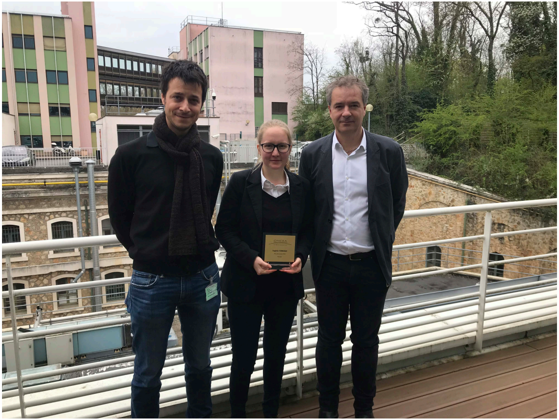
# Robust Model Order Selection Using Random Matrix Theory

Jean-Philippe Ovarlez[1,2], E. Terreaux[2] and F. Pascal[3]

[1]French Aerospace Lab, ONERA DEMR/MATS, France
[2]SONDRA CentraleSupélec, France
[3]L2S, CentraleSupélec, France

2018 First French-Finnish Statistical Learning Workshop

# Eugénie Terreaux PhD Days Prize ONERA 2018



# Motivations

The Model Order Selection is a fundamental problem in Signal Processing:

- Radar, Sonar: Direction of Arrival, Source Localization, STAP, Date of Arrival, Spectral Analysis (ARMA), etc.
- Hyperspectral: Unmixing, etc.
- Finance: portfolio optimization, efficient portfolio composition, etc.

In spite of the multitude of techniques available for solving this problem, most of them use information theoretic approaches, such as:

- the Akaike Information Criterion (AIC),
- the Minimum Description Length (MDL).

Recently, the Random Matrix Theory (RMT) based-approach has also been proposed.

**All these methods are classically based on white Gaussian noise assumption.**

# Contents

# Outline

# Problem formulation

Let $\{\mathbf{y}_i\}_{i\in[1,N]}$ be $N$ observations of size $m$ characterizing the $p < m$ mixed sources corrupted by additive noise:

$$\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j}\, \mathbf{m}_j + \mathbf{n}_i\,, \qquad i \in [1, N]\,, \tag{1}$$

which can be rewritten more compactly as:

$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{N}\,, \tag{2}$$

where

- $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{C}^{m\times N}$ are the observations,
- $\mathbf{M} \in \mathbb{C}^{m\times p}$ is the mixing matrix containing steering vectors of the $p$ sources,
- $\mathbf{S} \in \mathbb{C}^{p\times N}$ is the channel gain matrix,
- $\mathbf{N} \in \mathbb{C}^{m\times N}$ is the additive noise matrix, independent of the source signal.

**Generally, $p < m$ is unknown, $\mathbf{M}$ and $\mathbf{S}$ unknown and $\mathbf{N}$ is zero-mean white Gaussian noise.**

# Recall of Akaike Information Criterion and MDL

**Goal:** Given a set of $N$ observations $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$ and a family of models, (e.g., a parametrized family of pdf $\{f(\mathbf{Y}|\boldsymbol{\Theta}_k)\}_k$), select the model $k$ that best fits the data. Akaike [Akaike, 1974] proposal was to select the model which gives the minimum AIC defined as:

$$AIC(k) = -2 \log f\left(\mathbf{Y}|\hat{\boldsymbol{\Theta}}_k\right) + 2\,k\,,$$

where $\hat{\boldsymbol{\Theta}}$ is the MLE of vector $\boldsymbol{\Theta}_k$ and $k$ is the number of free adjusted parameters in vector $\boldsymbol{\Theta}_k$.

Schwartz [Schwarz, 1978] and Rissanen [Rissanen, 1978] approaches yield the same type of criterion, given by:

$$MDL = -\log f\left(\mathbf{Y}|\hat{\boldsymbol{\Theta}}_k\right) + \frac{1}{2}\,k\,\log N\,.$$

# Applications for Model Order Selection (1)

Let us consider the theoretical covariance matrix $\mathbf{R}$ of complex observations $\mathbf{y}_i$ with $\mathbf{n}_i$ white Gaussian noise:

$$\mathbf{R} = E\left[\mathbf{y}_i\,\mathbf{y}_i^H\right] = \mathbf{M}\,\mathbf{s}_i\,\mathbf{s}_i^H\,\mathbf{M}^H + \sigma^2\,\mathbf{I}_m = \mathbf{\Phi} + \sigma^2\,\mathbf{I}_m\,.$$

We assume here $\mathbf{M}$ full column rank ($\{\mathbf{m}_i\}_i$ linearly independent) and $\mathbf{s}_i\,\mathbf{s}_i^H$ non singular. We have:

- $\mathrm{rank}(\mathbf{\Phi}) = p$, the $m-p$ smallest eigenvalues of $\mathbf{\Phi}$ are equal to zero,
- $\mathrm{eig}(\mathbf{R}) = \left\{\lambda_1, \lambda_2, \ldots, \lambda_p, \sigma^2, \ldots, \sigma^2\right\}$.

We can define a family of covariance matrix $\mathbf{R}^{(k)} = \mathbf{\Phi}^{(k)} + \sigma^2\,\mathbf{I}_m$ as:

Model (k): $$\mathbf{R}^{(k)} = \sum_{i=1}^{k}\left(\lambda_i - \sigma^2\right)\,\mathbf{v}_i\,\mathbf{v}_i^H + \sigma^2\,\mathbf{I}_m\,,$$

where $\lambda_1$, ..., $\lambda_k$ are the $k$ highest eigenvalues of $\mathbf{R}^{(k)}$ and where $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are their corresponding eigenvectors. We can define also the vector $\mathbf{\Theta}_k$ of unknown parameters as:

$$\mathbf{\Theta}_k = \left(\lambda_1, \ldots, \lambda_k, \sigma^2, \mathbf{v}_1, \ldots, \mathbf{v}_k\right)^T\,.$$

# Applications for Model Order Selection (2)

$$
\begin{aligned}
-\log f\left(\mathbf{Y}|\hat{\boldsymbol{\Theta}}_k\right) &= -\log \prod_{i=1}^{N} \frac{1}{\pi^m \left|\mathbf{R}^{(k)}\right|} \exp\left(-\mathbf{y}_i^H \left(\mathbf{R}^{(k)}\right)^{-1} \mathbf{y}_i\right), \\
&\approx N \log\left|\mathbf{R}^{(k)}\right| + \mathrm{Tr}\left[\left(\mathbf{R}^{(k)}\right)^{-1} \sum_{i=1}^{N} \mathbf{y}_i\,\mathbf{y}_i^H\right], \\
&\approx N \log\left|\mathbf{R}^{(k)}\right| + \mathrm{Tr}\left[\left(\mathbf{R}^{(k)}\right)^{-1} N\,\hat{\mathbf{R}}\right], \qquad (3)
\end{aligned}
$$

where $\hat{\mathbf{R}} = \dfrac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i\,\mathbf{y}_i^H$. Maximizing (3) with respect to each parameter of $\boldsymbol{\Theta}_k$ [Anderson, 1963] leads to:

- $\hat{\lambda}_i = l_i$ and $\hat{\mathbf{v}}_i = \mathbf{u}_i, i \in [1, k]$, $\hat{\sigma}^2 = \dfrac{1}{m-k}\displaystyle\sum_{i=k+1}^{m} l_i$.

where $l_1 > l_2 \ldots > l_m$ and $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are the eigenvalues and eigenvectors of the Sample Covariance Matrix $\hat{\mathbf{R}}$.

# Applications for Model Order Selection (3)

The number of free parameters is obtained by counting the number of degrees of freedom spanned by $\boldsymbol{\Theta}_k = \left(\lambda_1, \dots, \lambda_k, \sigma^2, \mathbf{v}_1, \dots, \mathbf{v}_k\right)^T$ [Wax and Kailath, 1985]:

$k+1$ reals $+2\,k\,m$ reals $-\,2\,k$ normalizations $-\,2\,k\,(k-1)/2$ mutual orthogonalizations

Substituting the Maximum Likelihood Estimates in the log-likelihood (3) leads to:

$$AIC(k) = -2\,N \log \frac{\displaystyle\prod_{i=k+1}^{m} \lambda_i}{\left(\displaystyle\frac{1}{m-k}\sum_{i=k+1}^{m} \lambda_i\right)^{m-k}} + 2\,k(2\,m-k)\,,$$

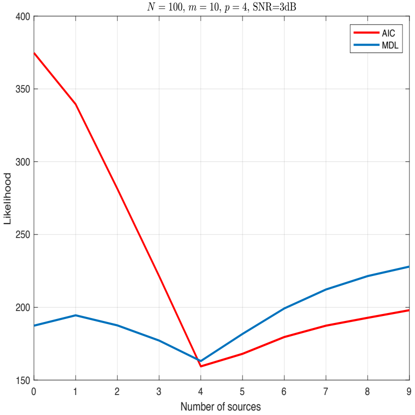
$$MDL(k) = -N \log \frac{\displaystyle\prod_{i=k+1}^{m} \lambda_i}{\left(\displaystyle\frac{1}{m-k}\sum_{i=k+1}^{m} \lambda_i\right)^{m-k}} + \frac{1}{2}\,k\,(2\,m-k)\,\log N\,.$$

# Applications for Model Order Selection (4)

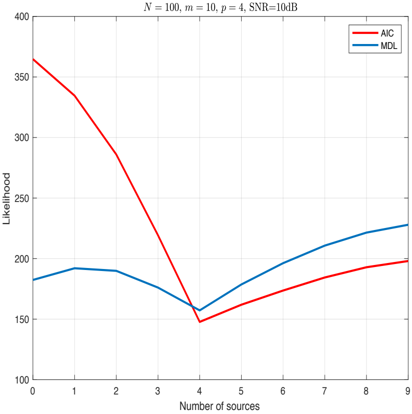
- AIC is shown to be not consistent and has a problem of over-estimation of the number of sources,
- MDL is consistent and is generally prefered to AIC,
- Both techniques are based on white Gaussian noise. They do not perform well for correlated noise or non-Gaussian noise,
- Both techniques may have some problems for high dimensional data.

# Examples (1)

SNR = 3dB

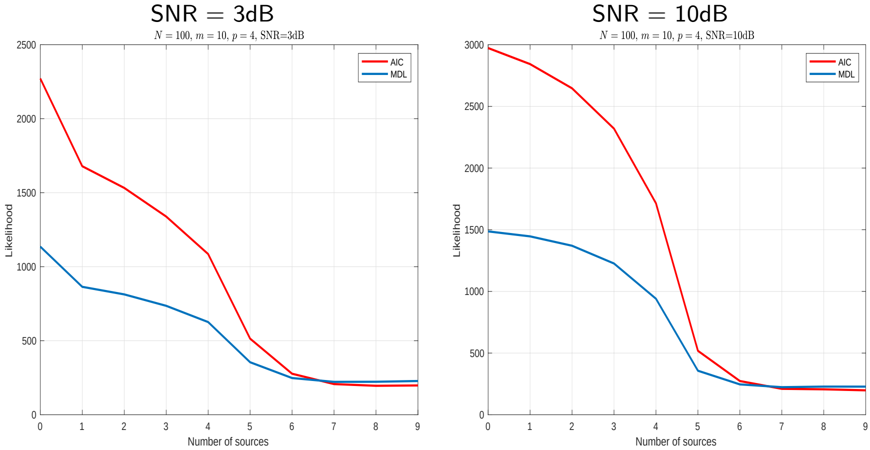



SNR = 10dB




AIC and MDL model order selection (white Gaussian noise, $N = 100$, $m = 10$).

# Examples (2)




AIC and MDL model order selection (correlated Gaussian noise, $\rho = 0.9$, $N = 100$, $m = 10$).

# Outline

# Some RMT-based results for detection schemes

The RMT (ex: [Couillet and Debbah, 2011]) allows 1) to understand the statistical behaviour of expressions involving estimate of large covariance matrices (ex: quadratic forms, ratios of the quadratic forms, SNIR Loss, performances of detection tests as ANMF, LR-ANMF, etc.) and 2) to correct it. At a finite distance (practical $m, N$ values), the corrected results are often valid.

- Sources localisation applications [F. Pascal, R. Couillet, ...]: the based-RMT Music algorithm (G-Music) is known to have higher performance than those of conventional algorithms when using all the eigenvalues of the covariance matrix.
- MIMO-STAP: the goal of A. Combernoux PHD thesis [Combernoux, 2016] was to analyse/improve the detection and filtering performances of low-rank detectors.
- Adaptive Radar Detection: when secondary data are correlated [Couillet et al., 2015].
- Hyperspectral Anomaly Detection - Unmixing: the goal of E. Terreaux PhD thesis [Terreaux et al., 2017] is to better analyse the rank of the anomalies space (model order selection) in Hyperspectral Imaging (high dimensional problem) for heterogeneous, correlated non-Gaussian environment.

# RMT key ideas (1)

Let $\{\mathbf{y}_i\}_{i\in[1,N]}$ be independent and distributed according to $\mathcal{CN}(\mathbf{0}_m, \mathbf{M})$. The Maximum Likelihood Estimate of $\mathbf{M}$ is the Sample Covariance Matrix given by

$$\widehat{\mathbf{M}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i\,\mathbf{y}_i^H = \frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H .$$

## Asymptotic Regime

If $N \to \infty$, then the strong law of large numbers says (or equivalently, in spectral norm):

$$\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\| \xrightarrow{a.s.} 0 .$$

## Random Matrix Regime

- No longer valid if $m$, $N \to \infty$ with $m/N \to c \in [0, \infty[$: $\left\|\widehat{\mathbf{M}} - \mathbf{M}\right\| \nrightarrow 0$,
- For practical large $m$, $N$ with $m \simeq N$, it can lead to dramatically wrong conclusions (even $m = N/100$).

# RMT key ideas (2)

Let $\{\mathbf{n}_i\}_{i\in[1,N]}$ be distributed according to $\mathcal{CN}\left(\mathbf{0}_m, \mathbf{C} = \sigma^2\,\mathbf{I}_m\right)$. We analyze the eigenvalues distribution of $\hat{\mathbf{C}} = \dfrac{1}{N}\sum_{i=1}^{N}\mathbf{n}_i\,\mathbf{n}_i^H = \dfrac{1}{N}\mathbf{N}\,\mathbf{N}^H$ where $c = m/N \in [0, \infty[$

## Random Matrix Regime

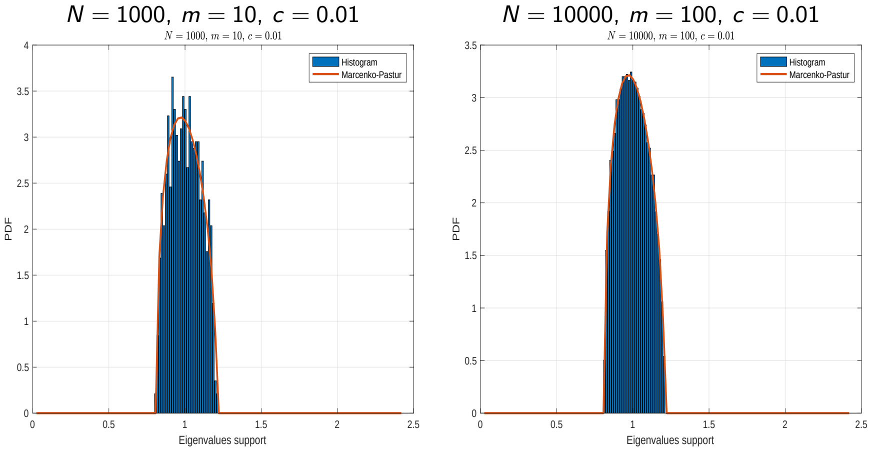
The distribution of the eigenvalues of $\hat{\mathbf{C}}$ tends almost surely toward the Marcenko-Pastur distribution

$$p(x) = \left(1 - \frac{1}{c}\right)_+ \delta(x) + \frac{1}{2\,\pi\,c\,x}\sqrt{(x-\lambda_-)_+\,(\lambda_+ - x)_+},$$

where $\lambda_- = \sigma^2\,\left(1-\sqrt{c}\right)^2$ and $\lambda_+ = \sigma^2\,\left(1+\sqrt{c}\right)^2$.

Not restricted to Gaussian statistics !

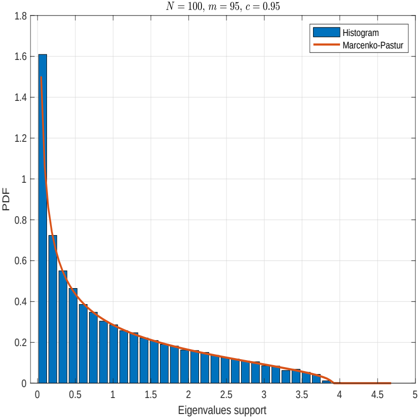
# RMT Examples (1): classical asymptotic regime



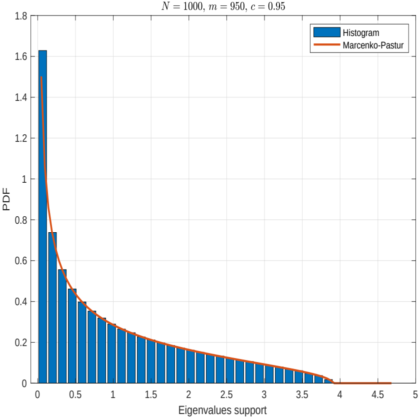

Eigenvalues support for **white** Gaussian noise ($\sigma^2 = 1$, $\mathbf{C} = \sigma^2\, \mathbf{I}_m$).

# RMT Examples (2): same RMT regime

$N = 100,\ m = 95,\ c = 0.95$

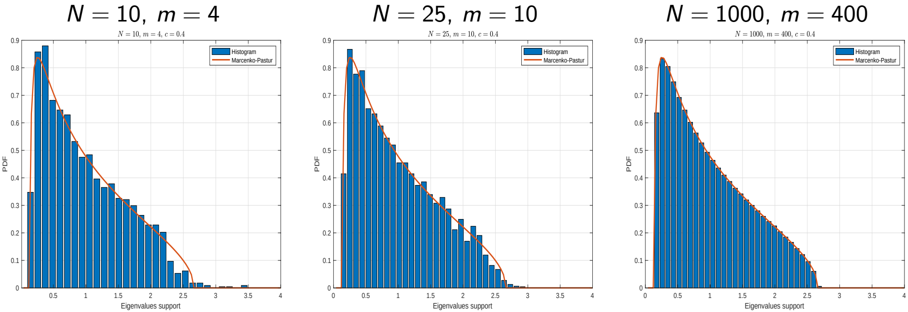



$N = 1000,\ m = 950,\ c = 0.95$




Eigenvalues support for **white** Gaussian noise ($\sigma^2 = 1$, $\mathbf{C} = \sigma^2\,\mathbf{I}_m$).

# RMT Examples (3): from where does start RMT regime ?

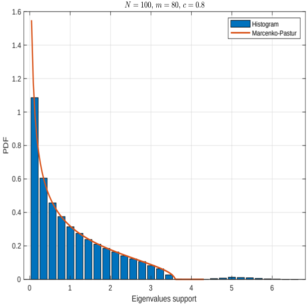


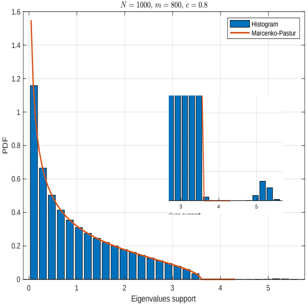

Eigenvalues support for **white** Gaussian noise ($\sigma^2 = 1$, $\mathbf{C} = \sigma^2\,\mathbf{I}_m$) and $c = 0.4$.

# Key ideas (3)

The behavior of the spectral measure brings information about the vast majority of the eigenvalues but is not affected by some individual eigenvalues behavior (like sources !). Whatever the perturbations (sources), the spectral measure converges toward Marcenko-Pastur distribution.

$N = 100,\ m = 80,\ c = 0.8$ $\quad$ $N = 1000,\ m = 800,\ c = 0.8$




SCM eigenvalues support for **white** Gaussian noise ($\sigma^2 = 1$, $\mathbf{C} = \sigma^2\, \mathbf{I}_m$) and sources.

# Source Detection with RMT

We consider $N$ observations $\left\{\mathbf{y}_k = \sqrt{\theta}\,\mathbf{u} + \mathbf{n}_k\right\}_{k\in[1,N]}$ with $\|\mathbf{u}\| = 1$. If the power $\theta$ of the source is **large enough**, then the limit of $\lambda_{max}\left(\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right)$ is strictly larger than the right edge of the bulk.

- if $\theta \leq \sigma^2\sqrt{c}$, then

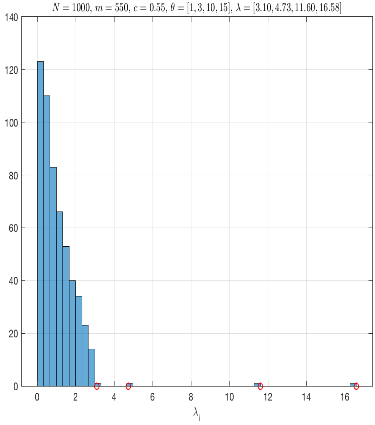
$$\lambda_{max}\left(\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right) \xrightarrow[N,m\to\infty]{a.s.} \sigma^2\,\left(1+\sqrt{c}\right)^2,$$

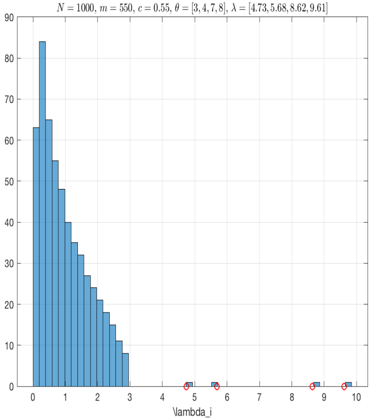
- if $\theta \geq \sigma^2\sqrt{c}$, then

$$\lambda_{max}\left(\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right) \xrightarrow[N,m\to\infty]{a.s.} \sigma^2\,(1+\theta)\left(1+\frac{c}{\theta}\right) \geq \sigma^2\,\left(1+\sqrt{c}\right)^2.$$

Above the threshold $\sigma^2\sqrt{c}$, $\lambda_{max}\left(\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right)$ asymptotically separates from the bulk.

# Source Detection with RMT



# Outline

# Model with correlated Gaussian noise

Let $\{\mathbf{y}_i\}_{i\in[1,N]}$ be $N$ observations of size $m$ characterizing the $p < m$ mixed sources corrupted by additive noise:

$$\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j}\,\mathbf{m}_j + \mathbf{C}^{1/2}\,\mathbf{n}_i\,, \qquad i \in [1, N]\,,$$

which can be rewritten more compactly as:

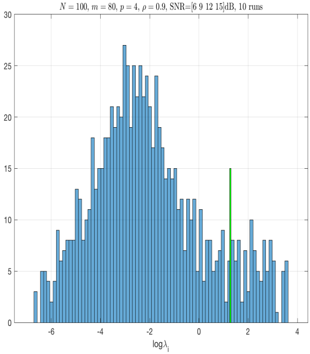
$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{N}\,,$$

where

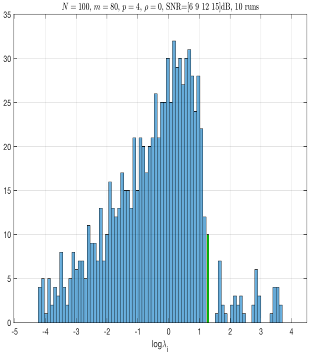
- $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{C}^{m\times N}$ are the observations,
- $\mathbf{M} \in \mathbb{C}^{m\times p}$ is the mixing matrix containing steering vectors of the $p$ sources,
- $\mathbf{S} \in \mathbb{C}^{p\times N}$ is the channel gain matrix,
- $\mathbf{N} \in \mathbb{C}^{m\times N}$ is the white Gaussian noise ($E\left[\mathbf{n}_i^H\,\mathbf{n}_i\right] = 1$), independent of the source signal,
- $\mathbf{C} \in \mathbb{C}^{m\times N}$ a **Toeplitz** Hermitian covariance matrix ($\mathrm{Tr}(\mathbf{C}) = m\,\sigma^2$).

# Problems due to the correlation

$c = 0.8,\ p = 4,\ \rho = 0.9$




$c = 0.8,\ p = 4,\ \rho = 0$




SCM eigenvalues support for Gaussian noise and 4 random sources (SNR= [6 9 12 15] dB. Left): **colored noise**. Right): **white noise**.

# Consistent Estimation for $\mathbf{C}$: Gaussian Case

## Proposition: [Terreaux et al., 2017]

*As $m, N \to \infty$ such that $m/N \to c \in [0, \infty[$, if $\mathbf{Y}$ does not contain sources, then:*

$$\left\| \mathcal{T}\left[\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right] - \mathbf{C} \right\| \xrightarrow{a.s.} 0\,,$$

*where $\mathcal{T}[\cdot]$ is the* **Toeplitz rectification** *operator:* $(\mathcal{T}[\mathbf{X}])_{ij} = \dfrac{1}{m}\displaystyle\sum_{k=1}^{m} \mathbf{X}_{k,k+|i-j|}$.

A consistent estimator $\hat{\mathbf{C}}$ of the background noise covariance matrix $\mathbf{C}$ characterizing the background noise is therefore defined through observations $\mathbf{Y}$ as $\hat{\mathbf{C}} = \mathcal{T}\left[\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right]$.

We can now **whiten** the observations $\mathbf{Y}$ by $\hat{\mathbf{C}}^{-1/2}\,\mathbf{Y}$.

# Behavior of whitened data: Gaussian Case

Let $\mathbf{Y}_w = \left(\mathcal{T}\left[\frac{1}{N}\mathbf{Y}\,\mathbf{Y}^H\right]\right)^{-1/2}\mathbf{Y}$ be the whitened data

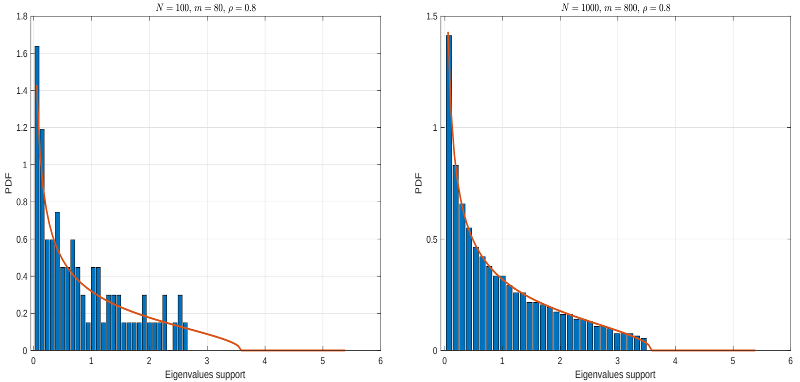
**Proposition: [Terreaux et al., 2017, Terreaux et al., 2018]**

*As $m, N \to \infty$ such that $m/N \to c \in [0, \infty[$, if $\mathbf{Y}_w$ does not contain sources, then:*

$$\left\|\frac{1}{N}\mathbf{Y}_w\,\mathbf{Y}_w^H - \frac{1}{N}\mathbf{N}\,\mathbf{N}^H\right\| \xrightarrow{a.s.} 0\,,$$

- Without sources, the spectral distribution of the whitened data covariance matrix of $\mathbf{Y}_w$ follows a Marchenko-Pastur distribution (same spectral distribution of unobservable covariance matrix of $\mathbf{N}$) characterized by its support $\left[\left(1-\sqrt{c}\right)^2, \left(1+\sqrt{c}\right)^2\right]$,
- All eigenvalues greater than $\left(1+\sqrt{c}\right)^2$ can be considered as sources,
- Detection occurs if $\text{SNR} = \dfrac{s_j^2\,\mathbf{m}_j^H\mathbf{m}_j}{m\,\sigma^2} \geq \sqrt{c}$.

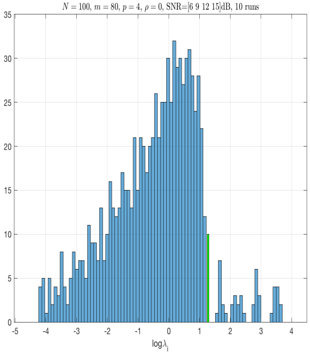
# Some RMT results: Gaussian Case with no-sources



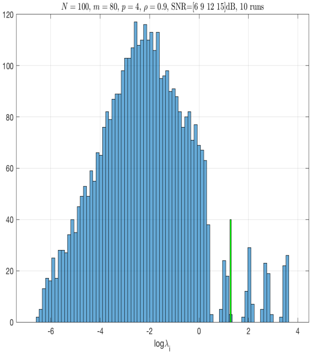

$$\left\| \frac{1}{N}\,\mathbf{Y}_w\,\mathbf{Y}_w^H - \frac{1}{N}\,\mathbf{N}\,\mathbf{N}^H \right\| \xrightarrow{a.s.} 0\,, \qquad E\left[\mathbf{N}\,\mathbf{N}^H\right] = \mathbf{I}_m$$

# Example: Gaussian noise and 4 random sources

$c = 0.8,\ p = 4,\ \rho = 0$ , 10 runs



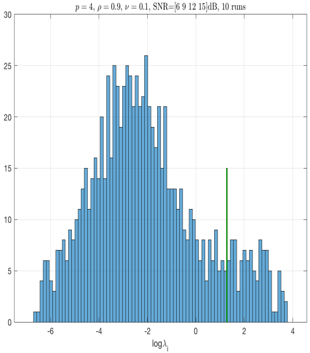

$c = 0.8,\ p = 4,\ \rho = 0.9$ , 10 runs

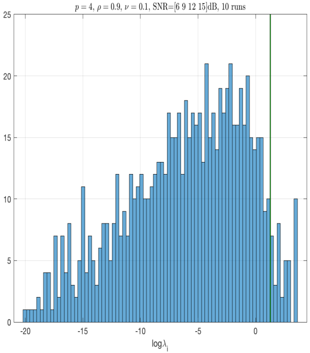



SCM eigenvalues support for Gaussian noise and sources (SNR= [6 9 12 15]dB). Left): **white noise**. Right): **whitened** colored noise .

# Problems in non-Gaussian case

$c = 0.8,\ p = 4,\ \rho = 0.9,\ \nu = 0.1$ $\qquad$ $c = 0.8,\ p = 4,\ \rho = 0.9,\ \nu = 0.1$




SCM eigenvalues support for K-distributed noise and sources (SNR= [6 9 12 15]dB).
Left): **colored K-distributed noise**. Right): **whitened colored K-distribution noise**.

# Outline

# Model with correlated Non Gaussian (CES) noise

Let $\{\mathbf{y}_i\}_{i\in[1,N]}$ be $N$ observations of size $m$ characterizing the $p < m$ mixed sources corrupted by additive noise:

$$\mathbf{y}_i = \sum_{j=1}^{p} s_{i,j}\,\mathbf{m}_j + \sqrt{\tau_i}\,\mathbf{C}^{1/2}\,\mathbf{n}_i\,, \qquad i \in [1,N]\,,$$

which can be rewritten more compactly as:

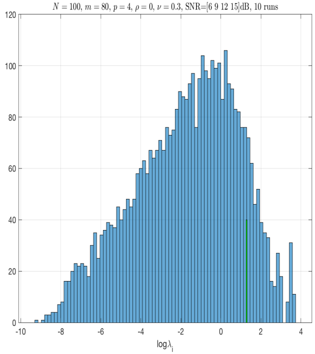
$$\mathbf{Y} = \mathbf{M}\,\mathbf{S} + \mathbf{C}^{1/2}\,\mathbf{N}\,\mathbf{T}^{1/2}\,,$$

where

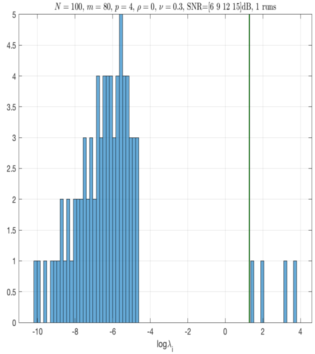
- $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{C}^{m\times N}$ are the observations,
- $\mathbf{M} \in \mathbb{C}^{m\times p}$ is the mixing matrix containing steering vectors of the $p$ sources,
- $\mathbf{S} \in \mathbb{C}^{p\times N}$ is the channel gain matrix, $\mathbf{T}$ is the texture diagonal matrix,
- $\mathbf{N} \in \mathbb{C}^{m\times N}$ is the white Gaussian noise ($E\left[\mathbf{n}_i^H\,\mathbf{n}_i\right] = 1$), independent of the source signal,
- $\mathbf{C} \in \mathbb{C}^{m\times N}$ a **Toeplitz** Hermitian covariance matrix ($\mathrm{Tr}(\mathbf{C}) = m\,\sigma^2$).

# Key idea 1: to use Robust Covariance Matrix Estimation

$N = 100,\ m = 80,\ c = 0.8,\ p = 4$



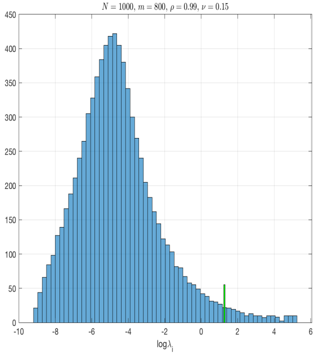

$N = 100,\ m = 80,\ c = 0.8,\ p = 4$



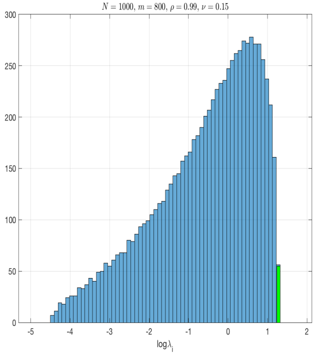

Eigenvalues support for **white K-distributed** noise ($\sigma^2 = 1$, $\nu = 0.3$) and 4 sources (SNR= [6 9 12 15]dB). Left): **SCM**. Right): **Tyler**.

# Key idea 2: to whiten the correlated data

$c = 0.8,\ \nu = 0.15,\ \rho = 0.99$




$c = 0.8,\ \nu = 0.15,\ \rho = 0.99$




Tyler eigenvalues support for correlated K-distributed noise ($\sigma^2 = 1$). Left): **unwhitened data.** Right): **whitened data** (right).

# Robust Consistent Estimation for $\mathbf{C}$: General case

Let $\hat{\mathbf{M}}_{FP} = \frac{m}{N}\sum_{i=1}^{N} \frac{\mathbf{y}_i\,\mathbf{y}_i^H}{\mathbf{y}_i^H\,\hat{\mathbf{M}}_{FP}^{-1}\,\mathbf{y}_i}$ be the Tyler M-estimator of $\mathbf{Y}$ scatter matrix.

## Proposition: [Terreaux et al., 2017]

*As $m, N \to \infty$ such that $m/N \to c \in [0, \infty[$, if $\mathbf{Y}$ does not contain sources, then:*

$$\left\| \mathcal{T}\left[\hat{\mathbf{M}}_{FP}\right] - \mathbf{C} \right\| \xrightarrow{a.s.} 0\,,$$

*where $\mathcal{T}[\cdot]$ is the* **Toeplitz rectification** *operator:* $(\mathcal{T}[\mathbf{X}])_{ij} = \frac{1}{m}\sum_{k=1}^{m} \mathbf{X}_{k,k+|i-j|}$.

A consistent estimator $\hat{\mathbf{C}}$ of the background noise covariance matrix $\mathbf{C}$ characterizing the background noise is therefore defined through observations $\mathbf{Y}$ as $\hat{\mathbf{C}} = \mathcal{T}\left[\hat{\mathbf{M}}_{FP}\right]$.

We can now **whiten** the observation $\mathbf{Y}$ by $\hat{\mathbf{C}}^{-1/2}\,\mathbf{Y}$.

# Behavior of whitened data: General case

Let $\mathbf{Y}_w = \left(\mathcal{T}\left[\hat{\mathbf{M}}_{FP}\right]\right)^{-1/2}\mathbf{Y}$ be the whitened data and $\hat{\mathbf{W}}_{FP}$ be the Tyler M-estimator of $\mathbf{Y}_w$.

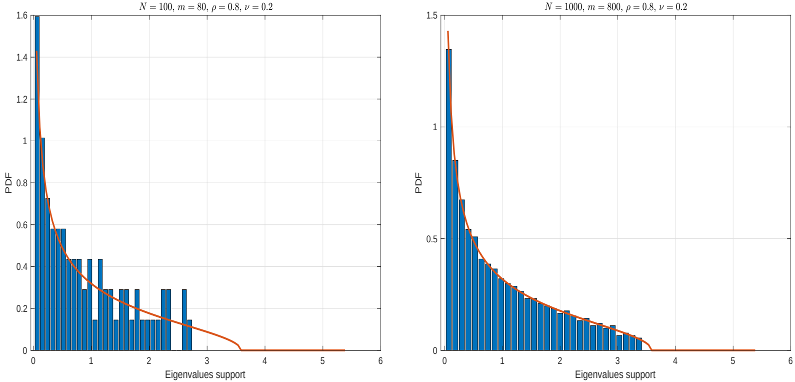
## Proposition: [Terreaux et al., 2017]

*As $m, N \to \infty$ such that $m/N \to c \in [0, \infty[$, if $\mathbf{Y}_w$ does not contain sources, then:*

$$\left\|\hat{\mathbf{W}}_{FP} - \frac{1}{N}\mathbf{N}\,\mathbf{N}^H\right\| \xrightarrow{a.s.} 0\,,$$

- Without sources, the spectral distribution of the whitened data scatter matrix of $\mathbf{Y}_w$ follows a Marchenko-Pastur distribution (same spectral distribution of unobservable covariance matrix of $\mathbf{N}$) characterized by its support $\left[\left(1-\sqrt{c}\right)^2, \left(1+\sqrt{c}\right)^2\right]$,
- All eigenvalues greater than $\left(1+\sqrt{c}\right)^2$ can be considered as sources,

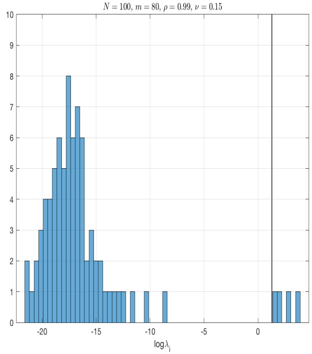
# Some RMT results: Non-Gaussian Case with no-sources

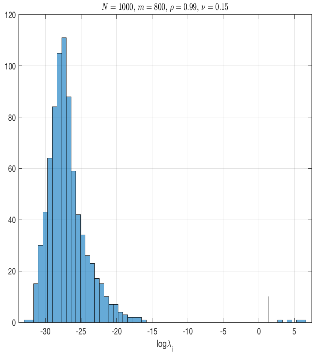



$$\left\| \hat{\mathbf{W}}_{FP} - \frac{1}{N} \mathbf{N}\,\mathbf{N}^H \right\| \xrightarrow{a.s.} 0\,, \qquad E\left[\mathbf{N}\,\mathbf{N}^H\right] = \mathbf{I}_m$$

# Example with sources

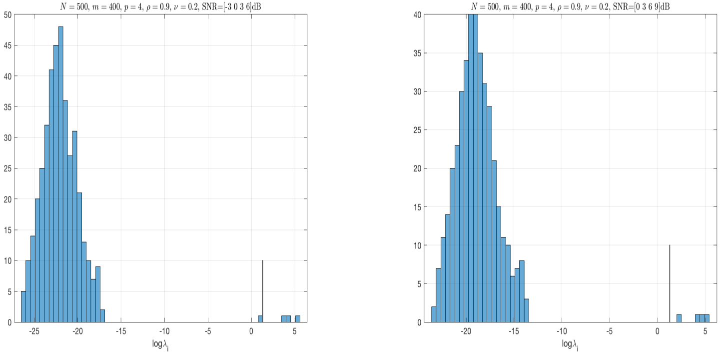
$c = 0.8,\ p = 4,\ \rho = 0.99,\ \nu = 0.15$ $\qquad$ $c = 0.8,\ p = 4,\ \rho = 0.99,\ \nu = 0.15$



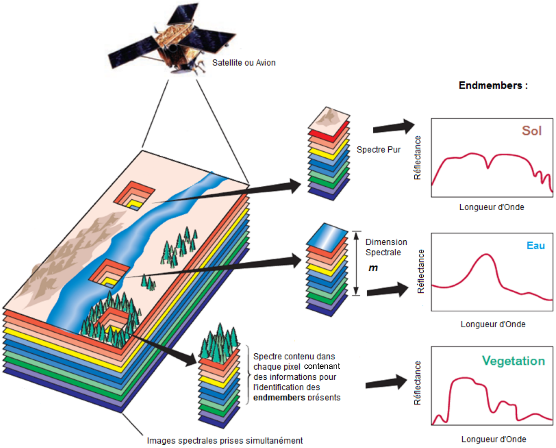

Tyler eigenvalues support for colored K-distributed noise and 4 sources (SNR= [3 6 9 10] dB). Left): $N = 100,\ m = 80$. Right): $N = 1000,\ m = 800$.

# SNR Impact




Tyler eigenvalues support for whitened observations (4 random sources and colored K-distributed noise). Left): **SNR= [-3 0 3 6] dB.** Right): **SNR= [0 3 6 9] dB.**

# Hyperspectral Imaging




## General Problems

- Estimation of the endmembers number
- Detection/Estimation of sources / Anomaly Detection
- Unmixing

## Considered Problems

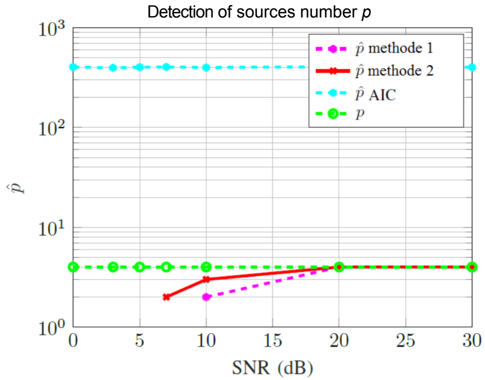
- Estimation of the number of endmembers
- Estimation of their spectrum

# Hyperspectral Imaging

With the set of observations $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$

- Estimation of the noise scatter matrix $\hat{\mathbf{C}} = \mathcal{T}\left[\hat{\mathbf{M}}\right]$ by Toeplitz rectification on:
  - Method 1: Maronna's M-estimators [Maronna, 1976] adapted to data statistic : $\hat{\mathbf{M}} = \frac{1}{N}\sum_{i=0}^{N-1} u\left(\frac{1}{m}\check{\mathbf{y}}_i^H\,\hat{\mathbf{M}}^{-1}\,\mathbf{y}_i\right)\,\mathbf{y}_i\,\mathbf{y}_i^H$.
  - Method 2: Tyler's M-estimator : $\hat{\mathbf{M}} = \frac{m}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_i\,\mathbf{y}_i^H}{\mathbf{y}_i^H\,\hat{\mathbf{M}}^{-1}\,\mathbf{y}_i}$.
- Whitening observations: $\mathbf{Y}_w = \left(\mathcal{T}\left[\hat{\mathbf{M}}\right]\right)^{-1/2}\mathbf{Y}$.
- Thresholding the eigenvalue distribution of the whitened data scatter matrix $\hat{\mathbf{W}}$:
  - Method 1: Threshold depending on the function $u(.)$ and data for Maronna's M-estimator $\hat{\mathbf{W}}$,
  - Method 2: Threshold independent of data for Tyler's M-estimator $\hat{\mathbf{W}}$,

# Hyperspectral Imaging




4 sources, $N = 2000$, $m = 900$, $\{\tau\}_{i \in [\![1, N]\!]}$ inverse-gamma

# Hyperspectral Imaging

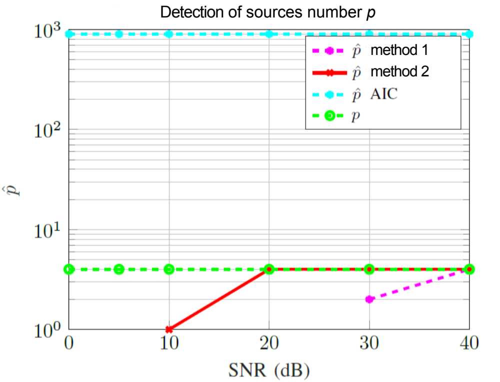
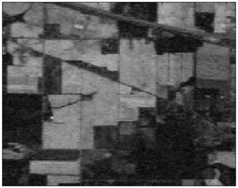
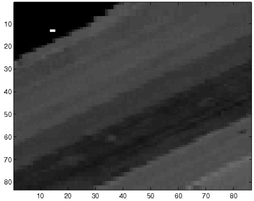



Fig 4 sources, $N = 2000$, $m = 900$, $\{\tau\}_{i \in [\![1, N]\!]}$ Student-t

# Hyperspectral Imaging

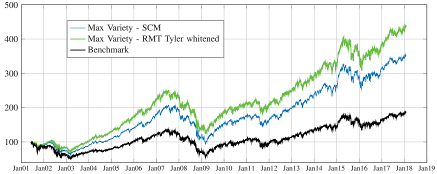


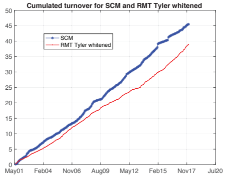


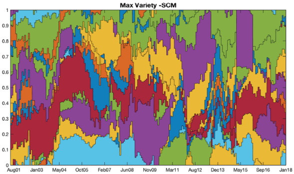

Estimation of the number of the most energetic endmembers

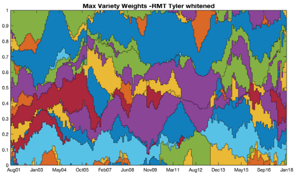
| | Images | Indian Pines | SalinasA | PaviaU | Cars |
|---|---|---|---|---|---|
| | $p$ | 16 | 9 | 9 | 6 |
| | $\hat{p}_{AIC}$ | 219 | 203 | 102 | 143 |
| | $\hat{p}_{Hysime}$ | **19** | 14 | 60 | 19 |
| Method 1 | $\hat{p}_{FP}$ | 11 | **9** | 1 | **3** |
| Method 2 | $\hat{p}_{TYL}$ | **13** | 2 | **10** | **13** |

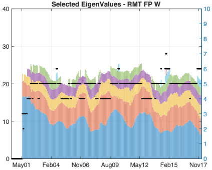
# Portfolio Performance Optimization [Jay et al., 2018]




| Max Variety Portfolios | Ann. Return | Ann. Volatility | Ratio (Ret / Vol) | Max DD |
|---|---|---|---|---|
| RMT Tyler Whithened | 9,71% | 12,9% | 0,75 | 50,41% |
| SCM | 8,51% | 13,80% | 0,62 | 55,02% |
| Benchmark | 4,92% | 15,19% | 0,32 | 58,36% |

# Outline

# Conclusions

This work has extended classical Model Order Selection techniques (AIC, MDL, etc.) for correlated and non-Gaussian additive noise.

- This extension was efficiently derived using latest results coming from RMT assuming **Toeplitz covariance structure assumption** for the noise covariance matrix,
- This quite *simple* technique can be easily applied on experimental data (radar, STAP, MIMO-STAP, SAR, HS, finance).

# The End

**Thank You !**

**I Wish You a Great Independence Day !**

Akaike, H. (1974).
A new look at the statistical model identification.
*IEEE Transactions on Automatic Control*, 19(6):716–723.

Anderson, T. W. (1963).
Asymptotic theory for principal component analysis.
*Ann. J. Math. Stat.*, 34:122–148.

Combernoux, A. (2016).
*Low rank detection and estimation using random matrix theory approaches for antenna array processing*.
Theses, Université Paris-Saclay.

Couillet, R. and Debbah, M. (2011).
*Random matrix methods for wireless communications*.
Cambridge University Press.

Couillet, R., Greco, M. S., Ovarlez, J.-P., and Pascal, F. (2015).
RMT for whitening space correlation and applications to radar detection.
In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 149–152.

Jay, E., Terreaux, E., Ovarlez, J.-P., and Pascal, F. (2018).
Improving portfolios global performance with robust covariance matrix estimation: Application to the maximum variety portfolio.
In *Proc. Eusipco 2018*.

Maronna, R. A. (1976).
Robust $M$-estimators of multivariate location and scatter.
*Annals of Statistics*, 4(1):51–67.

Rissanen, J. (1978).
Modeling by shortest data description.
*Automatica*, 14(5):465–471.

Schwarz, G. (1978).
Estimating the dimension of a model.
*The Annals of Statistics*, 6(2):461–464.

Terreaux, E., Ovarlez, J.-P., and Pascal, F. (2017).
Robust model order selection in large dimensional elliptically symmetric noise.
*arXiv preprint*.

Terreaux, E., Ovarlez, J.-P., and Pascal, F. (2018).
A Toeplitz-Tyler estimation of the model order in large dimensional regime.
In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4489–4493.

Wax, M. and Kailath, T. (1985).
Detection of signals by information theoretic criteria.
*IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392.