Provide scale-free estimators of C_N:

  $\rightarrow$ **Example**: Tyler's estimator (Tyler'81): if one observes $x_i = \tau_i z_i$ for unknown scalars τ_i,

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

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

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by the ML estimate for C_N.
  → **Example**: Maronna's estimator (Maronna'76) for elliptical x

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

with $u(s)$ such that

(i) $u(s)$ is continuous and non-increasing on $[0, \infty)$

(ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.

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

Reminders on robust estimation

→ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by the ML estimate for C_N.
  → **Example**: Maronna's estimator (Maronna'76) for elliptical x

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

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

Reminders on robust estimation

$\rightarrow$ The objectives of robust estimators:

- replace the SCM $\hat{S}_N$ by the ML estimate for C_N.
  $\rightarrow$ **Example**: Maronna's estimator (Maronna'76) for elliptical x

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

with $u(s)$ such that

(i) $u(s)$ is continuous and non-increasing on $[0, \infty)$

(ii) $\phi(s) = su(s)$ is non-decreasing, bounded by $\phi_\infty > 1$. Moreover, $\phi(s)$ increases where $\phi(s) < \phi_\infty$.

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

- existence is not too demanding
- uniqueness imposes strictly increasing $u(s)$ (inconsistent with Huber's estimate)
- consistency result: $\hat{C}_N \to C_N$ **if** $u(s)$ meets the ML estimator for C_N.

Robust RMT estimation

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

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

Setting and assumptions

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

Setting and assumptions

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

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

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

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

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

Setting and assumptions

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

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

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

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

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

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

RMT analysis of $\hat{C}_N$

- First remark: we can work with $C_N = I_N$ without generality restriction.

RMT analysis of $\hat{C}_N$

- First remark: we can work with $C_N = I_N$ without generality restriction.
- **Intuition:**
  - Denote

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

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

RMT analysis of $\hat{C}_N$

- First remark: we can work with $C_N = I_N$ without generality restriction.
- **Intuition:**
  - Denote

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

$\longrightarrow$ Intuitively, $\hat{C}_{(j)}$ and x_j are only "weakly" dependent.
  - We expect in particular:

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

for some deterministic equivalent γ_N.

RMT analysis of $\hat{C}_N$

- First remark: we can work with $C_N = I_N$ without generality restriction.
- **Intuition:**
  - Denote
$$\hat{C}_{(j)} = \frac{1}{n}\sum_{i\neq j}^{n} u\left(\frac{1}{N}x_i^* \hat{C}_N^{-1} x_i\right) x_i x_i^*$$
    $\longrightarrow$ Intuitively, $\hat{C}_{(j)}$ and x_j are only "weakly" dependent.
  - We expect in particular:
$$\frac{1}{N}x_i^* \hat{C}_{(i)}^{-1} x_i \simeq \tau_i \frac{1}{N}\operatorname{tr}\hat{C}_{(i)}^{-1} \simeq \tau_i \frac{1}{N}\operatorname{tr}\hat{C}_N^{-1} \simeq \tau_i \gamma_N$$
    for some deterministic equivalent γ_N.
- **Assuming this is correct, we then proceed as follows:**
  - *Algebraic manipulation*: For some function f (later called g^{-1}), write
$$\hat{C}_N = \frac{1}{n}\sum_{i=1}^{n} (u\circ f)\left(\frac{1}{N}x_i^* \hat{C}_{(i)}^{-1} x_i\right) x_i x_i^*$$
  - *Use conjecture* $\frac{1}{N}x_i^* \hat{C}_{(i)}^{-1} x_i \simeq \tau_i\gamma_N$ to get
$$\hat{C}_N \simeq \frac{1}{n}\sum_{i=1}^{n} (u\circ f)(\tau_i\gamma_N) x_i x_i^*$$
  - Use random matrix results to find a deterministic equivalent γ_N from $\gamma_N \simeq \frac{1}{N}\operatorname{tr}\hat{C}_N^{-1}$.

RMT analysis of $\hat{C}_N$: f and γ_N

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

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

  so that

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

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

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

RMT analysis of $\hat{C}_N$: f and γ_N

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

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

Hence

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

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

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

And we have:

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

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

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

Main result

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

Theorem (Asymptotic Equivalence)

Under the assumptions defined earlier, we have

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

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

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

Main result

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

Theorem (Asymptotic Equivalence)

Under the assumptions defined earlier, we have

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

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

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

- **Remarks:**
  - Corollary:

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

$\longrightarrow$ Important feature for detection and estimation.

Main result

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

Theorem (Asymptotic Equivalence)

Under the assumptions defined earlier, we have

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

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

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

- **Remarks:**
  - Corollary:

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

$\longrightarrow$ Important feature for detection and estimation.

- **Proof:** So far, we do not have a rigorous proof!

Proof of the "conjecture"

Proof of the "conjecture"

- **Technical trick:** Denote

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

and relabel terms such that

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

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

$$e_1 > 1 - \ell \text{ and } e_n < 1 + \ell \text{ for all large } n \text{ a.s.}$$

Proof of the "conjecture"

- **Technical trick:** Denote

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

and relabel terms such that

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

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

$$e_1 > 1 - \ell \text{ and } e_n < 1 + \ell \text{ for all large } n \text{ a.s.}$$

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

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

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

Proof

- **Specialization to e_n:**

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

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

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

Proof

- **Specialization to e_n:**

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

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

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

- **Random Matrix result:** We can prove precisely that:

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

(uniformity fundamental after relabeling)

Proof

- For all large n a.s., we then have (using growth of ψ)

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

Proof

- For all large n a.s., we then have (using growth of ψ)

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

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

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

Proof

- For all large n a.s., we then have (using growth of ψ)

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

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

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

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

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

Proof

- For all large n a.s., we then have (using growth of ψ)

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

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

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

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

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

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

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

Simulations

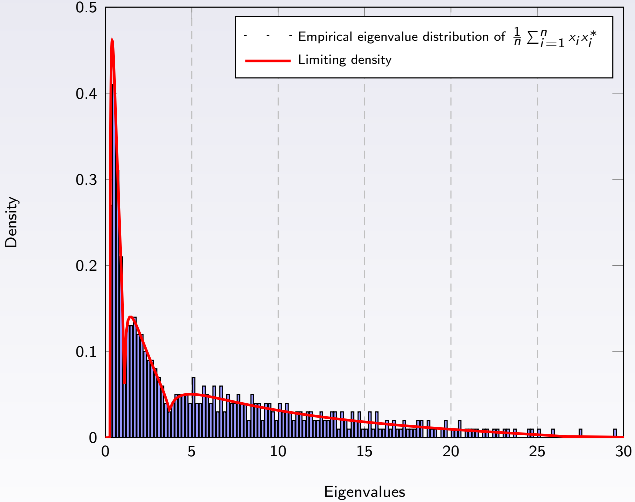
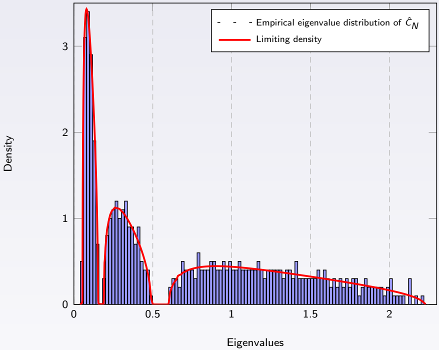


Figure: Histogram of the eigenvalues of $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^*$ for $n = 2500$, $N = 500$, $C_N = \mathrm{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

Simulations



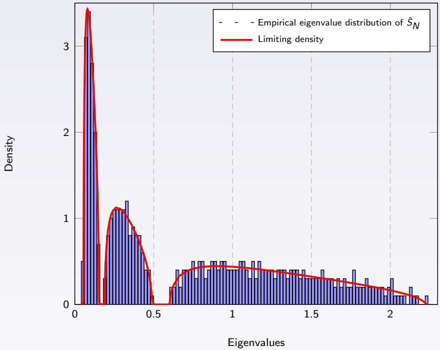


Figure: Histogram of the eigenvalues of $\hat{C}_N$ (left) and $\hat{S}_N$ (right) for $n = 2500$, $N = 500$, $C_N = \mathrm{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

Simulations

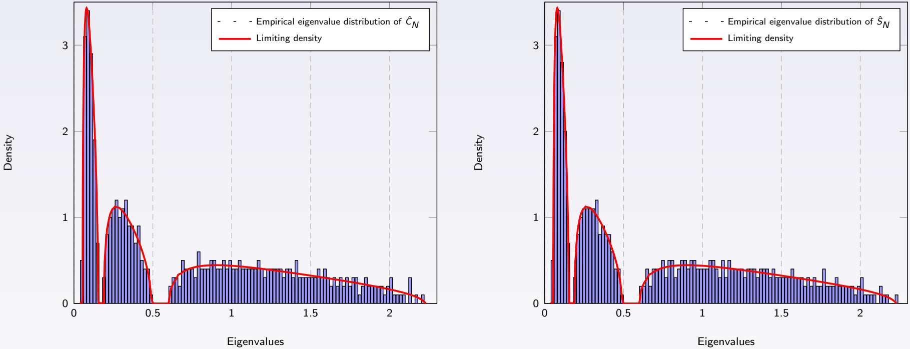


Figure: Histogram of the eigenvalues of $\hat{C}_N$ (left) and $\hat{S}_N$ (right) for $n = 2500$, $N = 500$, $C_N = \operatorname{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

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

Hint on potential applications

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

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

Hint on potential applications

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

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

  $\Rightarrow$ We shall see that we get **even better** than this...

Hint on potential applications

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

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

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

Outline

System Setting

- **Signal model:**

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

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

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

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

System Setting

- **Signal model:**

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

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

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

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

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

System Setting

- **Signal model:**

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

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

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

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

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

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

Some intuition

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

Some intuition

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

Some intuition

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

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

Theoretical results

Theorem (Extension to spiked robust model)

Under the same assumptions as in previous section,

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

where

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

with γ *the unique solution to*

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

and we recall

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

Theoretical results

Theorem (Extension to spiked robust model)

Under the same assumptions as in previous section,

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

where

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

with γ the unique solution to

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

and we recall

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

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

Localization of eigenvalues

Theorem (Eigenvalue localization)

Denote

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

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

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

Further denote

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

Localization of eigenvalues

Theorem (Eigenvalue localization)

Denote

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

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

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

Further denote

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

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

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

Simulation

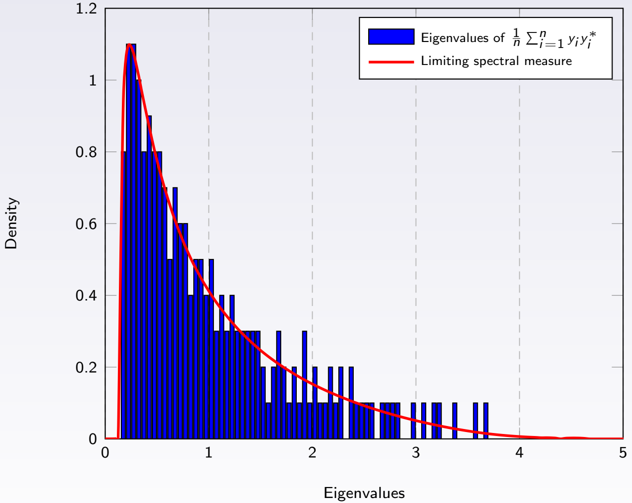


Figure: Histogram of the eigenvalues of $\frac{1}{n}\sum_i y_i y_i^*$ against the limiting spectral measure, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Sudent-t impulsions.

Simulation

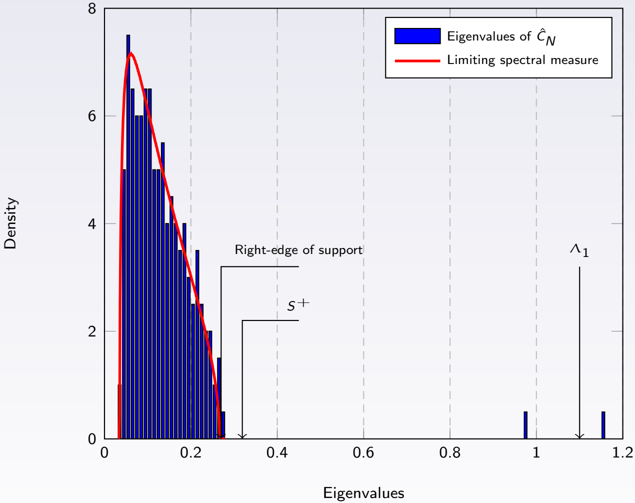


Figure: Histogram of the eigenvalues of $\hat{C}_N$ against the limiting spectral measure, for $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Comments

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

Comments

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

$\Rightarrow$ Induces a natural source detection algorithm.

Eigenvalue and eigenvector projection estimates

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

Eigenvalue and eigenvector projection estimates

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

Theorem (Estimation under known ν)

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

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

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

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

where

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

Eigenvalue and eigenvector projection estimates

Theorem (Estimation under unknown ν)

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

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

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

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

where

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

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

Application to G-MUSIC

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

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

Application to G-MUSIC

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

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

Corollary (Robust G-MUSIC)

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

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

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

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

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

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

where

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

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

Simulations: Single-shot in elliptical noise

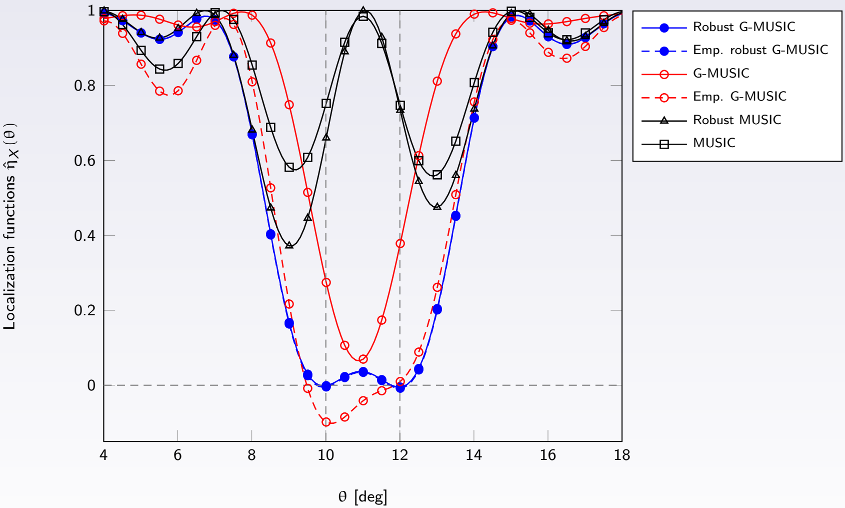


Figure: Random realization of the localization functions for the various MUSIC estimators, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 100$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$. Powers $p_1 = p_2 = 10^{0.5} = 5$ dB.

Simulations: Elliptical noise

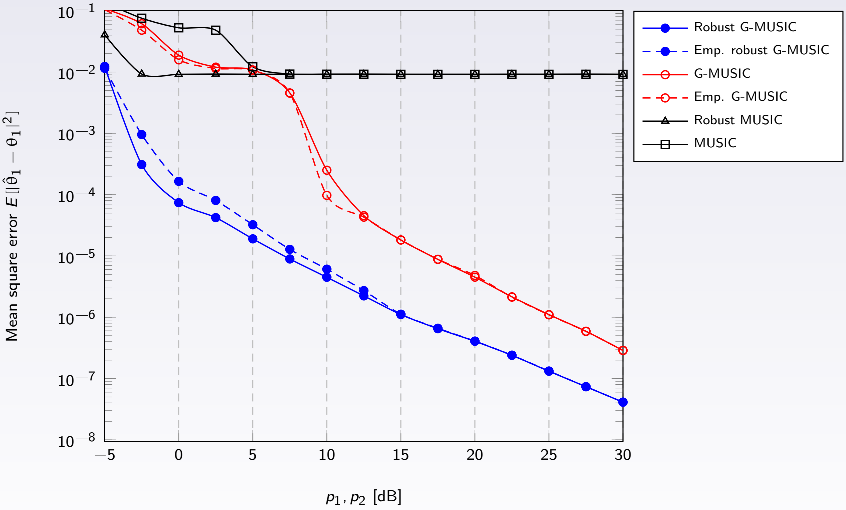


Figure: Means square error performance of the estimation of $\theta_1 = 10^\circ$, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 10$, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Simulations: Spurious impulses

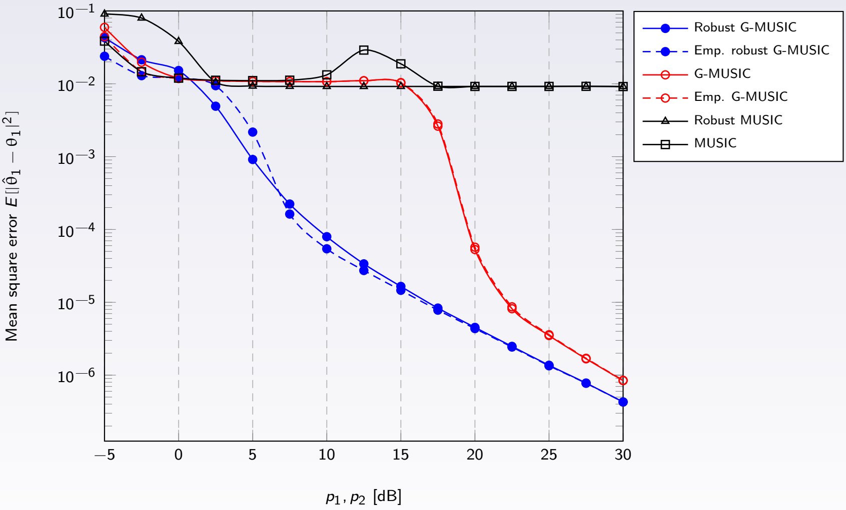


Figure: Means square error performance of the estimation of $\theta_1 = 10^\circ$, with $N = 20$, $n = 100$, two sources at 10° and 12°, sample outlier scenario $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Outline

Context

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

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

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

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

Context

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

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

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

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

Context

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

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

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

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

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

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

Main theoretical result

- **Which estimator is better?**

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

Main theoretical result

- **Which estimator is better?**

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

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

Main theoretical result

- **Which estimator is better?**

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

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

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

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

Pascal's estimator

Theorem (Pascal's estimator)

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

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

where

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

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

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

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

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

Chen's estimator

Theorem (Chen's estimator)

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

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

where

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

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

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

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

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

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

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

Asymptotic Model Equivalence

Theorem (Model Equivalence)

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

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

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

Asymptotic Model Equivalence

Theorem (Model Equivalence)

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

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

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

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

Asymptotic Model Equivalence

Theorem (Model Equivalence)

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

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

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

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

Optimal Shrinkage parameter

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

Optimal Shrinkage parameter

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

Optimal Shrinkage parameter

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

Theorem (Optimal Shrinkage)

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

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

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

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

Then, letting ε small enough,

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

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

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

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

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

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

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

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

Optimal Shrinkage Estimate

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

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

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

defined arbitrarily when no such solutions exist. Then

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

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

Simulations

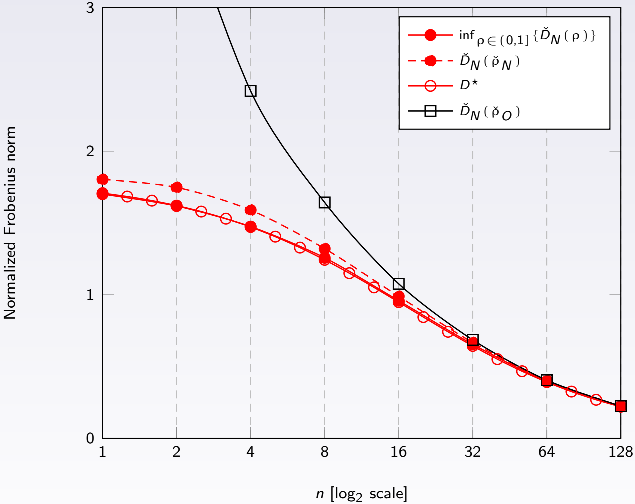


Figure: Performance of optimal shrinkage averaged over 10 000 Monte Carlo simulations, for $N = 32$, various values of n, $[C_N]_{ij} = r^{|i-j|}$ with $r = 0.7$; $\check{\rho}_N$ as above; $\check{\rho}_O$ the clairvoyant estimator proposed in (Chen'11).

Simulations

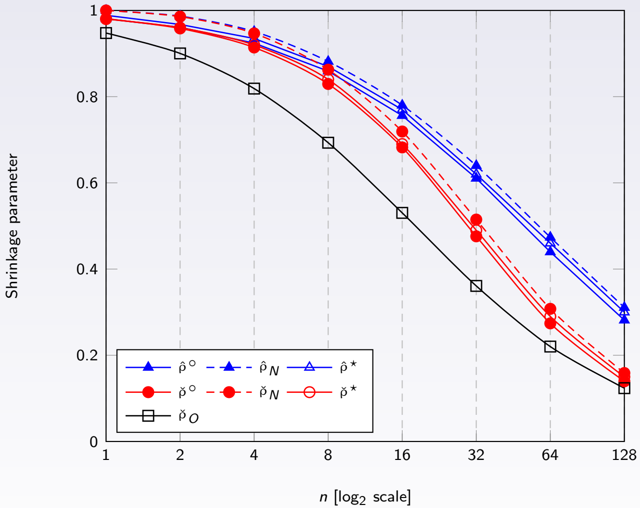


Figure: Shrinkage parameter ρ averaged over 10 000 Monte Carlo simulations, for $N = 32$, various values of n, $[C_N]_{ij} = r^{|i-j|}$ with $r = 0.7$; $\hat{\rho}_N$ and $\check{\rho}_N$ as above; $\check{\rho}_O$ the clairvoyant estimator proposed in (Chen'11); $\hat{\rho}^\circ = \operatorname{argmin}_{\{\rho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

Connection to Power Control

- Power control problem results in solving, for each $j = 1, \dots, n$

$$\lambda_j = \sigma^2 \left((1+\gamma_j^{-1}) \frac{1}{N} h_j^* \left(\frac{1}{N} \sum_{i=1}^{n} \frac{\lambda_i}{\sigma^2} h_i h_i^* + I_N \right)^{-1} h_j \right)^{-1}$$

with

- $h_i \in \mathbb{C}^N$ channel modeled as $h_i = \sqrt{r_i} x_i$, $x_i \sim \mathcal{CN}(0, I_N)$
- σ^2 power of additive noise
- γ_i target SINR for user i

Connection to Power Control

- Power control problem results in solving, for each $j = 1, \dots, n$

$$\lambda_j = \sigma^2 \left((1+\gamma_j^{-1}) \frac{1}{N} h_j^* \left(\frac{1}{N} \sum_{i=1}^{n} \frac{\lambda_i}{\sigma^2} h_i h_i^* + I_N \right)^{-1} h_j \right)^{-1}$$

  with
  - $h_i \in \mathbb{C}^N$ channel modeled as $h_i = \sqrt{r_i} x_i$, $x_i \sim \mathcal{CN}(0, I_N)$
  - σ^2 power of additive noise
  - γ_i target SINR for user i
- Can be rewritten as

$$\lambda_j = \frac{\sigma^2 \gamma_j}{r_j} \left(\frac{1}{N} x_j^* \left(\frac{1}{N} \sum_{i \neq j} \frac{\lambda_i r_i}{\sigma^2} x_i x_i^* + I_N \right)^{-1} x_j \right)^{-1}$$

Connection to Power Control

- Power control problem results in solving, for each $j = 1, \ldots, n$

$$\lambda_j = \sigma^2 \left((1+\gamma_j^{-1}) \frac{1}{N} h_j^* \left(\frac{1}{N} \sum_{i=1}^{n} \frac{\lambda_i}{\sigma^2} h_i h_i^* + I_N \right)^{-1} h_j \right)^{-1}$$

  with
  - $h_i \in \mathbb{C}^N$ channel modeled as $h_i = \sqrt{r_i} x_i$, $x_i \sim \mathcal{CN}(0, I_N)$
  - σ^2 power of additive noise
  - γ_i target SINR for user i
- Can be rewritten as

$$\lambda_j = \frac{\sigma^2 \gamma_j}{r_j} \left(\frac{1}{N} x_j^* \left(\frac{1}{N} \sum_{i \neq j} \frac{\lambda_i r_i}{\sigma^2} x_i x_i^* + I_N \right)^{-1} x_j \right)^{-1}$$

- With $d_j = \frac{\sigma^2 \gamma_j}{\lambda_j r_j}$, this is

$$d_j = \frac{1}{N} x_j^* \left(\frac{1}{N} \sum_{i \neq j} \frac{\gamma_i}{d_i} x_i x_i^* + I_N \right)^{-1} x_j .$$

Connection to Power Control

- Power control problem results in solving, for each $j = 1, \dots, n$

$$\lambda_j = \sigma^2 \left((1+\gamma_j^{-1}) \frac{1}{N} h_j^* \left(\frac{1}{N} \sum_{i=1}^{n} \frac{\lambda_i}{\sigma^2} h_i h_i^* + I_N \right)^{-1} h_j \right)^{-1}$$

  with
  - $h_i \in \mathbb{C}^N$ channel modeled as $h_i = \sqrt{r_i} x_i$, $x_i \sim \mathcal{CN}(0, I_N)$
  - σ^2 power of additive noise
  - γ_i target SINR for user i
- Can be rewritten as

$$\lambda_j = \frac{\sigma^2 \gamma_j}{r_j} \left(\frac{1}{N} x_j^* \left(\frac{1}{N} \sum_{i \neq j} \frac{\lambda_i r_i}{\sigma^2} x_i x_i^* + I_N \right)^{-1} x_j \right)^{-1}$$

- With $d_j = \frac{\sigma^2 \gamma_j}{\lambda_j r_j}$, this is

$$d_j = \frac{1}{N} x_j^* \left(\frac{1}{N} \sum_{i \neq j} \frac{\gamma_i}{d_i} x_i x_i^* + I_N \right)^{-1} x_j.$$

- Under assumption $\limsup_n \frac{1}{N} \sum_{i=1}^n \frac{\gamma_i}{1+\gamma_i} < 1$ we then have

$$\max_{1 \leqslant j \leqslant n} \left| \lambda_j - \frac{\sigma^2 \gamma_j}{r_j} \left(1 - \frac{1}{N} \sum_{i=1}^{n} \frac{\gamma_i}{1+\gamma_i} \right)^{-1} \right| \xrightarrow{\text{a.s.}} 0.$$

Outline

Context

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

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

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

Context

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

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

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

- **GLRT detection test:**

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

for some detection threshold Γ where

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

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

Context

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

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

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

- **GLRT detection test:**

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

for some detection threshold Γ where

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

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

$\longrightarrow$ In fact, originally found to be $\hat{C}_N(0)$ but

- only valid for $N < n$
- introducing ρ may bring improved for arbitrary N/n ratios.

Objectives and main results

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

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

$\Rightarrow$ Trivial result of little interest!

Objectives and main results

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

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

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

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

Objectives and main results

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

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

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

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

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

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

Objectives and main results

- **Initial observations:**
  - As $N, n \to \infty$, $N/n \to c > 0$, under $\mathcal{H}_0$,
$$T_N(\rho) \xrightarrow{\text{a.s.}} 0.$$
    $\Rightarrow$ Trivial result of little interest!
  - Natural question: for finite N, n and given Γ, find ρ such that
$$P\left(T_N(\rho) > \Gamma\right) = \min$$
  - Turns out the correct non-trivial object is, for $\gamma > 0$ fixed
$$P\left(\sqrt{N}T_N(\rho) > \gamma\right) = \min$$
- **Objectives:**
  - for each ρ, develop central limit theorem to evaluate
$$\lim_{\substack{N,n\to\infty \\ N/n\to c}} P\left(\sqrt{N}T_N(\rho) > \gamma\right)$$
  - determine limiting minimizing ρ
  - empirically estimate minimizing ρ

What do we need?

CLT over $\hat{C}_N$ statistics

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

What do we need?

CLT over $\hat{C}_N$ statistics

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

What do we need?

CLT over $\hat{C}_N$ statistics

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

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

$\Rightarrow$ This is our main result!

What do we need?

CLT over $\hat{C}_N$ statistics

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

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

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

Main results

Theorem (Fluctuation of bilinear forms)

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

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

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

False alarm performance

Theorem (Asymptotic detector performance)

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

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

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

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

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

False alarm performance

Theorem (Asymptotic detector performance)

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

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

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

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

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

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

Simulation

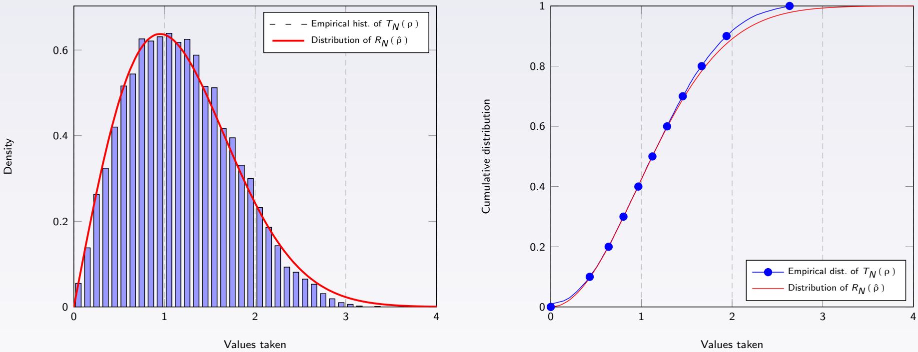


Figure: Histogram distribution function of the $\sqrt{N}T_N(\rho)$ versus $R_N(\hat{\rho})$, $N = 20$, $p = N^{-\frac{1}{2}}[1,\ldots,1]^T$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$, $\rho = 0.2$.

Simulation

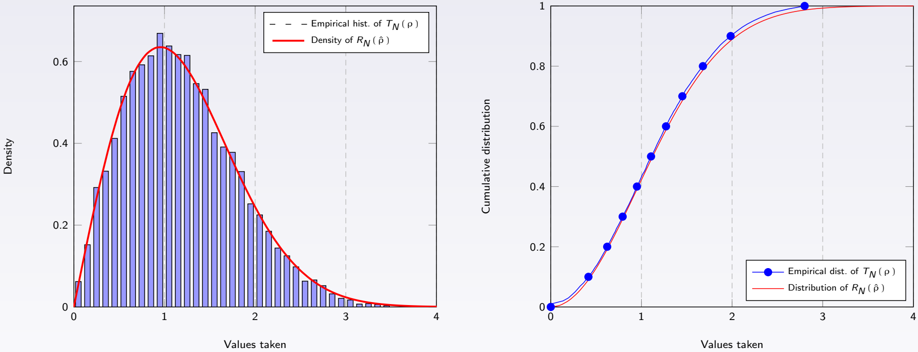


Figure: Histogram distribution function of the $\sqrt{N}T_N(\rho)$ versus $R_N(\hat{\rho})$, $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^{\mathsf{T}}$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$, $\rho = 0.2$.

Empirical estimation of optimal ρ

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

Empirical estimation of optimal ρ

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

Theorem (Empirical performance estimation)

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

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

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

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

Final result

Theorem (Optimality of empirical estimator)

Define

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

Then, for every $\gamma > 0$,

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

Simulations

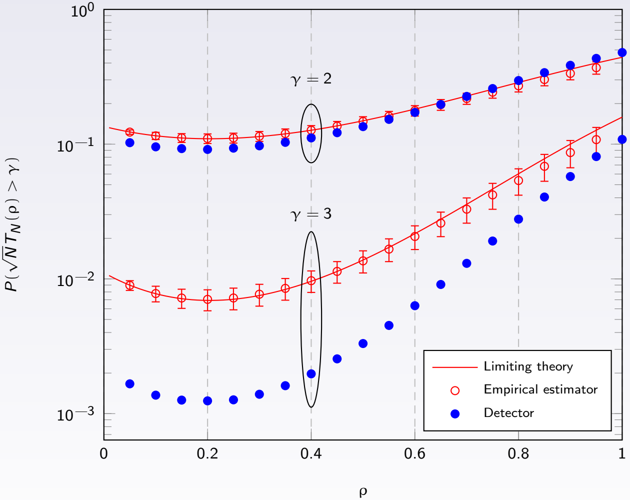


Figure: False alarm rate $P(\sqrt{N}T_N(\rho) > \gamma)$, $N = 20$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^{\mathsf{T}}$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$.

Simulations

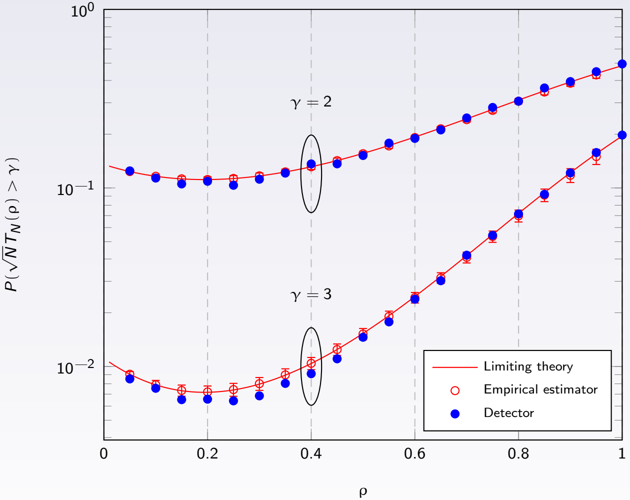


Figure: False alarm rate $P(\sqrt{N}T_N(\rho) > \gamma)$, $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, C_N Toeplitz from AR of order 0.7, $c_N = 1/2$.

Simulations

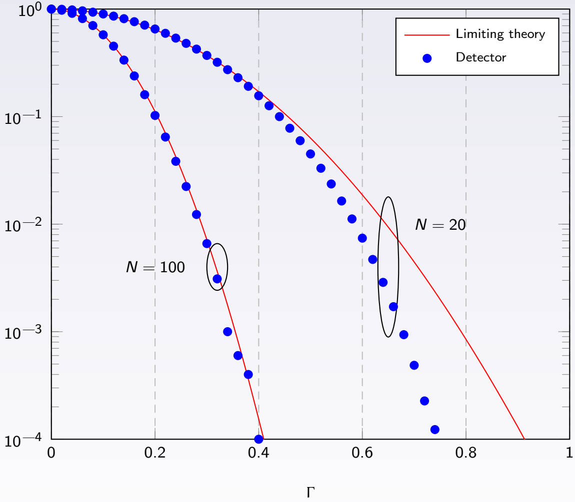


Figure: False alarm rate $P(T_N(\rho) > \Gamma)$ for $N = 20$ and $N = 100$, $p = N^{-\frac{1}{2}}[1, \ldots, 1]^\mathsf{T}$, $[C_N]_{ij} = 0.7^{|i-j|}$, $c_N = 1/2$.

Outline

Deterministic outliers

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

- $x_1, \ldots, x_{(1-\varepsilon_n)n}$ i.i.d. Gaussian zero mean covariance C_N
- $a_1, \ldots, a_{\varepsilon_n n}$ deterministic such that
  $0 < \min_i \liminf_n N^{-\frac{1}{2}} \|a_i\| \leqslant \max_i \limsup_n N^{-\frac{1}{2}} \|a_i\| < \infty$.

Theorem

Then, as $N, n \to \infty$,

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

with γ_n and $\alpha_{1,n}, \ldots, \alpha_{\varepsilon_n n,n}$ the unique positive solutions to the system of $\varepsilon_n n + 1$ equations ($i = 1, \ldots, \varepsilon_n n$)

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

$$\alpha_{i,n} = \frac{1}{N} a_i^* \left(\frac{(1-\varepsilon) v_c(\gamma_n)}{1 + c v_c(\gamma_n)\gamma_n} C_N + \frac{1}{n} \sum_{j \neq i}^{\varepsilon_n n} v(\alpha_{j,n}) a_j a_j^* \right)^{-1} a_i$$

and $v_c(x) = u\left(g^{-1}(x)\right)$, $g(x) = x/(1 - c\phi(x))$.

Comments

- Say $\varepsilon_n = 1/n \to 0$, then $\gamma_n \to \gamma$ with $\gamma = \phi^{-1}(1)/(1-c)$ and

$$\alpha_{1,n} = \left(\frac{\phi^{-1}(1)}{1-c} + o(1) \right) \frac{1}{N} a_1^* C_N^{-1} a_1.$$

- Rejection of outliers depends strongly on $\frac{1}{N} a_1^* C_N^{-1} a_1$ compared to 1.

Random outliers

Corollary

Assume now $a_i = D_N^{\frac{1}{2}} w_i$ *with* $\limsup_N \|D_N\| < \infty$. *Then,*

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

where

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

with γ_n *and* α_n *the unique positive solutions to*

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

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

- Now, for ε small, rejection depends on $\frac{1}{N} \mathrm{tr}\, D_N C_N^{-1}$.

Simulation example

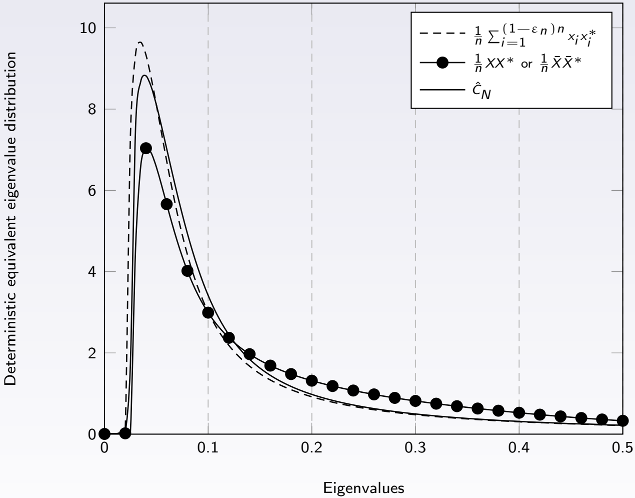


Figure: Limiting eigenvalue distributions. $[C_N]_{ij} = .9^{|i-j|}$, $D_N = I_N$, $\varepsilon = .05$.

The End

Thank you.