Almost sure location of the singular values of Gaussian large random matrices: the information plus noise model case

Philippe Loubaton, Pascal Vallet

Université de Paris-Est / Marne la Vallée, LIGM

EPFL/UMLV Workshop, 4/12/2009

Plan

1. Problem statement.
2. The asymptotic behaviour of the eigenvalue distribution of $\hat{R}_N$
3. Almost sure locations of the eigenvalues of $\hat{R}_N$.
4. Conclusion

Plan

The information plus noise model

Observation of N samples of a M-variate time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$, $N > M$.

- $\mathbf{y}_n = \mathbf{a}_n + \sigma \mathbf{w}_n$
- $(\mathbf{a}_n)_{n=1,\ldots,N}$ deterministic vectors, correspond to the useful signal
- $\sigma \mathbf{w}_n$ additive complex white Gaussian noise, $\mathbb{E}(\mathbf{w}_n \mathbf{w}_n^H) = \mathbf{I}_M$

Equivalent $M \times N$ matrix model

$$(\mathbf{y}_1, \ldots, \mathbf{y}_N) = \mathbf{Y} = \mathbf{A} + \sigma \mathbf{W}$$

$\mathbf{Y}$ non zero mean Gaussian random matrix with independent entries of variance σ^2

Typical applicative context : source localization

K source signals $(s_k)_{k=1,\dots,K}$ observed on a M sensors array, $K < M$

- $\mathbf{s}_n = (s_{1,n}, \dots, s_{K,n})^T$, $s_{k,n}$ k-th source signal at time n
- $\mathbf{a}_n = \mathbf{D}\mathbf{s}_n$
- $\mathbf{D}$ deterministic $M \times K$ directional vectors matrix, $K < M$

The associated Information plus Noise matrix model

- $\mathbf{Y} = \mathbf{DS} + \sigma\mathbf{W}$, $\mathbf{A} = \mathbf{DS}$
- $\mathrm{Rank}(\mathbf{A}) < M$

The empirical covariance matrix

The associated empirical covariance matrix.

$$\hat{\mathbf{R}}_N = \frac{\mathbf{Y}\mathbf{Y}^H}{N} = \frac{(\mathbf{A} + \sigma\mathbf{W})(\mathbf{A} + \sigma\mathbf{W})^H}{N}$$

Study of the location of the eigenvalues of matrix $\hat{\mathbf{R}}_N$

If $N \to \infty$, M fixed

- $\mathbf{R}_N = \frac{\mathbf{A}\mathbf{A}^H}{N}$
- $\hat{\mathbf{R}}_N - (\mathbf{R}_N + \sigma^2 \mathbf{I}_M) \to 0$
- Eigenvalues of $\hat{\mathbf{R}}_N \simeq$ eigenvalues of $\mathbf{R}_N + \sigma^2$
- Source localization context : σ^2 eigenvalue with multiplicity $M - K$, the K greatest eigenvalues $> \sigma^2$

If $N \to \infty$, M fixed

- $\mathbf{R}_N = \frac{\mathbf{A}\mathbf{A}^H}{N}$
- $\hat{\mathbf{R}}_N - (\mathbf{R}_N + \sigma^2 \mathbf{I}_M) \to 0$
- Eigenvalues of $\hat{\mathbf{R}}_N \simeq$ eigenvalues of $\mathbf{R}_N + \sigma^2$
- Source localization context : σ^2 eigenvalue with multiplicity $M - K$, the K greatest eigenvalues $> \sigma^2$

If M and N are of the same order of magnitude

- $M \to \infty, N \to \infty$
- $c_N = \frac{M}{N} \to$ non zero constant

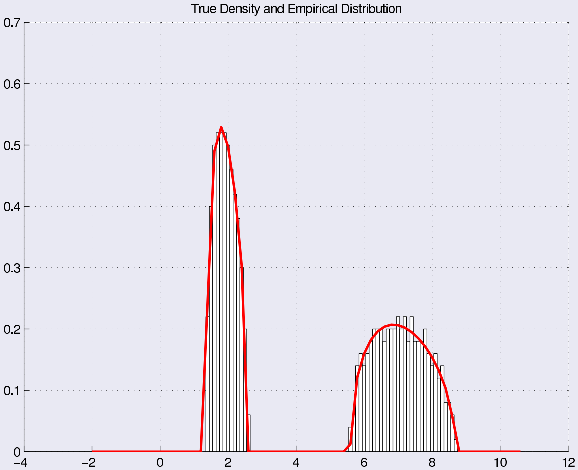
The histograms of the eigenvalues of $\hat{\mathbf{R}}_N$ have a deterministic behaviour which can be characterized : Dozier-Silverstein 2007.

Numerical illustration (I).

- $\sigma^2 = 2, M = 256$
- Eigenvalues of $\mathbf{R}_N$ 0 with multiplicity 128, 5 with multiplicity 128
- If $c = \frac{M}{N} \simeq 0$, eigenvalues of $\hat{\mathbf{R}}_N \simeq 2$ with multiplicity 128, 7 with multiplicity 128
- $c = \frac{M}{N}, c = 0.05, 0.2, 0.5$
- Representation of histograms of the eigenvalues of $\hat{\mathbf{R}}_N$

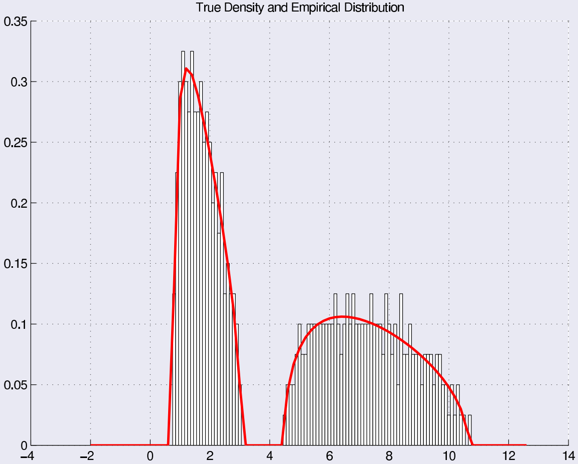
Numerical illustration (II).

$c = \frac{M}{N} = 0.05$



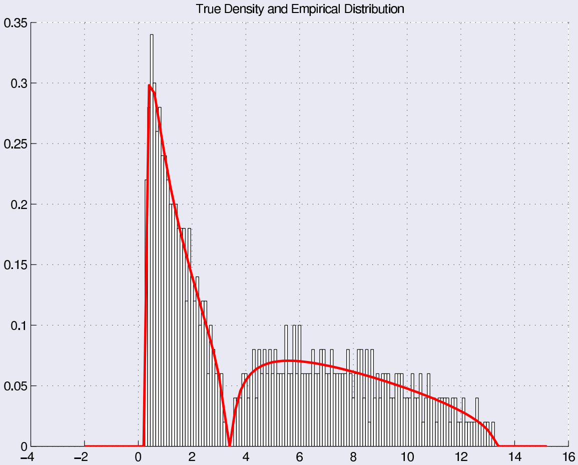
Illustrations numériques (III).

$c = \frac{M}{N} = 0.2$



Illustrations numériques (III).

$c = \frac{M}{N} = 0.5$



Plan

1. Problem statement.
2. The asymptotic behaviour of the eigenvalue distribution of $\hat{\mathbf{R}}_N$
3. Almost sure locations of the eigenvalues of $\hat{\mathbf{R}}_N$.
4. Conclusion

Mathematical formulation

The asymptotic regime

- $M \to \infty$, $N \to \infty$
- $c_N = \frac{M}{N} \to$ non zero constant

$(\hat{\lambda}_k)_{k=1,\ldots,M}$ eigenvalues of $\hat{\mathbf{R}}_N$, $(\lambda_k)_{k=1,\ldots,M}$ eigenvalues of $\mathbf{R}_N$, arranged in increasing order.

D.Z 2007 : It exists a deterministic probability measure μ_N carried by $\mathbb{R}^+$ such that

- $\frac{1}{M}\sum_{k=1}^{M} \delta(\lambda - \hat{\lambda}_k) - \mu_N \to 0$ weakly almost surely
- $\frac{1}{M}\sum_{k=1}^{M} \delta(\lambda - \hat{\lambda}_k)$: empirical eigenvalue distribution of $\hat{\mathbf{R}}_N$.

How to characterize μ_N

The Stieljès transform $m_N(z)$ of μ_N

- $m_N(z) = \int_{\mathbb{R}^+} \frac{\mu_N(d\lambda)}{\lambda - z}$ defined on $\mathbb{C} - \mathbb{R}^+$
- $\int_{\mathbb{R}^+} \phi(\lambda)\mu_N(d\lambda) = \frac{1}{\pi} \lim_{y \to 0^+} \mathrm{Im} \int_{\mathbb{R}^+} \phi(x) m_N(x + iy) dx$

Convergence of $\frac{1}{M}\sum_{k=1}^{M} \delta(\lambda - \hat{\lambda}_k)$ towards μ_N

Show that $\frac{1}{M}\sum_{k=1}^{M} \frac{1}{\hat{\lambda}_k - z} - m_N(z) \to 0$ a.s. for each $z \in \mathbb{C} - \mathbb{R}^+$.

How to characterize μ_N

The Stieljès transform $m_N(z)$ of μ_N

- $m_N(z) = \int_{\mathbb{R}^+} \frac{\mu_N(d\lambda)}{\lambda - z}$ defined on $\mathbb{C} - \mathbb{R}^+$
- $\int_{\mathbb{R}^+} \phi(\lambda)\mu_N(d\lambda) = \frac{1}{\pi} \lim_{y \to 0^+} \mathrm{Im} \int_{\mathbb{R}^+} \phi(x) m_N(x+iy) dx$

$m_N(z)$ is solution of the equation

$$\frac{m_N(z)}{1 + \sigma^2 c_N m_N(z)} = f_N(w_N(z))$$

- $w_N(z) = z(1 + \sigma^2 c_N m_N(z))^2 + \sigma^2(1 - c_N)(1 + \sigma^2 c_N m_N(z))$
- $f_N(w) = \frac{1}{M} \mathrm{Trace}(\mathbf{R}_N - w\mathbf{I}_M)^{-1} = \frac{1}{M} \sum_{k=1}^{M} \frac{1}{\lambda_k - z}$

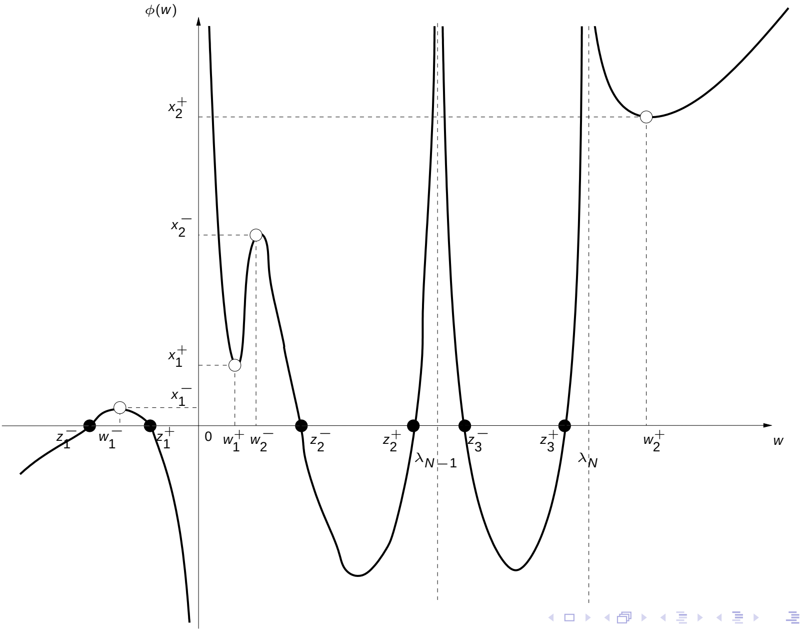
Properties of μ_N, $c_N = \frac{M}{N} < 1$.

- $\mu_N(d\lambda)$ absolutely continuous
- μ_N is compactly supported, $\mathcal{S}_N$ support of μ_N

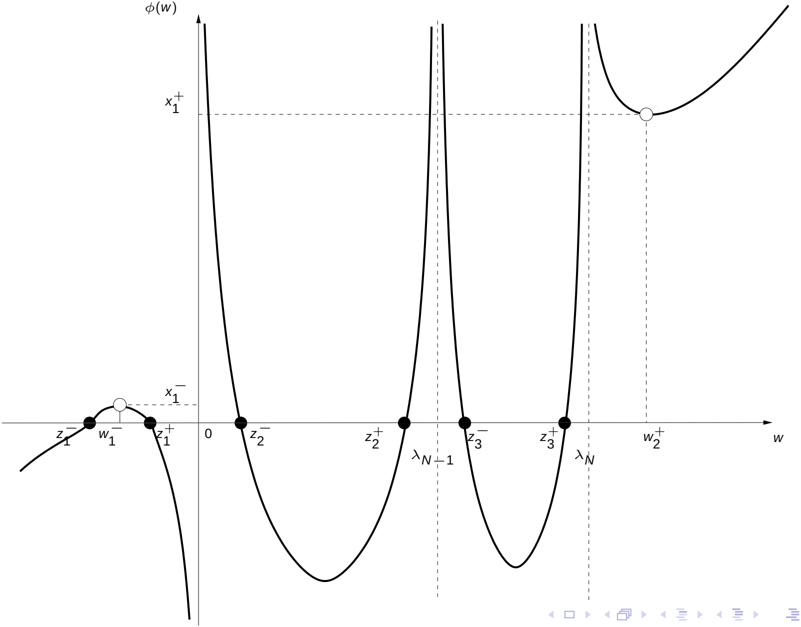
Properties of μ_N, $c_N = \frac{M}{N} < 1$.

Characterization of $\mathcal{S}_N$: reformulation of D.Z 2007 in Vallet-Loubaton-Mestre-2009

- Function $\phi_N(w)$ defined on $\mathbb{R}$ by $\phi_N(w) = w(1 - \sigma^2 c_N f_N(w))^2 + \sigma^2(1 - c_N)(1 - \sigma^2 c_N f_N(w))$
- ϕ_N has $2Q$ extrema whose preimages $w_{1,-}^{(N)} < w_{1,+}^{(N)} < \dots w_{Q,-}^{(N)} < w_{Q,+}^{(N)}$ satisfy $1 - \sigma^2 c_N f_N(w) > 0$. These extrema verify $x_{1,-}^{(N)} < x_{1,+}^{(N)} < \dots x_{Q,-}^{(N)} < x_{Q,+}^{(N)}$.
- $\mathcal{S}_N = [x_{1,-}^{(N)}, x_{1,+}^{(N)}] \cup \dots [x_{Q,-}^{(N)}, x_{Q,+}^{(N)}]$
- Each eigenvalue λ_l of $\mathbf{R}_N$ belongs to an interval $[w_{k,-}^{(N)}, w_{k,+}^{(N)}]$



$\phi(w)$
x_2^+
x_2^-
x_1^+
x_1^-
z_1^- w_1^- z_1^+ 0 w_1^+ w_2^- z_2^- z_2^+ z_3^- z_3^+ w_2^+ w
λ_{N-1}
λ_N



$\phi(w)$
x_1^+
x_1^-
z_1^-
w_1^-
z_1^+
0
z_2^-
z_2^+
λ_{N-1}
z_3^-
z_3^+
λ_N
w_2^+
w

Some definitions

- Each interval $[x_{k,-}^{(N)}, x_{k,+}^{(N)}]$ is called a cluster
- An eigenvalue $\lambda_l^{(N)}$ of $\mathbf{R}_N$ is said to be associated to cluster $[x_{k,-}^{(N)}, x_{k,+}^{(N)}]$ if $\lambda_l^{(N)} \in [w_{k,-}^{(N)}, w_{k,+}^{(N)}]$
- 2 eigenvalues of $\mathbf{R}_N$ are said to be separated if they are associated to different clusters

Plan

The new results I.

Technical hypothesis : $\sup_N \|\mathbf{R}_N\| < \infty$

Theorem 1.

Let $[a, b]$ such that $]a - \epsilon, b + \epsilon[\subset (\mathcal{S}_N)^c$ for each $N > N_0$. Then, almost surely, for N large enough, none of the eigenvalues of $\hat{\mathbf{R}}_N$ appears in $[a, b]$.

The new results II.

To simplify the statement of the second theorem, formulation adapted to the context of source localization.

- K sources and M sensors, $\mathbf{A} = \mathbf{DS}$
- $\mathrm{Rank}(\mathbf{R}_N) = K$

0 is eigenvalue of $\mathbf{R}_N$ with multiplicity $M - K$. 0 is of course associated to cluster $[x_{1,-}^{(N)}, x_{1,+}^{(N)}]$.

The new results II.

Theorem 2

Assume that it exists N_0 such that for each $N > N_0$, eigenvalue 0 is separated from the others and that

$$\sup_{N>N_0} x_{1,+}^{(N)} < \inf_{N>N_0} x_{2,-}^{(N)}$$

Consider $t_{1,-} < t_{1,+} < t_{2,-}$ such that

$$t_{1,-} < \inf_{N>N_0} x_{1,-}^{(N)} < \sup_{N>N_0} x_{1,+}^{(N)} < t_{1,+} < t_{2,-} < \inf_{N>N_0} x_{2,-}^{(N)}$$

Then, almost surely, for N large enough, $\hat{\lambda}_1^{(N)}, \ldots, \hat{\lambda}_{M-K}^{(N)} \in [t_{1,-}, t_{1,+}]$ and $\hat{\lambda}_{M-K+1}^{(N)} > t_{2,-}$.

The new results II.

Theorem 2

Assume that it exists N_0 such that for each $N > N_0$, eigenvalue 0 is separated from the others and that

$$\sup_{N>N_0} x_{1,+}^{(N)} < \inf_{N>N_0} x_{2,-}^{(N)}$$

Consider $t_{1,-} < t_{1,+} < t_{2,-}$ such that

$$t_{1,-} < \inf_{N>N_0} x_{1,-}^{(N)} < \sup_{N>N_0} x_{1,+}^{(N)} < t_{1,+} < t_{2,-} < \inf_{N>N_0} x_{2,-}^{(N)}$$

Then, almost surely, for N large enough, $\hat{\lambda}_1^{(N)}, \ldots, \hat{\lambda}_{M-K}^{(N)} \in [t_{1,-}, t_{1,+}]$ and $\hat{\lambda}_{M-K+1}^{(N)} > t_{2,-}$.

Existing related results.

- Bai and Silverstein 1998 in the context of the model $\mathbf{Y} = \mathbf{HW}$. $\mathbf{W}$ possibly non Gaussian
- Capitaine, Donati-Martin, and Feral 2009 in the context of the deformed Wigner model $\mathbf{Y} = \mathbf{A} + \mathbf{X}$, $\mathbf{X}$ Gaussian Wigner matrix, $\mathbf{A}$ deterministic hermitian matrix with constant rank.

Sketch of the proofs I.

Follow the Gaussian methods of Capitaine, Donati-Martin, and Feral 2009 based on ideas developed by Haagerup and Thorbjornsen 2005 in a different context.

Show that $\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{\hat{\lambda}_k - z}\right) = m_N(z) + \frac{\xi_N(z)}{N^2}$ where $\xi_N(z)$ is analytic on $\mathbb{C} - \mathbb{R}^+$, and satisfies

$$|\xi_N(z)| \leq (|z| + C)^l P(\frac{1}{|\mathrm{Im}(z)|})$$

P is a polynomial independent of N, C and l are independent of N. Use approaches developed by Pastur based on the Poincaré-Nash inequality and a Gaussian integration by parts formula (see Dumont-Hachem-Lasaulce-Loubaton-Najim 2010 in the context of a more general information plus noise model).

Sketch of the proofs II.

Using a useful Lemma in Haagerup and Thorbjornsen 2005 as well as the Stieljès inversion formula, we obtain that for each compactly supported $\mathcal{C}_\infty$ function ψ defined on $\mathbb{R}$, then

$$\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k)\right) = \int_{\mathcal{S}_N}\psi(\lambda)\mu_N(d\lambda) + O(\frac{1}{N^2})$$

Use this identity for well chosen functions ψ.

Proof of Theorem 2.

Assume Theorem 1 holds.

$\epsilon > 0$ such that $t_{1,+} + \epsilon < t_{2,-}$

- $\psi(\lambda) = 1$ on $[t_{1,-}, t_{1,+}]$
- $\psi(\lambda) = 0$ on $([t_{1,-} - \epsilon, t_{1,+} + \epsilon])^c$
- $\psi(\lambda)$ $\mathcal{C}_\infty$

Useful lemma

Under the hypotheses of Theorem 2,
$\mu_N([x_{1,-}^{(N)}, x_{1,+}^{(N)}]) = \mu_N([t_{1,-}, t_{1,+}]) = \frac{M-K}{M}$

We recall that

- $\psi(\lambda) = 1$ on $[x_{1,-}^{(N)}, x_{1,+}^{(N)}]$, $\psi(\lambda) = 0$ on the other components of $\mathcal{S}_N$
- $\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k)\right) = \int_{\mathcal{S}_N}\psi(\lambda)\mu_N(d\lambda) + O(\frac{1}{N^2})$

Therefore

$$\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k)\right) - \frac{M-K}{M} = O(\frac{1}{N^2})$$

Therefore

$$\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k)\right) - \frac{M-K}{M} = O(\frac{1}{N^2})$$

Use the Poincaré-Nash inequality to establish that

$$\mathrm{Var}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k)\right) = O(\frac{1}{N^4})$$

Therefore

$$\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k)\right) - \frac{M-K}{M} = O(\frac{1}{N^2})$$

Use the Poincaré-Nash inequality to establish that

$$\mathrm{Var}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k)\right) = O(\frac{1}{N^4})$$

This implies immediately that

$$\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\psi(\hat{\lambda}_k) - \frac{M-K}{M}\right)^2 = O(\frac{1}{N^4})$$

Define $E_N = \{\omega, \left| \frac{1}{M} \sum_{k=1}^{M} \psi(\hat{\lambda}_k) - \frac{M-K}{M} \right| > \frac{1}{N^{4/3}}$

Markov inequality + Borel-Cantelli lemma :

- $P(E_N) < \frac{1}{N^{4/3}}$
- $P(\limsup E_N) = 0$

Define $E_N = \{\omega, \left| \frac{1}{M} \sum_{k=1}^{M} \psi(\hat{\lambda}_k) - \frac{M-K}{M} \right| > \frac{1}{N^{4/3}}$

- Almost surely, for $N > N_1$, $\left| \frac{1}{M} \sum_{k=1}^{M} \psi(\hat{\lambda}_k) - \frac{M-K}{M} \right| < \frac{1}{N^{4/3}}$
- Almost surely, for $N > N_1$, $\left| \sum_{k=1}^{M} \psi(\hat{\lambda}_k) - (M-K) \right| = O(\frac{1}{N^{1/3}})$
- By Theorem 1, for each k, $\hat{\lambda}_k$ does not belong to $[t_{1,-} - \epsilon, t_{1,-}] \cup [t_{1,+}, t_{1,+} + \epsilon]$.
- Hence, $\psi(\hat{\lambda}_k) = 1$ (iff $\hat{\lambda}_k \in [t_{1,-}, t_{1,+}]$) or $\psi(\hat{\lambda}_k) = 0$ (iff $\hat{\lambda}_k$ does not belong to $[t_{1,-}, t_{1,+}]$)
- We finally obtain that $\sum_{k=1}^{M} \psi(\hat{\lambda}_k) = \mathrm{Card}\{k, \hat{\lambda}_k \in [t_{1,-}, t_{1,+}]\}$

Conclusion

- Almost surely, for N large enough,
  $\sum_{k=1}^{M} \psi(\hat{\lambda}_k) = \mathrm{Card}\{k, \hat{\lambda}_k \in [t_{1,-}, t_{1,+}]\} = (M - K)$
- The $M - K$ eigenvalues lying in $[t_{1,-}, t_{1,+}]$ are necessarily the $M - K$ smallest. Otherwise, the smallest one would belong to $[0, t_{1,-}]$, a contradiction by Theorem 1.

Plan

1. Problem statement.
2. The asymptotic behaviour of the eigenvalue distribution of $\hat{\mathbf{R}}_N$.
3. Almost sure locations of the eigenvalues of $\hat{\mathbf{R}}_N$.
4. **Conclusion**

Conclusion

These results have many statistical applications.

- Consistent estimation of direction of arrivals using subspace methods (Vallet-Loubaton-Mestre 2009)
- Information plus Noise spiked models ($\mathrm{Rank}(\mathbf{A})$ is fixed) : convergence of the largest eigenvalues, consistent estimation of the largest eigenvalues and the corresponding eigenvectors.
- ....