Estimation of High Dimentional Covariance Matrices

A. Kammoun
KAUST

Interest for the estimation of covariance matrices

- Many applied estimation methods rely on the second-order statistics of the observed random vector process:

$$\mathbf{y}_i = \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{x}_i, \, i = 1, \cdots, n \quad \text{and} \quad \mathbf{y}_i \in \mathbb{C}^N$$

$$\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n], \; \mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \Rightarrow \mathbf{Y} = \mathbf{\Sigma}^{\frac{1}{2}} \mathbf{X}$$

  Examples:
  - Array processing: Estimation of direction of arrivals
  - Inference of performance metrics (Capacity or SINR)

- Possible asymptotic regimes:
  - Number of observations is large as compared to the dimension of the received vector, $n \gg N$
  - Number of observations is of the same order of magnitude as the dimension of the received vector $n \propto N$. In the sequel, This regime will be referred to as : The Regime of Large Random Matrices

Shortcomings of the conventional estimation techniques

- Consider n observations $y_1, \cdots, y_n$ independent and identically distributed where

$$\mathbf{y}_i = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{x}_i, i = 1, \cdots, n \ \text{ and } \ \mathbf{y}_i \in \mathbb{C}^N$$

with the entries of $\mathbf{x}_i$ are zero mean indendent random variables such that $\mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}\right] = \mathbf{I}_N$. The sample covariance estimate $\widehat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$ is given by:

$$\widehat{\mathbf{\Sigma}} = \sum_{i=1}^{n} \mathbf{y}_i\mathbf{y}_i^{\mathrm{H}} = \frac{\mathbf{Y}\mathbf{Y}^{\mathrm{H}}}{n}$$

- If $n \gg N$, we know that $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \xrightarrow[n\to\infty]{a.s.} 0.$.
- If $n \propto N$ ($\frac{N}{n} \to c$), $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|$ does not converge to zero as N and n grow at infinity.

$\Rightarrow$ Need to develop new techniques adapted to this regime

Theory of Large Random Matrices

- The theory of large random matrices offers powerful tools to deal with the regime N and n growing to infinty at the same pace.
- Brief History of the use of this theory:
  - First results of random matrix theory were motivated by nuclear physics (Wigner 1950),
  - This theory has been firstly applied to the field of wireless communication in the mid-90 with the pioneering works of Telatar and Foshini.

  $\Rightarrow$ New approaches to analyze the performance of wireless communication systems.

  - In 2007, X. Mestre uses this theory to propose new consistant estimation methods,

  $\Rightarrow$ Practical interest of this theory to propose new effecient algorithms,

Results from the theory of large random matrices

- First results concern Wigner matrices: symmetric random matrices whose diagonal entries are 0 and whose upper-triangle entries are i.i.d taking the values ± 1.
- Let $\mathbf{W}_N \in \mathbb{C}^{N\times N}$ be a Wigner matrix.
  Then, as $N \to +\infty$, the distribution of the eigenvalues of $\frac{1}{\sqrt{N}}\mathbf{W}_N$ converges to the semi-circle law whose density is:

$$w(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & \\ 0 & \text{if } |x| > 2. \end{cases}$$

- This result was proven using the Moments method.

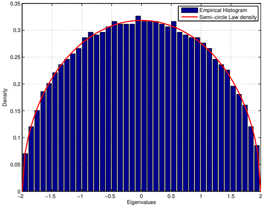


Figure: Histogram and Limit law ($N = 2000$)

Results from the theory of large random matrices

Generalization to singular values of rectangular matrices

- Limit Law of the eigenvalue distribution of

$$\mathbf{Y} = \mathbf{Y}_0 + \mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}$$

  where $\mathbf{T}$ is a real diagonal matrix independent of $\mathbf{X}$ and $\mathbf{Y}_0$ is a deterministic hermitian matrix.

- If $\mathbf{Y}_0 \neq 0$ or $\mathbf{T} \neq \mathbf{I}$, no closed form expression exists for the limiting spectrum.

- Instead, it can be characterized using the Stieltjes transform method.

Stieltjes Transform of a random variable

- Stieltjes tranform: Let X be a real-valued random variable with distribution $F_X(.)$ (with support $\mathcal{S}$). The stieltjes transform is the Cauchy transform of F_X defined for complex arguments outside the support $\mathcal{S}$ as:

$$m_X(z) = \mathbb{E}\left[\frac{1}{X-z}\right] = \int_{-\infty}^{+\infty} \frac{1}{\lambda - z} dF_X(\lambda)$$

- The Stieltjes transform is to the cauchy transform as the characteristic function is to the Fourier transform.
- Given $m_X(z)$, the probability density function of X can be retrieved as :

$$f_X(\lambda) = \lim_{\omega \to 0^+} \frac{1}{\pi} \Im\left[m_X(\lambda + j\omega)\right]$$

Stieltjes Transform of the covariance matrix

Stieltjes transform of a matrix is the Stieltjes transform of the empirical measure of its eigenvalues:

- Consider the model:

$$\mathbf{Y} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{X}$$

- Let $\lambda_1, \cdots, \lambda_N$ be the eigenvalues of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$. Denote by $dF^{\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$ the empirical measure for $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$:

$$dF^{\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} = \frac{1}{N}\sum_{k=1}^{N}\delta_{\lambda_k}$$

- The Stieltjes transform for $dF^{\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$ is therefore given by:

$$\widehat{m}_{\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}(z) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\lambda_k - z} = \frac{1}{N}\operatorname{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}$$

$\Rightarrow$ The study of the Stieltjes transform of the empirical measure amounts to studying the behaviour of the trace of the resolvent matrix:

$$\mathbf{Q}(z) \triangleq \left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} - zI_N\right)^{-1}$$

Stieltjes transform of covariance matrices

- Let $\alpha_1, \cdots, \alpha_N$ be the eigenvalues of $\mathbf{\Sigma}$.
  In the regime of large random matrices when $N, n \to +\infty$ with $\frac{N}{n} \to c$, the stietljes transform of the empirical measure converges almost surely to $\overline{m}(z)$ where $\overline{m}(z)$ is the unique solution to:

$$\overline{m} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\alpha_i(1 - c - cz\overline{m}) - z}.$$

- Let $\theta = \frac{1}{N} \sum_{k=1}^{N} f(\alpha_k)$, where f is a certain function. The conventional empirical estimate of θ is

$$\widehat{\theta}_N = \frac{1}{N} \sum_{i=1}^{N} f(\lambda_i)$$

  With the above result on hand, we can study the behaviour of $\widehat{\theta}_N$ since,

$$\frac{1}{N} \sum_{k=1}^{N} f(\lambda_i) = \frac{1}{2\pi \jmath} \oint_C f(z) \widehat{m}_{\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}}(z) dz \to \frac{1}{2\pi \jmath} \oint_C f(z) \overline{m}(z) dz \neq \theta$$

Moment method for the estimation of eigenvalues

Work published in [IEEE-SP 2012]

- System model:

$$\mathbf{Y} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{X}$$

where $\mathbf{\Sigma}$ has a finite number K of unknown eigenvalues $\beta_1, \cdots, \beta_K$ with unknown multiplicities $N_1, \cdots, N_K$.

- We consider the problem of estimating the eigenvalues along with their multiplicities in the regime

$$N, n \to +\infty \text{ with } \frac{N}{n} \to c.$$

Moment-method for the estimation of eigenvalues

- We assume that $\mathbf{\Sigma}$ has K eigenvalues $\beta_1, \cdots, \beta_K$ with unknown multiplicites $N_1, \cdots, N_K$,
- The moment method is based on a consistent estimation of the moments

$$\theta_\ell \triangleq \sum_{k=1}^{N} \frac{N_k}{N} \beta_k^\ell$$

- Write θ_ℓ as function of the non-observable Stieltjes transform:

$$\theta_\ell = \sum_{i=1}^{K} \frac{N_i}{N} \beta_i^k = \frac{1}{2\pi \jmath} \oint_{\mathcal{C}} \omega^\ell m_{\Sigma}(\omega) d\omega$$

Moment-method for the estimation of eigenvalues

- The limit $\underline{m}_N(z)$ of the observable Stieltjes transform of matrix $\frac{1}{N}\mathbf{Y}^{\mathsf{H}}\mathbf{Y}$ satisfies:

$$\underline{m}_N = -zm_\Sigma\left(-\frac{1}{\underline{m}}\right)\left(\frac{n}{N}m_\Sigma\left(-\frac{1}{\underline{m}_N(z)}\right)-\frac{N-n}{Nz}\right)$$

- Using the changing variable $\omega = -\frac{1}{\underline{m}_N(z)}$, we obtain,

$$\theta_k = \frac{(-1)^k}{2\jmath\pi}\oint_{\mathcal{C}}\frac{z\underline{m}'_N(z)}{\underline{m}_N^{k+1}(z)}\left(c\underline{m}_N(z)+\frac{1-c}{cz}\right).$$

Moment method for the estimation of the eigenvalues

- Recall that :

$$\theta_k = \frac{(-1)^k}{2\jmath\pi} \oint_{\mathcal{C}} \frac{z\underline{m}_N'(z)}{\underline{m}_N^{k+1}(z)} \left(c\underline{m}_N(z) + \frac{1-c}{cz} \right).$$

$\Rightarrow$ Compute consistent estimates $\widehat{\theta}_k$ of θ_k by substituting $\underline{m}_N(z)$ by the stieltjes transform corresponding to matrix $\frac{1}{N}\mathbf{Y}^{\mathsf{H}}\mathbf{Y}$.

- Consider the following system of equations

$$\begin{cases} \sum_{i=1}^{K} x_i & = 1 \\ \sum_{i=1}^{K} x_i y_i^k & = \widehat{\theta}_k \end{cases}$$

admits a unique solution $\left(\widehat{c}_1, \cdots, \widehat{c}_K, \widehat{\beta}_1, \cdots, \widehat{\beta}_K\right)$ which is a consistent estimator of $\left(\frac{N_1}{N}, \cdots, \frac{N_K}{N}\right)$ and $(\beta_1, \cdots, \beta_K)$, i.e, for each $1 \leq \ell \leq N$,

$$\widehat{c}_\ell - \frac{N_\ell}{N} \xrightarrow[n,N\to+\infty]{a.s.} 0 \text{ and } \widehat{\beta}_\ell - \beta_\ell \xrightarrow[n,N\to+\infty]{a.s.} 0$$

Moment method for the estimation of the eigenvalues

Algorithm for solving to the obtained system of equations:

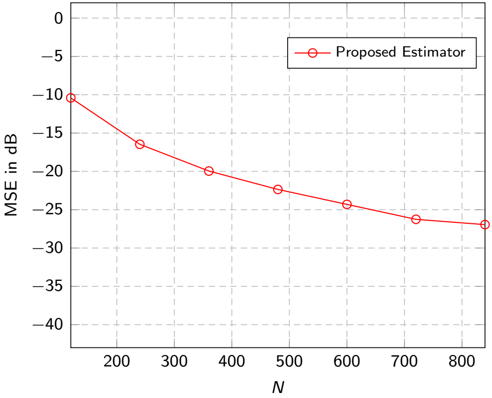
- Let $\mathbf{\Gamma} = \begin{bmatrix} \widehat{\theta}_0 & \widehat{\theta}_1 & \cdots & \widehat{\theta}_{K-1} \\ \widehat{\theta}_1 & \widehat{\theta}_2 & \cdots & \widehat{\theta}_K \\ \vdots & \ddots & \ddots & \vdots \\ \widehat{\theta}_{K-1} & \widehat{\theta}_K & \cdots & \widehat{\theta}_{2K-2} \end{bmatrix}$ and $\mathbf{b} = \left[\widehat{\theta}_L, \cdots, \widehat{\theta}_{2K-1}\right]^{\mathsf{T}}$,
- Consider $\mathbf{s} = -\Gamma^{-1}\mathbf{b}$, then, the solution $\widehat{\beta}_1, \cdots, \widehat{\beta}_L$ are solutions to the K order polynomial whose coefficients are given by $s_0, \cdots, s_K$.
- The vector $\widehat{\mathbf{c}} = \widehat{c}_1, \cdots, \widehat{c}_L$ is given by

$$\widehat{\mathbf{c}} = \mathbf{A}^{-1}\mathbf{d}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ \widehat{\beta}_1 & \cdots & \widehat{\beta}_K \\ \vdots & \vdots & \vdots \\ \widehat{\beta}_1^{K-1} & \cdots & \widehat{\beta}_K^{K-1} \end{bmatrix} \text{ and } \mathbf{d} = \left[\widehat{\beta}_0, \cdots, \widehat{\beta}_K\right]^{\mathsf{T}}$$

Numerical illustrations

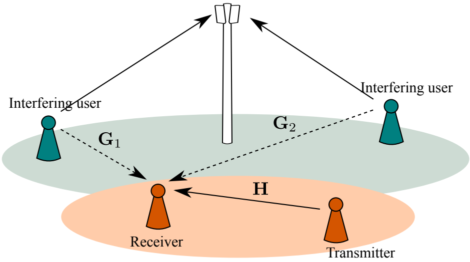


- $\beta_1 = 1, \beta_2 = 2, \beta_3 = 3$
- $\frac{N_1}{N} = 0.5, \ \frac{N_2}{N} = \frac{N_3}{N} = 0.25,$
- $\frac{N}{n} = \frac{3}{20}$

Consistent estimation of Functionals of the covariance matrix

Work published in [IEEE-IT,2013]

- In some scenarios, the parameter to be estimated is function on the eigenvalues and the eigenvectors of the covariance matrix,
- This scenario cannot be dealt with using the Stieltjes transform method.
- Example of scenario: estimation of the capacity under coloured interference:



Communication model and ergodic capacity

Communication equation: 2 phases

- A learning phase: $\mathbf{Y} = \sum_{k=1}^{I} \mathbf{G}_k \mathbf{X}_k + \sigma \mathbf{W} \triangleq \mathbf{G}\mathbf{X} + \sigma \mathbf{W}$
- A communication phase: $\mathbf{Y}_2 = \mathbf{H}\mathbf{X}_0 + \sum_{k=1}^{I} \mathbf{G}_k \mathbf{X}_{k,2} + \sigma \mathbf{W}_2$
- Assumption: The channel $\mathbf{H}$ is perfectly known,

Associated ergodic capacity:

$$C_{\mathrm{erg}} = \frac{1}{N} \log\det\left(\sigma^2 \mathbf{I} + \mathbf{G}\mathbf{G}^{\mathsf{H}} + \mathbf{H}\mathbf{H}^{\mathsf{H}}\right) - \frac{1}{N} \log\det\left(\sigma^2 \mathbf{I} + \mathbf{G}\mathbf{G}^{\mathsf{H}}\right)$$

Objective

Estimation of the ergodic capacity:

$$C_{\mathrm{erg}} = \frac{1}{N} \log\det\left(\sigma^2 \mathbf{I} + \mathbf{G}\mathbf{G}^{\mathsf{H}} + \mathbf{H}\mathbf{H}^{\mathsf{H}}\right) - \frac{1}{N} \log\det\left(\sigma^2 \mathbf{I} + \mathbf{G}\mathbf{G}^{\mathsf{H}}\right)$$

based on the $N \times n$ observations:

$$\mathbf{Y} = \mathbf{G}\mathbf{X} + \sigma \mathbf{W}$$

Regime of interest: n larger but of the same order of magnitude as n. Formally,

$$1 < \liminf \frac{n}{N} \leq \limsup \frac{n}{N} < +\infty.$$

Traditional estimator

Regime when $n \gg N$.

$$\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}} \xrightarrow[n\to+\infty, N\text{fixed}]{} \sigma^2\mathbf{I} + \mathbf{G}\mathbf{G}^{\mathrm{H}}$$

Hence,

$$\frac{1}{N}\log\det\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}} + \mathbf{H}\mathbf{H}^{\mathrm{H}}\right) - \frac{1}{N}\log\det\left(\sigma^2\mathbf{I} + \mathbf{G}\mathbf{G}^{\mathrm{H}} + \mathbf{H}\mathbf{H}^{\mathrm{H}}\right) \xrightarrow[n\to+\infty]{a.s} 0$$

$$\frac{1}{N}\log\det\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right) - \frac{1}{N}\log\det\left(\sigma^2\mathbf{I} + \mathbf{G}\mathbf{G}^{\mathrm{H}}\right) \xrightarrow[n\to+\infty]{a.s} 0$$

Define the parametrized traditional estimator as:

$$\hat{C}_{\mathrm{trad}}(y) = \frac{1}{N}\log\det\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}} + y\mathbf{H}\mathbf{H}^{\mathrm{H}}\right) - \frac{1}{N}\log\det\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)$$

If N is fixed and $n \to +\infty$, then:

$$\hat{C}_{\mathrm{trad}}(1) - C_{\mathrm{erg}} \to 0.$$

Deterministic equivalents

Fundamental equation: Let $y > 0$. The following equation in $\kappa = \kappa(y)$ admits a unique positive solution:

$$\kappa(y) = \frac{1}{n} \operatorname{tr} \left(\sigma^2 \mathbf{I} + \mathbf{G}\mathbf{G}^{\mathsf{H}}\right) \left(\frac{\sigma^2 \mathbf{I} + \mathbf{G}\mathbf{G}^{\mathsf{H}}}{1+\kappa} + y \mathbf{H}\mathbf{H}^{\mathsf{H}} \right)^{-1}$$

admits a unique positive solution,
Auxiliary quantities:

$$\mathbf{T}(y) = \left(y \mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{\sigma^2 \mathbf{I} + \mathbf{G}\mathbf{G}^{\mathsf{H}}}{1+\kappa} \right)^{-1}, \ \mathbf{Q}(y) = \left(y \mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{1}{n} \mathbf{Y}\mathbf{Y}^{\mathsf{H}} \right)^{-1}.$$

$$\frac{1}{N} \operatorname{tr} \mathbf{Q}(y) - \frac{1}{N} \operatorname{tr} \mathbf{T}(y) \xrightarrow[n,N \to +\infty]{a.s} 0.$$

Asymptotic results

For $y > 0$ and also for $y = 0$,

$$\frac{1}{N}\log\det\left(y\mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right) - \frac{1}{N}\log\det\left(y\mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{\sigma^2\mathbf{I} + \mathbf{G}\mathbf{G}^{\mathrm{H}}}{1+\kappa}\right) - \frac{n}{N}\log(1+\kappa) + \frac{n}{N}\frac{\kappa}{1+\kappa} \to 0.$$

$\Rightarrow$, under the regime of interest:

$$\hat{C}_{\mathrm{trad}}(y) - \left(\frac{1}{N}\log\det\left(y\mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{\sigma^2\mathbf{I} + \mathbf{G}\mathbf{G}^{\mathrm{H}}}{1+\kappa}\right) - \frac{1}{N}\log\det\left(\sigma^2\mathbf{I} + \mathbf{G}\mathbf{G}^{\mathrm{H}}\right)\right) - \frac{n}{N}\log(1+\kappa) + \frac{n}{N}\frac{\kappa}{1+\kappa} + \frac{n-N}{N}\log\left(\frac{n-N}{n}\right) - 1 \to 0.$$

Setting of y

To make the desired quantity appear in $\hat{C}_{\mathrm{trad}}(y)$, parameter y should be set so that :

$$y = \frac{1}{1+\kappa(y)}$$

We can prove that there is a unique y^* verifying the above equation. Moreover, a consistent estimate of $\hat{y}$ can be given as the solution to:

$$\widehat{y} = 1 - \frac{N}{n} + \frac{\widehat{y}}{n} \operatorname{tr} \mathbf{H}\mathbf{H}^{\mathsf{H}} \left(\widehat{y}\mathbf{H}\mathbf{H}^{\mathsf{H}} + \frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} \right)^{-1}$$

Convergence:

$$y^* - \widehat{y} \to 0.$$

Consistent estimator for C_{erg}

A consistent estimator for C_{erg} is given by:

$$\widehat{C}_G = \frac{1}{N}\log\det\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} + \hat{y}\mathbf{H}\mathbf{H}^{\mathsf{H}}\right) - \frac{1}{N}\log\det\left(\frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\right) + \frac{n-N}{N}\left(\log\left(\frac{n\widehat{y}}{n-N}\right)+1\right) - \frac{n}{N}\hat{y}$$

Fluctuations of the estimator: Let

$$\Theta_N = 2\log(ny^*) - \log\left((n-N)\left(n - \mathrm{tr}\left(\mathbf{I}_N + \mathbf{H}\mathbf{H}^{\mathsf{H}}\left(\mathbf{G}\mathbf{G}^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1}\right)^{-2}\right)\right),$$

Then,

$$0 < \liminf \Theta_N \leq \limsup \Theta_N < +\infty$$

and

$$\frac{N}{\Theta_N}\left(\widehat{C}_G - C_{\mathrm{erg}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Simulations

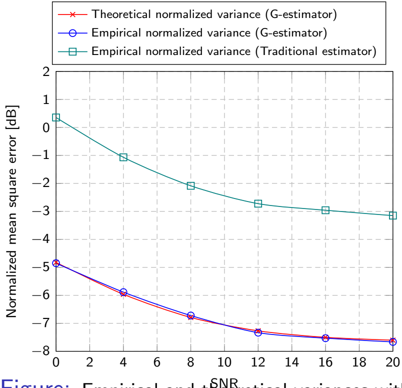


Figure: Empirical and theoretical variances with respect to the SNR

- $N = 4$ %Number of antennas of the receiver
- $n = 15$ %Number of available samples
- $s = 4$ %Number of antennas of the transmitter
- $n = 8$ %Number of interfering mono-antenna users

Simulations

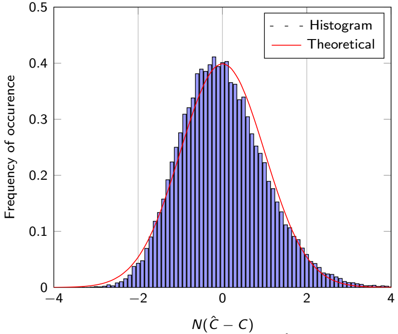


Figure: Histogram of $N(\hat{C} - C)$.

- $N = 4$ %Number of antennas of the interferred receiver
- $n = 15$ %Number of available samples
- $s = 4$ %Number of antennas of the transmitter
- $n = 8$ %Number of the interfering mono-antenna users
- $\mathrm{SNR} = 10\mathrm{dB}$

Performance analysis of robust scatter estimates

- In many signal processing applications, the noise is impulsive:
  - The clutter in radar is in general modeled using elliptical distributions,
  - Interference in wireless communication can be non Gaussian,

  $\Rightarrow$ Outliers, even in low numbers, can considerably degrade the performance of usual estimation and detection techniques,

- Commonly used model for impulsive noises, Elliptical distribution:

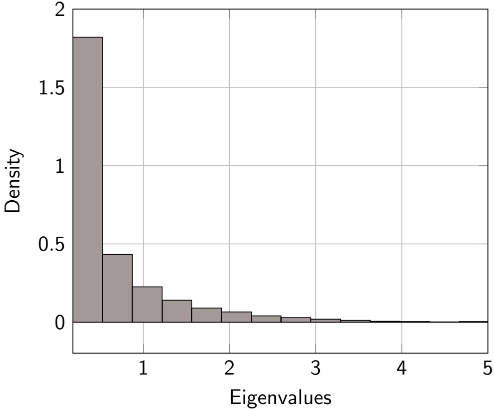
$$\mathbf{x}_i = \sqrt{\tau_i}\mathbf{z}_i \quad \text{où} \quad \begin{cases} \tau_i & \tau_i \text{ heavy tailed.} \\ \mathbf{z}_i & \text{Gaussienne.} \end{cases}$$

Sample covariance matrix under elliptical noises

- Remember that the sample covariance matrix is given by:

$$\widehat{\boldsymbol{\Sigma}}_N = \frac{1}{n}\sum_{i=1}^{n} \tau_i \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$$

- The support of the distribution of τ_i being non-compact, The eigenvalue spectrum support will be also unbounded.



Robust scatter matrix estimator

- Let $\mathbf{y}_1, \cdots, \mathbf{y}_n$, be n observations of size N and covariance $\mathbf{\Sigma}$. The robust scatter matrix estimator is solution of the following equation:

$$\widehat{\mathbf{C}}_N = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i^{\mathrm{H}} u\left(\frac{1}{N}\mathbf{y}_i^{\mathrm{H}} \widehat{\mathbf{C}}_N^{-1} \mathbf{y}_i\right),$$

  where u is a decreasing function such that $\phi = xu(x)$ is increasing.
- Such techniques can be traced back to the works of Huber (1964)
- They have been studied under the regime $N \ll n$.
- Very few results exists in the regime of large random matrices $N \propto n$ (Donoho, ElKaroui, 2013, not yet published).

Performance analysis of robust scatter estimates

- Recent works about the behaviour of the robust scatter methods concern the case where the received observations correspond to a purely noise signal. [Couillet, JMVA]

$$\mathbf{y}_i = \sqrt{\tau_i}\Sigma^{\frac{1}{2}}\mathbf{z}_i, i = 1, \cdots, n$$

- Result:

Theorem

Let $g : \mathbb{R}^+ \to \mathbb{R}^+, x \mapsto \frac{x}{1-c_N\phi(x)}$ and $v : x \mapsto \left(u \circ g^{-1}\right)(x)$, $\psi : x \mapsto xv(x)$. Let γ be the unique positive solution of the equation in γ:

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

Then:

$$\|\widehat{\mathbf{C}}_N - \widehat{\mathbf{S}}_N\| \xrightarrow{a.s} 0$$

where

$$\widehat{\mathbf{S}}_N = \frac{1}{n}\sum_{i=1}^{n} v(\tau_i\gamma_N)\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}}$$

Performance analysis of robust scatter estimates

- Consequence of this result:

$$\widehat{\mathbf{S}}_N = \frac{1}{n}\sum_{i=1}^{n} \tau_i v(\tau_i \gamma_N) \mathbf{z}_i \mathbf{z}_i^{\mathrm{H}}$$

- Function $v(\tau_i \gamma_N) = \frac{\Psi(\tau_i \gamma_N)}{\tau_i \gamma_N} \leq \frac{\Psi_\infty}{\tau_i \gamma_N}$

$$\widehat{\mathbf{S}}_N = \frac{1}{n}\sum_{i=1}^{n} \frac{\psi(\tau_i \gamma_N)}{\gamma_N} \mathbf{z}_i \mathbf{z}_i^{\mathrm{H}}$$

$\Rightarrow$ The support of the eigenvalue spectrum of $\widehat{\mathbf{S}}_N$ is compact.

Elements of the proof

- Define $\hat{\mathbf{C}}_{N,i}$ as:

$$\hat{\mathbf{C}}_{N,i} = \frac{1}{n}\sum_{j=1,i\neq j}^{n} \mathbf{y}_j\mathbf{y}_j^* u\left(\frac{1}{N}\mathbf{y}_j^{\mathrm{H}}\widehat{\mathbf{C}}_N^{-1}\mathbf{y}_j\right)$$

- Using standard matrix inversion lemma, we can prove that:

$$\frac{1}{N}\mathbf{y}_i^{\mathrm{H}}\widehat{\mathbf{C}}_N^{-1}\mathbf{y}_i = g^{-1}\left(\frac{1}{N}\mathbf{y}_i^{\mathrm{H}}\widehat{\mathbf{C}}_{N,i}^{-1}\mathbf{y}_i\right)$$

$\Rightarrow$

$$\hat{\mathbf{C}}_{N,i} = \frac{1}{n}\sum_{j=1,i\neq j}^{n} \mathbf{y}_j\mathbf{y}_j^* v\left(\frac{1}{N}\mathbf{y}_j^{\mathrm{H}}\widehat{\mathbf{C}}_{N,i}^{-1}\mathbf{y}_j\right)$$

- The dependence of $\hat{C}_{N,i}$ on $\mathbf{y}_j$ for $j \neq i$ is low. Therefore,

$$\frac{1}{N}\mathbf{y}_j^{\mathrm{H}}\widehat{\mathbf{C}}_{N,i}^{-1}\mathbf{y}_j \sim \frac{1}{N}\operatorname{tr}\boldsymbol{\Sigma}\widehat{\mathbf{C}}_{N,i}^{-1} \sim \tau_j\gamma_N$$

$\Rightarrow$ In the asymptotic regime, γ_N

$$\gamma_N \sim \frac{1}{N}\operatorname{tr}\left(\frac{1}{n}\sum_{j=1,j\neq i}^{n}\mathbf{y}_j\mathbf{y}_j^{\mathrm{H}} v(\tau_j\gamma_N)\right)^{-1}$$

Generalization to the signal plus noise

- System model: $\mathbf{y}_i = \mathbf{A}_N \mathbf{s}_i + \mathbf{x}_i$ where $\mathbf{x}_i = \sqrt{\tau_i}\mathbf{z}_i$.
- Result:

Convergence of the robust scatter estimator (work in preparation)

Let $\mathbf{B}_N = \mathbf{A}_N \mathbf{A}_N^*$. Define $\delta_1, \cdots, \delta_N$ the unique solutions to the following system of equations:

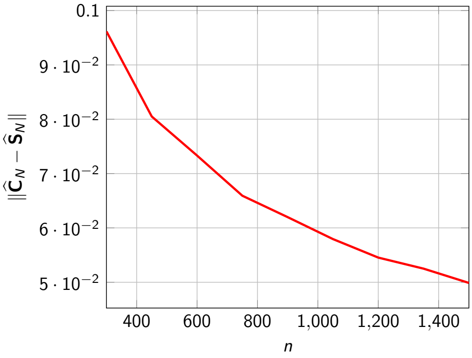
$$\delta_i = \frac{1}{N} \operatorname{tr}(\mathbf{B}_N + \tau_i \mathbf{I}_N) \left(\frac{1}{n} \sum_{j=1}^{n} \frac{v(\delta_j)(\mathbf{B}_N + \tau_j \mathbf{I}_N)}{1 + c_N \Psi(\delta_j)} \right)^{-1}$$

Then,

$$\|\widehat{\mathbf{C}}_N - \widehat{\mathbf{S}}_N\| \xrightarrow{a.s.} 0$$

where $\widehat{\mathbf{S}}_N = \frac{1}{n} \sum_{i=1}^{n} v(\delta_i) \mathbf{y}_i \mathbf{y}_i^{\mathsf{H}}$. Moreover, the support of the eigenvalue spectrum of $\widehat{\mathbf{S}}_N$ is compact.

Numerical results



- $t = 1$,
- $u(x) = \frac{1+t}{t+x}$
- $\frac{N}{n} = \frac{1}{3}$,

Perspectives

- In case the number of observations is limited, estimates of the covariance matrix are ill-conditioned.
  $\Rightarrow$ Study of the asymptotic behaviour of the regularized robust scatter matrix:

$$\widehat{\mathbf{C}}_N = \frac{1-\rho}{n}\sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i^{\mathrm{H}} u\left(\frac{1}{N}\mathbf{y}_i^{\mathrm{H}}\widehat{\mathbf{C}}_N^{-1}\mathbf{y}_i\right) + \rho \mathbf{I}_N$$

- Study of the behaviour of detection methods under elliptical noises, using the results on the information plus noise model

$\Rightarrow$ Fluctuations of the eigenvalues of robust scatter estimates of the covariance matrix,

- Study the minimum square error of robust estimation methods,

$\Rightarrow$ What is the best function u which achieves the minimum squared error ?