Exact separation of eigenvalues of large information plus noise complex Gaussian models

Philippe Loubaton, Pascal Vallet

Université de Paris-Est / Marne la Vallée, LIGM

11/10/2010

Plan

Plan

The information plus noise model

Introduced in Dozier-Silverstein-2007.

$M(N) \times N$ matrix $\mathbf{\Sigma}_N$

$$\mathbf{\Sigma}_N = \mathbf{B}_N + \sigma \mathbf{W}_N$$

- $\mathbf{B}_N$ deterministic matrix $\sup_N \|\mathbf{B}_N\| < +\infty$
- $\mathbf{W}_N$ zero mean complex Gaussian i.i.d. matrix $\mathbb{E}|\mathbf{W}_{N,i,j}|^2 = \frac{1}{N}$

Problem statement

Empirical covariance matrix $\hat{\mathbf{R}}_N = \mathbf{\Sigma}_N \mathbf{\Sigma}_N^*$

$(M, N) \to +\infty$, $c_N = \frac{M}{N} \to c < 1$

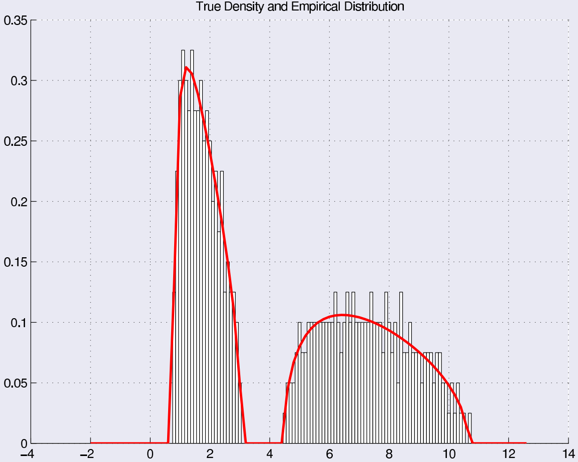
Prove the "Exact Separation" of the eigenvalues of $\hat{\mathbf{R}}_N$
Property introduced by Bai and Silverstein 1999 in the context of zero mean possibly non Gaussian correlated random matrices

Numerical illustration (I).

- $\sigma^2 = 2$
- Eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ 0 with multiplicity $\frac{M}{2}$, 5 with multiplicity $\frac{M}{2}$
- $c_N = \frac{M}{N}$, $c_N = 0.2$
- Representation of histograms of the eigenvalues of $\hat{\mathbf{R}}_N$

Numerical illustration (II).

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



Motivation

See the talk of P. Vallet tomorrow

- $\mathrm{Rank}(\mathbf{B}_N) = K(N) < M$
- $\mathbf{\Pi}_N$ orthogonal projection matrix on $(\mathrm{Range}(\mathbf{B}_N))^{\perp}$

Subspace estimation methods.

- Estimate consistently $\mathbf{a}_N^* \Pi_N \mathbf{a}_N$ from $\mathbf{\Sigma}_N$
- Needs to evaluate the location of the eigenvalues of $\hat{\mathbf{R}}_N$

Plan

1. Problem statement.
2. **Behaviour of the eigenvalue distribution of $\hat{\mathbf{R}}_N$.**
3. Exact separation of the eigenvalues of $\hat{\mathbf{R}}_N$.
4. Conclusion

The "asymptotic" limit eigenvalue distribution μ_N

Notation

$N \to +\infty$ stands for $(M,N) \to +\infty$, $c_N = \frac{M}{N} \to c < 1$

- $(\hat{\lambda}_{k,N})_{k=1,\dots,M}$ eigenvalues of $\hat{\mathbf{R}}_N$, $(\lambda_{k,N})_{k=1,\dots,M}$ eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$, arranged in decreasing order
- $Rank(\mathbf{B}_N) = K(N) < M$, $\lambda_{K+1,N} = \ldots = \lambda_{M,N} = 0$

Dozier-Silverstein 2007 : It exists a deterministic probability measure μ_N carried by $\mathbb{R}^+$ such that

- $\frac{1}{M}\sum_{k=1}^{M}\delta(\lambda - \hat{\lambda}_{k,N}) - \mu_N \to 0$ weakly almost surely

How to characterize μ_N

The Stieltjès transform $m_N(z)$ of μ_N

- $m_N(z) = \int_{\mathbb{R}^+} \frac{\mu_N(d\lambda)}{\lambda - z}$ defined on $\mathbb{C} - \mathbb{R}^+$

$m_N(z)$ is solution of the equation

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

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

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

$\mathcal{S}_N$ support of μ_N

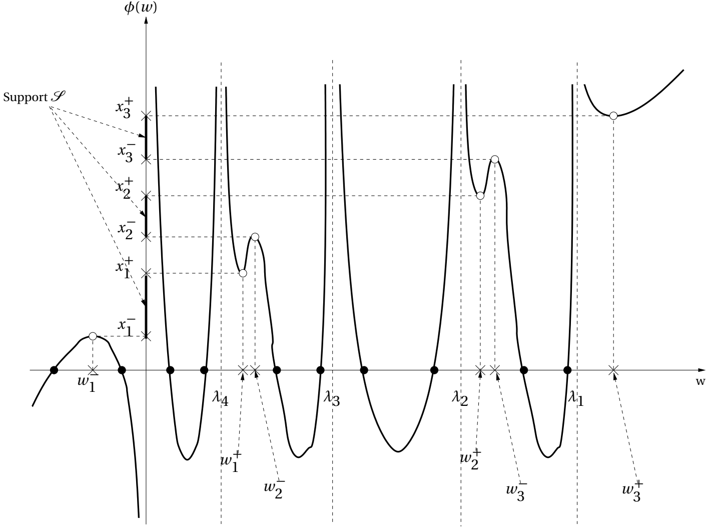
Dozier-Silverstein-2007

- For each $x \in \mathbb{R}$, $\lim_{z \to x, z \in \mathbb{C}^+} m_N(z) = m_N(x)$ exists
- $x \to m_N(x)$ continuous on $\mathbb{R}$, continuously differentiable on $\mathbb{R} \backslash \partial \mathcal{S}_N$
- $\mu_N(d\lambda)$ absolutely continuous, density $\frac{1}{\pi} \mathrm{Im}(m_N(x))$

Characterization of $\mathcal{S}_N$.

Reformulation of D-S 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 $2Q$ positive extrema with preimages $w_{1,-}^{(N)} < w_{1,+}^{(N)} < \ldots w_{Q,-}^{(N)} < w_{Q,+}^{(N)}$. These extrema verify $x_{1,-}^{(N)} < x_{1,+}^{(N)} < \ldots x_{Q,-}^{(N)} < x_{Q,+}^{(N)}$.
- $\mathcal{S}_N = [x_{1,-}^{(N)}, x_{1,+}^{(N)}] \cup \ldots [x_{Q,-}^{(N)}, x_{Q,+}^{(N)}]$
- Each eigenvalue $\lambda_{l,N}$ of $\mathbf{B}_N\mathbf{B}_N^*$ belongs to an interval $(w_{k,-}^{(N)}, w_{k,+}^{(N)})$



$\phi(w)$
Support $\mathscr{S}$
x_3^+
x_3^-
x_2^+
x_2^-
x_1^+
x_1^-
w_1
λ_4
λ_3
λ_2
λ_1
w
w_1^+
w_2^-
w_2^+
w_3^-
w_3^+

Some definitions

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

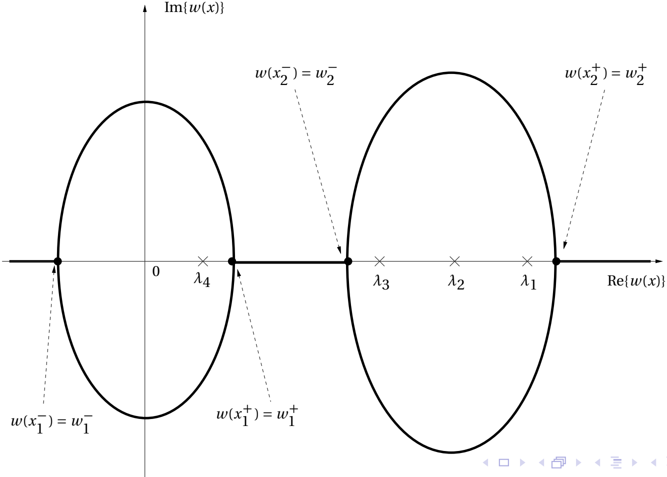
Some useful properties of $w_N(x)$

$$w_N(x) = x(1+\sigma^2 c_N m_N(x))^2 - \sigma^2(1-c_N)(1+\sigma^2 c_N m_N(x)).$$

- $\phi_N(w_N(x)) = x$ for each x
- $\mathrm{Int}(\mathcal{S}_N) = \{x, \mathrm{Im}(w_N(x)) > 0\}$
- $w_N(x)$ is real and increasing on each component of $\mathcal{S}_N^c$
- $w_N(x_{q,N}^-) = w_{q,N}^-, w_N(x_{q,N}^+) = w_{q,N}^+$
- $w_N(x)$ is continuous on $\mathbb{R}$ and continuously differentiable on $\mathbb{R} \backslash \partial \mathcal{S}_N$
- $|w_N'(x)| \simeq \frac{1}{|x - x_{q,N}^{-,+}|^{1/2}}$ if $x \simeq x_{q,N}^{-,+}$

Contours associated to function $x \to w_N(x)$ (I)

Illustration 2 clusters.



Contours associated to function $x \to w_N(x)$ (II)

$\mathcal{C}_q = \{w_N(x), x \in [x_{q,N}^-, x_{q,N}^+]\} \cup \{w_N(x)^*, x \in [x_{q,N}^-, x_{q,N}^+]\}$

- Encloses the eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ associated to cluster $[x_{q,N}^-, x_{q,N}^+]$
- Continuously differentiable path (except at $x_{q,N}^-, x_{q,N}^+$ where $|w_N'(x)| \simeq \frac{1}{|x - x_{q,N}^{-,+}|^{1/2}}$)

$g(w)$ continuous in a neighborhood of $\mathcal{C}_q$, $g(w^*) = g(w)^*$

$$\int_{\mathcal{C}_q^-} g(w)dw = 2i \int_{x_{q,N}^-}^{x_{q,N}^+} \mathrm{Im}\left(g(w_N(x))w_N'(x)\right) dx$$

Plan

The results.

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 $[a,b]$.

Theorem 2

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,

$$\begin{aligned} \mathrm{card}\{k : \hat{\lambda}_{k,N} < a\} &= \mathrm{card}\{k : \lambda_{k,N} < w_N(a)\} \\ \mathrm{card}\{k : \hat{\lambda}_{k,N} > b\} &= \mathrm{card}\{k : \lambda_{k,N} > w_N(b)\} \end{aligned}$$

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 i.i.d. Wigner matrix (or entries verifying the Poincaré-Nash inequality), $\mathbf{A}$ deterministic hermitian matrix with constant rank.
- No previous result in the context of the Information plus Noise model

Proof of Theorem 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 $\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{\hat{\lambda}_{k,N}-z}\right) = m_N(z) + \frac{\xi_N(z)}{N^2}$ 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 Poincaré-Nash inequality and the Gaussian integration by parts formula.

Proof of Theorem I.

Fundamental Lemma in Haagerup and Thorbjornsen 2005

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

for each $\psi \in \mathcal{C}_c^{\infty}(\mathbb{R},\mathbb{R})$.

Use this for well chosen functions ψ

Proof of Theorem 2.

$\eta > 0$ such that $a - \epsilon < a - \eta$

- $\psi(\lambda) = 1$ on $[0, a - \eta]$
- $\psi(\lambda) = 0$ if $\lambda \geq a$
- $\psi(\lambda) \in \mathcal{C}_c^\infty(\mathbb{R}, \mathbb{R})$

Theorem 1 with $[a - \eta, b]$ in place of $[a, b]$

Almost surely for N large enough

$$\mathrm{Tr}\,\psi(\hat{\mathbf{R}}_N) = \sum_{k=1}^{M} \psi(\hat{\lambda}_{k,N}) = \mathrm{card}\{k : \hat{\lambda}_{k,N} < a\}$$

Use Haagerup-Thorbjornsen Lemma

$$\mathbb{E}\left(\frac{1}{M}\operatorname{Tr}\psi(\hat{\mathbf{R}}_N)\right) = \mu_N([0, a-\eta]) + O(\frac{1}{N^2}) = \mu_N([0, a]) + O(\frac{1}{N^2})$$

Use Poincaré-Nash inequality and Haagerup-Thorbjornsen Lemma

$$\operatorname{Var}\left(\frac{1}{M}\operatorname{Tr}\psi(\hat{\mathbf{R}}_N)\right) = O(\frac{1}{N^4})$$

Markov inequality and Borel-Cantelli lemma

$\operatorname{Tr}\psi(\hat{\mathbf{R}}_N) - M\,\mu_N([0, a]) \to 0$ almost surely

Evaluate $M\mu_N([x_{q,N}^-, x_{q,N}^+])$

Show that $M\mu_N([x_{q,N}^-, x_{q,N}^+])$ = number of eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ associated to cluster $[x_{q,N}^-, x_{q,N}^+]$

$$\mu_N([x_{q,N}^-, x_{q,N}^+]) = \frac{1}{\pi}\int_{x_{q,N}^-}^{x_{q,N}^+} \mathrm{Im}\, m_N(x)\, dx$$

Evaluate the integral as a contour integral along path $\mathcal{C}_q$

- $m_N(x) = \frac{f_N(w_N(x))}{1-\sigma^2 c_N f_N(w_N(x))}$
- $\phi_N'(w_N(x))w_N'(x) = 1$ because $\phi_N(w_N(x)) = x$

Alternative expression of $\mu_N([x^-_{q,N}, x^+_{q,N}])$

$$\mu_N([x^-_{q,N}, x^+_{q,N}]) = \frac{1}{2i\pi} \int_{\mathcal{C}_q^-} \frac{f_N(w)\phi'_N(w)}{1 - \sigma^2 c_N \phi_N(w)} \, dw$$

Can be evaluted using the Residu Theorem

- $M \, \mu_N([x^-_{q,N}, x^+_{q,N}] =$ number of eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ enclosed by $\mathcal{C}_q$
- $M \, \mu_N([x^-_{q,N}, x^+_{q,N}] =$ number of eigenvalues of $\mathbf{B}_N\mathbf{B}_N^*$ associated to $[x^-_{q,N}, x^+_{q,N}]$

Plan

1. Problem statement.
2. Behaviour of the eigenvalue distribution of $\hat{\mathbf{R}}_N$.
3. Exact separation of the eigenvalues of $\hat{\mathbf{R}}_N$.
4. **Conclusion**

Possible extensions of the approach.

Non Gaussian model, but entries of W satisfy the Poincaré-Nash inequality.

$$\mathbb{E}\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{\hat{\lambda}_{k,N}-z}\right) = m_N(z) + \frac{1}{N}\int \frac{d\,\nu_N(\lambda)}{\lambda - z}\,d\lambda + \frac{\xi_N(z)}{N^2}$$

Support of $\nu_N \subset \mathcal{S}_N$? If yes, exact separation holds if and only for each q,

$$\nu_N([[x^-_{q,N}, x^+_{q,N}]]) = 0$$

Statistical applications

- Consistent estimation of direction of arrivals using subspace methods (Vallet-Loubaton-Mestre 2009)
- Information plus Noise spiked models ($\mathrm{Rank}(\mathbf{B}_N)$ is fixed) : easy to prove Benaych and Rao results on the behaviour of the largest eigenvalues
- ....