Random matrix methods for statistical signal processing.

Philippe Loubaton
Université Paris-Est Marne la Vallée
Laboratoire d'Informatique Gaspard Monge

14 December 2018

General context I.

The observation: N samples $(\mathbf{y}_n)_{n=1,\ldots,N}$ extracted from a M–dimensional time series ($\dim(\mathbf{y}_n) = M$).

Develop statiscal inference methods (detection, estimation,...) when M and N are large and of the same order of magnitude.

Relevance of this context

- M large becomes very comon due to the development of large sensor networks
- N is of the same order of magnitude than M when the duration of the observation is limited.

The high-dimensional asymptotic regime:

$M \to +\infty, N \to +\infty, c_N = \frac{M}{N} \to c_*, 0 < c_* < +\infty$

Observations can be collected into the large $M \times N$ random matrix $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$.

Example of new difficulties posed by the high-dimensional system regime.

Assume $(\mathbf{y}_n)_{n=1,\ldots,N}$ are i.i.d. zero mean Gaussian vectors with covariance matrix $\mathbf{R}$, and that we have to estimate $\theta = f(\mathbf{R})$.

Empirical covariance matrix $\hat{\mathbf{R}}_N = \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^* = \frac{\mathbf{Y}\mathbf{Y}^*}{N}$.

If M is fixed and $N \to +\infty$, $\|\hat{\mathbf{R}}_N - \mathbf{R}\| \to 0$ and $\hat{\theta}_N = f(\hat{\mathbf{R}}_N) \to \theta = f(\mathbf{R})$ when $N \to +\infty$. Estimator $\hat{\theta}_N$ is said to be consistent.

In the high-dimensional regime

- $\|\hat{\mathbf{R}}_N - \mathbf{R}_N\|$ does not converge towards 0
- $\hat{\theta}_N = f(\hat{\mathbf{R}}_N)$ does not behave as $\theta_N = f(\mathbf{R}_N)$

Large random matrix results allow to predict the asymptotic behaviour of $f(\hat{\mathbf{R}}_N)$ for certain functions f. This allows to build modified consistent estimates of $\theta = f(\mathbf{R})$.

Some early contributions I.

- The very first: V. Girko in the eighties, G-estimation theory, ”An introduction to statistical analysis of random arrays”, 700p., 1998.
  - Consistent estimation of $f(\mathbf{R})$ from $\hat{\mathbf{R}}$ in the Gaussian i.i.d. case, e.g. $\frac{1}{M}\,\mathrm{Tr}(\mathbf{R} + \sigma^2\mathbf{I})^{-1}$:

$$\frac{1}{M}\,\mathrm{Tr}(\mathbf{R} + \hat{\phi}(\omega^2)\mathbf{I})^{-1} \simeq \frac{\frac{1}{M}\mathrm{Tr}(\hat{\mathbf{R}} + \omega^2\mathbf{I})^{-1}}{1 - c_N + c_N\omega^2\frac{1}{M}\mathrm{Tr}(\hat{\mathbf{R}} + \omega^2\mathbf{I})^{-1}}$$

    $\hat{\phi}(\omega^2) = \sigma^2$ has a unique positive solution $\hat{\omega}_N^2$.
  - Consistent estimation of $f(\mathbf{B}\mathbf{B}^*)$ when $\mathbf{Y} = \mathbf{B} + \mathbf{V}$, $\mathbf{B}$ deterministic, $\mathbf{V}$ Gaussian matrix with i.i.d. entries (the Information plus Noise model)
- X. Mestre
  - Consistent estimation of the eigenvalues of $\mathbf{R}$ when the number of distinct eigenvalues of $\mathbf{R}$ remains fixed (2008)
  - Consistent estimation of direction of arrivals in the context of large sensor networks (2008)

Some early contributions II.

Detection of low rank signals in white noise

Johnstone (2001), Krichtman-Nadler and Nadler (2008, 2010), Nadakuditi-Edelman (2008), Nadakuditi-Silverstein (2010), Bianchi-Debbah-Maida-Najim (2009),...

- Hypothesis H_0, $\mathbf{Y} = \mathbf{V}$, $\mathbf{V}$ Gaussian matrix with i.i.d. entries
- Hypothesis H_1, $\mathbf{Y} = \mathbf{AS} + \mathbf{V}$, $\mathbf{A}$ deterministic $M \times K$ matrix, $\mathbf{S}$ random with i.i.d. entries or deterministic $K \times N$ matrix, K small compared to M and N, i.e. K does not scale with M, N, additive spiked model.

Corrections of log likelihood ratio procedures

Bai, Jiang, Yao, Zheng (2009)

Applications to source localization.

$$\underset{\text{Observation}}{\begin{bmatrix} \mathbf{y}_1 \cdots \mathbf{y}_N \end{bmatrix}} = \underset{\text{Channel}}{\begin{bmatrix} \mathbf{a}_1 \cdots \mathbf{a}_K \end{bmatrix}} \underset{\text{Source signals}}{\begin{bmatrix} \mathbf{s}^1 \\ \cdots \\ \mathbf{s}^K \end{bmatrix}} + \underset{\text{Noise}}{\begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_N \end{bmatrix}}$$

$$\underset{M \times N}{\mathbf{Y}} = \underset{M \times K}{\mathbf{A}} \quad \underset{K \times N}{\mathbf{S}} + \underset{M \times N}{\mathbf{V}}$$

- $\mathbf{V}$ random matrix with i.i.d. $\mathcal{N}(0, \sigma^2)$ entries
- K very small compared to M and N: K does not scale with M, N, additive spiked model
- $\mathbf{A}$ deterministic non observable, but $\mathbf{A} = (\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K))$, fonction $\theta \to \mathbf{a}(\theta)$ connue.
- We assume that $\mathbf{a}(\theta) = \frac{1}{\sqrt{M}}(1, e^{i\theta}, \ldots, e^{i(M-1)\theta})^T$
- $\mathbf{S}$ deterministic non observable, $\sup_N \|\frac{\mathbf{S}}{\sqrt{N}}\| < +\infty$, implies that $\sup_N \|\frac{\mathbf{AS}}{\sqrt{N}}\| < +\infty$
- Spectral norms of the signal and the noise of the same order of magnitude

Estimate $\theta_1, \ldots, \theta_K$ from $\mathbf{Y}$.

A popular approach: the subspace method.

$$\mathbf{Y} = \mathbf{AS} + \mathbf{V} = (\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K))\mathbf{S} \ + \mathbf{V}$$

Subspace approach

$\mathbf{\Pi}$ orthogonal projection matrix on the range space of $\mathbf{A}$,
$\mathbf{\Pi}^{\perp} = \mathbf{I} - \mathbf{\Pi}$:

$$\mathbf{a}(\theta)^* \mathbf{\Pi}^{\perp} \mathbf{a}(\theta) = 0$$

if and only if $\theta \in \{\theta_1, \ldots, \theta_K\}$
$\theta \to \mathbf{a}(\theta)^* \mathbf{\Pi}^{\perp} \mathbf{a}(\theta)$ localisation function.

- Estimate $\mathbf{\Pi}^{\perp}$ by $\hat{\mathbf{\Pi}}^{\perp}$
- Estimate the angles by the K most significant local minima of $\theta \to \mathbf{a}(\theta)^* \hat{\mathbf{\Pi}}^{\perp} \mathbf{a}(\theta)$, traditional estimated localisation function.

The traditional subspace approach, M fixed, $N \to +\infty$

$$\hat{\mathbf{R}}_N = \frac{\mathbf{Y}\mathbf{Y}^*}{N} \simeq \mathbb{E}(\hat{\mathbf{R}}_N) = \mathbf{A}\frac{\mathbf{S}\mathbf{S}^*}{N}\mathbf{A}^* + \sigma^2\mathbf{I} = \mathbf{R}_s + \sigma^2\mathbf{I}$$

Estimate $\mathbf{\Pi}^\perp$ by the orthogonal projection matrix $\hat{\mathbf{\Pi}}_t^\perp$ on the eigenspace associated to the smallest $M-K$ eigenvalues of $\hat{\mathbf{R}}_N$.

- $\|\mathbf{\Pi}^\perp - \hat{\mathbf{\Pi}}_t^\perp\| \to 0$
- Implies that the local minima of $\theta \to \mathbf{a}(\theta)^*\hat{\mathbf{\Pi}}_t^\perp\mathbf{a}(\theta)$ are consistent estimators in the regime M fixed and $N \to +\infty$.

Not valid in the high-dimensional regime

Notations

Spectral factorizations:

$$\frac{\mathbf{A}\mathbf{S}_N\mathbf{S}_N^*\mathbf{A}^*}{N} = \begin{bmatrix} \mathbf{u}_{1,N} & \cdots & \mathbf{u}_{K,N} \end{bmatrix} \begin{bmatrix} \lambda_{1,N} & & \\ & \ddots & \\ & & \lambda_{K,N} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{1,N} & \cdots & \mathbf{u}_{K,N} \end{bmatrix}$$

where $\lambda_{1,N} \geq \cdots \geq \lambda_{K,N}$.

$$\frac{\mathbf{Y}_N\mathbf{Y}_N^*}{N} = \begin{bmatrix} \hat{\mathbf{u}}_{1,N} & \cdots & \hat{\mathbf{u}}_{M,N} \end{bmatrix} \begin{bmatrix} \hat{\lambda}_{1,N} & & \\ & \ddots & \\ & & \hat{\lambda}_{M,N} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}}_{1,N} & \cdots & \hat{\mathbf{u}}_{M,N} \end{bmatrix}^*$$

where $\hat{\lambda}_{1,N} \geq \cdots \geq \hat{\lambda}_{M,N}$.

The high-dimensional regime

Can be deduced from Benaych-Georges and Nadakuditi, 2012

- Assume that $\lambda_{k,N} \to \rho_k$, for $k = 1, \ldots, K$, $\rho_1 > \ldots > \rho_K$
- Assume that $\rho_K > \sigma^2 \sqrt{c_*}$.
- Then, for $k = 1, \ldots, K$, for any sequence $\mathbf{b}_N$ of deterministic $M \times 1$ vectors such that $\sup_N \|\mathbf{b}_N\| < \infty$,

$$\mathbf{b}_N^* \mathbf{u}_{k,N} \mathbf{u}_{k,N}^* \mathbf{b}_N - \frac{\mathbf{b}_N^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{b}_N}{h(\hat{\lambda}_{k,N})} \xrightarrow[N \to \infty]{\text{a.s.}} 0$$

where $h(x)$ is a known function depending on the Stieljes transform of the MP distribution associated to matrix $\frac{\mathbf{V}\mathbf{V}^*}{N}$.

Modified subspace estimator (Vallet-L-Mestre-2015).

Assume that $\lim_{N \to +\infty} \lambda_{K,N} > \sigma^2 \sqrt{c_*}$

$$
\begin{aligned}
\mathbf{a}(\theta)^* \boldsymbol{\Pi}^{\perp} \mathbf{a}(\theta) &= \mathbf{a}(\theta)^* \left(\mathbf{I} - \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^* \right) \mathbf{a}(\theta) \\
&= \mathbf{a}(\theta)^* \left(\sum_{k=1}^{M} \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* - \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^* \right) \mathbf{a}(\theta) \\
&= \mathbf{a}(\theta)^* \left(\sum_{k=K+1}^{M} \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* + \sum_{k=1}^{K} \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* - \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^* \right) \mathbf{a}(\theta) \\
&\overset{N \text{ large}}{\simeq} \mathbf{a}(\theta)^* \left(\sum_{k=1}^{K} \left(1 - \frac{1}{h(\hat{\lambda}_k)} \right) \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* + \sum_{k=K+1}^{M} \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* \right) \mathbf{a}(\theta)
\end{aligned}
$$

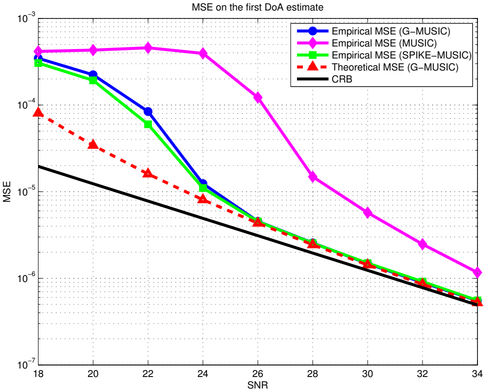
uniformly on $\theta \in [0, 2\pi]$.

Remarks.

- CLT on the DOA estimates show that $\mathbb{E}|\hat{\theta}_i - \theta_i|^2 = \mathcal{O}(\frac{1}{N^3})$ (connected to the choice of $\mathbf{a}(\theta)$).
- If the angles do not scale with M, N, vectors $\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)$ become orthogonal when M increases. Any reasonable approach provides consistent estimates, in particular the traditional subspace approach.
- In this context, the improved and traditional subspace approaches share the same asymptotic variance
- If $\theta_i - \theta_j = \mathcal{O}(\frac{1}{N})$, the improved estimators outperform the traditional ones.
- It is still possible to consistently estimate $\mathbf{a}(\theta)^* \mathbf{\Pi}^{\perp} \mathbf{a}(\theta)$ uniformly w.r.t. θ when $K \to +\infty$ at the same rate than M and N (Vallet-L-Mestre-2012)

Illustration

$K = 2, M = 40, N = 80, \theta_2 - \theta_1 = \frac{\pi}{N}$.



Mean square error of $\widehat{\theta}_1$ w.r.t. the signal to noise ratio.

Supervised detection of a known signal (Hiltunen-L, 2015,2017).

Observation		Canal	Signaux source		Bruit
$\begin{bmatrix} \mathbf{y}_1 \cdots \mathbf{y}_N \end{bmatrix}$	$=$	$\begin{bmatrix} \mathbf{a}_1 \cdots \mathbf{a}_L \end{bmatrix}$	$\begin{bmatrix} \mathbf{s}^1 \\ \cdots \\ \mathbf{s}^L \end{bmatrix}$	$+$	$\begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_N \end{bmatrix}$
$\mathbf{Y}$	$=$	$\mathbf{A}$	$\mathbf{S}$	$+$	$\mathbf{V}$
$M \times N$		$M \times L$	$L \times N$		$M \times N$

- $M + L < N$, $\mathbf{A}$ unknown deterministic matrix, $\mathbf{S}$ known matrix.
- $\mathbf{v}_1, \ldots, \mathbf{v}_N$ i.i.d. Gaussian random vectors $\mathcal{N}(0, \mathbf{R})$, $\mathbf{R}$ unknown.
- L may be of the same order of magnitude than M and N

- Hypothesis H_0, $\mathbf{Y} = \mathbf{V}$
- Hypothesis H_1, $\mathbf{Y} = \mathbf{AS} + \mathbf{V}$

Generalized Likelihood Ratio Test

Compare $\eta = \frac{1}{L} \log \det(\mathbf{I} + \mathbf{G}) = \frac{1}{L} \sum_{l=1}^{L} \log(1 + \lambda_l)$ to a threshold, where $\mathbf{G}$ is a random $L \times L$ matrix:

Under H_0,

$$\mathbf{G} = \left(\mathbf{V}_2/\sqrt{N}\right)^* \left(\mathbf{V}_1\mathbf{V}_1^*/N\right)^{-1} \left(\mathbf{V}_2/\sqrt{N}\right)$$

Under H_1,

$$\mathbf{G} = \left(\mathbf{A} + \mathbf{V}_2/\sqrt{N}\right)^* \left(\mathbf{V}_1\mathbf{V}_1^*/N\right)^{-1} \left(\mathbf{A} + \mathbf{V}_2/\sqrt{N}\right)$$

where $\mathbf{V}_1$ and $\mathbf{V}_2$ are 2 mutually independent $M \times (N - L)$ and $M \times L$ random matrices with $\mathcal{N}(0, 1)$ independent entries.

Find asymptotic approximations of the distribution of η_N under each hypothesis in order to evaluate the performance of the test.

The classical regime L, M fixed, $N \to +\infty$

- $\mathbf{V}_1$ is $M \times (N-L)$, $\frac{\mathbf{V}_1\mathbf{V}_1^*}{N} \to \mathbf{I}_M$
- $\mathbf{V}_2$ is $M \times L$, $\frac{\mathbf{V}_2}{\sqrt{N}} = \mathcal{O}_P(\frac{1}{\sqrt{N}})$

$$\left(\mathbf{V}_2/\sqrt{N}\right)^* (\mathbf{V}_1\mathbf{V}_1^*/N)^{-1} \left(\mathbf{V}_2/\sqrt{N}\right) \simeq \frac{\mathbf{V}_2\mathbf{V}_2^*}{N}$$

Under H_0, $N\eta_N$ converges towards a χ^2 distribution

$$\eta \simeq \frac{1}{L}\log\det(\mathbf{I} + \frac{\mathbf{V}_2\mathbf{V}_2^*}{N}) \simeq \frac{1}{L}\mathrm{Tr}\left(\frac{\mathbf{V}_2\mathbf{V}_2^*}{N}\right) = \frac{1}{NL}\chi^2(ML)$$

Under H_1, η_N converges towards a non zero mean Gaussian distribution

$$\eta \simeq \frac{1}{L}\log\det(\mathbf{I} + \mathbf{A}\mathbf{A}^*) + \frac{1}{\sqrt{N}}\mathcal{N}(0, \kappa)$$

L fixed, $M, N \to +\infty$, $c_N = \frac{M}{N} \to c_*$, $c_* < 1$

- $\mathbf{G}$ is a finite size matrix matrix under each hypothesis
- $\mathbf{V}_1$ is $M \times (N-L)$, $\frac{\mathbf{V}_1\mathbf{V}_1^*}{N}$ does not converge towards $\mathbf{I}_M$
- $\mathbf{V}_2$ is $M \times L$, $\frac{\mathbf{V}_2}{\sqrt{N}}$ is not a $\mathcal{O}_P(\frac{1}{\sqrt{N}})$

$$\left(\mathbf{V}_2/\sqrt{N}\right)^* \left(\mathbf{V}_1\mathbf{V}_1^*/N\right)^{-1} \left(\mathbf{V}_2/\sqrt{N}\right) \simeq \frac{1}{N}\mathrm{Tr}\left(\mathbf{V}_1\mathbf{V}_1^*/N\right)^{-1} \mathbf{I}_L \simeq \frac{c_N}{1-c_N}\mathbf{I}$$

Under H_0

$\eta_N \simeq \log(\frac{1}{1-c_N}) + \frac{1}{\sqrt{N}}\mathcal{N}(0, \frac{c_N}{1-c_N})$

Under H_1

$\eta_N \simeq \log(\frac{1}{1-c_N}) + \frac{1}{L}\log\det(\mathbf{I} + \mathbf{A}\mathbf{A}^*) + \frac{1}{\sqrt{N}}\mathcal{N}(0, \frac{c_N}{1-c_N} + \kappa_N)$

$L, M, N \to +\infty, d_N = \frac{L}{N} \to d_*, c_N = \frac{M}{N} \to c_*, c_* + d_* < 1$

The size of $\mathbf{G}$ converges towards $+\infty$

$\| \left(\mathbf{V}_2/\sqrt{N}\right)^* \left(\mathbf{V}_1\mathbf{V}_1^*/N\right)^{-1} \left(\mathbf{V}_2/\sqrt{N}\right) - \frac{1}{N}\mathrm{Tr}\left(\mathbf{V}_1\mathbf{V}_1^*/N\right)^{-1} \mathbf{I}_L \|$
does not converge towards 0.

Under H_0, using Zheng (2012), see also Bai-Silverstein's book

$$\eta_N \simeq \alpha_N + \frac{1}{L}\mathcal{N}\left(0, \frac{\log(1-c_N)(1-d_N)}{\log(1-c_N-d_N)}\right)$$

Under H_1

$$\eta_N \simeq (\alpha_N + \beta_N) + \frac{1}{L}\mathcal{N}\left(0, \frac{\log(1-c_N)(1-d_N)}{\log(1-c_N-d_N)} + \omega_N\right)$$

Illustrations I.

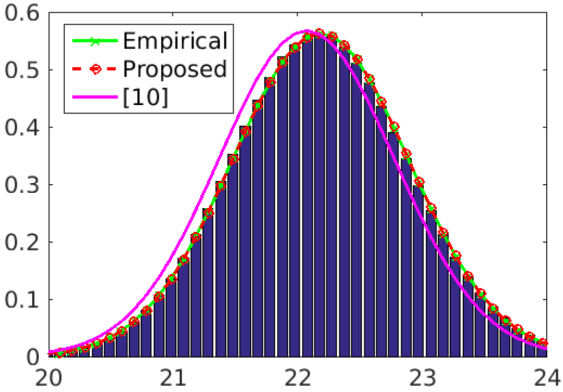


Figure: Gaussian approximation of η_N under H_1, $L = 20, M = 40, N = 80$.

Illustrations II.

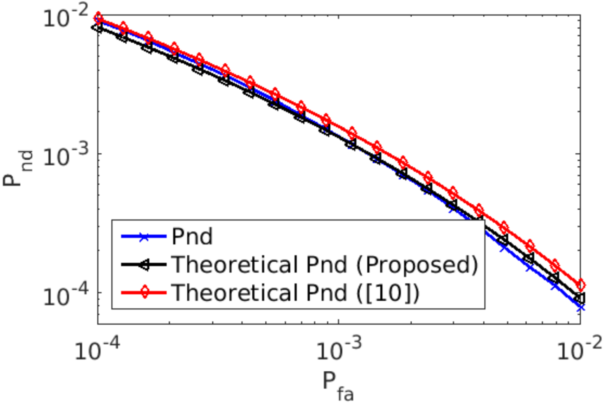


Figure: Experimental and theoretical ROC curves, $L = 20, M = 40, N = 80$.

Other kind of useful random matrix problems I.

In order to study estimation of noisy state space models:

$(v_n)_{n=1,\dots,N}$ i.i.d. $\mathcal{N}(0, \mathbf{R})$ vectors, L fixed integer.

Study the singular values of $\frac{\mathbf{V}_f \mathbf{V}_p^*}{N}$ where:

$$\mathbf{V}_p = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \dots & \mathbf{v}_N \\ \mathbf{v}_2 & \mathbf{v}_3 & \dots & \dots & \mathbf{v}_{N+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{v}_L & \mathbf{v}_{L+1} & \dots & \dots & \mathbf{v}_{N+L-1} \end{pmatrix}$$

$$\mathbf{V}_f = \begin{pmatrix} \mathbf{v}_{L+1} & \mathbf{v}_{L+2} & \dots & \dots & \mathbf{v}_{N+L} \\ \mathbf{v}_{L+2} & \mathbf{v}_{L+3} & \dots & \dots & \mathbf{v}_{N+L+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{v}_{2L} & \mathbf{v}_{2L+1} & \dots & \dots & \mathbf{v}_{N+2L-1} \end{pmatrix}$$

Same question, but we add to $\mathbf{V}_p$ and $\mathbf{V}_f$ low rank matrices having the same structure: behaviour of the largest singular values, and of the corresponding singular vectors.

Some references

- Z Bai, D Jiang, JF Yao, S Zheng, "Corrections to LRT on large-dimensional covariance matrix by RMT", The Annals of Statistics 37 (6B), 3822-3840
- F. Benaych-Georges and R. R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices", J. Multi. Var. Anal., vol. 111 (2012), pp. 120-135.
- P. Bianchi, M. Debbah, M. Maida, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory", *IEEE Trans. IT*, 57 (4), 2011.
- V.L. Girko, "An introduction to statistical analysis of random arrays", VSP, The Netherlands, 1998.
- S. Hiltunen, P. Loubaton, P. Chevalier, "Large system analysis of a GLRT for detection with large sensor arrays in temporally white noise", IEEE Trans. on Signal Processing, vol. 63, no. 20, pp. 5409 - 5423, October 15 2015

- S. Hiltunen, P. Loubaton, ”Asymptotic analysis of a GLR test for detection with large sensor arrays: new results”, Proc. ICASSP 2017, pp. 4506-4510, New Orleans, April 2017.
- I. Johnstone, On the distribution of the largest eigenvalue in principal components analysis. Ann. of Statist., 29(2001), 295–327.
- Z Li, G Pan, J Yao, ”On singular value distribution of large-dimensional autocovariance matrices”, Journal of Multivariate Analysis 137, 119-140.
- X. Mestre, ”Improved Estimation of Eigenvalues and Eigenvectors of Covariance Matrices Using Their Sample Estimates”, *IEEE Trans. IT*, 54 (11), 2008.
- X. Mestre and M.A. Lagunas, ”Modified subspace algorithms for DoA estimation with large arrays”, *IEEE Trans. SP*, 56 (2), 2008.

- R.R. Nadakuditi, A. Edelman, ”Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples”, *IEEE Transactions of Signal Processing*, vol. 56, no. 7, pp. 2625-2637, July 2008.
- R.R. Nadakuditi, J.W Silverstein, ”Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples”, *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 468 -480, June 2010.
- B. Nadler, ”Nonparametric Detection of Signals by Information Theoretic Criteria”, *IEEE Trans. SP*, 58 (5), 2010.
- P. Vallet, P. Loubaton, X. Mestre, ”Improved subspace estimation for multivariate observations of high dimension”, IEEE Transactions on Information Theory, vol. 58, no. 2, pp. 1043-1068, February 2012.

- P. Vallet, X. Mestre, P. Loubaton, ”Performance analysis of an improved MUSIC DoA estimator”, IEEE Trans. on Signal Processing, vol. 63, no. 23, pp. 6407-6422, December 1 2015
- S.Zheng, ”Central limit theorems for linear spectral statistics of large dimensional F-matrices”, *Ann. Inst. Henri Poincaré*, vol. 48, no. 2, pp. 444-476, 2012.

Estimation of certain parametric models, work in progress.

Structure of the observation.

- $\mathbf{y}_n = \mathbf{u}_n + \mathbf{v}_n$
- $(\mathbf{v}_n)_{n \in \mathbb{Z}}$ i.i.d. Gaussian $\mathcal{N}(0, \mathbf{R})$, $\mathbf{R}$ unknown
- $(\mathbf{u}_n)_{n \in \mathbb{Z}}$ M–dimensional stationary time series with rational spectrum, equivalent to:

$$
\begin{aligned}
\mathbf{x}_{n+1} &= \mathbf{A}\mathbf{x}_n + \mathbf{B}\nu_n \\
\mathbf{u}_n &= \mathbf{C}\mathbf{x}_n + \mathbf{D}\nu_n
\end{aligned}
$$

- $(\nu_n)_{n \in \mathbb{Z}}$ K-dimensional i.i.d. sequence $\mathcal{N}(0, \mathbf{I}_K)$, $K << M$.
- $\rho(\mathbf{A}) < 1$, $(\mathbf{x}_n)_{n \in \mathbb{Z}}$ is an order 1 P-dimensional autoregressive sequence called the state-space
- $\mathbf{C}, \mathbf{A}$ unique up to a similarity transform

Estimate $(\mathbf{C}, \mathbf{A})$ from $(\mathbf{y}_n)_{n=1,\ldots,N}$.

Traditional approach, M fixed, $N \to +\infty$, I

If $k \geq 1$, $\mathbb{E}(\mathbf{y}_{k+n}\mathbf{y}_n^*) = \mathbb{E}(\mathbf{u}_{k+n}\mathbf{u}_n^*) = \mathbf{C}\mathbf{A}^{k-1}\mathbf{G}$ for some matrix $\mathbf{G}$.

$\mathbf{C}_{f|p}^{(L)}$ $ML \times ML$ autocovariance matrix between the future and the past $\mathbf{C}_{f|p}^{(L)} = \mathbb{E}\left((\mathbf{y}_{n+1}^T, \ldots, \mathbf{y}_{n+L}^T)^T(\mathbf{y}_{n-(L-1)}^*, \ldots, \mathbf{y}_n^*)\right)$

For each $L \geq P = \dim(\mathbf{x}_n)$:

- $\mathrm{Rank}(\mathbf{C}_{f|p}^{(L)}) = P$
- Any minimal rank factorization of $\mathbf{C}_{f|p}^{(L)}$ can be written as $\mathbf{C}_{f|p}^{(L)} = \mathcal{O}\,\mathcal{C}$
- $\mathcal{O} = \begin{pmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{L-1} \end{pmatrix}, \mathcal{C} = (\mathbf{A}^{L-1}\mathbf{G}, \ldots, \mathbf{A}\mathbf{G}, \mathbf{G})$

From $\mathcal{O}$, easy to recover $\mathbf{C}$ and $\mathbf{A}$.

Traditional approach, M fixed, $N \to +\infty$, II

Estimate $\mathbf{C}_{f|p}^{(L)}$ from the observations

$$\mathbf{Y}_p = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \dots & \dots & \mathbf{y}_N \\ \mathbf{y}_2 & \mathbf{y}_3 & \dots & \dots & \mathbf{y}_{N+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \dots & \dots & \mathbf{y}_{N+L-1} \end{pmatrix}$$

$$\mathbf{Y}_f = \begin{pmatrix} \mathbf{y}_{L+1} & \mathbf{y}_{L+2} & \dots & \dots & \mathbf{y}_{N+L} \\ \mathbf{y}_{L+2} & \mathbf{y}_{L+3} & \dots & \dots & \mathbf{y}_{N+L+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{y}_{2L} & \mathbf{y}_{2L+1} & \dots & \dots & \mathbf{y}_{N+2L-1} \end{pmatrix}$$

Estimate $\mathbf{C}_{f|p}^{(L)}$ by $\hat{\mathbf{C}}_{f|p}^{(L)} = \frac{\mathbf{Y}_f \mathbf{Y}_p^*}{N}$ and use the truncated SVD of $\hat{\mathbf{C}}_{f|p}^{(L)}$ to estimate $\mathbf{C}$ and $\mathbf{A}$.

Consistent estimators because $\|\hat{\mathbf{C}}_{f|p}^{(L)} - \mathbf{C}_{f|p}^{(L)}\| \to 0$ when M fixed and $N \to +\infty$.

Behaviour of the scheme when $M, N \to +\infty$, L fixed $\frac{M}{N} \to c_*$?

Case where $\mathbf{u}_n = 0$, i.e. $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ are i.i.d. $\mathcal{N}(0, \mathbf{R})$, $\mathbf{C}_{f|p}^{(L)} = 0$

- Analysis of the behaviour of the singular values of $\hat{\mathbf{C}}_{f|p}^{(L)}$, or equivalently of the eigenvalues of $\hat{\mathbf{C}}_{f|p}^{(L)}(\hat{\mathbf{C}}_{f|p}^{(L)})^*$
- Under some technical extra assumptions: the empirical eigenvalue distribution has a deterministic behaviour that can be characterized using the Stieltjes transform approach, all the eigenvalues are concentrated in the neighbourhood of a union of intervals
- Li-Pan-Yao 2015 when $L = 1$ and $\mathbf{R} = \mathbf{I}$, L-Pastur-Tieplova in preparation in the general case

Case $\mathbf{u} \neq 0$: If P and K do not scale with M, N, $\mathbf{U}_f$ and $\mathbf{U}_p$ are finite rank matrices. Use a perturbation approach to study the largest eigenvalues of $\hat{\mathbf{C}}_{f|p}^{(L)}(\hat{\mathbf{C}}_{f|p}^{(L)})^*$