Large Random Matrices and High Dimensional Statistical Signal Processing

Summer School, Telecom ParisTech

Abla KAMMOUN[1]

King's Abdullah University of Technology and Science, Saudi Arabia

7-8 June 2016

Featured Applications of Random Matrix Theory (RMT)

Random matrix theory

Robust statistics

Since 2012
- Detection in impulsive noises

Machine learning

Since 2015
- Subspace clustering

Wireless Communication

Since 1990
- Performance analysis
- Optimal transceiver design

Signal Processing

Since 2007
- Estimation
- Detection

Outline

Outline

Motivation

- In many applications, the Gaussian distribution is not a good model to describe the underlying physics
  - Radar clutter,
  - Interference in indoor and outdoor mobile communication channels,
  - noise in imaging problems.

- Some observations, referred to as outliers, might present an atypical behaviour.
- A few number of outliers can impact severely the performances of traditional signal processing methods,

$\Longrightarrow$ It is of fundamental interest to develop robust signal processing methods

Outline

Complex elliptically symmetric (CES) distributions

- Form a natural extension of the complex normal distribution by allowing heavier or lighter tails than the complex normal distribution
- Many results for the complex normal distribution carry over to this broader class.
- Present better accuracy in modeling impulsive noises
- Are tractable and thus can be used to derive robust estimates from the maximum likelihood principle.

Complex elliptically symmetric (CES) distributions

Stochastic representation

- The basic block for constructing a random variable following a CES distribution is the standard normal distribution.

$$
\begin{tikzcd}
\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_p) \arrow[dr] & \perp\!\!\!\perp & \tau \text{ (scalar)} \arrow[dl] \\
& \mathbf{x} = \sqrt{\tau} \times \mathbf{C}_p^{\frac{1}{2}} \times \frac{\mathbf{z}}{\|\mathbf{z}\|} &
\end{tikzcd}
$$

- τ is a scalar used to model the impulsive behaviour of the noise,
- $\frac{\mathbf{z}}{\|\mathbf{z}\|}$ has a uniform distribution over the sphere,

$\Longrightarrow$ This representation is not unique

$$\mathbf{x} = \sqrt{\frac{\tau}{c}} \times (c\mathbf{C}_p)^{\frac{1}{2}} \times \frac{\mathbf{z}}{\|\mathbf{z}\|}$$

- Matrix $\mathbf{C}_p$ is called the scatter matrix. To remove scalar ambiguity, the scatter matrix is selected such that $\operatorname{tr} \mathbf{C}_p = p$.

Complex elliptically symmetric (CES) distributions

Definition

A continuous symmetric random vector $\mathbf{x} \in \mathbb{C}^p$ has a centered complex elliptically symmetric (CES) distribution if its p.d.f is of the form:

$$\boxed{f(\mathbf{x}) = C_{p,g}\left(\det(\mathbf{C}_p)^{-1}\right) g(\mathbf{x}^H \mathbf{C}_p^{-1} \mathbf{x})} \qquad (*)$$

where $\mathbf{C}_p$ is positive definite hermitian matrix, called the scatter matrix, $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ called the density generator and

$$C_{p,g} = 2(s_p \delta_{p,g})^{-1}$$

with $\delta_{p,g} = \int_0^\infty t^{p-1} g(t) dt$ and $s_p = 2\pi^p / \Gamma(p)$. We write that

$$\boxed{\mathbf{x} \sim CES(\mathbf{C}_p, g)} \qquad (1)$$

Link between two representations of CES random variables

- Consider τ with pdf $f_\tau(t) = t^{p-1} g(t) \delta_{p,g}^{-1}$.

$\Longrightarrow$ Then, $\mathbf{x} = \sqrt{\tau}\mathbf{C}_p^{\frac{1}{2}} \frac{\mathbf{z}}{\|\mathbf{z}\|}$ is a CES with pdf $f(x)$ given in (*).

Examples of CES distributions

- Gaussian distribution:

$$\mathbf{x} = \mathrm{Gam}(p, 1) \;\times\; \mathbf{C}_p^{\frac{1}{2}} \frac{\mathbf{z}}{\|\mathbf{z}\|}$$

- Compound Gaussian distribution: sub-family of CES distributions

$$\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_p) \quad \perp\!\!\!\perp \quad \tau \text{ (scalar)}$$

$$\boxed{\mathbf{x} = \sqrt{\tau} \times \mathbf{C}_p^{\frac{1}{2}} \times \mathbf{z}}$$

Stochastic compound Gaussian representation $\mathbf{x} = \sqrt{\tau}\, \mathbf{C}_p^{\frac{1}{2}} \times \mathbf{z} = \sqrt{\tau \|\mathbf{z}\|^2} \times \mathbf{C}_p^{\frac{1}{2}} \frac{\mathbf{z}}{\|\mathbf{z}\|}$

- $\tau > 0$ is called the texture and is independent of $\mathbf{z}$,
- $\mathbf{z}$ is called the speckle.

Compound Gaussian distribution

- Probability density function of Compound Gaussian distribution. If $\text{rank}(\mathbf{C}_N) = p$, the density exists and is given by:

$$f(\mathbf{x}) = \pi^{-p}\det(\mathbf{C}_N^{-1}) \int_0^\infty \tau^{-p} \exp\left(-\mathbf{x}^H \mathbf{C}_N^{-1} \mathbf{x}/\tau\right) f_\tau(\tau) d\tau$$

$\Longrightarrow$ The density generator associated with $\mathbf{x}$ is given by:

$$g(t) = \int_0^\infty \tau^{-p} \exp(-\frac{t}{\tau}) f_\tau(\tau) d\tau$$

- Most illustrative examples:

Distribution	Pdf	heavier tail
K-distribution	$\mathbf{x} = \sqrt{\text{Gam}(\nu, \nu^{-1})} \times \mathbf{z}$	Tail heavy when $\nu \nearrow$
Complex t-distribution	$\mathbf{x} = \sqrt{\frac{1}{\text{Gam}(\nu/2, 2\nu^{-1})}} \times \mathbf{z}$	Tail heavy when $\nu \searrow$
Generalized Gaussian distribution	$\mathbf{x} = \text{Gam}(\frac{p}{s}, b) \times \mathbf{z}$	Tail heavy when $s \searrow$

Outline

Maximum Likelihood estimator (MLE) for the scatter

- Problem statement: Given p-dimensional random vectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$ drawn from elliptical distribution with density generator g and scatter matrix $\mathbf{C}_p$.

Objective: estimate the scatter matrix $\mathbf{C}_p$

- Traditional estimator is the sample covariance matrix:

$$\widehat{\mathbf{S}}_p = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H$$

Its popularity owes to:
  - Simplicity,
  - Existing of a good understanding of its behaviour in the regimes : $\begin{cases} n \to \infty \text{ and } p \text{ fixed} \\ n, p \to \infty \text{ with } \frac{p}{n} \to c. \end{cases}$
  - Is the MLE estimator when observations are drawn from Gaussian distribution

The traditional estimator is not MLE when observations are not Gaussian.

Maximum Likelihood estimator (MLE) for the scatter

- Given $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^p$
  - independent and identically distributed
  - centered
  - $\mathbf{x}_i \sim CES(\mathbf{C}_p, g)$ with pdf $f(x)$.
- The Likelihood is given by:

$$L(\mathbf{C}_p) = \prod_{i=1}^{n} f(\mathbf{x}_i) \propto \left(\det(\mathbf{C}_p^{-1})\right)^n \prod_{i=1}^{n} g(\mathbf{x}_i^H \mathbf{C}_p^{-1} \mathbf{x}_i)$$

$\Longrightarrow$ The negative log-likelihood function

$$\boxed{\mathcal{L}_n(\mathbf{C}_p) \propto \sum_{i=1}^{n} -\log g\left(\mathbf{x}_i^H \mathbf{C}_p^{-1} \mathbf{x}_i\right) + n \log \det(\mathbf{C}_p)}$$

- The MLE estimator is given by:

$$\widehat{\mathbf{C}}_n = \underset{\substack{\mathbf{C}_p \in \mathbb{C}^p \\ \text{definite positive}}}{\operatorname{argmin}} \quad \mathcal{L}_n(\mathbf{C}_p)$$

$\Longrightarrow$

$$\boxed{\widehat{\mathbf{C}}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi\left(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i\right) \mathbf{x}_i \mathbf{x}_i^H, \qquad \varphi = -g'(t)/g(t)}$$

Maximum Likelihood estimator (MLE) for the scatter

- Gaussian case
  - $\mathbf{x}_1, \cdots, \mathbf{x}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_p) \implies \mathbf{x}_1, \cdots, \mathbf{x}_n \sim CES(\mathbf{C}_p, \exp(-t))$
  - Density generator is $g(t) = \exp(-t) \implies \varphi(t) = \exp(-t)/\exp(-t) = 1.$

$$\implies \widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^H$$

- CES case:

$$\boxed{\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} \varphi\left(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H, \qquad \varphi = -g'(t)/g(t)}$$

Natural questions:

- Existence,
- Uniqueness,
- Numerical evaluation
- Asymptotic behaviour

Examples of Maximum likelihood estimators

- Complex t-distribution $\mathbb{C}t_{p,\nu}$

$$\boxed{\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} \varphi\left(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H, \qquad \varphi = \frac{2p+\nu}{\nu+2t}}$$

- Generalized Gaussian distribution $\mathbb{C}GG_{p,s}$

$$\boxed{\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} \varphi\left(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H, \qquad \varphi = \frac{s}{b}t^{s-1}}$$

- Central Angular Gaussian distribution $\mathbb{C}AG_p(\mathbf{C}_N)$

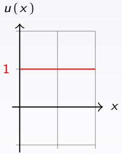
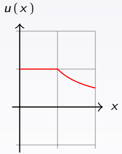
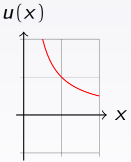
$$\boxed{\begin{aligned}\widehat{\mathbf{C}}_n &= \frac{1}{n}\sum_{i=1}^{n} \varphi\left(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H, \qquad \varphi(t) = \frac{p}{t}\\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i}\end{aligned}}$$

Outline

M-estimator of scatter matrix

- M-estimators of scatter are generalizations of the ML-estimators of the scatter matrix.
- They can be defined by allowing a general function u in place of φ not necessarily related to any elliptical density g,
- Motivation behind the M-estimator: the probability density function might not be known.

$$\boxed{\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u\left(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H}$$

SCM	Huber	Tyler
$u(x) = 1$	$u(x) = \begin{cases} K/e & \text{If} x \leqslant e \\ K/x & \text{If} x \geqslant e \end{cases}$	$u(x) = \frac{p}{x}$
		

M-estimators of scatter matrix

- Real-valued case: 1976-1991
  - Provide necessary and sufficient conditions for the existence and uniqueness of M-estimators
  - Extensively studied under the regime $n \to \infty$ and p fixed.
    - R. A. Maronna, "Robust M-estimators of multivariate location and scatter", 1976.
    - P.J. Huber, "Robust Statistics", 1981.
    - D. E. Tyler. " A Distribution-Free M-estimator of Multivariate Scatter", The Annals of Statistics, 1987
    - J. T. Kent, D. E. Tyler, "Redescending M-estimates of multivariate location and scatter", 1991.
    - J. T. Kent, D. E. Tyler, "Maximum likelihood estimation for the wrapped Cauchy distribution", 1988
    - D. E. Tyler, "Radial estimates and the test for sphericity, Biometrika 1982
- Generalized to the complex-case mainly in the engineering literature. 2003-Present
  - Provide necessary and sufficient conditions for the existence and uniqueness of M-estimators
  - Studied under the regime $n \to \infty$ and p fixed.
    - E. Ollila and V. Koivunen "Robust antenna array processing using M-estimators of pseudo-covariance", 2003
    - E. Ollila and V. Koivunen "Influence function and asymptotic efficiency of scatter matrix based array processors: Case MVDR beamformer", 2009
    - F. Pascal, P. Forster, J.-P. Ovarlez and P. Larzabal "Performance analysis of covariance matrix estimates in impulsive noise", 2008
    - M. Mahot, F. Pascal, P. Forster and J.-P. Ovarlez "Asymptotic properties of robust complex covariance matrix estimates" 2013
    - Y. Chitour, R. Couillet and F. Pascal "On the convergence of Maronna's M-estimators of scatter", 2015
- Studied under the regime $n, p \to \infty$ 2013-Present
  - R. Couillet, F. Pascal and J. W. Silverstein "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples" Journal of Multivariate Analysis, 2013

M-estimators of scatter matrix: Existence and uniqueness

- **Assumptions**
  - Let $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^p$
  - $x \mapsto u(x)$ is non-negative, continuous and non-increasing,
  - Consider $x \mapsto \phi(x) = xu(x)$ strictly increasing.
  - Let $K = \sup_{s \geqslant 0} \phi(s)$, then $p < K$.
  - There exists $a > 0$ such that for any hyperplane S satisfying $\dim(S) \leqslant p - 1$, we have $\frac{\#(\mathbf{x}_i \in S)}{n} \leqslant 1 - \frac{p}{K} - a$.
- **M-estimator of scatter:** Then, the solution of the following equation in $\widehat{\mathbf{C}}_n$

$$\widehat{\mathbf{C}}_n = \frac{1}{n} \sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^H$$

  exists and is unique.
- **Numerical evaluation:** Moreover, given any initial estimate $\boldsymbol{\Sigma}$ hermitian and positive definite, the following sequence $(\boldsymbol{\Sigma}_k)$ defined as:

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}$$

$$\boldsymbol{\Sigma}_{k+1} = \frac{1}{n} \sum_{i=1}^{n} u(\mathbf{x}_i^H \boldsymbol{\Sigma}_k^{-1} \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^H$$

  converges to $\widehat{\mathbf{C}}_n$.

M-estimators of scatter matrix: Asymptotic convergence

E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor "Complex elliptically symmetric distributions: Survey, new results and applications" IEEE Transactions on Signal Processing, 2012

- **Assumptions**
  - Let $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^p \sim CES(\mathbf{C}_p, g)$
  - $x \mapsto u(x)$ is non-negative, continuous and non-increasing,
  - Consider $x \mapsto \phi(x) = xu(x)$ strictly increasing.
  - Let $K = \sup_{s \geqslant 0} \phi(s)$, then $p < K$.
  - There exists $a > 0$ such that for any hyperplane S satisfying $\dim(S) \leqslant p - 1$, we have $\frac{\#(\mathbf{x}_i \in S)}{n} \leqslant 1 - \frac{p}{K} - a$.
  - Asymptotic regime: $n \to \infty$ with p fixed.
- **Convergence of the M-estimator of scatter:** Let $\mathbf{C}_\phi$ be the solution of the following equation:

$$\mathbf{C}_\phi = \mathbb{E}\left[u(\mathbf{x}^H \mathbf{C}_\phi^{-1} \mathbf{x}) \mathbf{x}\mathbf{x}^H\right]$$

Then,

$$\widehat{\mathbf{C}}_n \to \mathbf{C}_\phi$$

Moreover,

$$\mathbf{C}_\phi = \sigma \mathbf{C}_p$$

with σ being the unique solution to $\mathbb{E}_\tau\left[\phi\left(\frac{\tau}{\sigma}\right)\right] = p$.

M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H$$

M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i)\mathbf{x}_i \mathbf{x}_i^H$$

- Since p is assumed fixed, $\widehat{\mathbf{C}}_n$ can be considered as close to a deterministic matrix $\mathbf{C}_\phi$. By the strong law of large numbers, we expect that:

$$\frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i)\mathbf{x}_i \mathbf{x}_i^H \sim \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \mathbf{C}_\phi^{-1} \mathbf{x}_i)\mathbf{x}_i \mathbf{x}_i^H \sim \mathbb{E}\left[u(\mathbf{x}^H \mathbf{C}_\phi^{-1} \mathbf{x})\mathbf{x}\mathbf{x}^H\right]$$

M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H$$

- Since p is assumed fixed, $\widehat{\mathbf{C}}_n$ can be considered as close to a deterministic matrix $\mathbf{C}_\phi$. By the strong law of large numbers, we expect that:

$$\frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H \sim \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \mathbf{C}_\phi^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H \sim \mathbb{E}\left[u(\mathbf{x}^H \mathbf{C}_\phi^{-1}\mathbf{x})\mathbf{x}\mathbf{x}^H\right]$$

$\Longrightarrow$ Matrix $\mathbf{C}_\phi$ should thus satisfy:

$$\mathbf{C}_\phi = \mathbb{E}\left[u(\mathbf{x}^H \mathbf{C}_\phi^{-1}\mathbf{x})\mathbf{x}\mathbf{x}^H\right]$$

M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H$$

- Since p is assumed fixed, $\widehat{\mathbf{C}}_n$ can be considered as close to a deterministic matrix $\mathbf{C}_\phi$. By the strong law of large numbers, we expect that:

$$\frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H \sim \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \mathbf{C}_\phi^{-1} \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H \sim \mathbb{E}\left[u(\mathbf{x}^H \mathbf{C}_\phi^{-1} \mathbf{x})\mathbf{x}\mathbf{x}^H\right]$$

$\Longrightarrow$ Matrix $\mathbf{C}_\phi$ should thus satisfy:

$$\mathbf{C}_\phi = \mathbb{E}\left[u(\mathbf{x}^H \mathbf{C}_\phi^{-1} \mathbf{x})\mathbf{x}\mathbf{x}^H\right]$$

- Now, multiplying both sides by $\mathbf{C}_\phi^{-1}$, we obtain:

$$\mathbf{I}_p = \mathbb{E}\left[u(\mathbf{x}^H \mathbf{C}_\phi^{-1} \mathbf{x})\mathbf{x}\mathbf{x}^H \mathbf{C}_\phi^{-1}\right]$$

M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H$$

- Since p is assumed fixed, $\widehat{\mathbf{C}}_n$ can be considered as close to a deterministic matrix $\mathbf{C}_\phi$. By the strong law of large numbers, we expect that:

$$\frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H \sim \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \mathbf{C}_\phi^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H \sim \mathbb{E}\left[u(\mathbf{x}^H\mathbf{C}_\phi^{-1}\mathbf{x})\mathbf{x}\mathbf{x}^H\right]$$

$\Longrightarrow$ Matrix $\mathbf{C}_\phi$ should thus satisfy:

$$\mathbf{C}_\phi = \mathbb{E}\left[u(\mathbf{x}^H\mathbf{C}_\phi^{-1}\mathbf{x})\mathbf{x}\mathbf{x}^H\right]$$

- Now, multiplying both sides by $\mathbf{C}_\phi^{-1}$, we obtain:

$$\mathbf{I}_p = \mathbb{E}\left[u(\mathbf{x}^H\mathbf{C}_\phi^{-1}\mathbf{x})\mathbf{x}\mathbf{x}^H\mathbf{C}_\phi^{-1}\right]$$

- Taking the trace and replacing $\mathbf{C}_\phi$ by $\sigma\mathbf{C}_p$, we finally obtain:

$$p = \mathbb{E}\left[u(\mathbf{x}^H\mathbf{C}_\phi^{-1}\mathbf{x})\mathbf{x}^H\mathbf{C}_\phi^{-1}\mathbf{x}\right] = \mathbb{E}\left[\phi(\mathbf{x}^H\mathbf{C}_\phi^{-1}\mathbf{x})\right] = \mathbb{E}\left[\phi(\frac{\tau}{\sigma})\right]$$

M-estimators of scatter matrix: Fluctuations

M. Mahot, F. Pascal, P. Forster and J.-P, Ovarlez " Asymptotic Properties of Robust Complex Covariance Matrix Estimates" IEEE Transactions on Signal Processing, vol. 61, no 13, 2013.

- **Assumptions**
  - Let $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^p \sim CES(\mathbf{C}_N, g)$
  - $x \mapsto u(x)$ is non-negative, continuous and non-increasing,
  - Consider $x \mapsto \phi(x) = xu(x)$ strictly increasing.
  - Let $K = \sup_{s \geqslant 0} \phi(s)$, then $p < K$.
  - There exists $a > 0$ such that for any hyperplane S satisfying $\dim(S) \leqslant p - 1$, we have $\frac{\#(\mathbf{x}_i \in S)}{n} \leqslant 1 - \frac{p}{K} - a$.
  - Asymptotic regime: $n \to \infty$ with p fixed.

Then,

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{C}}_n - \mathbf{C}_\phi\right) \xrightarrow{d} \mathbb{CN}_{p^2}(\mathbf{0}, \mathbf{C}, \mathbf{P})$$

where the asymptotic covariance and pseudo-covariance matrices are given by:

$$\mathbf{C} = \vartheta_1 \left(\mathbf{C}_p^* \otimes \mathbf{C}_p\right) + \vartheta_2 \mathrm{vec}(\mathbf{C}_p)\mathrm{vec}(\mathbf{C}_p)^H$$
$$\mathbf{P} = \vartheta_1 \left(\mathbf{C}_p^* \otimes \mathbf{C}_p\right) \mathbf{K}_{p,p} + \vartheta_2 \mathrm{vec}(\mathbf{C}_p)\mathrm{vec}(\mathbf{C}_p)^T$$

with $\mathbf{K}$ being the commutation matrix. The constants ϑ_1 and ϑ_2 depends solely on the distribution of τ and function u.

Tyler estimators

D. E. Tyler " A distribution-free M-estimator of multivariate scatter" The Annals of statistics 1987
F. Pascal, P. Forster, J-P. Ovarlez and P. Larzabal "Performance Analysis of Covariance Matrix Estimates in Impulsive Noise" IEEE Transactions on Signal Processing 2008

- **Assumptions**
  - $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^{p\times 1}$ such that $n > p$.
  - $\mathrm{span}\left(\{\mathbf{x}_i\}\right) = \mathbb{C}^p$.
  - Tyler estimator is defined by $\widehat{\mathbf{C}}_n$ the solution of

$$\widehat{\mathbf{C}}_n = \sum_{i=1}^{n} \frac{p}{n} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i}$$

    such that $\frac{1}{p} \mathrm{tr}\, \widehat{\mathbf{C}}_n = 1$.

  where $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are independent and follow CES distribution with scatter $\mathbf{C}_p$.
- If $\mathbf{x}_1, \cdots \mathbf{x}_n$ are independent and follow CES distribution, then tyler estimator is the MLE of scatter of the random vectors $\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}, \cdots \frac{\mathbf{x}_n}{\|\mathbf{x}_n\|}$

Tyler estimators

F. Pascal, P. Forster, J-P. Ovarlez and P. Larzabal "Performance Analysis of Covariance Matrix Estimates in Impulsive Noise" IEEE Transactions on Signal Processing 2008

- **Assumptions**
  - $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^{p\times 1}$ independent and follow CES distribution with scatter $\mathbf{C}_p$ $(\frac{1}{p}\operatorname{tr}\mathbf{C}_p = 1)$
  - $n > p$.
  - $\operatorname{span}(\{\mathbf{x}_i\}) = \mathbb{C}^p$.
- Tyler estimator is consistent:

$$\widehat{\mathbf{C}}_n \xrightarrow[n\to\infty, p\text{fixed}]{a.s.} \mathbf{C}_p$$

- Fluctuations of Tyler estimators:

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{C}}_n - \mathbf{C}_p\right) \xrightarrow[n\to\infty, p \text{ fixed}]{d} \mathcal{CN}(\mathbf{0}_{p^2}, \mathbf{C}, \mathbf{P})$$

where

$$\mathbf{C} = \vartheta_1 \mathbf{C}_p^T \otimes \mathbf{C}_p + \vartheta_2 \mathrm{vec}(\mathbf{C}_p)\mathrm{vec}(\mathbf{C}_p)^H$$

$$\mathbf{P} = \vartheta_1 \left(\mathbf{C}_p^T \otimes \mathbf{C}_p\right)\mathbf{K}_{p,p} + \vartheta_2 \mathrm{vec}(\mathbf{C}_p)\mathrm{vec}(\mathbf{C}_p)^T$$

with $\mathbf{K}_p$ is the commutation matrix, $\vartheta_1 = \frac{p+1}{p}$ and $\vartheta_2 = -\frac{p+1}{p^2}$.

Comparison with fluctuations of SCM: Some comments

- Fluctuations of the sample covariance matrix (SCM) when $\mathbf{x}_1, \cdots, \mathbf{x}_n$ follow Gaussian distribution with zero mean and covariance $\mathbf{C}_p$:

$$\widehat{\mathbf{S}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H$$

Then:

$$\sqrt{n} vec\left(\widehat{\mathbf{S}}_n - \mathbf{C}_p\right) \xrightarrow{d} \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathrm{scm}}, \mathbf{P}_{\mathrm{scm}})$$

where

$$\mathbf{C}_{\mathrm{scm}} = \left(\mathbf{C}_p^* \otimes \mathbf{C}_p\right) \qquad \mathbf{P}_{\mathrm{scm}} = \left(\mathbf{C}_p^* \otimes \mathbf{C}_p\right)\mathbf{K}_{p,p}$$

- Recall, the fluctuations of the M-estimator of the scatter matrix are given by:

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{C}}_n - \mathbf{C}_\phi\right) \xrightarrow{d} \mathbb{C}\mathcal{N}_{p^2}(\mathbf{0}, \mathbf{C}, \mathbf{P})$$

where the asymptotic covariance and pseudo-covariance matrices are given by:

$$\mathbf{C} = \vartheta_1 \left(\mathbf{C}_p^* \otimes \mathbf{C}_p\right) + \vartheta_2 \mathrm{vec}(\mathbf{C}_p)\mathrm{vec}(\mathbf{C}_p)^H$$
$$\mathbf{P} = \vartheta_1 \left(\mathbf{C}_p^* \otimes \mathbf{C}_p\right)\mathbf{K}_{p,p} + \vartheta_2 \mathrm{vec}(\mathbf{C}_p)\mathrm{vec}(\mathbf{C}_p)^T$$

Comparison with fluctuations of SCM: Some comments

- Let H be any continuous map defined on the set of Hermitian positive definite matrix. Assume that $H(\mathbf{V}) = H(\alpha\mathbf{V})$ for all $\alpha > 0$. By the Delta method and using the fact $H'(\mathbf{C}_p)$ is orthogonal to $\mathrm{vec}(\mathbf{C}_p)$, we get:

$$\sqrt{n}\left(H(\widehat{\mathbf{C}}_n) - H(\mathbf{C}_\phi)\right) \xrightarrow{d} \mathbb{C}\mathcal{N}(\mathbf{0}, \alpha_M, \beta_M)$$

where
$$\alpha_M = \vartheta_1 H'(\mathbf{C}_p)^T \left(\mathbf{C}_p^T \otimes \mathbf{C}_p\right) H'(\mathbf{C}_p)$$
$$\beta_M = \vartheta_1 H'(\mathbf{C}_p)^T \left(\mathbf{C}_p^T \otimes \mathbf{C}_p\right) \mathbf{K}_{p,p} H'(\mathbf{C}_p$$

- Applying the Delta method for the SCM, we get:

$$\sqrt{n}\left(H(\widehat{\mathbf{S}}_n) - H(\mathbf{C}_p)\right) \xrightarrow{d} \mathbb{C}\mathcal{N}(\mathbf{0}, \alpha_{\mathrm{scm}}, \beta_{\mathrm{scm}})$$

where
$$\alpha_{\mathrm{scm}} = H'(\mathbf{C}_p)^T \left(\mathbf{C}_p^T \otimes \mathbf{C}_p\right) H'(\mathbf{C}_p)$$
$$\beta_{\mathrm{scm}} = H'(\mathbf{C}_p)^T \left(\mathbf{C}_p^T \otimes \mathbf{C}_p\right) \mathbf{K}_{p,p} H'(\mathbf{C}_p$$

Comparison with fluctuations of SCM: Some comments

- For functions satisfying $H(\mathbf{V}) = H(\alpha \mathbf{V})$,

  $H(\widehat{\mathbf{C}}_n)$ with n observations has the same fluctuations of $H(\mathrm{SCM})$ with $n\vartheta_1$ observations.
- These functions are often encountered in practice:
  - ANMF statistics
  - SINR at the MVDR
  - etc

CLT on the SCM allows us to retrieve CLT on Tyler and M-scatter estimators

Outline

Regularized robust estimators

F. Pascal, Y. Chitour and Y. Quek "Generalized robust shrinkage estimator and its application to STAP detection problem" IEEE Transactions on signal processing 2013,
Y. Chen and A. Wiesel and A. O. Hero "Robust shrinkage estimation of high-dimensional covariance matrices" IEEE Transactions on Signal Processing 2011

Motivation

- M-estimators of scatter are limited to $p < n$, otherwise do not exist,
- They might be not well-conditioned if n is not sufficiently high, or the true covariance matrix has low rank.

$\Rightarrow$ Using M-estimators of scatter can be not desirable in scenarios where we need to compute the inverse.

Regularized robust estimators: Interpretation

E. Ollila and D. E. Tyler "Regularized M-estimators of scatter matrix" IEEE Transactions on signal processing" 2014

- Recall that the ML estimator of scatter should minimize the negative log-likelihood function:

$$\mathcal{L}_n(\mathbf{\Sigma}) = \sum_{i=1}^{n} -\log g(\mathbf{x}_i^H \mathbf{\Sigma}^{-1} \mathbf{x}_i) - n \log \det(\mathbf{\Sigma}^{-1})$$

- To increase the stability of the solution, it is common to introduce $n\rho\mathcal{P}(\mathbf{\Sigma})$ to the cost function. $\Longrightarrow$ The penalized cost function becomes:

$$\mathcal{L}_n(\mathbf{\Sigma}) = \sum_{i=1}^{n} -\log g(\mathbf{x}_i^H \mathbf{\Sigma}^{-1} \mathbf{x}_i) + n \log \det(\mathbf{\Sigma}) + \rho n\mathcal{P}(\mathbf{\Sigma})$$

  where $\rho \geqslant 0$ is a regularization (shrinkage) parameter.
- This procedure can be extended to M-estimators of scatter, by using a general function α ($u = \alpha'$) in place of $-\log g$, leading to:

$$\mathcal{L}_n(\mathbf{\Sigma}) = \sum_{i=1}^{n} \alpha(\mathbf{x}_i^H \mathbf{\Sigma}^{-1} \mathbf{x}_i) - n \log \det(\mathbf{\Sigma}^{-1}) + \rho n\mathcal{P}(\mathbf{\Sigma})$$

Regularized robust estimators: Interpretation

E. Ollila and D. E. Tyler "Regularized M-estimators of scatter matrix" IEEE Transactions on signal processing" 2014

- To ensure stability of $\boldsymbol{\Sigma}^{-1}$, we consider the penalty function:

$$\mathcal{P}(\boldsymbol{\Sigma}) = n \operatorname{tr} \boldsymbol{\Sigma}^{-1}$$

$\Longrightarrow$ The penalized cost function becomes:

$$\mathcal{L}_\rho(\boldsymbol{\Sigma}) = \sum_{i=1}^{n} \alpha(\mathbf{x}_i^H \boldsymbol{\Sigma}^{-1} \mathbf{x}_i) - n \log(\det \boldsymbol{\Sigma}^{-1}) + n\rho \operatorname{tr} \boldsymbol{\Sigma}^{-1} \qquad (*)$$

- A critical point of (*) is a solution to:

$$\widehat{\mathbf{C}}_n = \frac{1}{n} \sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^H + \rho \mathbf{I}_p$$

Regularized robust estimators: Existence and uniqueness

E. Ollila and D. E. Tyler "Regularized M-estimators of scatter matrix" IEEE Transactions on signal processing" 2014

- **Existence and uniqueness**

Theorem

Assume that:

- $t \mapsto \alpha(t)$ *is bounded below (does not tend to* $-\infty$*)*
- *Function* $\alpha(t)$ *is nondecreasing and continuous for* $0 < t < \infty$*. Also, the map* $r(x) = \alpha(e^x)$ *is convex for* $-\infty < x\infty$*.*
- *Function* $\alpha(t)$ *is differentiable*

Then a solution to the following equation

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H + \rho\mathbf{I}_p$$

exists and is unique. Moreover it coincides with the minimum of $\mathcal{L}_\rho(\boldsymbol{\Sigma})$*.*

- The regularized M-estimates do not require any condition on the sample $\mathbf{x}_1, \cdots, \mathbf{x}_n$

$\Longrightarrow$ This is in contrast to the non-regularized estimates which do not exist when $p < n$.

Regularized robust estimators: Numerical evaluation

E. Ollila and D. E. Tyler "Regularized M-estimators of scatter matrix" IEEE Transactions on signal processing" 2014

Theorem

Assume that:

- $t \mapsto \alpha(t)$ *is bounded below (does not tend to* $-\infty$*)*
- *Function* $\alpha(t)$ *is nondecreasing and continuous for* $0 < t < \infty$*. Also, the map* $r(x) = \alpha(e^x)$ *is convex for* $-\infty < x\infty$*.*
- *Function* $\alpha(t)$ *is differentiable*
- *Function* $u = \alpha^{'}$ *is non-increasing*

Then, the iterations

$$\boldsymbol{\Sigma}_{k+1} = \frac{1}{n}\sum_{i=1}^{n} u(\mathbf{x}_i^H \boldsymbol{\Sigma}_k^{-1} \mathbf{x_i})\mathbf{x}_i\mathbf{x}_i^H + \rho \mathbf{I}_p$$

converges to the regularized M-estimate $\widehat{\mathbf{C}}_n$*.*

Regularized Tyler-estimator: Existence

E. Ollila and D. E. Tyler "Regularized M-estimators of scatter matrix" IEEE Transactions on signal processing" 2014
F. Pascal, Y. Chitour and Y. Quek "Generalized robust shrinkage estimator and its application to STAP detection problem" IEEE Transactions on Signal processing, 2014

- Regularized-Tyler estimator obtained by setting $\alpha(x) = (1-\rho)p\log(x)$ or equivalently $u(x) = \frac{(1-\rho)p}{x}$.

$$\widehat{\mathbf{C}}_n(\rho) = (1-\rho)\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_p$$

- Since α is not bounded below, the previous result concerning the existence and uniqueness of the regularized M-estimate does not hold.

Regularized Tyler estimator: Existence

E. Ollila and D. E. Tyler "Regularized M-estimators of scatter matrix" IEEE Transactions on signal processing" 2014
F. Pascal, Y. Chitour and Y. Quek "Generalized robust shrinkage estimator and its application to STAP detection problem" IEEE Transactions on Signal processing, 2014

Theorem

Assume that
Cond. A *:For any subspace S of $\mathbb{C}^p$ such that $1 \leqslant \dim(S) < p$, the inequality $\frac{\#\{\mathbf{x}_i \in S\}}{n} < \frac{\dim(S)}{p(1-\rho)}$ holds. Then:*

$$\widehat{\mathbf{C}}_n(\rho) = (1-\rho)\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_p$$

exists and is unique.

Comments. Assume that $\mathbf{x}_i$ are linearly independent. Then:

$$\frac{\#\{\mathbf{x}_i \in S\}}{n} \leqslant \frac{p-1}{n}$$

Hence, **Cond A** is equivalent to $\rho \geqslant 1 - \frac{n}{p}$.

Regularized Tyler estimator: Numerical evaluation

Theorem

Assume that

Cond. A *:For any subspace S of $\mathbb{C}^p$ such that $1 \leqslant \dim(S) < p$, the inequality $\frac{\#\{\mathbf{x}_i \in S\}}{n} < \frac{\dim(S)}{p(1-\rho)}$ holds. Then, the iterations:*

$$\boldsymbol{\Sigma}_{k+1} = (1-\rho)\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\boldsymbol{\Sigma}_k^{-1}\mathbf{x}_i} + \rho\mathbf{I}_p$$

converges to the regularized Tyler estimator $\widehat{\mathbf{C}}_n(\rho)$.

Normalized regularized Tyler estimator

Y. I. Abramovich and N. K. Spencer "Diagonally loadded normalized sample matrix inversion for outlier-resistant adaptive filtering", ICASSP 2007
Y. Chen, A. Wiesel and A. O. Hero "Robust shrinkage estimation of high-dimensional covariance matrix matrices" IEEE Transactions on signal processing, 2011

Theorem

Let $0 < \rho < 1$ be a regularization coefficient. The following iterations:

$$\widetilde{\boldsymbol{\Sigma}}_{k+1} = (1-\rho)\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\boldsymbol{\Sigma}_k^{-1}\mathbf{x}_i} + \rho\mathbf{I}_p$$

$$\boldsymbol{\Sigma}_{k+1} = \frac{\widetilde{\boldsymbol{\Sigma}}_{k+1}}{tr\widetilde{\boldsymbol{\Sigma}}_{k+1}/p}$$

converge to the a unique limit for any positive definite initial matrix $\widetilde{\boldsymbol{\Sigma}}_0$. Moreover this limit is solution to the following fixed-point equation:

$$\check{\mathbf{C}}_n(\rho) = \frac{\check{\mathbf{B}}_n(\rho)}{\frac{1}{p}tr\check{\mathbf{B}}_n(\rho)}$$

with

$$\check{\mathbf{B}}_n(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H\check{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_p.$$

Regularized Tyler estimator: Convergence

A. Kammoun, R. Couillet, F. Pascal and M.-S. Alouini "Convergence and fluctuations of regularized tyler estimators" IEEE Transactions on Signal Processing 2016

- **Assumptions**: Take $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^{p\times 1}$, i.e, $\mathbf{x}_i = \mathbf{C}_p^{\frac{1}{2}}\mathbf{w}_i$ with:
  - $\mathbf{w}_1, \cdots, \mathbf{w}_p \in \mathbb{C}^{p\times 1}$ independent complex standard Gaussian vectors with zero mean $\mathbf{0}_{p\times 1}$ and covariance $\mathbf{I}_p$,
  - We assume that $\frac{1}{p}\operatorname{tr}\mathbf{C}_p = 1$ and that $\mathbf{C}_p$ is positive definite.
- **Regularized Tyler estimator**: Let $\rho \in \left(0, \max(1-\frac{n}{p}, 1)\right)$. Consider $\widehat{\mathbf{C}}_n(\rho)$ solution to the following fixed point equation:

$$\widehat{\mathbf{C}}_n(\rho) = (1-\rho)\frac{p}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_p.$$

Let $\boldsymbol{\Sigma}_0(\rho)$ be the unique solution to the following equation:

$$\boldsymbol{\Sigma}_0(\rho) = p(1-\rho)\mathbb{E}\left[\frac{\mathbf{x}\mathbf{x}^H}{\mathbf{x}^H\boldsymbol{\Sigma}_0^{-1}(\rho)\mathbf{x}}\right] + \rho\mathbf{I}_p$$

Then, for any $\kappa > 0$,

$$\sup_{\rho\in[\kappa,1]}\left\|\widehat{\mathbf{C}}_n(\rho) - \boldsymbol{\Sigma}_0(\rho)\right\| \xrightarrow[n\to\infty, p\text{fixed}]{a.s.} 0.$$

Regularized Tyler estimator: Characterization of $\mathbf{\Sigma}_0(\rho)$

A. Kammoun, R. Couillet, F. Pascal and M.-S. Alouini "Convergence and fluctuations of regularized tyler estimators" IEEE Transactions on Signal Processing 2016

Evaluation of $\mathbf{\Sigma}_0(\rho)$.

- Recall that $\mathbf{\Sigma}_0(\rho)$ satisfies:

$$\mathbf{\Sigma}_0(\rho) = p(1-\rho)\mathbb{E}\left[\frac{\mathbf{x}\mathbf{x}^H}{\mathbf{x}^H\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{x}}\right] + \rho\mathbf{I}_p$$

- Multiplying both sides by $\mathbf{C}_p^{-\frac{1}{2}}$, we obtain:

$$p(1-\rho)\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^H}{\mathbf{w}^H\mathbf{C}_p^{\frac{1}{2}}\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{C}_p^{\frac{1}{2}}\mathbf{w}}\right] + \rho\mathbf{C}_p^{-1} = \mathbf{C}_p^{-\frac{1}{2}}\mathbf{\Sigma}_0(\rho)\mathbf{C}_p^{-\frac{1}{2}}$$

- Let $\mathbf{C}_p^{\frac{1}{2}}\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{C}_p^{\frac{1}{2}} = \mathbf{V}\mathbf{D}\mathbf{V}^H$ the eigenvalue decomposition of $\mathbf{C}_p^{\frac{1}{2}}\mathbf{\Sigma}_0^{-1}(\rho)\mathbf{C}_p^{\frac{1}{2}}$
- Then,

$$p(1-\rho)\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^H}{\mathbf{w}^H\mathbf{D}\mathbf{w}}\right] + \rho\mathbf{V}^H\mathbf{C}_p^{-1}\mathbf{V} = \mathbf{D}^{-1}$$

$\Longrightarrow \mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^H}{\mathbf{w}^H\mathbf{D}\mathbf{w}}\right]$ is diagonal implying that matrix $\mathbf{\Sigma}_0$ and $\mathbf{C}_p$ share the same eigenvector space

Regularized Tyler estimator: Characterization of $\boldsymbol{\Sigma}_0(\rho)$

A. Kammoun, R. Couillet, F. Pascal and M.-S. Alouini "Convergence and fluctuations of regularized tyler estimators" IEEE Transactions on Signal Processing 2016

Lemma

Consider $\mathbf{w} = \left[w_1, \cdots, w_p\right]^T$ *be a standard complex Gaussian vector. Let* $\mathbf{D}$ *be a diagonal matrix with positive diagonal elements* $d_1, \cdots, d_p$*. Then:*

$$\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^H}{\mathbf{w}^H\mathbf{D}\mathbf{w}}\right] = \mathrm{diag}(\alpha_1, \cdots, \alpha_p)$$

with α_i *for* $i = 1, \cdots, p$ *being given by:*

$$\alpha_i = \frac{1}{2^p p}\frac{1}{d_i \prod_{j=1}^p d_j} F_D^{(p)}\left(p, 1, \cdots, \underset{\substack{\uparrow\\ i\text{-th}\\ \text{position}}}{2}, 1, \cdots, 1, p+1, \frac{d_1 - \frac{1}{2}}{d_1}, \cdots, \frac{d_p - \frac{1}{2}}{d_p}\right), \qquad (2)$$

where $F_D^{(p)}$ *is the Lauricella's type D hypergeometric function.*

Regularized Tyler estimator: Convergence

A. Kammoun, R. Couillet, F. Pascal and M.-S. Alouini "Convergence and fluctuations of regularized tyler estimators" IEEE Transactions on Signal Processing 2016

- Recall that:

$$p(1-\rho)\mathbb{E}\left[\frac{\mathbf{w}\mathbf{w}^H}{\mathbf{w}^H\mathbf{D}\mathbf{w}}\right]+\rho\mathbf{V}^H\mathbf{C}_p^{-1}\mathbf{V}=\mathbf{D}^{-1}$$

- Let $\mathbf{D}=\mathrm{diag}(d_1,\cdots,d_p)$. Denote by $\alpha_i(\{d_i\}_{i=1}^p)=\mathbb{E}\left[\frac{|w_i|^2}{\mathbf{w}^H\mathbf{D}\mathbf{w}}\right]$. Then,

$$p(1-\rho)\alpha_i(\{d_i\}_{i=1}^p)+\frac{\rho}{\lambda_i}=\frac{1}{d_i}$$

where $\alpha_i(\{d_i\}_{i=1}^p)$ is computed as previously.

Regularized Tyler estimator: Convergence

A. Kammoun, R. Couillet, F. Pascal and M.-S. Alouini "Convergence and fluctuations of regularized tyler estimators" IEEE Transactions on Signal Processing 2016

Matrix $\mathbf{\Sigma}_0$ can be computed as:

- Let $\mathbf{C}_N = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ be the eigenvalue decomposition of $\mathbf{C}_N$ with $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_p)$ and $\lambda_1 \geqslant \lambda_2 \cdots \geqslant \lambda_p$.
- Start from $d_1^{(0)}, \cdots, d_p^{(0)}$. Compute iteratively:

$$d_i^{(t+1)} = \frac{1}{\frac{\rho}{\lambda_i} + p(1-\rho)\alpha_i(\mathrm{diag}(\mathbf{d}^{(t)}))}$$

  until convergence. Let $d_{1,\infty}^{(0)}, \cdots, d_{p,\infty}^{(0)}$ be the obtained values after applying several iterations.
- Set $s_{i,\infty} = \lambda_i d_{i,\infty}$. Then

$$\mathbf{\Sigma}_0 = \mathbf{V}\mathrm{diag}\left(\left[s_{1,\infty}, \cdots, s_{p,\infty}\right]\right)\mathbf{V}^H$$

Regularized Tyler estimator: Hint on the proof

A. Kammoun, R. Couillet, F. Pascal and M.-S. Alouini "Convergence and fluctuations of regularized tyler estimators" IEEE Transactions on Signal Processing 2016

The proof is based on controlling the random elements:

$$\epsilon_i(\rho) = \frac{\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i - \mathbf{x}_i^H \mathbf{\Sigma}_0^{-1}(\rho)\mathbf{x}_i}{\sqrt{\mathbf{x}_i^H \mathbf{\Sigma}_0^{-1}(\rho)\mathbf{x}_i}\sqrt{\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i}}$$

and showing that they converge to zero almost surely.

- First observe that by the strong law of large numbers,

$$\mathbf{\Sigma}_0(\rho) \simeq p(1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \mathbf{\Sigma}_0^{-1}(\rho)\mathbf{x}_i} + \rho \mathbf{I}_p$$

- After some computations based on the resolvent Lemma, we can prove that:

$$\max_{1\leqslant i\leqslant n} \epsilon_i(\rho)\left(1 - \sqrt{\|\mathbf{I}_p - \rho\mathbf{\Sigma}_0^{-1}\| - o(1)}\right) \leqslant o(1)$$

  where $o(1)$ is a term converging almost surely to zero, thus showing the convergence of $\max_{1\leqslant i\leqslant n} \epsilon_i(\rho)$ to zero.

- Again, using some linear algebra manipulations, we can show that amost surely,

$$\left\|\widehat{\mathbf{C}}_n - \mathbf{\Sigma}_0(\rho)\right\| \leqslant \max_{1\leqslant i\leqslant n} \epsilon_i(\rho) + o(1)$$

  thus showing that $\left\|\widehat{\mathbf{C}}_n - \mathbf{\Sigma}_0(\rho)\right\| \xrightarrow{\text{a.s.}} 0$.

Regularized Tyler estimators: Fluctuations

A. Kammoun, R. Couillet, F. Pascal and M.-S. Alouini "Convergence and fluctuations of regularized tyler estimators" IEEE Transactions on Signal Processing 2016

- Similarly to the M-scatter estimator, we can establish a CLT on $\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{C}}_n(\rho)-\mathbf{\Sigma}_0(\rho)\right)$. We prove that:

$$\sqrt{n}\mathrm{vec}\left(\widehat{\mathbf{C}}_n(\rho)-\mathbf{\Sigma}_0(\rho)\right)\xrightarrow[n\to\infty,p\text{fixed}]{d}\mathbb{C}\mathcal{N}\left(\mathbf{0},\mathbf{M}_1,\mathbf{M}_2\right)$$

with $\mathbf{M}_1$ and $\mathbf{M}_2$ are given by:

$$\begin{aligned}\mathbf{M}_1=&\left(\left(\mathbf{\Sigma}_0^{\frac{1}{2}}\right)^{\mathsf{T}}\otimes\mathbf{\Sigma}_0^{\frac{1}{2}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}})^{-1}\left(\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathsf{T}}\otimes\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)\\&\times\tilde{\mathbf{M}}_1\left(\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathsf{T}}\otimes\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}})^{-1}\left(\left(\mathbf{\Sigma}_0^{\frac{1}{2}}\right)^{\mathsf{T}}\otimes\mathbf{\Sigma}_0^{\frac{1}{2}}\right)\\\mathbf{M}_2=&\left(\left(\mathbf{\Sigma}_0^{\frac{1}{2}}\right)^{\mathsf{T}}\otimes\mathbf{\Sigma}_0^{\frac{1}{2}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}})^{-1}\left(\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathsf{T}}\otimes\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)\\&\times\tilde{\mathbf{M}}_2\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\otimes\left(\mathbf{\Sigma}_0^{-\frac{1}{2}}\right)^{\mathsf{T}}\right)(\mathbf{I}_{N^2}-\tilde{\mathbf{F}}^{\mathsf{T}})^{-1}\left(\mathbf{\Sigma}_0^{\frac{1}{2}}\otimes\left(\mathbf{\Sigma}_0^{\frac{1}{2}}\right)^{\mathsf{T}}\right)\end{aligned}$$

- The formula is quite involved and does not allow to use results based on the SCM.

Regularized Tyler estimators: Open Questions

The same analysis might be employed to study in the regime $n \to \infty$ with p fixed:

- Normalized regularized Tyler estimators:
  $\check{\mathbf{C}}_n(\rho) = \frac{\check{\mathbf{B}}_n(\rho)}{\frac{1}{p}\operatorname{tr}\check{\mathbf{B}}_n(\rho)}$ with $\check{\mathbf{B}}_n(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H\check{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_p$.
- Regularized M-scatter estimators

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^n u(\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^H + \rho\mathbf{I}_p$$

- Regularized Tyler estimators with priori

$$\widehat{\mathbf{C}}_n(\rho) = (1-\rho)\frac{p}{n}\sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i^H}{\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{T}$$

  where $\mathbf{T}$ is a priori matrix.
- Normalized Regularized Tyler estimators with priori
  $\check{\mathbf{C}}_n(\rho) = \frac{\check{\mathbf{B}}_n(\rho)}{\frac{1}{p}\operatorname{tr}\check{\mathbf{B}}_n(\rho)}$ with $\check{\mathbf{B}}_n(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H\check{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\frac{\mathbf{T}}{\operatorname{tr}\check{\mathbf{C}}_n^{-1}\mathbf{T}}$.

Y. Sun, P. Babu and D. P. Palomar "Regularized Tyler's scatter estimator: Existence, uniqueness and algorithms" IEEE Transactions on Signal Processing, October, 2014

Outline

Outline

Outline

Illustrative random models covered by RMT

Let $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ denote the matrix observation. Then, the following models are encountered in practice:

- White space and in time:

$$\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \qquad \mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_p)$$

- Space and time correlation

$$\mathbf{X} = \mathbf{C}_p^{\frac{1}{2}} \mathbf{W} \mathbf{T}^{\frac{1}{2}}$$

- Information-plus-noise model: non-centered observations with /or without correlation

$$\mathbf{X} = \mathbf{C}_p^{\frac{1}{2}} \mathbf{W} \mathbf{T}^{\frac{1}{2}} + \mathbf{A}$$

Key Assumption: n, p grow to ∞ with $\frac{p}{n} \to c$

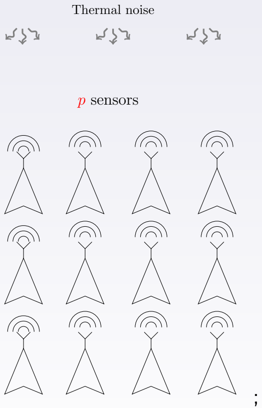
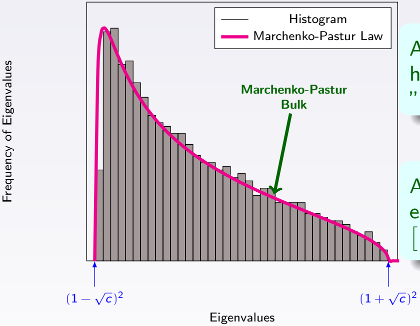


Illustration: Marchenko-Pastur Law

If $\mathbf{C}_p = \mathbf{I}_p$,



As n, p tends to infinity with $\frac{p}{n} \to c$, the histogram can be approximated by a "Deterministic" curve !

As n, p tends to infinity with $\frac{p}{n} \to c$, all the eigenvalues are contained in the interval $\left[(1-\sqrt{c})^2, (1+\sqrt{c})^2\right]$

Spiked random models

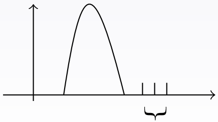
Assume that we have:

$$\mathbf{Y} = \underbrace{\mathbf{X}}_{\substack{\text{High rank random} \\ \text{signal}}} + \beta \underbrace{\mathbf{L}}_{\substack{\text{Signal lying in a finite} \\ \text{dimensional subspace of rank K}}}$$

where $\beta \in \{0, 1\}$. $\Rightarrow$ Information plus noise model where the information lie in a low-rank subspace.

- If $\beta = 0$, $\mathbf{Y} = \mathbf{X}$, the histogram of $\frac{\mathbf{Y}\mathbf{Y}^H}{n}$ is composed of a bulk

- If $\beta = 1$, at most K eigenvalues will appear outside the bulk



Outline

Estimation of covariance matrices

- Problem relevant to several signal processing applications:
  - Estimation of direction of arrival
  - Estimation of the noise power
  - Detection of signals using Information theoretic criteria

- Consider $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^p$, n observations of size p independent and identically distributed of unknown covariance $\mathbf{C}_p$.
- Let $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ the matrix of observations.
- Inference problem: Consider $\theta = f(\mathbf{C}_p)$, where f is a certain functional.

Objective: Estimate parameter θ from the observation matrix $\mathbf{X}$

- Classical Approach to handle this inference problem
  1. Form the sample covariance matrix $\widehat{\mathbf{S}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^H$.
  2. Substitute the unknown covariance matrix $\mathbf{C}_p$ by $\widehat{\mathbf{S}}_n$.

An estimator of θ is given by: $\widehat{\theta} = f(\widehat{\mathbf{S}}_n)$

Estimation of covariance matrices

- As n increases while p taken fixed, then, from the law of large numbers

$$\widehat{\mathbf{S}}_n \xrightarrow{\text{a.s.}} \mathbf{C}_p$$

- The convergence is in operator norm, in that:

$$\|\widehat{\mathbf{S}}_n - \mathbf{C}_p\| \xrightarrow{\text{a.s.}} 0.$$

- As a result, by the continuous mapping theorem:

$$\widehat{\theta} \xrightarrow{\text{a.s.}} \theta.$$

$\Longrightarrow$ $\boxed{\text{The conventional estimator } \widehat{\theta} \text{ is consistent in the regime } n \to \infty \text{ and } p \text{ fixed.}}$

Estimation of covariance matrices

- In practice, it might be the case that $n \propto p$ if not $n < p$.
  - Large antenna arrays vs limited number of observations $\Longrightarrow$ Large dimensional inputs
  - Systems with fast dynamics.
- This situation is modeled by the following assumption: $n \to \infty$ and $p \to \infty$ with $\frac{p}{n} \to c$

Random matrix theory regime: $n, p \to \infty$ with $\frac{p}{n} \to c \in (0, \infty)$

- Under this assumption, we still have by the law of large numbers:

$$\left[\widehat{\mathbf{S}}_n\right]_{i,j} \xrightarrow{\text{a.s.}} \left[\mathbf{C}_p\right]_{i,j}$$

But $\boxed{\left\|\widehat{\mathbf{S}}_n - \mathbf{C}_p\right\| \text{ does not converge to zero.}}$

$\Longrightarrow$ $\boxed{\text{Then, } \widehat{\theta} \text{ is not a consistent estimator of } \theta \text{ in the RMT regime}}$

G-estimation

- Classical signal processing methods.

$$\mathbf{x}_1, \cdots, \mathbf{x}_n \longrightarrow \boxed{\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H} \xrightarrow{\widehat{\mathbf{S}}_N} \boxed{\widehat{\theta} = f(\widehat{\mathbf{S}}_N)}$$

Sample covariance matrix

Estimator of θ

- Improved methods using RMT. In general, we proceed into two steps:
  1. Understand the behaviour of the conventional estimators under the RMT regime

$$\widehat{\theta} - \theta \xrightarrow{\text{a.s.}} \text{Bias}$$

  2. Form a new improved estimator, often referred to as G-estimator that is consistent in the RMT regime:

$$\widehat{\theta}_G - \theta \xrightarrow{\text{a.s.}} 0.$$

$$\mathbf{x}_1, \cdots, \mathbf{x}_n \longrightarrow \boxed{\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H} \xrightarrow{\widehat{\mathbf{S}}_N} \boxed{\widehat{\theta} = g(\widehat{\mathbf{S}}_N)}$$

Sample covariance matrix

Estimator of θ

Outline

Random Matrix regime of M-scatter estimator

Definition

For $\mathbf{x}_1, \cdots, \mathbf{x}_n$ with $n > p$, $\widehat{\mathbf{C}}_n$ is the solution of:

$$\widehat{\mathbf{C}}_n = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{p} \mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i \right) \mathbf{x}_i \mathbf{x}_i^H$$

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

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

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

Note that taking $\tilde{\mathbf{x}}_i = \mathbf{C}_n^{-\frac{1}{2}} \mathbf{x}_i$, and setting $\widetilde{\mathbf{C}}_n \triangleq \mathbf{C}_n^{-\frac{1}{2}} \widehat{\mathbf{C}}_n^{\frac{1}{2}} \widetilde{\mathbf{C}}_n^{-\frac{1}{2}}$,

$$\widetilde{\mathbf{C}}_n = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{p} \mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1} \mathbf{x}_i \right) \mathbf{x}_i \mathbf{x}_i^H$$

$\Longrightarrow$ Without loss of generality, we can assume $\mathbf{C}_p = \mathbf{I}_p$.

Model description

- Assumption on the data (Elliptical model): $\mathbf{x}_1, \cdots, \mathbf{x}_n$ independent

$$\mathbf{x}_i = \sqrt{\tau_i}\mathbf{C}_p^{\frac{1}{2}}\mathbf{w}_i$$

  - $\mathbf{w}_i \in \mathbb{C}^p$, unitarily invariant with $\|\mathbf{w}_i\|^2 = p$,
  - $\mathbf{C}_p \succeq 0$ with $\limsup_N \|\mathbf{C}_p\| < \infty$
  - $\tau_i > 0$ independent of $\mathbf{w}_i$,
  - If $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{\tau_i}$ there exists $m > 0$ such that: $\tilde{\nu}_n([0, m)) < 1 - \phi_\infty^{-1}$ for all large n a.s.
  - $\int \tau\nu_n(d\tau) = 1$

- Assumption on the tail: For each $a > b > 0$, we have:

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

- Random matrix regime: As $n, p \to \infty$, $c_n = \frac{p}{n} \to c \in (0, 1)$.

Main challenges

- Contrary to the classical regime, we do not have that $\hat{\mathbf{C}}_n$ converges to some deterministic limit. In particular:

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H \quad \text{do not converge to} \quad \mathbb{E}\left[u\left(\frac{1}{p}\mathbf{x}^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}\right)\mathbf{x}\mathbf{x}^H\right]}$$

- Major issues with $\hat{\mathbf{C}}_n$
  - No closed-form expression
  - Sum of non-independent rank-one matrices $u\left(\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$
  - No explicit relation between $\hat{\mathbf{C}}_n$ and the random vectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$.

$\Longrightarrow$ Classical random matrix theory tools cannot be directly used.

Heuristic approach

- Rewriting of $\hat{\mathbf{C}}_n$:
  - Denote:

$$\hat{\mathbf{C}}_{(j)} = \frac{1}{n}\sum_{i\neq j}^{n} u\left(\frac{1}{p}\mathbf{x}_i^H \hat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

Heuristic approach

- Rewriting of $\hat{\mathbf{C}}_n$:
  - Denote:

$$\hat{\mathbf{C}}_{(j)} = \frac{1}{n}\sum_{i\neq j}^{n} u\left(\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

  - Using the identity $(\mathbf{A}+t\mathbf{v}\mathbf{v}^H)^{-1} = \mathbf{A}^{-1}/(1+t\mathbf{v}^H\mathbf{A}^{-1}\mathbf{v})$. Then:

$$\tau_i\tilde{d}_i \triangleq \frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i}{1+c_n u\left(\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i}$$

$\Longrightarrow$ $\tau_i d_i \triangleq \frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i = \dfrac{\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}_i}{1-c_n\phi\left(\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)} = g\left(\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)$ with:

$$g(x) = \frac{x}{1-c_n\phi(x)}$$

Heuristic approach

- Rewriting of $\widehat{\mathbf{C}}_n$:
  - Denote:

$$\widehat{\mathbf{C}}_{(j)} = \frac{1}{n}\sum_{i\neq j}^{n} u\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

  - Using the identity $(\mathbf{A}+t\mathbf{v}\mathbf{v}^H)^{-1} = \mathbf{A}^{-1}/(1+t\mathbf{v}^H\mathbf{A}^{-1}\mathbf{v})$. Then:

$$\tau_i\tilde{d}_i \triangleq \frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i}{1+c_n u\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i}$$

$$\Longrightarrow \tau_i d_i \triangleq \tfrac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i}{1-c_n\phi\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)} = g\left(\tfrac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right) \text{ with:}$$

$$g(x) = \frac{x}{1-c_n\phi(x)}$$

  - g is monontous and increasing, then $\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i = g^{-1}\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)$

Heuristic approach

- Rewriting of $\widehat{\mathbf{C}}_n$:
  - Denote:

$$\widehat{\mathbf{C}}_{(j)} = \frac{1}{n}\sum_{i\neq j}^{n} u\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

  - Using the identity $(\mathbf{A}+t\mathbf{v}\mathbf{v}^H)^{-1} = \mathbf{A}^{-1}/(1+t\mathbf{v}^H\mathbf{A}^{-1}\mathbf{v})$. Then:

$$\tau_i\tilde{d}_i \triangleq \frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i}{1+c_n u\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i}$$

$$\Longrightarrow \tau_i d_i \triangleq \tfrac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i = \frac{\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i}{1-c_n\phi\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)} = g\left(\tfrac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right) \text{ with:}$$

$$g(x) = \frac{x}{1-c_n\phi(x)}$$

  - g is monontous and increasing, then $\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i = g^{-1}\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)$
  - Using this equality in the expression of $\widehat{\mathbf{C}}_n$,

$$\widehat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_n^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H = \frac{1}{n}\sum_{i=1}^{n} u\circ g^{-1}\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$
$$= \frac{1}{n}\sum_{i=1}^{n} v\left(\frac{1}{p}\mathbf{x}_i^H\widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

  with $v = u\circ g^{-1}$.

Heuristic approach

- Recall that:

$$\hat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} v\left(\frac{1}{p}\mathbf{x}_i^H \hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

- Intuitively $\hat{\mathbf{C}}_{(i)}^{-1}$ and $\mathbf{x}_i$ are weakly dependent.

Heuristic approach

- Recall that:

$$\hat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} v\left(\frac{1}{p}\mathbf{x}_i^H \hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

- Intuitively $\hat{\mathbf{C}}_{(i)}^{-1}$ and $\mathbf{x}_i$ are weakly dependent.

$\Longrightarrow$ We expect in particular using the convergence of quadratic forms lemma that:

$$\frac{1}{p}\mathbf{x}_i^H \hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i \simeq \tau_i \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_{(i)}^{-1} \simeq \tau_i \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n^{-1}$$

$\Longrightarrow$ We thus have:

$$\hat{\mathbf{C}}_n \simeq \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n^{-1}\right)\mathbf{x}_i\mathbf{x}_i^H$$

Heuristic approach

- Recall that:

$$\hat{\mathbf{C}}_n = \frac{1}{n}\sum_{i=1}^{n} v\left(\frac{1}{p}\mathbf{x}_i^H \hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^H$$

- Intuitively $\hat{\mathbf{C}}_{(i)}^{-1}$ and $\mathbf{x}_i$ are weakly dependent.

$\Longrightarrow$ We expect in particular using the convergence of quadratic forms lemma that:

$$\frac{1}{p}\mathbf{x}_i^H \hat{\mathbf{C}}_{(i)}^{-1}\mathbf{x}_i \simeq \tau_i \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_{(i)}^{-1} \simeq \tau_i \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n^{-1}$$

$\Longrightarrow$ We thus have:

$$\hat{\mathbf{C}}_n \simeq \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n^{-1}\right)\mathbf{x}_i\mathbf{x}_i^H$$

- We assume that $\frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n^{-1} \simeq \gamma_n$

$$\Longrightarrow \hat{\mathbf{C}}_n \simeq \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\mathbf{x}_i\mathbf{x}_i^H$$

Heuristic approach

- **Determination of γ_n.**
  - Recall that:

$$\widehat{\mathbf{C}}_n \simeq \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i \gamma\right) \mathbf{x}_i \mathbf{x}_i^H = \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i \gamma\right) \tau_i \mathbf{w}_i \mathbf{w}_i^H$$

  - Moreover, $\gamma \simeq \frac{1}{p} \operatorname{tr} \widehat{\mathbf{C}}_n^{-1}$. Hence,

$$\gamma \simeq \frac{1}{p} \operatorname{tr} \left(\frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i \gamma\right) \tau_i \mathbf{w}_i \mathbf{w}_i^H \right)^{-1}$$

Heuristic approach

- **Determination of γ_n.**
  - Recall that:

$$\hat{\mathbf{C}}_n \simeq \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\mathbf{x}_i\mathbf{x}_i^H = \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\tau_i\mathbf{w}_i\mathbf{w}_i^H$$

  - Moreover, $\gamma \simeq \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n^{-1}$. Hence,

$$\gamma \simeq \frac{1}{p}\operatorname{tr}\left(\frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\tau_i\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}$$

  - Let $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_n]$ and $\mathbf{D} = \mathrm{diag}\{\tau_i\gamma\}_{i=1}^n$. Then:

$$\gamma \sim \frac{1}{p}\operatorname{tr}\left(\mathbf{WDW}^H\right)^{-1} = m_{\frac{1}{n}\mathbf{WDW}^H}(0)$$

  where m is the stieltjes transform associated with $\mathbf{WDW}^H$ at 0.

Heuristic approach

- **Determination of γ_n.**
  - Recall that:

$$\hat{\mathbf{C}}_n \simeq \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\mathbf{x}_i\mathbf{x}_i^H = \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\tau_i\mathbf{w}_i\mathbf{w}_i^H$$

  - Moreover, $\gamma \simeq \frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n^{-1}$. Hence,

$$\gamma \simeq \frac{1}{p}\operatorname{tr}\left(\frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\tau_i\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}$$

  - Let $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_n]$ and $\mathbf{D} = \mathrm{diag}\{\tau_i\gamma\}_{i=1}^n$. Then:

$$\gamma \sim \frac{1}{p}\operatorname{tr}\left(\mathbf{W}\mathbf{D}\mathbf{W}^H\right)^{-1} = m_{\frac{1}{n}\mathbf{W}\mathbf{D}\mathbf{W}^H}(0)$$

    where m is the stieltjes transform associated with $\mathbf{W}\mathbf{D}\mathbf{W}^H$ at 0.
  - Using standard results from random matrix theory,

$$m_{\frac{1}{n}\mathbf{W}\mathbf{D}\mathbf{W}^H}(0) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_i v(\tau_i\gamma)}{1 + c\tau_i v(\tau_i\gamma) m_{\frac{1}{n}\mathbf{W}\mathbf{D}\mathbf{W}^H}(0)}\right)^{-1}$$

Heuristic approach

- **Determination of γ_n.**
  - Recall that:

$$\widehat{\mathbf{C}}_n \simeq \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\mathbf{x}_i\mathbf{x}_i^H = \frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\tau_i\mathbf{w}_i\mathbf{w}_i^H$$

  - Moreover, $\gamma \simeq \frac{1}{p}\operatorname{tr}\widehat{\mathbf{C}}_n^{-1}$. Hence,

$$\gamma \simeq \frac{1}{p}\operatorname{tr}\left(\frac{1}{n}\sum_{i=1}^{n} v\left(\tau_i\gamma\right)\tau_i\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}$$

  - Let $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_n]$ and $\mathbf{D} = \mathrm{diag}\{\tau_i\gamma\}_{i=1}^{n}$. Then:

$$\gamma \sim \frac{1}{p}\operatorname{tr}\left(\mathbf{WDW}^H\right)^{-1} = m_{\frac{1}{n}\mathbf{WDW}^H}(0)$$

    where m is the stieltjes transform associated with $\mathbf{WDW}^H$ at 0.
  - Using standard results from random matrix theory,

$$m_{\frac{1}{n}\mathbf{WDW}^H}(0) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_i v(\tau_i\gamma)}{1 + c\tau_i v(\tau_i\gamma) m_{\frac{1}{n}\mathbf{WDW}^H}(0)}\right)^{-1}$$

$\Longrightarrow$ We define γ as a solution of the fixed-point equation:

$$\gamma = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_i v(\tau_i\gamma)}{1 + c\tau_i v(\tau_i\gamma)\gamma}\right)^{-1} = \gamma\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\psi(\tau_i\gamma)}{1 + c\psi(\tau_i\gamma)}\right)^{-1}, \quad \psi(x) = xv(x)$$

Main result

R. Couillet, F. Pascal, J. W. Silverstein " The random matrix regime of Maronna's M-estimator with elliptically distributed samples " Elsevier Journal of Multivariate Analysis

Theorem

Asymptotic equivalence Let γ be the unique positive solution of :

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

Then, under the assumptions defined earlier, we have:

$$\|\widehat{\mathbf{C}}_n - \widehat{\mathbf{S}}_n\| \xrightarrow{\text{a.s.}} 0 \text{ where } \widehat{\mathbf{S}}_n = \frac{1}{n} \sum_{i=1}^{n} v(\tau_i \gamma) \mathbf{x}_i \mathbf{x}_i^H$$

- **Takeaways**
  - Propagation to $\widehat{\mathbf{S}}_n$ of first order results on $\widehat{\mathbf{C}}_n$
  - $\widehat{\mathbf{S}}_n$ can be studied using classical results from random matrix theory, contrary to $\widehat{\mathbf{C}}_n$.

$\Longrightarrow$ Important consequences for array signal processing applications.

Main steps of the proof

- Let $d_i = \frac{1}{p}\mathbf{w}_i^H \widehat{\mathbf{C}}_{(i)}^{-1}\mathbf{w}_i$. The proof consists in showing that

$$\max_{1\leqslant i\leqslant n} |e_i - 1| \xrightarrow{\text{a.s.}} 0, \qquad \text{with} \qquad e_i = \frac{v(\tau_i d_i)}{v(\tau_i \gamma)}$$

- We relabel $e_1, \cdots, e_n$ such that:

$$e_1 \leqslant \cdots \leqslant e_n.$$

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

$$e_1 \geqslant 1 - \ell \ i.o. \quad \text{and} \quad e_n \leqslant 1 + \ell \ i.o.$$

Main steps in the proof

- We proceed by contradiction. We assume that $e_n \geqslant 1 + \ell$ $i.o.$

$$e_n = \frac{v(\tau_n d_n)}{v(\tau_n \gamma_n)}$$

Main steps in the proof

- We proceed by contradiction. We assume that $e_n \geqslant 1 + \ell$ *i.o.*

$$e_n = \frac{v(\tau_n d_n)}{v(\tau_n \gamma_n)}$$

$$v(\tau_n \gamma_n) e_n = v(\tau_n d_n) = v\left(\tau_n \frac{1}{p} \mathbf{w}_n^H \widehat{\mathbf{C}}_{(n)}^{-1} \mathbf{w}_n\right)$$

Main steps in the proof

- We proceed by contradiction. We assume that $e_n \geqslant 1 + \ell$ *i.o.*

$$e_n = \frac{v(\tau_n d_n)}{v(\tau_n \gamma_n)}$$

$$\begin{aligned} v(\tau_n \gamma_n) e_n = v(\tau_n d_n) &= v\left(\tau_n \frac{1}{p} \mathbf{w}_n^H \widehat{\mathbf{C}}_{(n)}^{-1} \mathbf{w}_n\right) \\ &= v\left(\tau_n \frac{1}{p} \mathbf{w}_n^H \left(\frac{1}{n} \sum_{i<n} \tau_i v(\tau_i d_i) \mathbf{w}_i \mathbf{w}_i^H\right)^{-1} \mathbf{w}_n\right) \end{aligned}$$

Main steps in the proof

- We proceed by contradiction. We assume that $e_n \geqslant 1 + \ell \; i.o.$

$$e_n = \frac{v(\tau_n d_n)}{v(\tau_n \gamma_n)}$$

$$\begin{aligned} v(\tau_n\gamma_n)e_n = v(\tau_n d_n) &= v\left(\tau_n \frac{1}{p}\mathbf{w}_n^H \widehat{\mathbf{C}}_{(n)}^{-1}\mathbf{w}_n\right) \\ &= v\left(\tau_n \frac{1}{p}\mathbf{w}_n^H \left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i d_i)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right) \\ &= v\left(\tau_n \frac{1}{p}\mathbf{w}_n^H \left(\frac{1}{n}\sum_{i<n}\tau_i e_i v(\tau_i \gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right) \end{aligned}$$

Main steps in the proof

- We proceed by contradiction. We assume that $e_n \geqslant 1+\ell$ *i.o.*

$$e_n = \frac{v(\tau_n d_n)}{v(\tau_n \gamma_n)}$$

$$\begin{aligned}
v(\tau_n\gamma_n)e_n = v(\tau_n d_n) &= v\left(\tau_n \frac{1}{p}\mathbf{w}_n^H \widehat{\mathbf{C}}_{(n)}^{-1}\mathbf{w}_n\right) \\
&= v\left(\tau_n \frac{1}{p}\mathbf{w}_n^H \left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i d_i)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right) \\
&= v\left(\tau_n \frac{1}{p}\mathbf{w}_n^H \left(\frac{1}{n}\sum_{i<n}\tau_i e_i v(\tau_i \gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right) \\
&\leqslant v\left(\tau_n e_n^{-1} \frac{1}{p}\mathbf{w}_n^H \left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i \gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right)
\end{aligned}$$

Main steps in the proof

- Recall that:

$$v(\tau_n\gamma_n)e_n \leqslant v\left(\tau_n e_n^{-1}\frac{1}{p}\mathbf{w}_n^H\left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i\gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right)$$

Main steps in the proof

- Recall that:

$$v(\tau_n\gamma_n)e_n \leqslant v\left(\tau_n e_n^{-1}\frac{1}{p}\mathbf{w}_n^H\left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i\gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right)$$

- Using standard approaches from random matrix theory, we can prove that:

$$\max_{1\leqslant j\leqslant n}\left|\frac{1}{p}\mathbf{w}_j^H\left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i\gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_j-\gamma_n\right| \xrightarrow{\text{a.s.}} 0.$$

Main steps in the proof

- Recall that:

$$v(\tau_n \gamma_n) e_n \leqslant v\left(\tau_n e_n^{-1} \frac{1}{p} \mathbf{w}_n^H \left(\frac{1}{n}\sum_{i<n} \tau_i v(\tau_i \gamma_n) \mathbf{w}_i \mathbf{w}_i^H\right)^{-1} \mathbf{w}_n\right)$$

- Using standard approaches from random matrix theory, we can prove that:

$$\max_{1\leqslant j\leqslant n}\left|\frac{1}{p}\mathbf{w}_j^H \left(\frac{1}{n}\sum_{i<n} \tau_i v(\tau_i \gamma_n) \mathbf{w}_i \mathbf{w}_i^H\right)^{-1} \mathbf{w}_j - \gamma_n\right| \xrightarrow{\text{a.s.}} 0.$$

$\Longrightarrow$

$$v(\tau_n \gamma_n) e_n \leqslant v(\tau_n e_n^{-1}(\gamma_n - \epsilon_n)), \text{ with } \epsilon_n \to 0.$$

Main steps in the proof

- Recall that:

$$v(\tau_n\gamma_n)e_n \leqslant v\left(\tau_n e_n^{-1}\frac{1}{p}\mathbf{w}_n^H\left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i\gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right)$$

- Using standard approaches from random matrix theory, we can prove that:

$$\max_{1\leqslant j\leqslant n}\left|\frac{1}{p}\mathbf{w}_j^H\left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i\gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_j-\gamma_n\right|\xrightarrow{\text{a.s.}}0.$$

$\Longrightarrow$

$$v(\tau_n\gamma_n)e_n \leqslant v(\tau_n e_n^{-1}(\gamma_n-\epsilon_n)),\ \text{with } \epsilon_n\to 0.$$

- Use $\psi(x)=xv(x)$,

$$\psi(\tau_n\gamma_n)\leqslant\psi(\tau_n e_n^{-1}\gamma_n)\left(1-\epsilon_n\gamma_n^{-1}\right)^{-1}$$

Main steps in the proof

- Recall that:

$$v(\tau_n\gamma_n)e_n \leqslant v\left(\tau_n e_n^{-1}\frac{1}{p}\mathbf{w}_n^H\left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i\gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_n\right)$$

- Using standard approaches from random matrix theory, we can prove that:

$$\max_{1\leqslant j\leqslant n}\left|\frac{1}{p}\mathbf{w}_j^H\left(\frac{1}{n}\sum_{i<n}\tau_i v(\tau_i\gamma_n)\mathbf{w}_i\mathbf{w}_i^H\right)^{-1}\mathbf{w}_j-\gamma_n\right|\xrightarrow{\text{a.s.}} 0.$$

$\Longrightarrow$

$$v(\tau_n\gamma_n)e_n \leqslant v(\tau_n e_n^{-1}(\gamma_n-\epsilon_n)),\ \text{with } \epsilon_n\to 0.$$

- Use $\psi(x)=xv(x)$,

$$\psi(\tau_n\gamma_n)\leqslant\psi(\tau_n e_n^{-1}\gamma_n)\left(1-\epsilon_n\gamma_n^{-1}\right)^{-1}$$

- Assume that $e_n\geqslant 1+\ell$ i.o, and $\tau_n\in[a,b]$. Consider a sequence over which $e_n>1+\ell$, $\tau_n\to\tau_0$ and $\gamma_n\to\gamma_0$,

$$\psi(\tau_0\gamma_0)\leqslant\psi\left(\frac{\tau_0\gamma_0}{1+\ell}\right),\quad \text{Contradiction}$$

Simulations

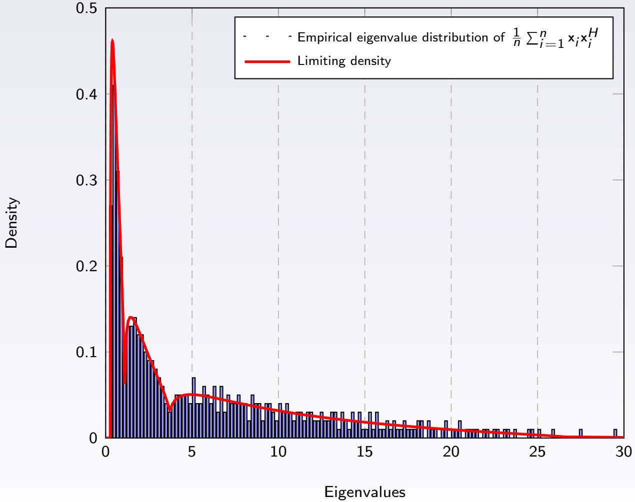


Figure: Histogram of the eigenvalues of $\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^H$ for $n = 2500$, $N = 500$, $C_p = \mathrm{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

Simulations

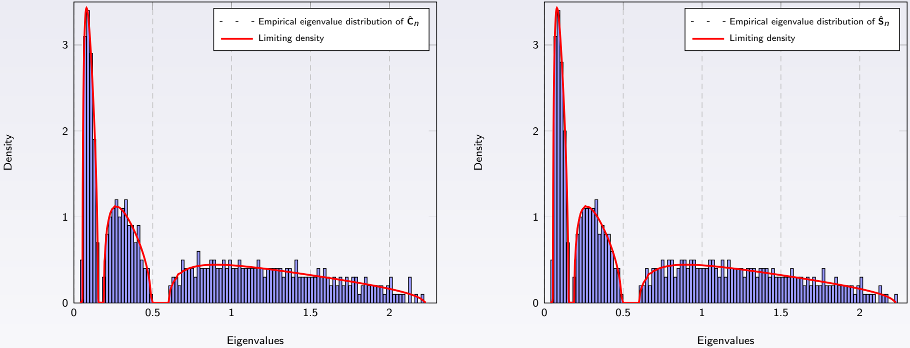
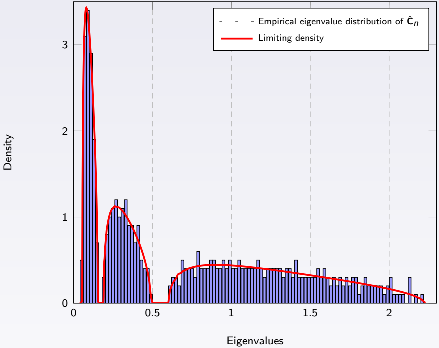


Figure: Histogram of the eigenvalues of $\hat{\mathbf{C}}_n$ (left) and $\hat{\mathbf{S}}_n$ (right) for $n = 2500$, $N = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

Simulations



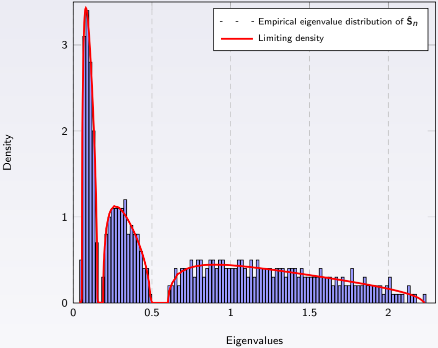


Figure: Histogram of the eigenvalues of $\hat{\mathbf{C}}_n$ (left) and $\hat{\mathbf{S}}_n$ (right) for $n = 2500$, $N = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, τ_1 with $\Gamma(.5, 2)$-distribution.

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

Extension to information plus-noise model

A. Kammoun and M.-S. Alouini " The random matrix regime of Maronna's M-estimator for observations corrupted by elliptical noises" Submitted to Journal of Multivariate Analysis

- Information plus noise model:

$$\mathbf{x}_i = \mathbf{A}\mathbf{s}_i + \sqrt{\tau_i}\mathbf{w}_i$$

  - $\mathbf{A} \in \mathbb{C}^{p\times K}$ is of full rank.
  - Let $\mathbf{B}_n = \mathbf{A}\mathbf{A}^H$, we assume $\liminf \frac{1}{n}\operatorname{tr}\mathbf{B}_n > 0$
  - $\mathbf{s}_1, \cdots, \mathbf{s}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, independent standard Gaussian distributed vectors.
  - Same assumptions on functions ϕ and u,
  - Random matrix theory regime: $n, p, K \to \infty$

Then:

$$\|\widehat{\mathbf{C}}_n - \widehat{\mathbf{S}}_n\| \xrightarrow{\text{a.s.}} 0,$$

where

$$\widehat{\mathbf{S}}_n = \frac{1}{n}\sum_{i=1}^{n} v(\delta_i)\mathbf{x}_i\mathbf{x}_i^H$$

with $\delta_1, \cdots, \delta_n$ being the unique solutions to:

$$\delta_i = \frac{1}{p}\operatorname{tr}\left(\mathbf{B}_n + \tau_i\mathbf{I}_p\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{v(\delta_j)(\mathbf{B}_n + \tau_j\mathbf{I}_p)}{1 + c\psi(\delta_j)}\right)^{-1}$$

Illustration

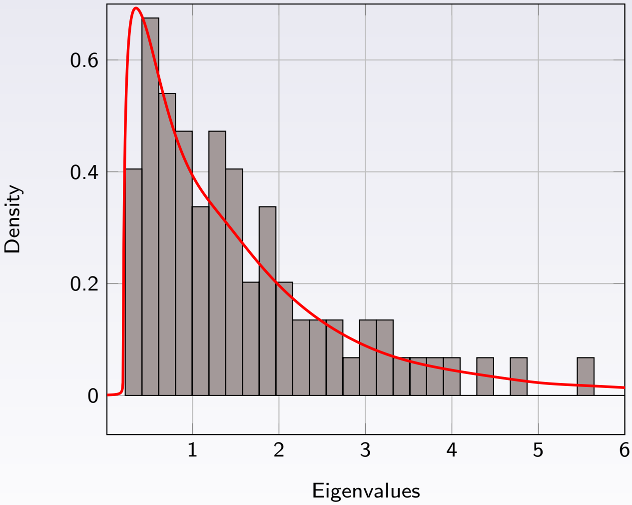
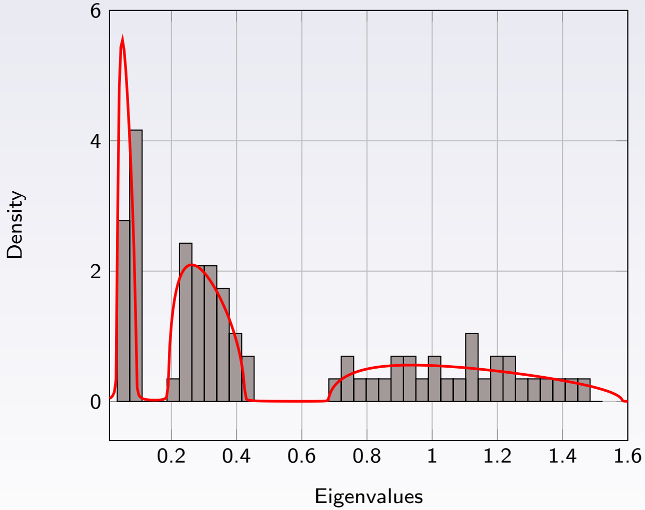


Figure: Histogram of the eigenvalues of $\hat{\mathbf{S}}_n$ against the limiting spectral measure with $n = 800$, $p = 80$ $\mathbf{B}_p = \mathrm{diag}(0.5\mathbf{I}_30, \mathbf{I}_{30}, 0\mathbf{I}_20)$

Illustration



;

Figure: Histogram of the eigenvalues of $\hat{C}_N$ against the limiting spectral measure for $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.1$, $N = 80$, $n = 800$

Extension to spiked models

- **Assumption:** $\mathbf{x}_1, \cdots, \mathbf{x}_n \in \mathbb{C}^p$ independent:

$$\mathbf{x}_i = \sum_{\ell=1}^{K} \sqrt{p_\ell}\mathbf{a}_\ell s_{\ell i} + \sqrt{\tau_i}\mathbf{w}_i$$

  - $\mathbf{w}_i \in \mathbb{C}^p$ independent standard Gaussian vectors,
  - τ_i deterministic scalars.
  - $p_1 \geqslant \cdots \geqslant p_K \geqslant 0$
  - $\mathbf{a}_1, \cdots, \mathbf{a}_K \in \mathbb{C}^{p\times 1}$ deterministic with $\sum_{\ell=1}^{K} p_\ell \mathbf{a}_\ell \mathbf{a}_\ell^H = \mathrm{diag}\{p_i\}_{i=1}^{K}$.
  - $\frac{1}{n}\sum_{i=1}^{n} \delta_{\tau_i} \to \nu$ weakly.
  - $s_{1,i}, \cdots, s_{Kn}$ independent with zero mean and unit variance.

Theorem

Under the previous assumptions, as $n \to \infty$,

$$\|\widehat{\mathbf{C}}_n - \widehat{\mathbf{S}}_n\| \xrightarrow{\text{a.s.}} 0,$$

with

$$\widehat{\mathbf{S}}_n = \frac{1}{n}\sum_{i=1}^{n} v(\tau_i\gamma)\mathbf{x}_i\mathbf{x}_i^H$$

Outline

Eigenvalue Localization

R. Couillet "Robust spiked random matrices and a robust G-MUSIC estimator " Journal of Multivariate Analysis 2014

Theorem (Eigenvalue localization)

Let $\hat{\lambda}_1, \cdots, \hat{\lambda}_p$ be the eigenvalues of $\widehat{\mathbf{C}}_n$. 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

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

Eigenvalue Localization

R. Couillet "Robust spiked random matrices and a robust G-MUSIC estimator " Journal of Multivariate Analysis 2014

Theorem (Eigenvalue localization)

Let $\hat{\lambda}_1, \cdots, \hat{\lambda}_p$ be the eigenvalues of $\widehat{\mathbf{C}}_n$. 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

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

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

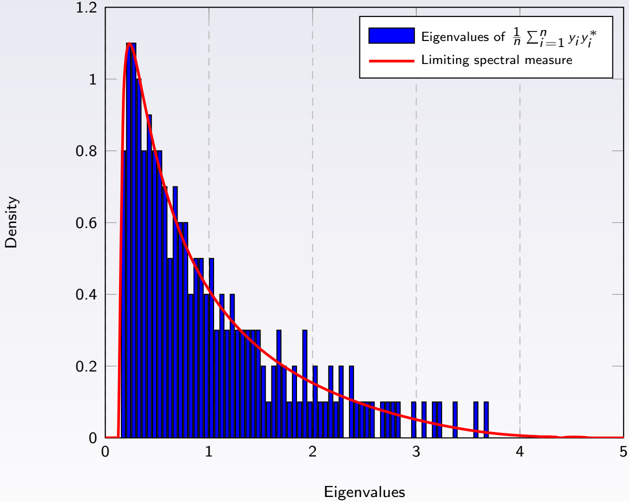


Figure: Histogram of the eigenvalues of $\frac{1}{n}\sum_i y_i y_i^*$ against the limiting spectral measure, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Sudent-t impulsions.

Simulation

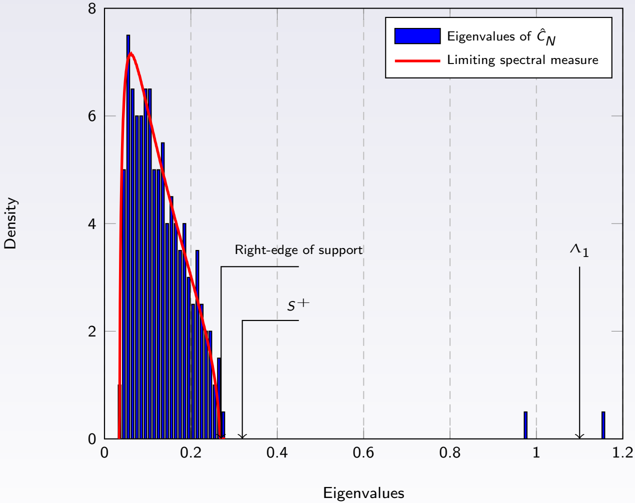


Figure: Histogram of the eigenvalues of $\hat{C}_n$ against the limiting spectral measure, for $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Comments

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

Comments

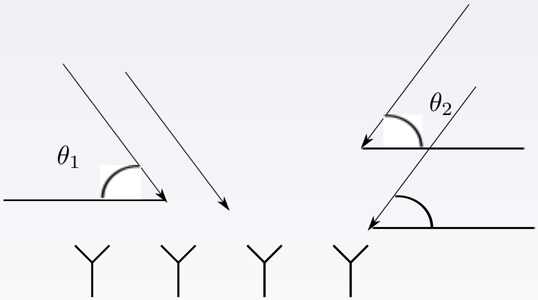
- **SCM vs robust:** Spikes invisible in SCM due to impulsive noise, reborn in robust estimate of scatter.
- **Largest eigenvalues:**
  - $\lambda_i(\widehat{\mathbf{C}}_n) > S^+ \Rightarrow$ Presence of a source!
  - $\lambda_i(\widehat{\mathbf{C}}_n) \in (\sup(\text{Support}), S^+) \Rightarrow$ May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

$\Rightarrow$ Induces a natural source detection algorithm.

Outline

Music algorithm

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



Source Localization using Music Algorithm

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

$$\mathbf{x}_i = \sum_{k=1}^{K} \sqrt{p}_k \mathbf{a}_p(\theta_k) s_{k,t} + \sqrt{\tau}_i \mathbf{w}_i, t = 1, \cdots, n$$

- $\mathbf{A}_p = \left[\mathbf{a}_p(\theta_1), \cdots, \mathbf{a}_p(\theta_K)\right]$ with $\mathbf{a}_p(\theta) = \begin{bmatrix} 1 \\ e^{\imath\pi \sin\theta} \\ \cdots \\ e^{\imath(p-1)\pi\sin\theta} \end{bmatrix}$
- Objective: infer $\theta_1, \cdots, \theta_K$ from the n observations
- Let $\mathbf{X}_n = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$, then,

$$\mathbf{X}_n = \mathbf{A}_p \operatorname{diag}\{p_i\}_{i=1}^{K} \mathbf{S} + \mathbf{W}\operatorname{diag}\{\tau_i\}_{i=1}^{n}$$

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

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

MUSIC algorithms

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

$$\hat{\eta} = \mathbf{a}_p(\theta)\hat{\Pi}\mathbf{a}_p(\theta)$$

  where $\hat{\Pi}$ is the orthogonal projection matrix on the eigenspace associated to the $p-K$ largest eigenvalues of $\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^H$.
- G-MUSIC: Angles are estimated as local minima of : $\hat{\eta}_G$ where $\hat{\eta}_G$ is a consistent estimator of $\eta(\theta)$ that is based on the sample covariance matrix $\frac{1}{n}\mathbf{X}_n\mathbf{X}_n^H$.

$$\xrightarrow{\mathbf{x}_1,\cdots,\mathbf{x}_n} \boxed{\widehat{\mathbf{S}}_n} \xrightarrow{\widehat{\mathbf{S}}_n} \boxed{g(\widehat{\mathbf{S}}_n)}$$

Sample covariance matrix — Consistent Estimator $\eta(\theta)$

- Robust G-MUSIC: Angles are estimated as local minima of : $\hat{\eta}_{R,G}$ where $\hat{\eta}_{R,G}$ is a consistent estimator of η that is based on the M-scatter estimator $\hat{\mathbf{C}}_n$.

$$\xrightarrow{\mathbf{x}_1,\cdots,\mathbf{x}_n} \boxed{\widehat{\mathbf{C}}_n} \xrightarrow{\widehat{\mathbf{C}}_n} \boxed{g(\widehat{\mathbf{C}}_n)}$$

M-scatter estimator — Consistent Estimator $\eta(\theta)$

Eigenvalue and eigenvector projection estimates

- $\mathbf{u}_k$ eigenvector of k-th largest eigenvalue of $\mathbf{A}\mathbf{A}^H \sum_{i=1}^K p_i \mathbf{a}_i(\theta)\mathbf{a}_i(\theta)^H$
- $\hat{\mathbf{u}}_k$ eigenvector of k-th largest eigenvalue of $\hat{\mathbf{C}}_n$

Eigenvalue and eigenvector projection estimates

- $\mathbf{u}_k$ eigenvector of k-th largest eigenvalue of $\mathbf{A}\mathbf{A}^H \sum_{i=1}^K p_i \mathbf{a}_i(\theta)\mathbf{a}_i(\theta)^H$
- $\hat{\mathbf{u}}_k$ eigenvector of k-th largest eigenvalue of $\hat{\mathbf{C}}_n$

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 $p_j > p_-$*

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

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

Eigenvalue and eigenvector projection estimates

Theorem (Estimation under unknown ν)

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

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

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

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

where

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

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

Application to G-MUSIC

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

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

Application to G-MUSIC

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

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

Corollary (Robust G-MUSIC)

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

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

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

Then, for each $p_j > p_-$,

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

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

where

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

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

Simulations: Single-shot in elliptical noise

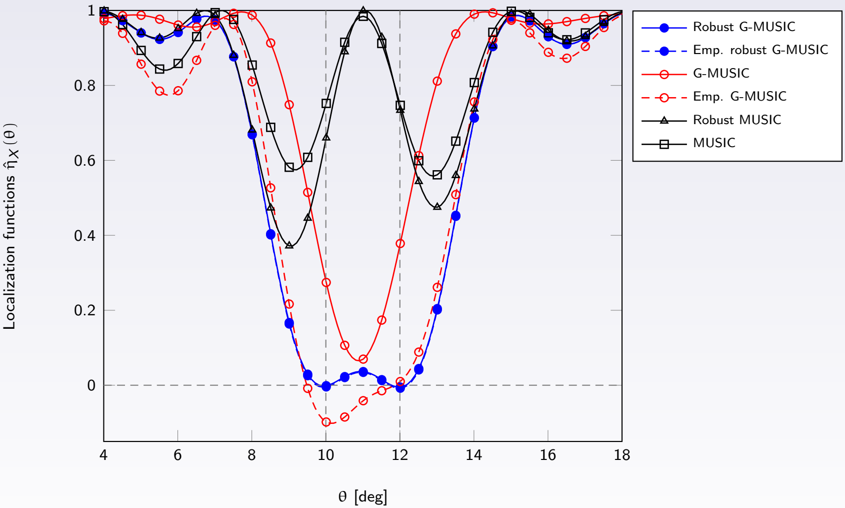


Figure: Random realization of the localization functions for the various MUSIC estimators, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 100$, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$. Powers $p_1 = p_2 = 10^{0.5} = 5$ dB.

Simulations: Elliptical noise

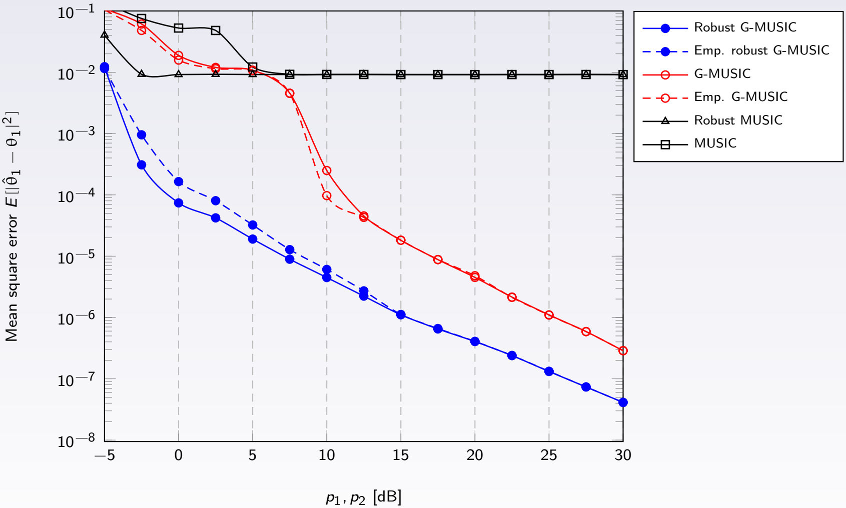


Figure: Means square error performance of the estimation of $\theta_1 = 10^\circ$, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 10$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

Outline

Regularized estimators

- Recall in Gaussian settings, the Ledoit-Wolf estimator is given by:

$$(1-\rho)\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^H + \rho\mathbf{I}_p, \quad \text{for some} \quad \rho \in [0,1] .$$

- Regularized Tyler estimators in elliptical distributed data settings

$$\hat{\mathbf{C}}_n(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H\hat{\mathbf{C}}_N^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_p, \quad \rho \in \left(\max\left\{0, \frac{p-n}{n}\right\}, 1\right]$$

$$\check{\mathbf{C}}_n(\rho) = \frac{\check{\mathbf{B}}_n(\rho)}{\frac{1}{p}\operatorname{tr}\check{\mathbf{B}}_n(\rho)}, \quad \check{\mathbf{B}}_n(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H\check{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho\mathbf{I}_p, \quad \rho \in (0,1]$$

Main theoretical result

- **Which estimator is better ?** There is no clear answer to this question
- **Result using RMT** It can be proven that they are equivalent in the asymptotic random matrix regime.
- **Assumptions**
  - Elliptical model:

$$\mathbf{x}_i = \sqrt{\tau_i}\mathbf{C}_p^{\frac{1}{2}}\mathbf{w}_i$$

    with $\mathbf{w}_1, \cdots, \mathbf{w}_n$ independent standard Gaussian random vectors.
  - Matrix $\mathbf{C}_p$ satisfies $\frac{1}{p}\operatorname{tr}\mathbf{C}_p = 1$ with $\limsup \|\mathbf{C}_p\| < \infty$
  - $\nu_p = \frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i}(\mathbf{C}_p) \to \nu$, weakly with $\nu \neq \delta_0$ almost everywhere.

Regularized Tyler estimator

R. Couillet and M. R. McKay "Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators" Journal Of Multivariate Analysis 2014

Theorem

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

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

where

$$\hat{\mathbf{C}}_n(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H \hat{\mathbf{C}}_n(\rho)^{-1} x_i} + \rho I_p$$

$$\hat{\mathbf{S}}_n(\rho) = \frac{1}{\hat{\gamma}(\rho)} \frac{1-\rho}{1-(1-\rho)c} \frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_n^{\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^H \mathbf{C}_n^{\frac{1}{2}} + \rho I_p$$

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

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

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

Normalized Regularized Tyler Estimator

Theorem (Normalized Regularized Tyler estimator)

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

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

$$\text{where}\check{\mathbf{C}}_n(\rho) = \frac{\check{\mathbf{B}}_n(\rho)}{\frac{1}{N} tr \check{\mathbf{B}}_N(\rho)}, \quad \check{\mathbf{B}}_N(\rho) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H \check{\mathbf{C}}_n(\rho)^{-1}\mathbf{x}_i} + \rho I_p$$

$$\check{S}_n(\rho) = \frac{1-\rho}{1-\rho+T_\rho}\frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_n^{\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^H \mathbf{C}_n^{\frac{1}{2}} + \frac{T_\rho}{1-\rho+T_\rho} I_p$$

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* $\check{\gamma}$

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

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

Asymptotic model equivalence

R. Couillet and M. R. McKay "Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators" Journal Of Multivariate Analysis 2014

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*

$$\frac{\hat{\mathbf{S}}_n(\hat{\rho})}{\frac{1}{\hat{\gamma}(\hat{\rho})}\frac{1-\hat{\rho}}{1-(1-\hat{\rho})c} + \hat{\rho}} = \check{\mathbf{S}}_n(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_n^{\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^* \mathbf{C}_p^{\frac{1}{2}} + \rho I_p.$$

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

R. Couillet and M. R. McKay "Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators" Journal Of Multivariate Analysis 2014

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*

$$\frac{\hat{\mathbf{S}}_n(\hat{\rho})}{\frac{1}{\hat{\gamma}(\hat{\rho})}\frac{1-\hat{\rho}}{1-(1-\hat{\rho})c} + \hat{\rho}} = \check{\mathbf{S}}_n(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_n^{\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^* \mathbf{C}_p^{\frac{1}{2}} + \rho I_p.$$

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

R. Couillet and M. R. McKay "Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators" Journal Of Multivariate Analysis 2014

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*

$$\frac{\hat{\mathbf{S}}_n(\hat{\rho})}{\frac{1}{\hat{\gamma}(\hat{\rho})}\frac{1-\hat{\rho}}{1-(1-\hat{\rho})c}+\hat{\rho}} = \check{\mathbf{S}}_n(\check{\rho}) = (1-\rho)\frac{1}{n}\sum_{i=1}^{n}\mathbf{C}_n^{\frac{1}{2}}\mathbf{w}_i\mathbf{w}_i^*\mathbf{C}_p^{\frac{1}{2}} + \rho I_p.$$

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

$$\hat{D}_n(\rho) = \frac{1}{p} tr \left(\left(\frac{\hat{\mathbf{C}}_p(\rho)}{\frac{1}{p} tr \hat{\mathbf{C}}_n(\rho)} - \mathbf{C}_p \right)^2 \right), \quad \check{D}_n(\rho) = \frac{1}{p} tr \left(\left(\check{\mathbf{C}}_n(\rho) - \mathbf{C}_p \right)^2 \right).$$

Denote $D^\star = c \frac{M_2 - 1}{c + M_2 - 1}$, $\rho^\star = \frac{c}{c + M_2 - 1}$, $M_2 = \lim_n \frac{1}{p} \sum_{i=1}^N \lambda_i^2(\mathbf{C}_n)$ 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,

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

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

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

$$\frac{\hat{\rho}_n}{\frac{1}{p}\operatorname{tr}\hat{\mathbf{C}}_n(\hat{\rho}_n)} = \frac{c_n}{\frac{1}{p}\operatorname{tr}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\|\mathbf{x}_i\|^2}\right)^2\right] - 1}$$

$$\frac{\check{\rho}_n \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{x}_i^H \check{\mathbf{C}}_n(\check{\rho}_n)^{-1}\mathbf{x}_i}{\|\mathbf{x}_i\|^2}}{1 - \check{\rho}_n + \check{\rho}_n \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{x}_i^H \check{\mathbf{C}}_N(\check{\rho}_n)^{-1}\mathbf{x}_i}{\|\mathbf{x}_i\|^2}} = \frac{c_n}{\frac{1}{p}\operatorname{tr}\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\|\mathbf{x}_i\|^2}\right)^2\right] - 1}$$

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

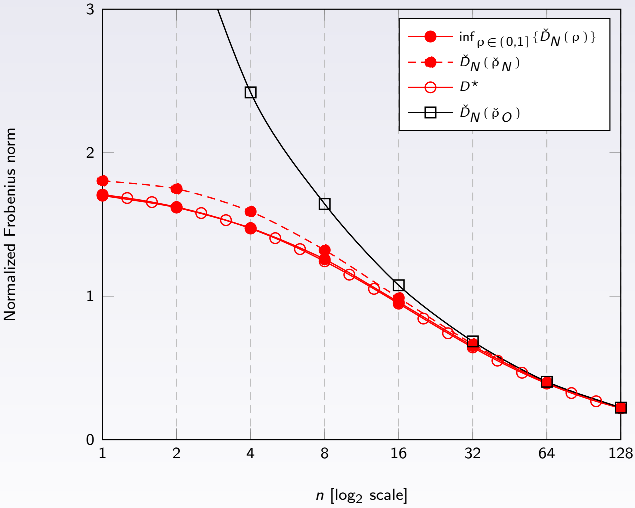


Figure: Performance of optimal shrinkage averaged over 10 000 Monte Carlo simulations, for $N = 32$, various values of n, $[C_N]_{ij} = r^{|i-j|}$ with $r = 0.7$; $\check{\rho}_N$ as above; $\check{\rho}_O$ the clairvoyant estimator proposed in (Chen'11).

Second-order statistics of the regularized Tyler estimator

- Motivation: From above, we know that:

$$\sup_{\rho} \|\hat{\mathbf{C}}_n(\rho) - \hat{\mathbf{S}}_n(\rho)\| \xrightarrow{\text{a.s.}} 0,$$

- First order implications :
  - $\left|\mathbf{a}^H \hat{\mathbf{C}}_n(\rho)\mathbf{b} - \mathbf{a}^H \hat{\mathbf{S}}_n(\rho)\mathbf{b}\right| \xrightarrow{\text{a.s.}} 0$
  - $\left|\frac{1}{n}\operatorname{tr} f(\hat{\mathbf{C}}_n(\rho)) - \frac{1}{n}\operatorname{tr} f(\hat{\mathbf{S}}_n(\rho))\right| \xrightarrow{\text{a.s.}} 0.$
- Does not imply propagation to $\hat{\mathbf{S}}_n(\rho)$ of second-order results on $\hat{\mathbf{C}}_n(\rho)$.
- From simulations it seems that:

$$n^{\frac{1}{2}-\epsilon} \|\hat{\mathbf{C}}_n(\rho) - \hat{\mathbf{S}}_n(\rho)\| \xrightarrow{\text{a.s.}} 0.$$

$\Longrightarrow$ $n^{\frac{1}{2}-\epsilon}\left(\mathbf{a}^H \hat{\mathbf{C}}_n \mathbf{b} - \mathbf{a}^H \hat{\mathbf{S}}_n \mathbf{b}\right) \to 0,$ Weak result

- Since $\sqrt{n}\mathbf{a}^H \hat{\mathbf{S}}_n \mathbf{b} - \sqrt{n}\mathbf{a}^H \mathbb{E}\left[\hat{\mathbf{S}}_n\right]\mathbf{b} \to \mathcal{N}(0, \sigma^2)$, we expect that:

$$\sqrt{n}\mathbf{a}^H \hat{\mathbf{C}}_n \mathbf{b} - \sqrt{n}\mathbf{a}^H \mathbb{E}\left[\hat{\mathbf{C}}_n\right]\mathbf{b} \to \mathcal{N}(0, \sigma^2)$$

Fluctuations of quadratic forms of $\widehat{\mathbf{C}}_n(\rho)$

R. Couillet, A. Kammoun, F. Pascal " Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals" Journal of multivariate Analysis 2015.

Theorem

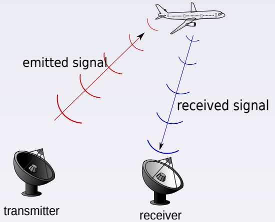
Let $\mathbf{a}, \mathbf{b} \in \mathbb{C}^p$ *with* $\|\mathbf{a}\| = \|\mathbf{b}\| = 1$. *Then, as* $n \to \infty$ *with* $p/n \to c \in (0, \infty)$ *for all* $\epsilon > 0$, $k \in \mathbb{Z}$,

$$\sup_{\rho \in \mathcal{R}_\kappa} p^{1-\epsilon} \left| \mathbf{a}^H \widehat{\mathbf{C}}_n^k \mathbf{b} - \mathbf{a}^H \widehat{\mathbf{S}}_n^k \mathbf{b} \right| \xrightarrow{\text{a.s.}} 0,$$

- **Comments**:
  - The proof is involved and relies on Martingale computations
  - The fluctuations of quadratic forms associated with $\widehat{\mathbf{C}}_n$ as the same as that associated with $\widehat{\mathbf{S}}_n$
  - This result has important implications in array signal processing applications wherein such quadratic forms naturally arise.

Outline

Application: Radar detection



$$\begin{cases} H_0 : \text{received signal} = \text{noise} \\ H_1 : \text{received signal} = \text{target signal} + \text{noise} \end{cases}$$

$$\Rightarrow \begin{cases} H_0 : \mathbf{x} = \mathbf{y} & \text{No target} \\ H_1 : \mathbf{x} = a\mathbf{p} + \mathbf{y} & \text{Presence of target} \\ \|\mathbf{p}\| = 1 \end{cases}$$

- Clutter model: Compound Gaussian distribution

$$\mathbf{y}_i = \sqrt{\tau_i}\mathbf{z}_i \quad \text{where} \quad \begin{cases} \tau_i & \tau_i \text{ heavy-tailed} \\ \mathbf{z}_i & \text{Gaussian } \mathbf{C}_p \end{cases}$$

- If $\mathbf{C}_p$ is known up to a scale factor, the GLRT principle leads to the following detector (Normalized matched filter)

$$T_p \triangleq \frac{\left|\mathbf{x}^H \mathbf{C}_p^{-1} \mathbf{p}\right|}{\sqrt{\mathbf{p}^* \mathbf{C}_p^{-1} \mathbf{p}}\sqrt{\mathbf{x}^H \mathbf{C}_p^{-1} \mathbf{x}}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \Gamma$$

Adaptive Normalized Matched Filter (ANMF) detector

- In practice, matrix $\mathbf{C}_p$ is unknown.

$\Longrightarrow$ We assume that we have n observations containing only noise. These data are often called secondary data.

- Matrix $\mathbf{C}_p$ is estimated using the regularized Tyler estimator $\widehat{\mathbf{C}}_n(\rho)$ given by:

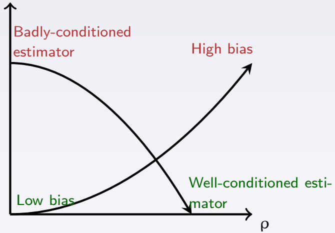
$$\widehat{\mathbf{C}}_n(\rho) = \frac{(1-\rho)}{n} \sum_{i=1}^{n} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H \widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}_i} + \rho \mathbf{I}_p$$

- The statistics T_p is replaced by:

$$\widehat{\mathbf{T}}_n = \triangleq \frac{\left|\mathbf{x}^H \widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{p}\right|}{\sqrt{\mathbf{p}^H \widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{p}}\sqrt{\mathbf{x}^H \widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{x}}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \Gamma$$

Choice of the regularization parameter

- The parameter ρ plays an important role in the performances of the methods using the regularized Tyler estimator.



? How to find the ρ that satisfies good balance between bias and conditioning of the estimator

- Classical methods for choosing ρ

Select the value that minimizes a certain loss function.

- Drawbacks of this approach: Generic values that are not necessarily adapted to the underlying application.

Optimal design of the ANMF detector

- Select the parameter ρ and the threshold Γ such that:
  - Maintain the probability of false alarm at a fixed rate P_{fa}

$$P_{\mathrm{fa}} = \mathbb{P}\left[\widehat{T}_n \geqslant \Gamma | H_0\right]$$

  - Maximize the probability of detection:

$$P_{\mathrm{d}} \triangleq \mathbb{P}\left[\widehat{T}_n \geqslant \Gamma | H_1\right]$$

$\Longrightarrow$ Need to study the fluctuations of $\widehat{T}_n$.

Asymptotic False alarm probability and probability of detection

- **Initial observations**
  - If the threshold Γ is taken fixed, then as $n, p \to \infty$ with $\frac{p}{n} \to c$ and under H_0

$$\widehat{T}_n \xrightarrow{\text{a.s.}} 0.$$

$\Longrightarrow$ Trivial result of little interest!

  - Select the parameters such that we avoid empty statements, $P_{\mathrm{fa}} \to 0$ and $P_{\mathrm{d}} \to 1$.
    - Consider $\Gamma = \frac{r}{\sqrt{p}}$.

$$P_{\mathrm{fa}} = \mathbb{P}\left(\widehat{T}_n \geqslant \Gamma\right) = \mathbb{P}\left(\sqrt{p}\widehat{T}_n \geqslant r | H_0\right) \quad \text{of order 1.}$$

    - Evaluate the detection probability: If $\|p\| = 1$ and $a = \mathcal{O}(1)$.

$$P_{\mathrm{d}} = \mathbb{P}\left(\widehat{T}_n \geqslant \Gamma\right) = \mathbb{P}\left(\sqrt{p}\widehat{T}_n \geqslant r | H_1\right) \quad \text{of order 1.}$$

- **Techniques of calculation** Since

$$\sup_{\rho \in \mathcal{R}_\kappa} p^{1-\epsilon} \left| \mathbf{a}^H \widehat{\mathbf{C}}_n^{-1} \mathbf{b} - \mathbf{a}^H \widehat{\mathbf{S}}_n^{-1} \mathbf{b} \right| \to 0$$

the fluctuations of $\widehat{T}_n$ are the same as $\widetilde{T}_n$ where $\widehat{\mathbf{C}}_n$ is replaced by $\widehat{S}_n$:

$$\widetilde{T}_n = \frac{\left| \mathbf{x}^H \widehat{\mathbf{S}}_n^{-1}(\rho) \mathbf{p} \right|}{\sqrt{\mathbf{p}^H \widehat{\mathbf{S}}_n^{-1}(\rho) \mathbf{p}} \sqrt{\mathbf{x}^H \widehat{\mathbf{S}}_n^{-1}(\rho) \mathbf{x}}}$$

Asymptotic false alarm and detection probabilities

- Recall

$$\sqrt{p}\,\widetilde{T}_n = \frac{\left|\sqrt{p}\mathbf{x}^H\widehat{\mathbf{S}}_n^{-1}(\rho)\mathbf{p}\right|}{\sqrt{\mathbf{p}^H\widehat{\mathbf{S}}_n^{-1}(\rho)\mathbf{p}}\sqrt{\mathbf{x}^H\widehat{\mathbf{S}}_n^{-1}(\rho)\mathbf{x}}}$$

- The fluctuations are due to $\left|\sqrt{p}\mathbf{x}^H\widehat{\mathbf{S}}_n^{-1}(\rho)\mathbf{p}\right|$.

$\Longrightarrow$ As $n, p \to \infty$ and condition on τ

- Under H_0: $\sqrt{p}\,\widehat{T}_n \sim \|X\|$ with $X \sim N(0, \sigma_p^2\mathbf{I}_2)$.
- Under H_1: $\sqrt{p}\,\widehat{T}_n \sim \|X\|$ with $X \sim N(\mu_p, \sigma_p^2\mathbf{I}_2)$

$$\sigma_p^2 = \frac{1}{2}\frac{\mathbf{p}^H\mathbf{C}_p\mathbf{Q}_p^2(\tilde{\rho})\mathbf{p}}{\mathbf{p}^H\mathbf{Q}_p(\tilde{\rho})\mathbf{p}}\left(\frac{1}{p}\operatorname{tr}\mathbf{C}_p\mathbf{Q}_p(\tilde{\rho})\left(1 - c(1-\tilde{\rho})^2 m(-\tilde{\rho})^2\frac{1}{p}\operatorname{tr}\mathbf{C}_p^2\mathbf{Q}_p^2(\tilde{\rho})\right)\right)^{-1}$$

$$\mu_p = \left[\frac{a}{\sqrt{\tau}}\frac{\sqrt{\mathbf{p}^H\mathbf{Q}_p(\rho)\mathbf{p}}}{\sqrt{\frac{1}{p}\operatorname{tr}\mathbf{C}_p\mathbf{Q}_p(\rho)}}, 0\right]^T$$

where $m(-\rho)$ is the unique solution to:

$$m(-\rho) = \left(\rho + \frac{c(1-\rho)}{p}\operatorname{tr}\mathbf{C}_p\left(\mathbf{I}_p + (1-\rho)m(-\rho)\mathbf{C}_p\right)^{-1}\right)^{-1}$$

with $\mathbf{Q}_p(\rho) = \left(\mathbf{I}_p + (1-\rho)m(-\rho)\mathbf{C}_p\right)^{-1}$ and $\tilde{\rho} = \rho\left(\rho + \frac{1}{\gamma_N(\rho)}\frac{1-\rho}{1-(1-\rho)c}\right)^{-1}$.

False alarm performance

Theorem (Asymptotic detector performance)

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

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| P\left(\widehat{T}_n(\rho) > \frac{\gamma}{\sqrt{p}} \right) - \exp\left(-\frac{\gamma^2}{2\sigma_p^2(\underline{\rho})} \right) \right| \to 0$$

where $\rho \mapsto \tilde{\rho} = \rho \left(\rho + \frac{1}{\gamma(\rho)} \frac{1-\rho}{1-(1-\rho)c} \right)$ *and*

$$\sigma_p^2(\tilde{\rho}) \triangleq \frac{1}{2} \frac{\mathbf{p}^H \mathbf{C}_p \mathbf{Q}_p^2(\tilde{\rho}) \mathbf{p}}{\mathbf{p}^H \mathbf{Q}_p(\hat{\rho}) \mathbf{p} \cdot \frac{1}{p} tr \mathbf{C}_p \mathbf{Q}_p(\tilde{\rho}) \cdot \left(1 - c(1-\tilde{\rho})^2 m(-\tilde{\rho})^2 \frac{1}{p} tr \mathbf{C}_p^2 \mathbf{Q}_p^2(\tilde{\rho}) \right)}$$

with $\mathbf{Q}_p(\tilde{\rho}) \triangleq (\mathbf{I}_p + (1-\tilde{\rho}) m(-\tilde{\rho}) \mathbf{C}_p)^{-1}$.

False alarm performance

Theorem (Asymptotic detector performance)

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

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| P\left(\widehat{T}_n(\rho) > \frac{\gamma}{\sqrt{p}} \right) - \exp\left(-\frac{\gamma^2}{2\sigma_p^2(\underline{\rho})} \right) \right| \to 0$$

where $\rho \mapsto \tilde{\rho} = \rho \left(\rho + \frac{1}{\gamma(\rho)} \frac{1-\rho}{1-(1-\rho)c} \right)$ *and*

$$\sigma_p^2(\tilde{\rho}) \triangleq \frac{1}{2} \frac{\mathbf{p}^H \mathbf{C}_p \mathbf{Q}_p^2(\tilde{\rho}) \mathbf{p}}{\mathbf{p}^H \mathbf{Q}_p(\hat{\rho}) \mathbf{p} \cdot \frac{1}{p} tr \mathbf{C}_p \mathbf{Q}_p(\tilde{\rho}) \cdot \left(1 - c(1-\tilde{\rho})^2 m(-\tilde{\rho})^2 \frac{1}{p} tr \mathbf{C}_p^2 \mathbf{Q}_p^2(\tilde{\rho}) \right)}$$

with $\mathbf{Q}_p(\tilde{\rho}) \triangleq (\mathbf{I}_p + (1-\tilde{\rho}) m(-\tilde{\rho}) \mathbf{C}_p)^{-1}$.

- Limiting Rayleigh distribution
  $\Rightarrow$ Weak convergence to Rayleigh variable $R_N(\hat{\rho})$

Detection probability performance

Theorem (Detection probability)

As $p, n \to \infty$ with $c_n = \frac{p}{n} \to c \in (0, \infty)$,

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| \mathbb{P}\left[\widehat{T}_n(\rho) > \frac{r}{\sqrt{p}} | H_1 \right] - \mathbb{E}\left[Q_1 \left(g(\rho), \frac{r}{\sigma_p(\rho)} \right) \right] \right| \to 0,$$

where the expectation is taken over the distribution of τ, $\sigma_p(\rho)$ has the same expression as in the previous theorem and

$$g(\rho) = \frac{\sqrt{1 - c(1-\tilde{\rho})^2 m(-\tilde{\rho}) \frac{1}{p} tr \mathbf{C}_p^2 \mathbf{Q}_p^2(\tilde{\rho})}}{\sqrt{\mathbf{p}^H \mathbf{C}_p \mathbf{Q}_p^2(\tilde{\rho}) \mathbf{p}}} \sqrt{\frac{2}{\tau}} a \left| \mathbf{p}^H \mathbf{Q}_p(\tilde{\rho}) \mathbf{p} \right|.$$

and Q_1 is the Marcum Q-function.

Optimal design of the ANMF detector

- To maximize the probability of detection, ρ should be set such that:

$$\rho^* = \arg\max_{\rho} \frac{\mu_p}{\sigma_p} = \arg\max_{\rho} f_p(\rho) \tag{3}$$

with

$$f_p(\rho) = \frac{\left(1 - c(1-\tilde{\rho})^2 m(-\tilde{\rho}) \frac{1}{p} \operatorname{tr} \mathbf{C}_p^2 \mathbf{Q}_p^2(\tilde{\rho})\right) \left(\mathbf{p}^H \mathbf{Q}_p(\tilde{\rho}) \mathbf{p}\right)^2}{\mathbf{p}^H \mathbf{C}_p \mathbf{Q}_p^2(\tilde{\rho}) \mathbf{p}}$$

- For false alarm probability η, test statistics threshold r set to:

$$r = \sigma_p(\rho) \sqrt{-2 \log \eta}$$

$\Rightarrow$ **Optimal design**

- Select ρ^* solution of (3)
- Select the threshold $r = \sigma_N(\rho^*) \sqrt{-2 \log \eta}$

✗ Optimal values depend on unknown $\mathbf{C}_N$. We need to build consistent estimates

Optimal design of the ANMF detector

Consistent estimates of $\sigma_p(\rho)$ and $f_p(\rho)$

- Define $\hat{f}_p(\rho)$ and $\hat{\sigma}_p^2(\rho)$:

$$\hat{f}_p(\rho) = \left(\mathbf{p}^H \widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{p}\right)^2 \left(\frac{1}{p}\operatorname{tr}\widehat{\mathbf{C}}_n(\rho) - \rho\right) \frac{(1-c_n+c_n\rho)^2}{\left(\mathbf{p}^H\widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{p} - \rho\mathbf{p}^H\widehat{\mathbf{C}}_n^{-2}(\rho)\mathbf{p}\right)}$$

$$\hat{\sigma}_n^2(\rho) = \frac{1}{2}\frac{1-\rho\frac{\mathbf{p}^H\widehat{\mathbf{C}}_n^{-2}(\rho)\mathbf{p}}{\mathbf{p}^H\widehat{\mathbf{C}}_n^{-1}(\rho)\mathbf{p}}}{(1-c_n+c_n\rho)(1-\rho)}$$

- Then, we have:

$$\sup_{\rho\in\mathcal{R}_\kappa}\left|\hat{\sigma}_p(\rho)-\sigma_p(\rho)\right| \xrightarrow{\text{a.s.}} 0,$$

$$\sup_{\rho\in\mathcal{R}_\kappa}\left|\hat{f}_p(\rho)-f_p(\rho)\right| \xrightarrow{\text{a.s.}} 0$$

Optimal design of the ANMF detector

- Select ρ such that:

$$\hat{\rho} = \arg\max_{\rho} \hat{f}_p(\rho)$$

- Select the threshold such that:

$$\hat{r} = \hat{\sigma}_p(\rho)\sqrt{-2\log\eta}$$

Numerical illustration

Experiment Setting

- N_a number of samples
- N_p number of pulses
- $\mathbf{p} = \mathbf{a}_{N_p}(f_d) \otimes \mathbf{a}_{N_a}(f_s)$ where $\mathbf{a}_k(f) = \{\exp(\jmath 2\pi(\ell-1)f)\}_{\ell=1}^{k}$
- $\mathbf{C}_N \propto \left(\mathbf{I}_N + \sum_{i=1}^{N_\phi} \sigma^2 \mathbf{A}_{N_p}(f_{d_i}, f_{s_i}) \mathbf{A}_{N_p}(f_{d_i}, f_{s_i})^H\right)$ with $\mathbf{A}_{N_p}(f_{d_i}, f_{s_i}) = \mathbf{a}_{N_p}(f_{d_i}) \otimes \mathbf{a}_{N_a}(f_{s_i})$
- K-distributed clutter with zero meand and shape $\nu = 4.5$
- Compare with $\hat{\rho}_{\text{ollila}}$ given by:

$$\hat{\rho}_{\text{ollila}} = \frac{N \operatorname{tr} \widehat{\mathbf{C}}_0 - 1}{N \operatorname{tr} \widehat{\mathbf{C}}_0 - 1 + n(N+1)\left(N^{-1} \operatorname{tr} \widehat{\mathbf{C}}_0^{-2} - 1\right)}$$

where $\widehat{\mathbf{C}}_0$ is the conventional Tyler estimator.

ROC curve

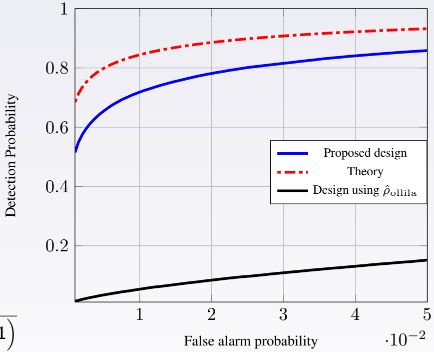


Fig. 1: ROC curves for non Gaussian clutters when $N = 128$ ($N_a = 10$, $N_p = 25$), $n = 128$, $f_s = 0.5$, $f_d = 0.2$, $a = 0.3$

Conclusion

- Provide recent results on robust statistics in classical and RMT regimes
- Discuss the impact of these results on several applications : Source localization and radar detection
- Show how to adapt the tools from random matrix theory to the field of robust statsitics.

Open questions

Classical regime:

- Extend the asymptotic convergence tools to other robust-scatter estimators : regularized robust-scatter estimators, normalized robust scatter estimators
- Exploit these results to understand the impact of the normalization on the performances of normalized robust-scatter estimators.

Random matrix theory regime

- Study linear statistics of the robust-scatter M estimator.
- Equivalence between $\widehat{\mathbf{C}}_n$ and $\widehat{\mathbf{S}}_n$ suggests that good performances should be also obtained when considering the following estimator:

$$\overline{\mathbf{S}} = \frac{(1-\rho)}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i\mathbf{x}_i^H}{\frac{1}{p}\mathbf{x}_i^H\mathbf{x}_i} + \rho\mathbf{I}_p.$$

- Study the performances of methods using $\overline{\mathbf{S}}$ instead of $\widehat{\mathbf{C}}_n$
- Extend these results to observations not necessarily following CES distributions, example arbitrary deterministic outliers.

D. Morales-Jimenez, R. Couillet and M. R. Mckay "Large Dimensional Analysis of Robust M-estimators of Covariance with outliers" IEEE Transactions on Signal Processing 2015.

For both regimes

- Study the joint mean and scatter robust estimators.

$\Longrightarrow$ Application hyperspectral imaging.