Random Matrices in Wireless Communications

Course 3: Beyond the spectrum: signal detection and spiked models

Romain Couillet
ST-Ericsson, Supélec, FRANCE
romain.couillet@supelec.fr

Supélec

Outline

Outline

Outline

Definition of a support

P. Billingsley, "Probability and measure," Wiley New York, 2008.

- According to Billingsley, a support of the probability measure P is any subspace A of Ω such that $P(A) = 1$. If F is the probability distribution of the random variable X, then $F(x) = P(\{\omega : X(\omega) \leq x\})$
- In random matrix theory, we call "the support of F" the smallest subspace of Ω of probability one (in the inclusion sense).
- *Example:* the distribution of density $f(x) = I_{\{0 \leq x \leq 1\}} + \delta(x - 10)$ has support $[0, 1]$.

Definition of a support

P. Billingsley, "Probability and measure," Wiley New York, 2008.

- According to Billingsley, a support of the probability measure P is any subspace A of Ω such that $P(A) = 1$. If F is the probability distribution of the random variable X, then $F(x) = P(\{\omega : X(\omega) \leq x\})$
- In random matrix theory, we call "the support of F" the smallest subspace of Ω of probability one (in the inclusion sense).
- *Example:* the distribution of density $f(x) = I_{\{0 \leq x \leq 1\}} + \delta(x - 10)$ has support $[0, 1]$.

Why go beyond the spectrum?

- Limiting spectral results only say where “most” of the eigenvalues are 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.
  - in general, for any $F_N^{(0)}$, if $F_N - F_N^{(0)} \Rightarrow 0$, then $F_N^{(0)} \Rightarrow F$
  - this is true for instance if F_N and $F_N^{(0)}$ differ by $o(N)$ eigenvalues.
- 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 $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$?

Why go beyond the spectrum?

- Limiting spectral results only say where "most" of the eigenvalues are 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.
  - in general, for any $F_N^{(0)}$, if $F_N - F_N^{(0)} \Rightarrow 0$, then $F_N^{(0)} \Rightarrow F$
  - this is true for instance if F_N and $F_N^{(0)}$ differ by $o(N)$ eigenvalues.
- We know that, for $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ with i.i.d. zero mean variance $1/n$,

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

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 $\mathbf{X}_N\mathbf{X}_N^H$?

Why go beyond the spectrum?

- Limiting spectral results only say where "most" of the eigenvalues are 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.
  - in general, for any $F_N^{(0)}$, if $F_N - F_N^{(0)} \Rightarrow 0$, then $F_N^{(0)} \Rightarrow F$
  - this is true for instance if F_N and $F_N^{(0)}$ differ by $o(N)$ eigenvalues.
- 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 $\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$?

Why go beyond the spectrum?

- Limiting spectral results only say where “most” of the eigenvalues are 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.
  - in general, for any $F_N^{(0)}$, if $F_N - F_N^{(0)} \Rightarrow 0$, then $F_N^{(0)} \Rightarrow F$
  - this is true for instance if F_N and $F_N^{(0)}$ differ by $o(N)$ eigenvalues.
- 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 ?

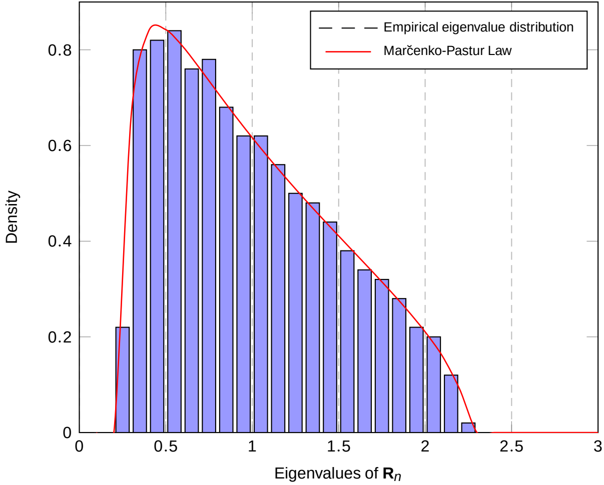

Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$

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 $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have i.i.d. entries with zero mean, variance $1/n$ and finite 4^{th} order moment. Let $\mathbf{T}_N \in \mathbb{C}^{N\times N}$ be nonrandom and uniformly bounded with N. The e.s.d. of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ converges weakly and almost surely to some F, as $N, n \to \infty$. Let F_N° be the distribution whose Stieltjes transform $m_{F_N^{\circ}}(z)$ is solution of

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

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

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

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, for all large N, $\mathbf{B}_N$ has no eigenvalue in $[a, b]$, with probability 1"

- If $\mathbf{T}_N$ is not identity,
  - for any large 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 add the resulting support to some space $\mathcal{A} \subset \mathbb{R}$.
  - do the previous for all $N \geq N_0$ and for the asymptotic $\lim_N F^{\mathbf{T}_N}$. This forms $\mathcal{A}$.
  - take $[a, b]$ outside $\mathcal{A}$, the result shows, for all N large, there is no eigenvalue there.
- this is very different from taking $[a, b]$ only outside the support of F !!!
- this is essential to understand spiked models, discussed in Section 23.

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, for all large N, $\mathbf{B}_N$ has no eigenvalue in $[a, b]$, with probability 1"

- If $\mathbf{T}_N$ is not identity,
  - for any large 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 add the resulting support to some space $\mathcal{A} \subset \mathbb{R}$.
  - do the previous for all $N \geq N_0$ and for the asymptotic $\lim_N F^{\mathbf{T}_N}$. This forms $\mathcal{A}$.
  - take $[a, b]$ outside $\mathcal{A}$, the result shows, for all N large, there is no eigenvalue there.
- this is very different from taking $[a, b]$ only outside the support of F !!!
- this is essential to understand spiked models, discussed in Section 23.

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, for all large N, $\mathbf{B}_N$ has no eigenvalue in $[a, b]$, with probability 1"

- If $\mathbf{T}_N$ is not identity,
  - for any large 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 add the resulting support to some space $\mathcal{A} \subset \mathbb{R}$.
  - do the previous for all $N \geq N_0$ and for the asymptotic $\lim_N F^{\mathbf{T}_N}$. This forms $\mathcal{A}$.
  - take $[a, b]$ outside $\mathcal{A}$, the result shows, for all N large, there is no eigenvalue there.
- this is very different from taking $[a, b]$ only outside the support of F !!!
- this is essential to understand spiked models, discussed in Section 23.

No eigenvalue outside which 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,” Journal of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has been shown yet 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,

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

- Unknown but worth digging are:
  - information plus noise models

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

- *Important remark*: $\mathbf{T}$ and $\mathbf{R}$ need not be deterministic as long as they have limiting distributions with probability 1 (thanks to Fubini/Tonelli's theorem).

No eigenvalue outside which 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,” Journal of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has been shown yet 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,

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

- Unknown but worth digging are:
  - information plus noise models

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

- *Important remark*: $\mathbf{T}$ and $\mathbf{R}$ need not be deterministic as long as they have limiting distributions with probability 1 (thanks to Fubini/Tonelli's theorem).

No eigenvalue outside which 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,” Journal of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has been shown yet 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,

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

- Unknown but worth digging are:
  - information plus noise models

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

- *Important remark*: **T** and **R** need not be deterministic as long as they have limiting distributions with probability 1 (thanks to Fubini/Tonelli's theorem).

No eigenvalue outside which 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," Journal of Multivariate Analysis vol. 100, no. 1, pp. 37-57, 2009.

- It has been shown yet 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,

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

- Unknown but worth digging are:
  - information plus noise models

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

- *Important remark*: **T** and **R** need not be deterministic as long as they have limiting distributions with probability 1 (thanks to Fubini/Tonelli's theorem).

Sketch of Proof

- Proof entirely relies on the Stieltjes transform!
- Up to now, we know

$$|m_{\mathbf{B}_N}(z) - m_{F_N}(z)| \xrightarrow{\text{a.s.}} 0$$

- This is not enough, we need in fact to show: for $z = x + i\sqrt{k}v_N$, $v_N = \frac{1}{N^{1/68}}$, $k = 1, \ldots, 34$,

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

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

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

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.

Sketch of Proof

- Proof entirely relies on the Stieltjes transform!
- Up to now, we know

$$|m_{\mathbf{B}_N}(z) - m_{F_N}(z)| \xrightarrow{\text{a.s.}} 0$$

- This is not enough, we need in fact to show: for $z = x + i\sqrt{k}v_N$, $v_N = \frac{1}{N^{1/68}}$, $k = 1, \ldots, 34$,

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

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

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

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.

Sketch of Proof

- Proof entirely relies on the Stieltjes transform!
- Up to now, we know

$$|m_{\mathbf{B}_N}(z) - m_{F_N}(z)| \xrightarrow{\text{a.s.}} 0$$

- This is not enough, we need in fact to show: for $z = x + i\sqrt{k}v_N$, $v_N = \frac{1}{N^{1/68}}$, $k = 1, \ldots, 34$,

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

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

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

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
  - if $\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^H$ 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 straight away
    - we need further study of the spectrum

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 $\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}$ 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 straight away
    - we need further study of the spectrum

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 straight away
    - we need further study of 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 the Stieltjes transform of the e.s.d. of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathrm{H}}\mathbf{T}_N^{\frac{1}{2}}$ satisfies $m_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} m_F(z)$, with

$$m_F(z) = \left(-z - \frac{n}{N}\int \frac{t}{1+tm_N(z)}dH(t)\right)^{-1}$$

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

- This can be inverted into

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

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

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 the Stieltjes transform of the e.s.d. of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ satisfies $m_{\mathbf{B}_N}(z) \xrightarrow{\text{a.s.}} m_F(z)$, with

$$m_F(z) = \left(-z - \frac{n}{N}\int \frac{t}{1 + t m_N(z)} dH(t)\right)^{-1}$$

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

- This can be inverted into

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

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 $\Im[m_F(z)]$ along this line. Now we can do better.
- It is shown that

$$\lim_{\substack{z \to x \in \mathbb{R}^* \\ z \in \mathbb{C}^+}} m_F(z) = m_0(x)$$

exists. We also have,

  - for x_0 inside the support, the density $f(x)$ of F in x is $\frac{1}{\pi}\Im[m_0(x)]$ with $m_0(x)$ the unique solution $m \in \mathbb{C}^+$ of

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

  - let $m_0 \in \mathbb{R}^*$ and x_F the equivalent to z_F on the real line. Then "x_0 outside the support of F" is equivalent to "$x_F'(m_F(x_0)) > 0$, $m_F(x_0) \neq 0$, $-1/m_F(x_0)$ outside the support of H".
- This provides another way to determine the support!. For $m \in (-\infty, 0)$, evaluate $x_F(m)$. Whenever x_F decreases, the image is outside the support. The rest is inside.

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 $\Im[m_F(z)]$ along this line. Now we can do better.
- It is shown that

$$\lim_{\substack{z \to x \in \mathbb{R}^* \\ z \in \mathbb{C}^+}} m_F(z) = m_0(x)$$

  exists. We also have,
  - for x_0 inside the support, the density $f(x)$ of F in x is $\frac{1}{\pi}\Im[m_0(x)]$ with $m_0(x)$ the unique solution $m \in \mathbb{C}^+$ of

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

  - let $m_0 \in \mathbb{R}^*$ and x_F the equivalent to z_F on the real line. Then "x_0 outside the support of F" is equivalent to "$x_F'(m_F(x_0)) > 0$, $m_F(x_0) \neq 0$, $-1/m_F(x_0)$ outside the support of H".
- This provides another way to determine the support!. For $m \in (-\infty, 0)$, evaluate $x_F(m)$. Whenever x_F decreases, the image is outside the support. The rest is inside.

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 $\Im[m_F(z)]$ along this line. Now we can do better.
- It is shown that
$$\lim_{\substack{z \to x \in \mathbb{R}^* \\ z \in \mathbb{C}^+}} m_F(z) = m_0(x)$$
exists. We also have,
  - for x_0 inside the support, the density $f(x)$ of F in x is $\frac{1}{\pi}\Im[m_0(x)]$ with $m_0(x)$ the unique solution $m \in \mathbb{C}^+$ of
$$x = -\frac{1}{m} - c \int \frac{t}{1+tm} dH(t)$$
  - let $m_0 \in \mathbb{R}^*$ and x_F the equivalent to z_F on the real line. Then "x_0 outside the support of F" is equivalent to "$x_F'(m_F(x_0)) > 0$, $m_F(x_0) \neq 0$, $-1/m_F(x_0)$ outside the support of H".
- This provides another way to determine the support!. For $m \in (-\infty, 0)$, evaluate $x_F(m)$. Whenever x_F decreases, the image is outside the support. The rest is inside.

Another way to determine the spectrum: spectrum to analyze

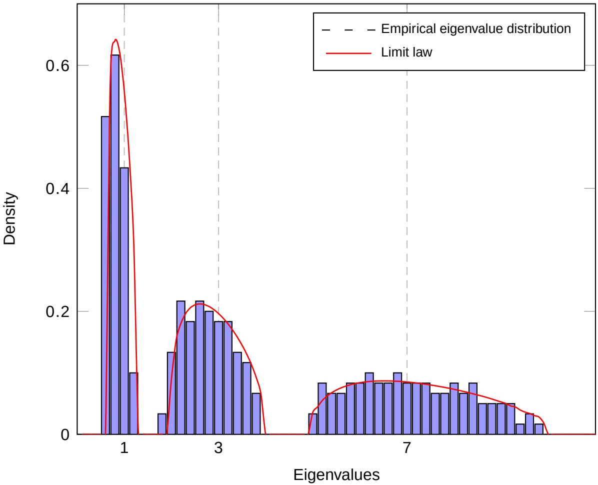

Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathrm{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7.

Another way to determine the spectrum: inverse function method

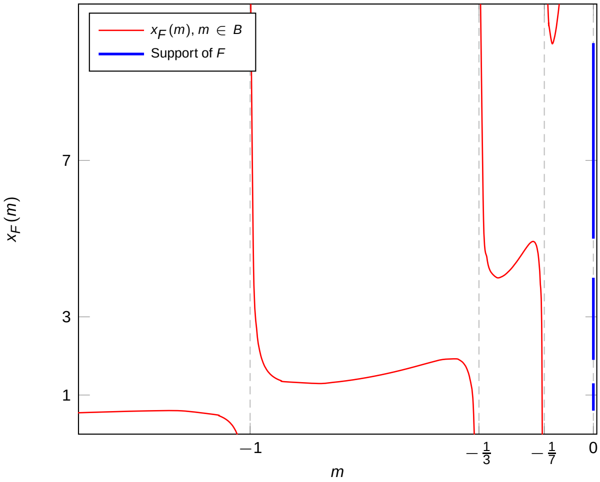

Figure: Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathrm{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7. The support of F is read on the vertical axis, whenever 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 $m_1^- < m_1^+ \leq m_2^- < m_2^+ \leq \ldots \leq m_Q^- < m_Q^+$.

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.
- 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.
- only the case of sample covariance matrix model is yet known
- inverse Stieltjes transform does not exist for more involved models

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.
- only the case of sample covariance matrix model is yet known
- inverse Stieltjes transform does not exist for more involved models

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.
- only the case of sample covariance matrix model is yet known
- inverse Stieltjes transform does not exist for more involved models

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 eigenvalues 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 $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^H \mathbf{T}_N^{\frac{1}{2}}$ 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 $c(1 - H(0)) \leq 1$, or $c(1 - H(0)) > 1$ but $[a, b]$ is not contained in $[0, x_0]$, then there exists N_0 such that for all $N \geq N_0$,*

$$P(\lambda_{i_N} > b, \lambda_{i_N+1} < a) = 1$$

where i_N is the unique integer such that

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

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 eigenvalues 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 $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^H \mathbf{T}_N^{\frac{1}{2}}$ *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* $c(1 - H(0)) \leq 1$*, or* $c(1 - H(0)) > 1$ *but* $[a, b]$ *is not contained in* $[0, x_0]$*, then there exists* N_0 *such that for all* $N \geq N_0$*,*

$$P(\lambda_{i_N} > b, \lambda_{i_N+1} < a) = 1$$

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, something extra "signal" must have been transmitted.
- the quantity of eigenvalues in each cluster gives an exact estimate of the multiplicity of the population!
- see Part 4., essential for eigen-inference.
- again, exact separation is only known for the sample covariance matrix model.
- if we need it for another model, we need to prove it!

Consequence of exact separation

- if eigenvalues are found outside the expected clusters, something extra “signal” must have been transmitted.
- the quantity of eigenvalues in each cluster gives an exact estimate of the multiplicity of the population!
- see Part 4., essential for eigen-inference.
- again, exact separation is only known for the sample covariance matrix model.
- if we need it for another model, we need to prove it!

Outline

Eigenvalues outside the limiting support ?

- 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.
- Is that in contradiction to previous Theorem?

Eigenvalues outside the limiting support ?

- 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.
- Is that in contradiction to previous Theorem?

Eigenvalues outside the support

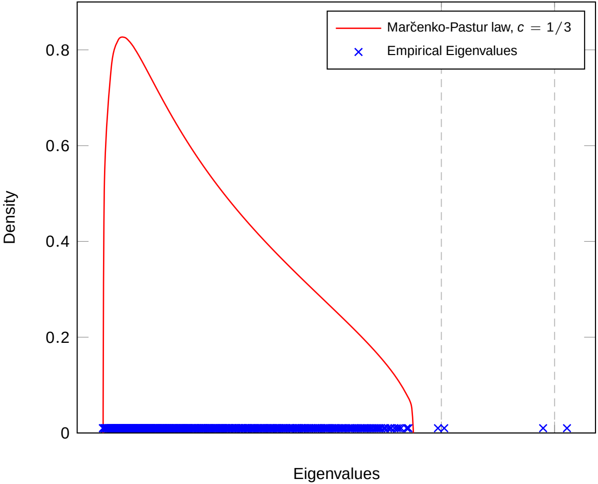

Figure: Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N} \Rightarrow 1_{[1,\infty)}$, ...Dimensions: $N = 500$, $n = 1500$.

Eigenvalues outside the support

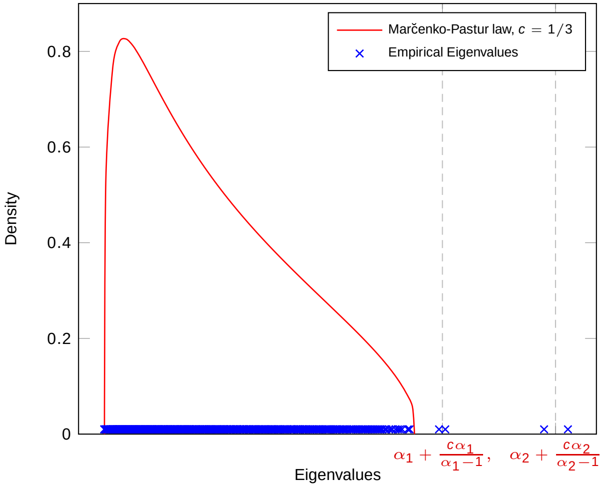

Figure: Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $F^{\mathbf{T}_N} \Rightarrow 1_{[1,\infty)}$, but $\mathbf{T}_N$ is a diagonal of ones but for the first four entries set to $\{\alpha_1, \alpha_1, \alpha_2, \alpha_2\}$, $\alpha_1 = 2$, $\alpha_2 = 3$.Dimensions: $N = 500$, $n = 1500$.

Spiked models

- No contradiction with "no eigenvalue" theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additiional clusters of positive measure in F_N°.
- However, for practical purposes, the presence of "spikes" determine the presence of a signal!
  - What about the absence of spikes?

Spiked models

- No contradiction with “no eigenvalue” theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additiional clusters of positive measure in F_N°.
- However, for practical purposes, the presence of “spikes” determine the presence of a signal!
  - What about the absence of spikes?

Spiked models

- No contradiction with “no eigenvalue” theorem, since the finitely numerous eigenvalues of $\mathbf{T}_N$ will form additiional clusters of positive measure in F_N°.
- However, for practical purposes, the presence of “spikes” determine 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*

$$\mathbf{T}_N = \mathrm{diag}(\underbrace{\alpha_1,\ldots,\alpha_1}_{k_1},\ldots,\underbrace{\alpha_M,\ldots,\alpha_M}_{k_M},\underbrace{1,\ldots,1}_{N-\sum_{i=1}^M k_i})$$

with $\alpha_1 > \ldots > \alpha_M > 0$, $c = \lim_N N/n$. *Call* $M_0 = \#\{j|\alpha_j > 1+\sqrt{c}\}$. *For* $c<1$, *call* M_1 *the integer such that* $(M-M_1) = \#\{j|\alpha_j < 1-\sqrt{c}$. *Denote* $\lambda_1,\ldots,\lambda_N$ *the eigenvalues of* $\mathbf{B}_N$. *We then have*

- *for* $1\le j\le M_0$, $1\le i\le k_j$,

$$\lambda_{k_1+\ldots+k_{j-1}+i} \xrightarrow{\text{a.s.}} \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}$$

\>\>\>

<<<

- for the other eigenvalues, we must discriminate upon c,
  - if $c < 1$,
    - for $M_1 + 1 \leq j \leq M$, $1 \leq i \leq k_j$,

$$\lambda_{N-k_j-\ldots-k_M+i} \xrightarrow{\text{a.s.}} \alpha_j + \frac{c\alpha_j}{\alpha_j - 1}$$

    - for the eigenvalues of $\mathbf{T}_N$ inside $[1-\sqrt{c}, 1+\sqrt{c}]$,

$$\lambda_{k_1+\ldots+k_{M_0}+1} \xrightarrow{\text{a.s.}} (1+\sqrt{c})^2$$

$$\lambda_{N-k_{M_1+1}-\ldots-k_M} \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$$

  - if $c > 1$,

$$\lambda_n \xrightarrow{\text{a.s.}} (1-\sqrt{c})^2$$

$$\lambda_{n+1} = \ldots = \lambda_N = 0$$

  - if $c = 1$,

$$\lambda_{\min(n,N)} \xrightarrow{\text{a.s.}} 0$$

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 in 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.
- (*in my opinion*) that just means "mere observation" is not the method to go for. A lot more might be said from the finite size sample eigenvalues than by looking at the position of their extremes.
- lastly, if all population spikes are "projected" to the edge of the spectrum, we should have a closer look to the edge, that must be "denser" than it should!

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 in 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.
- (*in my opinion*) that just means "mere observation" is not the method to go for. A lot more might be said from the finite size sample eigenvalues than by looking at the position of their extremes.
- lastly, if all population spikes are "projected" to the edge of the spectrum, we should have a closer look to the edge, that must be "denser" than it should!

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 in 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.
- (*in my opinion*) that just means “mere observation” is not the method to go for. A lot more might be said from the finite size sample eigenvalues than by looking at the position of their extremes.
- lastly, if all population spikes are “projected” to the edge of the spectrum, we should have a closer look to the edge, that must be “denser” than it should!

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 in 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.
- (*in my opinion*) that just means “mere observation” is not the method to go for. A lot more might be said from the finite size sample eigenvalues than by looking at the position of their extremes.
- lastly, if all population spikes are “projected” to the edge of the spectrum, we should have a closer look to the edge, that must be “denser” than it should!

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 in 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.
- (*in my opinion*) that just means “mere observation” is not the method to go for. A lot more might be said from the finite size sample eigenvalues than by looking at the position of their extremes.
- lastly, if all population spikes are “projected” to the edge of the spectrum, we should have a closer look to the edge, that must be “denser” than it should!

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 in 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.
- (*in my opinion*) that just means "mere observation" is not the method to go for. A lot more might be said from the finite size sample eigenvalues than by looking at the position of their extremes.
- lastly, if all population spikes are "projected" to the edge of the spectrum, we should have a closer look to the edge, that must be "denser" than it should!

Outline

The finite-size reality

- From the previous section, it seems smart to
  - look at the extreme eigenvalues and compare them to the theoretical limits
  - take the ratio between largest and smallest eigenvalues to determine if this fit the support size
- however,
  - all theorems imply very large N.
  - the N_0 above which theorems are valid is unknown (only in "i.o." terms)
- in reality, we deal with not-too-large dimensions, so we need to estimate errors using asymptotic methods.
- the solution is to zoom on the relevant part of the spectrum and see what happens there.

The finite-size reality

- From the previous section, it seems smart to
  - look at the extreme eigenvalues and compare them to the theoretical limits
  - take the ratio between largest and smallest eigenvalues to determine if this fit the support size
- however,
  - all theorems imply very large N.
  - the N_0 above which theorems are valid is unknown (only in "i.o." terms)
- in reality, we deal with not-too-large dimensions, so we need to estimate errors using asymptotic methods.
- the solution is to zoom on the relevant part of the spectrum and see what happens there.

The finite-size reality

- From the previous section, it seems smart to
  - look at the extreme eigenvalues and compare them to the theoretical limits
  - take the ratio between largest and smallest eigenvalues to determine if this fit the support size
- however,
  - all theorems imply very large N.
  - the N_0 above which theorems are valid is unknown (only in "i.o." terms)
- in reality, we deal with not-too-large dimensions, so we need to estimate errors using asymptotic methods.
- the solution is to zoom on the relevant part of the spectrum and see what happens there.

The finite-size reality

- From the previous section, it seems smart to
  - look at the extreme eigenvalues and compare them to the theoretical limits
  - take the ratio between largest and smallest eigenvalues to determine if this fit the support size
- however,
  - all theorems imply very large N.
  - the N_0 above which theorems are valid is unknown (only in "i.o." terms)
- in reality, we deal with not-too-large dimensions, so we need to estimate errors using asymptotic methods.
- the solution is to zoom on the relevant part of the spectrum and see what happens there.

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

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 $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, 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

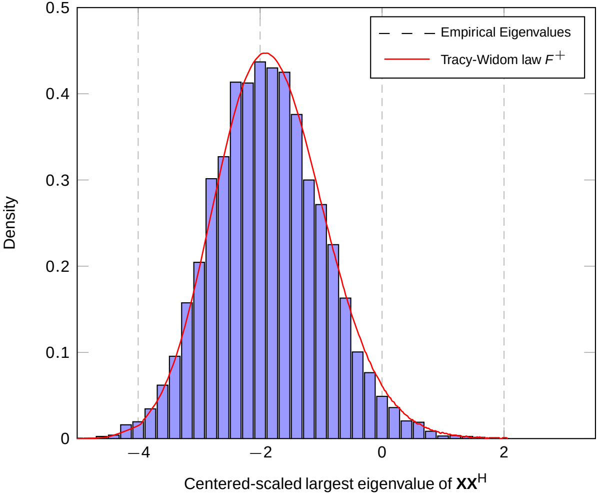

Figure: Distribution of $N^{\frac{2}{3}}c^{-\frac{1}{2}}(1+\sqrt{c})^{-\frac{4}{3}}\left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the distribution of X^+ (distributed as Tracy-Widom law) for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{XX}^{\mathsf{H}}$. Empirical distribution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

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives an 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?

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives an 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?

Comments on the Tracy-Widom law

- deeper result than limit eigenvalue result
- gives an 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?

Generalization of the Tracy-Widom law

J. Baik, "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 $\mathbf{T}_N = \mathrm{diag}(t_1, \ldots, t_N)$. 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 $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{T}^{\frac{1}{2}}$, 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.

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

  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 $(0, 1 + \sqrt{c})$.
- 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.

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

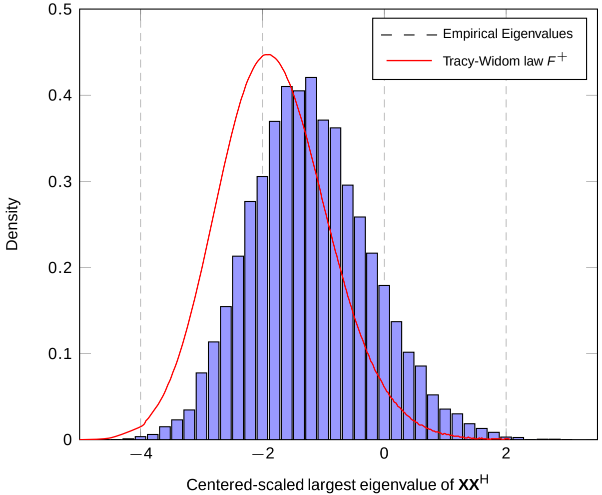

Figure: Distribution of $N^{\frac{2}{3}} c^{-\frac{1}{2}} (1+\sqrt{c})^{-\frac{4}{3}} \left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the distribution of X^+ (distributed as Tracy-Widom law) for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{T}^{\frac{1}{2}} \mathbf{XX}^{\mathsf{H}}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

What is left to be done?

- given previous slides, it is difficult to detect a population spike!
- if the $N \times n$ observation matrix is divided into chunks, it works better
- there is yet no proposition of an optimal way to detect population spikes.
- even if results were more conclusive, they require multiple realizations of large matrices to be usable... but one often has access to only one-shot results.
- given the “extra-tail” of the Tracy-Widom for smaller N, can that be exploited?

What is left to be done?

- given previous slides, it is difficult to detect a population spike!
- if the $N \times n$ observation matrix is divided into chunks, it works better
- there is yet no proposition of an optimal way to detect population spikes.
- even if results were more conclusive, they require multiple realizations of large matrices to be usable... but one often has access to only one-shot results.
- given the “extra-tail” of the Tracy-Widom for smaller N, can that be exploited?

What is left to be done?

- given previous slides, it is difficult to detect a population spike!
- if the $N \times n$ observation matrix is divided into chunks, it works better
- there is yet no proposition of an optimal way to detect population spikes.
- even if results were more conclusive, they require multiple realizations of large matrices to be usable... but one often has access to only one-shot results.
- given the “extra-tail” of the Tracy-Widom for smaller N, can that be exploited?

What is left to be done?

- given previous slides, it is difficult to detect a population spike!
- if the $N \times n$ observation matrix is divided into chunks, it works better
- there is yet no proposition of an optimal way to detect population spikes.
- even if results were more conclusive, they require multiple realizations of large matrices to be usable... but one often has access to only one-shot results.
- given the "extra-tail" of the Tracy-Widom for smaller N, can that be exploited?

Outline

Signal sensing scenario

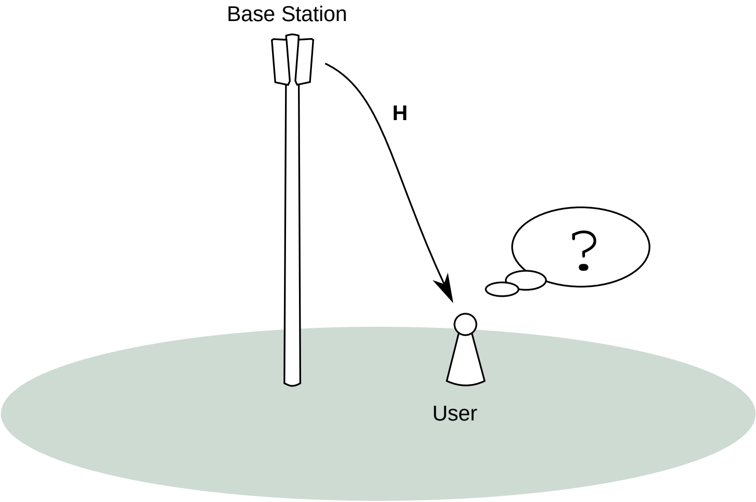

Figure: Signal detection scenario

System model

Consider N sensors that need to decide on the presence of a (single-source) signal embedded into white noise. With $\mathbf{y}_k \in \mathbb{C}^N$ the signal received at time k, we have the situations

$$\mathbf{y}_k = \begin{cases} \mathbf{h}x_k + \sigma\mathbf{w}_k & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \sigma\mathbf{w}_k & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

for a signal x_k, a flat fading channel $\mathbf{h}$ and noise $\mathbf{w}_k$ at time k. If $\mathbf{h}$ is stable for n time instants, this can be brought into the matrix $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_n]$,

$$\mathbf{Y} = \begin{cases} \begin{bmatrix} h_1 & \sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots \\ h_N & 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \\ w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \begin{bmatrix} \sigma & \cdots & 0 \\ \vdots & \ddots & \cdots \\ 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

System model

Consider N sensors that need to decide on the presence of a (single-source) signal embedded into white noise. With $\mathbf{y}_k \in \mathbb{C}^N$ the signal received at time k, we have the situations

$$\mathbf{y}_k = \begin{cases} \mathbf{h}x_k + \sigma \mathbf{w}_k & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \sigma \mathbf{w}_k & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

for a signal x_k, a flat fading channel $\mathbf{h}$ and noise $\mathbf{w}_k$ at time k. If $\mathbf{h}$ is stable for n time instants, this can be brought into the matrix $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_n]$,

$$\mathbf{Y} = \begin{cases} \begin{bmatrix} h_1 & \sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots \\ h_N & 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \\ w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \begin{bmatrix} \sigma & \cdots & 0 \\ \vdots & \ddots & \cdots \\ 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

Hypothesis test

- the problem here consists in finding the most plausible hypothesis among $\mathcal{H}_0$ and $\mathcal{H}_1$.
  - while nothing is known about the channel
  - while σ might not be known
  - while maybe more than a signal source is present...
- strategies of approach,
  - full blown finite-size Bayesian hypothesis testing give $\mathbf{Y}$
  - exploiting the asymptotic eigen-structure of $\mathbf{Y}$, assuming it is large enough

Hypothesis test

- the problem here consists in finding the most plausible hypothesis among $\mathcal{H}_0$ and $\mathcal{H}_1$.
  - while nothing is known about the channel
  - while σ might not be known
  - while maybe more than a signal source is present...
- strategies of approach,
  - full blown finite-size Bayesian hypothesis testing give $\mathbf{Y}$
  - exploiting the asymptotic eigen-structure of $\mathbf{Y}$, assuming it is large enough

Maximum entropy decision

R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Detection," *submitted to IEEE Trans. on Signal Processing*

- We need to determine the value of

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})} \tag{1}$$

- To cope with the absence of knowledge, we use the maximum-entropy principle to assign Bayesian *a priori* to unknown variables. Under knowledge of the SNR only,
  - $\mathbf{h}$ (possibly even $\mathbf{H} \in \mathbb{C}^{N\times M}$ for M hypothetical transmitters) is assigned a Gaussian distribution.
  - the signal and noise matrices have Gaussian i.i.d. entries.

Maximum entropy decision

R. Couillet, M. Debbah, “A Bayesian Framework for Collaborative Multi-Source Signal Detection,” *submitted to IEEE Trans. on Signal Processing*

- We need to determine the value of

$$C(\mathbf{Y}) = \frac{P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})} \tag{1}$$

- To cope with the absence of knowledge, we use the maximum-entropy principle to assign Bayesian *a priori* to unknown variables. Under knowledge of the SNR only,
  - $\mathbf{h}$ (possibly even $\mathbf{H} \in \mathbb{C}^{N\times M}$ for M hypothetical transmitters) is assigned a Gaussian distribution.
  - the signal and noise matrices have Gaussian i.i.d. entries.

Evaluating $C(\mathbf{Y})$

- Under hypothesis $\mathcal{H}_0$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{NL}} e^{-\frac{1}{\sigma^2}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

- Under hypothesis $\mathcal{H}_1$, assuming $M = 1$,

$$P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^N \lambda_i}}{N\pi^{nN}\sigma^{2(N-1)(n-1)}} \sum_{l=1}^N \frac{e^{\frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1 \\ i\neq l}}^N (\lambda_l - \lambda_i)} J_{N-n-1}(\sigma^2, \lambda_l)$$

with

$$J_k(x,y) = \int_x^{+\infty} t^k e^{-t-\frac{y}{t}} dt \text{ and } \lambda_i \text{ the eigenvalues of } \mathbf{Y}\mathbf{Y}^{\mathsf{H}}$$

- Finally,

$$C_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{1}{N} \sum_{l=1}^N \frac{\sigma^{2(N+n-1)} e^{\sigma^2 + \frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1 \\ i\neq l}}^N (\lambda_l - \lambda_i)} J_{N-n-1}(\sigma^2, \lambda_l)$$

Evaluating $C(\mathbf{Y})$

- Under hypothesis $\mathcal{H}_0$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{NL}} e^{-\frac{1}{\sigma^2}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

- Under hypothesis $\mathcal{H}_1$, assuming $M = 1$,

$$P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^{N}\lambda_i}}{N\pi^{nN}\sigma^{2(N-1)(n-1)}} \sum_{l=1}^{N} \frac{e^{\frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\neq l}}^{N}(\lambda_l - \lambda_i)} J_{N-n-1}(\sigma^2, \lambda_l)$$

with

$$J_k(x,y) = \int_x^{+\infty} t^k e^{-t-\frac{y}{t}}\,dt \text{ and } \lambda_i \text{ the eigenvalues of } \mathbf{Y}\mathbf{Y}^{\mathsf{H}}$$

- Finally,

$$C_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{1}{N}\sum_{l=1}^{N} \frac{\sigma^{2(N+n-1)} e^{\sigma^2 + \frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\neq l}}^{N}(\lambda_l - \lambda_i)} J_{N-n-1}(\sigma^2, \lambda_l)$$

Evaluating $C(\mathbf{Y})$

- Under hypothesis $\mathcal{H}_0$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y}) = \frac{1}{(\pi\sigma^2)^{NL}} e^{-\frac{1}{\sigma^2}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

- Under hypothesis $\mathcal{H}_1$, assuming $M = 1$,

$$P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{e^{\sigma^2 - \frac{1}{\sigma^2}\sum_{i=1}^N \lambda_i}}{N\pi^{nN}\sigma^{2(N-1)(n-1)}} \sum_{l=1}^N \frac{e^{\frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\neq l}}^N (\lambda_l - \lambda_i)} J_{N-n-1}(\sigma^2, \lambda_l)$$

with

$$J_k(x,y) = \int_x^{+\infty} t^k e^{-t-\frac{y}{t}}\,dt \text{ and } \lambda_i \text{ the eigenvalues of } \mathbf{Y}\mathbf{Y}^{\mathsf{H}}$$

- Finally,

$$C_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{1}{N}\sum_{l=1}^N \frac{\sigma^{2(N+n-1)} e^{\sigma^2 + \frac{\lambda_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\neq l}}^N (\lambda_l - \lambda_i)} J_{N-n-1}(\sigma^2, \lambda_l)$$

Generalization

- Generalization to $M \geq 1$ (M known)

$$C_{\mathbf{Y}|I_M}(\mathbf{Y}) = \frac{\sigma^{2M(N+n-M)}(N-M)!e^{M^2\sigma^2}}{N!M^{(M-1-2n)M/2}\prod_{j=1}^{M-1} j!} \sum_{\mathbf{a}\subset[1,N]} \frac{e^{\frac{\sum_{i=1}^{M}\lambda_{a_i}}{\sigma^2}}}{\prod_{a_i}\prod_{\substack{j\neq a_1\\ \dots \\ j\neq a_i}}(\lambda_{a_i}-\lambda_j)} \sum_{\mathbf{b}\in\mathcal{P}(M)} (-1)^{\mathrm{sgn}(\mathbf{b})+M} \prod_{l=1}^{M} J_{N-n-2+b_l}(M\sigma^2,$$

in which $J_k^{(x)} = J_k(2\sigma^2, 2x)$.

- If M is unknown

$$P(\mathbf{Y}|I_0) = \sum_{i=1}^{M_{max}} P(\mathbf{Y}|\text{“}M=i\text{”}, I_0)\cdot P(\text{“}M=i\text{”}|I_0)$$

- If σ is unknown, one possibility:

$$P_{\mathbf{Y}|I'_M} = \frac{1}{\sigma_+^2-\sigma_-^2}\int_{\sigma_-^2}^{\sigma_+^2} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y},\sigma^2)d\sigma^2$$

but here, long-standing problem with the maximum entropy principle and the determination of $P_{\sigma^2|I_M}$.

Generalization

- Generalization to $M \geq 1$ (M known)

$$C_{\mathbf{Y}|I_M}(\mathbf{Y}) = \frac{\sigma^{2M(N+n-M)}(N-M)!e^{M^2\sigma^2}}{N!M^{(M-1-2n)M/2}\prod_{j=1}^{M-1} j!} \sum_{\mathbf{a}\subset[1,N]} \frac{e^{\frac{\sum_{i=1}^{M}\lambda_{a_i}}{\sigma^2}}}{\prod_{a_i}\prod_{\substack{j\neq a_1 \\ \dots \\ j\neq a_i}}(\lambda_{a_i}-\lambda_j)} \sum_{\mathbf{b}\in\mathcal{P}(M)} (-1)^{\mathrm{sgn}(\mathbf{b})+M} \prod_{l=1}^{M} J_{N-n-2+b_l}(M\sigma^2,$$

in which $J_k^{(x)} = J_k(2\sigma^2, 2x)$.

- If M is unknown

$$P(\mathbf{Y}|I_0) = \sum_{i=1}^{M_{max}} P(\mathbf{Y}|\text{“}M=i\text{”}, I_0)\cdot P(\text{“}M=i\text{”}|I_0)$$

- If σ is unknown, one possibility:

$$P_{\mathbf{Y}|I'_M} = \frac{1}{\sigma_+^2 - \sigma_-^2}\int_{\sigma_-^2}^{\sigma_+^2} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y},\sigma^2)d\sigma^2$$

but here, long-standing problem with the maximum entropy principle and the determination of $P_{\sigma^2|I_M}$.

Results against classical energy detector

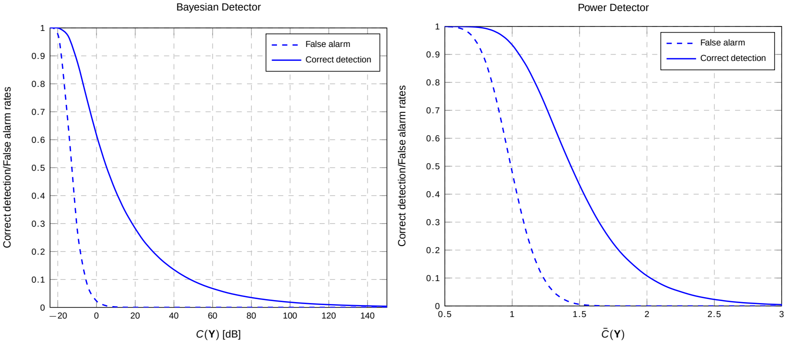

Figure: Power detection performance in SIMO - $M = 1$, $N = 4$, $L = 8$, SNR $= -3$ dB. On the left, Bayesian detector; on the right, classical power detector.

Bayesian detection gain

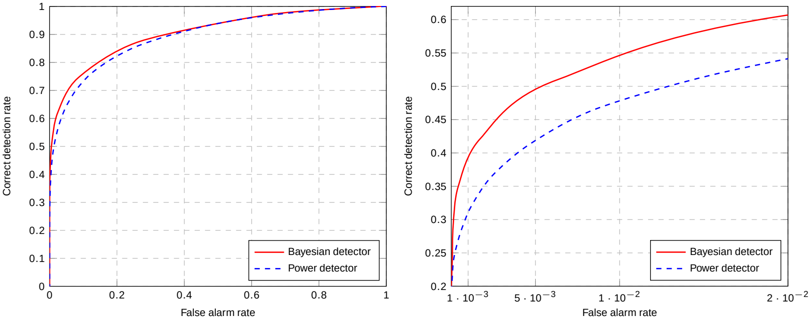

Figure: CDR against FAR for SIMO transmission - $M = 1$, $N = 4$, $L = 8$, SNR $= -3$ dB. On the left, full FAR range; on the right, FAR range of practical interest.

When σ is unknown

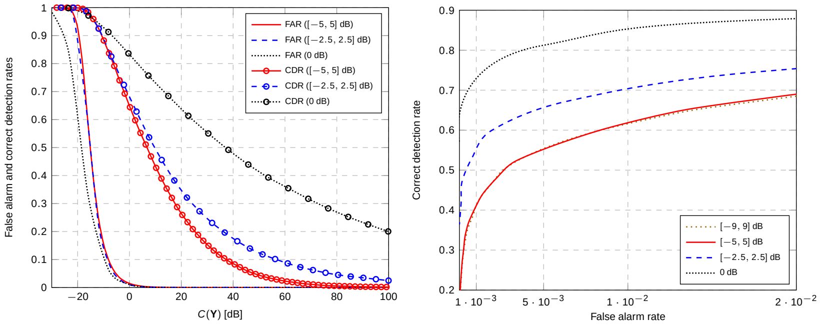

Figure: FAR and CDR, for different *a priori* information: exact SNR (0 dB), short range SNR ([−2.5, 2.5] dB) and large range SNR ([−5, 5] dB) SNR, $M = 1$, $N = 4$, $L = 8$, true SNR $= 0$ dB

Comments on the results

- On the plus side,
  - for this model, only eigenvalues matter.
  - for known σ, M, Bayesian approach outperforms energy detection.
  - however, for increasing M, it can be shown energy detection is close-to-optimal.
  - for unknown σ, M, new results.
  - for unknown σ, saturation observed: detection is always possible!
- However,
  - formulas are painful to derive
  - extension to more structured channels is very tough
  - for applications, formulas are difficult to exploit.

Comments on the results

- On the plus side,
  - for this model, only eigenvalues matter.
  - for known σ, M, Bayesian approach outperforms energy detection.
  - however, for increasing M, it can be shown energy detection is close-to-optimal.
  - for unknown σ, M, new results.
  - for unknown σ, saturation observed: detection is always possible!
- However,
  - formulas are painful to derive
  - extension to more structured channels is very tough
  - for applications, formulas are difficult to exploit.

Comments on the results

- On the plus side,
  - for this model, only eigenvalues matter.
  - for known σ, M, Bayesian approach outperforms energy detection.
  - however, for increasing M, it can be shown energy detection is close-to-optimal.
  - for unknown σ, M, new results.
  - for unknown σ, saturation observed: detection is always possible!
- However,
  - formulas are painful to derive
  - extension to more structured channels is very tough
  - for applications, formulas are difficult to exploit.

Outline

Reminder of the hypothesis testing problem

- *Reminder:* we want to test the hypothesis $\mathcal{H}_0$ against $\mathcal{H}_1$,

$$\mathbf{Y} = \begin{cases} \begin{bmatrix} h_1 & \sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots \\ h_N & 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \\ w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \begin{bmatrix} \sigma & \cdots & 0 \\ \vdots & \ddots & \cdots \\ 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

- we wish now to simplify the previous results using asymptotic compact-form results.

Reminder of the hypothesis testing problem

- *Reminder:* we want to test the hypothesis $\mathcal{H}_0$ against $\mathcal{H}_1$,

$$\mathbf{Y} = \begin{cases} \begin{bmatrix} h_1 & \sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots \\ h_N & 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \\ w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, information plus noise, hypothesis } \mathcal{H}_1 \\ \begin{bmatrix} \sigma & \cdots & 0 \\ \vdots & \ddots & \cdots \\ 0 & \cdots & \sigma \end{bmatrix} \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \cdots & \vdots \\ w_{N1} & \cdots & w_{Nn} \end{bmatrix} & \text{, pure noise, hpothesis } \mathcal{H}_0 \end{cases}$$

- we wish now to simplify the previous results using asymptotic compact-form results.

Exploiting the conditioning number

L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, “Cooperative spectrum sensing using random matrix theory,” International Symposium on Wireless Pervasive Computing, pp. 334-338 , 2008.

- under either hypothesis,
  - if $\mathcal{H}_0$, for N large, we expect $F_{\mathbf{YY}^{\mathsf{H}}}$ close to the Marčenko-Pastur law, of support $[\sigma^2\left(1-\sqrt{c}\right)^2, \sigma^2\left(1+\sqrt{c}\right)^2]$.
  - if $\mathcal{H}_1$, if population spike more than $1+\sqrt{\frac{N}{n}}$, largest eigenvalue is further away.
- the conditioning number of $\mathbf{YY}^{\mathsf{H}}$ is therefore asymptotically, as $N, n \to \infty$, $N/n \to c$,
  - if $\mathcal{H}_0$,

$$\mathrm{cond}(\mathbf{Y}) \triangleq \frac{\lambda_{\max}}{\lambda_{\min}} \to \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  - if $\mathcal{H}_1$,

$$\mathrm{cond}(\mathbf{Y}) \to t_1 + \frac{ct_1}{t_1-1} > \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  with $t_1 = \sum_{k=1}^{N} |h_k|^2 + \sigma^2$
- the conditioning number is independent of σ. We then have the decision criterion, whether or not σ is known,

$$\text{decide} \begin{cases} \mathcal{H}_0: & \text{if } \mathrm{cond}(\mathbf{YY}^{\mathsf{H}}) \le \frac{\left(1-\sqrt{\frac{N}{n}}\right)^2}{\left(1+\sqrt{\frac{N}{n}}\right)^2} + \varepsilon \\ \mathcal{H}_1: & \text{otherwise.} \end{cases}$$

for some security margin ε.

Exploiting the conditioning number

L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, “Cooperative spectrum sensing using random matrix theory,” International Symposium on Wireless Pervasive Computing, pp. 334-338 , 2008.

- under either hypothesis,
  - if $\mathcal{H}_0$, for N large, we expect $F_{\mathbf{YY}^{\mathsf{H}}}$ close to the Marčenko-Pastur law, of support $[\sigma^2\left(1-\sqrt{c}\right)^2, \sigma^2\left(1+\sqrt{c}\right)^2]$.
  - if $\mathcal{H}_1$, if population spike more than $1+\sqrt{\frac{N}{n}}$, largest eigenvalue is further away.
- the conditioning number of $\mathbf{YY}^{\mathsf{H}}$ is therefore asymptotically, as $N, n \to \infty$, $N/n \to c$,
  - if $\mathcal{H}_0$,

$$\mathrm{cond}(\mathbf{Y}) \triangleq \frac{\lambda_{\max}}{\lambda_{\min}} \to \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  - if $\mathcal{H}_1$,

$$\mathrm{cond}(\mathbf{Y}) \to t_1 + \frac{ct_1}{t_1-1} > \frac{\left(1-\sqrt{c}\right)^2}{\left(1+\sqrt{c}\right)^2}$$

  with $t_1 = \sum_{k=1}^{N} |h_k|^2 + \sigma^2$
- the conditioning number is independent of σ. We then have the decision criterion, whether or not σ is known,

$$\mathrm{decide} \begin{cases} \mathcal{H}_0: & \text{if } \mathrm{cond}(\mathbf{YY}^{\mathsf{H}}) \leq \frac{\left(1-\sqrt{\frac{N}{n}}\right)^2}{\left(1+\sqrt{\frac{N}{n}}\right)^2} + \varepsilon \\ \mathcal{H}_1: & \text{otherwise.} \end{cases}$$

  for some security margin ε.

Comments on the method

- Advantages are clear,
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known
- Drawbacks are however numerous,
  - however, previous statement stands only for very large N (it is expected that the dimension N for which asymptotic results arise be a function of σ!)
  - purely *ad-hoc* method that does not rely on any performance analysis

Comments on the method

- Advantages are clear,
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known
- Drawbacks are however numerous,
  - however, previous statement stands only for very large N (it is expected that the dimension N for which asymptotic results arise be a function of σ!)
  - purely *ad-hoc* method that does not rely on any performance analysis

Comments on the method

- Advantages are clear,
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known
- Drawbacks are however numerous,
  - however, previous statement stands only for very large N (it is expected that the dimension N for which asymptotic results arise be a function of σ!)
  - purely *ad-hoc* method that does not rely on any performance analysis

A deeper study of the problem

P. Bianchi, M. Debbah, M. Maïda, J. Najim, “Performance of Statistical Tests for Source Detection using Random Matrix Theory,”, arXiv preprint, 2009.

- We use the alternative generalized likelihood ratio test (GLRT) decision criterion, i.e.

$$C(\mathbf{Y}) = \frac{\sup_{\sigma^2,\mathbf{h}} P_{\mathbf{Y}|\mathbf{h},\sigma^2}(\mathbf{Y},\mathbf{h},\sigma^2)}{\sup_{\sigma^2} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y}|\sigma^2)}$$

- Denote

$$T_N = \frac{\lambda_{\min}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\frac{1}{N}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

  To guarantee a maximum false alarm ratio of α,

$$\text{decide} \begin{cases} \mathcal{H}_1: & \text{if } \left(1-\frac{1}{N}\right)^{(1-N)n} T_N^{-n}\left(1-\frac{\mathbf{T}_N}{N}\right)^{(1-N)n} > \xi_N \\ \mathcal{H}_0: & \text{otherwise.} \end{cases}$$

  for some threshold ξ_N that can be explicitly given as a function of α.
- This test is shown to perform better than the conditioning number test.
- But requires the knowledge of σ.

A deeper study of the problem

P. Bianchi, M. Debbah, M. Maïda, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory,", arXiv preprint, 2009.

- We use the alternative generalized likelihood ratio test (GLRT) decision criterion, i.e.

$$C(\mathbf{Y}) = \frac{\sup_{\sigma^2,\mathbf{h}} P_{\mathbf{Y}|\mathbf{h},\sigma^2}(\mathbf{Y},\mathbf{h},\sigma^2)}{\sup_{\sigma^2} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y}|\sigma^2)}$$

- Denote

$$T_N = \frac{\lambda_{\min}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\frac{1}{N}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

  To guarantee a maximum false alarm ratio of α,

$$\text{decide} \begin{cases} \mathcal{H}_1 : & \text{if } \left(1-\frac{1}{N}\right)^{(1-N)n} T_N^{-n}\left(1-\frac{T_N}{N}\right)^{(1-N)n} > \xi_N \\ \mathcal{H}_0 : & \text{otherwise.} \end{cases}$$

  for some threshold ξ_N that can be explicitly given as a function of α.
- This test is shown to perform better than the conditioning number test.
- But requires the knowledge of σ.

A deeper study of the problem

P. Bianchi, M. Debbah, M. Maïda, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory,", arXiv preprint, 2009.

- We use the alternative generalized likelihood ratio test (GLRT) decision criterion, i.e.

$$C(\mathbf{Y}) = \frac{\sup_{\sigma^2,\mathbf{h}} P_{\mathbf{Y}|\mathbf{h},\sigma^2}(\mathbf{Y},\mathbf{h},\sigma^2)}{\sup_{\sigma^2} P_{\mathbf{Y}|\sigma^2}(\mathbf{Y}|\sigma^2)}$$

- Denote

$$T_N = \frac{\lambda_{\min}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\frac{1}{N}\operatorname{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}$$

  To guarantee a maximum false alarm ratio of α,

$$\text{decide} \begin{cases} \mathcal{H}_1 : & \text{if } \left(1-\frac{1}{N}\right)^{(1-N)n} T_N^{-n} \left(1-\frac{\mathbf{T}_N}{N}\right)^{(1-N)n} > \xi_N \\ \mathcal{H}_0 : & \text{otherwise.} \end{cases}$$

  for some threshold ξ_N that can be explicitly given as a function of α.
- This test is shown to perform better than the conditioning number test.
- But requires the knowledge of σ.

Conclusions and openings

- Summary of this course,
  - theoretical results on “no eigenvalues outside the support” have large consequences in terms of signal detection.
  - results on spiked models exhibit fundamental limits linked to ratio N/n
  - even when looking closer to limit eigenvalue behaviour, nothing better is achievable.
  - finite-size considerations may solve the problem but are results are too complex
- perspectives,
  - a better way of using the N samples must be thought about.
  - statistical results are no good for one-shot treatments, large deviations may not be the solution.
  - what else to read in the eigenvalues?

Conclusions and openings

- Summary of this course,
  - theoretical results on “no eigenvalues outside the support” have large consequences in terms of signal detection.
  - results on spiked models exhibit fundamental limits linked to ratio N/n
  - even when looking closer to limit eigenvalue behaviour, nothing better is achievable.
  - finite-size considerations may solve the problem but are results are too complex
- perspectives,
  - a better way of using the N samples must be thought about.
  - statistical results are no good for one-shot treatments, large deviations may not be the solution.
  - what else to read in the eigenvalues?

Conclusions and openings

- Summary of this course,
  - theoretical results on “no eigenvalues outside the support” have large consequences in terms of signal detection.
  - results on spiked models exhibit fundamental limits linked to ratio N/n
  - even when looking closer to limit eigenvalue behaviour, nothing better is achievable.
  - finite-size considerations may solve the problem but are results are too complex
- perspectives,
  - a better way of using the N samples must be thought about.
  - statistical results are no good for one-shot treatments, large deviations may not be the solution.
  - what else to read in the eigenvalues?

Conclusions and openings

- Summary of this course,
  - theoretical results on “no eigenvalues outside the support” have large consequences in terms of signal detection.
  - results on spiked models exhibit fundamental limits linked to ratio N/n
  - even when looking closer to limit eigenvalue behaviour, nothing better is achievable.
  - finite-size considerations may solve the problem but are results are too complex
- perspectives,
  - a better way of using the N samples must be thought about.
  - statistical results are no good for one-shot treatments, large deviations may not be the solution.
  - what else to read in the eigenvalues?