Large Random Matrices and High Dimensional Statistical Signal Processing
Summer School, Telecom ParisTech

Abla KAMMOUN

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

- **Wireless Communication**
  - Since 1990
    - Performance analysis
    - Optimal transceiver design

- **Signal Processing**
  - Since 2007
    - Estimation
    - Detection

- **Machine learning**
  - Since 2015
    - Subspace clustering
Outline

Part I: Robust statistics/
  - Motivation
  - Distribution models
  - Maximum Likelihood estimators
  - M-estimators of scatters
  - Regularized Robust estimators

Part II: Random Matrix Theory for robust estimation/
  - Review of random matrix theory results
    - Detection
    - Estimation
  - M-scatter estimator in the large random matrix regime
    - Eigenvalue localization
    - Source localization
  - Regularized estimators
  - Application: Radar detection
Outline

Part I: Robust statistics/ Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
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,

⇒ It is of fundamental interest to develop robust signal processing methods
Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
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.

\[ z \sim \mathcal{CN}(0, I_p) \]

- \( \tau \) is a scalar used to model the impulsive behaviour of the noise,
- \( \frac{z}{\|z\|} \) has a uniform distribution over the sphere,
- This representation is not unique

\[ x = \sqrt{\tau} \times C_p^{\frac{1}{2}} \times \frac{z}{\|z\|} \]

- Matrix \( C_p \) is called the scatter matrix. To remove scalar ambiguity, the scatter matrix is selected such that \( \text{tr} \ C_p = p \).
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:

$$f(\mathbf{x}) = C_{p,g} \left( \det(C_p)^{-1} \right) g(\mathbf{x}^H C_p^{-1} \mathbf{x})$$

(*)

where $C_p$ 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

$$\mathbf{x} \sim CES(C_p, g)$$

(1)

Link between two representations of CES random variables

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

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

- Gaussian distribution:

\[ x = \text{Gam}(p, 1) \times C_{\frac{1}{p}}^{1} \frac{z}{\|z\|} \]

- Compound Gaussian distribution: sub-family of CES distributions

\[ z \sim \mathcal{C}\mathcal{N}(0, I_p) \quad \| \quad \tau \text{ (scalar)} \]

\[ x = \sqrt{\tau} \times C_{\frac{1}{p}}^{1} \times z \]

Stochastic compound Gaussian representation:

\[ x = \sqrt{\tau} \times C_{\frac{1}{p}}^{1} \times z = \sqrt{\tau \|z\|^2} \times C_{\frac{1}{p}}^{1} \frac{z}{\|z\|} \]

- \( \tau > 0 \) is called the texture and is independent of \( z \),
- \( z \) is called the speckle.
Part I: Robust statistics/ Distribution models

Compound Gaussian distribution

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

\[
f(x) = \pi^{-p} \det(C_N^{-1}) \int_0^\infty \tau^{-p} \exp\left(-x^H C_N^{-1} x / \tau\right) f_\tau(\tau) \, d\tau
\]

- The density generator associated with \( x \) is given by:

\[
g(t) = \int_0^\infty \tau^{-p} \exp\left(-\frac{t}{\tau}\right) f_\tau(\tau) \, d\tau
\]

- Most illustrative examples:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Pdf</th>
<th>heavier tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-distribution</td>
<td>( x = \sqrt{\text{Gam}(\nu, \nu^{-1})} \times z )</td>
<td>Tail heavy when ( \nu \uparrow )</td>
</tr>
<tr>
<td>Complex t-distribution</td>
<td>( x = \sqrt{\frac{1}{\text{Gam}(\nu/2, 2\nu^{-1})}} \times z )</td>
<td>Tail heavy when ( \nu \downarrow )</td>
</tr>
<tr>
<td>Generalized Gaussian distribution</td>
<td>( x = \text{Gam}(\frac{p}{s}, b) \times z )</td>
<td>Tail heavy when ( s \downarrow )</td>
</tr>
</tbody>
</table>
Outline

Part I: Robust statistics/
  Motivation
  Distribution models
  **Maximum Likelihood estimators**
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
Problem statement: Given $p$-dimensional random vectors $x_1, \cdots, x_n$ drawn from elliptical distribution with density generator $g$ and scatter matrix $C_p$.

Objective: estimate the scatter matrix $C_p$.

Traditional estimator is the sample covariance matrix:

$$\hat{S}_p = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H$$

Its popularity owes to:

- Simplicity,
- Existing of a good understanding of its behaviour in the regimes:
  - $n \to \infty$ and $p$ fixed
  - $n, p \to \infty$ with $\frac{p}{n} \to c$
- 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 $x_1, \cdots, x_n \in \mathbb{C}^p$
  - independent and identically distributed
  - centered
  - $x_i \sim CES(C_p, g)$ with pdf $f(x)$.
- The Likelihood is given by:

$$L(C_p) = \prod_{i=1}^{n} f(x_i) \propto \left( \det(C_p^{-1}) \right)^n \prod_{i=1}^{n} g(x_i^H C_p^{-1} x_i)$$

- The negative log-likelihood function

$$\mathcal{L}_n(C_p) \propto \sum_{i=1}^{n} -\log g \left( x_i^H C_p^{-1} x_i \right) + n \log \det(C_p)$$

- The MLE estimator is given by:

$$\hat{C}_n = \arg \min_{C_p \in \mathbb{C}^p} \text{definite positive} \mathcal{L}_n(C_p)$$

$$\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi \left( x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H, \quad \varphi = -g'(t)/g(t)$$
Maximum Likelihood estimator (MLE) for the scatter

- **Gaussian case**
  - $x_1, \cdots, x_n \sim \mathcal{CN}(0, C_p) \implies x_1, \cdots, x_n \sim CES(C_p, \exp(-t))$
  - Density generator is $g(t) = \exp(-t) \implies \varphi(t) = \exp(-t)/\exp(-t) = 1.$

  
  $$\implies \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H$$

- **CES case:**

  $$\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi \left( x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H, \quad \varphi = -g'(t)/g(t)$$

**Natural questions:**
- Existence,
- Uniqueness,
- Numerical evaluation
- Asymptotic behaviour
Examples of Maximum likelihood estimators

- **Complex t-distribution** \( \mathbb{C}t_{p,\nu} \)

  \[
  \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi \left( x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H, \quad \varphi = \frac{2p + \nu}{\nu + 2t}
  \]

- **Generalized Gaussian distribution** \( \mathbb{C}GG_{p,s} \)

  \[
  \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi \left( x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H, \quad \varphi = \frac{s}{b} t^{s-1}
  \]

- **Central Angular Gaussian distribution** \( \mathbb{C}AG_p(C_N) \)

  \[
  \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi \left( x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H, \quad \varphi (t) = \frac{p}{t}
  \]

  \[
  = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\frac{1}{p} x_i^H \hat{C}_n^{-1} x_i}
  \]
Part I: Robust statistics/

Motivation
Distribution models
Maximum Likelihood estimators
**M-estimators of scatters**
Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/

Review of random matrix theory results
  Detection
  Estimation

M-scatter estimator in the large random matrix regime
  Eigenvalue localization
  Source localization

Regularized estimators
Application: Radar detection
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 $\varphi$ not necessarily related to any elliptical density $g$.
- Motivation behind the M-estimator: the probability density function might not be known.

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

<table>
<thead>
<tr>
<th>SCM</th>
<th>Huber</th>
<th>Tyler</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u(x) = 1$</td>
<td>$u(x) = \begin{cases} K/e &amp; \text{if } x \leq e \ K/x &amp; \text{if } x \geq e \end{cases}$</td>
<td>$u(x) = \frac{p}{x}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$u(x)$</th>
<th>$u(x)$</th>
<th>$u(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Graph SCM" /></td>
<td><img src="image" alt="Graph Huber" /></td>
<td><img src="image" alt="Graph Tyler" /></td>
</tr>
</tbody>
</table>
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.
    - 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
    - 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**
M-estimators of scatter matrix: Existence and uniqueness

**Assumptions**

- Let $x_1, \ldots, x_n \in \mathbb{C}^p$
- $x \mapsto u(x)$ is non-negative, continuous and non-increasing,
- Consider $x \mapsto \phi(x) = xu(x)$ strictly increasing.
- Let $K = \sup_{s \geq 0} \phi(s)$, then $p < K$.
- There exists $a > 0$ such that for any hyperplane $S$ satisfying $\dim(S) \leq p - 1$, we have $\frac{\#(x_i \in S)}{n} \leq 1 - \frac{p}{K} - a$.

**M-estimator of scatter:** Then, the solution of the following equation in $\hat{C}_n$

$$\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_n^{-1} x_i) x_i x_i^H$$

exists and is unique.

**Numerical evaluation:** Moreover, given any initial estimate $\Sigma$ hermitian and positive definite, the following sequence $(\Sigma_k)$ defined as:

$$\Sigma_0 = \Sigma$$
$$\Sigma_{k+1} = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \Sigma_k^{-1} x_i) x_i x_i^H$$

converges to $\hat{C}_n$. 
M-estimators of scatter matrix: Asymptotic convergence


Assumptions

- Let \( x_1, \ldots, x_n \in \mathbb{C}^p \sim CES(C_p, g) \)
- \( x \mapsto u(x) \) is non-negative, continuous and non-increasing,
- Consider \( x \mapsto \phi(x) = x u(x) \) strictly increasing.
- Let \( K = \sup_{s \geq 0} \phi(s) \), then \( p < K \).
- There exists \( a > 0 \) such that for any hyperplane \( S \) satisfying \( \dim(S) \leq p - 1 \), we have
  \[
  \frac{\#(x_i \in S)}{n} \leq 1 - \frac{p}{K} - a.
  \]
- Asymptotic regime: \( n \to \infty \) with \( p \) fixed.

Convergence of the M-estimator of scatter:

Let \( C_\phi \) be the solution of the following equation:

\[
C_\phi = \mathbb{E} \left[ u(x^H C_\phi^{-1} x) xx^H \right]
\]

Then,

\[
\hat{C}_n \to C_\phi
\]

Moreover,

\[
C_\phi = \sigma C_p
\]

with \( \sigma \) 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:

\[ \widehat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \widehat{C}_n^{-1} x_i) x_i x_i^H \]
M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

\[ \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_n^{-1} x_i)x_i^H \]

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

\[ \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_n^{-1} x_i)x_i^H \sim \frac{1}{n} \sum_{i=1}^{n} u(x_i^H C_\phi^{-1} x_i)x_i^H \sim \mathbb{E} \left[ u(x^H C_\phi^{-1} x)xx^H \right] \]
M-estimators of scatter matrix: Heuristic arguments

Recall that the robust estimator is solution to:

\[ \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_n^{-1} x_i) x_i x_i^H \]

Since \( p \) is assumed fixed, \( \hat{C}_n \) can be considered as close to a deterministic matrix \( C_\phi \). By the strong law of large numbers, we expect that:

\[ \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_n^{-1} x_i) x_i x_i^H \sim \frac{1}{n} \sum_{i=1}^{n} u(x_i^H C_\phi^{-1} x_i) x_i x_i^H \sim \mathbb{E} \left[ u(x^H C_\phi^{-1} x) xx^H \right] \]

\[ \implies \text{Matrix } C_\phi \text{ should thus satisfy:} \]

\[ C_\phi = \mathbb{E} \left[ u(x^H C_\phi^{-1} x) xx^H \right] \]
M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

\[
\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_{n}^{-1} x_i) x_i x_i^H
\]

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

\[
\frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_{n}^{-1} x_i) x_i x_i^H \sim \frac{1}{n} \sum_{i=1}^{n} u(x_i^H C_{\phi}^{-1} x_i) x_i x_i^H \sim \mathbb{E} \left[ u(x^H C_{\phi}^{-1} x) xx^H \right]
\]

\( \implies \) Matrix \( C_{\phi} \) should thus satisfy:

\[
C_{\phi} = \mathbb{E} \left[ u(x^H C_{\phi}^{-1} x) xx^H \right]
\]

- Now, multiplying both sides by \( C_{\phi}^{-1} \), we obtain:

\[
I_p = \mathbb{E} \left[ u(x^H C_{\phi}^{-1} x) xx^H C_{\phi}^{-1} \right]
\]
M-estimators of scatter matrix: Heuristic arguments

- Recall that the robust estimator is solution to:

\[ \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_n^{-1} x_i) x_i x_i^H \]

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

\[
\frac{1}{n} \sum_{i=1}^{n} u(x_i^H \hat{C}_n^{-1} x_i) x_i x_i^H \sim \frac{1}{n} \sum_{i=1}^{n} u(x_i^H C_\phi^{-1} x_i) x_i x_i^H \sim \mathbb{E} \left[ u(x^H C_\phi^{-1} x) x x^H \right]
\]

\[ \implies \text{Matrix } C_\phi \text{ should thus satisfy:} \]

\[ C_\phi = \mathbb{E} \left[ u(x^H C_\phi^{-1} x) x x^H \right] \]

- Now, multiplying both sides by \( C_\phi^{-1} \), we obtain:

\[ I_p = \mathbb{E} \left[ u(x^H C_\phi^{-1} x) x x^H C_\phi^{-1} \right] \]

- Taking the trace and replacing \( C_\phi \) by \( \sigma C_p \), we finally obtain:

\[ p = \mathbb{E} \left[ u(x^H C_\phi^{-1} x) x^H C_\phi^{-1} x \right] = \mathbb{E} \left[ \phi \left( \frac{x^H C_\phi^{-1} x}{\sigma} \right) \right] \]
M-estimators of scatter matrix: Fluctuations


Assumptions

- Let \( x_1, \ldots, x_n \in \mathbb{C}^p \sim CES(\mathcal{C}_N, g) \)
- \( x \mapsto u(x) \) is non-negative, continuous and non-increasing,
- Consider \( x \mapsto \phi(x) = xu(x) \) strictly increasing.
- Let \( K = \sup_{s \geq 0} \phi(s) \), then \( p < K \).
- There exists \( a > 0 \) such that for any hyperplane \( S \) satisfying \( \dim(S) \leq p - 1 \), we have \( \frac{\#(x_i \in S)}{n} \leq 1 - \frac{p}{K} - a. \)
- Asymptotic regime: \( n \to \infty \) with \( p \) fixed.

Then,

\[
\sqrt{n} \text{vec} \left( \hat{\mathbf{C}}_n - \mathbf{C}_\phi \right) \xrightarrow{d} \mathcal{CN}_{p^2}(\mathbf{0}, \mathbf{C}, \mathbf{P})
\]

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

\[
\mathbf{C} = \vartheta_1 (\mathbf{C}_p^* \otimes \mathbf{C}_p) + \vartheta_2 \text{vec}(\mathbf{C}_p)\text{vec}(\mathbf{C}_p)^H
\]

\[
\mathbf{P} = \vartheta_1 (\mathbf{C}_p^* \otimes \mathbf{C}_p) \mathbf{K}_{p,p} + \vartheta_2 \text{vec}(\mathbf{C}_p)\text{vec}(\mathbf{C}_p)^T
\]

with \( \mathbf{K} \) being the commutation matrix. The constants \( \vartheta_1 \) and \( \vartheta_2 \) depends solely on the distribution of \( \tau \) and function \( u \).
Tyler estimators

D. E. Tyler ” A distribution-free M-estimator of multivariate scatter” The Annals of statistics 1987

Assumptions

- $x_1, \ldots, x_n \in \mathbb{C}^{p \times 1}$ such that $n > p$.
- $\text{span}\{x_i\} = \mathbb{C}^p$.
- Tyler estimator is defined by $\hat{C}_n$ the solution of

$$\hat{C}_n = \sum_{i=1}^{n} \frac{p}{n} \frac{x_i x_i^H}{x_i^H \hat{C}_n^{-1} x_i}$$

such that $\frac{1}{p} \text{tr} \hat{C}_n = 1$.

where $x_1, \ldots, x_n$ are independent and follow CES distribution with scatter $C_p$.

- If $x_1, \ldots, x_n$ are independent and follow CES distribution, then tyler estimator is the MLE of scatter of the random vectors $\frac{x_1}{\|x_1\|}, \ldots, \frac{x_n}{\|x_n\|}$.
Tyler estimators


- **Assumptions**
  - $x_1, \cdots, x_n \in \mathbb{C}^{p \times 1}$ independent and follow CES distribution with scatter $C_p$ ($\frac{1}{p} \text{tr} C_p = 1$)
  - $n > p$.
  - $\text{span } \{x_i\} = \mathbb{C}^p$.

- Tyler estimator is consistent:
  \[
  \hat{C}_n \xrightarrow{\text{a.s.}} \xrightarrow{n \to \infty, p \text{fixed}} C_p
  \]

- Fluctuations of Tyler estimators:
  \[
  \sqrt{n} \text{vec} \left( \hat{C}_n - C_p \right) \xrightarrow{d} \xrightarrow{n \to \infty, p \text{ fixed}} C N \left( \mathbf{0}_{p^2}, C, P \right)
  \]
  where
  \[
  C = \varrho_1 C_p^T \otimes C_p + \varrho_2 \text{vec}(C_p) \text{vec}(C_p)^H
  \]
  \[
  P = \varrho_1 \left( C_p^T \otimes C_p \right) K_{p,p} + \varrho_2 \text{vec}(C_p) \text{vec}(C_p)^T
  \]
  with $K_p$ is the commutation matrix, $\varrho_1 = \frac{p+1}{p}$ and $\varrho_2 = -\frac{p+1}{p^2}$.
Comparison with fluctuations of SCM: Some comments

- Fluctuations of the sample covariance matrix (SCM) when \( x_1, \cdots, x_n \) follow Gaussian distribution with zero mean and covariance \( C_p \):

\[
\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H
\]

Then:

\[
\sqrt{n} \text{vec} \left( \hat{S}_n - C_p \right) \xrightarrow{d} \mathcal{CN}(0, C_{\text{scm}}, P_{\text{scm}})
\]

where

\[
C_{\text{scm}} = (C_p^* \otimes C_p) \quad P_{\text{scm}} = (C_p^* \otimes C_p) K_{p,p}
\]

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

\[
\sqrt{n} \text{vec} \left( \hat{C}_n - C_\phi \right) \xrightarrow{d} \mathcal{CN}_2(0, C, P)
\]

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

\[
C = \vartheta_1 (C_p^* \otimes C_p) + \vartheta_2 \text{vec}(C_p) \text{vec}(C_p)^H
\]

\[
P = \vartheta_1 (C_p^* \otimes C_p) K_{p,p} + \vartheta_2 \text{vec}(C_p) \text{vec}(C_p)^T
\]
Comparison with fluctuations of SCM: Some comments

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

$$\sqrt{n} \left( H(\hat{C}_n) - H(C_\phi) \right) \overset{d}{\rightarrow} \mathcal{CN}(0, \alpha_M, \beta_M)$$

where

$$\alpha_M = \vartheta_1 H'(C_p)^T \left( C_p^T \otimes C_p \right) H'(C_p)$$

$$\beta_M = \vartheta_1 H'(C_p)^T \left( C_p^T \otimes C_p \right) K_{p,p} H'(C_p)$$

Applying the Delta method for the SCM, we get:

$$\sqrt{n} \left( H(\hat{S}_n) - H(C_p) \right) \overset{d}{\rightarrow} \mathcal{CN}(0, \alpha_{scm}, \beta_{scm})$$

where

$$\alpha_{scm} = H'(C_p)^T \left( C_p^T \otimes C_p \right) H'(C_p)$$

$$\beta_{scm} = H'(C_p)^T \left( C_p^T \otimes C_p \right) K_{p,p} H'(C_p)$$
Comparison with fluctuations of SCM: Some comments

For functions satisfying $H(V) = H(\alpha V)$,

$H(\hat{C}_n)$ with $n$ observations has the same fluctuations of $H(\text{SCM})$ with $n\theta_1$ 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

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  **Regularized Robust estimators**

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  **Regularized estimators**
  Application: Radar detection
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.

⇒ 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:

\[ L_n(\Sigma) = \sum_{i=1}^{n} - \log g(x_i^H \Sigma^{-1} x_i) - n \log \det(\Sigma^{-1}) \]

- To increase the stability of the solution, it is common to introduce \( n \rho P(\Sigma) \) to the cost function. \( \iff \) The penalized cost function becomes:

\[ L_n(\Sigma) = \sum_{i=1}^{n} - \log g(x_i^H \Sigma^{-1} x_i) + n \log \det(\Sigma) + \rho n P(\Sigma) \]

where \( \rho \geq 0 \) is a regularization (shrinkage) parameter.

- This procedure can be extended to M-estimators of scatter, by using a general function \( \alpha(u = \alpha') \) in place of \( -\log g \), leading to:

\[ L_n(\Sigma) = \sum_{i=1}^{n} \alpha(x_i^H \Sigma^{-1} x_i) - n \log \det(\Sigma^{-1}) + \rho n P(\Sigma) \]
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 $\Sigma^{-1}$, we consider the penalty function:

$$P(\Sigma) = n \text{tr} \Sigma^{-1}$$

The penalized cost function becomes:

$$\mathcal{L}_\rho(\Sigma) = \sum_{i=1}^{n} \alpha(x_i^H \Sigma^{-1} x_i) - n \log(\det \Sigma^{-1}) + n \rho \text{tr} \Sigma^{-1} \quad (*)$$

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

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

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

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

▶ The regularized M-estimates do not require any condition on the sample $x_1, \cdots, x_n$

⇒ 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

\[
\Sigma_{k+1} = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H \Sigma_k^{-1} x_i) x_i x_i^H + \rho I_p
\]

converges to the regularized M-estimate \( \hat{C}_n \).
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 $u(x) = \frac{(1-\rho)p}{x}$.

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

- Since $\alpha$ 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
\[ \text{Cond. A : For any subspace } S \text{ of } \mathbb{C}^p \text{ such that } 1 \leq \dim(S) < p, \text{ the inequality } \frac{\# \{ x_i \in S \}}{n} < \frac{\dim(S)}{p(1-\rho)} \]
holds. Then:
\[ \hat{C}_n(\rho) = (1-\rho) \frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\hat{C}_n^{-1}(\rho) x_i} + \rho I_p \]
exists and is unique.

Comments. Assume that \( x_i \) are linearly independent. Then:
\[ \frac{\# \{ x_i \in S \}}{n} \leq \frac{p-1}{n} \]
Hence, \textbf{Cond A} is equivalent to \( \rho \geq 1 - \frac{n}{p} \).
Regularized Tyler estimator: Numerical evaluation

**Theorem**

Assume that

**Cond. A** : For any subspace $S$ of $\mathbb{C}^p$ such that $1 \leq \dim(S) < p$, the inequality

$$\frac{\#\{x_i \in S\}}{n} < \frac{\dim(S)}{p(1-\rho)}$$

holds. Then, the iterations:

$$\Sigma_{k+1} = (1 - \rho) \frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{x_i^H \Sigma_k^{-1} x_i} + \rho I_p$$

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

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:

$$
\tilde{\Sigma}_{k+1} = (1 - \rho) \frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{x_i^H \Sigma_k^{-1} x_i} + \rho I_p
$$

$$
\Sigma_{k+1} = \frac{\tilde{\Sigma}_{k+1}}{tr \tilde{\Sigma}_{k+1}/p}
$$

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

$$
\tilde{\mathcal{C}}_n(\rho) = \frac{\tilde{\mathcal{B}}_n(\rho)}{\frac{1}{p} tr \tilde{\mathcal{B}}_n(\rho)}
$$

with

$$
\tilde{\mathcal{B}}_n(\rho) = (1 - \rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\frac{1}{p} x_i^H \tilde{\mathcal{C}}_n^{-1}(\rho) x_i} + \rho I_p.
$$
Regularized Tyler estimator: Convergence


Assumptions: Take \( x_1, \cdots, x_n \in \mathbb{C}^{p \times 1} \), i.e, \( x_i = C_p^{1/2} w_i \) with:

- \( w_1, \cdots, w_p \in \mathbb{C}^{p \times 1} \) independent complex standard Gaussian vectors with zero mean \( 0_{p \times 1} \) and covariance \( I_p \),
- We assume that \( \frac{1}{p} \text{tr} C_p = 1 \) and that \( C_p \) is positive definite.

Regularized Tyler estimator: Let \( \rho \in \left( 0, \max\left( 1 - \frac{n}{p}, 1 \right) \right) \). Consider \( \hat{C}_n(\rho) \) solution to the following fixed point equation:

\[
\hat{C}_n(\rho) = (1 - \rho) \frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{x_i^H \hat{C}_n^{-1}(\rho) x_i} + \rho I_p.
\]

Let \( \Sigma_0(\rho) \) be the unique solution to the following equation:

\[
\Sigma_0(\rho) = p(1 - \rho) \mathbb{E} \left[ \frac{xx^H}{x^H \Sigma_0^{-1}(\rho) x} \right] + \rho I_p
\]

Then, for any \( \kappa > 0 \),

\[
\sup_{\rho \in [\kappa, 1]} \left\| \hat{C}_n(\rho) - \Sigma_0(\rho) \right\| \xrightarrow{a.s.} n \to \infty, p \text{fixed} 0.
\]
Regularized Tyler estimator: Characterization of $\Sigma_0(\rho)$


Evaluation of $\Sigma_0(\rho)$.

- Recall that $\Sigma_0(\rho)$ satisfies:

$$\Sigma_0(\rho) = p(1 - \rho)\mathbb{E} \left[ \frac{xx^H}{x^H \Sigma_0^{-1}(\rho)x} \right] + \rho I_p$$

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

$$p(1 - \rho)\mathbb{E} \left[ \frac{ww^H}{w^H C_p^\frac{1}{2} \Sigma_0^{-1}(\rho) C_p^\frac{1}{2} w} \right] + \rho C_p^{-1} = C_p^{-\frac{1}{2}} \Sigma_0(\rho) C_p^{-\frac{1}{2}}$$

- Let $C_p^\frac{1}{2} \Sigma_0^{-1}(\rho) C_p^\frac{1}{2} = VDV^H$ the eigenvalue decomposition of $C_p^\frac{1}{2} \Sigma_0^{-1}(\rho) C_p^\frac{1}{2}$

- Then,

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

$\Rightarrow \mathbb{E} \left[ \frac{ww^H}{w^H Dw} \right]$ is diagonal implying that matrix $\Sigma_0$ and $C_p$ share the same eigenvector space.
Regularized Tyler estimator: Characterization of $\Sigma_0(\rho)$


Lemma

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

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

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

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

where $F_D^{(p)}$ is the Lauricella's type $D$ hypergeometric function.
Regularized Tyler estimator: Convergence


Recall that:

\[ p(1 - \rho)\mathbb{E}\left[ \frac{ww^H}{w^HDw} \right] + \rho V^H C_p^{-1}V = D^{-1} \]

Let \( D = \text{diag}(d_1, \ldots, d_p) \). Denote by \( \alpha_i(\{d_i\}_{i=1}^p) = \mathbb{E}\left[ \frac{|w_i|^2}{w^HDw} \right] \). 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


Matrix $\Sigma_0$ can be computed as:

- Let $C_N = V\Lambda V^H$ be the eigenvalue decomposition of $C_N$ with $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_p)$ and $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p$.

- Start from $d_1^{(0)}, \cdots, d_p^{(0)}$. Compute iteratively:

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

until convergence. Let $d_1^{(0)}, \cdots, d_p^{(0)}$ be the obtained values after applying several iterations.

- Set $s_{i,\infty} = \lambda_i d_{i,\infty}$. Then

$$\Sigma_0 = V \text{diag} \left( [s_{1,\infty}, \cdots, s_{p,\infty}] \right) V^H$$
Regularized Tyler estimator: Hint on the proof


The proof is based on controlling the random elements:

\[ e_i(\rho) = \frac{x_i^H \hat{C}_n^{-1}(\rho)x_i - x_i^H \Sigma_0^{-1}(\rho)x_i}{\sqrt{x_i^H \Sigma_0^{-1}(\rho)x_i} \sqrt{x_i^H \hat{C}_n^{-1}(\rho)x_i}} \]

and showing that they converge to zero almost surely.

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

\[ \Sigma_0(\rho) \simeq p(1 - \rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{x_i^H \Sigma_0^{-1}(\rho)x_i} + \rho I_p \]

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

\[ \max_{1 \leq i \leq n} e_i(\rho) \left( 1 - \sqrt{\| I_p - \rho \Sigma_0^{-1} \|} - o(1) \right) \leq o(1) \]

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

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

\[ \left\| \hat{C}_n - \Sigma_0(\rho) \right\| \leq \max_{1 \leq i \leq n} e_i(\rho) + o(1) \]

thus showing that \( \left\| \hat{C}_n - \Sigma_0(\rho) \right\| \xrightarrow{\text{a.s.}} 0. \)
Regularized Tyler estimators: Fluctuations


Similarly to the M-scatter estimator, we can establish a CLT on $\sqrt{n}\text{vec} \left( \hat{C}_n(\rho) - \Sigma_0(\rho) \right)$. We prove that:

$$\sqrt{n}\text{vec} \left( \hat{C}_n(\rho) - \Sigma_0(\rho) \right) \xrightarrow{d} \mathbb{C}N \left( 0, M_1, M_2 \right)$$

with $M_1$ and $M_2$ are given by:

$$M_1 = \left( \left( \Sigma_0^{\frac{1}{2}} \right)^T \otimes \Sigma_0^{\frac{1}{2}} \right) \left( I_{N^2} - \tilde{F} \right)^{-1} \left( \left( \Sigma_0^{\frac{1}{2}} \right)^T \otimes \Sigma_0^{\frac{1}{2}} \right)$$

$$M_2 = \left( \left( \Sigma_0^{\frac{1}{2}} \right)^T \otimes \Sigma_0^{\frac{1}{2}} \right) \left( I_{N^2} - \tilde{F}^T \right)^{-1} \left( \left( \Sigma_0^{\frac{1}{2}} \right)^T \otimes \Sigma_0^{\frac{1}{2}} \right)$$

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:**
  \[
  \hat{C}_n(\rho) = \frac{\hat{B}_n(\rho)}{\frac{1}{p}\text{tr}\hat{B}_n(\rho)} \quad \text{with} \quad \hat{B}_n(\rho) = (1 - \rho)\frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{1 - \rho \hat{C}_n^{-1}(\rho)x_i} + \rho I_p.
  \]

- **Regularized M-scatter estimators**
  \[
  \hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u(x_i^H\hat{C}_n^{-1}x_i)x_i x_i^H + \rho I_p.
  \]

- **Regularized Tyler estimators with priori**
  \[
  \hat{C}_n(\rho) = (1 - \rho)\frac{p}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{x_i^H\hat{C}_n^{-1}(\rho)x_i} + \rho T
  \]
  where $T$ is a priori matrix.

- **Normalized Regularized Tyler estimators with priori**
  \[
  \check{C}_n(\rho) = \frac{\check{B}_n(\rho)}{\frac{1}{p}\text{tr}\check{B}_n(\rho)} \quad \text{with} \quad \check{B}_n(\rho) = (1 - \rho)\frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{1 - \rho \hat{C}_n^{-1}(\rho)x_i} + \rho \frac{T}{\text{tr}\hat{C}_n^{-1}T}.
  \]

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

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
Outline

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
Part I: Robust statistics/
   Motivation
   Distribution models
   Maximum Likelihood estimators
   M-estimators of scatters
   Regularized Robust estimators

Part II: Random Matrix Theory for robust estimation/
   Review of random matrix theory results
     Detection
     Estimation
     M-scatter estimator in the large random matrix regime
     Eigenvalue localization
     Source localization
   Regularized estimators
   Application: Radar detection
Illustrative random models covered by RMT

Let \( X = [x_1, \cdots, x_n] \) denote the matrix observation. Then, the following models are encountered in practice:

- **White space and in time:**
  \[
  X = [x_1, \cdots, x_n], \quad x_j \sim \mathcal{CN}(0, I_p)
  \]

- **Space and time correlation**
  \[
  X = C_1^{\frac{1}{2}} W T^{\frac{1}{2}}
  \]

- **Information-plus-noise model: non-centered observations with /or without correlation**
  \[
  X = C_2^{\frac{1}{2}} W T^{\frac{1}{2}} + A
  \]

**Key Assumption:** \( n, p \) grow to \( \infty \) with \( \frac{p}{n} \to c \)
Illustration: Marchenko-Pastur Law

If $C_p = I_p$,

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

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

Assume that we have:

\[ Y = \underbrace{X + \beta L}_{\text{High rank random signal}} \]

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

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

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

at most \( K \) isolated eigenvalues
Outline

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
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 $x_1, \cdots, x_n \in \mathbb{C}^p$, $n$ observations of size $p$ independent and identically distributed of unknown covariance $C_p$.

- Let $X = [x_1, \cdots, x_n]$ the matrix of observations.

- Inference problem: Consider $\theta = f(C_p)$, where $f$ is a certain functional.

**Objective:** Estimate parameter $\theta$ from the observation matrix $X$

- Classical Approach to handle this inference problem
  1. Form the sample covariance matrix $\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H$.
  2. Substitute the unknown covariance matrix $C_p$ by $\hat{S}_n$.

An estimator of $\theta$ is given by: $\hat{\theta} = f(\hat{S}_n)$
Estimation of covariance matrices

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

$$\hat{S}_n \overset{a.s.}{\to} C_p$$

- The convergence is in operator norm, in that:

$$\|\hat{S}_n - C_p\| \overset{a.s.}{\to} 0.$$

- As a result, by the continuous mapping theorem:

$$\hat{\theta} \overset{a.s.}{\to} \theta.$$

The conventional estimator $\hat{\theta}$ is consistent in the regime $n \to \infty$ and $p$ 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 $\Rightarrow$ 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:
  \[
  \left[\hat{S}_n\right]_{i,j} \xrightarrow{a.s.} \left[C_p\right]_{i,j}
  \]

But $\|\hat{S}_n - C_p\|$ does not converge to zero.

Then, $\hat{\theta}$ is not a consistent estimator of $\theta$ in the RMT regime
G-estimation

- Classical signal processing methods.

\[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H \rightarrow \hat{S}_N \]

Sample covariance matrix

\[ \hat{\theta} = f(\hat{S}_N) \]

Estimator of \( \theta \)

- Improved methods using RMT. In general, we proceed into two steps:

  1. Understand the behaviour of the conventional estimators under the RMT regime

\[ \hat{\theta} - \theta \xrightarrow{a.s.} \text{Bias} \]

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

\[ \hat{\theta}_G - \theta \xrightarrow{a.s.} 0. \]

\[ \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H \rightarrow \hat{S}_N \]

Sample covariance matrix

\[ \hat{\theta} = g(\hat{S}_N) \]

Estimator of \( \theta \)
Outline

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
Random Matrix regime of M-scatter estimator

Definition
For $x_1, \cdots, x_n$ with $n > p$, $\hat{C}_n$ is the solution of:

$$\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H$$

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

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

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

Note that taking $\tilde{x}_i = C_n^{-\frac{1}{2}} x_i$, and setting $\tilde{C}_n \triangleq C_n^{-\frac{1}{2}} \hat{C}_n^{-\frac{1}{2}} C_n^{-\frac{1}{2}}$,

$$\tilde{C}_n = \frac{1}{n} \sum_{i=1}^{n} u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H$$

$\Rightarrow$ Without loss of generality, we can assume $C_p = I_p$. 
Model description

- Assumption on the data (Elliptical model): \( x_1, \ldots, x_n \) independent

\[
x_i = \sqrt{\tau_i} C_p^{1/2} w_i
\]

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

- Assumption on the tail: For each \( a > b > 0 \), we have:

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

- 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 $\hat{C}_n$ converges to some deterministic limit. In particular:

$$\frac{1}{n} \sum_{i=1}^{n} u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H$$

do not converge to
$$\mathbb{E} \left[ u \left( \frac{1}{p} x^H \hat{C}_n^{-1} x \right) xx^H \right]$$

- Major issues with $\hat{C}_n$
  - No closed-form expression
  - Sum of non-independent rank-one matrices $u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H$
  - No explicit relation between $\hat{C}_n$ and the random vectors $x_1, \ldots, x_n$.

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

- Rewriting of \( \hat{C}_n \):
  - Denote:

\[
\hat{C}(j) = \frac{1}{n} \sum_{i \neq j} u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H
\]
### Heuristic approach

- **Rewriting of \( \widehat{C}_n \):**
  
  Denote:
  
  \[ \widehat{C}_{(j)} = \frac{1}{n} \sum_{i \neq j} u \left( \frac{1}{p} x_i^H \widehat{C}_n^{-1} x_i \right) x_i x_i^H \]

- **Using the identity** \((A + tvv^H)^{-1} = A^{-1}/(1 + tv^H A^{-1} v)\). Then:

  \[ \tau_i d_i \triangleq \frac{1}{p} x_i^H \widehat{C}_n^{-1} x_i = \frac{\frac{1}{p} x_i^H \widehat{C}_{(i)}^{-1} x_i}{1 + c_n u \left( \frac{1}{p} x_i^H \widehat{C}_n^{-1} x_i \right) \frac{1}{p} x_i^H \widehat{C}_{(i)}^{-1} x_i} \]

  \[ \implies \tau_i d_i \triangleq \frac{1}{p} x_i^H \widehat{C}_{(i)}^{-1} x_i = \frac{\frac{1}{p} x_i^H \widehat{C}_n^{-1} x_i}{1 - c_n \phi \left( \frac{1}{p} x_i^H \widehat{C}_{(i)}^{-1} x_i \right)} = g \left( \frac{1}{p} x_i^H \widehat{C}_n^{-1} x_i \right) \text{ with:} \]

  \[ g(x) = \frac{x}{1 - c_n \phi(x)} \]
Heuristic approach

- Rewriting of $\hat{C}_n$:
  - Denote:
    \[
    \hat{C}_{(j)} = \frac{1}{n} \sum_{i \neq j}^n u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H
    \]

- Using the identity $(A + tvv^H)^{-1} = A^{-1} / (1 + tv^H A^{-1} v)$. Then:
  \[
  \tau_i \tilde{d}_i \triangleq \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i = \frac{\frac{1}{p} x_i^H \hat{C}_{(i)} x_i}{1 + c_n u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) \frac{1}{p} x_i^H \hat{C}_{(i)} x_i}
  \]
  \[
  \Rightarrow \tau_i d_i \triangleq \frac{1}{p} x_i^H \hat{C}_{(i)} x_i = \frac{\frac{1}{p} x_i^H \hat{C}_n^{-1} x_i}{1 - c_n \phi \left( \frac{1}{p} x_i^H \hat{C}_{(i)} x_i \right)} = g \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) \text{ with:}
  \]
  \[
  g(x) = \frac{x}{1 - c_n \phi(x)}
  \]
  - $g$ is monontous and increasing, then $\frac{1}{p} x_i^H \hat{C}_n^{-1} x_i = g^{-1} \left( \frac{1}{p} x_i^H \hat{C}_{(i)} x_i \right)$
Heuristic approach

- Rewriting of $\hat{C}_n$:
  
  Denote:
  
  $$\hat{C}(j) = \frac{1}{n} \sum_{i \neq j} u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H$$

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

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

  $$\implies \tau_i d_i \triangleq \frac{1}{p} x_i^H \hat{C}(i) x_i = \frac{\frac{1}{p} x_i^H \hat{C}_n^{-1} x_i}{1 - c_n \Phi \left( \frac{1}{p} x_i^H \hat{C}(i) x_i \right)} = g \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right)$$

  with:

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

- $g$ is monontous and increasing, then $\frac{1}{p} x_i^H \hat{C}_n^{-1} x_i = g^{-1} \left( \frac{1}{p} x_i^H \hat{C}(i) x_i \right)$

- Using this equality in the expression of $\hat{C}_n$,

  $$\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} u \left( \frac{1}{p} x_i^H \hat{C}_n^{-1} x_i \right) x_i x_i^H$$

  $$= \frac{1}{n} \sum_{i=1}^{n} v \left( \frac{1}{p} x_i^H \hat{C}(i) x_i \right) x_i x_i^H$$

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

- Recall that:

\[
\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} v \left( \frac{1}{p} x_i^H \hat{C}^{-1}_{(i)} x_i \right) x_i x_i^H
\]

- Intuitively \( \hat{C}^{-1}_{(i)} \) and \( x_i \) are weakly dependent.
Heuristic approach

- Recall that:

\[
\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} v \left( \frac{1}{p} x_i^H \hat{C}_{(i)}^{-1} x_i \right) x_i x_i^H
\]

- Intuitively \( \hat{C}_{(i)}^{-1} \) and \( x_i \) are weakly dependent.

\[\implies\] We expect in particular using the convergence of quadratic forms lemma that:

\[
\frac{1}{p} x_i^H \hat{C}_{(i)}^{-1} x_i \simeq \tau_i \frac{1}{p} \text{tr} \hat{C}_{(i)}^{-1} \simeq \tau_i \frac{1}{p} \text{tr} \hat{C}_n^{-1}
\]

\[\implies\] We thus have:

\[
\hat{C}_n \simeq \frac{1}{n} \sum_{i=1}^{n} v \left( \tau_i \frac{1}{p} \text{tr} \hat{C}_n^{-1} \right) x_i x_i^H
\]
Heuristic approach

Recall that:

\[
\hat{C}_n = \frac{1}{n} \sum_{i=1}^{n} v \left( \frac{1}{p} x_i^H \hat{C}_{(i)}^{-1} x_i \right) x_i x_i^H
\]

Intuitively \( \hat{C}_{(i)}^{-1} \) and \( x_i \) are weakly dependent.

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

\[
\frac{1}{p} x_i^H \hat{C}_{(i)}^{-1} x_i \sim \tau_i \frac{1}{p} \text{tr} \hat{C}_{(i)}^{-1} \sim \tau_i \frac{1}{p} \text{tr} \hat{C}_n^{-1}
\]

We thus have:

\[
\hat{C}_n \sim \frac{1}{n} \sum_{i=1}^{n} v \left( \tau_i \frac{1}{p} \text{tr} \hat{C}_n^{-1} \right) x_i x_i^H
\]

We assume that \( \frac{1}{p} \text{tr} \hat{C}_n^{-1} \sim \gamma_n \)

\[
\Rightarrow \hat{C}_n \sim \frac{1}{n} \sum_{i=1}^{n} v \left( \tau_i \gamma \right) x_i x_i^H
\]
Heuristic approach

- **Determination of $\gamma_n$.**
  - Recall that:
    \[ \hat{C}_n \simeq \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) x_i x_i^H = \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) \tau_i w_i w_i^H \]
  - Moreover, $\gamma \simeq \frac{1}{p} \text{tr} \hat{C}_n^{-1}$. Hence,
    \[ \gamma \simeq \frac{1}{p} \text{tr} \left( \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) \tau_i w_i w_i^H \right)^{-1} \]
Heuristic approach

- **Determination of** $\gamma_n$.
  - Recall that:
    \[
    \hat{C}_n \simeq \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) x_i x_i^H = \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) \tau_i w_i w_i^H
    \]
  - Moreover, $\gamma \simeq \frac{1}{p} \text{tr}\hat{C}_n^{-1}$. Hence,
    \[
    \gamma \simeq \frac{1}{p} \text{tr} \left( \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) \tau_i w_i w_i^H \right)^{-1}
    \]
  - Let $W = [w_1, \cdots, w_n]$ and $D = \text{diag}\{\tau_i \gamma\}_{i=1}^{n}$. Then:
    \[
    \gamma \sim \frac{1}{p} \text{tr} \left( WDW^H \right)^{-1} = m_{\frac{1}{n}WDW^H(0)}
    \]
    where $m$ is the stieltjes transform associated with $WDW^H$ at 0.
Heuristic approach

▶ **Determination of $\gamma_n$.**

▶ Recall that:

$$\hat{C}_n \simeq \frac{1}{n} \sum_{i=1}^{n} \nu (\tau_i \gamma) x_i x_i^H = \frac{1}{n} \sum_{i=1}^{n} \nu (\tau_i \gamma) \tau_i w_i w_i^H$$

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

$$\gamma \simeq \frac{1}{p} \text{tr} \left( \frac{1}{n} \sum_{i=1}^{n} \nu (\tau_i \gamma) \tau_i w_i w_i^H \right)^{-1}$$

▶ Let $W = [w_1, \cdots, w_n]$ and $D = \text{diag} \{ \tau_i \gamma \}_{i=1}^{n}$. Then:

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

where $m$ is the stieltjes transform associated with $WDW^H$ at 0.

▶ Using standard results from random matrix theory,

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

- **Determination of $\gamma_n$.**
  - Recall that:
    \[
    \hat{C}_n \simeq \frac{1}{n} \sum_{i=1}^{n} \nu (\tau_i \gamma) x_i x_i^H = \frac{1}{n} \sum_{i=1}^{n} \nu (\tau_i \gamma) \tau_i w_i w_i^H
    \]
  - Moreover, $\gamma \simeq \frac{1}{p} \text{tr} \hat{C}_n^{-1}$. Hence,
    \[
    \gamma \simeq \frac{1}{p} \text{tr} \left( \frac{1}{n} \sum_{i=1}^{n} \nu (\tau_i \gamma) \tau_i w_i w_i^H \right)^{-1}
    \]
  - Let $W = [w_1, \ldots, w_n]$ and $D = \text{diag} \{\tau_i \gamma\}_{i=1}^{n}$. Then:
    \[
    \gamma \sim \frac{1}{p} \text{tr} \left( WD W^H \right)^{-1} = m_{\frac{1}{n} WD W^H} (0)
    \]
    where $m$ is the stieltjes transform associated with $WD W^H$ at 0.
  - Using standard results from random matrix theory,
    \[
    m_{\frac{1}{n} WD W^H} (0) = \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\tau_i \nu (\tau_i \gamma)}{1 + c \tau_i \nu (\tau_i \gamma) m_{\frac{1}{n} WD W^H} (0)} \right)^{-1}
    \]
  - We define $\gamma$ as a solution of the fixed-point equation:
    \[
    \gamma = \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\tau_i \nu (\tau_i \gamma)}{1 + c \tau_i \nu (\tau_i \gamma) \gamma} \right)^{-1} = \gamma \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\psi (\tau_i \gamma)}{1 + c \psi (\tau_i \gamma)} \right)^{-1}, \quad \psi(x) = xv(x)
    \]
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 $\gamma$ 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:

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

where $\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) x_i x_i^H$

**Takeaways**

- Propagation to $\hat{S}_n$ of first order results on $\hat{C}_n$
- $\hat{S}_n$ can be studied using classical results from random matrix theory, contrary to $\hat{C}_n$.
- Important consequences for array signal processing applications.
Main steps of the proof

- Let \( d_i = \frac{1}{p} w_i^H \hat{C}^{-1} w_i \). The proof consists in showing that

\[
\max_{1 \leq i \leq n} |e_i - 1| \xrightarrow{\text{a.s.}} 0, \quad \text{with} \quad e_i = \frac{\nu(\tau_i d_i)}{\nu(\tau_i \gamma)}
\]

- We relabel \( e_1, \cdots, e_n \) such that:

\[
e_1 \leq \cdots \leq e_n.
\]

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

\[
e_1 \geq 1 - \ell \ i.o. \quad \text{and} \quad e_n \leq 1 + \ell \ i.o.
\]
Main steps in the proof

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

$$e_n = \frac{\nu(\tau_n d_n)}{\nu(\tau_n \gamma_n)}$$
Main steps in the proof

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

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

$$v(\tau_n \gamma_n) e_n = v(\tau_n d_n) = v\left(\tau_n \frac{1}{p} w_n^H \hat{C}_{(n)}^{-1} w_n\right)$$
Main steps in the proof

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

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

$$\nu(\tau_n \gamma_n) e_n = \nu(\tau_n d_n) = \nu \left( \tau_n \frac{1}{p} w_n^H \hat{C}_n^{-1} w_n \right)$$

$$= \nu \left( \tau_n \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i d_i) w_i w_i^H \right)^{-1} w_n \right)$$
Main steps in the proof

- We proceed by contradiction. We assume that $e_n \geq 1 + \ell \text{ i.o.}$

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

$$\nu(\tau_n \gamma_n) e_n = \nu(\tau_n d_n) = \nu\left(\frac{1}{p} w_n^H \hat{C}^{-1}_{(n)} w_n\right)$$

$$= \nu\left(\tau_n \frac{1}{p} w_n^H \left(\frac{1}{n} \sum_{i < n} \tau_i \nu(\tau_i d_i) w_i w_i^H\right)^{-1} w_n\right)$$

$$= \nu\left(\tau_n \frac{1}{p} w_n^H \left(\frac{1}{n} \sum_{i < n} \tau_i e_i \nu(\tau_i \gamma_n) w_i w_i^H\right)^{-1} w_n\right)$$
Main steps in the proof

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

\[ e_n = \frac{\nu(\tau_n d_n)}{\nu(\tau_n \gamma_n)} \]

\[ \nu(\tau_n \gamma_n) e_n = \nu(\tau_n d_n) = \nu \left( \tau_n \frac{1}{p} w_n^H \hat{C}_{(n)}^{-1} w_n \right) \]

\[ = \nu \left( \tau_n \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i d_i) w_i w_i^H \right)^{-1} w_n \right) \]

\[ = \nu \left( \tau_n \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i e_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_n \right) \]

\[ \leq \nu \left( \tau_n e_n^{-1} \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_n \right) \]
Main steps in the proof

- Recall that:

\[ \nu(\tau_n \gamma_n)e_n \leq \nu \left( \tau_n e_n^{-1} \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_n \right) \]
Main steps in the proof

▪ Recall that:
\[
\nu(\tau_n \gamma_n) e_n \leq \nu \left( \tau_n e_n^{-1} \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_n \right)
\]

▪ Using standard approaches from random matrix theory, we can prove that:
\[
\max_{1 \leq j \leq n} \left| \frac{1}{p} w_j^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_j - \gamma_n \right| \xrightarrow{\text{a.s.}} 0.
\]
Main steps in the proof

- Recall that:

\[ \nu(\tau_n \gamma_n) e_n \leq \nu \left( \tau_n e_n^{-1} \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_n \right) \]

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

\[ \max_{1 \leq j \leq n} \left| \frac{1}{p} w_j^H \left( \frac{1}{n} \sum_{i<n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_j - \gamma_n \right| \xrightarrow{\text{a.s.}} 0. \]

\[ \Rightarrow \]

\[ \nu(\tau_n \gamma_n) e_n \leq \nu(\tau_n e_n^{-1} (\gamma_n - e_n)), \text{ with } e_n \to 0. \]
Main steps in the proof

- Recall that:

\[ v(\tau_n \gamma_n) e_n \leq v \left( \tau_n e_n^{-1} \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i<n} \tau_i v(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_n \right) \]

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

\[ \max_{1 \leq j \leq n} \left| \frac{1}{p} w_j^H \left( \frac{1}{n} \sum_{i<n} \tau_i v(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_j - \gamma_n \right| \xrightarrow{a.s.} 0. \]

\[ \Rightarrow \]

\[ v(\tau_n \gamma_n) e_n \leq v(\tau_n e_n^{-1}(\gamma_n - e_n)), \text{ with } e_n \to 0. \]

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

\[ \psi(\tau_n \gamma_n) \leq \psi(\tau_n e_n^{-1} \gamma_n) \left( 1 - e_n \gamma_n^{-1} \right)^{-1} \]
Main steps in the proof

- Recall that:

\[ \nu(\tau_n \gamma_n) e_n \leq \nu \left( \tau_n e_n^{-1} \frac{1}{p} w_n^H \left( \frac{1}{n} \sum_{i < n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_n \right) \]

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

\[ \max_{1 \leq j \leq n} \left| \frac{1}{p} w_j^H \left( \frac{1}{n} \sum_{i < n} \tau_i \nu(\tau_i \gamma_n) w_i w_i^H \right)^{-1} w_j - \gamma_n \right| \xrightarrow{a.s.} 0. \]

\[ \Rightarrow \]

\[ \nu(\tau_n \gamma_n) e_n \leq \nu(\tau_n e_n^{-1} (\gamma_n - e_n)), \text{ with } e_n \to 0. \]

- Use \( \psi(x) = x \nu(x) \),

\[ \psi(\tau_n \gamma_n) \leq \psi(\tau_n e_n^{-1} \gamma_n) \left( 1 - e_n \gamma_n^{-1} \right)^{-1} \]

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

\[ \psi(\tau_0 \gamma_0) \leq \psi \left( \frac{\tau_0 \gamma_0}{1 + \ell} \right), \text{ Contradiction} \]
Simulations

Figure: Histogram of the eigenvalues of $\frac{1}{n} \sum_{i=1}^{n} x_i x_i^H$ for $n = 2500$, $N = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_1$ with $\Gamma(0.5, 2)$-distribution.
Simulations

Figure: Histogram of the eigenvalues of $\hat{\mathbf{C}}_n$ (left) and $\hat{\mathbf{S}}_n$ (right) for $n = 2500$, $N = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_1$ with $\Gamma(.5, 2)$-distribution.
Simulations

Figure: Histogram of the eigenvalues of $\hat{C}_n$ (left) and $\hat{S}_n$ (right) for $n = 2500$, $N = 500$, $C_p = \text{diag}(I_{125}, 3I_{125}, 10I_{250})$, $\tau_1$ with $\Gamma(.5, 2)$-distribution.

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

\[ \text{Empirical eigenvalue distribution of } \hat{C}_n \quad \text{Limiting density} \]
\[ \text{Empirical eigenvalue distribution of } \hat{S}_n \quad \text{Limiting density} \]
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:

\[ x_i = A s_i + \sqrt{\tau_i} w_i \]

- \( A \in \mathbb{C}^{p \times K} \) is of full rank.
- Let \( B_n = AA^H \), we assume \( \liminf_{n \to \infty} \frac{1}{n} \text{tr} B_n > 0 \).
- \( s_1, \ldots, s_n \sim \mathcal{CN}(0, I_K) \), independent standard Gaussian distributed vectors.
- Same assumptions on functions \( \phi \) and \( u \).
- Random matrix theory regime: \( n, p, K \to \infty \)

Then:

\[ \| \hat{C}_n - \hat{S}_n \| \xrightarrow{a.s.} 0, \]

where

\[ \hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} \nu(\delta_i) x_i x_i^H \]

with \( \delta_1, \ldots, \delta_n \) being the unique solutions to:

\[ \delta_i = \frac{1}{p} \text{tr} \left( B_n + \tau_i I_p \right) \left( \frac{1}{n} \sum_{j=1}^{n} \frac{\nu(\delta_j) (B_n + \tau_j I_p)}{1 + c\psi(\delta_j)} \right)^{-1} \]
Illustration

Figure: Histogram of the eigenvalues of $\hat{S}_n$ against the limiting spectral measure with $n = 800$, $p = 80$ $B_p = \text{diag}(0.5I_{30}, I_{30}, 0I_{20})$
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:** $x_1, \cdots, x_n \in \mathbb{C}^p$ independent:

$$x_i = \sum_{\ell=1}^{K} \sqrt{p_{\ell}} a_{\ell} s_{\ell i} + \sqrt{\tau_i} w_i$$

- $w_i \in \mathbb{C}^p$ independent standard Gaussian vectors,
- $\tau_i$ deterministic scalars.
- $p_1 \geq \cdots \geq p_K \geq 0$
- $a_1, \cdots, a_K \in \mathbb{C}^{p \times 1}$ deterministic with $\sum_{\ell=1}^{K} p_{\ell} a_{\ell} a_{\ell}^H = \text{diag}\{p_i\}_{i=1}^{K}$.
- $\frac{1}{n} \sum_{i=1}^{n} \delta_{\tau_i} \rightarrow \nu$ weakly.
- $s_{1,i}, \cdots, s_{Kn}$ independent with zero mean and unit variance.

**Theorem**

*Under the previous assumptions, as $n \rightarrow \infty$,*

$$\|\hat{C}_n - \hat{S}_n\| \overset{a.s.}{\rightarrow} 0,$$

*with*

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} \nu(\tau_i \gamma) x_i x_i^H$$
Outline

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
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 $\hat{C}_n$. 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

$$
p_\pm \triangleq \lim_{x \downarrow S^\pm} -c \left( \int \frac{\delta(x)v_c(t\gamma)}{1 + \delta(x)tv_c(t\gamma)} \nu(dt) \right)^{-1}, \quad S^\pm \triangleq \frac{\phi_\infty(1 + \sqrt{c})^2}{\gamma(1 - c\phi_\infty)}.
$$
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 $\hat{C}_n$. 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 $p_j > p_-$, $\hat{\lambda}_j \xrightarrow{a.s.} \Lambda_j > S^+$, otherwise $\limsup_n \hat{\lambda}_j \leq S^+$ a.s., with $\Lambda_j$ unique positive solution to

$$
-c \left( \delta(\Lambda_j) \int \frac{v_c(\tau \gamma)}{1 + \delta(\Lambda_j)\tau v_c(\tau \gamma)} \nu(d\tau) \right)^{-1} = p_j.
$$
Figure: Histogram of the eigenvalues of $\frac{1}{n} \sum_{i=1}^{n} y_i y_i^*$ against the limiting spectral measure, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-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**:
  - $\lambda_i(\hat{C}_n) > S^+ \Rightarrow$ Presence of a source!
  - $\lambda_i(\hat{C}_n) \in (\sup(\text{Support}), S^+) \Rightarrow$ May be due to a source or to a noise impulse.
  - $\lambda_i(\hat{C}_n) < \sup(\text{Support}) \Rightarrow$ As usual, nothing can be said.

$\Rightarrow$ Induces a natural source detection algorithm.
Outline

Part I: Robust statistics/
   Motivation
   Distribution models
   Maximum Likelihood estimators
   M-estimators of scatters
   Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
   Review of random matrix theory results
      Detection
      Estimation
   M-scatter estimator in the large random matrix regime
      Eigenvalue localization
      Source localization
   Regularized estimators
   Application: Radar detection
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:

$$x_i = \sum_{k=1}^{K} \sqrt{p_k} a_p(\theta_k) s_{k,t} + \sqrt{\tau_i} w_i, \; t = 1, \ldots, n$$

- $A_p = [a_p(\theta_1), \ldots, a_p(\theta_K)]$ with $a_p(\theta) = \begin{bmatrix} 1 \\ e^{1\pi\sin\theta} \\ \vdots \\ e^{1(p-1)\pi\sin\theta} \end{bmatrix}$

- Objective: infer $\theta_1, \ldots, \theta_K$ from the $n$ observations
- Let $X_n = [x_1, \ldots, x_n]$, then,

$$X_n = A_p \text{diag}\{p_i\}_{i=1}^{K} S + W \text{diag}\{\tau_i\}_{i=1}^{n}$$

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

$$\eta(\theta) \triangleq a_N(\theta)^* \Pi a_N(\theta) = 0$$
MUSIC algorithms

- Traditional MUSIC algorithm: Angles are estimated as local minima of:
  \[
  \hat{\eta} = a_p(\theta) \hat{\Pi} a_p(\theta)
  \]
  where \( \hat{\Pi} \) is the orthogonal projection matrix on the eigenspace associated to the \( p - K \) largest eigenvalues of \( \frac{1}{n} X_n X_n^H \).

- 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 \( \frac{1}{n} X_n X_n^H \).

- Robust G-MUSIC: Angles are estimated as local minima of: \( \hat{\eta}_{R,G} \) where \( \hat{\eta}_{R,G} \) is a consistent estimator of \( \eta(\theta) \) that is based on the M-scatter estimator \( \hat{C}_n \).
Eigenvalue and eigenvector projection estimates

- $u_k$ eigenvector of $k$-th largest eigenvalue of $AA^H \sum_{i=1}^{K} p_i a_i(\theta)a_i(\theta)^H$
- $\hat{u}_k$ eigenvector of $k$-th largest eigenvalue of $\hat{C}_n$
Eigenvalue and eigenvector projection estimates

- \( u_k \) eigenvector of \( k \)-th largest eigenvalue of \( AA^H \sum_{i=1}^{K} p_i a_i(\theta) a_i(\theta)^H \)
- \( \hat{u}_k \) eigenvector of \( k \)-th largest eigenvalue of \( \hat{C}_n \)

**Theorem (Estimation under known \( \nu \))**

1. Power estimation. *For each \( p_j > p_- \),*

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

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

\[
\sum_{k, p_k=p_j} a^H u_k u_k^H b - \sum_{k, p_k=p_j} w_k a^H \hat{u}_k \hat{u}_k^H b \xrightarrow{a.s.} 0
\]

\[
w_k = \frac{\int \frac{\nu_c(t \gamma)}{\left(1 + \delta(\hat{\lambda}_k) t \nu_c(t \gamma)\right)^2} \nu(dt)}{\int \frac{\nu_c(t \gamma)}{1 + \delta(\hat{\lambda}_k) t \nu_c(t \gamma)} \nu(dt)} \left(1 - \frac{1}{c} \int \frac{\delta(\hat{\lambda}_k)^2 t^2 \nu_c(t \gamma)^2}{\left(1 + \delta(\hat{\lambda}_k) t \nu_c(t \gamma)\right)^2} \nu(dt)\right).
\]
Part II. Random Matrix Theory for robust estimation/ M-scatter estimator in the large random matrix regime

**Eigenvalue and eigenvector projection estimates**

**Theorem (Estimation under unknown $\nu$)**

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

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

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

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

   where

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

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

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

\[
a(\theta) = p^{-\frac{1}{2}} \left[ \exp(2\pi i dj \sin(\theta)) \right]_{j=0}^{N-1}.
\]
Application to G-MUSIC

Assume the model \( a_i = a(\theta_i) \) with
\[
a(\theta) = p^{-\frac{1}{2}} \left[ \exp(2\pi i dj \sin(\theta)) \right]_{j=0}^{N-1}.
\]

Corollary (Robust G-MUSIC)

Define \( \hat{\eta}_{RG}(\theta) \) and \( \hat{\eta}_{RG}^{emp}(\theta) \) as
\[
\begin{align*}
\hat{\eta}_{RG}(\theta) &= 1 - \sum_{k=1}^{\{|j, p_j > p_-\}|} w_k a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta), \\
\hat{\eta}_{RG}^{emp}(\theta) &= 1 - \sum_{k=1}^{\{|j, p_j > p_-\}|} \hat{w}_k a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta).
\end{align*}
\]

Then, for each \( p_j > p_- \),
\[
\hat{\theta}_j \xrightarrow{a.s.} \theta_j, \quad \hat{\theta}_j^{emp} \xrightarrow{a.s.} \theta_j
\]

where
\[
\begin{align*}
\hat{\theta}_j &\triangleq \arg\min_{\theta \in \mathbb{R}^K_j} \{ \hat{\eta}_{RG}(\theta) \} \\
\hat{\theta}_j^{emp} &\triangleq \arg\min_{\theta \in \mathbb{R}^K_j} \{ \hat{\eta}_{RG}^{emp}(\theta) \}.
\end{align*}
\]
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 $10^\circ$ and $12^\circ$, 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 $10^\circ$ and $12^\circ$, Student-t impulses with parameter $\beta = 10$, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$. 

$P_1, P_2$ [dB]
Outline

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
Regularized estimators

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

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

- Regularized Tyler estimators in elliptical distributed data settings

\[
\hat{C}_n(\rho) = (1 - \rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\frac{1}{p} x_i H \hat{C}_n^{-1}(\rho) x_i} + \rho I_p, \quad \rho \in \left(\max \left\{ 0, \frac{p - n}{n} \right\}, 1 \right)
\]

\[
\tilde{C}_n(\rho) = \frac{\tilde{B}_n(\rho)}{\frac{1}{p} \text{tr} \tilde{B}_n(\rho)}, \quad \tilde{B}_n(\rho) = (1 - \rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\frac{1}{p} x_i H \tilde{C}_n^{-1}(\rho) x_i} + \rho I_p, \quad \rho \in (0, 1]
\]
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:**
    \[ x_i = \sqrt{\tau_i} C_p^{\frac{1}{2}} w_i \]
    with \( w_1, \cdots, w_n \) independent standard Gaussian random vectors.
  - **Matrix** \( C_p \) satisfies \( \frac{1}{p} \text{tr} C_p = 1 \) with \( \limsup \| C_p \| < \infty \).
  - \( \nu_p = \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i(C_p)} \to \nu \), 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 $\hat{R}_\varepsilon = [\varepsilon + \max\{0, 1 - c^{-1}\}, 1]$. Then, as $p, n \to \infty$, $p/n \to c \in (0, \infty)$,

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

where

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

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

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 \( \tilde{\mathcal{R}} = [\varepsilon, 1] \). Then, as \( p, n \to \infty \), \( p/n \to c \in (0, \infty) \),

\[
\sup_{\rho \in \tilde{\mathcal{R}}} \left\| \tilde{\mathcal{C}}_n(\rho) - \tilde{\mathcal{S}}_n(\rho) \right\| \xrightarrow{a.s.} 0
\]

where

\[
\tilde{\mathcal{C}}_n(\rho) = \frac{\tilde{\mathcal{B}}_n(\rho)}{1/n \text{tr} \tilde{\mathcal{B}}_N(\rho)}, \quad \tilde{\mathcal{B}}_N(\rho) = (1 - \rho) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\rho} + \rho I_p
\]

\[
\tilde{\mathcal{S}}_n(\rho) = \frac{1 - \rho}{1 - \rho + T_\rho} \frac{1}{n} \sum_{i=1}^{n} C_{1/2} \frac{w_i w_i^H}{\rho} C_{1/2} + \frac{T_\rho}{1 - \rho + T_\rho} I_p
\]

in which \( T_\rho = \rho \tilde{\gamma}(\rho) F(\tilde{\gamma}(\rho); \rho) \) 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 \( \tilde{\gamma}(\rho) \) is the unique positive solution to the equation in \( \tilde{\gamma} \),

\[
1 = \frac{1}{p} \sum_{i=1}^{N} \tilde{\gamma} \rho + \frac{\lambda_i(C_n)}{(1 - \rho)c + F(\tilde{\gamma}; \rho) \lambda_i(C_n)}.
\]

Moreover, \( \rho \mapsto \tilde{\gamma}(\rho) \) 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 $\tilde{\rho} \in (0, 1]$ such that

$$\frac{\hat{S}_n(\hat{\rho})}{\hat{\gamma}(\hat{\rho})} + \hat{\rho} = \tilde{S}_n(\tilde{\rho}) = (1 - \rho) \frac{1}{n} \sum_{i=1}^{n} C_n^{1/2} w_i w_i^* C_{\tilde{\rho}}^{1/2} + \rho I_p.$$ 

Besides, $(0, 1] \rightarrow (\max\{0, 1 - c^{-1}\}, 1]$, $\rho \mapsto \hat{\rho}$ and $(0, 1] \rightarrow (0, 1]$, $\rho \mapsto \tilde{\rho}$ 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

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

Besides, $(0, 1] \rightarrow (\max\{0, 1 - c^{-1}\}, 1]$, $\rho \mapsto \hat{\rho}$ and $(0, 1] \rightarrow (0, 1]$, $\rho \mapsto \check{\rho}$ 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 $\bar{\rho} \in (0, 1]$ such that

$$
\frac{\hat{S}_n(\hat{\rho})}{1 - \hat{\rho}} + \hat{\rho} = \bar{S}_n(\bar{\rho}) = (1 - \rho) \frac{1}{n} \sum_{i=1}^{n} C_n^2 w_i w_i^* C_n^2 + \rho I_p.
$$

Besides, $\{0, 1] \rightarrow (\max\{0, 1 - c^{-1}\}, 1]$, $\rho \mapsto \hat{\rho}$ and $\{0, 1] \rightarrow (0, 1]$, $\rho \mapsto \bar{\rho}$ 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 $\rho$ in the theorems is essential to find optimal values of $\rho$. 
Optimal Shrinkage parameter

- Chen sought for a Frobenius norm minimizing $\rho$ but got stuck by implicit nature of $\hat{C}_N(\rho)$.
Optimal Shrinkage parameter

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

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

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

$$
\frac{\hat{\rho}^*}{1 - \hat{\rho}^*} + \check{\rho}^* = \frac{T_{\check{\rho}^*}}{1 - \check{\rho}^* + T_{\check{\rho}^*}} = \rho^*.
$$

*Then, letting $\epsilon$ small enough,*

$$
\inf_{\rho \in \hat{\mathcal{R}}_\epsilon} \hat{D}_n(\rho) \xrightarrow{a.s.} D^*, \quad \inf_{\rho \in \check{\mathcal{R}}_\epsilon} \check{D}_n(\rho) \xrightarrow{a.s.} D^*
$$

$$
\hat{D}_b(\hat{\rho}^*) \xrightarrow{a.s.} D^*, \quad \check{D}_n(\check{\rho}^*) \xrightarrow{a.s.} D^*.
$$
Estimating $\hat{\rho}^*$ and $\check{\rho}^*$

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

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

$$\hat{\rho}_n \frac{1}{p} \text{tr} \hat{C}_n(\hat{\rho}_n) = c_n \frac{1}{p} \text{tr} \left( \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H \right)^2 \right) - \frac{1}{n} \sum_{i=1}^{n} x_i^H \check{C}_n(\check{\rho}_n) - \frac{1}{n} x_i \|x_i\|^2 = c_n \frac{1}{p} \text{tr} \left( \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H \right)^2 \right) - 1 \check{\rho}_n \frac{1}{n} \sum_{i=1}^{n} x_i^H \check{C}_n(\check{\rho}_n) - \frac{1}{n} x_i \|x_i\|^2 = c_n \frac{1}{p} \text{tr} \left( \left( \frac{1}{n} \sum_{i=1}^{n} x_i x_i^H \right)^2 \right) - 1 \check{\rho}_n \frac{1}{n} \sum_{i=1}^{n} x_i^H \check{C}_n(\check{\rho}_n) - \frac{1}{n} x_i \|x_i\|^2 $$

defined arbitrarily when no such solutions exist. Then

$$\hat{\rho}_n \stackrel{a.s.}{\rightarrow} \hat{\rho}^*$$

$$\check{\rho}_n \stackrel{a.s.}{\rightarrow} \check{\rho}^*$$

$$\hat{D}_n(\hat{\rho}_n) \stackrel{a.s.}{\rightarrow} D^*$$

$$\check{D}_n(\check{\rho}_n) \stackrel{a.s.}{\rightarrow} D^*$$
Estimating $\hat{\rho}^*$ and $\tilde{\rho}^*$

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

- Theorem only useful if $\hat{\rho}^*$ and $\check{\rho}^*$ 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

\[
\frac{1}{p} \text{tr} \hat{C}_n(\hat{\rho}_n) = \frac{c_n}{1} \text{tr} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\|x_i\|^2} \right)^2 \right] - 1
\]

\[
\check{\rho}_n \frac{1}{n} \sum_{i=1}^{n} \frac{x_i^H \hat{C}_n(\check{\rho}_n)^{-1} x_i}{\|x_i\|^2} = \frac{c_n}{1} \text{tr} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{\|x_i\|^2} \right)^2 \right] - 1
\]

defined arbitrarily when no such solutions exist. Then

\[
\hat{\rho}_n \xrightarrow{a.s.} \hat{\rho}^*, \quad \check{\rho}_n \xrightarrow{a.s.} \check{\rho}^*,
\]

\[
\hat{D}_n(\hat{\rho}_n) \xrightarrow{a.s.} D^*, \quad \check{D}_n(\check{\rho}_n) \xrightarrow{a.s.} \check{D}^*.
\]
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 $r = 0.7$; $\tilde{\rho}_N$ as above; $\tilde{\rho}_O$ the clairvoyant estimator proposed in (Chen’11).
Second-order statistics of the regularized Tyler estimator

- Motivation: From above, we know that:
  \[
  \sup_{\rho} \| \hat{C}_n(\rho) - \hat{S}_n(\rho) \| \xrightarrow{a.s.} 0,
  \]

- First order implications:
  \[
  \begin{align*}
  &\left| a^H \hat{C}_n(\rho) b - a^H \hat{S}_n(\rho) b \right| \xrightarrow{a.s.} 0 \\
  &\left| \frac{1}{n} \text{tr } f(\hat{C}_n(\rho)) - \frac{1}{n} \text{tr } f(\hat{S}_n(\rho)) \right| \xrightarrow{a.s.} 0.
  \end{align*}
  \]

- Does not imply propagation to \( \hat{S}_n(\rho) \) of second-order results on \( \hat{C}_n(\rho) \).

- From simulations it seems that:
  \[
  n^{\frac{1}{2} - \epsilon} \| \hat{C}_n(\rho) - \hat{S}_n(\rho) \| \xrightarrow{a.s.} 0.
  \]

\[
\Rightarrow n^{\frac{1}{2} - \epsilon} \left( a^H \hat{C}_n b - a^H \hat{S}_n b \right) \to 0, \quad \text{Weak result}
\]

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


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

$$\sup_{\rho \in \mathbb{R}_+} p^{1-\epsilon} \left| a^H \hat{C}_n^k b - a^H \hat{S}_n^k b \right| \overset{\text{a.s.}}{\to} 0,$$

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

Part I: Robust statistics/
  Motivation
  Distribution models
  Maximum Likelihood estimators
  M-estimators of scatters
  Regularized Robust estimators

Part II. Random Matrix Theory for robust estimation/
  Review of random matrix theory results
    Detection
    Estimation
  M-scatter estimator in the large random matrix regime
    Eigenvalue localization
    Source localization
  Regularized estimators
  Application: Radar detection
Application: Radar detection

\begin{align*}
\{ & H_0 : \text{received signal} = \text{noise} \\
& H_1 : \text{received signal} = \text{target signal} + \text{noise} \\
\} \Rightarrow \begin{cases} \\
H_0 : x = y & \text{No target} \\
H_1 : x = ap + y & \text{Presence of target} \\
\|p\| = 1 & \\
\end{cases}
\end{align*}

- Clutter model: Compound Gaussian distribution

\[ y_i = \sqrt{\tau_i} z_i \quad \text{where} \quad \begin{cases} \\
\tau_i & \text{heavy-tailed} \\
z_i & \text{Gaussian } C_p \\
\end{cases} \]

- If \( C_p \) is known up to a scale factor, the GLRT principle leads to the following detector (Normalized matched filter)

\[ T_p \triangleq \frac{|x^H C_p^{-1} p|}{\sqrt{p^* C_p^{-1} p} \sqrt{x^H C_p^{-1} x}} \overset{\mathcal{F}_1}{\underset{\mathcal{F}_0}{\gtrless}} \Gamma \]
Adaptive Normalized Matched Filter (ANMF) detector

- In practice, matrix $C_p$ is unknown.

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

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

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

- The statistics $T_p$ is replaced by:

$$
\hat{T}_n \triangleq \frac{|x^H \hat{C}_n^{-1}(\rho) p|}{\sqrt{p^H \hat{C}_n^{-1}(\rho) p} \sqrt{x^H \hat{C}_n^{-1}(\rho) x}} \overset{\mathbb{H}_1}{\underset{\mathbb{H}_0}{\gtrless}} \Gamma
$$
Choice of the regularization parameter

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

How to find the $\rho$ that satisfies good balance between bias and conditioning of the estimator:

- Classical methods for choosing $\rho$
  
  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 $\rho$ and the threshold $\Gamma$ such that:
  - Maintain the probability of false alarm at a fixed rate $P_{fa}$
    \[
P_{fa} = \mathbb{P} \left[ \hat{T}_n \geq \Gamma \middle| H_0 \right]
    \]
  - Maximize the probability of detection:
    \[
P_d \triangleq \mathbb{P} \left[ \hat{T}_n \geq \Gamma \middle| H_1 \right]
    \]
- Need to study the fluctuations of $\hat{T}_n$. 
Asymptotic False alarm probability and probability of detection

- **Initial observations**
  - If the threshold $\Gamma$ is taken fixed, then as $n, p \to \infty$ with $\frac{p}{n} \to c$ and under $H_0$
    \[ \hat{T}_n \xrightarrow{a.s.} 0. \]
  - Trivial result of little interest!
  - Select the parameters such that we avoid empty statements, $P_{fa} \to 0$ and $P_{d} \to 1$.
    - Consider $\Gamma = \frac{r}{\sqrt{p}}$.
      \[ P_{fa} = \mathbb{P}(\hat{T}_n \geq \Gamma) = \mathbb{P}(\sqrt{p}\hat{T}_n \geq r|H_0) \text{ of order } 1. \]
    - Evaluate the detection probability: If $\|p\| = 1$ and $a = O(1)$.
      \[ P_{d} = \mathbb{P}(\hat{T}_n \geq \Gamma) = \mathbb{P}(\sqrt{p}\hat{T}_n \geq r|H_1) \text{ of order } 1. \]

- **Techniques of calculation** Since
  \[ \sup_{\rho \in \mathbb{R}_k} p^{1-e} \left| a^H\hat{C}_n^{-1}b - a^H\hat{S}_n^{-1}b \right| \to 0 \]
  the fluctuations of $\hat{T}_n$ are the same as $\tilde{T}_n$ where $\hat{C}_n$ is replaced by $\hat{S}_n$:
  \[ \tilde{T}_n = \frac{\left| x^H\hat{S}_n^{-1}(\rho)p \right|}{\sqrt{p^H\hat{S}_n^{-1}(\rho)p} \sqrt{x^H\hat{S}_n^{-1}(\rho)x}} \]
Asymptotic false alarm and detection probabilities

- Recall

\[ \sqrt{p} \tilde{T}_n = \frac{\sqrt{p} x^H \hat{S}_n^{-1}(\rho) p}{\sqrt{p^H \hat{S}_n^{-1}(\rho) p \sqrt{x^H \hat{S}_n^{-1}(\rho) x}}}. \]

- The fluctuations are due to \( \sqrt{p} x^H \hat{S}_n^{-1}(\rho) p \).

\[ \implies \text{As } n, p \to \infty \text{ and condition on } \tau \]

- Under \( H_0 \): \( \sqrt{p} \tilde{T}_n \sim \|X\| \) with \( X \sim N(0, \sigma^2 p I_2) \).
- Under \( H_1 \): \( \sqrt{p} \tilde{T}_n \sim \|X\| \) with \( X \sim N(\mu_p, \sigma^2 p I_2) \)

\[ \sigma^2_p = \frac{1}{2} \frac{p^H C_p Q_p^2(\tilde{\rho}) p}{p^H Q_p(\tilde{\rho}) p} \left( \frac{1}{p} \text{tr} C_p Q_p(\tilde{\rho}) \left( 1 - c(1 - \tilde{\rho})^2 \mu_p \right)^2 \frac{1}{p} \text{tr} C_p^2 Q_p^2(\tilde{\rho}) \right)^{-1} \]

\[ \mu_p = \left[ \frac{a}{\sqrt{\tau}} \frac{\sqrt{p^H Q_p(\rho) p}}{\sqrt{\frac{1}{p} \text{tr} C_p Q_p(\rho)}}, 0 \right]^T \]

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

\[ m(-\rho) = \left( \rho + \frac{c(1 - \rho)}{p} \text{tr} C_p (I_p + (1 - \rho) m(-\rho) C_p)^{-1} \right)^{-1} \]

with \( Q_p(\rho) = (I_p + (1 - \rho) m(-\rho) C_p)^{-1} \) and \( \tilde{\rho} = \rho \left( \rho + \frac{1}{\gamma N(\rho)} \frac{1}{1 - (1 - \rho)c} \right)^{-1} \).
False alarm performance

Theorem (Asymptotic detector performance)

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

\[
\sup_{\rho \in \mathbb{R}_+} \left| P \left( \hat{T}_n(\rho) > \frac{\gamma}{\sqrt{p}} \right) - \exp \left( -\frac{\gamma^2}{2\sigma_p^2(\rho)} \right) \right| \to 0
\]

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

\[
\sigma_p^2(\tilde{\rho}) \triangleq \frac{1}{2} \frac{p^H C_p Q_p^2(\tilde{\rho}) p}{p^H Q_p(\tilde{\rho}) p \cdot \frac{1}{p} \text{tr} C_p Q_p(\tilde{\rho}) \cdot \left( 1 - c (1 - \tilde{\rho})^2 m(-\tilde{\rho})^2 \frac{1}{p} \text{tr} C_p^2 Q_p^2(\tilde{\rho}) \right)}
\]

with \( Q_p(\tilde{\rho}) \triangleq (I_p + (1 - \tilde{\rho}) m(-\tilde{\rho}) C_p)^{-1} \).
False alarm performance

**Theorem (Asymptotic detector performance)**

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

$$\sup_{\rho \in \mathbb{R}_K} \left| P \left( \hat{T}_n(\rho) > \frac{\gamma}{\sqrt{p}} \right) - \exp \left( -\frac{\gamma^2}{2\sigma^2_p(\rho)} \right) \right| \to 0$$

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

$$\sigma^2_p(\tilde{\rho}) \triangleq \frac{1}{2} \frac{p^H C_p Q_p^2(\tilde{\rho}) p}{p^H Q_p(\tilde{\rho}) p \cdot \frac{1}{p} \operatorname{tr} C_p Q_p(\tilde{\rho}) \cdot \left( 1 - c(1 - \tilde{\rho})^2 m(-\tilde{\rho})^2 \frac{1}{p} \operatorname{tr} C_p^2 Q_p^2(\tilde{\rho}) \right)}$$

with $Q_p(\tilde{\rho}) \triangleq \left( I_p + (1 - \tilde{\rho}) m(-\tilde{\rho}) C_p \right)^{-1}$.

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

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

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

$$
g(\rho) = \sqrt{\frac{1 - c(1 - \tilde{\rho})^2 m(-\tilde{\rho})}{p} \frac{1}{p} tr C_p^2 Q_2^2(\tilde{\rho})} \frac{2}{\sqrt{\tau}} a \left| p^H Q_\rho(\tilde{\rho}) p \right|.
$$

and $Q_1$ is the Marcum Q-function.
Optimal design of the ANMF detector

- To maximize the probability of detection, \( \rho \) should be set such that:

\[
\rho^* = \arg \max_{\rho} \frac{\mu_p}{\sigma_p} = \arg \max_{\rho} f_p(\rho)
\]

with

\[
f_p(\rho) = \frac{(1 - c(1 - \tilde{\rho})^2 m(-\tilde{\rho}) \frac{1}{p} \text{tr} C_p^2 Q_p^2(\tilde{\rho})) (p^H Q_p(\tilde{\rho}) p)^2}{p^H C_p Q_p^2(\tilde{\rho}) p}
\]

- For false alarm probability \( \eta \), test statistics threshold \( r \) set to:

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

⇒ **Optimal design**

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

\( \checkmark \) Optimal values depend on unknown \( C_N \). 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 $\hat{\sigma}^2_p(\rho)$:

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

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

- Then, we have:

$$
\sup_{\rho \in \mathbb{R}_+} |\hat{\sigma}_p(\rho) - \sigma_p(\rho)| \overset{a.s.}{\longrightarrow} 0,
$$

$$
\sup_{\rho \in \mathbb{R}_+} |\hat{f}_p(\rho) - f_p(\rho)| \overset{a.s.}{\longrightarrow} 0
$$

Optimal design of the ANMF detector

- Select $\rho$ 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
- $p = a_{N_p}(f_d) \otimes a_{N_a}(f_s)$ where
  $a_k(f) = \{\exp(j2\pi(\ell - 1)f)\}_{\ell=1}^k$
- $C_N \propto \left( I_N + \sum_{i=1}^{N_p} \sigma^2 A_{N_p}(f_{d_i}, f_{s_i}) A_{N_p}(f_{d_i}, f_{s_i})^H \right)$ with
  $A_{N_p}(f_{d_i}, f_{s_i}) = a_{N_p}(f_{d_i}) \otimes a_{N_a}(f_{s_i})$
- K-distributed clutter with zero mean and shape $\nu = 4.5$
- Compare with $\hat{\rho}_{ollila}$ given by:
  $$\hat{\rho}_{ollila} = \frac{N \text{tr} \widehat{C}_0 - 1}{N \text{tr} \widehat{C}_0 - 1 + n(N + 1) \left(N^{-1} \text{tr} \widehat{C}_0^{-2} - 1\right)}$$
  where $\widehat{C}_0$ is the conventional Tyler estimator.

ROC curve

![ROC curve](image)

**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 statistics.
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 $\hat{C}_n$ and $\hat{S}_n$ suggests that good performances should be also obtained when considering the following estimator:

$$\bar{S} = \frac{(1 - \rho)}{n} \sum_{i=1}^{n} \frac{x_i x_i^H}{1 + \rho x_i^H x_i} + \rho I_p.$$

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


**For both regimes**
- Study the joint mean and scatter robust estimators.
- Application hyperspectral imaging.