A Toeplitz-Tyler Estimation of the Model Order in Large Dimensional Regime

Eugenie Terreaux[1], Jean-Philippe Ovarlez[1,2], Frederic Pascal[3]

[1] SONDRA, CentraleSupélec, University Paris-Saclay, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France
[2] ONERA, DEMR/TSI, Chemin de la Hunière, F-91190 Gif sur Yvette, France
[3] L2S, CentraleSupélec, CNRS, University Paris-Saclay, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France

Introduction and Motivations

Let us consider a set of N **observations** $\{\mathbf{y}_i\}_{i \in [\![1,N[\![}$ where each $\mathbf{y}_i$ is a **multidimensional** m**-vector**.

GOAL: Estimate the model order p for high dimensional and Complex Elliptically Symmetric (CES) distributed signal

- **Large number of data**: N and m are of same order with possibly $N > m$ $(N, m) \to \infty$ with $m/N \to c \neq 0$
- **CES noise**: $\{\mathbf{n}_i = \sqrt{\tau_i}\,\mathbf{C}^{1/2}\,\mathbf{x}_i\}_{i \in [\![1,N]\!]}$

. $\mathbf{x}_i$ is a m-vector uniformly distributed on the sphere of dimension m,
. τ_i is a positive scalar random variable representing the texture,
. $\mathbf{C}$ is a Hermitian Toeplitz scatter matrix.

Statistical Model:

$$\mathbf{y}_i = \sum_{j=1}^{p} a_{i,j}\,\mathbf{m}_j + \sqrt{\tau_i}\,\mathbf{C}^{1/2}\,\mathbf{x}_i\,, \quad i \in [\![1,N]\!]$$

. $\{\mathbf{m}_j\}$ unknown independent m-steering vector of p sources
. $a_{i,j}$ amplitude of the j-th source in the i-th observation.

Proposed Algorithm:

. **Estimation** of the noise scatter matrix $\mathbf{C}$
. **Whitening** the received signal
. **Robust Estimation** of the scatter matrix of the whitened observations
. **Thresholding** the eigenvalues of the estimated scatter matrix and **estimation of the model order** p

Assumptions:

. $\sum_{i=1}^{N} \delta_{\tau_i}$ satisfies $\int \tau \mu_N(d\tau) \xrightarrow{a.s.} 1$.

. $\{c_k\}_{k \in [0,m-1]}$ are absolutely summable coefficients, such that $c_0 \neq 0$.

. $\frac{1}{N}\sum \delta_{\lambda_i(\mathbf{C})}$ converges almost surely toward the true measure ν, λ_i the $i-th$ largest eigenvalue of $\mathbf{C}$

Signal Whitening

- Tyler M-estimator of the scatter matrix $\mathbf{C}$ enforced to be Toeplitz-structured with the operator $\mathcal{T}$ [1] and [2]

Definitions and Notations:

. Tyler M-estimator $\hat{\mathbf{C}}$ is the unique solution of $\mathbf{\Sigma} = \frac{m}{N}\sum_{i=1}^{N}\frac{\mathbf{y}_i\mathbf{y}_i^H}{\mathbf{y}_i^H\mathbf{\Sigma}^{-1}\mathbf{y}_i}$

$$\mathcal{T}\left(\hat{\mathbf{C}}\right) = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m}\hat{c}_{i,i} & \frac{1}{m}\sum_{i=1}^{m-1}\hat{c}_{i,i+1} & \cdots & \frac{1}{m}\hat{c}_{1,m} \\ \frac{1}{m}\sum_{i=2}^{m}\hat{c}_{i,i-1} & \ddots & \ddots & \\ \vdots & & \ddots & \\ \frac{1}{m}\hat{c}_{m,1} & & & \frac{1}{m}\sum_{i=1}^{m}\hat{c}_{i,i} \end{pmatrix}$$

Theorem 1: Consistent estimator of $\mathbf{C}$

$$\left\| \mathcal{T}\left(\hat{\mathbf{C}}\right) - \mathbf{C} \right\| \xrightarrow{a.s.} 0$$

- Whitening: $\hat{\mathbf{Y}}_w = \hat{\mathbf{C}}^{-1/2}\,\mathbf{Y} = \hat{\mathbf{C}}^{-1/2}\,\mathbf{M}\,\mathbf{A} + \hat{\mathbf{C}}^{-1/2}\,\mathbf{C}^{-1/2}\,\mathbf{X}\,\mathbf{T}^{1/2}$

Signal Subspace Rank Estimation

- Tyler estimation of the white scatter matrix: $\hat{\mathbf{W}} = \frac{m}{N}\sum_{i=1}^{N}\frac{\hat{\mathbf{y}}_{wi}\hat{\mathbf{y}}_{wi}^H}{\hat{\mathbf{y}}_{wi}^H\hat{\mathbf{W}}^{-1}\hat{\mathbf{y}}_{wi}}$
- Sample Covariance Matrix: $\hat{\mathbf{S}}_w = \frac{1}{N}\mathbf{X}\mathbf{X}^H$

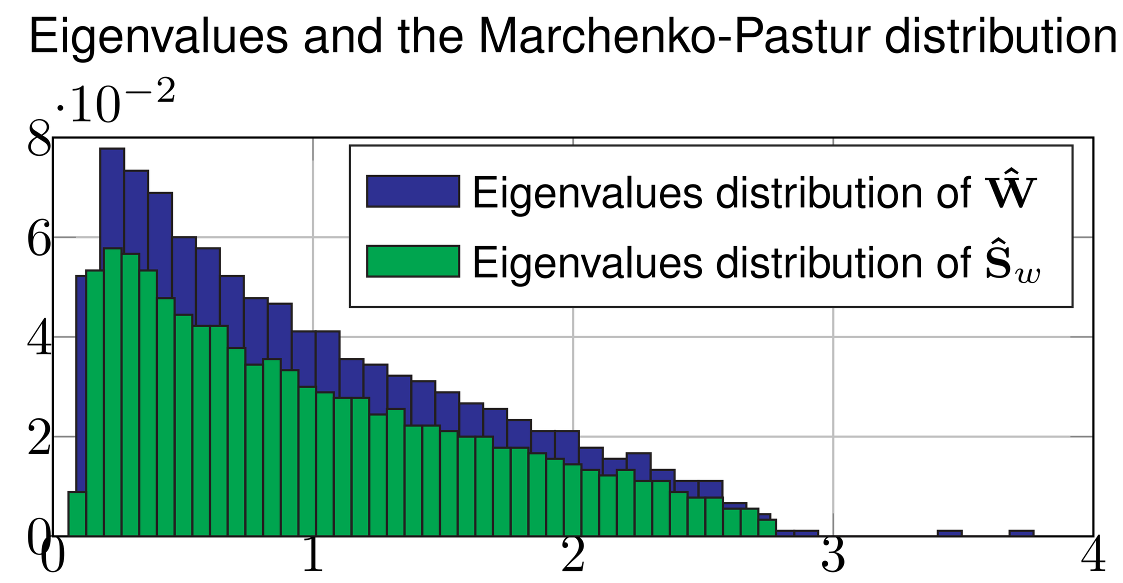
Theorem 2: Convergence of $\hat{\mathbf{W}}$

$$\left\| \hat{\mathbf{W}} - \hat{\mathbf{S}}_w \right\| \xrightarrow{a.s.} 0$$

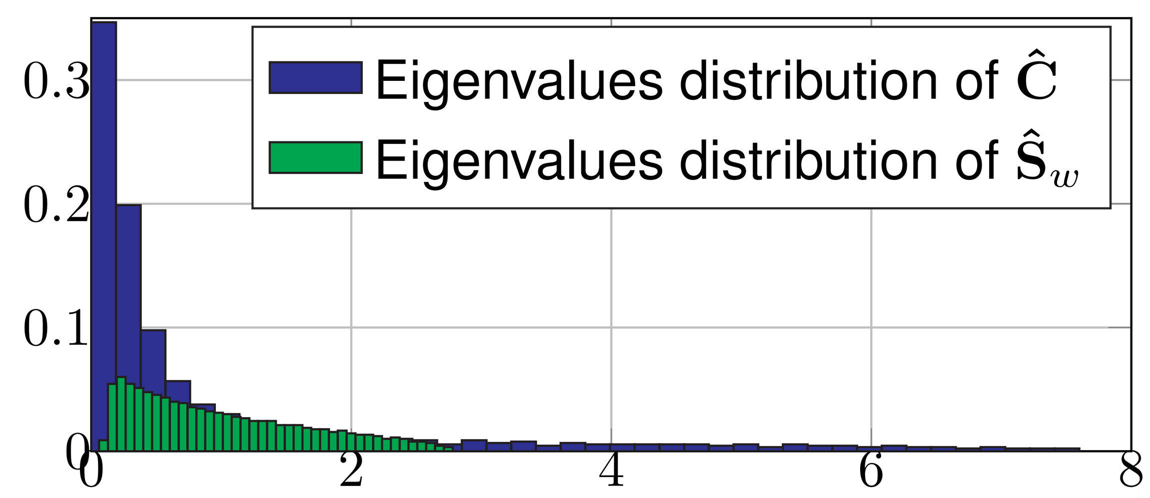
- Distribution of the eigenvalues of $\hat{\mathbf{S}}_w \xrightarrow{a.s.}$ Marchenko-Pastur distribution
- Marchenko Pastur distribution: **compact** and **known** support $\longrightarrow$ **threshold** $(1+\sqrt{c})^2$ on the eigenvalues of $\hat{\mathbf{W}}$ to separate noise and sources and **estimation** of p

Results and Simulations

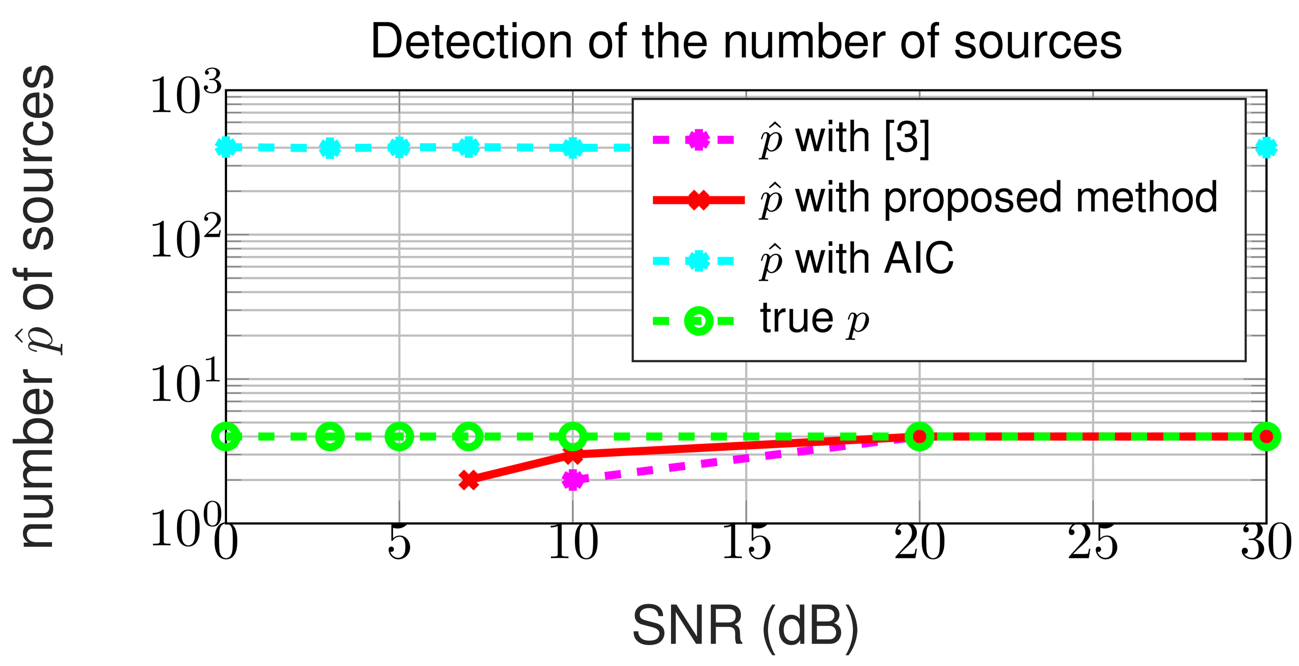
- Distribution of the eigenvalues for $p = 4$, $N = 2000$, $m = 900$, τ_i, $i \in [\![0,N]\!]$ Inverse Gamma distributed, 4 eigenvalues are upon the Marchenko-Pastur distribution:



- For unwhitened signal, that is for $\hat{\mathbf{C}}$, no sources can be detected upon the Marchenko-Pastur distribution :



- Comparison between different methods for different SNR and $p = 4$, $N = 2000$, $m = 900$, τ_i, $i \in [\![0,N]\!]$ inverse gamma distributed



- Our algorithm gives interesting and encouraging results.

Conclusion

This method can be generally applied for any model order selection problems as radar clutter rank estimation, sources localization or any hyperspectral problems such as anomaly detection or linear or non-linear unmixing techniques.

[1] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," The Annals of Statistics, vol. 15, no. 1, pp. 234–251, March 1987.
[2] T. Zhang, C. Xiuyuan, and A. Singer, "Marcenko-Pastur law for Tyler's and Maronna's M-estimators," Journal of Multivariate Analysis, 149, pp.114-123, 2016
[3] E. Terreaux, J. P. Ovarlez, and F. Pascal, "Robust model order selection in large dimensional elliptically symmetric noise," arXiv preprint, https://arxiv.org/abs/1710.06735, 2017.