Some Applications of Large Random Matrices to Statistical Signal Processing

Ph. Loubaton

LIGM, Université Paris-Est Marne la Vallée.

GDR ISIS, 16 Mai 2013

The model considered in the following

Observation: M–dimensional time series $\mathbf{y}_n$ observed from $n = 1, \ldots, N$.

- $\mathbf{y}_n = \mathbf{A}\mathbf{s}_n + \mathbf{v}_n$
- $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_K)$ deterministic unknown rank $K < M$ matrix
- $\mathbf{s}_n = (s_{1,n}, \ldots, s_{K,n})^T$, $((s_{k,n})_{n \in \mathbb{Z}})_{k=1,K}$ are $K < M$ non observable deterministic "source signals"
- $(\mathbf{v}_n)_{n \in \mathbb{Z}}$ additive complex white Gaussian noise such that $\mathbb{E}(\mathbf{v}_n \mathbf{v}_n^H) = \sigma^2 \mathbf{I}_M$

In matrix form

- $\mathbf{Y}_N = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ observation $M \times N$ matrix
- $\mathbf{S}_N = (\mathbf{s}_1, \ldots, \mathbf{s}_N)$ signal $K \times N$ matrix, $\mathrm{Rank}(\mathbf{S}_N) = K$.
- $\mathbf{Y}_N = \mathbf{A}\mathbf{S}_N + \mathbf{V}_N$ Information + Noise model with rank deficient Information component.

Class of problems to be addressed.

Covariance matrices of the model.

- $\mathbf{Y}_N = \mathbf{A}\mathbf{S}_N + \mathbf{V}_N$
- Empirical covariance matrix $\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N}$
- "True" covariance matrix $\mathbb{E}\left(\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N}\right) = \mathbf{A}\frac{\mathbf{S}_N \mathbf{S}_N^*}{N}\mathbf{A}^* + \sigma^2 \mathbf{I}_M$

Extract informations on $\frac{\mathbf{A}\mathbf{S}_N \mathbf{S}_N^* \mathbf{A}^*}{N}$ from $\mathbf{Y}_N$.

- Classical problems if $M << N$, M fixed and $N \rightarrow +\infty$ because when $N \rightarrow +\infty$,

$$\| \frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N} - \left(\mathbf{A}\frac{\mathbf{S}_N \mathbf{S}_N^*}{N}\mathbf{A}^* + \sigma^2 \mathbf{I}_M\right) \| \rightarrow 0$$

- We consider the case where M and N are of the same order of magnitude, $M \rightarrow +\infty, N \rightarrow +\infty$ in such a way that $c_N = \frac{M}{N} \rightarrow c_*$, $0 < c_* < +\infty$. We assume $c_* < 1$.

Examples I.

The asymptotic regime

- $M = M(N)$, $N \to +\infty$ in such a way that $c_N = \frac{M(N)}{N} \to c_*$, $0 < c_* < 1$
- Written as $N \to +\infty$

Detection of the signal component

- Presence / Absence of signal.
- Consistent estimation of the number of sources K

Subspace estimation

- $\mathbf{\Pi}_N$ orthogonal projection on the column space of $\mathbf{A}$, $\mathbf{\Pi}_N^\perp$ the orthogonal projection on $[\mathrm{sp}(\mathbf{A})]^\perp$
- Consistent estimation of $\mathbf{d}_1^* \mathbf{\Pi}_N^\perp \mathbf{d}_2$, $\mathbf{d}_1, \mathbf{d}_2$ deterministic vectors.

Examples II.

The asymptotic regime

- $M = M(N)$, $N \to +\infty$ in such a way that $c_N = \frac{M(N)}{N} \to c_*$, $0 < c_* < 1$
- Written as $N \to +\infty$

Inference on the eigenvalues/eigenvectors of $\frac{\mathbf{A}\mathbf{S}_N\mathbf{S}_N^*\mathbf{A}^*}{N}$.

- If K does not scale with N, estimate the eigenvalues and associated eigenvectors
- If K scales with N, estimate linear statistics of the non zero eigenvalues $\lambda_1 \geq \lambda_2 \ldots, \lambda_K$ of $\frac{\mathbf{A}\mathbf{S}_N\mathbf{S}_N^*\mathbf{A}^*}{N}$

$$\frac{1}{K} \sum_{i=1}^{K} \psi(\lambda_k)$$

Another (more) popular model not addressed in this talk.

Zero mean correlated model.

- $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N) = \mathbf{R}^{1/2}\mathbf{X}$
- $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ is a complex Gaussian i.i.d. random matrix, $\mathbf{R}$ covariance matrix of the $(\mathbf{y}_n)_{n=1,\ldots,N}$
- The above information plus noise model reduces to the zero mean correlated model when the source signals are mutually independent i.i.d. gaussian sequences, $\mathbf{R} = \mathbf{A}\mathbf{A}^* + \sigma^2 \mathbf{I}_M$

Extract informations on $\mathbf{R}$ from $\mathbf{Y}$.

Brief history of the field.

- First works on large random matrices in the 1950's (E. Wigner)
- A huge number of works devoted to theoretical physics (Brezin, Dyson, Mehta, Pastur and colleagues,...)
- Small community of researchers of probability theory until the 1990's (Bai, Silverstein, Girko, Pastur and colleagues,..)
- Great interest in the probability theory community since the 1990's (stimulated by the free probability theory)
- First papers using large random matrices in the context of digital communications en 1999 (Tse, Verdu-Shamai,...)
- First papers devoted to statistics of large random matrices around 2005 (El-Karoui, Yao, Bai, Silverstein,...)
- First papers devoted to applications to statistical signal processing in 2008 (Mestre-Lagunas, Nadakuditi-Edelman)

The case where $\mathbf{Y} = \mathbf{V}$.

$$\mathbf{V} = \begin{pmatrix} V_{11} & V_{12} & \dots & V_{1N} \\ V_{21} & V_{22} & \dots & V_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ V_{M1} & V_{M2} & \dots & V_{MN} \end{pmatrix}$$

$(V_{ij})_{1 \leq i \leq M, 1 \leq j \leq N}$ i.i.d. complex Gaussian random variables $\mathcal{CN}(0, \sigma^2)$.

$\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ columns of $\mathbf{V}$, $\mathbb{E}(\mathbf{v}_n \mathbf{v}_n^*) = \sigma^2 \mathbf{I}_M$

Empirical covariance matrix:

$$\frac{\mathbf{V}\mathbf{V}^*}{N} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{v}_n \mathbf{v}_n^*$$

Behaviour of the empirical distribution of the eigenvalues of $\frac{\mathbf{V}\mathbf{V}^*}{N}$ for large M and N.

- $\hat{\lambda}_{1,N} \geq \hat{\lambda}_{2,N} \geq \ldots \geq \hat{\lambda}_{M,N}$ eigenvalues of $\frac{\mathbf{V}\mathbf{V}^*}{N}$
- Empirical eigenvalue distribution: $\hat{\mu}_N = \frac{1}{M}\sum_{i=1}^{M} \delta(\lambda - \hat{\lambda}_{i,N})$

How behave the histograms of the eigenvalues $(\hat{\lambda}_{i,N})_{i=1,\ldots,M}$ of $\frac{\mathbf{V}\mathbf{V}^*}{N}$ when M and N increase.

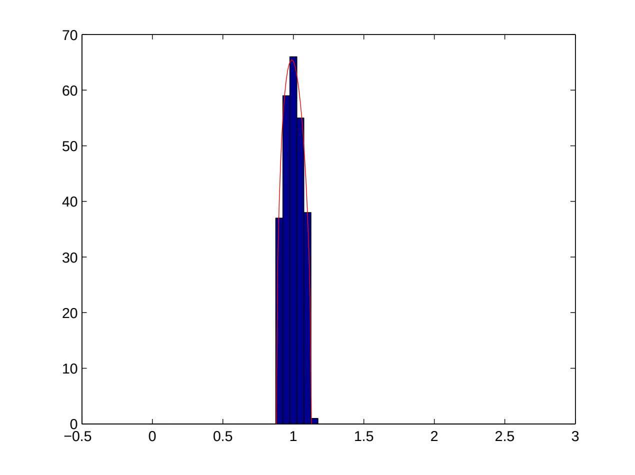
Well known case: M fixed, N increases i.e. $c_N = \frac{M}{N}$ small

$\frac{\mathbf{V}\mathbf{V}^*}{N} \simeq \mathbb{E}(\mathbf{v}_n\mathbf{v}_n^*) = \sigma^2 \mathbf{I}_M$ by the law of large numbers.

If $N >> M$, the eigenvalues of $\frac{\mathbf{V}\mathbf{V}^*}{N}$ are concentrated around σ^2.

Illustration.

$M = 256, c_N = \frac{M}{N} = \frac{1}{256}, \sigma^2 = 1$



If M et N are of the same order of magnitude.

- The entry (i,j) of $\frac{\mathbf{V}\mathbf{V}^*}{N} \simeq \sigma^2 \delta_{i-j}$ but
- $\|\frac{\mathbf{V}\mathbf{V}^*}{N} - \sigma^2 \mathbf{I}_M\|$ does not converge torwards 0.

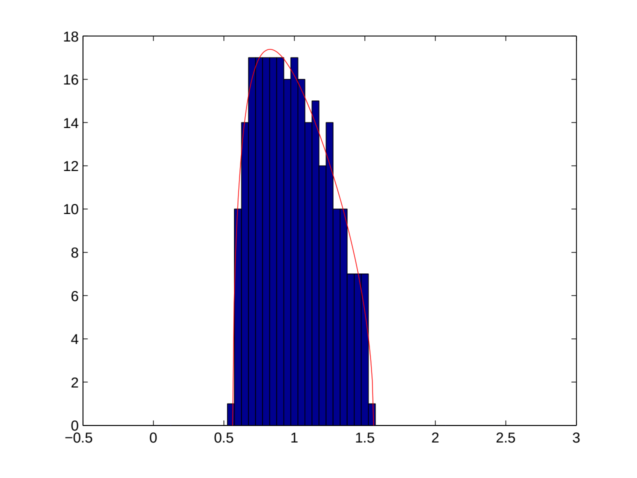
The histograms of the eigenvalues of $\frac{\mathbf{V}\mathbf{V}^*}{N}$ tend to concentrate around the probability density of the so-called Marcenko-Pastur distribution: if $c_N \leq 1$

$$
\begin{aligned}
p_{c_N}(\lambda) &= \frac{1}{2\pi c_N \lambda} \sqrt{[\sigma^2(1+\sqrt{c_N})^2 - \lambda][\lambda - \sigma^2(1-\sqrt{c_N})^2]} \\
&\quad \text{if } \lambda \in [\sigma^2(1-\sqrt{c_N})^2, \sigma^2(1+\sqrt{c_N})^2] \\
&= 0 \ \text{ otherwise}
\end{aligned}
$$

Result still true in the non Gaussian case

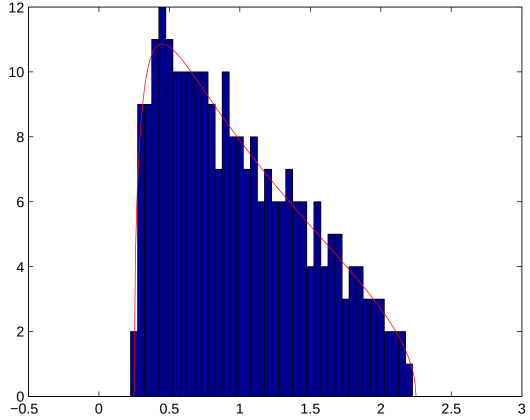
Illustrations I.

$M = 256, c_N = \frac{M}{N} = \frac{1}{16}, \sigma^2 = 1$



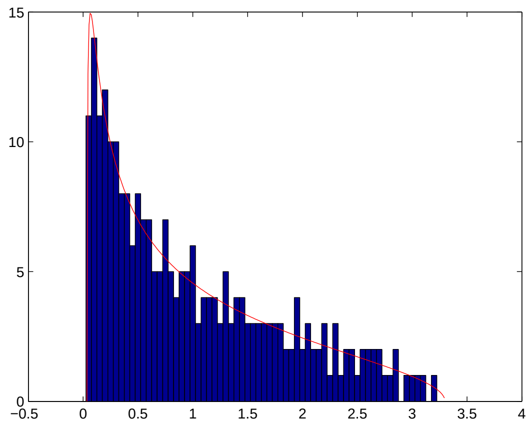
Illustrations II.

$M = 256, c_N = \frac{M}{N} = \frac{1}{4}, \sigma^2 = 1$



Illustrations III.

$M = 256, c_N = \frac{M}{N} = 2/3, \sigma^2 = 1$



More formally

$$\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_{k,N}) - \int \psi(\lambda) p_{c_N}(\lambda)\, d\lambda \to 0$$

How can it be proved ?

- Stieltjes transform of a measure μ: $z \longrightarrow \int_{\mathbb{R}} \frac{d\mu(\lambda)}{\lambda - z}$
- $m_N(z)$ Stieltjes transform of the MP distribution solution of the equation

$$m_N(z) = \frac{1}{-z\left[1+\sigma^2 c_N m_N(z)\right] + \sigma^2(1-c_N)}$$

- Show that the Stieltjes transform $\hat{m}_N(z)$ of the empirical eigenvalue distribution $\hat{\mu}_N$ converges for each z towards the Stieljes transform $m_N(z)$ of the MP distribution.

Another view of $\hat{m}_N(z)$

- $\hat{m}_N(z) = \int_{\mathbb{R}} \frac{d\hat{\mu}(\lambda)}{\lambda - z} = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{\hat{\lambda}_{i,N} - z}$
- Resolvent of $\frac{\mathbf{V}\mathbf{V}^*}{N}$:

$$z \to \mathbf{Q}_N(z) = \left(\frac{\mathbf{V}\mathbf{V}^*}{N} - z\mathbf{I} \right)^{-1}$$

- $\hat{m}_N(z) = \frac{1}{M} \mathrm{Tr} \mathbf{Q}_N(z)$

More powerfull result concerning $\mathbf{Q}_N(z)$

- $(\mathbf{Q}_N(z))_{i,j} - m_N(z)\delta(i-j) \to 0$, $\mathbf{Q}_N(z) \simeq m_N(z)\mathbf{I}_M$
- Illustration $z = 0$ $(c_N < 1)$,

$$\left(\frac{\mathbf{V}\mathbf{V}^*}{N} \right)^{-1}_{i,j} - \frac{\delta(i-j)}{\sigma^2(1 - c_N)} \to 0$$

Finer convergence results.

Convergence of the extreme eigenvalues

$$\hat{\lambda}_{1,N} - \sigma^2(1+\sqrt{c_N})^2 \xrightarrow[N,M\to\infty]{a.s.} 0$$
$$\hat{\lambda}_{M,N} - \sigma^2(1-\sqrt{c_N})^2 \xrightarrow[N,M\to\infty]{a.s.} 0$$

Implies the following almost sure location property of the $(\hat{\lambda}_{i,N})_{i=1,\ldots,M}$.

- For each $\epsilon > 0$, almost surely, all the eigenvalues belong to $[\sigma^2(1-\sqrt{c_N})^2 - \epsilon, \sigma^2(1+\sqrt{c_N})^2 + \epsilon]$ for N large enough.
- Important property valid in the context of other models based on i.i.d. complex Gaussian matrices (Bai-Silverstein 1999 for the zero mean correlated case, Loubaton-Vallet EJP 2011 for the Information plus Noise model, Male PTRF 2012).

Fluctuations of the extreme eigenvalues.

A Central Limit Theorem holds for the largest eigenvalue $\hat{\lambda}_{1,N}$. When correctly centered and rescaled, $\hat{\lambda}_{1,N}$ converges to a **Tracy-Widom** distribution:

$$\frac{N^{2/3}}{\sigma^2} \times \frac{\hat{\lambda}_{1,N} - \sigma^2(1+\sqrt{c_N})^2}{(1+\sqrt{c_N})\left(\frac{1}{\sqrt{c_N}}+1\right)^{1/3}} \xrightarrow[N,M\to\infty]{\mathcal{L}} \mu_{TW} \ .$$

The function μ_{TW} stands for **Tracy-Widom** distribution.

A similar result holds for $\hat{\lambda}_{M,N}$, the smallest eigenvalue.

Fluctuations of the linear statistics of the $(\hat{\lambda}_{i,N})_{i=1,\ldots,M}$.

- $\mathbb{E}\left(\frac{1}{M}\sum_{i=1}^{M}\psi(\hat{\lambda}_{i,N})\right) - \int \psi(\lambda) p_{c_N}(\lambda)\, d\lambda \;= \mathcal{O}(\frac{1}{N^2})$
- $N\left[\left(\frac{1}{M}\sum_{i=1}^{M}\psi(\hat{\lambda}_{i,N})\right) - \int \psi(\lambda) p_{c_N}(\lambda)\, d\lambda\right] \rightarrow \mathcal{N}(0,\Delta)$

The $(\hat{\lambda}_{i,N})_{i=1,\ldots,M}$ do not behave at all as realizations of independent random variables.

3. The case K does not scale with N.
   - Behaviour of the largest eigenvalues and related eigenvectors.
   - Applications.

The model.

We recall that:

$$\begin{array}{ccccc} \text{Rcv signal} & & \text{Channel} & \text{Src signal} & & \text{Noise} \\ \begin{bmatrix} \mathbf{y}_1 \cdots \mathbf{y}_N \end{bmatrix} & = & \begin{bmatrix} \mathbf{a}_1 \cdots \mathbf{a}_K \end{bmatrix} & \begin{bmatrix} \mathbf{s}^1 \\ \cdots \\ \mathbf{s}^K \end{bmatrix} & + & \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_N \end{bmatrix} \\ \mathbf{Y}_N & = & \mathbf{A}_N & \mathbf{S}_N & + & \mathbf{V}_N \\ M \times N & & M \times K & K \times N & & M \times N \end{array}$$

$\mathbf{Y}_N$ = Matrix with Gaussian iid elements + fixed rank perturbation.

Asymptotic regime: $N \to \infty$, $M/N \to c_*$, and K is fixed.

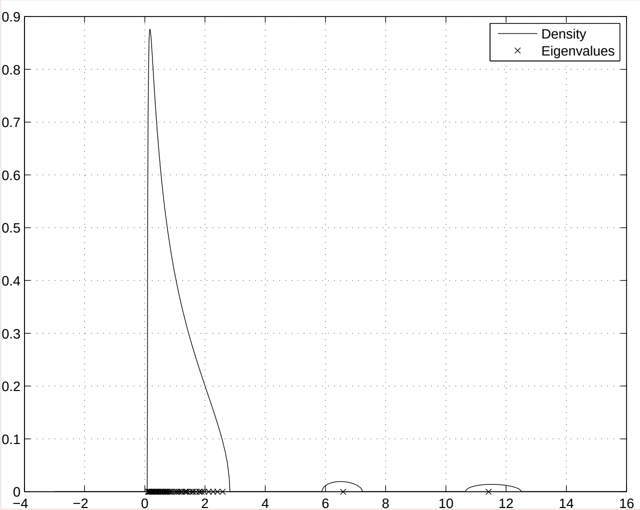
Results to be used when **number of sources** K **is** $\ll M$.

Impact of the fixed rank term on the eigenvalues and eigenvectors of $\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N}$

- $\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N}$ and $\frac{\mathbf{V}_N \mathbf{V}_N^*}{N}$ have the same (Marčenko Pastur) asymptotic eigenvalue distribution.
- $\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N}$ might have at most K **isolated eigenvalues** outside the support of the MP distribution.

Illustration

$c = \frac{M}{N} = 0.5, N = 100,\ K = 2,\ \sigma^2 = 1$



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

Main result on the eigenvalues

Theorem 1: Benaych-Georges and Nadakuditi, 2011

- Assume that $\lambda_{k,N} \to \rho_k$ for $k = 1, \ldots, K$.
- Let $i \leq K$ be the maximum index for which $\rho_i > \sigma^2 \sqrt{c_*}$ ($\lambda_{k,N} > \sigma^2 \sqrt{c_N}$ for $k \leq i$ and N large enough). Then for $k = 1, \ldots, i$,

$$\hat{\lambda}_{k,N} - \frac{\left(\sigma^2 c_N + \lambda_{k,N}\right)\left(\lambda_{k,N} + \sigma^2\right)}{\lambda_{k,N}} \xrightarrow[N\to\infty]{\text{a.s.}} 0$$

$\gamma_{k,N} = \frac{\left(\sigma^2 c_N + \lambda_{k,N}\right)\left(\lambda_{k,N} + \sigma^2\right)}{\lambda_{k,N}} > \sigma^2(1 + \sqrt{c_N})^2$ while

$$\hat{\lambda}_{i+1,N} - \sigma^2(1 + \sqrt{c_N})^2 \xrightarrow[N\to\infty]{\text{a.s.}} 0$$

- Finally, $\hat{\lambda}_{k,N} - \gamma_{k,N} = \mathcal{O}_P(1/\sqrt{N})$ for $k \leq i$.

Comments on Theorem I.

The almost sure location of the eigenvalues of $\frac{\mathbf{V}\mathbf{V}^*}{N}$ around the support of the MP distribution plays a fundamental role.

It is possible to estimate consistently the $(\lambda_{k,N})_{k=1,\ldots,i}$ from the $(\hat{\lambda}_{k,N})_{k=1,\ldots,i}$

- $\lambda \rightarrow \frac{(\sigma^2 c_N + \lambda)(\lambda + \sigma^2)}{\lambda}$ is invertible for $\lambda > \sigma^2 \sqrt{c_N}$ and its inverse is the Stieljes transform $m_{MP,N}$ of the MP distribution.
- Therefore, $\lambda_{k,N} = m_{MP,N}(\gamma_{k,N})$
- As $\hat{\lambda}_{k,N} - \gamma_{k,N} \rightarrow 0$, we have

$$
\begin{aligned}
\lambda_{k,N} - m_{MP,N}(\hat{\lambda}_{k,N}) &\rightarrow 0 \\
\lambda_{k,N} - m_{MP,N}(\hat{\lambda}_{k,N}) &= \mathcal{O}_P(1/\sqrt{N})
\end{aligned}
$$

Main result on the eigenvectors

Theorem 2: Benaych-Georges and Nadakuditi, 2011

- Assume the setting of Theorem 1. Assume in addition that $\rho_1 > \rho_2 > \cdots > \rho_i \ (> \sigma^2 \sqrt{c_*})$.
- For $k = 1, \ldots, i$, let

$$\mathbf{\Pi}_{k,N} = \mathbf{u}_{k,N}\mathbf{u}_{k,N}^* \quad \text{and} \quad \widehat{\mathbf{\Pi}}_{k,N} = \hat{\mathbf{u}}_{k,N}\hat{\mathbf{u}}_{k,N}^*.$$

  Then for any sequence $\mathbf{b}_N$ of deterministic $M \times 1$ vectors such that $\sup_N \|\mathbf{b}_N\| < \infty$,

$$\mathbf{b}_N^* \widehat{\mathbf{\Pi}}_{k,N} \mathbf{b}_N - h(\gamma_{k,N}) \mathbf{b}_N^* \mathbf{\Pi}_{k,N} \mathbf{b}_N \xrightarrow[N \to \infty]{\text{a.s.}} 0$$

  where $h(x)$ is a function depending on the Stieljes transform of the MP distribution, and $0 < h(\gamma_{k,N}) < 1$.
- Finally, $\mathbf{b}_N^* \widehat{\mathbf{\Pi}}_{k,N} \mathbf{b}_N - h(\gamma_{k,N}) \mathbf{b}_N^* \mathbf{\Pi}_{k,N} \mathbf{b}_N = \mathcal{O}_P(1/\sqrt{N})$

Comments on Theorem II.

It is possible to estimate consistently $\mathbf{b}_N^* \mathbf{\Pi}_{k,N} \mathbf{b}_N$ for $k = 1, \ldots, i$.

- $\mathbf{b}_N^* \mathbf{\Pi}_{k,N} \mathbf{b}_N - \frac{1}{h(\gamma_{k,N})} \mathbf{b}_N^* \widehat{\mathbf{\Pi}}_{k,N} \mathbf{b}_N \rightarrow 0$
- As $\hat{\lambda}_{k,N} - \gamma_{k,N} \rightarrow 0$, we have

$$\begin{aligned} \mathbf{b}_N^* \mathbf{\Pi}_{k,N} \mathbf{b}_N - \frac{1}{h(\hat{\lambda}_{k,N})} \mathbf{b}_N^* \widehat{\mathbf{\Pi}}_{k,N} \mathbf{b}_N &\rightarrow 0 \\ \mathbf{b}_N^* \mathbf{\Pi}_{k,N} \mathbf{b}_N - \frac{1}{h(\hat{\lambda}_{k,N})} \mathbf{b}_N^* \widehat{\mathbf{\Pi}}_{k,N} \mathbf{b}_N &= \mathcal{O}_P(1/\sqrt{N}) \end{aligned}$$

Reformulation and comments on Theorem II.

$\hat{\mathbf{u}}_{k,N}$ is not a good estimate of $\mathbf{u}_{k,N}$.

- If $\hat{\mathbf{u}}_{k,N}^* \mathbf{u}_{k,N} > 0$, equivalent to

$$\mathbf{b}^* \left(\hat{\mathbf{u}}_{k,N} - \sqrt{h(\gamma_{k,N})}\, \mathbf{u}_{k,N} \right) \rightarrow 0$$

  for each $\mathbf{b}$.
- $\mathbf{b} = \mathbf{u}_{k,N}$ yields to $\mathbf{u}_{k,N}^* \hat{\mathbf{u}}_{k,N} - \sqrt{h(\gamma_{k,N})} \rightarrow 0$
- $0 < h(\gamma_{k,N}) < 1$ can be written as

$$h(\gamma_{k,N}) = \frac{1 - \left(\sigma^2 \sqrt{c_N} / \lambda_{k,N}\right)^2}{1 + \sigma^2 c_N / \lambda_{k,N}}$$

- $(\mathbf{u}_{k,N})_m - \frac{1}{\sqrt{h(\hat{\lambda}_{k,N})}} (\hat{\mathbf{u}}_{k,N})_m = \mathcal{O}_P(1/\sqrt{N})$ for each m

3. The case K does not scale with N.
   - Behaviour of the largest eigenvalues and related eigenvectors.
   - Applications.

Testing $K = 0$ versus $K = 1$ (I).

Nadakuditi-Edelmann (IEEE-SP 2008), Nadler (IEEE-SP 2010), Bianchi-Debbah-Maeda-Najim (IEEE-IT 2011) when $(s_n)_{n=1,\ldots,N}$ is an i.i.d. complex Gaussian sequence.

Hypothesis test: $\begin{cases} \mathbf{H0} & : & \mathbf{Y}_N = \mathbf{V}_N & \text{(Noise)} \\ \mathbf{H1} & : & \mathbf{Y}_N = \mathbf{a}\,\mathbf{s}_N + \mathbf{V}_N & \text{(Info+Noise)} \end{cases}$

Remark on the signal to noise ratio.

- Hypothesis: $\lambda_N = \lambda_{max}\left((\mathbf{a}\mathbf{s}_N\mathbf{s}_N^*\mathbf{a}^*)/N\right) = \|\mathbf{a}\|^2 \, \frac{1}{N}\sum_{n=1}^N |s_n|^2 \to \rho$
- Signal to noise ratio: $\frac{\lambda_N}{M\sigma^2} \to 0$
- $\frac{\lambda_N}{\sigma^2}$ represents the signal to noise ratio at the matched filter output defined as $\mathbf{a}^*\mathbf{y}_n$

Testing $K = 0$ versus $K = 1$ (II).

Generalized Likelihood Ratio Test (GLRT)

$$T_N = \frac{\hat{\lambda}_{1,N}}{\frac{1}{M} \mathrm{tr}\left(\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N} \right)}$$

Analysis of T_N under each hypothesis.

- Possible to evaluate the distribution of T_N under each hypothesis.
- Asymptotic analysis of T_N is much more informative.

Testing $K = 0$ versus $K = 1$ (III).

- Under either **H0** or **H1**, $\frac{1}{M} \mathrm{tr}\left(\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N}\right) \xrightarrow[N\to\infty]{\text{a.s.}} \sigma^2$.
- Under **H1** (consequence of main result on eigenvalues):
  - If $\rho > \sigma^2 \sqrt{c_*}$ $\quad (\frac{\lambda_N}{\sigma^2} > \sqrt{c}_N)$, then

$$\hat{\lambda}_{1,N} \simeq \gamma_N = \frac{\left(\sigma^2 c_N + \lambda_N\right)\left(\lambda_N + \sigma^2\right)}{\lambda_N} > \sigma^2(1+\sqrt{c_N})^2,$$
$$\hat{\lambda}_{2,N} \simeq \sigma^2(1+\sqrt{c_N})^2.$$

  - If $\rho < \sigma^2 \sqrt{c_*}$ $\quad (\frac{\lambda_N}{\sigma^2} < \sqrt{c}_N)$, then

$$\hat{\lambda}_{1,N} \simeq \sigma^2(1+\sqrt{c_N})^2$$

Testing $K = 0$ versus $K = 1$ (IV).

We therefore have

- Under **H0**,

$$T_N \simeq (1 + \sqrt{c_N})^2.$$

- Under **H1**,
  - If $\rho > \sigma^2\sqrt{c_*}$ $\quad (\frac{\lambda_N}{\sigma^2} > \sqrt{c}_N)$, then

$$T_N \simeq \frac{\left(\sigma^2 c_N + \lambda_N\right)\left(\lambda_N + \sigma^2\right)}{\sigma^2 \, \lambda_N} > (1 + \sqrt{c_N})^2$$

  - If $\rho < \sigma^2\sqrt{c_*}$ $\quad (\frac{\lambda_N}{\sigma^2} < \sqrt{c}_N)$, then

$$T_N \simeq (1 + \sqrt{c_N})^2.$$

$\rho > \sigma^2\sqrt{c_*}$ provides the **limit of detectability** by the GLRT.

- False Alarm Probability can be evaluated with the help of the Tracy-Widom law.

Subspace estimation and applications to source localization.

- Mestre-Lagunas (IEEE-SP 2008) when the source signals are i.i.d. gaussian independent sequences (use of the zero-mean correlated model).
- In the context of Information plus Noise models, see Vallet-Loubaton-Mestre (IEEE-IT 2012) and Hachem-Loubaton-Mestre-Najim-Vallet (J. Multivariate Analysis 2013).

Subspace estimation

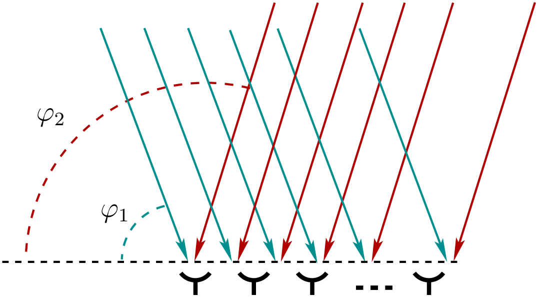
- $\mathbf{\Pi}_N$ orthogonal projection on the column space of $\mathbf{A}$, $\mathbf{\Pi}_N^{\perp}$ the orthogonal projection on $[\mathrm{sp}(\mathbf{A})]^{\perp}$
- Consistent estimation of $\mathbf{b}_N^* \mathbf{\Pi}_N^{\perp} \mathbf{b}_N$, $\mathbf{b}_N$ deterministic vector.

Source localization.

Problem

K radio sources send their signals to a uniform array of M antennas during N signal snapshots.

Estimate arrival angles $\varphi_1, \ldots, \varphi_K$



Example with two sources

Source localization with a subspace method (MUSIC)

Model.

- $\mathbf{Y}_N = \mathbf{A}_N \mathbf{S}_N + \mathbf{V}_N$
- $\mathbf{A}_N = \left[\mathbf{a}_N(\varphi_1) \quad \cdots \quad \mathbf{a}_N(\varphi_K) \right]$ with $\mathbf{a}_N(\varphi) = \frac{1}{\sqrt{M}} \begin{bmatrix} 1 \\ e^{\imath\varphi} \\ \vdots \\ e^{\imath(M-1)\varphi} \end{bmatrix}$

MUSIC algorithm principle

- $\mathbf{a}_N(\varphi)^* \mathbf{\Pi}_N^{\perp} \mathbf{a}_N(\varphi) = 0 \quad \Leftrightarrow \quad \varphi \in \{\varphi_1, \ldots, \varphi_K\}$
- Estimate $\mathbf{a}_N(\varphi)^* \mathbf{\Pi}_N^{\perp} \mathbf{a}_N(\varphi)$ for each φ, and evaluate the arguments of the local minima of the estimate w.r.t. φ.
- Traditional estimate : $\mathbf{a}_N(\varphi)^* \left(\sum_{k=K+1}^{M} \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \right) \mathbf{a}_N(\varphi)$.

Application of Theorem II.

Modified MUSIC estimator: application of Theorem 2

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

$$\mathbf{a}_N(\varphi)^* \mathbf{\Pi}_N \mathbf{a}_N(\varphi) - \sum_{k=1}^{K} \frac{\mathbf{a}_N(\varphi)^* \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \mathbf{a}_N(\varphi)}{h(\hat{\lambda}_{k,N})} \xrightarrow[N \to \infty]{\text{a.s.}} 0$$

uniformly on $\varphi \in [0, \pi]$.

Modification of the traditional estimator

$$\mathbf{a}(\varphi)^* \mathbf{\Pi}^{\perp} \mathbf{a}(\varphi) = \mathbf{a}(\varphi)^* \left(\sum_{k=1}^{M} \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* - \mathbf{\Pi} \right) \mathbf{a}(\varphi)$$

$$\overset{N \text{ large}}{\simeq} \mathbf{a}(\varphi)^* \left(\sum_{k=1}^{K} \left(1 - \frac{1}{h(\hat{\lambda}_k)} \right) \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* + \sum_{k=K+1}^{M} \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^* \right) \mathbf{a}(\varphi)$$

On the condition $\lim_{N\to+\infty} \lambda_{K,N} > \sigma^2\sqrt{c_*}$.

$$K = 2, \lim_{N\to+\infty} \frac{\mathbf{S}_N\mathbf{S}_N^*}{N} \to \mathbf{I}_2$$

If $M(\varphi_2 - \varphi_1) \to +\infty$.

- $\mathbf{a}_N(\varphi_1)^*\mathbf{a}_N(\varphi_2) \to 0$
- $\lambda_{i,N} \to 1$ for $i = 1, 2$.
- Reduces to the detectability condition $\sigma^2\sqrt{c_*} < 1$

If $M(\varphi_2 - \varphi_1) \to \alpha$

- $|\mathbf{a}_N(\varphi_1)^*\mathbf{a}_N(\varphi_2)| \to \frac{\sin \alpha/2}{\alpha/2}$
- $\lambda_{1,N} \to 1 + \frac{\sin(\alpha/2)}{\alpha/2}$
- $\lambda_{2,N} \to 1 - \frac{\sin(\alpha/2)}{\alpha/2}$
- $\sigma^2\sqrt{c_*} < 1 - \frac{\sin(\alpha/2)}{\alpha/2}$

Asymptotic behaviour of the estimates, $K = 2$.

If $M(\varphi_2 - \varphi_1) \to +\infty$

- Traditional estimates consistent, same performance that the improved ones.
- $M\left(\hat{\varphi}_k - \varphi_k\right) \to 0$
- $MN^{1/2}\left(\hat{\varphi}_k - \varphi_k\right) \to \mathcal{N}(0, \delta_N)$
- $\delta_N = 6\left[\frac{\frac{\lambda_N}{\sigma^2}+1}{\left(\frac{\lambda_N}{\sigma^2}\right)^2 - c_N}\right]$

If $M(\varphi_2 - \varphi_1) \to \alpha$

- $\hat{\varphi}_k^{(t)} - \varphi_k = \mathcal{O}_P(\frac{1}{M})$
- $\hat{\varphi}_k - \varphi_k = \mathcal{O}_P(\frac{1}{MN^{1/2}})$

Asymptotic behaviour of the estimates, $K = 2$.

Conclusion

- If the angles are far enough (w.r.t. $\frac{1}{M}$), the traditional estimate and the improved estimate have equivalent performance.
- On the contrary, the improved estimate outperforms the traditional estimate for close angles.

On the condition K does not scale with N.

Means in practice that $\frac{K}{M}$ should be small enough.

If this is not the case.

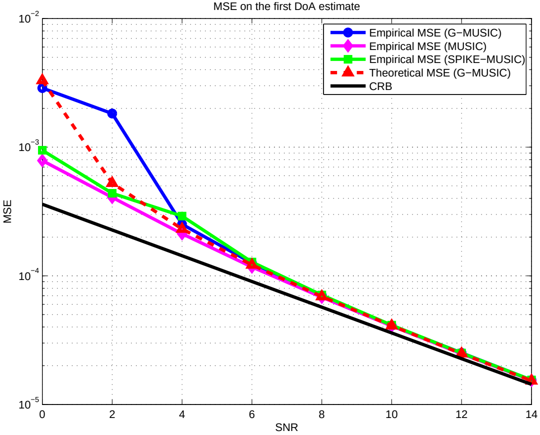
- Regime in which K may scale with (M, N)
- Possible to estimate $\mathbf{a}_N(\varphi)^* \mathbf{\Pi}_N^{\perp} \mathbf{a}_N(\varphi)$ consistently for each φ by

$$\mathbf{a}_N(\varphi)^* \left[\sum_{k=1}^{M} \hat{\alpha}_{k,N} \, \hat{\mathbf{u}}_{k,N} \hat{\mathbf{u}}_{k,N}^* \right] \mathbf{a}_N(\varphi)$$

  for some coefficients $(\hat{\alpha}_{k,N})_{k=1,\ldots,M}$ depending on the eigenvalues $(\hat{\lambda}_{k,N})_{k=1,\ldots,M}$
- Needs to study deeply the properties of the asymptotic eigenvalue distribution of matrix $\frac{\mathbf{Y}_N \mathbf{Y}_N^*}{N}$ in order to locate the eigenvalues.

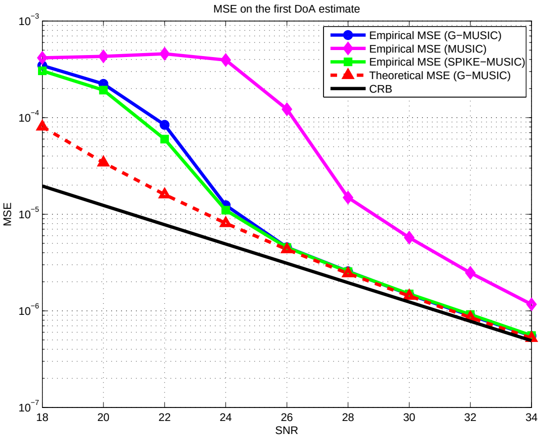
See Vallet-Loubaton-Mestre (IEEE-IT-2012).

$K = 2,\ M = 20, N = 40,\ \varphi_2 - \varphi_1 = \frac{\pi}{4}$.



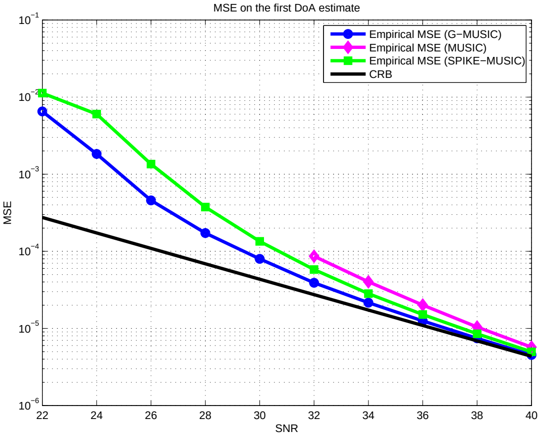
The minimum mean square error of the various estimates of φ_1 w.r.t. $10\log_{10}(\frac{1}{\sigma^2})$.

$K = 2,\ M = 40, N = 80,\ \varphi_2 - \varphi_1 = \frac{\pi}{2M}.$



The minimum mean square error of the various estimates of φ_1 w.r.t. $10\log_{10}(\frac{1}{\sigma^2})$.

$K = 5,\ M = 20, N = 40,\ \varphi_{k+1} - \varphi_k = \frac{2\pi}{35}.$



The minimum mean square error of the various estimates of φ_1 w.r.t. $10\log_{10}(\frac{1}{\sigma^2})$.

1. Problem statement

2. The case $K = 0$. The Marcenko-Pastur distribution

3. The case K does not scale with N.

4. Other problems.
   - Wideband models.

4. Other problems.
   - Wideband models.

Wideband single source model.

Narrowband single source model: $\mathbf{y}_n = \mathbf{a}s_n + \mathbf{v}_n$.

$\mathbf{y}_n = \sum_{p=0}^{P-1} \mathbf{a}_p s_{n-p} + \mathbf{v}_n$

- $(s_n)_{n=1,\ldots,N}$ non observable deterministic sequence
- $(\mathbf{a}_p)_{p=0,\ldots,P-1}$ unknown deterministic M-dimensional vectors

Usual to introduce $ML \times N$ matrix $\mathbf{Y}_N^{(L)}$ defined by

$$\mathbf{Y}_N^L = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \\ \mathbf{y}_2 & \mathbf{y}_3 & \cdots & \mathbf{y}_{N+1} \\ \mathbf{y}_3 & \cdots & \cdots & \mathbf{y}_{N+2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_L & \mathbf{y}_{L+1} & \cdots & \mathbf{y}_{N+L-1} \end{pmatrix}$$

Block-Hankel information plus noise model.

Expression of $\mathbf{Y}_N^L$

$\mathbf{Y}_N^{(L)} = \mathbf{A}_N^{(L)} \mathbf{S}_N^{(L)} + \mathbf{V}_N^{(L)}$

- $\mathbf{A}_N^{(L)}$ is a $ML \times (P+L-1)$ matrix, $\mathbf{S}_N^{(L)}$ is a $(P+L-1) \times N$ matrix
- The entries of $\mathbf{V}_N^{(L)}$ are no longer i.i.d.

Important questions.

- How behaves the empirical eigenvalue distribution of $\frac{\mathbf{V}_N^{(L)} \mathbf{V}_N^{(L)*}}{N}$
- What can be said on the greatest eigenvalues of $\frac{\mathbf{Y}_N^{(L)} \mathbf{Y}_N^{(L)*}}{N}$

Some preliminary results for $K = 0$.

$$\mathbf{V}_N^L = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_N \\ \mathbf{v}_2 & \mathbf{v}_3 & \ldots & \mathbf{v}_{N+1} \\ \mathbf{v}_3 & \ldots & \ldots & \mathbf{v}_{N+2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_L & \mathbf{v}_{L+1} & \ldots & \mathbf{v}_{N+L-1} \end{pmatrix}$$

ML and N converge to $+\infty$ at the same rate, $\frac{ML}{N} \to d_*$

- If $M \to +\infty$:
- The empirical eigenvalue distribution of $\frac{\mathbf{V}_N^{(L)}\mathbf{V}_N^{(L)*}}{N}$ behaves like the Marcenko-Pastur distribution with parameters $d_N = \frac{ML}{N}$. The number of "independent random entries" of $\mathbf{V}_N^{(L)}$ is $\mathcal{O}(MN) >> N$, nice averaging effects.
- If $\frac{L}{M} \to 0$, the eigenvalues do not escape from $(1 - \sqrt{d}_N)^2, (1 + \sqrt{d}_N)^2)$

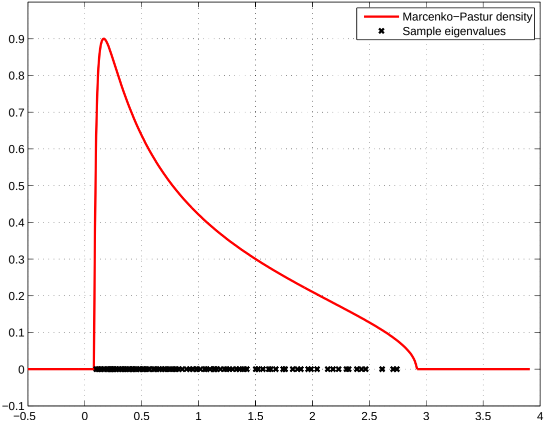
Some preliminary results for $K = 0$.

$$\mathbf{V}_N^L = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_N \\ \mathbf{v}_2 & \mathbf{v}_3 & \ldots & \mathbf{v}_{N+1} \\ \mathbf{v}_3 & \ldots & \ldots & \mathbf{v}_{N+2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_L & \mathbf{v}_{L+1} & \ldots & \mathbf{v}_{N+L-1} \end{pmatrix}$$

ML and N converge to $+\infty$ at the same rate, $\frac{ML}{N} \rightarrow d_*$

- If M if fixed. Much more complicated situation because the number of "independent random entries" of $\mathbf{V}_N^{(L)}$ is $\mathcal{O}(N)$, not enough to observe nice averaging effects.
- The empirical eigenvalue distribution has a non bounded limit distribution difficult to study

$M = 20, L = 5, N = 2ML, \sigma^2 = 1.$



$M = 20, L = 60, N = 2ML, \sigma^2 = 1.$