Crash Course on Random Matrix Theory

Part II: Advanced notions and applications to signal processing

Morning Session: Advanced notions of RMT

Romain Couillet and Mérouane Debbah

SUPELEC

27-01-2012

Outline

Outline

Outline

Why go beyond the spectrum?

- Limiting spectral results only say where the "mass" of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $f_N(x) = \frac{1}{N}\sum_{k=1}^{N} \delta(x - a_k)$.

Why go beyond the spectrum?

- Limiting spectral results only say where the "mass" of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $f_N(x) = \frac{1}{N}\sum_{k=1}^{N}\delta(x - a_k)$.
  - $f_N^{(0)}(x) = \frac{1}{N}\delta(x) + \frac{1}{N}\sum_{k=1}^{N-1}\delta(x - a_k)$ also converges to F.

Why go beyond the spectrum?

- Limiting spectral results only say where the “mass” of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $f_N(x) = \frac{1}{N}\sum_{k=1}^{N}\delta(x-a_k)$.
  - $f_N^{(0)}(x) = \frac{1}{N}\delta(x) + \frac{1}{N}\sum_{k=1}^{N-1}\delta(x-a_k)$ also converges to F.
  - if F_N and $F_N^{(0)}$ are discrete and differ by $o(N)$ bounded masses, $F_N^{(0)} \Rightarrow F$.

Why go beyond the spectrum?

- Limiting spectral results only say where the "mass" of eigenvalues lies asymptotically. Say $F_N \Rightarrow F$, with $f_N(x) = \frac{1}{N}\sum_{k=1}^{N}\delta(x-a_k)$.
  - $f_N^{(0)}(x) = \frac{1}{N}\delta(x) + \frac{1}{N}\sum_{k=1}^{N-1}\delta(x-a_k)$ also converges to F.
  - if F_N and $F_N^{(0)}$ are discrete and differ by $o(N)$ bounded masses, $F_N^{(0)} \Rightarrow F$.
- We know that, for $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ with i.i.d. zero mean variance $1/n$,

$$F^{\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}} \Rightarrow F_c$$

with F_c is the compactly supported Marčenko-Pastur law of parameter $c = \lim_N \frac{N}{n}$.

Question: for very large N, where are the eigenvalues of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$?

Are there eigenvalues outside the support ?

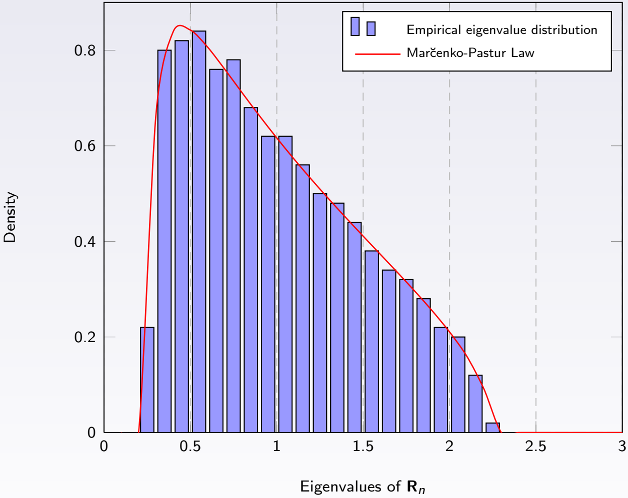


Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$

No eigenvalue outside the support of sample covariance matrices

Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," The Annals of Probability, vol. 26, no.1 pp. 316-345, 1998.

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ with i.i.d. entries with zero mean, variance $1/n$ and 4^{th} order moment of order $O(1/n^2)$. Let $\mathbf{T}_N \in \mathbb{C}^{N\times N}$ be nonrandom and bounded in norm and with $F^{\mathbf{T}_N} \Rightarrow H$. We know that

$$F^{\mathbf{B}_N} \Rightarrow F \text{ almost surely}, \quad \mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}.$$

Let F_N be the distribution with $m_N(z)$ solution of

$$\underline{m}_N = -\left(z - \frac{N}{n}\int \frac{\tau}{1+\tau \underline{m}_N} dF^{\mathbf{T}_N}(\tau)\right)^{-1}, \quad \underline{m}_N(z) = \frac{N}{n}m_N(z) + \frac{N-n}{n}\frac{1}{z}.$$

Choose $N_0 \in \mathbb{N}$ and $[a,b]$, $a > 0$, outside the union of the supports of F and F_N for all $N \geqslant N_0$. Denote $\mathcal{L}_N$ the set of eigenvalues of $\mathbf{B}_N$. Then,

$$P(\mathcal{L}_N \cap [a,b] \neq \emptyset \text{ i.o.}) = 0.$$

How to read the result?

- If $\mathbf{T}_N = \mathbf{I}_N$ for all N, then this result is equivalent to

  "For $[a, b]$ outside the support of the Marčenko-Pastur law, with probability 1, $\mathbf{B}_N$ has no eigenvalue in $[a, b]$ for all large N"

How to read the result?

- If $\mathbf{T}_N = \mathbf{I}_N$ for all N, then this result is equivalent to

  "For $[a, b]$ outside the support of the Marčenko-Pastur law, with probability 1, $\mathbf{B}_N$ has no eigenvalue in $[a, b]$ for all large N"
- If $\mathbf{T}_N$ is not identity,
  - call S the support of the limiting F.
  - for some N_0, take the l.s.d. of $\mathbf{B}_N$ as if $\lim_N F^{\mathbf{T}_N} = F^{\mathbf{T}_{N_0}}$, and call its support S_{N_0}.
  - do the previous for all $N \geqslant N_0$. Call $\mathcal{A} = S \cup \bigcap_{N \geqslant N_0} S_N$.
  - take $[a, b]$ outside $\mathcal{A}$, and pick a random sequence $\mathbf{B}_1, \mathbf{B}_2, \ldots$. The result shows that, for all N large, there is no eigenvalue of $\mathbf{B}_N$ in $[a, b]$.

How to read the result?

- If $\mathbf{T}_N = \mathbf{I}_N$ for all N, then this result is equivalent to

  "For $[a,b]$ outside the support of the Marčenko-Pastur law, with probability 1, $\mathbf{B}_N$ has no eigenvalue in $[a,b]$ for all large N"

- If $\mathbf{T}_N$ is not identity,
  - call S the support of the limiting F.
  - for some N_0, take the l.s.d. of $\mathbf{B}_N$ as if $\lim_N F^{\mathbf{T}_N} = F^{\mathbf{T}_{N_0}}$, and call its support S_{N_0}.
  - do the previous for all $N \geqslant N_0$. Call $\mathcal{A} = S \cup \bigcap_{N \geqslant N_0} S_N$.
  - take $[a,b]$ outside $\mathcal{A}$, and pick a random sequence $\mathbf{B}_1, \mathbf{B}_2, \ldots$. The result shows that, for all N large, there is no eigenvalue of $\mathbf{B}_N$ in $[a,b]$.
- this is very different from taking $[a,b]$ only outside the support of F only!
- this is essential to understand spiked models, discussed later.

No eigenvalue outside the support: which models?

J. W. Silverstein, P. Debashis, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has already been shown that (for all large N) there is no eigenvalues outside the support of
  - *Marčenko-Pastur law*: $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Sample covariance matrix*: $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{T}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment.

No eigenvalue outside the support: which models?

J. W. Silverstein, P. Debashis, “No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix,” J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has already been shown that (for all large N) there is no eigenvalues outside the support of
  - *Marčenko-Pastur law*: $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Sample covariance matrix*: $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{T}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment.

J. W. Silverstein, Z.D. Bai, Y.Q. Yin, “A note on the largest eigenvalue of a large dimensional sample covariance matrix,” Journal of Multivariate Analysis, vol. 26, no. 2, pp. 166-168, 1988.

- If 4^{th} order moment is infinite,

$$\limsup_N \lambda_{\max}^{\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \infty$$

No eigenvalue outside the support: which models?

J. W. Silverstein, P. Debashis, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has already been shown that (for all large N) there is no eigenvalues outside the support of
  - *Marčenko-Pastur law*: $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Sample covariance matrix*: $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ and $\mathbf{X}^{\mathsf{H}}\mathbf{T}\mathbf{X}$, $\mathbf{X}$ i.i.d. with zero mean, variance $1/N$, finite 4^{th} order moment.
  - *Doubly-correlated matrix*: $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment.

J. W. Silverstein, Z.D. Bai, Y.Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," Journal of Multivariate Analysis, vol. 26, no. 2, pp. 166-168, 1988.

- If 4^{th} order moment is infinite,

$$\limsup_N \lambda_{\max}^{\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \infty$$

J. Silverstein, Z. Bai, "No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices" to appear in Random Matrices: Theory and Applications.

- Only recently, information plus noise models, $\mathbf{X}$ with i.i.d. zero mean, variance $1/N$, finite 4^{th} order moment.

$$(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}}$$

Sketch of Proof

- Proof entirely relies on the Stieltjes transform.
- Up to now, we know $|m_{\mathbf{B}_N}(z) - m_N(z)| \xrightarrow{\text{a.s.}} 0$ for $z \in \mathbb{C} \setminus \mathbb{R}^-$.

Sketch of Proof

- Proof entirely relies on the Stieltjes transform.
- Up to now, we know $|m_{\mathbf{B}_N}(z) - m_N(z)| \xrightarrow{\text{a.s.}} 0$ for $z \in \mathbb{C} \setminus \mathbb{R}^-$.
- This is not enough, we need in fact to show: for $z = x + i\sqrt{k}v_N$, $v_N = N^{-1/68}$, $k = 1, \ldots, 34$,

$$\max_{1\leqslant k\leqslant 34} \sup_{x\in[a,b]} \left| m_{\mathbf{B}_N}(x + ik^{\frac{1}{2}}v_N) - m_N((x + ik^{\frac{1}{2}}v_N) \right| = o(v_N^{67}).$$

Sketch of Proof

- Proof entirely relies on the Stieltjes transform.
- Up to now, we know $|m_{\mathbf{B}_N}(z) - m_N(z)| \xrightarrow{\text{a.s.}} 0$ for $z \in \mathbb{C} \setminus \mathbb{R}^-$.
- This is not enough, we need in fact to show: for $z = x + i\sqrt{k}v_N$, $v_N = N^{-1/68}$, $k = 1, \ldots, 34$,

$$\max_{1 \leqslant k \leqslant 34} \sup_{x \in [a,b]} \left| m_{\mathbf{B}_N}(x + ik^{\frac{1}{2}}v_N) - m_N((x + ik^{\frac{1}{2}}v_N) \right| = o(v_N^{67}).$$

- Expanding the Stieltjes transforms and considering only the imaginary parts, this is

$$\max_{1 \leqslant k \leqslant 34} \sup_{x \in [a,b]} \left| \int \frac{d(F^{\mathbf{B}_N}(\lambda) - F_N(\lambda))}{(x-\lambda)^2 + kv_N^2} \right| = o(v_N^{66})$$

almost surely. Taking successive differences over the 34 values of k, we end up with

$$\sup_{x \in [a,b]} \left| \int \frac{(v_N^2)^{33} d(F^{\mathbf{B}_N}(\lambda) - F_N(\lambda))}{\prod_{k=1}^{34}((x-\lambda)^2 + kv_N^2)} \right| = o(v_N^{66})$$

Consider $a' < a$ and $b' > b$ such that $[a', b']$ is outside the support of F. We then have

$$\sup_{x \in [a,b]} \left| \int \frac{1_{\mathbb{R}^+ \setminus [a',b']}(\lambda) d(F\mathbf{B}_N(\lambda) - F_N(\lambda))}{\prod_{k=1}^{34}((x-\lambda)^2 + kv_N^2)} + \sum_{\lambda_j \in [a',b']} \frac{v_N^{68}}{\prod_{k=1}^{34}((x-\lambda_j)^2 + kv_N^2)} \right| = o(1)$$

almost surely. If, there is one eigenvalue of all $\mathbf{B}_{\phi(N)}$ in $[a, b]$, then one term of the sum is $1/34! > 0$. So the integral must away from zero. But the integral tends to 0. Contradiction.

What's the link with wireless communications?

Assume N sensors wish to detect the presence of a signal. They scan successive samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then

What's the link with wireless communications?

Assume N sensors wish to detect the presence of a signal. They scan successive samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then

- if $\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$ has eigenvalues outside the support: with high probability, a signal was transmitted.

What's the link with wireless communications?

Assume N sensors wish to detect the presence of a signal. They scan successive samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then

- if $\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$ has eigenvalues outside the support: with high probability, a signal was transmitted.
- if $\mathbf{R}_n$ has all eigenvalues inside the *expected* noise support, what can we say?
  - we cannot conclude so far
  - we need to further study the spectrum

Outline

Stieltjes transform inversion for covariance matrix models

J. W. Silverstein, S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.

- We know for the model $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ that, if $F^{\mathbf{T}_N} \Rightarrow H$, the Stieltjes transform of the e.s.d. of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N\mathbf{X}_N$ satisfies $m_{\underline{\mathbf{B}}_N}(z) \xrightarrow{\text{a.s.}} m_{\underline{F}}(z)$, with

$$m_{\underline{F}}(z) = \left(-z - c\int \frac{t}{1+tm_{\underline{F}}(z)}dH(t)\right)^{-1}$$

which is unique on the set $\{z \in \mathbb{C}^+, m_{\underline{F}}(z) \in \mathbb{C}^+\}$.

Stieltjes transform inversion for covariance matrix models

J. W. Silverstein, S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.

- We know for the model $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ that, if $F^{\mathbf{T}_N} \Rightarrow H$, the Stieltjes transform of the e.s.d. of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$ satisfies $m_{\underline{\mathbf{B}}_N}(z) \xrightarrow{\text{a.s.}} m_{\underline{F}}(z)$, with

$$m_{\underline{F}}(z) = \left(-z - c\int \frac{t}{1+tm_{\underline{F}}(z)}dH(t)\right)^{-1}$$

  which is unique on the set $\{z \in \mathbb{C}^+, m_{\underline{F}}(z) \in \mathbb{C}^+\}$.

- This can be inverted into

$$z_{\underline{F}}(m) = -\frac{1}{m} - c\int \frac{t}{1+tm}dH(t)$$

  for $m \in \mathbb{C}^+$.

Stieltjes transform inversion and spectrum characterization

Remember that we can evaluate the spectrum density by taking a complex line close to $\mathbb{R}$ and evaluating $\Im[m_{\underline{F}}(z)]$ along this line. Now we can do better.

Stieltjes transform inversion and spectrum characterization

Remember that we can evaluate the spectrum density by taking a complex line close to $\mathbb{R}$ and evaluating $\Im[m_{\underline{F}}(z)]$ along this line. Now we can do better.

It is shown that

$$\lim_{\substack{z\to x\in\mathbb{R}^* \\ z\in\mathbb{C}^+}} m_{\underline{F}}(z) = m_0(x) \quad \text{exists.}$$

We also have,

- for x_0 inside the support, the density $\underline{f}(x)$ of $\underline{F}$ in x_0 is $\frac{1}{\pi}\Im[m_0]$ with m_0 the unique solution $m \in \mathbb{C}^+$ of
$$[z_{\underline{F}}(m) =]\ x_0 = -\frac{1}{m} - c\int \frac{t}{1+tm} dH(t)$$
- let $m_0 \in \mathbb{R}^*$ and $x_{\underline{F}}$ the equivalent to $z_{\underline{F}}$ on the real line. Then "x_0 outside the support of $\underline{F}$" is equivalent to "$x'_{\underline{F}}(m_{\underline{F}}(x_0)) > 0$, $m_{\underline{F}}(x_0) \neq 0$, $-1/m_{\underline{F}}(x_0)$ outside the support of H".

Stieltjes transform inversion and spectrum characterization

Remember that we can evaluate the spectrum density by taking a complex line close to $\mathbb{R}$ and evaluating $\Im[m_{\underline{F}}(z)]$ along this line. Now we can do better.

It is shown that

$$\lim_{\substack{z \to x \in \mathbb{R}^* \\ z \in \mathbb{C}^+}} m_{\underline{F}}(z) = m_0(x) \quad \text{exists.}$$

We also have,

- for x_0 inside the support, the density $\underline{f}(x)$ of $\underline{F}$ in x_0 is $\frac{1}{\pi}\Im[m_0]$ with m_0 the unique solution $m \in \mathbb{C}^+$ of
$$[z_{\underline{F}}(m) =]\ x_0 = -\frac{1}{m} - c\int \frac{t}{1+tm} dH(t)$$
- let $m_0 \in \mathbb{R}^*$ and $x_{\underline{F}}$ the equivalent to $z_{\underline{F}}$ on the real line. Then "x_0 outside the support of $\underline{F}$" is equivalent to "$x_{\underline{F}}'(m_{\underline{F}}(x_0)) > 0$, $m_{\underline{F}}(x_0) \neq 0$, $-1/m_{\underline{F}}(x_0)$ outside the support of H".

This provides another way to determine the support!. For $m \in (-\infty, 0)$, evaluate $x_{\underline{F}}(m)$. Whenever $x_{\underline{F}}$ decreases, the image is outside the support. The rest is inside.

Another way to determine the spectrum: spectrum to analyze

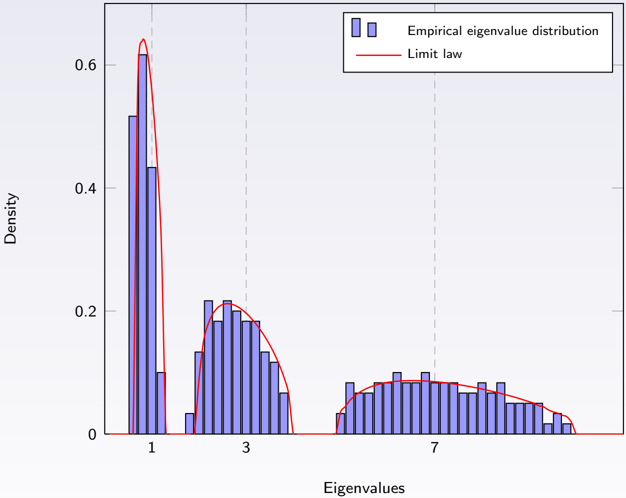


Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7.

Another way to determine the spectrum: inverse function method

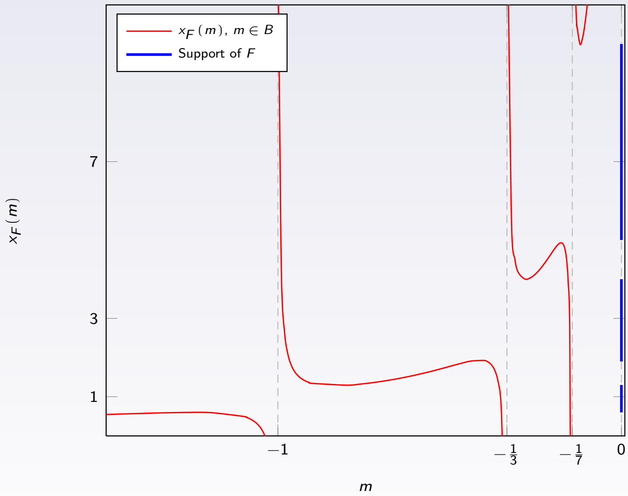


Figure: Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7. The support of F is read on the vertical axis, whenever m_F is decreasing.

Cluster boundaries in sample covariance matrix models

Xavier Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," IEEE Transactions on Information Theory, vol. 54, no. 11, Nov. 2008.

Theorem

Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have i.i.d. entries of zero mean, variance $1/n$, and $\mathbf{T}_N$ be diagonal such that $F^{\mathbf{T}_N} \Rightarrow H$, as $n, N \to \infty$, $N/n \to c$, where H' has K masses in $t_1, \ldots, t_K$ with multiplicity $n_1, \ldots, n_K$ respectively. Then the l.s.d. of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ has support $\mathcal{S}$ given by

$$\mathcal{S} = [x_1^-, x_1^+] \cup [x_2^-, x_2^+] \cup \ldots \cup [x_Q^-, x_Q^+]$$

with $x_q^- = x_F(m_q^-)$, $x_q^+ = x_F(m_q^+)$, and

$$x_F(m) = -\frac{1}{m} - c\frac{1}{n}\sum_{k=1}^{K} n_k \frac{t_k}{1+t_k m}$$

with $2Q$ the number of real-valued solutions counting multiplicities of $x_F'(m) = 0$ denoted in order $m_1^- < m_1^+ \leqslant m_2^- < m_2^+ \leqslant \ldots \leqslant m_Q^- < m_Q^+$.

Comments on spectrum characterization

Previous results allows to determine

- the spectrum boundaries
- the number Q of clusters
- as a consequence, the total separation or not of the spectrum in K clusters.

Comments on spectrum characterization

Previous results allows to determine

- the spectrum boundaries
- the number Q of clusters
- as a consequence, the total separation or not of the spectrum in K clusters.

Mestre goes further: to determine local separability of the spectrum,

- identify the K inflexion points, i.e. the K solutions $m_1, \ldots, m_K$ to

$$x_F''(m) = 0$$

- check whether $x_F'(m_i) > 0$ and $x_F'(m_{i+1}) > 0$
- if so, the cluster in between corresponds to a single population eigenvalue.

Outline

Further than the "no eigenvalues" result

Z. D. Bai, J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.

- The result on "no eigenvalue outside the support"
  - says where eigenvalues are not to be found
  - does not say, as we feel, that (if cluster separation) in cluster k, there are exactly n_k eigenvalues.

Further than the "no eigenvalues" result

Z. D. Bai, J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.

- The result on "no eigenvalue outside the support"
  - says where eigenvalues are not to be found
  - does not say, as we feel, that (if cluster separation) in cluster k, there are exactly n_k eigenvalues.
- This is in fact the case,

Theorem

Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$ *with l.s.d.* F*,* $\mathbf{X}_N$ *i.i.d., zero mean, variance* $1/n$*, finite* 4^{th} *moment,* $F^{\mathbf{T}_N} \Rightarrow H$*, and* $\frac{N}{n} \to c$*. Consider* $0 < a < b$ *such that* $[a,b]$ *is outside the support of* F*. Denote additionally* λ_k*'s and* τ_k*'s the ordered eigenvalues of* $\mathbf{B}_N$ *and* $\mathbf{T}_N$*. Then we have*

1. *If* $c(1-H(0)) > 1$*, then the smallest eigenvalue* x_0 *of the support of* F *is positive and* $\lambda_N \to x_0$ *almost surely, as* $N \to \infty$*.*
2. *If* $c(1-H(0)) \leqslant 1$*, or* $c(1-H(0)) > 1$ *but* $[a,b]$ *is not contained in* $[0,x_0]$*, then, almost surely, there exists* N_0 *such that for all* $N \geqslant N_0$*,*

$$\lambda_{i_N} > b, \quad \lambda_{i_N+1} < a$$

where i_N *is the unique integer such that*

$$\tau_{i_N} > -1/m_F(b)$$
$$\tau_{i_N+1} < -1/m_F(a).$$

Consequence of exact separation

- If eigenvalues are found outside the expected clusters, some extra “signal” must have been transmitted.
- The quantity of eigenvalues in each cluster gives an exact estimate of the multiplicity of the population!
- This is essential for eigen-inference.

Consequence of exact separation

- If eigenvalues are found outside the expected clusters, some extra “signal” must have been transmitted.
- The quantity of eigenvalues in each cluster gives an exact estimate of the multiplicity of the population!
- This is essential for eigen-inference.
- Exact separation is only known for the sample covariance matrix model so far.
- Recently, extension to information-plus-noise model.

Outline

Deeper into the spectrum

- In order to derive statistical detection tests, we need more information on the extreme eigenvalues.

Deeper into the spectrum

- In order to derive statistical detection tests, we need more information on the extreme eigenvalues.
- We will study the fluctuations of the extreme eigenvalues (second order statistics)
- However, the Stieltjes transform method is not adapted here!

Distribution of the largest eigenvalues of $\mathbf{XX}^{\mathsf{H}}$

C. A. Tracy, H. Widom, "On orthogonal and symplectic matrix ensembles," Communications in Mathematical Physics, vol. 177, no. 3, pp. 727-754, 1996.
K. Johansson, "Shape Fluctuations and Random Matrices," Comm. Math. Phys. vol. 209, pp. 437-476, 2000.

Theorem

Let $\mathbf{X} \in \mathbb{C}^{N\times n}$ have i.i.d. Gaussian entries of zero mean and variance $1/n$. Denoting λ_N^+ the largest eigenvalue of $\mathbf{XX}^{\mathsf{H}}$, then

$$N^{\frac{2}{3}} \frac{\lambda_N^+ - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{\frac{1}{2}}} \Rightarrow X^+ \sim F^+$$

with $c = \lim_N N/n$ and F^+ the Tracy-Widom distribution given by

$$F^+(t) = \exp\left(-\int_t^\infty (x-t)^2 q^2(x) dx\right)$$

with q the Painlevé II function that solves the differential equation

$$q''(x) = xq(x) + 2q^3(x)$$
$$q(x) \sim_{x\to\infty} \mathrm{Ai}(x)$$

in which $\mathrm{Ai}(x)$ is the Airy function.

The law of Tracy-Widom

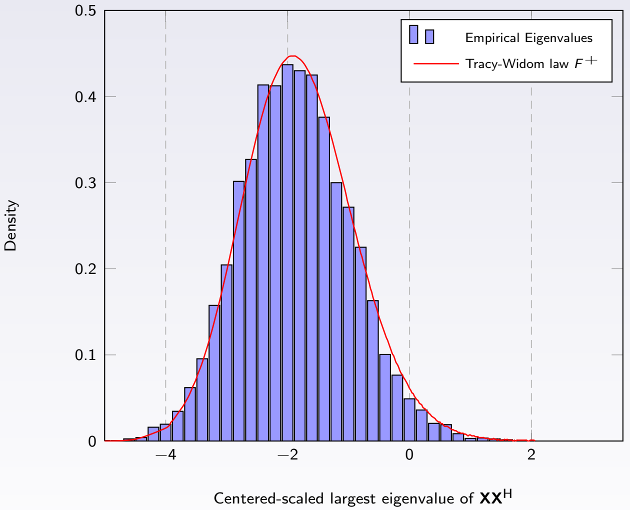


Figure: Distribution of $N^{\frac{2}{3}} c^{-\frac{1}{2}} (1+\sqrt{c})^{-\frac{4}{3}} \left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the distribution of X^+ (distributed as Tracy-Widom law) for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{XX}^{\mathsf{H}}$. Empirical distribution taken over $10,000$ Monte-Carlo simulations.

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \dots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

with $K(x, y)$ the kernel Laguerre polynomial.

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \ldots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

  with $K(x, y)$ the kernel Laguerre polynomial.
- *Fredholm determinants*: we can write hole probability as a Fredholm determinant.

$$P\left(N^{2/3}\left(\lambda_i - (1+\sqrt{c})^2\right) \in A, i = 1, \ldots, N\right) = 1 + \sum_{k \geqslant 1} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K_N(x_i, x_j) \prod dx_i$$

$$\triangleq \det(\mathbf{I}_N - \mathcal{K}_N).$$

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \ldots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

  with $K(x,y)$ the kernel Laguerre polynomial.

- *Fredholm determinants*: we can write hole probability as a Fredholm determinant.

$$P\left(N^{2/3}\left(\lambda_i - (1+\sqrt{c})^2\right) \in A, i = 1, \ldots, N\right) = 1 + \sum_{k \geqslant 1} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K_N(x_i, x_j) \prod dx_i$$
$$\triangleq \det(\mathbf{I}_N - \mathcal{K}_N).$$

- *kernel theory*: show that K_N converges to a Airy kernel.

$$K_N(x,y) \to K_{\mathrm{Airy}}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}.$$

Techniques of proof

Method of proof requires very different tools:

- *orthogonal (Laguerre) polynomials*: to write joint *unordered* eigenvalue distribution as a kernel determinant.

$$\rho_N(\lambda_1, \ldots, \lambda_p) = \det_{i,j=1}^{p} K_N(\lambda_i, \lambda_j)$$

  with $K(x, y)$ the kernel Laguerre polynomial.

- *Fredholm determinants*: we can write hole probability as a Fredholm determinant.

$$P\left(N^{2/3}\left(\lambda_i - (1+\sqrt{c})^2\right) \in A, i = 1, \ldots, N\right) = 1 + \sum_{k \geqslant 1} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K_N(x_i, x_j) \prod dx_i$$

$$\triangleq \det(\mathbf{I}_N - \mathcal{K}_N).$$

- *kernel theory*: show that K_N converges to a Airy kernel.

$$K_N(x, y) \to K_{\mathrm{Airy}}(x, y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x - y}.$$

- *differential equation tricks*: hole probability in $[t, \infty)$ gives right-most eigenvalue distribution, which is simplified as solution of a Painelvé differential equation: the Tracy-Widom distribution.

$$F^+(t) = e^{-\int_t^\infty (x-t) q(x)^2 dx}, \quad q'' = tq + 2q^3, \; q(x) \sim_{x\to\infty} \mathrm{Ai}(x).$$

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives a hint on convergence speed
- fairly biased on the left: even fewer eigenvalues outside the support.

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives a hint on convergence speed
- fairly biased on the left: even fewer eigenvalues outside the support.
- can be shown to hold for other distributions than Gaussian under mild assumptions

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives a hint on convergence speed
- fairly biased on the left: even fewer eigenvalues outside the support.
- can be shown to hold for other distributions than Gaussian under mild assumptions
- Now, what about largest eigenvalue of a spiked model?

Related bibliography

- I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," Annals of Statistics, vol. 99, no. 2, pp. 295-327, 2001.
- K. Johansson, "Shape Fluctuations and Random Matrices," Comm. Math. Phys. vol. 209, pp. 437-476, 2000.
- C. A. Tracy and H. Widom, "On orthogonal and symplectic matrix ensembles," Communications in Mathematical Physics, vol. 177, no. 3, pp. 727-754, 1996.
- G. W. Anderson, A. Guionnet, O. Zeitouni, "An introduction to random matrices", Cambridge studies in advanced mathematics, vol. 118, 2010.
- F. Bornemann, "On the numerical evaluation of distributions in random matrix theory: A review," Markov Process. Relat. Fields, vol. 16, pp. 803-866, 2010.
- Y. Q. Yin, Z. D. Bai, P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix," Probability Theory and Related Fields, vol. 78, no. 4, pp. 509-521, 1988.
- J. W. Silverstein, Z.D. Bai and Y.Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," Journal of Multivariate Analysis, vol. 26, no. 2, pp. 166-168. 1988.
- C. A. Tracy, H. Widom, "On orthogonal and symplectic matrix ensembles," Communications in Mathematical Physics, vol. 177, no. 3, pp. 727-754, 1996.
- J. W. Silverstein, S. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices," Journal of Multivariate Analysis, vol. 54, no. 2, pp. 295-309, 1995.
- Z. D. Bai, J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," The Annals of Probability, vol. 26, no.1 pp. 316-345, 1998.
- Z. D. Bai, J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," The Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.
- J. W. Silverstein, P. Debashis, "No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix," J. of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

Outline

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \ldots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.

Introduction of the problem

- *Reminder:* for a sequence $\mathbf{x}_1, \dots, x_n \in \mathbb{C}^N$ of independent random variables,

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}$$

  is an n-consistent estimator of $\mathbf{R} = E[\mathbf{x}_1 \mathbf{x}_1^{\mathrm{H}}]$.
- If n, N have comparable sizes, this no longer holds.
- Typically, n, N-consistent estimators of the full $\mathbf{R}$ matrix perform very badly.
- If only the eigenvalues of $\mathbf{R}$ are of interest, things can be done. The process of retrieving information about eigenvalues, eigenspace projections, or functional of these is called eigen-inference.

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis,"
http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called G-estimators (Generalized estimators). Among those, we find
  - *G_1-estimator of generalized variance.* For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1}\left[\log\det(\mathbf{R}_n) + \log \frac{n(n-1)^N}{(n-N)\prod_{k=1}^{N}(n-k)}\right]$$

with α_n any sequence such that $\alpha_n^{-2}\log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1}\log\det(\mathbf{R}) \to 0$$

in probability.

Girko and the G-estimators

V. Girko, "Ten years of general statistical analysis,"
http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf

- Girko has come up with more than 50 N, n-consistent estimators, called G-estimators (Generalized estimators). Among those, we find
  - *G_1-estimator of generalized variance.* For

$$G_1(\mathbf{R}_n) = \alpha_n^{-1}\left[\log\det(\mathbf{R}_n) + \log\frac{n(n-1)^N}{(n-N)\prod_{k=1}^N (n-k)}\right]$$

with α_n any sequence such that $\alpha_n^{-2}\log(n/(n-N)) \to 0$, we have

$$G_1(\mathbf{R}_n) - \alpha_n^{-1}\log\det(\mathbf{R}) \to 0$$

in probability.

- However, Girko's proofs are rarely readable, if existent.

Outline

Position of the problem

- it has long been difficult to analytically invert the simplest $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N^{\frac{1}{2}}$ model to recover the diagonal entries of $\mathbf{T}_N$. Indeed, we only have the deterministic equivalent result

$$\underline{m}_N(z) = \left(-z + c\int \frac{t}{1+t\underline{m}_N(z)}dF^{\mathbf{T}_N}(t)\right)^{-1}$$

with $\underline{m}_N$ the deterministic equivalent of the Stieltjes transform for $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N\mathbf{X}_N$.

Position of the problem

- it has long been difficult to analytically invert the simplest $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$ model to recover the diagonal entries of $\mathbf{T}_N$. Indeed, we only have the deterministic equivalent result

$$\underline{m}_N(z) = \left(-z + c \int \frac{t}{1 + t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t) \right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of the Stieltjes transform for $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N$.
- when $\mathbf{T}_N$ has eigenvalues $t_1, \ldots, t_K$ with multiplicity $n_1, \ldots, n_K$, this is

$$\underline{m}_N(z) = \left(-z + \frac{1}{N} \sum_{k=1}^{K} n_k \frac{t_k}{1 + t_k \underline{m}_N(z)} \right)^{-1}$$

- an N, n-consistent estimator for the t_k's was never found until recently...

Position of the problem

- it has long been difficult to analytically invert the simplest $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ model to recover the diagonal entries of $\mathbf{T}_N$. Indeed, we only have the deterministic equivalent result

$$\underline{m}_N(z) = \left(-z + c\int \frac{t}{1+t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t)\right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of the Stieltjes transform for $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$.
- when $\mathbf{T}_N$ has eigenvalues $t_1,\ldots,t_K$ with multiplicity $n_1,\ldots,n_K$, this is

$$\underline{m}_N(z) = \left(-z + \frac{1}{N}\sum_{k=1}^{K} n_k \frac{t_k}{1+t_k\underline{m}_N(z)}\right)^{-1}$$

- an N,n-consistent estimator for the t_k's was never found until recently...
- however, moment-based methods and free probability approaches provide simple solutions to estimate consistently all moments of $F^{\mathbf{T}_N}$.

Reminder on moment-based approaches

- For free random matrices **A** and **B**, we have the cumulant/moment relationships,

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1 \\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A})C_{|V_2|}(\mathbf{B})$$

- this allows one to compute all moments of sum and product distributions

$$\mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$$

- in addition, we have results for the information-plus-noise model

$$\mathbf{B}_N = \frac{1}{n}(\mathbf{R}_N + \sigma\mathbf{X}_N)(\mathbf{R}_N + \sigma\mathbf{X}_N)^{\mathsf{H}}$$

  whose e.s.d. converges weakly and almost surely to μ_B such that

$$\mu_B = \left((\mu_\Gamma \boxslash \mu_c) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $\Gamma_N = \mathbf{R}_N\mathbf{R}_N^{\mathsf{H}}$.

Reminder on moment-based approaches

- For free random matrices **A** and **B**, we have the cumulant/moment relationships,

$$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$$

$$M_n(\mathbf{AB}) = \sum_{(\pi_1,\pi_2)\in NC(n)} \prod_{\substack{V_1\in\pi_1 \\ V_2\in\pi_2}} C_{|V_1|}(\mathbf{A}) C_{|V_2|}(\mathbf{B})$$

- this allows one to compute all moments of sum and product distributions

$$\mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$$

- in addition, we have results for the information-plus-noise model

$$\mathbf{B}_N = \frac{1}{n}(\mathbf{R}_N + \sigma\mathbf{X}_N)(\mathbf{R}_N + \sigma\mathbf{X}_N)^{\mathrm{H}}$$

  whose e.s.d. converges weakly and almost surely to μ_B such that

$$\mu_B = \left((\mu_\Gamma \boxslash \mu_c) \boxplus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with μ_c the Marčenko-Pastur law and $\Gamma_N = \mathbf{R}_N\mathbf{R}_N^{\mathrm{H}}$.
- all basic matrix operations needed in wireless communications are accessible for convenient matrices (Gaussian, Vandermonde etc.)
- all operations are merely polynomial operations on the moments. As a consequence, for $\mathbf{B}_N = f(\mathbf{R}_N)$,

From free convolution to free deconvolution

Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.

- we have the further result that

Polynomial Relations

The k^{th} moment of the l.s.d. of $\mathbf{B}_N$ is a polynomial of the k-first moments of the l.s.d. of $\mathbf{R}_N$

From free convolution to free deconvolution

Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.

- we have the further result that

Polynomial Relations

The k^{th} moment of the l.s.d. of $\mathbf{B}_N$ is a polynomial of the k-first moments of the l.s.d. of $\mathbf{R}_N$

- we can therefore invert the problem and express the k^{th} moment of $\mathbf{R}_N$ as the first k moments of $\mathbf{B}_N$. This entails deconvolution operations,

$$\mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} = \mu_{\mathbf{AB}} \boxslash \mu_{\mathbf{B}}$$

and for the information-plus-noise model, $\mathbf{B}_N = \frac{1}{n} (\mathbf{R}_N + \sigma \mathbf{X}_N) (\mathbf{R}_N + \sigma \mathbf{X}_N)^{\mathsf{H}}$

$$\mu_\Gamma = \left((\mu_B \boxtimes \mu_c) \boxminus \delta_{\sigma^2} \right) \boxslash \mu_c$$

From free convolution to free deconvolution

Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv preprint math.PR/0702342, 2007.

- we have the further result that

Polynomial Relations

The k^{th} moment of the l.s.d. of $\mathbf{B}_N$ is a polynomial of the k-first moments of the l.s.d. of $\mathbf{R}_N$

- we can therefore invert the problem and express the k^{th} moment of $\mathbf{R}_N$ as the first k moments of $\mathbf{B}_N$. This entails deconvolution operations,

$$\mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}}$$

$$\mu_{\mathbf{A}} = \mu_{\mathbf{AB}} \boxslash \mu_{\mathbf{B}}$$

and for the information-plus-noise model, $\mathbf{B}_N = \frac{1}{n}(\mathbf{R}_N + \sigma\mathbf{X}_N)(\mathbf{R}_N + \sigma\mathbf{X}_N)^{\mathsf{H}}$

$$\mu_{\Gamma} = \left((\mu_B \boxtimes \mu_c) \boxminus \delta_{\sigma^2}\right) \boxslash \mu_c$$

- for more involved models, the polynomial relations can be iterated and even automatically generated.

Example of polynomial relation

- Consider the information-plus-noise model

$$\mathbf{Y} = \mathbf{D} + \mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$, $\mathbf{D} \in \mathbb{C}^{N \times n}$, $\mathbf{X} \in \mathbb{C}^{N \times n}$ with i.i.d. entries of mean 0 and variance 1. Denote

$$M_k = \lim_{n \to \infty} \frac{1}{n} \mathrm{tr}\left(\frac{1}{N} \mathbf{Y}\mathbf{Y}^{\mathrm{H}} \right)^k$$

$$D_k = \lim_{n \to \infty} \frac{1}{n} \mathrm{tr}\left(\frac{1}{N} \mathbf{D}\mathbf{D}^{\mathrm{H}} \right)^k$$

Example of polynomial relation

- Consider the information-plus-noise model

$$\mathbf{Y} = \mathbf{D} + \mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$, $\mathbf{D} \in \mathbb{C}^{N \times n}$, $\mathbf{X} \in \mathbb{C}^{N \times n}$ with i.i.d. entries of mean 0 and variance 1. Denote

$$M_k = \lim_{n \to \infty} \frac{1}{n} \mathrm{tr} \left(\frac{1}{N} \mathbf{Y}\mathbf{Y}^{\mathrm{H}} \right)^k$$

$$D_k = \lim_{n \to \infty} \frac{1}{n} \mathrm{tr} \left(\frac{1}{N} \mathbf{D}\mathbf{D}^{\mathrm{H}} \right)^k$$

- For that model, we have the relations

$$M_1 = D_1 + 1$$
$$M_2 = D_2 + (2 + 2c) D_1 + (1 + c)$$
$$M_3 = D_3 + (3 + 3c) D_2 + 3c {D_1}^2 + (1 + 3c + c^2)$$

Example of polynomial relation

- Consider the information-plus-noise model

$$\mathbf{Y} = \mathbf{D} + \mathbf{X}$$

with $\mathbf{Y} \in \mathbb{C}^{N \times n}$, $\mathbf{D} \in \mathbb{C}^{N \times n}$, $\mathbf{X} \in \mathbb{C}^{N \times n}$ with i.i.d. entries of mean 0 and variance 1. Denote

$$M_k = \lim_{n \to \infty} \frac{1}{n} \mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\right)^k$$

$$D_k = \lim_{n \to \infty} \frac{1}{n} \mathrm{tr}\left(\frac{1}{N}\mathbf{D}\mathbf{D}^{\mathsf{H}}\right)^k$$

- For that model, we have the relations

$$\begin{aligned}
M_1 &= D_1 + 1 \\
M_2 &= D_2 + (2 + 2c)D_1 + (1 + c) \\
M_3 &= D_3 + (3 + 3c)D_2 + 3c{D_1}^2 + (1 + 3c + c^2)
\end{aligned}$$

hence

$$\begin{aligned}
D_1 &= M_1 - 1 \\
D_2 &= M_2 - (2 + 2c)M_1 + (1 + c) \\
D_3 &= M_3 - (3 + 3c)M_2 - 3c{M_1}^2 + (6c^2 + 18c + 6)M_1 - (4c^2 + 12c + 4)
\end{aligned}$$

Finite size statistical inference

A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Finite Dimensional Statistical Inference," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2457-2473, 2011.

- it might happen that, instead of one large matrix realization, we have access to several smaller such matrices. In that case, we seek an estimate for

$$E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\right)^{k}\right]$$

  instead of their limits.

Finite size statistical inference

A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Finite Dimensional Statistical Inference," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2457-2473, 2011.

- it might happen that, instead of one large matrix realization, we have access to several smaller such matrices. In that case, we seek an estimate for

$$E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k\right]$$

  instead of their limits.
- we have further combinatorics theorems for all previous elementary models.

Finite size statistical inference

A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Finite Dimensional Statistical Inference," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2457-2473, 2011.

- it might happen that, instead of one large matrix realization, we have access to several smaller such matrices. In that case, we seek an estimate for

$$E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k\right]$$

  instead of their limits.
- we have further combinatorics theorems for all previous elementary models.
- *example:* the previous relations extend to

$$\begin{aligned}
M_1 &= D_1 + 1\\
M_2 &= D_2 + (2+2c)D_1 + (1+c)\\
M_3 &= D_3 + (3+3c)D_2 + 3cD_1^2 + (3+9c+3c^2+3N^{-2})D_1 + (1+3c+c^2+N^{-2})
\end{aligned}$$

Current and further studies

- in addition to estimating the average moments themselves, we can evaluate the variance of the empirical moments

$$E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k - E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k\right]\right]$$

Current and further studies

- in addition to estimating the average moments themselves, we can evaluate the variance of the empirical moments

$$E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k} - E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^{k}\right]\right]$$

- if the moments have Gaussian distributions (left to be proven for models other than sample covariance matrix), the full behaviour of the empirical moments is known.

Current and further studies

- in addition to estimating the average moments themselves, we can evaluate the variance of the empirical moments

$$E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{YY}^{\mathrm{H}}\right)^{k} - E\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{N}\mathbf{YY}^{\mathrm{H}}\right)^{k}\right]\right]$$

- if the moments have Gaussian distributions (left to be proven for models other than sample covariance matrix), the full behaviour of the empirical moments is known.
- statistical maximum-likelihood/MMSE methods can then be used.

Related bibliography

- N. R. Rao, A. Edelman, "The polynomial method for random matrices," Foundations of Computational Mathematics, *accepted for publication*.
- N. R. Rao, J. A. Mingo, R. Speicher, A. Edelman, "Statistical eigen-inference from large Wishart matrices," Annals of Statistics, vol. 36, no. 6, pp. 2850-2885, 2008.
- A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Finite Dimensional Statistical Inference," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2457-2473, 2011.
- Ø. Ryan, M. Debbah, "Multiplicative free convolution and information-plus-noise type matrices," Arxiv math.PR/0702342, 2007.
- Ø. Ryan, M. Debbah, "Free deconvolution for signal processing applications," IEEE International Symposium on Information Theory, pp. 1846-1850, 2007.
- Ø. Ryan, M. Debbah, "Asymptotic Behavior of Random Vandermonde Matrices With Entries on the Unit Circle," IEEE Trans. on Information Theory, vol. 55, no. 7, pp. 3115-3147, 2009.

Outline

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- It has long been thought the inverse problem of estimating $t_1, \ldots, t_K$ from the Stieltjes transform method was not possible.
- Only trials were iterative convex optimization methods.

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- It has long been thought the inverse problem of estimating $t_1, \ldots, t_K$ from the Stieltjes transform method was not possible.
- Only trials were iterative convex optimization methods.
- The problem was partially solved by Mestre in 2008!

A long standing problem

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

- Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N}$ is formed of a finite number of masses $t_1, \ldots, t_K$.
- It has long been thought the inverse problem of estimating $t_1, \ldots, t_K$ from the Stieltjes transform method was not possible.
- Only trials were iterative convex optimization methods.
- The problem was partially solved by Mestre in 2008!
- His technique uses elegant complex analysis tools. The description of this technique is the subject of this course.

Reminders

- Consider the sample covariance matrix model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$.
- Up to now, we saw:
  - that there is no eigenvalue outside the support with probability 1 for all large N.
  - that for all large N, when the spectrum is divided into clusters, the number of empirical eigenvalues in each cluster is exactly as we expect.

Reminders

- Consider the sample covariance matrix model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$.
- Up to now, we saw:
  - that there is no eigenvalue outside the support with probability 1 for all large N.
  - that for all large N, when the spectrum is divided into clusters, the number of empirical eigenvalues in each cluster is exactly as we expect.
- these results are of crucial importance for the following.

Inverse problem for sample covariance matrix

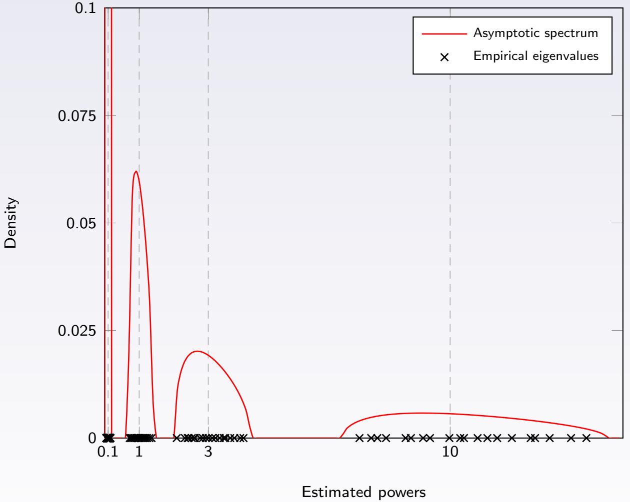


Figure: Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $N/n = 10$, $M/N = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

Eigen-inference for the sample covariance matrix model

X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.

Theorem

Consider the model $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$*, with* $\mathbf{X}_N \in \mathbb{C}^{N \times n}$*, i.i.d. with entries of zero mean, variance* $1/n$*, and* $\mathbf{T}_N \in \mathbb{R}^{N \times N}$ *is diagonal with* K *distinct entries* $t_1, \ldots, t_K$ *of multiplicity* $N_1, \ldots, N_K$ *of same order as* n*. Let* $k \in \{1, \ldots, K\}$*. Then, if the cluster associated to* t_k *is separated from the clusters associated to* $k-1$ *and* $k+1$*, as* $N, n \to \infty$*,* $N/n \to c$*,*

$$\hat{t}_k = \frac{n}{N_k} \sum_{m \in \mathcal{N}_k} (\lambda_m - \mu_m)$$

is an N, n*-consistent estimator of* t_k*, where* $\mathcal{N}_k = \{N - \sum_{i=k}^K N_i + 1, \ldots, N - \sum_{i=k+1}^K N_i\}$*,* $\lambda_1, \ldots, \lambda_N$ *are the eigenvalues of* $\mathbf{B}_N$ *and* $\mu_1, \ldots, \mu_N$ *are the* N *solutions of*

$$m_{\mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N}(\mu) = 0$$

or equivalently, $\mu_1, \ldots, \mu_N$ *are the eigenvalues of* $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N} \sqrt{\boldsymbol{\lambda}} \sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$*.*

A trick to compute the μ_k's

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources", IEEE Transactions on Information Theory, vol. 57, no. 4, pp. 2420-2439, 2011.

Lemma

Let $\mathbf{A} \in \mathbb{C}^{n \times N}$ be diagonal with entries $\lambda_1, \ldots, \lambda_N$ and $\mathbf{y} \in \mathbb{C}^N$. Then the eigenvalues of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ are the N real solutions in x of

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - x} = 1$$

A trick to compute the μ_k's

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources", IEEE Transactions on Information Theory, vol. 57, no. 4, pp. 2420-2439, 2011.

Lemma

Let $\mathbf{A} \in \mathbb{C}^{n\times N}$ be diagonal with entries $\lambda_1, \ldots, \lambda_N$ and $\mathbf{y} \in \mathbb{C}^N$. Then the eigenvalues of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ are the N real solutions in x of

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - x} = 1$$

Taking $\mathbf{A} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $y_i^2 = \frac{1}{n}\lambda_i$, the eigenvalues of $\mathbf{A} - \mathbf{y}\mathbf{y}^{\mathrm{H}}$ are the solutions of

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i - x} = 1$$

which is equivalent to

$$m_{\mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N\mathbf{X}_N}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i - x} = 0$$

A trick to compute the μ_k's

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources", IEEE Transactions on Information Theory, vol. 57, no. 4, pp. 2420-2439, 2011.

Lemma

Let $\mathbf{A} \in \mathbb{C}^{n\times N}$ be diagonal with entries $\lambda_1, \ldots, \lambda_N$ and $\mathbf{y} \in \mathbb{C}^N$. Then the eigenvalues of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ are the N real solutions in x of

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - x} = 1$$

Taking $\mathbf{A} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $y_i^2 = \frac{1}{n}\lambda_i$, the eigenvalues of $\mathbf{A} - \mathbf{y}\mathbf{y}^{\mathsf{H}}$ are the solutions of

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i - x} = 1$$

which is equivalent to

$$m_{\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i - x} = 0$$

The μ_k's are then the eigenvalues of a matrix that is function of $\lambda_1, \ldots, \lambda_N$.

Proof of the lemma

Let $\mathbf{A} \in \mathbb{C}^{N\times N}$ be Hermitian and $\mathbf{y} \in \mathbb{C}^N$. If μ is an eigenvalue of $(\mathbf{A} - \mathbf{y}\mathbf{y}^*)$ with eigenvector $\mathbf{x}$, we have

$$\begin{aligned}
(\mathbf{A} - \mathbf{y}\mathbf{y}^*)\mathbf{x} &= \mu\mathbf{x} \\
(\mathbf{A} - \mu I)x &= \mathbf{y}^* x \mathbf{y} \\
x &= \mathbf{y}^* x (\mathbf{A} - \mu I)^{-1}\mathbf{y} \\
\mathbf{y}^* x &= \mathbf{y}^* x \mathbf{y}^* (\mathbf{A} - \mu I)^{-1}\mathbf{y} \\
1 &= \mathbf{y}^* (\mathbf{A} - \mu I)^{-1}\mathbf{y}
\end{aligned}$$

Take $\mathbf{A}$ diagonal with entries $\lambda_1, \ldots, \lambda_N$, we then have

$$\sum_{i=1}^{N} \frac{y_i^2}{\lambda_i - \mu} = 1 \quad (1)$$

Remarks on Mestre's result

Assuming cluster separation, the result consists in

- taking the empirical *ordered* λ_i's inside the cluster (note that exact separation ensures there are N_k of these!)
- getting the *ordered* eigenvalues $\mu_1, \ldots, \mu_N$ of

$$\operatorname{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$$

  with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$. Keep only those of index inside $\mathcal{N}_k$.
- take the difference and scale.

How to obtain this result?

- Major trick requires tools from complex analysis

How to obtain this result?

- Major trick requires tools from complex analysis
- Silverstein's Stieltjes transform identity: for the *conjugate* model $\underline{\mathbf{B}}_N = \mathbf{X}_N^* \mathbf{T}_N \mathbf{X}_N$,

$$\underline{m}_N(z) = \left(-z - c \int \frac{t}{1 + t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t) \right)^{-1}$$

with $\underline{m}_N$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}$. This is the only random matrix result we need.

How to obtain this result?

- Major trick requires tools from complex analysis
- Silverstein's Stieltjes transform identity: for the *conjugate* model $\underline{\mathbf{B}}_N = \mathbf{X}_N^* \mathbf{T}_N \mathbf{X}_N$,

$$\underline{m}_N(z) = \left(-z - c \int \frac{t}{1 + t\underline{m}_N(z)} dF^{\mathbf{T}_N}(t) \right)^{-1}$$

  with $\underline{m}_N$ the deterministic equivalent of $m_{\underline{\mathbf{B}}_N}$. This is the only random matrix result we need.
- Before going further, we need some reminders from complex analysis.

Reminders of complex analysis

- Cauchy integration formula

Theorem

Let $U \subset \mathbb{C}$ be an open set and $f : U \to \mathbb{C}$ be holomorphic on U. Let $\gamma \subset U$ be a continuous contour (i.e. closed path). Then, for a **inside** *the surface formed by γ, we have*

$$\frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z-a} dz = f(a)$$

while for a **outside** *the surface formed by γ,*

$$\frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z-a} dz = 0.$$

Limiting spectrum of the sample covariance matrix

J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," J. of Multivariate Analysis, vol. 54, no. 2, pp. 175-192, 1995.

Reminder:

- If $F^{\mathbf{T}_N} \Rightarrow F^T$, then $m_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} m_F(z)$ such that

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1}$$

Limiting spectrum of the sample covariance matrix

J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," J. of Multivariate Analysis, vol. 54, no. 2, pp. 175-192, 1995.

Reminder:

- If $F^{\mathbf{T}_N} \Rightarrow F^T$, then $m_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} m_F(z)$ such that

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1}$$

  or equivalently

$$m_{F^T}\left(-1/m_{\underline{F}}(z)\right) = -z m_{\underline{F}}(z) m_F(z)$$

  with $m_{\underline{F}}(z) = c m_F(z) + (c-1)\frac{1}{z}$ and $N/n \to c$.

Reminders of complex analysis (2)

- Residue calculus

Theorem

Let γ be a contour on $\mathbb{C}$. For f holomorphic inside γ but on a discrete number of points, to compute the expression

$$\frac{1}{2\pi i}\oint_\gamma f(z)dz$$

one must

1. *determine the poles of f lying inside the surface formed by γ, i.e. those values a such that*
$$\lim_{z\to a}|f(z)|=\infty$$
2. *determine the order of each pole, i.e. the smallest k such that*
$$\lim_{z\to a}|(z-a)^k f(z)|<\infty$$
3. *compute the residues of f at the poles, i.e. evaluate the value*
$$\mathrm{Res}(f,a)\triangleq\lim_{z\to a}\frac{d^{k-1}}{dz^{k-1}}\left[(z-a)^k f(z)\right]$$
4. *the integral is then the sum of all residues.*
$$\frac{1}{2\pi i}\oint_\gamma f(z)dz=\sum_{a\in\{\text{ poles of } f\}}\mathrm{Res}(f,a)$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

- After the variable change $\omega = -1/m_{\underline{F}}(z)$,

$$t_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_F(z) \frac{m'_{\underline{F}}(z)}{m^2_{\underline{F}}(z)} dz,$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

- After the variable change $\omega = -1/m_{\underline{F}}(z)$,

$$t_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_F(z) \frac{m'_{\underline{F}}(z)}{m^2_{\underline{F}}(z)} dz,$$

- When the system dimensions are large,

$$m_F(z) \simeq m_{\mathbf{B}_N}(z) \triangleq \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k - z}, \quad \text{with} \quad (\lambda_1, \dots, \lambda_N) = \mathrm{eig}(\mathbf{B}_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}).$$

Complex integration

- From Cauchy integral formula, denoting $\mathcal{C}_k$ a contour enclosing **only** t_k,

$$t_k = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{\omega - t_k} d\omega = \frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N_k}\sum_{j=1}^{K} N_j \frac{\omega}{\omega - t_j} d\omega = \frac{N}{2\pi i N_k}\oint_{\mathcal{C}_k} \omega m_T(\omega) d\omega.$$

- After the variable change $\omega = -1/m_{\underline{F}}(z)$,

$$t_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_F(z) \frac{m'_{\underline{F}}(z)}{m^2_{\underline{F}}(z)} dz,$$

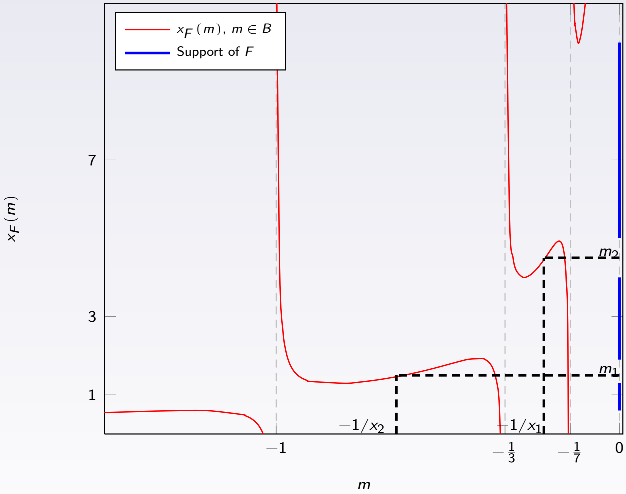
- When the system dimensions are large,

$$m_F(z) \simeq m_{\mathbf{B}_N}(z) \triangleq \frac{1}{N}\sum_{k=1}^{N} \frac{1}{\lambda_k - z}, \quad \text{with} \quad (\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{B}_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}).$$

- Dominated convergence arguments then show

$$t_k - \hat{t}_k \xrightarrow{\text{a.s.}} 0 \quad \text{with} \quad \hat{t}_k = \frac{N}{N_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_{\underline{F},k}} z m_{\mathbf{B}_N}(z) \frac{m'_{\underline{\mathbf{B}}_N}(z)}{m^2_{\underline{\mathbf{B}}_N}(z)} dz$$

Understanding the contour change



- **IF** $\mathcal{C}_{\underline{F},k}$ encloses cluster k with real points $m_1 < m_2$
- **THEN** $-1/m_1 = x_1 < t_k < x_2 = -1/m_2$ and $\mathcal{C}_k$ encloses t_k.

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.
- remember that

$$m_{\mathbf{B}_N}(w) = \frac{n}{N} m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\frac{1}{w}$$

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \ldots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \ldots, \mu_N$ to $\underline{\hat{m}}_N(z) = 0$.
- remember that

$$m_{\mathbf{B}_N}(w) = \frac{n}{N} m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\frac{1}{w}$$

- residue calculus, denote $f(w) = \left(\frac{n}{N} w m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\right) \frac{m'_{\underline{\mathbf{B}}_N}(w)}{m_{\underline{\mathbf{B}}_N}(w)^2}$,
  - the λ_k's are poles of order 1 and

$$\lim_{z \to \lambda_k} (z - \lambda_k) f(z) = -\frac{n}{N} \lambda_k$$

  - the μ_k's are also poles of order 1 and by L'Hospital's rule

$$\lim_{z \to \mu_k} (z - \lambda_k) f(z) = \lim_{z \to \mu_k} \frac{n}{N} \frac{(z - \mu_k) z m'_{\underline{\mathbf{B}}_N}(z)}{m_{\underline{\mathbf{B}}_N}(z)} = \frac{n}{N} \mu_k$$

Poles and residues

- we find two sets of poles (outside zeros):
  - $\lambda_1, \dots, \lambda_N$, the eigenvalues of $\mathbf{B}_N$.
  - the solutions $\mu_1, \dots, \mu_N$ to $\hat{\underline{m}}_N(z) = 0$.
- remember that

$$m_{\mathbf{B}_N}(w) = \frac{n}{N} m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\frac{1}{w}$$

- residue calculus, denote $f(w) = \left(\frac{n}{N} w m_{\underline{\mathbf{B}}_N}(w) + \frac{n-N}{N}\right) \frac{m'_{\underline{\mathbf{B}}_N}(w)}{m_{\underline{\mathbf{B}}_N}(w)^2}$,
  - the λ_k's are poles of order 1 and

$$\lim_{z\to\lambda_k} (z-\lambda_k) f(z) = -\frac{n}{N}\lambda_k$$

  - the μ_k's are also poles of order 1 and by L'Hospital's rule

$$\lim_{z\to\mu_k} (z-\lambda_k) f(z) = \lim_{z\to\mu_k} \frac{n}{N} \frac{(z-\mu_k) z m'_{\underline{\mathbf{B}}_N}(z)}{m_{\underline{\mathbf{B}}_N}(z)} = \frac{n}{N}\mu_k$$

  - So, finally

$$\hat{t}_k = \frac{n}{N_k} \sum_{m\in\text{contour}} (\lambda_m - \mu_m)$$

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.
- Since the μ_i are rank-1 perturbations of the λ_i, they have the interleaving property

$$\lambda_1 < \mu_2 < \lambda_2 < \ldots < \mu_N < \lambda_N$$

Which poles in the contour?

- we now need to determine which poles are in the contour of interest.
- Since the μ_i are rank-1 perturbations of the λ_i, they have the interleaving property

$$\lambda_1 < \mu_2 < \lambda_2 < \ldots < \mu_N < \lambda_N$$

- what about μ_1? the trick is to use the fact that

$$\frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{z} dz = 0$$

which leads to

$$\frac{1}{2\pi i} \oint_{\partial \Gamma_k} \frac{m'_{\underline{F}}(w)}{m_{\underline{F}}(w)^2} dw = 0$$

the empirical version of which is

$$\#\{i : \lambda_i \in \Gamma_k\} - \#\{i : \mu_i \in \Gamma_k\}$$

Since their difference tends to 0, there are as many λ_k's as μ_k's in the contour, hence μ_1 is asymptotically in the integration contour.

Related bibliography

- X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," IEEE Transactions on Signal Processing, vol. 56, no.11, 2008.
- X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," IEEE trans. on Information Theory, vol. 54, no. 11, pp. 5113-5129, 2008.
- R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources", IEEE Transactions on Information Theory, vol. 57, no. 4, pp. 2420-2439, 2011.
- P. Vallet, P. Loubaton and X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case," arxiv preprint 1002.3234, 2010.

Outline

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. H of $\mathbf{T}_N$ is a Dirac in 1.

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. H of $\mathbf{T}_N$ is a Dirac in 1.
- No contradiction with "no eigenvalue" theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additional clusters of positive measure in F_N.

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. H of $\mathbf{T}_N$ is a Dirac in 1.
- No contradiction with "no eigenvalue" theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additional clusters of positive measure in F_N.
- However, for practical purposes, the presence of "spikes" determines the presence of a signal!

Spiked models

- We can create sample covariance matrix models $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with l.s.d. F ($\mathbf{X}_N$ as usual) for which
  - some sample eigenvalues are found outside the support of F
  - the l.s.d. H of $\mathbf{T}_N$ is a Dirac in 1.
- No contradiction with "no eigenvalue" theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additional clusters of positive measure in F_N.
- However, for practical purposes, the presence of "spikes" determines the presence of a signal!

What about the absence of spikes?

Absence of spikes $\overset{?}{\Rightarrow}$ No signal

J. Baik, J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," Journal of Multivariate Analysis, vol. 97, no. 6, pp. 1382-1408, 2006.

Theorem

Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$*, where* $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ *has i.i.d., zero mean and variance* $1/n$ *entries, and* $\mathbf{T}_N \in \mathbb{R}^{N \times N}$ *diagonal given by*

$$\mathbf{T}_N = diag(\underbrace{1+\omega_1, \ldots, 1+\omega_1}_{k_1}, \ldots, \underbrace{1+\omega_M, \ldots, 1+\omega_M}_{k_M}, \underbrace{1, \ldots, 1}_{N-\sum_{i=1}^M k_i})$$

with $\omega_1 > \ldots > \omega_M > -1$*,* $c = \lim_N N/n$*. We then have*

- *if* $\omega_j > \sqrt{c}$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_j + c\frac{1+\omega_j}{\omega_j}$
- *if* $\omega_{k_j} \in (0, \sqrt{c}]$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$
- *if* $\omega_{k_j} \in [-\sqrt{c}, 0)$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$

Absence of spikes $\overset{?}{\Rightarrow}$ No signal

J. Baik, J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," Journal of Multivariate Analysis, vol. 97, no. 6, pp. 1382-1408, 2006.

Theorem

Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, *where* $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ *has i.i.d., zero mean and variance* $1/n$ *entries, and* $\mathbf{T}_N \in \mathbb{R}^{N \times N}$ *diagonal given by*

$$\mathbf{T}_N = diag(\underbrace{1+\omega_1, \ldots, 1+\omega_1}_{k_1}, \ldots, \underbrace{1+\omega_M, \ldots, 1+\omega_M}_{k_M}, \underbrace{1, \ldots, 1}_{N-\sum_{i=1}^M k_i})$$

with $\omega_1 > \ldots > \omega_M > -1$, $c = \lim_N N/n$. *We then have*

- *if* $\omega_j > \sqrt{c}$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_j + c\frac{1+\omega_j}{\omega_j}$
- *if* $\omega_{k_j} \in (0, \sqrt{c}]$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$
- *if* $\omega_{k_j} \in [-\sqrt{c}, 0)$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$
- *for the other eigenvalues, we discriminate over c:*
  - *if* $\omega_{k_j} < -\sqrt{c}$, $c < 1$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_j + c\frac{1+\omega_j}{\omega_j}$
  - *if* $\omega_{k_j} < -\sqrt{c}$, $c > 1$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$

Absence of spikes $\overset{?}{\Rightarrow}$ No signal

J. Baik, J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," Journal of Multivariate Analysis, vol. 97, no. 6, pp. 1382-1408, 2006.

Theorem

Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, *where* $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ *has i.i.d., zero mean and variance* $1/n$ *entries, and* $\mathbf{T}_N \in \mathbb{R}^{N \times N}$ *diagonal given by*

$$\mathbf{T}_N = diag(\underbrace{1+\omega_1, \ldots, 1+\omega_1}_{k_1}, \ldots, \underbrace{1+\omega_M, \ldots, 1+\omega_M}_{k_M}, \underbrace{1, \ldots, 1}_{N - \sum_{i=1}^{M} k_i})$$

with $\omega_1 > \ldots > \omega_M > -1$, $c = \lim_N N/n$. *We then have*

- *if* $\omega_j > \sqrt{c}$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_j + c\frac{1+\omega_j}{\omega_j}$
- *if* $\omega_{k_j} \in (0, \sqrt{c}]$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$
- *if* $\omega_{k_j} \in [-\sqrt{c}, 0)$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$
- *for the other eigenvalues, we discriminate over* c*:*
  - *if* $\omega_{k_j} < -\sqrt{c}$, $c < 1$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} 1 + \omega_j + c\frac{1+\omega_j}{\omega_j}$
  - *if* $\omega_{k_j} < -\sqrt{c}$, $c > 1$, $\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$

Proof: See Section "Research Today: Advanced Statistic Inference"

Eigenvalues outside the support

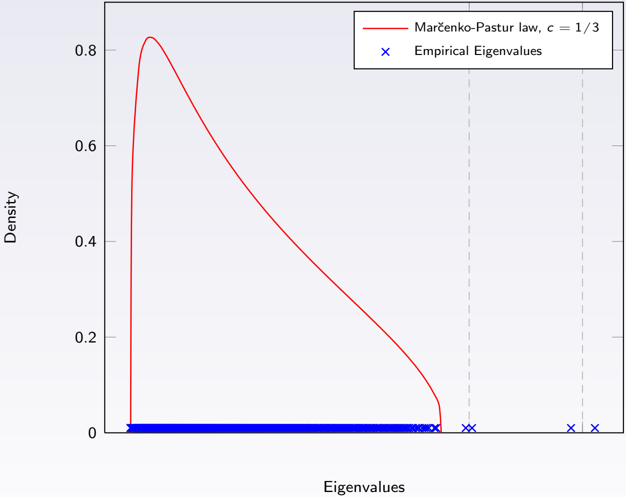


Figure: Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N} \Rightarrow 1_{[1,\infty)}$, ...Dimensions: $N = 500$, $n = 1500$.

Eigenvalues outside the support

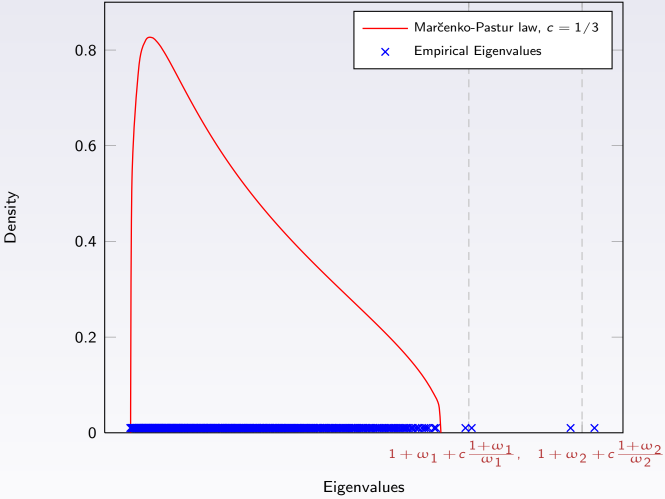


Figure: Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N} \Rightarrow 1_{[1,\infty)}$, but $\mathbf{T}_N$ is a diagonal of ones but for the first four entries set to $\{1+\omega_1, 1+\omega_1, 1+\omega_2, 1+\omega_2\}$, $\omega_1 = 1$, $\omega_2 = 2$.Dimensions: $N = 500$, $n = 1500$.

Interpretation of the result

- if c is large, or alternatively, if some "population spikes" are small, part to all of the population spikes are attracted by the support!

Interpretation of the result

- if c is large, or alternatively, if some "population spikes" are small, part to all of the population spikes are attracted by the support!
- if so, no way to decide on the existence of the spikes *from looking at the largest eigenvalues*
- in telecommunication words, signals might be missed using largest eigenvalues methods.

Interpretation of the result

- if c is large, or alternatively, if some "population spikes" are small, part to all of the population spikes are attracted by the support!
- if so, no way to decide on the existence of the spikes *from looking at the largest eigenvalues*
- in telecommunication words, signals might be missed using largest eigenvalues methods.
- as a consequence,
  - the more the sensors (N),
  - the larger $c = \lim N/n$,
  - the more probable we miss a spike

Interpretation of the result

- if c is large, or alternatively, if some "population spikes" are small, part to all of the population spikes are attracted by the support!
- if so, no way to decide on the existence of the spikes *from looking at the largest eigenvalues*
- in telecommunication words, signals might be missed using largest eigenvalues methods.
- as a consequence,
  - the more the sensors (N),
  - the larger $c = \lim N/n$,
  - the more probable we miss a spike
  - THAT LOOKS LIKE A PARADOX.

Generalization of the Tracy-Widom law

J. Baik, G. Ben Arous, S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," The Annals of Probability, vol. 33, no. 5, pp. 1643-1697, 2005.

Theorem

Let $\mathbf{X} \in \mathbb{C}^{N\times n}$ have i.i.d. Gaussian entries of zero mean and variance $1/n$ and $\mathbf{T}_N = diag(t_1, \ldots, t_N)$. Assume, for some fixed r, $t_{r+1} = \ldots = t_N = 1$ and $t_1 = \ldots = t_k$ while $t_{k+1}, \ldots, t_r$ lie in a compact subset of $(0, t_1)$. Assume further $c = \lim N/n < 1$. Denoting λ_N^+ the largest eigenvalue of $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$, we have

- *If $t_1 < 1 + \sqrt{\frac{N}{n}}$,*

$$N^{\frac{2}{3}} \frac{\lambda_N^+ - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}} c^{\frac{1}{2}}} \Rightarrow X^+ \sim F^+$$

  with F^+ the Tracy-Widom distribution.

- *If $t_1 > 1 + \sqrt{\frac{N}{n}}$,*

$$\left(t_1^2 - \frac{t_1^2 c}{(t_1-1)^2}\right)^{\frac{1}{2}} n^{\frac{1}{2}} \left[\lambda_N^+ - (t_1 + \frac{t_1 c}{t_1 - 1})\right] \Rightarrow X_k \sim G_k$$

  for some function G_k that is the distribution of the largest eigenvalue of the $k \times k$ GUE.

$$G_k(x) = \frac{1}{Z_k} \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \prod_{1\leqslant i<j\leqslant k} |\xi_i - \xi_j|^2 \prod_{i=1}^{k} e^{-\frac{1}{2}\xi_i^2} d\xi_1 \ldots d\xi_k$$

  In particular, $G_1(x) = \mathrm{erf}(x)$

Comments on the result

- there exists a “phase transition” when the largest population eigenvalues move from inside to outside $(0, 1+\sqrt{c})$.

Comments on the result

- there exists a "phase transition" when the largest population eigenvalues move from inside to outside $(0, 1 + \sqrt{c})$.
- more importantly, for $t_1 < 1 + \sqrt{c}$, we still have the same Tracy-Widom,
  - no way to see the spike even when zooming in
  - in fact, simulation suggests that convergence rate to the Tracy-Widom is slower with spikes.

Presence of a spike in previous model

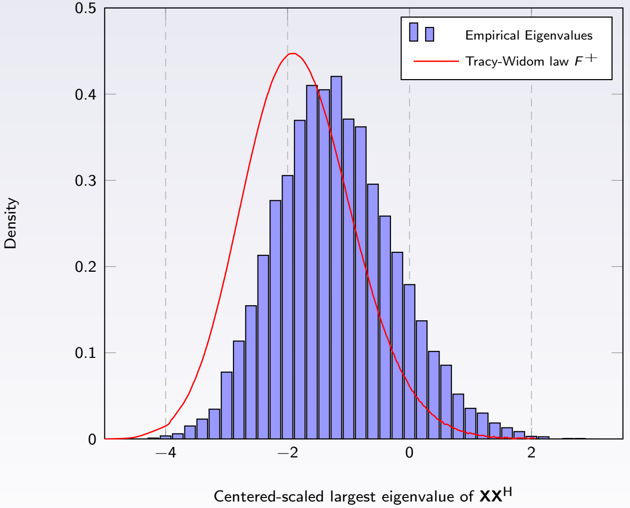


Figure: Distribution of $N^{\frac{2}{3}}c^{-\frac{1}{2}}(1+\sqrt{c})^{-\frac{4}{3}}\left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the distribution of X^+ (distributed as Tracy-Widom law) for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{T}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ with $\mathbf{T}$ diagonal with all entries 1 but for $T_{11} = 1.5$. Empirical distribution taken over 10,000 Monte-Carlo simulations.

Related bibliography

- J. W. Silverstein, J. Baik, “Eigenvalues of large sample covariance matrices of spiked population models” Journal of Multivariate Analysis, vol. 97, no. 6, pp. 1382-1408, 2006.
- J. Baik, G. Ben Arous, S. Péché, “Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices,” The Annals of Probability, vol. 33, no. 5, pp. 1643-1697, 2005.
- W. Hachem, P. Loubaton, X. Mestre, J. Najim, P. Vallet, “A Subspace Estimator for Fixed Rank Perturbations of Large Random Matrices,” arxiv preprint 1106.1497, 2011.
- R. Couillet, W. Hachem, “Local failure detection and diagnosis in large sensor networks”, (submitted to) IEEE Transactions on Information Theory, arXiv preprint 1107.1409.
- F. Benaych-Georges, R. Rao, “The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices,” Advances in Mathematics, vol. 227, no. 1, pp. 494-521, 2011.

Outline

Outline

Eigenvalue and eigenvectors statistics: Method

- Consider the model

$$\boldsymbol{\Sigma} = (\mathbf{I}_N + \mathbf{P})^{\frac{1}{2}}\mathbf{X}$$

with, for simplicity

  - $\mathbf{X}$ standard Gaussian
  - $\mathbf{P} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{C}^{N\times r}$, $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_r)$, $\omega_1 > \ldots > \omega_r > 0$.

- We study the convergence properties of
  - $\lambda_1 > \ldots > \lambda_r$, the r largest eigenvalues of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}$
  - $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i$, with $\hat{\mathbf{u}}_i$ the eigenvector associated to λ_i.

Eigenvalue and eigenvectors statistics: Method

- Consider the model

$$\boldsymbol{\Sigma} = (\mathbf{I}_N + \mathbf{P})^{\frac{1}{2}}\mathbf{X}$$

  with, for simplicity
  - $\mathbf{X}$ standard Gaussian
  - $\mathbf{P} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{C}^{N\times r}$, $\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_r)$, $\omega_1 > \ldots > \omega_r > 0$.
- We study the convergence properties of
  - $\lambda_1 > \ldots > \lambda_r$, the r largest eigenvalues of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}$
  - $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i$, with $\hat{\mathbf{u}}_i$ the eigenvector associated to λ_i.
- Systematic study based on two ingredients:
  - random matrix tools (the Stieltjes transform method)
  - complex analysis (complex contour integration)

First order limits: Eigenvalues

- We start with a study of the limiting extreme eigenvalues.

First order limits: Eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\begin{aligned}\det(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) &= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}]) \\ &= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).\end{aligned}$$

First order limits: Eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\begin{aligned}\det(\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) &= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}]) \\ &= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).\end{aligned}$$

- if x eigenvalue of $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}}$ but not of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, then for n large, $x > (1+\sqrt{c})^2$ (edge of MP law support) and

$$\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N+\mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}) = \det(\mathbf{I}_r + x\mathbf{\Omega}\mathbf{U}^*(\mathbf{I}_N + \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}) = 0$$

with $\mathbf{P} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} \in \mathbb{C}^{N\times r}$.

First order limits: Eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\det(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) = \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}])$$
$$= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).$$

- if x eigenvalue of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}$ but not of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, then for n large, $x > (1+\sqrt{c})^2$ (edge of MP law support) and

$$\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N+\mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}) = \det(\mathbf{I}_r + x\boldsymbol{\Omega}\mathbf{U}^*(\mathbf{I}_N + \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}) = 0$$

with $\mathbf{P} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} \in \mathbb{C}^{N\times r}$.

- due to unitary invariance of $\mathbf{X}$,

$$\mathbf{U}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U} \xrightarrow{\text{a.s.}} \int (t-x)^{-1} dF^{MP}(t)\mathbf{I}_r \triangleq m(x)\mathbf{I}_r$$

with F^{MP} the MP law, and $m(x)$ the Stieltjes transform of the MP law (often known for $r = 1$ as trace lemma).

First order limits: Eigenvalues

- We start with a study of the limiting extreme eigenvalues.
- Let $x > 0$, then

$$\det(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}} - x\mathbf{I}_N) = \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N + x[\mathbf{I}_N - (\mathbf{I}_N + \mathbf{P})^{-1}])$$
$$= \det(\mathbf{I}_N + \mathbf{P})\det(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}).$$

- if x eigenvalue of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}}$ but not of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, then for n large, $x > (1 + \sqrt{c})^2$ (edge of MP law support) and

  $$\det(\mathbf{I}_N + x\mathbf{P}(\mathbf{I}_N + \mathbf{P})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}) = \det(\mathbf{I}_r + x\boldsymbol{\Omega}\mathbf{U}^*(\mathbf{I}_N + \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}})^{-1}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}) = 0$$

  with $\mathbf{P} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} \in \mathbb{C}^{N \times r}$.
- due to unitary invariance of $\mathbf{X}$,

  $$\mathbf{U}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U} \xrightarrow{\text{a.s.}} \int (t - x)^{-1} dF^{MP}(t)\mathbf{I}_r \triangleq m(x)\mathbf{I}_r$$

  with F^{MP} the MP law, and $m(x)$ the Stieltjes transform of the MP law (often known for $r = 1$ as trace lemma).
- finally, we have that the *limiting* solutions ρ_k satisfy $\rho_k m(\rho_k) + (1 + \omega_k)\omega_k^{-1} = 0$.
- replacing $m(x)$, this is finally:

$$\lambda_k \xrightarrow{\text{a.s.}} \rho_k \triangleq 1 + \omega_k + c(1 + \omega_k)\omega_k^{-1}, \ \textit{if}\ \omega_k > \sqrt{c}$$

First order limits: Eigenvector projections (2)

- We now study the limiting behaviour of eigenvector projections.

First order limits: Eigenvector projections (2)

- We now study the limiting behaviour of eigenvector projections.
- Consider ω_i and its corresponding eigenvector $\mathbf{u}_i$, then, from Cauchy-integration formula

$$\mathbf{u}_i^\mathsf{H}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^\mathsf{H}\mathbf{u}_i = \frac{-1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{u}_i^\mathsf{H}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\mathsf{H} - z\mathbf{I}_N)^{-1}\mathbf{u}_i dz$$
$$= \frac{-1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{u}_i^\mathsf{H}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}(\mathbf{X}\mathbf{X}^\mathsf{H} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i dz + \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \hat{\mathbf{a}}_1^\mathsf{H}\widehat{\mathbf{H}}^{-1}\hat{\mathbf{a}}_2 dz$$

with $\mathcal{C}_i$ enclosing ρ_i only and

$$\begin{cases} \widehat{H} & = \mathbf{I}_r + z\boldsymbol{\Omega}(\mathbf{I}_r+\boldsymbol{\Omega})^{-1}\mathbf{U}^\mathsf{H}(\mathbf{X}\mathbf{X}^\mathsf{H} - z\mathbf{I}_N)^{-1}\mathbf{U} \\ \hat{a}_1^\mathsf{H} & = z\mathbf{u}_1^*(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}(\mathbf{X}\mathbf{X}^\mathsf{H} - z\mathbf{I}_N)^{-1}\mathbf{U} \\ \hat{a}_2 & = \boldsymbol{\Omega}(\mathbf{I}_r+\boldsymbol{\Omega})^{-1}\mathbf{U}^\mathsf{H}(\mathbf{X}\mathbf{X}^\mathsf{H} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i. \end{cases}$$

First order limits: Eigenvector projections (2)

- We now study the limiting behaviour of eigenvector projections.
- Consider ω_i and its corresponding eigenvector $\mathbf{u}_i$, then, from Cauchy-integration formula

$$\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i = \frac{-1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{u}_i^{\mathsf{H}}(\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{u}_i dz$$
$$= \frac{-1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{u}_i^{\mathsf{H}}(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i dz + \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \hat{\mathbf{a}}_1^{\mathsf{H}}\widehat{\mathbf{H}}^{-1}\hat{\mathbf{a}}_2 dz$$

with $\mathcal{C}_i$ enclosing ρ_i only and

$$\begin{cases} \widehat{H} & = \mathbf{I}_r + z\mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{U} \\ \hat{\mathbf{a}}_1^{\mathsf{H}} & = z\mathbf{u}_1^*(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{U} \\ \hat{\mathbf{a}}_2 & = \mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i. \end{cases}$$

- For large n, the first term has no pole, while the second converges to

$$T_i \triangleq \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i} \mathbf{a}_1^{\mathsf{H}}\mathbf{H}^{-1}\mathbf{a}_2 dz, \text{ with } \begin{cases} \mathbf{H} & = \mathbf{I}_r + zm(z)\mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1} \\ \mathbf{a}_1^{\mathsf{H}} & = zm(z)\mathbf{u}_1^*(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}\mathbf{U} \\ \mathbf{a}_2 & = m(z)\mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{I}_N + \mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i \end{cases}$$

which after development is $T_i = \sum_{\ell=1}^{r} \frac{1}{1+\omega_\ell} \frac{1}{2\pi\imath} \oint_{\mathcal{C}_i} \frac{zm^2(z)}{\frac{1+\omega_\ell}{\omega_\ell} + zm(z)} dz$.

First order limits: Eigenvector projections (2)

- We now study the limiting behaviour of eigenvector projections.
- Consider ω_i and its corresponding eigenvector $\mathbf{u}_i$, then, from Cauchy-integration formula

$$\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i = \frac{-1}{2\pi\imath}\oint_{\mathcal{C}_i}\mathbf{u}_i^{\mathsf{H}}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{u}_i dz$$
$$= \frac{-1}{2\pi\imath}\oint_{\mathcal{C}_i}\mathbf{u}_i^{\mathsf{H}}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i dz + \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i}\hat{\mathbf{a}}_1^{\mathsf{H}}\widehat{\mathbf{H}}^{-1}\hat{\mathbf{a}}_2 dz$$

with $\mathcal{C}_i$ enclosing ρ_i only and

$$\begin{cases} \widehat{H} & = \mathbf{I}_r + z\boldsymbol{\Omega}(\mathbf{I}_r+\boldsymbol{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{U} \\ \hat{\mathbf{a}}_1^{\mathsf{H}} & = z\mathbf{u}_1^*(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\mathbf{U} \\ \hat{\mathbf{a}}_2 & = \boldsymbol{\Omega}(\mathbf{I}_r+\boldsymbol{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i. \end{cases}$$

- For large n, the first term has no pole, while the second converges to

$$T_i \triangleq \frac{1}{2\pi\imath}\oint_{\mathcal{C}_i}\mathbf{a}_1^{\mathsf{H}}\mathbf{H}^{-1}\mathbf{a}_2 dz, \text{ with } \begin{cases} \mathbf{H} & = \mathbf{I}_r + zm(z)\boldsymbol{\Omega}(\mathbf{I}_r+\boldsymbol{\Omega})^{-1} \\ \mathbf{a}_1^{\mathsf{H}} & = zm(z)\mathbf{u}_1^*(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{U} \\ \mathbf{a}_2 & = m(z)\boldsymbol{\Omega}(\mathbf{I}_r+\boldsymbol{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{I}_N+\mathbf{P})^{-\frac{1}{2}}\mathbf{u}_i \end{cases}$$

which after development is $T_i = \sum_{\ell=1}^{r}\frac{1}{1+\omega_\ell}\frac{1}{2\pi\imath}\oint_{\mathcal{C}_i}\frac{zm^2(z)}{\frac{1+\omega_\ell}{\omega_\ell}+zm(z)}dz$.

- Using residue calculus, the sole pole is in ρ_i and we find $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i \xrightarrow{\text{a.s.}} \zeta_i \triangleq \frac{1-c\omega_i^{-2}}{1+c\omega_i^{-1}}$.

Fluctuations

- The objective is to find second order behaviour for the joint variable

$$\left(\left(\sqrt{N}(\lambda_i - \rho_i)\right)_{i=1}^{r}, \left(\sqrt{N}(\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i - \zeta_i)\right)_{i=1}^{r}\right)$$

Fluctuations

- The objective is to find second order behaviour for the joint variable

$$\left(\left(\sqrt{N}(\lambda_i-\rho_i)\right)_{i=1}^r,\left(\sqrt{N}(\mathbf{u}_i^{\sf H}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\sf H}\mathbf{u}_i-\zeta_i)\right)_{i=1}^r\right)$$

- Outline of the method:
  - Complex integration framework for the quantities $\sqrt{N}(\lambda_i-\rho_i)$ and $\sqrt{N}(\mathbf{u}_i^{\sf H}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\sf H}\mathbf{u}_i-\zeta_i)$:

$$\sqrt{N}(\lambda_i-\rho_i)-\left[-\frac{\rho_i}{h'(\rho_i)}\mathbf{u}_i^{\sf H}(m(\rho_i)\mathbf{I}_N-(\mathbf{XX}^{\sf H}-\rho_i\mathbf{I}_N)^{-1})\mathbf{u}_i\right]\xrightarrow{\text{a.s.}}0$$

$$\begin{aligned}\sqrt{N}(\mathbf{u}_i^{\sf H}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\sf H}\mathbf{u}_i-\zeta_i)-\Bigg[&\frac{h(\rho_i)(1+h(\rho_i))h''(\rho_i)}{h'(\rho_i)^3}\mathbf{u}_i^{\sf H}(m(\rho_i)\mathbf{I}_N-(\mathbf{XX}^{\sf H}-\rho_i\mathbf{I}_N)^{-1})\mathbf{u}_i\\&-\frac{h(\rho_i)(1+h(\rho_i))}{h'(\rho_i)^2}\mathbf{u}_i^{\sf H}(m'(\rho_i)\mathbf{I}_N-(\mathbf{XX}^{\sf H}-\rho_i\mathbf{I}_N)^{-2})\mathbf{u}_i\Bigg]\xrightarrow{\text{a.s.}}0\end{aligned}$$

    with $h(x)=xm(x)$.
  - Joint fluctuations of Stieltjes transforms:

$$\left(\mathbf{u}_i^{\sf H}(m(\rho_i)\mathbf{I}_N-(\mathbf{XX}^{\sf H}-\rho_i\mathbf{I}_N)^{-1})\mathbf{u}_i,\mathbf{u}_j^{\sf H}(m'(\rho_j)\mathbf{I}_N-(\mathbf{XX}^{\sf H}-\rho_j\mathbf{I}_N)^{-2})\mathbf{u}_j\right)\Rightarrow\mathcal{N}(0,R(\rho_i)\delta_i^j)$$

    with

$$R(\rho)=\begin{bmatrix}m'(\rho)-m(\rho)^2 & m''(\rho)/2-m(\rho)m'(\rho)\\ m''(\rho)/2-m(\rho)m'(\rho) & m^{(3)}(\rho)/6-m'(\rho)^2\end{bmatrix}$$

Joint fluctuations

R. Couillet, W. Hachem, "Local failure detection and diagnosis in large sensor networks", (submitted to) IEEE Transactions on Information Theory, arXiv preprint 1107.1409.

- Replacing $m(\rho_i)$, this finally proves the following theorem:

Theorem

Under the conditions above, assuming $\omega_i > \sqrt{c}$ for each $i \in \{1, \ldots, r\}$,

$$\left(\left(\sqrt{N}(\lambda_i - \rho_i)\right)_{i=1}^r, \left(\sqrt{N}(\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i - \zeta_i)\right)_{i=1}^r\right) \Rightarrow \mathcal{N}\left(0, \begin{bmatrix} C(\rho_1) & & \\ & \ddots & \\ & & C(\rho_r) \end{bmatrix}\right)$$

where

$$C(\rho_i) \triangleq \begin{bmatrix} \frac{c^2(1+\omega_i)^2}{(c+\omega_i)^2(\omega_i^2-c)}\left(c\frac{(1+\omega_i)^2}{(c+\omega_i)^2}+1\right) & \frac{(1+\omega_i)^3c^2}{(\omega_i+c)^2\omega_i} \\ \frac{(1+\omega_i)^3c^2}{(\omega_i+c)^2\omega_i} & \frac{c(1+\omega_i)^2(\omega_i^2-c)}{\omega_i^2} \end{bmatrix}.$$

Simulation

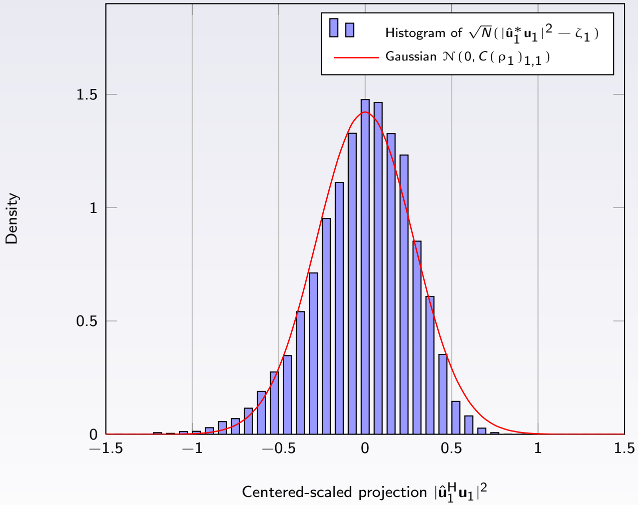


Figure: Empirical and theoretical distribution of the fluctuations of $\hat{\mathbf{u}}_1$ with $r = 1$, $X_{ij} \sim \mathcal{CN}(0, 1/n)$, $N/n = 1/8$, $N = 64$ and $\omega_1 = 1$.

Outline

Reminder: fluctuations of functionals of the spectrum

J. W. Silverstein, Z. D. Bai, "CLT of linear spectral statistics of large dimensional sample covariance matrices" Annals of Probability 32(1A) (2004), pp. 553-605.

Theorem

$$\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}, \quad \underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N$$

as usual with $\mathbf{X}_N$ *Gaussian,* $F^{\mathbf{T}_N} = diag(\{\tau_i\}) \Rightarrow H$, $|\mathbf{T}_N|$, $\tau_1 \geqslant \ldots \geqslant \tau_N$. *Denote* F *and* F_N *the l.s.d. and det. eq. of* $F^{\mathbf{B}_N}$, *and*

$$G_N \triangleq N \left[F^{\mathbf{B}_N} - F_N \right].$$

For $f_1, \ldots, f_k$ *well behaved, then*

$$\left(\int f_1(x) dG_N(x), \ldots, \int f_k(x) dG_N(x) \right) \Rightarrow (X_{f_1}, \ldots, X_{f_k})$$

of zero mean and covariance $\mathrm{Cov}(X_f, X_g)$, $(f, g) \in \{f_1, \ldots, f_k\}^2$, *such that*

$$\mathrm{Cov}(X_f, X_g) = -\frac{1}{2\pi i} \oint\oint \frac{f(z_1) g(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} \underline{m}'(z_1) \underline{m}'(z_2) dz_1 dz_2$$

for $\underline{m}(z)$ *the Stieltjes transform of the l.s.d. of* $\underline{\mathbf{B}}_N$. *The integration contours are positively defined with winding number one and enclose the support of* F.

The delta-method

- The central limit of random matrix-based estimates follow from basic fluctuation results, using the delta method.

The delta-method

- The central limit of random matrix-based estimates follow from basic fluctuation results, using the delta method.

Theorem

Let $X_1, X_2, \ldots \in \mathbb{R}^n$ *be a random sequence such that*

$$a_n(X_n - \mu) \Rightarrow X \sim \mathcal{N}(0, \mathbf{V})$$

for some sequence $a_1, a_2, \ldots \uparrow \infty$*. Then for* $f : \mathbb{R}^n \to \mathbb{R}^N$*, a function differentiable at* μ

$$a_n(f(X_n) - f(\mu)) \Rightarrow \mathbf{J}(f)X$$

with $\mathbf{J}(f)$ *the Jacobian matrix of* f*.*

Example of application: fluctuations of Mestre's estimator

J. Yao, R. Couillet, J. Najim, M. Debbah, "Fluctuations of an Improved Population Eigenvalue Estimator in Sample Covariance Matrix Models", (submitted to) IEEE Transactions on Information Theory.

Theorem

$$\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}, \quad \mathbf{T}_N = \mathit{diag}(\{t_k\}_{k=1}^K) \text{ with large multiplicities.}$$

Assume asymptotic cluster separability. Then, as N, n grow large

$$\left(n(\hat{t}_k - t_k)\right)_{k=1}^K \Rightarrow \mathcal{CN}(0, \mathbf{\Theta}), \text{ with}$$

$$\Theta_{k,k'} \triangleq -\frac{1}{4\pi^2 c^2 c_i c_j} \oint_{\mathcal{C}_k} \oint_{\mathcal{C}_{k'}} \left[\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} \right] \frac{dz_1 dz_2}{\underline{m}(z_1)\underline{m}(z_2)}$$

where $\mathcal{C}_k$ is the support enclosing cluster k.

Example of application: fluctuations of Mestre's estimator (2)

An estimator of the variance is also given in the following result.

Theorem

We also have

$$\hat{\Theta}_{k,k'} - \Theta_{k,k'} \xrightarrow{\text{a.s.}} 0$$

as $N, n \to \infty$, *where*

$$\hat{\Theta}_{k,k'} \triangleq \frac{n^2}{N_k N_{k'}} \left[\sum_{\substack{i \in \mathcal{N}_k \\ j \in \mathcal{N}_{k'}}} \frac{-1}{(\mu_i - \mu_j)^2 m'_{\mathbf{B}_N}(\mu_i) m'_{\mathbf{B}_N}(\mu_j)} + \delta_{kk'} \sum_{i \in \mathcal{N}_k} \left(\frac{m'''_{\mathbf{B}_N}(\mu_i)}{6 m'_{\mathbf{B}_N}(\mu_i)^3} - \frac{m''_{\mathbf{B}_N}(\mu_i)^2}{4 m'_{\mathbf{B}_N}(\mu_i)^4} \right) \right]$$

μ_i, *ordered eigenvalues of diag*$(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$; $\boldsymbol{\lambda}$, *ordered vector of eigenvalues of* $\mathbf{B}_N$.

Related bibliography

- J. Yao, R. Couillet, J. Najim, M. Debbah, "Fluctuations of an Improved Population Eigenvalue Estimator in Sample Covariance Matrix Models", (submitted to) IEEE Transactions on Information Theory.
- J. W. Silverstein, Z. D. Bai, "CLT of linear spectral statistics of large dimensional sample covariance matrices" Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.
- R. Couillet, W. Hachem, "Local failure detection and diagnosis in large sensor networks", (submitted to) IEEE Transactions on Information Theory, arXiv Preprint 1107.1409.
- F. Benaych-Georges, R. Rao, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," Advances in Mathematics, vol. 227, no. 1, pp. 494-521, 2011.
- J. Baik, G. Ben Arous, S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," The Annals of Probability, vol. 33, no. 5, pp. 1643-1697, 2005.