Random Matrix Theory for Signal Processing Applications

Romain Couillet[1], **Mérouane Debbah**[2]
[1]EDF Chair on System Sciences and the Energy Challenge, Supélec, Gif sur Yvette, France
[2]Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif sur Yvette, FRANCE
{*romain.couillet,merouane.debbah*}*@supelec.fr*

ICASSP 2011, Prague, Czech Republic.

Outline

Outline

Outline

Definitions

Random Matrix

A random matrix is a matrix $\mathbf{X} \in \mathbb{C}^{N \times n}$ with random entries X_{ij} following a given probability distribution.

- In many problems (with symmetrical structures), interest is on:
  - eigenvalue distribution
  - eigenvector projections.
- Pioneering works due to Wishart on matrices

$$\mathbf{X}\mathbf{X}^{\mathrm{H}}$$

with $X_{ij} \sim \mathcal{CN}(0,1)$

Definitions

Random Matrix

A random matrix is a matrix $\mathbf{X} \in \mathbb{C}^{N \times n}$ with random entries X_{ij} following a given probability distribution.

- In many problems (with symmetrical structures), interest is on:
  - eigenvalue distribution
  - eigenvector projections.
- Pioneering works due to Wishart on matrices

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

with $X_{ij} \sim \mathcal{CN}(0,1)$

Definitions

Random Matrix

A random matrix is a matrix $\mathbf{X} \in \mathbb{C}^{N \times n}$ with random entries X_{ij} following a given probability distribution.

- In many problems (with symmetrical structures), interest is on:
  - eigenvalue distribution
  - eigenvector projections.
- Pioneering works due to Wishart on matrices

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

with $X_{ij} \sim \mathcal{CN}(0, 1)$

Wishart matrices

J. Wishart, “The generalized product moment distribution in samples from a normal multivariate population”, Biometrika, vol. 20A, pp. 32-52, 1928.

- Wishart describes the distribution of $\mathbf{R}_n = \mathbf{X}\mathbf{X}^{\mathsf{H}} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$, $\mathbf{x}_i \in \mathbb{C}^N \sim \mathcal{CN}(0, \mathbf{R})$,

$$P_{\mathbf{R}_n}(\mathbf{B}) = \frac{\pi^{N(N-1)/2}}{\det \mathbf{R}^n \prod_{i=1}^{N} (n-i)!} e^{-\operatorname{tr}(\mathbf{R}^{-1}\mathbf{B})} \det \mathbf{B}^{n-N}$$

- Joint and marginal eigenvalue distributions:

$$P_{(\lambda_i)}(\lambda_1, \ldots, \lambda_N) = \frac{\det(\{e^{-r_j^{-1}\lambda_i}\}_N)}{\Delta(\mathbf{R}^{-1})} \Delta(\mathbf{L}) \prod_{j=1}^{N} \frac{\lambda_j^{n-N}}{j!(n-j)!}$$

with $r_1 \geq \ldots \geq r_N$ the eigenvalues of $\mathbf{R}$ and $\mathbf{L} = \operatorname{diag}(\lambda_1 \geq \ldots \geq \lambda_N)$ and

$$p_\lambda(\lambda) = \frac{1}{M} \sum_{k=0}^{N-1} \frac{k!}{(k+n-N)!} [L_k^{n-N}]^2 \lambda^{n-N} e^{-\lambda}$$

where L_n^k are the Laguerre polynomials

$$L_n^k(\lambda) = \frac{e^\lambda}{k!\lambda^n} \frac{d^k}{d\lambda^k}(e^{-\lambda}\lambda^{n+k}).$$

Extension to more generic matrices

T. Ratnarajah and R. Vaillancourt and M. Alvo, "Eigenvalues and condition numbers of complex random matrices," SIAM Journal on Matrix Analysis and Applications, vol. 26, no. 2, pp. 441-456, 2005.

- Extensions to:
  - correlated Gaussian involve heavy tools (Schur polynomials)
  - non-Gaussian is virtually impossible!
- Solution is to assume increasing matrix dimensions: $N, n \to \infty$
  - deterministic limiting behaviour is often observed
  - loose assumptions on entry distributions (e.g. rotational symmetry, independent entries)
  - robust framework for very generic models are known:
    - Stieltjes transform methods (more efficient than Fourier transform)
    - moments/free probability methods (extension of classical probability for non-commutative variables)
    - physical methods for large systems (replica method)

This tutorial will introduce the major used methods but concentrates on the powerful Stieltjes transform method.

Extension to more generic matrices

T. Ratnarajah and R. Vaillancourt and M. Alvo, "Eigenvalues and condition numbers of complex random matrices," SIAM Journal on Matrix Analysis and Applications, vol. 26, no. 2, pp. 441-456, 2005.

- Extensions to:
  - correlated Gaussian involve heavy tools (Schur polynomials)
  - non-Gaussian is virtually impossible!
- Solution is to assume increasing matrix dimensions: $N, n \to \infty$
  - deterministic limiting behaviour is often observed
  - loose assumptions on entry distributions (e.g. rotational symmetry, independent entries)
  - robust framework for very generic models are known:
    - Stieltjes transform methods (more efficient than Fourier transform)
    - moments/free probability methods (extension of classical probability for non-commutative variables)
    - physical methods for large systems (replica method)

This tutorial will introduce the major used methods but concentrates on the powerful Stieltjes transform method.

Extension to more generic matrices

T. Ratnarajah and R. Vaillancourt and M. Alvo, "Eigenvalues and condition numbers of complex random matrices," SIAM Journal on Matrix Analysis and Applications, vol. 26, no. 2, pp. 441-456, 2005.

- Extensions to:
  - correlated Gaussian involve heavy tools (Schur polynomials)
  - non-Gaussian is virtually impossible!
- Solution is to assume increasing matrix dimensions: $N, n \to \infty$
  - deterministic limiting behaviour is often observed
  - loose assumptions on entry distributions (e.g. rotational symmetry, independent entries)
  - robust framework for very generic models are known:
    - Stieltjes transform methods (more efficient than Fourier transform)
    - moments/free probability methods (extension of classical probability for non-commutative variables)
    - physical methods for large systems (replica method)

This tutorial will introduce the major used methods but concentrates on the powerful Stieltjes transform method.

Extension to more generic matrices

T. Ratnarajah and R. Vaillancourt and M. Alvo, “Eigenvalues and condition numbers of complex random matrices,” SIAM Journal on Matrix Analysis and Applications, vol. 26, no. 2, pp. 441-456, 2005.

- Extensions to:
  - correlated Gaussian involve heavy tools (Schur polynomials)
  - non-Gaussian is virtually impossible!
- Solution is to assume increasing matrix dimensions: $N, n \to \infty$
  - deterministic limiting behaviour is often observed
  - loose assumptions on entry distributions (e.g. rotational symmetry, independent entries)
  - robust framework for very generic models are known:
    - Stieltjes transform methods (more efficient than Fourier transform)
    - moments/free probability methods (extension of classical probability for non-commutative variables)
    - physical methods for large systems (replica method)

This tutorial will introduce the major used methods but concentrates on the powerful Stieltjes transform method.

Extension to more generic matrices

T. Ratnarajah and R. Vaillancourt and M. Alvo, "Eigenvalues and condition numbers of complex random matrices," SIAM Journal on Matrix Analysis and Applications, vol. 26, no. 2, pp. 441-456, 2005.

- Extensions to:
  - correlated Gaussian involve heavy tools (Schur polynomials)
  - non-Gaussian is virtually impossible!
- Solution is to assume increasing matrix dimensions: $N, n \to \infty$
  - deterministic limiting behaviour is often observed
  - loose assumptions on entry distributions (e.g. rotational symmetry, independent entries)
  - robust framework for very generic models are known:
    - Stieltjes transform methods (more efficient than Fourier transform)
    - moments/free probability methods (extension of classical probability for non-commutative variables)
    - physical methods for large systems (replica method)

This tutorial will introduce the major used methods but concentrates on the powerful Stieltjes transform method.

Outline

Large dimensional data

Let $\mathbf{w}_1, \mathbf{w}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{w}_1 \mathbf{w}_1^{\mathrm{H}}] \in \mathbb{C}^{N \times N}$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} = \mathbf{W}\mathbf{W}^{\mathrm{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}}$$

  is a "good" estimate of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

What about the global behaviour? What about the eigenvalue distribution?

Large dimensional data

Let $\mathbf{w}_1, \mathbf{w}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{w}_1 \mathbf{w}_1^{\mathrm{H}}] \in \mathbb{C}^{N \times N}$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} = \mathbf{W}\mathbf{W}^{\mathrm{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}}$$

  is a "good" estimate of $\mathbf{R}$.

- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

What about the global behaviour? What about the eigenvalue distribution?

Large dimensional data

Let $\mathbf{w}_1, \mathbf{w}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{w}_1 \mathbf{w}_1^{\mathrm{H}}] \in \mathbb{C}^{N \times N}$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} = \mathbf{W}\mathbf{W}^{\mathrm{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}}$$

  is a “good” estimate of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

What about the global behaviour? What about the eigenvalue distribution?

Large dimensional data

Let $\mathbf{w}_1, \mathbf{w}_2 \ldots \in \mathbb{C}^N$ be independently drawn from an N-variate process of mean zero and covariance $\mathbf{R} = \mathrm{E}[\mathbf{w}_1 \mathbf{w}_1^{\mathrm{H}}] \in \mathbb{C}^{N \times N}$.

Law of large numbers

As $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} = \mathbf{W}\mathbf{W}^{\mathrm{H}} \xrightarrow{\text{a.s.}} \mathbf{R}$$

In reality, one cannot afford $n \to \infty$.

- if $n \gg N$,

$$\mathbf{R}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}}$$

  is a “good” estimate of $\mathbf{R}$.
- if $N/n = O(1)$, and if both (n, N) are large, we can still say, for all (i, j),

$$(\mathbf{R}_n)_{ij} \xrightarrow{\text{a.s.}} (\mathbf{R})_{ij}$$

What about the global behaviour? What about the eigenvalue distribution?

Empirical and limit spectra of Wishart matrices

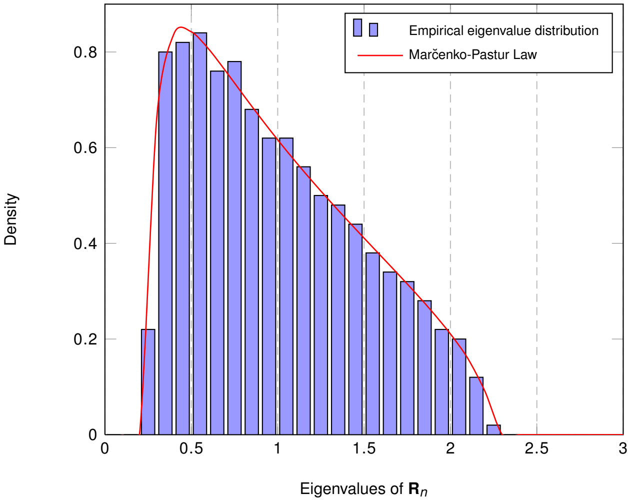


Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

The Marčenko-Pastur Law

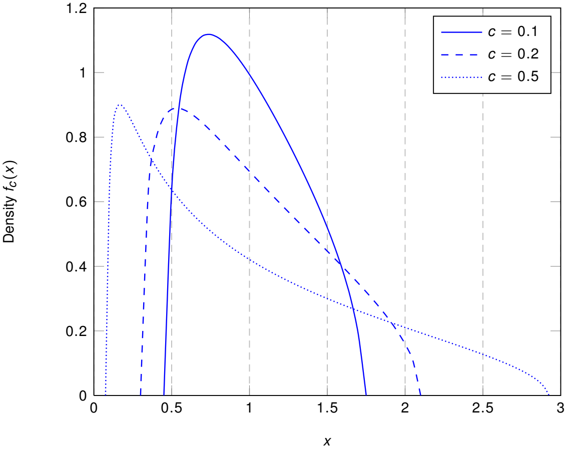


Figure: Marčenko-Pastur law for different limit ratios $c = \lim N/n$.

The Marčenko-Pastur law

Let $\mathbf{W} \in \mathbb{C}^{N \times n}$ have i.i.d. elements, of zero mean and variance $1/n$.
Eigenvalues of the matrix

$$n \left\{ \underbrace{\begin{bmatrix} & \mathbf{W}^{\mathsf{H}} & \end{bmatrix}}_{N} \right. \begin{bmatrix} \\ \mathbf{W} \\ \\ \end{bmatrix}$$

when $N, n \to \infty$ with $N/n \to c$ **IS NOT IDENTITY!**

Remark: If the entries are Gaussian, the matrix is called a Wishart matrix with n degrees of freedom. The **exact** distribution is known in the finite case.

Deriving the Marčenko-Pastur law

- We wish to determine the density $f_c(\lambda)$ of the asymptotic law, defined by

$$f_c(\lambda) = \lim_{\substack{N\to\infty \\ n\to\infty \\ N/n\to c}} \sum_{i=1}^{N} \delta\left(\lambda - \lambda_i(\mathbf{R}_n)\right)$$

- With $N/n \to c$, the moments of this distribution are given by

$$\begin{aligned}
M_1^N &= \frac{1}{N}\operatorname{tr}\mathbf{R}_n = \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\mathbf{R}_n) \to \int \lambda f_c(\lambda)d\lambda = 1 \\
M_2^N &= \frac{1}{N}\operatorname{tr}\mathbf{R}_n^2 = \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\mathbf{R}_n)^2 \to \int \lambda^2 f_c(\lambda)d\lambda = 1 + c \\
M_3^N &= \frac{1}{N}\operatorname{tr}\mathbf{R}_n^3 = \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\mathbf{R}_n)^3 \to \int \lambda^3 f_c(\lambda)d\lambda = c^2 + 3c + 1 \\
\cdots &= \cdots
\end{aligned}$$

- These moments correspond to a *unique* distribution function (under mild assumptions), which has density the Marčenko-Pastur law

$$f(x) = (1 - \frac{1}{c})^+\delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi c x}, \text{ with } a = (1-\sqrt{c})^2,\ b = (1+\sqrt{c})^2.$$

Deriving the Marčenko-Pastur law

- We wish to determine the density $f_c(\lambda)$ of the asymptotic law, defined by

$$f_c(\lambda) = \lim_{\substack{N\to\infty \\ n\to\infty \\ N/n\to c}} \sum_{i=1}^{N} \delta\left(\lambda - \lambda_i(\mathbf{R}_n)\right)$$

- With $N/n \to c$, the moments of this distribution are given by

$$\begin{aligned} M_1^N &= \frac{1}{N}\operatorname{tr}\mathbf{R}_n = \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\mathbf{R}_n) \to \int \lambda f_c(\lambda)d\lambda = 1 \\ M_2^N &= \frac{1}{N}\operatorname{tr}\mathbf{R}_n^2 = \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\mathbf{R}_n)^2 \to \int \lambda^2 f_c(\lambda)d\lambda = 1+c \\ M_3^N &= \frac{1}{N}\operatorname{tr}\mathbf{R}_n^3 = \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\mathbf{R}_n)^3 \to \int \lambda^3 f_c(\lambda)d\lambda = c^2+3c+1 \\ \cdots &= \cdots \end{aligned}$$

- These moments correspond to a *unique* distribution function (under mild assumptions), which has density the Marčenko-Pastur law

$$f(x) = (1-\frac{1}{c})^+\delta(x) + \frac{\sqrt{(x-a)^+(b-x)^+}}{2\pi cx}, \text{ with } a=(1-\sqrt{c})^2,\ b=(1+\sqrt{c})^2.$$

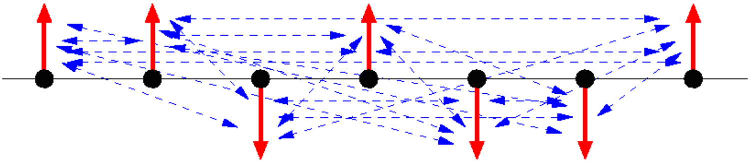
Outline

Wigner and semi-circle law

Schrödinger's equation

$$H\Phi_i = E_i\Phi_i$$

where Φ_i is the wave function,
E_i is the energy level,
H is the Hamiltonian.



Magnetic interactions between the spins of electrons

The birth of large dimensional random matrix theory

Eugene Paul Wigner, 1902-1995

The birth of large dimensional random matrix theory

E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," The annals of mathematics, vol. 62, pp. 546-564, 1955.

$$\mathbf{X}_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 0 & +1 & +1 & +1 & -1 & -1 & \cdots \\ +1 & 0 & -1 & +1 & +1 & +1 & \cdots \\ +1 & -1 & 0 & +1 & +1 & +1 & \cdots \\ +1 & +1 & +1 & 0 & +1 & +1 & \cdots \\ -1 & +1 & +1 & +1 & 0 & -1 & \cdots \\ -1 & +1 & +1 & +1 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

As the matrix dimension increases, what can we say about the eigenvalues (energy levels)?

Semi-circle law, Full circle law...

- If $\mathbf{X}_N \in \mathbb{C}^{N\times N}$ is Hermitian with i.i.d. entries of mean 0, variance $1/N$ above the diagonal, then $F^{\mathbf{X}_N} \xrightarrow{\text{a.s.}} F$ where F has density f the semi-circle law

$$f(x) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}$$

- Shown from the method of moments

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr} \mathbf{X}_N^{2k} = \frac{1}{k+1} C_k^{2k}$$

which are exactly the moments of $f(x)$!

- If $\mathbf{X}_N \in \mathbb{C}^{N\times N}$ has i.i.d. 0 mean, variance $1/N$ entries, then asymptotically its complex eigenvalues distribute uniformly on the complex unit circle.

Semi-circle law, Full circle law...

- If $\mathbf{X}_N \in \mathbb{C}^{N\times N}$ is Hermitian with i.i.d. entries of mean 0, variance $1/N$ above the diagonal, then $F^{\mathbf{X}_N} \xrightarrow{\text{a.s.}} F$ where F has density f the semi-circle law

$$f(x) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}$$

- Shown from the method of moments

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr} \mathbf{X}_N^{2k} = \frac{1}{k+1} C_k^{2k}$$

  which are exactly the moments of $f(x)$!
- If $\mathbf{X}_N \in \mathbb{C}^{N\times N}$ has i.i.d. 0 mean, variance $1/N$ entries, then asymptotically its complex eigenvalues distribute uniformly on the complex unit circle.

Semi-circle law

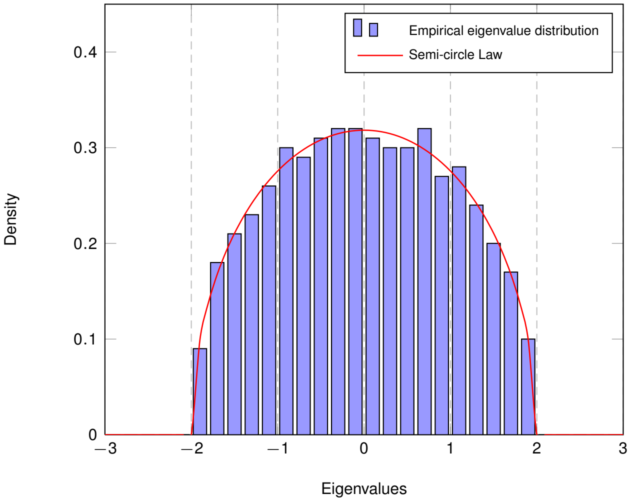


Figure: Histogram of the eigenvalues of Wigner matrices and the semi-circle law, for $N = 500$

Circular law

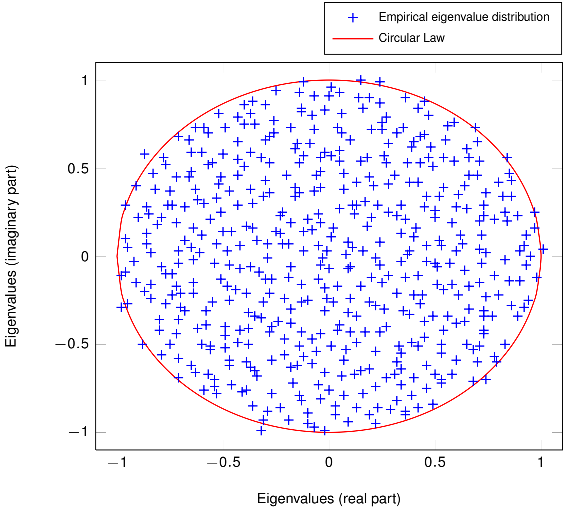


Figure: Eigenvalues of $\mathbf{X}_N$ with i.i.d. standard Gaussian entries, for $N = 500$.

More involved matrix models

- much study has surrounded the Marčenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have a closed-form expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

To study these models, the method of moments is not enough!
A consistent powerful mathematical framework is required.

More involved matrix models

- much study has surrounded the Marčenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have a closed-form expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

To study these models, the method of moments is not enough!
A consistent powerful mathematical framework is required.

More involved matrix models

- much study has surrounded the Marčenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have a closed-form expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

To study these models, the method of moments is not enough!
A consistent powerful mathematical framework is required.

More involved matrix models

- much study has surrounded the Marčenko-Pastur law, the Wigner semi-circle law etc.
- for practical purposes, we often need more general matrix models
  - products and sums of random matrices
  - i.i.d. models with correlation/variance profile
  - distribution of inverses etc.
- for these models, it is often impossible to have a closed-form expression of the limiting distribution.
- sometimes we do not have a limiting convergence.

To study these models, the method of moments is not enough!
A consistent powerful mathematical framework is required.

Outline

Eigenvalue distribution and moments

- The Hermitian matrix $\mathbf{R}_N \in \mathbb{C}^{N\times N}$ has successive *empirical* moments M_k^N, $k = 1, 2, \ldots$,

$$M_k^N = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^k$$

- In classical probability theory, for A, B independent,

$$c_k(A+B) = c_k(A) + c_k(B)$$

with $c_k(X)$ the cumulants of X. The cumulants c_k are connected to the moments m_k by,

$$m_k = \sum_{\pi \in \mathcal{P}(k)} \prod_{V\in\pi} c_{|V|}$$

A natural extension of classical probability for non-commutative random variables exist, called

Free Probability

Eigenvalue distribution and moments

- The Hermitian matrix $\mathbf{R}_N \in \mathbb{C}^{N\times N}$ has successive *empirical* moments M_k^N, $k = 1, 2, \ldots$,

$$M_k^N = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^k$$

- In classical probability theory, for A, B independent,

$$c_k(A+B) = c_k(A) + c_k(B)$$

with $c_k(X)$ the cumulants of X. The cumulants c_k are connected to the moments m_k by,

$$m_k = \sum_{\pi\in\mathcal{P}(k)} \prod_{V\in\pi} c_{|V|}$$

A natural extension of classical probability for non-commutative random variables exist, called

Free Probability

Eigenvalue distribution and moments

- The Hermitian matrix $\mathbf{R}_N \in \mathbb{C}^{N\times N}$ has successive *empirical* moments M_k^N, $k = 1, 2, \ldots$,

$$M_k^N = \frac{1}{N}\sum_{i=1}^{N} \lambda_i^k$$

- In classical probability theory, for A, B independent,

$$c_k(A+B) = c_k(A) + c_k(B)$$

with $c_k(X)$ the cumulants of X. The cumulants c_k are connected to the moments m_k by,

$$m_k = \sum_{\pi\in\mathcal{P}(k)} \prod_{V\in\pi} c_{|V|}$$

A natural extension of classical probability for non-commutative random variables exist, called

Free Probability

Free probability

Free probability applies to asymptotically large random matrices. We denote the moments without superscript.

- To connect the moments of $\mathbf{A}+\mathbf{B}$ to those of $\mathbf{A}$ and $\mathbf{B}$, independence is not enough. $\mathbf{A}$ and $\mathbf{B}$ must be asymptotically free,
  - two Gaussian matrices are free
  - a Gaussian matrix and any deterministic matrix are free
  - unitary (Haar distributed) matrices are free
  - a Haar matrix and a Gaussian matrix are free etc.
- Similarly as in classical probability, we define free cumulants C_k,

$$C_1 = M_1$$
$$C_2 = M_2 - M_1^2$$
$$C_3 = M_3 - 3M_1M_2 + 2M_1^2$$

R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," Mem. A.M.S., vol. 627, 1998.

- Combinatorial description by non-crossing partitions,

$$M_n = \sum_{\pi \in NC(n)} \prod_{V \in \pi} C_{|V|}$$

Free probability

Free probability applies to asymptotically large random matrices. We denote the moments without superscript.

- To connect the moments of $\mathbf{A} + \mathbf{B}$ to those of $\mathbf{A}$ and $\mathbf{B}$, independence is not enough. $\mathbf{A}$ and $\mathbf{B}$ must be asymptotically free,
  - two Gaussian matrices are free
  - a Gaussian matrix and any deterministic matrix are free
  - unitary (Haar distributed) matrices are free
  - a Haar matrix and a Gaussian matrix are free etc.
- Similarly as in classical probability, we define free cumulants C_k,

$$
\begin{aligned}
C_1 &= M_1 \\
C_2 &= M_2 - M_1^2 \\
C_3 &= M_3 - 3M_1M_2 + 2M_1^2
\end{aligned}
$$

R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," Mem. A.M.S., vol. 627, 1998.

- Combinatorial description by non-crossing partitions,

$$M_n = \sum_{\pi \in NC(n)} \prod_{V \in \pi} C_{|V|}$$

Free probability

Free probability applies to asymptotically large random matrices. We denote the moments without superscript.

- To connect the moments of $\mathbf{A}+\mathbf{B}$ to those of $\mathbf{A}$ and $\mathbf{B}$, independence is not enough. $\mathbf{A}$ and $\mathbf{B}$ must be asymptotically free,
  - two Gaussian matrices are free
  - a Gaussian matrix and any deterministic matrix are free
  - unitary (Haar distributed) matrices are free
  - a Haar matrix and a Gaussian matrix are free etc.
- Similarly as in classical probability, we define free cumulants C_k,

$$\begin{aligned} C_1 &= M_1 \\ C_2 &= M_2 - M_1^2 \\ C_3 &= M_3 - 3M_1M_2 + 2M_1^2 \end{aligned}$$

R. Speicher, “Combinatorial theory of the free product with amalgamation and operator-valued free probability theory,” Mem. A.M.S., vol. 627, 1998.

- Combinatorial description by non-crossing partitions,

$$M_n = \sum_{\pi \in NC(n)} \prod_{V \in \pi} C_{|V|}$$

Non-crossing partitions

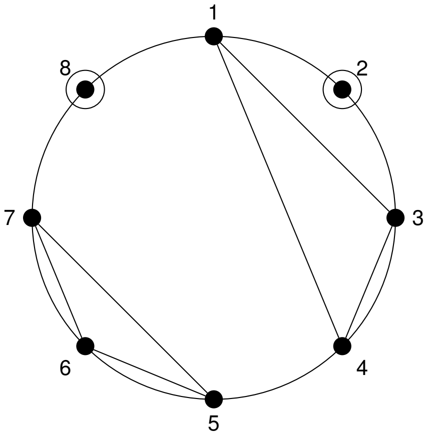


Figure: Non-crossing partition $\pi = \{\{1,3,4\},\{2\},\{5,6,7\},\{8\}\}$ of $NC(8)$.

Moments of sums and products of random matrices

- Combinatorial calculus of all moments

Theorem

For free random matrices $\mathbf{A}$ *and* $\mathbf{B}$, *we have the relationship,*

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

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

in conjunction with free moment-cumulant formula, gives all moments of sum and product.

Theorem

If F *is a compactly supported distribution function, then* F *is determined by its moments.*

- In the absence of support compactness, some conditions (e.g. Carleman) have to be checked. This is in particular the case of Vandermonde matrices.

Moments of sums and products of random matrices

- Combinatorial calculus of all moments

Theorem

For free random matrices $\mathbf{A}$ *and* $\mathbf{B}$*, we have the relationship,*

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

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

in conjunction with free moment-cumulant formula, gives all moments of sum and product.

Theorem

If F is a compactly supported distribution function, then F is determined by its moments.

- In the absence of support compactness, some conditions (e.g. Carleman) have to be checked. This is in particular the case of Vandermonde matrices.

Free convolution

- In classical probability theory, for independent A, B,

$$\mu_{A+B}(x) = \mu_A(x) * \mu_B(x) \triangleq \int \mu_A(t)\mu_B(x-t)dt$$

- In free probability, for free **A**, **B**, we use the notations

$$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}},\ \mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxslash \mu_{\mathbf{B}}$$

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

Theorem

Convolution of the information-plus-noise model Let $\mathbf{W}_N \in \mathbb{C}^{N\times n}$ have i.i.d. Gaussian entries of mean 0 and variance 1, $\mathbf{A}_N \in \mathbb{C}^{N\times n}$, such that $\mu_{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathrm{H}}} \Rightarrow \mu_A$, as $n/N \to c$. Then the eigenvalue distribution of

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)^{\mathrm{H}}$$

converges weakly and almost surely to μ_B such that

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

with μ_c the Marčenko-Pastur law with ratio c.

Free convolution

- In classical probability theory, for independent A, B,

$$\mu_{A+B}(x) = \mu_A(x) * \mu_B(x) \overset{\Delta}{=} \int \mu_A(t)\mu_B(x-t)dt$$

- In free probability, for free **A**, **B**, we use the notations

$$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxminus \mu_{\mathbf{B}},\ \mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}},\ \mu_{\mathbf{A}} = \mu_{\mathbf{A}+\mathbf{B}} \boxslash \mu_{\mathbf{B}}$$

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

Theorem

Convolution of the information-plus-noise model Let $\mathbf{W}_N \in \mathbb{C}^{N\times n}$ *have i.i.d. Gaussian entries of mean* 0 *and variance* 1*,* $\mathbf{A}_N \in \mathbb{C}^{N\times n}$*, such that* $\mu_{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathrm{H}}} \Rightarrow \mu_A$*, as* $n/N \to c$*. Then the eigenvalue distribution of*

$$\mathbf{B}_N = \frac{1}{n}\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)^{\mathrm{H}}$$

converges weakly and almost surely to μ_B *such that*

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

with μ_c *the Marčenko-Pastur law with ratio c.*

Similarities between classical and free probability

	Classical Probability	Free probability
Moments	$m_k = \int x^k dF(x)$	$M_k = \int x^k dF(x)$
Cumulants	$m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{\lvert V \rvert}$	$M_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{V \in \pi} C_{\lvert V \rvert}$
Independence	classical independence	freeness
Additive convolution	$f_{A+B} = f_A * f_B$	$\mu_{\mathbf{A}+\mathbf{B}} = \mu_{\mathbf{A}} \boxplus \mu_{\mathbf{B}}$
Multiplicative convolution	f_{AB}	$\mu_{\mathbf{AB}} = \mu_{\mathbf{A}} \boxtimes \mu_{\mathbf{B}}$
Sum Rule	$c_k(A+B) = c_k(A) + c_k(B)$	$C_k(\mathbf{A}+\mathbf{B}) = C_k(\mathbf{A}) + C_k(\mathbf{B})$
Central Limit	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \to \mathcal{N}(0,1)$	$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \Rightarrow$ semi-circle law

Bibliography on Free Probability related work

- D. Voiculescu, “Addition of certain non-commuting random variables,” Journal of functional analysis, vol. 66, no. 3, pp. 323-346, 1986.
- R. Speicher, “Combinatorial theory of the free product with amalgamation and operator-valued free probability theory,” Mem. A.M.S., vol. 627, 1998.
- R. Seroul, D. O’Shea, “Programming for Mathematicians,” Springer, 2000.
- H. Bercovici, V. Pata, “The law of large numbers for free identically distributed random variables,” The Annals of Probability, pp. 453-465, 1996.
- A. Nica, R. Speicher, “On the multiplication of free N-tuples of noncommutative random variables,” American Journal of Mathematics, pp. 799-837, 1996.
- Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.
- N. R. Rao, A. Edelman, “The polynomial method for random matrices,” Foundations of Computational Mathematics, vol. 8, no. 6, pp. 649-702, 2008.
- 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

The Stieltjes transform

Definition

Let F be a real distribution function. The Stieltjes transform m_F of F is the function defined, for $z \in \mathbb{C} \setminus \mathbb{R}$, as

$$m_F(z) = \int \frac{1}{\lambda - z} dF(\lambda)$$

For $a < b$ real, denoting $z = x + iy$, we have the inverse formula

$$F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x + iy)]$$

Knowing the Stieltjes transform is knowing the eigenvalue distribution!

The Stieltjes transform

Definition

Let F be a real distribution function. The Stieltjes transform m_F of F is the function defined, for $z \in \mathbb{C} \setminus \mathbb{R}$, as

$$m_F(z) = \int \frac{1}{\lambda - z} dF(\lambda)$$

For $a < b$ real, denoting $z = x + iy$, we have the inverse formula

$$F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x + iy)]$$

Knowing the Stieltjes transform is knowing the eigenvalue distribution!

Remark on the Stieltjes transform

- If F is the eigenvalue distribution of a Hermitian matrix $\mathbf{X}_N \in \mathbb{C}^{N\times N}$, we might denote $m_{\mathbf{X}} \triangleq m_F$, and

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N} \operatorname{tr} (\mathbf{X}_N - z\mathbf{I}_N)^{-1}$$

- For compactly supported eigenvalue distribution,

$$m_F(z) = -\frac{1}{z} \int \frac{1}{1 - \frac{\lambda}{z}} = -\sum_{k=0}^{\infty} M_k^N z^{-k-1}$$

The Stieltjes transform is doubly more powerful than the moment approach!

- conveys more information than any K-finite sequence $M_1, \ldots, M_K$.
- is not handicapped by the support compactness constraint.

- however, Stieltjes transform methods, while stronger, are more painful to work with.

Remark on the Stieltjes transform

- If F is the eigenvalue distribution of a Hermitian matrix $\mathbf{X}_N \in \mathbb{C}^{N\times N}$, we might denote $m_{\mathbf{X}} \triangleq m_F$, and

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N} \operatorname{tr} (\mathbf{X}_N - z\mathbf{I}_N)^{-1}$$

- For compactly supported eigenvalue distribution,

$$m_F(z) = -\frac{1}{z} \int \frac{1}{1 - \frac{\lambda}{z}} = -\sum_{k=0}^{\infty} M_k^N z^{-k-1}$$

The Stieltjes transform is doubly more powerful than the moment approach!

- conveys more information than any K-finite sequence $M_1, \ldots, M_K$.
- is not handicapped by the support compactness constraint.

- however, Stieltjes transform methods, while stronger, are more painful to work with.

Remark on the Stieltjes transform

- If F is the eigenvalue distribution of a Hermitian matrix $\mathbf{X}_N \in \mathbb{C}^{N\times N}$, we might denote $m_{\mathbf{X}} \triangleq m_F$, and

$$m_{\mathbf{X}}(z) = \int \frac{1}{\lambda - z} dF(\lambda) = \frac{1}{N} \operatorname{tr}\left(\mathbf{X}_N - z\mathbf{I}_N\right)^{-1}$$

- For compactly supported eigenvalue distribution,

$$m_F(z) = -\frac{1}{z} \int \frac{1}{1 - \frac{\lambda}{z}} = -\sum_{k=0}^{\infty} M_k^N z^{-k-1}$$

The Stieltjes transform is doubly more powerful than the moment approach!

- conveys more information than any K-finite sequence $M_1, \ldots, M_K$.
- is not handicapped by the support compactness constraint.

- however, Stieltjes transform methods, while stronger, are more painful to work with.

Stieltjes transform proof of the Marčenko-Pastur law

- We wish to prove that the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X} \in \mathbb{C}^{N \times n}$, with entries $\mathcal{CN}(0, 1/n)$ tends to the MP law.
- From a matrix inversion lemma

$$\left[(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

with $\mathbf{X}^{\mathsf{H}} = \begin{bmatrix}\mathbf{y} & \mathbf{Y}^{\mathsf{H}}\end{bmatrix}$.
- From the *trace lemma*

$$\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}$$

for all large n.
- From the *rank-1 perturbation lemma*,

$$\frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1}.$$

- Since the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ is the same as that of $\mathbf{X}^{\mathsf{H}}\mathbf{X}$ but for some zeros

$$\frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1} = \frac{1}{n}\operatorname{tr}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} + \frac{N-n}{n}\frac{1}{z}.$$

- Replacing and summing over all diagonal components,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \simeq \frac{1}{1 - \frac{N}{n} - z - z\frac{N}{n}\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}}$$

Stieltjes transform proof of the Marčenko-Pastur law

- We wish to prove that the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X} \in \mathbb{C}^{N\times n}$, with entries $\mathcal{CN}(0,1/n)$ tends to the MP law.
- From a matrix inversion lemma

$$\left[(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

  with $\mathbf{X}^{\mathsf{H}} = \left[\mathbf{y} \quad \mathbf{Y}^{\mathsf{H}}\right]$.
- From the *trace lemma*

$$\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}$$

  for all large n.
- From the *rank-1 perturbation lemma*,

$$\frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1}.$$

- Since the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ is the same as that of $\mathbf{X}^{\mathsf{H}}\mathbf{X}$ but for some zeros

$$\frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1} = \frac{1}{n}\operatorname{tr}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} + \frac{N-n}{n}\frac{1}{z}.$$

- Replacing and summing over all diagonal components,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \simeq \frac{1}{1 - \frac{N}{n} - z - z\frac{N}{n}\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}}$$

Stieltjes transform proof of the Marčenko-Pastur law

- We wish to prove that the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X} \in \mathbb{C}^{N\times n}$, with entries $\mathcal{CN}(0, 1/n)$ tends to the MP law.
- From a matrix inversion lemma

$$\left[(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

  with $\mathbf{X}^{\mathsf{H}} = \left[\mathbf{y} \quad \mathbf{Y}^{\mathsf{H}}\right]$.
- From the *trace lemma*

$$\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}$$

  for all large n.
- From the *rank-1 perturbation lemma*,

$$\frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1}.$$

- Since the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ is the same as that of $\mathbf{X}^{\mathsf{H}}\mathbf{X}$ but for some zeros

$$\frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1} = \frac{1}{n}\operatorname{tr}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} + \frac{N-n}{n}\frac{1}{z}.$$

- Replacing and summing over all diagonal components,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \simeq \frac{1}{1 - \frac{N}{n} - z - z\frac{N}{n}\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}}$$

Stieltjes transform proof of the Marčenko-Pastur law

- We wish to prove that the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X} \in \mathbb{C}^{N \times n}$, with entries $\mathcal{CN}(0, 1/n)$ tends to the MP law.
- From a matrix inversion lemma

$$\left[(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

  with $\mathbf{X}^{\mathsf{H}} = \begin{bmatrix}\mathbf{y} & \mathbf{Y}^{\mathsf{H}}\end{bmatrix}$.
- From the *trace lemma*

$$\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}$$

  for all large n.
- From the *rank-1 perturbation lemma*,

$$\frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1}.$$

- Since the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ is the same as that of $\mathbf{X}^{\mathsf{H}}\mathbf{X}$ but for some zeros

$$\frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1} = \frac{1}{n}\operatorname{tr}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} + \frac{N-n}{n}\frac{1}{z}.$$

- Replacing and summing over all diagonal components,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \simeq \frac{1}{1 - \frac{N}{n} - z - z\frac{N}{n}\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}}$$

Stieltjes transform proof of the Marčenko-Pastur law

- We wish to prove that the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, $\mathbf{X} \in \mathbb{C}^{N\times n}$, with entries $\mathcal{CN}(0, 1/n)$ tends to the MP law.
- From a matrix inversion lemma

$$\left[(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\right]_{11} = \frac{1}{-z - z\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y}}$$

with $\mathbf{X}^{\mathsf{H}} = \begin{bmatrix}\mathbf{y} & \mathbf{Y}^{\mathsf{H}}\end{bmatrix}$.
- From the *trace lemma*

$$\mathbf{y}^{\mathsf{H}}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}\mathbf{y} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1}$$

for all large n.
- From the *rank-1 perturbation lemma*,

$$\frac{1}{n}\operatorname{tr}(\mathbf{Y}^{\mathsf{H}}\mathbf{Y} - z\mathbf{I}_n)^{-1} \simeq \frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1}.$$

- Since the spectrum of $\mathbf{X}\mathbf{X}^{\mathsf{H}}$ is the same as that of $\mathbf{X}^{\mathsf{H}}\mathbf{X}$ but for some zeros

$$\frac{1}{n}\operatorname{tr}(\mathbf{X}^{\mathsf{H}}\mathbf{X} - z\mathbf{I}_n)^{-1} = \frac{1}{n}\operatorname{tr}(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} + \frac{N-n}{n}\frac{1}{z}.$$

- Replacing and summing over all diagonal components,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1} \simeq \frac{1}{1 - \frac{N}{n} - z - z\frac{N}{n}\frac{1}{N}\operatorname{tr}\left(\mathbf{X}\mathbf{X}^{\mathsf{H}} - z\mathbf{I}_N\right)^{-1}}$$

Stieltjes transform proof of the Marčenko-Pastur law (2)

- This is a second order polynomial of the type

$$m_F(z) = \frac{1}{1 - c - z - zcm_F(z)}$$

with solution

$$m_F(z) = \frac{1-c}{2cz} - \frac{1}{2c} - \frac{\sqrt{(1-c-z)^2 - 4cz}}{2cz}$$

- Using the *Stieltjes inversion formula*

$$f(x) \triangleq F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x+iy)]$$

we finally obtain

$$f(x) = (1-c^{-1})^+ \delta(x) + \frac{1}{2\pi cx} \sqrt{(x-a)^+(b-x)^+}$$

with $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$, of support $[a, b]$.

Other asymptotic results using the Stieltjes transform

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

Theorem

Let $\underline{\mathbf{B}}_N = \mathbf{X}_N \mathbf{T}_N \mathbf{X}_N^{\mathsf{H}} \in \mathbb{C}^{N\times N}$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *has i.i.d. entries of mean* 0 *and variance* $1/N$, $F^{\mathbf{T}_N} \Rightarrow F^T$, $n/N \to c$. *Then,* $F^{\underline{\mathbf{B}}_N} \Rightarrow \underline{F}$ *almost surely,* $\underline{F}$ *having Stieltjes transform*

$$m_{\underline{F}}(z) = \left(c \int \frac{t}{1 + t m_{\underline{F}}(z)} dF^T(t) - z \right)^{-1} = \left[\frac{1}{N} \operatorname{tr} \mathbf{T}_N \left(m_{\underline{F}}(z) \mathbf{T}_N + \mathbf{I}_N \right)^{-1} - z \right]^{-1}$$

which has a unique solution $m_{\underline{F}}(z) \in \mathbb{C}^+$ *if* $z \in \mathbb{C}^+$, *and* $m_{\underline{F}}(z) > 0$ *if* $z < 0$.

- in general, no explicit expression for $\underline{F}$.
- Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with asymptotic distribution F,

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

Spectrum of the sample covariance matrix model $\mathbf{B}_N = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}$, with $\mathbf{X}_N^{\mathsf{H}} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbf{x}_i$ i.i.d. with zero mean and covariance $\mathbf{T}_N = \mathrm{E}[\mathbf{x}_1 \mathbf{x}_1^{\mathsf{H}}]$.

Other asymptotic results using the Stieltjes transform

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

Theorem

Let $\underline{\mathbf{B}}_N = \mathbf{X}_N\mathbf{T}_N\mathbf{X}_N^{\mathrm{H}} \in \mathbb{C}^{N\times N}$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *has i.i.d. entries of mean* 0 *and variance* $1/N$, $F^{\mathbf{T}_N} \Rightarrow F^T$, $n/N \to c$. *Then,* $F^{\underline{\mathbf{B}}_N} \Rightarrow \underline{F}$ *almost surely,* $\underline{F}$ *having Stieltjes transform*

$$m_{\underline{F}}(z) = \left(c\int \frac{t}{1+tm_{\underline{F}}(z)}dF^T(t) - z\right)^{-1} = \left[\frac{1}{N}\operatorname{tr}\mathbf{T}_N\left(m_{\underline{F}}(z)\mathbf{T}_N + \mathbf{I}_N\right)^{-1} - z\right]^{-1}$$

which has a unique solution $m_{\underline{F}}(z) \in \mathbb{C}^+$ *if* $z \in \mathbb{C}^+$, *and* $m_{\underline{F}}(z) > 0$ *if* $z < 0$.

- in general, no explicit expression for $\underline{F}$.
- Stieltjes transform of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N^{\mathrm{H}}\mathbf{X}_N\mathbf{T}_N^{\frac{1}{2}}$ with asymptotic distribution F,

$$m_F = cm_{\underline{F}} + (c-1)\frac{1}{z}$$

Spectrum of the sample covariance matrix model $\mathbf{B}_N = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}$, with $\mathbf{X}_N^{\mathrm{H}} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbf{x}_i$ i.i.d. with zero mean and covariance $\mathbf{T}_N = \mathrm{E}[\mathbf{x}_1\mathbf{x}_1^{\mathrm{H}}]$.

Getting F' from m_F

- Remember that

$$f(x) \triangleq F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x+iy)]$$

- to plot the density $f(x)$, span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.

Example (Sample covariance matrix)

For N multiple of 3, let $dF^T(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathrm{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

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

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Getting F' from m_F

- Remember that

$$f(x) \triangleq F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x+iy)]$$

- to plot the density $f(x)$, span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.

Example (Sample covariance matrix)

For N multiple of 3, let $dF^T(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

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

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Getting F' from m_F

- Remember that

$$f(x) \triangleq F'(x) = \lim_{y \to 0} \frac{1}{\pi} \Im[m_F(x+iy)]$$

- to plot the density $f(x)$, span $z = x + iy$ on the line $\{x \in \mathbb{R}, y = \varepsilon\}$ parallel but close to the real axis, solve $m_F(z)$ for each z, and plot $\Im[m_F(z)]$.

Example (Sample covariance matrix)

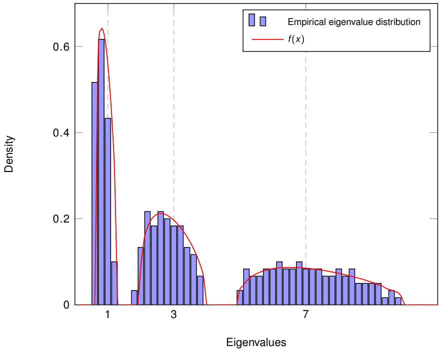
For N multiple of 3, let $dF^T(x) = \frac{1}{3}\delta(x-1) + \frac{1}{3}\delta(x-3) + \frac{1}{3}\delta(x-K)$ and let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathrm{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ with $F^{\mathbf{B}_N} \to F$, then

$$m_F = c m_{\underline{F}} + (c-1)\frac{1}{z}$$

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

We take $c = 1/10$ and alternatively $K = 7$ and $K = 4$.

Spectrum of the sample covariance matrix



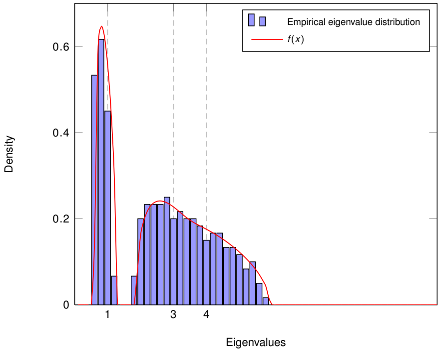


Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$, $N = 3000$, $n = 300$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

The Shannon Transform

A. M. Tulino, S. Verdù, “Random matrix theory and wireless communications,” Now Publishers Inc., 2004.

Definition

Let F be a probability distribution, m_F its Stieltjes transform, then the Shannon-transform $\mathcal{V}_F$ of F is defined as

$$\mathcal{V}_F(x) \triangleq \int_0^\infty \log(1 + x\lambda) dF(\lambda) = \int_x^\infty \left(\frac{1}{t} - m_F(-t) \right) dt$$

If F is the distribution function of the eigenvalues of $\mathbf{X}\mathbf{X}^\mathsf{H} \in \mathbb{C}^{N \times N}$,

$$\mathcal{V}_F(x) = \frac{1}{N} \log\det \left(\mathbf{I}_N + x\mathbf{X}\mathbf{X}^\mathsf{H} \right).$$

Note that this last relation is fundamental to wireless communication purposes!

Outline

No eigenvalues outside the support!

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

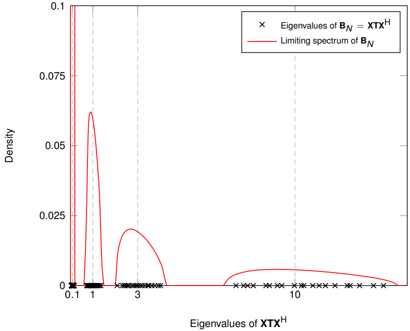
- We showed that the eigenvalue distribution $F^{\mathbf{B}_N}$ of $\mathbf{B}_N = \mathbf{XTX}^{\mathsf{H}}$, $F^{\mathbf{T}_N} \Rightarrow F^T$:
  - is similar to a deterministic F_N
  - sometimes converges WEAKLY to F with $\mathrm{Supp}(F)$ made of compact sets.
- There is more:

For all N_0, there is no eigenvalue of $\mathbf{B}_N$ outside $\mathrm{Supp}(F) \cup \bigcup_{N \geq N_0} \mathrm{Supp}(F_N)$, for all large N.

No eigenvalues outside the support!

Z. Bai, J. Silverstein, “No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices,” Annals of Prob., vol. 26, no.1 pp. 316-345, 1998.

- We showed that the eigenvalue distribution $F^{\mathbf{B}_N}$ of $\mathbf{B}_N = \mathbf{XTX}^H$, $F^{\mathbf{T}_N} \Rightarrow F^T$:
  - is similar to a deterministic F_N
  - sometimes converges WEAKLY to F with $\mathrm{Supp}(F)$ made of compact sets.
- There is more:

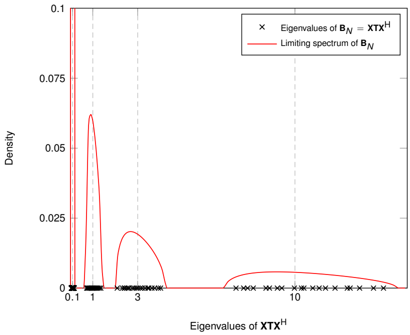


For all N_0, there is no eigenvalue of $\mathbf{B}_N$ outside $\mathrm{Supp}(F) \cup \bigcup_{N \geq N_0} \mathrm{Supp}(F_N)$, for all large N.

No eigenvalues outside the support!

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

- We showed that the eigenvalue distribution $F^{\mathbf{B}_N}$ of $\mathbf{B}_N = \mathbf{XTX}^H$, $F^{\mathbf{T}_N} \Rightarrow F^T$:
  - is similar to a deterministic F_N
  - sometimes converges WEAKLY to F with $\mathrm{Supp}(F)$ made of compact sets.
- There is more:



For all N_0, there is no eigenvalue of $\mathbf{B}_N$ outside $\mathrm{Supp}(F) \cup \bigcup_{N \geq N_0} \mathrm{Supp}(F_N)$, for all large N.

The spiked model

- For $\mathbf{T}$ composed of finitely many eigenvalues with large multiplicities (e.g. $\mathbf{T} = \mathbf{I}_N$), no eigenvalue of $\mathbf{B}_N$ outside $\mathrm{Supp}(F)$.
- If, for r fixed, $\mathbf{T}$ is a rank-r perturbation of $\mathbf{I}_N$,

$$\mathrm{diag}(\underbrace{1,\ldots,1}_{\text{multiplicity } (N-r)},1+\omega_1,\ldots,1+\omega_r)$$

then, depending on whether $\omega_i > \sqrt{N/n}$,

Figure: Eigenvalues of $\mathbf{B}_N = \mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$, $\mathbf{T}$ diagonal of 1's but for the last four entries set to $\{3,3,2,2\}$. On top, $N = 500$, $n = 1500$. At bottom, $N = 500$, $n = 400$. Theoretical limit eigenvalues of $\mathbf{B}_N$ are stressed.

The spiked model

- For $\mathbf{T}$ composed of finitely many eigenvalues with large multiplicities (e.g. $\mathbf{T} = \mathbf{I}_N$), no eigenvalue of $\mathbf{B}_N$ outside $\mathrm{Supp}(F)$.
- If, for *r* fixed, $\mathbf{T}$ is a rank-r perturbation of $\mathbf{I}_N$,

$$\mathrm{diag}(\underbrace{1,\ldots,1}_{\text{multiplicity } (N-r)}, 1+\omega_1,\ldots,1+\omega_r)$$

then, depending on whether $\omega_i > \sqrt{N/n}$,

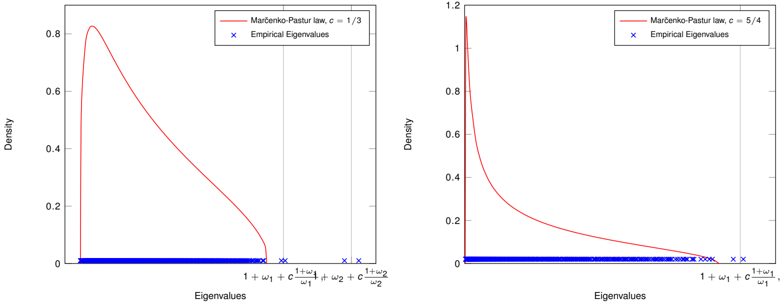


Figure: Eigenvalues of $\mathbf{B}_N = \mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{T}^{\frac{1}{2}}$, $\mathbf{T}$ diagonal of 1's but for the last four entries set to $\{3, 3, 2, 2\}$. On top, $N = 500$, $n = 1500$. At bottom, $N = 500$, $n = 400$. Theoretical limit eigenvalues of $\mathbf{B}_N$ are stressed.

Limits for the spiked models

J. Baik and 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.
D. Paul, “Asymptotics of sample eigenstructure for a large dimensional spiked covariance model,” Statistica Sinica, vol. 17, no. 4, pp. 1617, 2007.

- Assume **T** as above with:
  - $\omega_1 > \cdots > \omega_r > 0$ the population spikes
  - $\mathbf{u}_1, \ldots, \mathbf{u}_r \in \mathbb{C}^N$, the associated population eigenvectors
  - $\hat{\lambda}_1 > \ldots > \hat{\lambda}_r$ the largest eigenvalues of $\mathbf{B}_N$
  - $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_r$ the associated sample eigenvalues
- Then, with $\lim N/n = c$, we have the first order limits:

$$\hat{\lambda}_k \xrightarrow{\text{a.s.}} \begin{cases} 1 + \omega_k + c\frac{1+\omega_k}{\omega_k} & , \omega_k > \sqrt{c} \\ (1+\sqrt{c})^2 & , \omega_k \leq \sqrt{c} \end{cases} .$$

$$|\hat{\mathbf{u}}_k^* \mathbf{u}_k|^2 \xrightarrow{\text{a.s.}} \begin{cases} \frac{1-c\omega_k^{-2}}{1+c\omega_k^{-1}} & , \omega_k > \sqrt{c} \\ 0 & , \omega_k \leq \sqrt{c} \end{cases} .$$

Limits for the spiked models

J. Baik and 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.
D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," Statistica Sinica, vol. 17, no. 4, pp. 1617, 2007.

- Assume **T** as above with:
  - $\omega_1 > \cdots > \omega_r > 0$ the population spikes
  - $\mathbf{u}_1, \ldots, \mathbf{u}_r \in \mathbb{C}^N$, the associated population eigenvectors
  - $\hat{\lambda}_1 > \ldots > \hat{\lambda}_r$ the largest eigenvalues of $\mathbf{B}_N$
  - $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_r$ the associated sample eigenvalues
- Then, with $\lim N/n = c$, we have the first order limits:

$$\hat{\lambda}_k \xrightarrow{\text{a.s.}} \begin{cases} 1 + \omega_k + c\frac{1+\omega_k}{\omega_k} & , \omega_k > \sqrt{c} \\ (1+\sqrt{c})^2 & , \omega_k \leq \sqrt{c} \end{cases} .$$

$$|\hat{\mathbf{u}}_k^* \mathbf{u}_k|^2 \xrightarrow{\text{a.s.}} \begin{cases} \frac{1-c\omega_k^{-2}}{1+c\omega_k^{-1}} & , \omega_k > \sqrt{c} \\ 0 & , \omega_k \leq \sqrt{c} \end{cases} .$$

Second order limits for the spiked models

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.
J. Baik and G. Ben Arous and S. Péché, "Phase transition of the largest eigenvalue for non-null complex sample covariance matrices," The Annals of Prob., vol. 33, no. 5, pp. 1643-1697, 2005.
R. Couillet and W. Hachem, "Local failure detection and identification in large sensor networks," *submitted to* IEEE Transaction on Information Theory, 2011.

- As well as the second order limits in the Gaussian case:
  - If $\omega_k > \sqrt{c}$

$$\sqrt{N}\begin{pmatrix} |\mathbf{u}_k^* \hat{\mathbf{u}}_k|^2 - \left[\frac{1-c\omega_k^{-2}}{1+c\omega_k^{-1}}\right] \\ \hat{\lambda}_k - \left[1+\omega_k + c\frac{1+\omega_k}{\omega_k}\right] \end{pmatrix} \Rightarrow \mathcal{CN}\left(0, \begin{bmatrix} \frac{c^2(1+\omega_k)^2}{(c+\omega_k)^2(\omega_k^2-c)}\left(c\frac{(1+\omega_k)^2}{(c+\omega_k)^2}+1\right) & \frac{(1+\omega_k)^3 c^2}{(\omega_k+c)^2\omega_k} \\ \frac{(1+\omega_k)^3 c^2}{(\omega_k+c)^2\omega_k} & \frac{c(1+\omega_k)^2(\omega_k^2-c)}{\omega_k^2} \end{bmatrix}\right)$$

  - If $\omega_k < \sqrt{c}$

$$N^{\frac{2}{3}}\frac{\hat{\lambda}_k - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}}\sqrt{c}} \Rightarrow T_2$$

with T_2 the complex Tracy-Widom distribution.

Second order limits for the spiked models

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.
J. Baik and G. Ben Arous and S. Péché, “Phase transition of the largest eigenvalue for non-null complex sample covariance matrices,” The Annals of Prob., vol. 33, no. 5, pp. 1643-1697, 2005.
R. Couillet and W. Hachem, “Local failure detection and identification in large sensor networks,” *submitted to* IEEE Transaction on Information Theory, 2011.

- As well as the second order limits in the Gaussian case:
  - If $\omega_k > \sqrt{c}$

$$\sqrt{N}\begin{pmatrix} |\mathbf{u}_k^* \hat{\mathbf{u}}_k|^2 - \left[\frac{1-c\omega_k^{-2}}{1+c\omega_k^{-1}}\right] \\ \hat{\lambda}_k - \left[1+\omega_k+c\frac{1+\omega_k}{\omega_k}\right] \end{pmatrix} \Rightarrow \mathcal{CN}\left(0, \begin{bmatrix} \frac{c^2(1+\omega_k)^2}{(c+\omega_k)^2(\omega_k^2-c)}\left(c\frac{(1+\omega_k)^2}{(c+\omega_k)^2}+1\right) & \frac{(1+\omega_k)^3c^2}{(\omega_k+c)^2\omega_k} \\ \frac{(1+\omega_k)^3c^2}{(\omega_k+c)^2\omega_k} & \frac{c(1+\omega_k)^2(\omega_k^2-c)}{\omega_k^2} \end{bmatrix}\right)$$

  - If $\omega_k < \sqrt{c}$

$$N^{\frac{2}{3}}\frac{\hat{\lambda}_k-(1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}}\sqrt{c}} \Rightarrow T_2$$

  with T_2 the complex Tracy-Widom distribution.

Second order statistics, $\omega_k < \sqrt{c}$

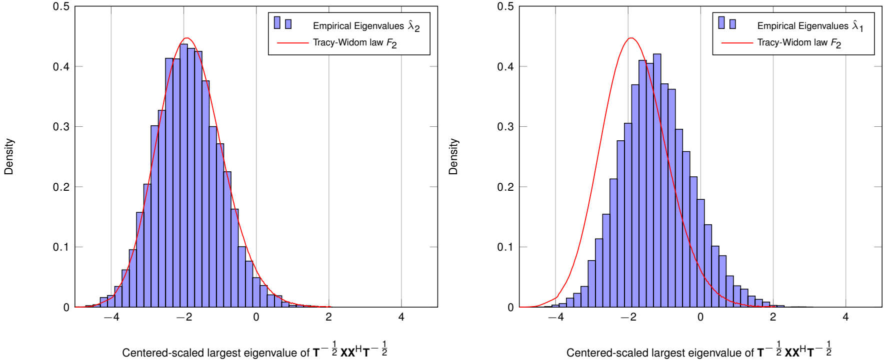


Figure: Distribution of $N^{\frac{2}{3}}c^{-\frac{1}{2}}(1+\sqrt{c})^{-\frac{4}{3}}\left[\hat{\lambda}_k - (1+\sqrt{c})^2\right]$ against the Tracy-Widom law for $N = 500$, $n = 1500$, $c = 1/3$, $\mathbf{T} = \mathrm{diag}(1, \ldots, 1, 1.5)$ $(0.5 < \sqrt{c})$. Empirical distribution taken over $10,000$ Monte-Carlo simulations.

Second order statistics, $\omega_k > \sqrt{c}$

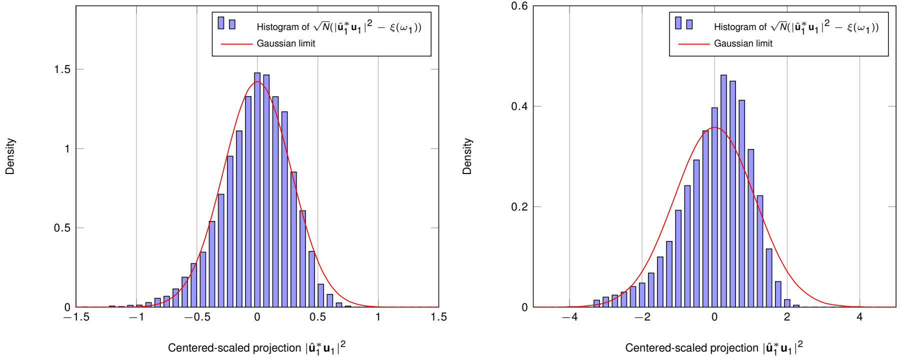


Figure: Empirical and theoretical distribution of the fluctuations of $\hat{\mathbf{u}}_1$ if X has i.i.d. $\mathcal{CN}(0, 1/n)$ entries, $N/n = 1/8$, $N = 64$, $\omega_1 = 1$ (left) or $\omega_1 = 0.5$ (right).

Outline

Models studied with analytic tools

- *Stieltjes transform:* models involving i.i.d. matrices
  - sample covariance matrix models, $\mathbf{XTX}^{\mathsf{H}}$ and $\mathbf{T}^{\frac{1}{2}}\mathbf{X}^{\mathsf{H}}\mathbf{XT}^{\frac{1}{2}}$
  - doubly correlated models, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$. With $\mathbf{X}$ Gaussian, Kronecker model.
  - doubly correlated models with external matrix, $\mathbf{R}^{\frac{1}{2}}\mathbf{XTX}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} + \mathbf{A}$.
  - variance profile, $\mathbf{XX}^{\mathsf{H}}$, where $\mathbf{X}$ has i.i.d. entries with mean 0, variance $\sigma^2_{i,j}$.
  - Ricean channels, $\mathbf{XX}^{\mathsf{H}} + \mathbf{A}$, where $\mathbf{X}$ has a variance profile.
  - sum of doubly correlated i.i.d. matrices, $\sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$.
  - information-plus-noise models $(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}}$
  - frequency-selective doubly-correlated channels $(\sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})(\sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})$
  - sum of frequency-selective doubly-correlated channels $\sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}}\mathbf{H}_k\mathbf{T}_k\mathbf{H}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$, where $\mathbf{H}_k = \sum_{l=1}^L \mathbf{R}_{kl}'^{\frac{1}{2}}\mathbf{X}_{kl}\mathbf{T}_{kl}'\mathbf{X}_{kl}^{\mathsf{H}}\mathbf{R}_{kl}'^{\frac{1}{2}}$.
- *R- and S-transforms:* models involving a column subset $\mathbf{W}$ of unitary matrices
  - doubly correlated Haar matrix $\mathbf{R}^{\frac{1}{2}}\mathbf{WTW}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$
  - sum of simply correlated Haar matrices $\sum_{k=1}^K \mathbf{W}_k\mathbf{T}_k\mathbf{W}_k^{\mathsf{H}}$

In most cases, $\mathbf{T}$ and $\mathbf{R}$ can be taken random, but independent of $\mathbf{X}$. More involved random matrices, such as Vandermonde matrices, were not yet studied.

Models studied with analytic tools

- *Stieltjes transform:* models involving i.i.d. matrices
  - sample covariance matrix models, $\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}$ and $\mathbf{T}^{\frac{1}{2}}\mathbf{X}^{\mathsf{H}}\mathbf{X}\mathbf{T}^{\frac{1}{2}}$
  - doubly correlated models, $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$. With $\mathbf{X}$ Gaussian, Kronecker model.
  - doubly correlated models with external matrix, $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} + \mathbf{A}$.
  - variance profile, $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, where $\mathbf{X}$ has i.i.d. entries with mean 0, variance $\sigma_{i,j}^2$.
  - Ricean channels, $\mathbf{X}\mathbf{X}^{\mathsf{H}} + \mathbf{A}$, where $\mathbf{X}$ has a variance profile.
  - sum of doubly correlated i.i.d. matrices, $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$.
  - information-plus-noise models $(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}}$
  - frequency-selective doubly-correlated channels $(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})$
  - sum of frequency-selective doubly-correlated channels $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{H}_k\mathbf{T}_k\mathbf{H}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$, where $\mathbf{H}_k = \sum_{l=1}^{L} \mathbf{R}_{kl}'^{\frac{1}{2}}\mathbf{X}_{kl}\mathbf{T}_{kl}'\mathbf{X}_{kl}^{\mathsf{H}}\mathbf{R}_{kl}'^{\frac{1}{2}}$.
- *R- and S-transforms:* models involving a column subset $\mathbf{W}$ of unitary matrices
  - doubly correlated Haar matrix $\mathbf{R}^{\frac{1}{2}}\mathbf{W}\mathbf{T}\mathbf{W}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$
  - sum of simply correlated Haar matrices $\sum_{k=1}^{K} \mathbf{W}_k\mathbf{T}_k\mathbf{W}_k^{\mathsf{H}}$

In most cases, $\mathbf{T}$ and $\mathbf{R}$ can be taken random, but independent of $\mathbf{X}$. More involved random matrices, such as Vandermonde matrices, were not yet studied.

Models studied with analytic tools

- *Stieltjes transform:* models involving i.i.d. matrices
  - sample covariance matrix models, $\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}$ and $\mathbf{T}^{\frac{1}{2}}\mathbf{X}^{\mathsf{H}}\mathbf{X}\mathbf{T}^{\frac{1}{2}}$
  - doubly correlated models, $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$. With $\mathbf{X}$ Gaussian, Kronecker model.
  - doubly correlated models with external matrix, $\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}} + \mathbf{A}$.
  - variance profile, $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, where $\mathbf{X}$ has i.i.d. entries with mean 0, variance $\sigma_{i,j}^2$.
  - Ricean channels, $\mathbf{X}\mathbf{X}^{\mathsf{H}} + \mathbf{A}$, where $\mathbf{X}$ has a variance profile.
  - sum of doubly correlated i.i.d. matrices, $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$.
  - information-plus-noise models $(\mathbf{X}+\mathbf{A})(\mathbf{X}+\mathbf{A})^{\mathsf{H}}$
  - frequency-selective doubly-correlated channels $(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})(\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k\mathbf{R}_k^{\frac{1}{2}})$
  - sum of frequency-selective doubly-correlated channels $\sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}}\mathbf{H}_k\mathbf{T}_k\mathbf{H}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$, where $\mathbf{H}_k = \sum_{l=1}^{L} \mathbf{R}_{kl}'^{\frac{1}{2}}\mathbf{X}_{kl}\mathbf{T}_{kl}'\mathbf{X}_{kl}^{\mathsf{H}}\mathbf{R}_{kl}'^{\frac{1}{2}}$.
- *R- and S-transforms:* models involving a column subset $\mathbf{W}$ of unitary matrices
  - doubly correlated Haar matrix $\mathbf{R}^{\frac{1}{2}}\mathbf{W}\mathbf{T}\mathbf{W}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}$
  - sum of simply correlated Haar matrices $\sum_{k=1}^{K} \mathbf{W}_k\mathbf{T}_k\mathbf{W}_k^{\mathsf{H}}$

In most cases, $\mathbf{T}$ and $\mathbf{R}$ can be taken random, but independent of $\mathbf{X}$. More involved random matrices, such as Vandermonde matrices, were not yet studied.

Models studied with moments/free probability

- asymptotic results
  - most of the above models with Gaussian $\mathbf{X}$.
  - products $\mathbf{V}_1\mathbf{V}_1^{\mathrm{H}}\mathbf{T}_1\mathbf{V}_2\mathbf{V}_2^{\mathrm{H}}\mathbf{T}_2\ldots$ of Vandermonde and deterministic matrices
  - *conjecture:* any probability space of matrices invariant to row or column permutations.
- marginal studies, not yet fully explored
  - rectangular free convolution: singular values of rectangular matrices
  - finite size models. Instead of almost sure convergence of $m_{\mathbf{X}_N}$ as $N \to \infty$, we can study finite size behaviour of $\mathrm{E}[m_{\mathbf{X}_N}]$.

Models studied with moments/free probability

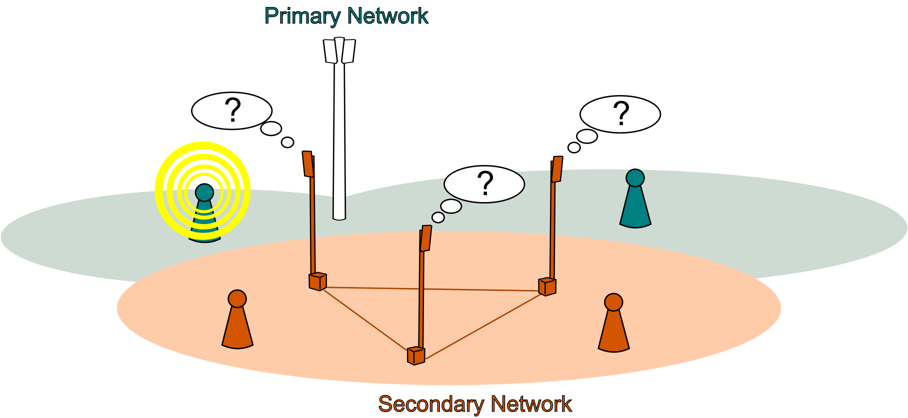
- asymptotic results
  - most of the above models with Gaussian $\mathbf{X}$.
  - products $\mathbf{V}_1\mathbf{V}_1^{\mathrm{H}}\mathbf{T}_1\mathbf{V}_2\mathbf{V}_2^{\mathrm{H}}\mathbf{T}_2\ldots$ of Vandermonde and deterministic matrices
  - *conjecture:* any probability space of matrices invariant to row or column permutations.
- marginal studies, not yet fully explored
  - rectangular free convolution: singular values of rectangular matrices
  - finite size models. Instead of almost sure convergence of $m_{\mathbf{X}_N}$ as $N \to \infty$, we can study finite size behaviour of $\mathrm{E}[m_{\mathbf{X}_N}]$.

Related bibliography

- R. B. Dozier, J. W. Silverstein, “On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices,” Journal of Multivariate Analysis, vol. 98, no. 4, pp. 678-694, 2007.
- J. W. Silverstein, Z. D. Bai, “On the empirical distribution of eigenvalues of a class of large dimensional random matrices,” Journal of Multivariate Analysis, vol. 54, no. 2, pp. 175-192, 1995.
- 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.
- F. Benaych-Georges, “Rectangular random matrices, related free entropy and free Fisher’s information,” Arxiv preprint math/0512081, 2005.
- Ø. Ryan, M. Debbah, “Multiplicative free convolution and information-plus-noise type matrices,” Arxiv preprint math.PR/0702342, 2007.
- V. L. Girko, “Theory of Random Determinants,” Kluwer, Dordrecht, 1990.
- R. Couillet, M. Debbah, J. W. Silverstein, “A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels,” *submitted to IEEE Trans. on Information Theory*.
- V. L. Girko, “Theory of Random Determinants,” Kluwer, Dordrecht, 1990.
- W. Hachem, Ph. Loubaton, J. Najim, “Deterministic Equivalents for Certain Functionals of Large Random Matrices”, Annals of Applied Probability, vol. 17, no. 3, 2007.
- M. J. M. Peacock, I. B. Collings, M. L. Honig, “Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions,” IEEE Trans. on Information Theory, vol. 54, no. 5, pp. 2123, 2008.
- D. Petz, J. Réffy, “On Asymptotics of large Haar distributed unitary matrices,” Periodica Math. Hungar., vol. 49, pp. 103-117, 2004.
- Ø. Ryan, A. Masucci, S. Yang, M. Debbah, “Finite dimensional statistical inference,” *submitted to IEEE Trans. on Information Theory, Dec. 2009.*

Outline

Signal Sensing in Cognitive Radios



Outline

Problem formulation

- Assume the scenario of
  - an *hypothetical* signal source $\sqrt{P}\mathbf{x} \in \mathbb{C}^n$ of power P
  - a transfer channel $\mathbf{H} \in \mathbb{C}^{N\times n}$
  - a sensor network of n sensors
  - additive noise $\sigma\mathbf{w} \in \mathbb{C}^N$ of variance $\sigma^2\mathbf{I}_N$.
- We consider the following *hypothesis test*

$$\mathbf{y}^{(m)} = \begin{cases} \sigma\mathbf{w}^{(m)} & ,(\mathcal{H}_0) \\ \sqrt{P}\mathbf{H}\mathbf{x}^{(m)} + \sigma\mathbf{w}^{(m)} & ,(\mathcal{H}_1) \end{cases}$$

- We wish to confront the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ given the data matrix $\mathbf{Y} \triangleq [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$.
- We consider, in a Bayesian framework, the Neyman-Pearson test ratio

$$C(\mathbf{Y}) \triangleq \frac{P_{\mathcal{H}_1|\mathbf{Y},I}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y},I}(\mathbf{Y})}$$

with prior information I on $\mathbf{H}, \mathbf{x}^{(m)}, \sigma, \ldots$.

Problem formulation

- Assume the scenario of
  - an *hypothetical* signal source $\sqrt{P}\mathbf{x} \in \mathbb{C}^n$ of power P
  - a transfer channel $\mathbf{H} \in \mathbb{C}^{N\times n}$
  - a sensor network of n sensors
  - additive noise $\sigma\mathbf{w} \in \mathbb{C}^N$ of variance $\sigma^2\mathbf{I}_N$.
- We consider the following *hypothesis test*

$$\mathbf{y}^{(m)} = \begin{cases} \sigma\mathbf{w}^{(m)} & , (\mathcal{H}_0) \\ \sqrt{P}\mathbf{H}\mathbf{x}^{(m)} + \sigma\mathbf{w}^{(m)} & , (\mathcal{H}_1) \end{cases}$$

- We wish to confront the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ given the data matrix $\mathbf{Y} \triangleq [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$.
- We consider, in a Bayesian framework, the Neyman-Pearson test ratio

$$C(\mathbf{Y}) \triangleq \frac{P_{\mathcal{H}_1|\mathbf{Y},I}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y},I}(\mathbf{Y})}$$

with prior information I on $\mathbf{H}, \mathbf{x}^{(m)}, \sigma, \ldots$.

Problem formulation

- Assume the scenario of
  - an *hypothetical* signal source $\sqrt{P}\mathbf{x} \in \mathbb{C}^n$ of power P
  - a transfer channel $\mathbf{H} \in \mathbb{C}^{N\times n}$
  - a sensor network of n sensors
  - additive noise $\sigma\mathbf{w} \in \mathbb{C}^N$ of variance $\sigma^2\mathbf{I}_N$.
- We consider the following *hypothesis test*

$$\mathbf{y}^{(m)} = \begin{cases} \sigma\mathbf{w}^{(m)} & , (\mathcal{H}_0) \\ \sqrt{P}\mathbf{H}\mathbf{x}^{(m)} + \sigma\mathbf{w}^{(m)} & , (\mathcal{H}_1) \end{cases}$$

- We wish to confront the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ given the data matrix $\mathbf{Y} \triangleq [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$.
- We consider, in a Bayesian framework, the Neyman-Pearson test ratio

$$C(\mathbf{Y}) \triangleq \frac{P_{\mathcal{H}_1|\mathbf{Y},I}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y},I}(\mathbf{Y})}$$

with prior information I on $\mathbf{H}, \mathbf{x}^{(m)}, \sigma, \ldots$.

Problem formulation

- Assume the scenario of
  - an *hypothetical* signal source $\sqrt{P}\mathbf{x} \in \mathbb{C}^n$ of power P
  - a transfer channel $\mathbf{H} \in \mathbb{C}^{N\times n}$
  - a sensor network of n sensors
  - additive noise $\sigma\mathbf{w} \in \mathbb{C}^N$ of variance $\sigma^2\mathbf{I}_N$.
- We consider the following *hypothesis test*

$$\mathbf{y}^{(m)} = \begin{cases} \sigma\mathbf{w}^{(m)} & ,(\mathcal{H}_0) \\ \sqrt{P}\mathbf{H}\mathbf{x}^{(m)} + \sigma\mathbf{w}^{(m)} & ,(\mathcal{H}_1) \end{cases}$$

- We wish to confront the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ given the data matrix $\mathbf{Y} \triangleq [\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$.
- We consider, in a Bayesian framework, the Neyman-Pearson test ratio

$$C(\mathbf{Y}) \triangleq \frac{P_{\mathcal{H}_1|\mathbf{Y},I}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y},I}(\mathbf{Y})}$$

with prior information I on $\mathbf{H},\mathbf{x}^{(m)},\sigma,\ldots$.

A Bayesian framework for cognitive radios

- We assume prior statistical and deterministic knowledge I on $\mathbf{H}, \sigma, P$
- Using the maximum entropy principle (MaxEnt), a prior $P_{(\mathbf{H},\sigma,P)}(\mathbf{H}, \sigma, P)$ can be derived

$$P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y}) = \int_{(\mathbf{H},\sigma,P)} P_{\mathbf{Y}|\mathcal{H}_i,I,\mathbf{H},\sigma,P}(\mathbf{Y}) P_{(\mathbf{H},\sigma,P)}(\mathbf{H}, \sigma, P) d(\mathbf{H}, \sigma, P)$$

- In the following,
  - we derive the case $P = 1$, σ known and the knowledge about $\mathbf{H}$ conveys unitary invariance
    - $\mathrm{E}[\mathrm{tr}\, \mathbf{H}\mathbf{H}^{\mathsf{H}}]$ known: this is what we assume here;
    - $\mathrm{E}[\mathbf{H}\mathbf{H}^{\mathsf{H}}] = \mathbf{Q}$ unknown but such that $\mathrm{E}[\mathrm{tr}\, \mathbf{Q}]$ is known;
    - $\mathrm{rank}(\mathbf{H}\mathbf{H}^{\mathsf{H}})$ known.
  - we compare alternative methods when $P = 1$ and σ are unknown.

A Bayesian framework for cognitive radios

- We assume prior statistical and deterministic knowledge I on $\mathbf{H}, \sigma, P$
- Using the maximum entropy principle (MaxEnt), a prior $P_{(\mathbf{H},\sigma,P)}(\mathbf{H},\sigma,P)$ can be derived

$$P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y}) = \int_{(\mathbf{H},\sigma,P)} P_{\mathbf{Y}|\mathcal{H}_i,I,\mathbf{H},\sigma,P}(\mathbf{Y}) P_{(\mathbf{H},\sigma,P)}(\mathbf{H},\sigma,P) d(\mathbf{H},\sigma,P)$$

- In the following,
  - we derive the case $P = 1$, σ known and the knowledge about $\mathbf{H}$ conveys unitary invariance
    - $\mathrm{E}[\mathrm{tr}\,\mathbf{H}\mathbf{H}^{\mathrm{H}}]$ known: this is what we assume here;
    - $\mathrm{E}[\mathbf{H}\mathbf{H}^{\mathrm{H}}] = \mathbf{Q}$ unknown but such that $\mathrm{E}[\mathrm{tr}\,\mathbf{Q}]$ is known;
    - $\mathrm{rank}(\mathbf{H}\mathbf{H}^{\mathrm{H}})$ known.
  - we compare alternative methods when $P = 1$ and σ are unknown.

A Bayesian framework for cognitive radios

- We assume prior statistical and deterministic knowledge I on $\mathbf{H}, \sigma, P$
- Using the maximum entropy principle (MaxEnt), a prior $P_{(\mathbf{H},\sigma,P)}(\mathbf{H}, \sigma, P)$ can be derived

$$P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y}) = \int_{(\mathbf{H},\sigma,P)} P_{\mathbf{Y}|\mathcal{H}_i,I,\mathbf{H},\sigma,P}(\mathbf{Y}) P_{(\mathbf{H},\sigma,P)}(\mathbf{H}, \sigma, P) d(\mathbf{H}, \sigma, P)$$

- In the following,
  - we derive the case $P = 1$, σ known and the knowledge about **H** conveys unitary invariance
    - $\mathrm{E}[\mathrm{tr}\, \mathbf{H}\mathbf{H}^{\mathsf{H}}]$ known: this is what we assume here;
    - $\mathrm{E}[\mathbf{H}\mathbf{H}^{\mathsf{H}}] = \mathbf{Q}$ unknown but such that $\mathrm{E}[\mathrm{tr}\, \mathbf{Q}]$ is known;
    - $\mathrm{rank}(\mathbf{H}\mathbf{H}^{\mathsf{H}})$ known.
  - we compare alternative methods when $P = 1$ and σ are unknown.

Evaluation of $P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y})$

- by MaxEnt, **X**, **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.
- **Under** $\mathcal{H}_0$:
  - $\mathbf{Y} = \sigma \mathbf{W}$

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

- **Under** $\mathcal{H}_1$:
  - $\mathbf{Y} = \begin{bmatrix}\sqrt{P}\mathbf{H} & \sigma\mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}\geq 0} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma})P_{\mathbf{\Sigma}}(\mathbf{\Sigma})d\mathbf{\Sigma}$$

with $\mathbf{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$.
From unitary invariance of **H**, denoting $\mathbf{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$, $\operatorname{diag}(\mathbf{G}) = (g_1,\ldots,g_n,\sigma^2,\ldots,\sigma^2)$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times(\sigma^2,\infty)^n} P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}(\mathbf{Y},\mathbf{U},\mathbf{G})P_{\mathbf{U}}(\mathbf{U})P_{(g_1,\ldots,g_n)}(g_1,\ldots,g_n)d\mathbf{U}dg_1\ldots dg_n$$

where
- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
- $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);
- if **H** is Gaussian, $P_{(g_1-\sigma^2,\ldots,g_n-\sigma^2)}$ is the joint eigenvalue distribution of a central Wishart;

Evaluation of $P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y})$

- by MaxEnt, **X**, **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.
- **Under $\mathcal{H}_0$:**
  - $\mathbf{Y} = \sigma \mathbf{W}$

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

- **Under $\mathcal{H}_1$:**
  - $\mathbf{Y} = \begin{bmatrix}\sqrt{P}\mathbf{H} & \sigma \mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}\geq 0} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma})P_{\mathbf{\Sigma}}(\mathbf{\Sigma})d\mathbf{\Sigma}$$

with $\mathbf{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$.
From unitary invariance of **H**, denoting $\mathbf{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$, $\operatorname{diag}(\mathbf{G}) = (g_1,\ldots,g_n,\sigma^2,\ldots,\sigma^2)$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times(\sigma^2,\infty)^n} P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}(\mathbf{Y},\mathbf{U},\mathbf{G})P_{\mathbf{U}}(\mathbf{U})P_{(g_1,\ldots,g_n)}(g_1,\ldots,g_n)d\mathbf{U}dg_1\ldots dg_n$$

where
- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
- $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);
- if **H** is Gaussian, $P_{(g_1-\sigma^2,\ldots,g_n-\sigma^2)}$ is the joint eigenvalue distribution of a central Wishart;

Evaluation of $P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y})$

- by MaxEnt, **X**, **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.
- **Under $\mathcal{H}_0$:**
  - $\mathbf{Y} = \sigma\mathbf{W}$

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

- **Under $\mathcal{H}_1$:**
  - $\mathbf{Y} = \begin{bmatrix}\sqrt{P}\mathbf{H} & \sigma\mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\boldsymbol{\Sigma}\geq 0} P_{\mathbf{Y}|\boldsymbol{\Sigma},\mathcal{H}_1}(\mathbf{Y},\boldsymbol{\Sigma})P_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma})d\boldsymbol{\Sigma}$$

with $\boldsymbol{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$.

From unitary invariance of **H**, denoting $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$, $\operatorname{diag}(\mathbf{G}) = (g_1,\ldots,g_n,\sigma^2,\ldots,\sigma^2)$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times(\sigma^2,\infty)^n} P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}(\mathbf{Y},\mathbf{U},\mathbf{G})P_{\mathbf{U}}(\mathbf{U})P_{(g_1,\ldots,g_n)}(g_1,\ldots,g_n)d\mathbf{U}dg_1\ldots dg_n$$

where

- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
- $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);
- if **H** is Gaussian, $P_{(g_1-\sigma^2,\ldots,g_n-\sigma^2)}$ is the joint eigenvalue distribution of a central Wishart;

Evaluation of $P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y})$

- by MaxEnt, **X**, **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.
- **Under $\mathcal{H}_0$:**
  - $\mathbf{Y} = \sigma \mathbf{W}$

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

- **Under $\mathcal{H}_1$:**
  - $\mathbf{Y} = \begin{bmatrix}\sqrt{P}\mathbf{H} & \sigma\mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}\geq 0} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma})P_{\mathbf{\Sigma}}(\mathbf{\Sigma})d\mathbf{\Sigma}$$

with $\mathbf{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$.
From unitary invariance of **H**, denoting $\mathbf{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$, $\operatorname{diag}(\mathbf{G}) = (g_1,\ldots,g_n,\sigma^2,\ldots,\sigma^2)$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times(\sigma^2,\infty)^n} P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}(\mathbf{Y},\mathbf{U},\mathbf{G})P_{\mathbf{U}}(\mathbf{U})P_{(g_1,\ldots,g_n)}(g_1,\ldots,g_n)d\mathbf{U}dg_1\ldots dg_n$$

where

- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
- $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);
- if **H** is Gaussian, $P_{(g_1-\sigma^2,\ldots,g_n-\sigma^2)}$ is the joint eigenvalue distribution of a central Wishart;

Evaluation of $P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y})$

- by MaxEnt, **X**, **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.
- **Under $\mathcal{H}_0$:**
  - $\mathbf{Y} = \sigma \mathbf{W}$

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

- **Under $\mathcal{H}_1$:**
  - $\mathbf{Y} = \begin{bmatrix}\sqrt{P}\mathbf{H} & \sigma \mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\boldsymbol{\Sigma}\geq 0} P_{\mathbf{Y}|\boldsymbol{\Sigma},\mathcal{H}_1}(\mathbf{Y},\boldsymbol{\Sigma}) P_{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}) d\boldsymbol{\Sigma}$$

with $\boldsymbol{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N$.
From unitary invariance of **H**, denoting $\boldsymbol{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$, $\operatorname{diag}(\mathbf{G}) = (g_1,\ldots,g_n,\sigma^2,\ldots,\sigma^2)$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times(\sigma^2,\infty)^n} P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}(\mathbf{Y},\mathbf{U},\mathbf{G}) P_{\mathbf{U}}(\mathbf{U}) P_{(g_1,\ldots,g_n)}(g_1,\ldots,g_n) d\mathbf{U} dg_1\ldots dg_n$$

where
- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
- $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);
- if **H** is Gaussian, $P_{(g_1-\sigma^2,\ldots,g_n-\sigma^2)}$ is the joint eigenvalue distribution of a central Wishart;

Evaluation of $P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y})$

- by MaxEnt, **X**, **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.
- **Under $\mathcal{H}_0$:**
  - $\mathbf{Y} = \sigma \mathbf{W}$

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

- **Under $\mathcal{H}_1$:**
  - $\mathbf{Y} = \begin{bmatrix}\sqrt{P}\mathbf{H} & \sigma \mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}\geq 0} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma})P_{\mathbf{\Sigma}}(\mathbf{\Sigma})d\mathbf{\Sigma}$$

with $\mathbf{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$.
From unitary invariance of **H**, denoting $\mathbf{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$, $\operatorname{diag}(\mathbf{G}) = (g_1,\ldots,g_n,\sigma^2,\ldots,\sigma^2)$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times(\sigma^2,\infty)^n} P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}(\mathbf{Y},\mathbf{U},\mathbf{G})P_{\mathbf{U}}(\mathbf{U})P_{(g_1,\ldots,g_n)}(g_1,\ldots,g_n)d\mathbf{U}dg_1\ldots dg_n$$

where
- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
- $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);
- if **H** is Gaussian, $P_{(g_1-\sigma^2,\ldots,g_n-\sigma^2)}$ is the joint eigenvalue distribution of a central Wishart;

Evaluation of $P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y})$

- by MaxEnt, **X**, **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.
- **Under $\mathcal{H}_0$:**
  - $\mathbf{Y} = \sigma \mathbf{W}$

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

- **Under $\mathcal{H}_1$:**
  - $\mathbf{Y} = \begin{bmatrix}\sqrt{P}\mathbf{H} & \sigma\mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}\geq 0} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma})P_{\mathbf{\Sigma}}(\mathbf{\Sigma})d\mathbf{\Sigma}$$

with $\mathbf{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N$.
From unitary invariance of **H**, denoting $\mathbf{\Sigma} = \mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$, $\operatorname{diag}(\mathbf{G}) = (g_1,\ldots,g_n,\sigma^2,\ldots,\sigma^2)$

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times(\sigma^2,\infty)^n} P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}(\mathbf{Y},\mathbf{U},\mathbf{G})P_{\mathbf{U}}(\mathbf{U}){\color{red}P_{(g_1,\ldots,g_n)}(g_1,\ldots,g_n)}d\mathbf{U}dg_1\ldots dg_n$$

where

- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
- $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);
- if **H** is Gaussian, ${\color{red}P_{(g_1-\sigma^2,\ldots,g_n-\sigma^2)}}$ is the joint eigenvalue distribution of a central Wishart;

Result in the Gaussian case, $n = 1$

R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Sensing", IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5186-5195, 2010.

Theorem (Neyman-Pearson test)

The ratio $C(\mathbf{Y})$ when the receiver knows $n = 1$, $P = 1$, $\mathrm{E}[\frac{1}{N}\operatorname{tr}\mathbf{H}\mathbf{H}^{\mathrm{H}}] = 1$ and σ^2, reads

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

with $\lambda_1, \ldots, \lambda_N$ the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and where

$$J_k(x, y) \triangleq \int_x^{+\infty} t^k e^{-t-\frac{y}{t}}\, dt.$$

- non trivial dependency on $\lambda_1, \ldots, \lambda_N$
- contrary to energy detector, $\sum_i \lambda_i$ is not a sufficient statistic;
- integration over σ^2 (or P when $P \neq 1$) is difficult.

Result in the Gaussian case, $n = 1$

R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Sensing", IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5186-5195, 2010.

Theorem (Neyman-Pearson test)

The ratio $C(\mathbf{Y})$ when the receiver knows $n = 1$, $P = 1$, $\mathrm{E}[\frac{1}{N}\operatorname{tr}\mathbf{H}\mathbf{H}^{\mathrm{H}}] = 1$ and σ^2, reads

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

with $\lambda_1, \ldots, \lambda_N$ the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and where

$$J_k(x, y) \triangleq \int_x^{+\infty} t^k e^{-t-\frac{y}{t}} dt.$$

- non trivial dependency on $\lambda_1, \ldots, \lambda_N$
- contrary to energy detector, $\sum_i \lambda_i$ is not a sufficient statistic;
- integration over σ^2 (or P when $P \neq 1$) is difficult.

Result in the Gaussian case, $n = 1$

R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Sensing", IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5186-5195, 2010.

Theorem (Neyman-Pearson test)

The ratio $C(\mathbf{Y})$ when the receiver knows $n = 1$, $P = 1$, $\mathrm{E}[\frac{1}{N}\operatorname{tr}\mathbf{H}\mathbf{H}^{\mathsf{H}}] = 1$ and σ^2, reads

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

with $\lambda_1, \ldots, \lambda_N$ the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and where

$$J_k(x, y) \triangleq \int_x^{+\infty} t^k e^{-t-\frac{y}{t}}\, dt.$$

- non trivial dependency on $\lambda_1, \ldots, \lambda_N$
- contrary to energy detector, $\sum_i \lambda_i$ is not a sufficient statistic;
- integration over σ^2 (or P when $P \neq 1$) is difficult.

Comparison to energy detector

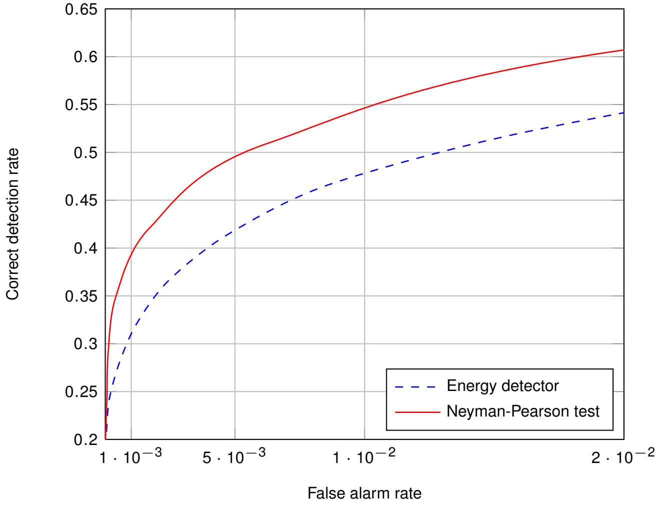


Figure: ROC curve for single-source detection, $K = 1$, $N = 4$, $M = 8$, $\mathrm{SNR} = -3$ dB, FAR range of practical interest, with signal power $P = 0$ dBm, either known or unknown at the receiver.

Comparison to energy detector

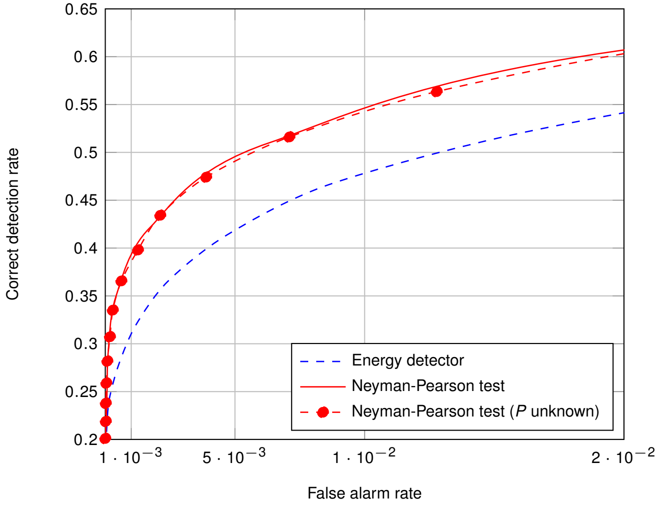


Figure: ROC curve for single-source detection, $K = 1$, $N = 4$, $M = 8$, $\mathrm{SNR} = -3$ dB, FAR range of practical interest, with signal power $P = 0$ dBm, either known or unknown at the receiver.

Outline

Reminder: the Marčenko-Pastur Law

If $\mathcal{H}_0$, then the eigenvalues of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} = \sigma^2 \frac{1}{N}\mathbf{W}\mathbf{W}^{\mathsf{H}}$ asymptotically distribute as

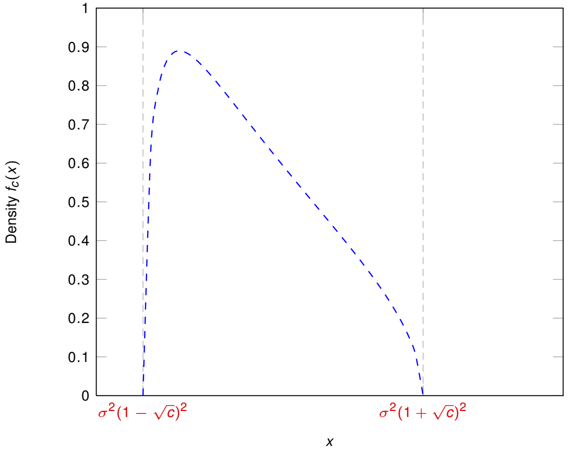


Figure: Marčenko-Pastur law with $c = \lim N/L$.

Alternative Tests in Large Random Matrix Theory

Reminder:

Theorem

$$P(\text{no eigenvalues outside } [\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2] \text{ for all large } N) = 1$$

- If $\mathcal{H}_0$,

$$\frac{\lambda_{\max}(\frac{1}{N}\mathbf{YY}^{\mathrm{H}})}{\lambda_{\min}(\frac{1}{N}\mathbf{YY}^{\mathrm{H}})} \xrightarrow{\text{a.s.}} \frac{(1+\sqrt{c})^2}{(1-\sqrt{c})^2}$$

- independent of the SNR!

Alternative Tests in Large Random Matrix Theory

Reminder:

Theorem

$$P(\textit{no eigenvalues outside } [\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2] \textit{ for all large } N) = 1$$

- If $\mathcal{H}_0$,

$$\frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\lambda_{\min}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})} \xrightarrow{\text{a.s.}} \frac{(1+\sqrt{c})^2}{(1-\sqrt{c})^2}$$

- independent of the SNR!

Alternative Tests in Large Random Matrix Theory

Reminder:

Theorem

$$P(\textit{no eigenvalues outside } [\sigma^2(1-\sqrt{c})^2, \sigma^2(1+\sqrt{c})^2] \textit{ for all large } N) = 1$$

- If $\mathcal{H}_0$,

$$\frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\lambda_{\min}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})} \xrightarrow{\text{a.s.}} \frac{(1+\sqrt{c})^2}{(1-\sqrt{c})^2}$$

- independent of the SNR!

Conditioning Number Test

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

Conditioning number test

$$C_{\text{cond}}(\mathbf{Y}) = \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\lambda_{\min}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}$$

- if $C_{\text{cond}}(\mathbf{Y}) > \tau$, presence of a signal.
- if $C_{\text{cond}}(\mathbf{Y}) < \tau$, absence of signal.

- but this is *ad-hoc*! how good does it compare to optimal?
- can we find non *ad-hoc* approaches?

Conditioning Number Test

L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative spectrum sensing using random matrix theory," International Symposium on Wireless Pervasive Computing, Santorini, Greece, 2008.

Conditioning number test

$$C_{\text{cond}}(\mathbf{Y}) = \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}{\lambda_{\min}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})}$$

- if $C_{\text{cond}}(\mathbf{Y}) > \tau$, presence of a signal.
- if $C_{\text{cond}}(\mathbf{Y}) < \tau$, absence of signal.

- but this is *ad-hoc*! how good does it compare to optimal?
- can we find non *ad-hoc* approaches?

Generalized Likelihood Ratio Test

Bianchi, J. Najim, M. Maida, M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," Proceedings of IEEE Statistical Signal Processing Workshop, 2009.

Generalized Likelihood Ratio Test

- Alternative test to Neyman-Pearson test,

$$C_{\rm GLRT}(\mathbf{Y}) = \frac{\sup_{\mathbf{H},\sigma^2} P_{\mathcal{H}_1|\mathbf{Y},\mathbf{H},\sigma^2}(\mathbf{Y})}{\sup_{\sigma^2} P_{\mathcal{H}_0|\mathbf{Y},\sigma^2}(\mathbf{Y})}$$

- based on ratios of maximum likelihood
- clearly sub-optimal but avoid the need for priors.

- GLRT test

$$C_{\rm GLRT}(\mathbf{Y}) = \left(\left(1-\frac{1}{N}\right)^{N-1} \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\rm H})}{\frac{1}{N}\sum_{i=1}^N \lambda_i}\left(1-\frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\rm H})}{\sum_{i=1}^N \lambda_i}\right)^{N-1}\right)^{-L}.$$

- Contrary to the *ad-hoc* conditioning number test, GLRT based on

$$\frac{\lambda_{\max}}{\frac{1}{N}\operatorname{tr}(\mathbf{Y}\mathbf{Y}^{\rm H})}$$

Generalized Likelihood Ratio Test

Bianchi, J. Najim, M. Maida, M. Debbah, “Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing,” Proceedings of IEEE Statistical Signal Processing Workshop, 2009.

Generalized Likelihood Ratio Test

- Alternative test to Neyman-Pearson test,

$$C_{\mathrm{GLRT}}(\mathbf{Y}) = \frac{\sup_{\mathbf{H},\sigma^2} P_{\mathcal{H}_1|\mathbf{Y},\mathbf{H},\sigma^2}(\mathbf{Y})}{\sup_{\sigma^2} P_{\mathcal{H}_0|\mathbf{Y},\sigma^2}(\mathbf{Y})}$$

- based on ratios of maximum likelihood
- clearly sub-optimal but avoid the need for priors.

- GLRT test

$$C_{\mathrm{GLRT}}(\mathbf{Y}) = \left(\left(1 - \frac{1}{N}\right)^{N-1} \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}})}{\frac{1}{N}\sum_{i=1}^{N}\lambda_i} \left(1 - \frac{\lambda_{\max}(\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}})}{\sum_{i=1}^{N}\lambda_i}\right)^{N-1} \right)^{-L}.$$

- Contrary to the *ad-hoc* conditioning number test, GLRT based on

$$\frac{\lambda_{\max}}{\frac{1}{N}\mathrm{tr}(\mathbf{Y}\mathbf{Y}^{\mathrm{H}})}$$

Neyman-Pearson Test against Asymptotic Tests

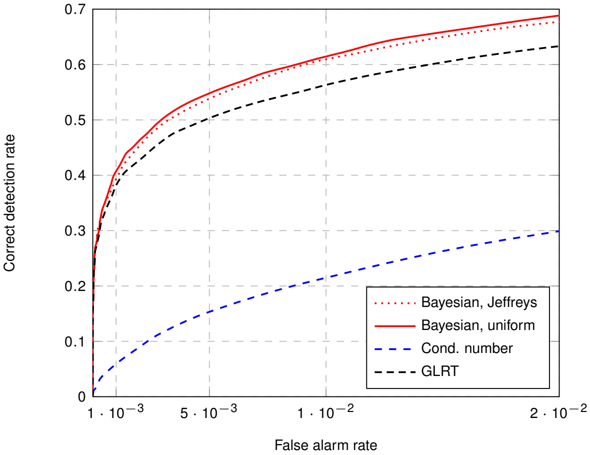
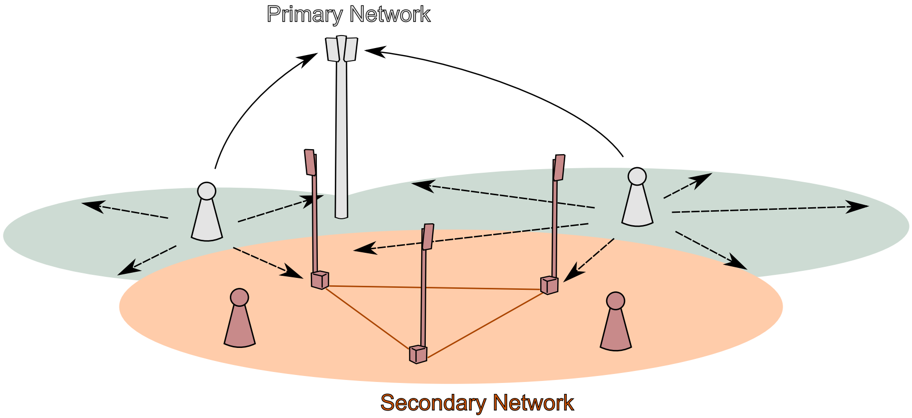


Figure: ROC curve for *a priori* unknown σ^2 of the Bayesian method, conditioning number method and GLRT method, $M = 1$, $N = 4$, $L = 8$, SNR $= 0$ dB. For the Bayesian method, both uniform and Jeffreys prior, with exponent $\alpha = 1$, are provided.

Outline

Application Context: Coverage range in Femtocells



Problem Statement

- We now consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}$$

and wish to infer $P_1, \ldots, P_K$.

- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{X}_k + \sigma\mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix}\mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K\end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma\mathbf{W} = \left[\mathbf{H}\mathbf{P}^{\frac{1}{2}} \quad \sigma\mathbf{I}_N\right]\begin{bmatrix}\mathbf{X} \\ \mathbf{W}\end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^{\mathsf{T}}\ \mathbf{W}^{\mathsf{T}})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

Most information about $P_1, \ldots, P_K$ is contained in the eigenvalues of $\mathbf{B}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

Problem Statement

- We now consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}$$

and wish to infer $P_1, \ldots, P_K$.

- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{X}_k + \sigma\mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma\mathbf{W} = \left[\mathbf{H}\mathbf{P}^{\frac{1}{2}} \quad \sigma\mathbf{I}_N\right] \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^{\mathsf{T}}\ \mathbf{W}^{\mathsf{T}})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

Most information about $P_1, \ldots, P_K$ is contained in the eigenvalues of $\mathbf{B}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

Problem Statement

- We now consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{x}_k^{(m)} + \sigma \mathbf{w}^{(m)}$$

and wish to infer $P_1, \ldots, P_K$.

- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{X}_k + \sigma \mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma \mathbf{W} = \left[\mathbf{H}\mathbf{P}^{\frac{1}{2}} \quad \sigma \mathbf{I}_N\right] \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^{\mathsf{T}} \ \mathbf{W}^{\mathsf{T}})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

Most information about $P_1, \ldots, P_K$ is contained in the eigenvalues of $\mathbf{B}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

Problem Statement

- We now consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{x}_k^{(m)} + \sigma \mathbf{w}^{(m)}$$

  and wish to infer $P_1, \ldots, P_K$.
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{X}_k + \sigma \mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma \mathbf{W} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^{\mathsf{T}}\ \mathbf{W}^{\mathsf{T}})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

Most information about $P_1, \ldots, P_K$ is contained in the eigenvalues of $\mathbf{B}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

Problem Statement

- We now consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}$$

  and wish to infer $P_1, \ldots, P_K$.
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{X}_k + \sigma\mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix}\mathbf{X}_1\\ \vdots \\ \mathbf{X}_K\end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma\mathbf{W} = \left[\mathbf{H}\mathbf{P}^{\frac{1}{2}} \quad \sigma\mathbf{I}_N\right]\begin{bmatrix}\mathbf{X}\\ \mathbf{W}\end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^{\mathsf{T}}\ \mathbf{W}^{\mathsf{T}})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

Most information about $P_1, \ldots, P_K$ is contained in the eigenvalues of $\mathbf{B}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

Problem Statement

- We now consider the model

$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{x}_k^{(m)} + \sigma \mathbf{w}^{(m)}$$

and wish to infer $P_1, \ldots, P_K$.

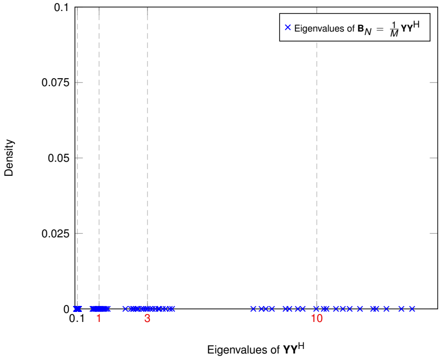
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{P_k} \mathbf{H}_k \mathbf{X}_k + \sigma \mathbf{W} = \underbrace{\left[\sqrt{P_1}\mathbf{H}_1 \quad \cdots \quad \sqrt{P_K}\mathbf{H}_K\right]}_{\triangleq \mathbf{H}\mathbf{P}^{\frac{1}{2}}} \underbrace{\begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_K \end{bmatrix}}_{\triangleq \mathbf{X}} + \sigma \mathbf{W} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}.$$

- If $\mathbf{H}$, $(\mathbf{X}^{\mathsf{T}} \ \mathbf{W}^{\mathsf{T}})$ are unitarily invariant, $\mathbf{Y}$ is unitarily invariant.

Most information about $P_1, \ldots, P_K$ is contained in the eigenvalues of $\mathbf{B}_N \triangleq \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$.

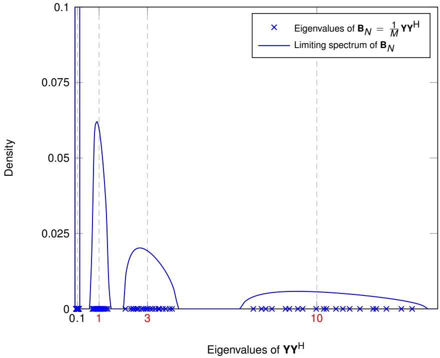
From small to large system analysis



The classical approach requires to evaluate $P_{P_1,\ldots,P_K|Y}$

- assuming Gaussian parameters, this is similar to previous calculus
- leads to a very involved expression
- prohibitively expensive to evaluate even for small N, n_k, M

From small to large system analysis



Assuming dimensions N, n_k, M grow large, large dimensional random matrix theory provides

- a link between:
  - **the "observation":** the limiting spectral distribution (l.s.d.) of $\mathbf{B}_N$;
  - **the "hidden parameters":** the powers $P_1, \ldots, P_K$, i.e. the l.s.d. of $\mathbf{P}$.
- consistent estimators of the hidden parameters.

Outline

Optimal ML/MMSE estimators

R. Couillet and M. Guillaud, "Performance of Statistical Inference Methods for the Energy Estimation of Multiple Sources," *Invited Paper*, IEEE International Communications Conference, Nice, France, 2011.

- conditional probability

Theorem

Assume $P_1, \ldots, P_K$ have multiplicity $n_1 = \ldots = n_K = 1$. Then, denoting $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ the eigenvalues of $\mathbf{B}_N$

$$P_{\mathbf{Y}|P_1,\ldots,P_K}(\mathbf{Y}) = \frac{C(-1)^{Nn+1} e^{N\sigma^2 \sum_{i=1}^n \frac{1}{P_i}}}{\sigma^{2(N-n)(M-n)} \prod_{i=1}^n P_i^{M-n+1} \Delta(\mathbf{P})} \sum_{\mathbf{a}\in\mathcal{S}_n^N} (-1)^{|\mathbf{a}|} \operatorname{sgn}(\mathbf{a}) e^{\frac{M}{\sigma^2}|\boldsymbol{\lambda}[\bar{\mathbf{a}}]|}$$

$$\times \frac{\Delta(\operatorname{diag}(\boldsymbol{\lambda}[\bar{\mathbf{a}}]))}{\Delta(\operatorname{diag}(\boldsymbol{\lambda}))} \sum_{\mathbf{b}\in\mathcal{S}_n} \operatorname{sgn}(\mathbf{b}) \prod_{i=1}^n J_{N-M-1}\left(\frac{N\sigma^2}{P_{b_i}}, \frac{NM\lambda_{a_i}}{P_{b_i}}\right).$$

- ML/MMSE estimators

$$\hat{P}^{(\mathrm{ML})} = \arg \max_{P_1,\ldots,P_K} P_{\mathbf{Y}|P_1,\ldots,P_K}(\mathbf{Y})$$

$$\hat{P}^{(\mathrm{MMSE})} = \int_{[0,\infty)^K} (P_1,\ldots,P_K) P_{P_1,\ldots,P_K|\mathbf{Y}}(P_1,\ldots,P_K) dP_1 \ldots dP_K$$

Optimal ML/MMSE estimators

R. Couillet and M. Guillaud, “Performance of Statistical Inference Methods for the Energy Estimation of Multiple Sources,” *Invited Paper*, IEEE International Communications Conference, Nice, France, 2011.

- conditional probability

Theorem

Assume $P_1, \ldots, P_K$ have multiplicity $n_1 = \ldots = n_K = 1$. Then, denoting $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ the eigenvalues of $\mathbf{B}_N$

$$P_{\mathbf{Y}|P_1,\ldots,P_K}(\mathbf{Y}) = \frac{C(-1)^{Nn+1} e^{N\sigma^2 \sum_{i=1}^{n} \frac{1}{P_i}}}{\sigma^{2(N-n)(M-n)} \prod_{i=1}^{n} P_i^{M-n+1} \Delta(\mathbf{P})} \sum_{\mathbf{a}\in\mathcal{S}_n^N} (-1)^{|\mathbf{a}|} \operatorname{sgn}(\mathbf{a}) e^{\frac{M}{\sigma^2}|\boldsymbol{\lambda}[\bar{\mathbf{a}}]|}$$

$$\times \frac{\Delta(\operatorname{diag}(\boldsymbol{\lambda}[\bar{\mathbf{a}}]))}{\Delta(\operatorname{diag}(\boldsymbol{\lambda}))} \sum_{\mathbf{b}\in\mathcal{S}_n} \operatorname{sgn}(\mathbf{b}) \prod_{i=1}^{n} J_{N-M-1}\left(\frac{N\sigma^2}{P_{b_i}}, \frac{NM\lambda_{a_i}}{P_{b_i}}\right).$$

- ML/MMSE estimators

$$\underline{\hat{P}}^{(\mathrm{ML})} = \arg\max_{P_1,\ldots,P_K} P_{\mathbf{Y}|P_1,\ldots,P_K}(\mathbf{Y})$$

$$\underline{\hat{P}}^{(\mathrm{MMSE})} = \int_{[0,\infty)^K} (P_1, \ldots, P_K) P_{P_1,\ldots,P_K|\mathbf{Y}}(P_1, \ldots, P_K) dP_1 \ldots dP_K$$

Outline

Reminder on free deconvolution

- Free probability provides tools to compute

$$p_k = \frac{1}{K}\sum_{i=1}^{K}\lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K}P_i^k$$

as a function of

$$b_k = \frac{1}{N}\sum_{i=1}^{N}\lambda(\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}})^k$$

- One can obtain all the successive sum powers of $P_1, \ldots, P_K$.
- From that, we can infer on the values of each P_k!
- The tools come from the relations,
  - cumulant to moment (and also moment to cumulant),

$$M_n = \sum_{\pi \in NC(n)} \prod_{V \in \pi} C_{|V|}$$

  - Sums of cumulants for *asymptotically free* **A** and **B** (of measure $\mu_A \boxplus \mu_B$),

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

  - Products of cumulants for *asymptotically free* **A** and **B** (of measure $\mu_A \boxtimes \mu_B$),

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

  - Moments of information plus noise models $\mathbf{B}_N = \frac{1}{n}\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)^{\mathrm{H}}$,

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

with μ_c the Marčenko-Pastur law with ratio c.

Reminder on free deconvolution

- Free probability provides tools to compute

$$p_k = \frac{1}{K}\sum_{i=1}^{K}\lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K}P_i^k$$

as a function of

$$b_k = \frac{1}{N}\sum_{i=1}^{N}\lambda(\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}})^k$$

- One can obtain all the successive sum powers of $P_1, \dots, P_K$.
- From that, we can infer on the values of each P_k!
- The tools come from the relations,
  - cumulant to moment (and also moment to cumulant),

$$M_n = \sum_{\pi \in NC(n)} \prod_{V \in \pi} C_{|V|}$$

  - Sums of cumulants for *asymptotically free* $\mathbf{A}$ and $\mathbf{B}$ (of measure $\mu_A \boxplus \mu_B$),

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

  - Products of cumulants for *asymptotically free* $\mathbf{A}$ and $\mathbf{B}$ (of measure $\mu_A \boxtimes \mu_B$),

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

  - Moments of information plus noise models $\mathbf{B}_N = \frac{1}{n}\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)\left(\mathbf{A}_N + \sigma\mathbf{W}_N\right)^{\mathrm{H}}$,

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

with μ_c the Marčenko-Pastur law with ratio c.

Reminder on free deconvolution

- Free probability provides tools to compute

$$p_k = \frac{1}{K}\sum_{i=1}^{K}\lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K}P_i^k$$

as a function of

$$b_k = \frac{1}{N}\sum_{i=1}^{N}\lambda(\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}})^k$$

- One can obtain all the successive sum powers of $P_1,\ldots,P_K$.
- From that, we can infer on the values of each P_k!
- The tools come from the relations,
  - cumulant to moment (and also moment to cumulant),

$$M_n = \sum_{\pi\in NC(n)}\prod_{V\in\pi} C_{|V|}$$

  - Sums of cumulants for *asymptotically free* **A** and **B** (of measure $\mu_A \boxplus \mu_B$),

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

  - Products of cumulants for *asymptotically free* **A** and **B** (of measure $\mu_A \boxtimes \mu_B$),

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

  - Moments of information plus noise models $\mathbf{B}_N = \frac{1}{n}(\mathbf{A}_N + \sigma\mathbf{W}_N)(\mathbf{A}_N + \sigma\mathbf{W}_N)^{\mathsf{H}}$,

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

with μ_c the Marčenko-Pastur law with ratio c.

Free deconvolution approach

- one can deconvolve $\mathbf{YY}^{\mathsf{H}}$ in three steps,
  - an information-plus-noise model with "deterministic matrix" $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{YY}^{\mathsf{H}} = (\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^{\mathsf{H}}$$

  - from $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{XX}^{\mathsf{H}}$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{HH}^{\mathsf{H}}$

$$\mathbf{PHH}^{\mathsf{H}}$$

Free deconvolution approach

- one can deconvolve $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ in three steps,
  - an information-plus-noise model with "deterministic matrix" $\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{Y}\mathbf{Y}^{\mathsf{H}} = (\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^{\mathsf{H}}$$

  - from $\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{H}\mathbf{H}^{\mathsf{H}}$

$$\mathbf{P}\mathbf{H}\mathbf{H}^{\mathsf{H}}$$

Free deconvolution approach

- one can deconvolve $\mathbf{YY}^H$ in three steps,
  - an information-plus-noise model with "deterministic matrix" $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^H\mathbf{P}^{\frac{1}{2}}\mathbf{H}^H$,

$$\mathbf{YY}^H = (\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})(\mathbf{HP}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W})^H$$

  - from $\mathbf{HP}^{\frac{1}{2}}\mathbf{XX}^H\mathbf{P}^{\frac{1}{2}}\mathbf{H}^H$, up to a Gram matrix commutation, we can deconvolve the signal $\mathbf{X}$,

$$\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^H\mathbf{P}^{\frac{1}{2}}\mathbf{XX}^H$$

  - from $\mathbf{P}^{\frac{1}{2}}\mathbf{HH}^H\mathbf{P}^{\frac{1}{2}}$, a new matrix commutation allows one to deconvolve $\mathbf{HH}^H$

$$\mathbf{PHH}^H$$

Free deconvolution approach

- In terms of distributions

$$\mu^{\infty}_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} = \left(\left(\mu^{\infty}_{\mathbf{B}_N} \boxslash \mu_{\frac{1}{c}}\right) \boxminus \delta_{\sigma^2}\right) \boxtimes \mu_{\frac{1}{c}}$$

$$\mu^{\infty}_{\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}} = \mu^{\infty}_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}} \boxslash \mu_{\frac{1}{cc_0}}$$

$$\mu^{\infty}_{\mathbf{P}} = \mu^{\infty}_{\mathbf{P}\mathbf{H}^{\mathsf{H}}\mathbf{H}} \boxslash \mu_{\frac{1}{c_0}}$$

- Numerically, with $b_m \triangleq \frac{1}{N}\mathrm{E}\left[\operatorname{tr}\mathbf{B}_N^m\right]$ and $p_m \triangleq \sum_{k=1}^{K}\frac{n_k}{n}P_k^m$

$$\begin{aligned}
b_1 &= N^{-1}np_1 + 1\\
b_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right)p_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right)p_1{}^2 + \left(2N^{-1}n + 2M^{-1}n\right)p_1 + \left(1 + NM^{-1}\right)\\
b_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right)p_3\\
&+ \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right)p_2p_1\\
&+ \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right)p_1{}^3\\
&+ \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right)p_2\\
&+ \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right)p_1{}^2\\
&+ \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right)p_1 .
\end{aligned}$$

Free deconvolution approach

- In terms of distributions

$$\mu^{\infty}_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} = \left(\left(\mu^{\infty}_{\mathbf{B}_N} \boxslash \mu_{\frac{1}{c}}\right) \boxminus \delta_{\sigma^2}\right) \boxtimes \mu_{\frac{1}{c}}$$

$$\mu^{\infty}_{\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}} = \mu^{\infty}_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}} \boxslash \mu_{\frac{1}{cc_0}}$$

$$\mu^{\infty}_{\mathbf{P}} = \mu^{\infty}_{\mathbf{P}\mathbf{H}^{\mathsf{H}}\mathbf{H}} \boxslash \mu_{\frac{1}{c_0}}$$

- Numerically, with $b_m \triangleq \frac{1}{N}\mathrm{E}\left[\operatorname{tr}\mathbf{B}_N^m\right]$ and $p_m \triangleq \sum_{k=1}^{K} \frac{n_k}{n} P_k^m$

$$\begin{aligned}
b_1 &= N^{-1} n p_1 + 1 \\
b_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right) p_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right) p_1{}^2 + \left(2N^{-1}n + 2M^{-1}n\right) p_1 + \left(1 + NM^{-1}\right) \\
b_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right) p_3 \\
&+ \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right) p_2 p_1 \\
&+ \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right) p_1{}^3 \\
&+ \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right) p_2 \\
&+ \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right) p_1{}^2 \\
&+ \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right) p_1 .
\end{aligned}$$

Newton-Girard inversion

- Once the p_i^m are obtained, in the particular case $n_1 = \ldots = n_K$, Newton-Girard formulas give $P_1, \ldots, P_K$ as the solutions of

$$X^K - \Pi_1 X^{K-1} + \Pi_2 X^{K-2} - \ldots + (-1)^K \Pi_K = 0$$

with $\Pi_1, \ldots, \Pi_n$ recursively computed from

$$(-1)^K K \Pi_K + \sum_{i=1}^{K} (-1)^{K+i} p_i \Pi_{K-i} = 0.$$

- fast method but with major limitations!
  - polynomial solutions can be purely complex
  - moment estimates propagate errors to higher order moments (2nd estimate 10^3 worse than 1st!)
  - modifying Newton-Girard formulas boils down to ad-hoc methods...
  - ML and MMSE methods are prohibitively expensive.

Newton-Girard inversion

- Once the p_i^m are obtained, in the particular case $n_1 = \ldots = n_K$, Newton-Girard formulas give $P_1, \ldots, P_K$ as the solutions of
$$X^K - \Pi_1 X^{K-1} + \Pi_2 X^{K-2} - \ldots + (-1)^K \Pi_K = 0$$
with $\Pi_1, \ldots, \Pi_n$ recursively computed from
$$(-1)^K K \Pi_K + \sum_{i=1}^{K} (-1)^{K+i} p_i \Pi_{K-i} = 0.$$
- fast method but with major limitations!
  - polynomial solutions can be purely complex
  - moment estimates propagate errors to higher order moments (2nd estimate 10^3 worse than 1st!)
  - modifying Newton-Girard formulas boils down to ad-hoc methods...
  - ML and MMSE methods are prohibitively expensive.

Outline

Limiting spectrum of the sample covariance matrix

- Recall the model

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}$$

  very similar to a sample covariance matrix.

- *for simplicity of analysis*, consider the *sample covariance matrix* model

$$\mathbf{Y} \triangleq \mathbf{T}^{\frac{1}{2}}\mathbf{X} \in \mathbb{C}^{N\times n},\ \mathbf{B}_N = \frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} \in \mathbb{C}^{N\times N},\ \underline{\mathbf{B}}_N = \frac{1}{n}\mathbf{Y}^{\mathsf{H}}\mathbf{Y} \in \mathbb{C}^{n\times n}$$

  where $\mathbf{T} \in \mathbb{C}^{N\times N}$ has eigenvalues $t_1, \ldots, t_K$, t_k with multiplicity N_k and $\mathbf{X} \in \mathbb{C}^{N\times n}$ is i.i.d. zero mean, variance 1.

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

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

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

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

Limiting spectrum of the sample covariance matrix

- Recall the model

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}$$

very similar to a sample covariance matrix.

- *for simplicity of analysis*, consider the *sample covariance matrix* model

$$\mathbf{Y} \triangleq \mathbf{T}^{\frac{1}{2}} \mathbf{X} \in \mathbb{C}^{N \times n}, \ \mathbf{B}_N = \frac{1}{n} \mathbf{Y}\mathbf{Y}^{\mathsf{H}} \in \mathbb{C}^{N \times N}, \ \underline{\mathbf{B}}_N = \frac{1}{n} \mathbf{Y}^{\mathsf{H}}\mathbf{Y} \in \mathbb{C}^{n \times n}$$

where $\mathbf{T} \in \mathbb{C}^{N \times N}$ has eigenvalues $t_1, \ldots, t_K$, t_k with multiplicity N_k and $\mathbf{X} \in \mathbb{C}^{N \times n}$ is i.i.d. zero mean, variance 1.

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

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

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

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

Limiting spectrum of the sample covariance matrix

- Recall the model

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}$$

  very similar to a sample covariance matrix.

- *for simplicity of analysis*, consider the *sample covariance matrix* model

$$\mathbf{Y} \triangleq \mathbf{T}^{\frac{1}{2}} \mathbf{X} \in \mathbb{C}^{N \times n}, \ \mathbf{B}_N = \frac{1}{n}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} \in \mathbb{C}^{N \times N}, \ \underline{\mathbf{B}}_N = \frac{1}{n}\mathbf{Y}^{\mathsf{H}}\mathbf{Y} \in \mathbb{C}^{n \times n}$$

  where $\mathbf{T} \in \mathbb{C}^{N \times N}$ has eigenvalues $t_1, \ldots, t_K$, t_k with multiplicity N_k and $\mathbf{X} \in \mathbb{C}^{N \times n}$ is i.i.d. zero mean, variance 1.

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

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

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

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

Complex integration

- From Cauchy integral formula, with $\mathcal{C}_k$ a contour enclosing **only** t_k (negatively oriented),

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

- 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,

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

- Dominated convergence arguments then show

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

Complex integration

- From Cauchy integral formula, with $\mathcal{C}_k$ a contour enclosing **only** t_k (negatively oriented),

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

- 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,

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

- Dominated convergence arguments then show

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

Complex integration

- From Cauchy integral formula, with $\mathcal{C}_k$ a contour enclosing **only** t_k (negatively oriented),

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

- 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,

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

- Dominated convergence arguments then show

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

Complex integration

- From Cauchy integral formula, with $\mathcal{C}_k$ a contour enclosing **only** t_k (negatively oriented),

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

- 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,

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

- Dominated convergence arguments then show

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

Complex integration

- From Cauchy integral formula, with $\mathcal{C}_k$ a contour enclosing **only** t_k (negatively oriented),

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

- 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,

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

- Dominated convergence arguments then show

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

Complex integration

- From Cauchy integral formula, with $\mathcal{C}_k$ a contour enclosing **only** t_k (negatively oriented),

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

- 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,

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

- Dominated convergence arguments then show

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

Where does the contour go?

Intuition:

- $m_{\underline{F}}(z)$ is defined outside the support of $\underline{F}$
- on the real axis, $m'_{\underline{F}}(z) = \int \frac{1}{(t-z)^2} d\underline{F}(t) > 0$
- it therefore has a local growing inverse outside the support of $\underline{F}$
- notice that $m_{\underline{F}}(z)$ has a closed-form inverse

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

It can be shown that $z_{\underline{F}}(m)$, $m < 0$, is growing *if and only if* its image is outside the support of $\underline{F}$.

Where does the contour go?

Intuition:

- $m_{\underline{F}}(z)$ is defined outside the support of $\underline{F}$
- on the real axis, $m'_{\underline{F}}(z) = \int \frac{1}{(t-z)^2} d\underline{F}(t) > 0$
- it therefore has a local growing inverse outside the support of $\underline{F}$
- notice that $m_{\underline{F}}(z)$ has a closed-form inverse

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

It can be shown that $z_{\underline{F}}(m)$, $m < 0$, is growing *if and only if* its image is outside the support of $\underline{F}$.

Where does the contour go?

Intuition:

- $m_{\underline{F}}(z)$ is defined outside the support of $\underline{F}$
- on the real axis, $m'_{\underline{F}}(z) = \int \frac{1}{(t-z)^2} d\underline{F}(t) > 0$
- it therefore has a local growing inverse outside the support of $\underline{F}$
- notice that $m_{\underline{F}}(z)$ has a closed-form inverse

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

It can be shown that $z_{\underline{F}}(m)$, $m < 0$, is growing *if and only if* its image is outside the support of $\underline{F}$.

Inverse formula for the Stieltjes transform

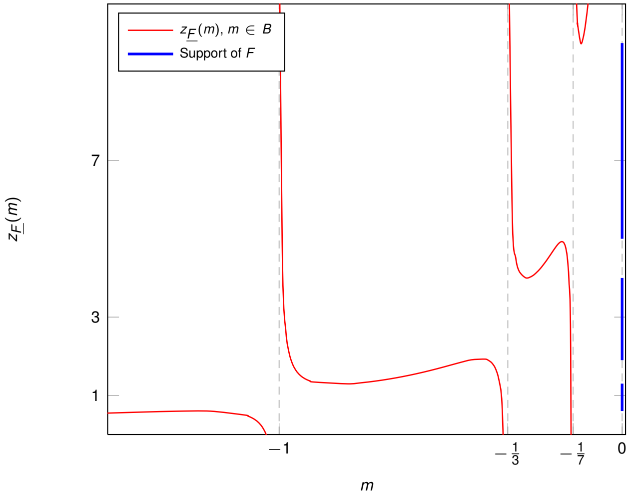


Figure: $z_{\underline{F}}(m)$, with $\underline{F}$ the l.s.d. of $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N$ with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3 and 7. The support of F is read on the vertical axis, whenever $x_{\underline{F}}(m)$ is not increasing.

Playing with the asymptotes...

- denote x_k^-, x_k^+ two points on either side of cluster k in $\underline{F}$ such that $x_k^- = z_{\underline{F}}(m_k^-)$ and $x_k^+ = z_{\underline{F}}(m_k^+)$.
- from the asymptotes, we observe that

$$t_{k-1} < -\frac{1}{m_k^-} < t_k < -\frac{1}{m_k^+} < t_{k+1}$$

- we can therefore take a contour $\mathcal{C}_{\underline{F},k}$ that crosses the real line at $-\frac{1}{m_k^-}$ and at $-\frac{1}{m_k^+}$ and is outside the real line everywhere else.

Playing with the asymptotes...

- denote x_k^-, x_k^+ two points on either side of cluster k in $\underline{F}$ such that $x_k^- = z_{\underline{F}}(m_k^-)$ and $x_k^+ = z_{\underline{F}}(m_k^+)$.
- from the asymptotes, we observe that

$$t_{k-1} < -\frac{1}{m_k^-} < t_k < -\frac{1}{m_k^+} < t_{k+1}$$

- we can therefore take a contour $\mathcal{C}_{\underline{F},k}$ that crosses the real line at $-\frac{1}{m_k^-}$ and at $-\frac{1}{m_k^+}$ and is outside the real line everywhere else.

Playing with the asymptotes...

- denote x_k^-, x_k^+ two points on either side of cluster k in $\underline{F}$ such that $x_k^- = z_{\underline{F}}(m_k^-)$ and $x_k^+ = z_{\underline{F}}(m_k^+)$.
- from the asymptotes, we observe that

$$t_{k-1} < -\frac{1}{m_k^-} < t_k < -\frac{1}{m_k^+} < t_{k+1}$$

- we can therefore take a contour $\mathcal{C}_{\underline{F},k}$ that crosses the real line at $-\frac{1}{m_k^-}$ and at $-\frac{1}{m_k^+}$ and is outside the real line everywhere else.

Termination

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.

- If remains to compute the integral from residue calculus.

$$\hat{t}_k = \frac{N}{N_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_{\underline{F},k}} z m_{\mathbf{B}_N}(z) \frac{m'_{\underline{\mathbf{B}}_N}(z)}{m^2_{\underline{\mathbf{B}}_N}(z)} dz.$$

- From exact separation (Bai and Silverstein, 1998), $\mathcal{C}_{\underline{F},k}$ encloses exactly the "expected" eigenvalues, almost surely for all large N.
- The integral gives the estimator

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

with $\mathcal{N}_k$ the indexes of cluster k and $\mu_1 \leq \ldots \leq \mu_N$ are the ordered eigenvalues of the matrix $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$.

Termination

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.

- If remains to compute the integral from residue calculus.

$$\hat{t}_k = \frac{N}{N_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_{\underline{F},k}} z m_{\mathbf{B}_N}(z) \frac{m'_{\underline{\mathbf{B}}_N}(z)}{m^2_{\underline{\mathbf{B}}_N}(z)} dz.$$

- From exact separation (Bai and Silverstein, 1998), $\mathcal{C}_{\underline{F},k}$ encloses exactly the "expected" eigenvalues, almost surely for all large N.
- The integral gives the estimator

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

with $\mathcal{N}_k$ the indexes of cluster k and $\mu_1 \leq \ldots \leq \mu_N$ are the ordered eigenvalues of the matrix $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^{\mathsf{T}}$.

Application to the current model

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

- Extending $\mathbf{Y}$ with zeros, our model is a "double sample covariance matrix"

$$\underbrace{\mathbf{Y}}_{(N+n)\times M} = \underbrace{\begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma \mathbf{I}_N \\ 0 & 0 \end{bmatrix}}_{(N+n)\times(N+n)} \underbrace{\begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix}}_{(N+n)\times M}.$$

- Limiting distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$

Theorem (l.s.d. of $\mathbf{B}_N$)

Let $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ *with eigenvalues* $\lambda_1, \ldots, \lambda_N$. *Denote* $m_{\underline{\mathbf{B}}_N}(z) \triangleq \frac{1}{M}\sum_{k=1}^{M} \frac{1}{\lambda_k - z}$, *with* $\lambda_i = 0$ *for* $i > N$. *Then, for* $M/N \to c$, $N/n_k \to c_k$, $N/n \to c_0$, *for any* $z \in \mathbb{C}^+$,

$$m_{\underline{\mathbf{B}}_N}(z) \xrightarrow{\text{a.s.}} m_{\underline{F}}(z)$$

with $m_{\underline{F}}(z)$ *the unique solution in* $\mathbb{C}^+$ *of*

$$\frac{1}{m_{\underline{F}}(z)} = -\sigma^2 + \frac{1}{f(z)}\left[\frac{c_0 - 1}{c_0} + m_P\left(-\frac{1}{f(z)}\right)\right], \quad \text{with } f(z) = (c-1)m_{\underline{F}}(z) - czm_{\underline{F}}(z)^2.$$

Application to the current model (2)

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

- estimator calculus

Theorem (Estimator of $P_1, \dots, P_K$)

Let $\mathbf{B}_N \in \mathbb{C}^{N \times N}$ be defined as above and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$, $\lambda_1 < \dots < \lambda_N$. Assume that asymptotic cluster separability condition is fulfilled for some k. Then, as N, n, $M \to \infty$,

$$\hat{P}_k - P_k \xrightarrow{\text{a.s.}} 0,$$

where

$$\hat{P}_k = \frac{NM}{n_k(M-N)} \sum_{i \in \mathcal{N}_k} (\eta_i - \mu_i)$$

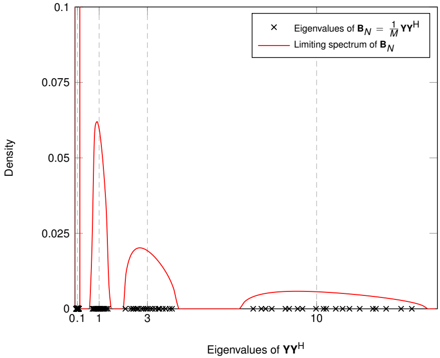
with $\mathcal{N}_k$ the set indexing the eigenvalues in cluster k of F, $\eta_1 < \dots < \eta_N$ the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ and $\mu_1 < \dots < \mu_N$ the eigenvalues of $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$.

Remarks

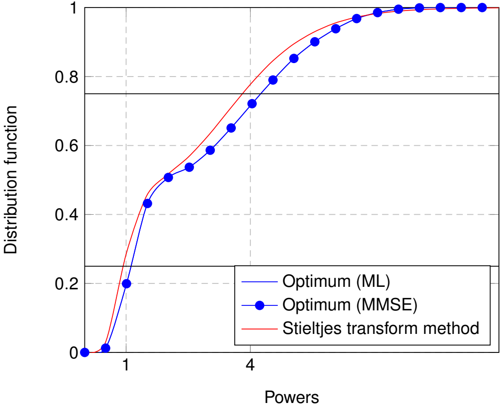
- solution is computationally simple, explicit, and the final formula compact.
- cluster separability condition is fundamental. This requires
  - for all other parameters fixed, the P_k cannot be too close to one another: source separation problem.
  - for all other parameters fixed, σ^2 must be kept low: low SNR undecidability problem.
  - for all other parameters fixed, M/N cannot be too low: sample deficiency issue (not such an issue though).
  - for all other parameters fixed, N/n cannot be too low: diversity issue.
- exact spectrum separability is an essential ingredient (known for very few models to this day).

Remarks

- solution is computationally simple, explicit, and the final formula compact.
- cluster separability condition is fundamental. This requires
  - for all other parameters fixed, the P_k cannot be too close to one another: source separation problem.
  - for all other parameters fixed, σ^2 must be kept low: low SNR undecidability problem.
  - for all other parameters fixed, M/N cannot be too low: sample deficiency issue (not such an issue though).
  - for all other parameters fixed, N/n cannot be too low: diversity issue.
- exact spectrum separability is an essential ingredient (known for very few models to this day).



Stieltjes transform method vs. optimum



MSE	P_1	P_2
Opt. MMSE	0.1239	0.1278
Stieltjes	0.1514	0.1332

Figure: Distribution function for the detection of two power sources, $P_1 = 1$, $P_2 = 4$, $n_1 = n_2 = 1$, $M = N = 16$. Optimum against Stieltjes transform method.

Stieltjes transform method vs. conventional method

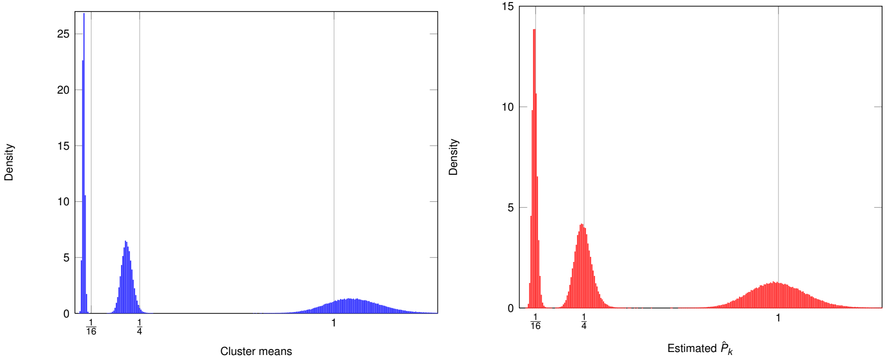


Figure: Histogram of the cluster-mean approach and of $\hat{P}_k$ for $k \in \{1,2,3\}$, $P_1 = 1/16$, $P_2 = 1/4$, $P_3 = 1$, $n_1 = n_2 = n_3 = 4$ antennas per user, $N = 24$ sensors, $M = 128$ samples and $\mathrm{SNR} = 20$ dB.

Performance comparison

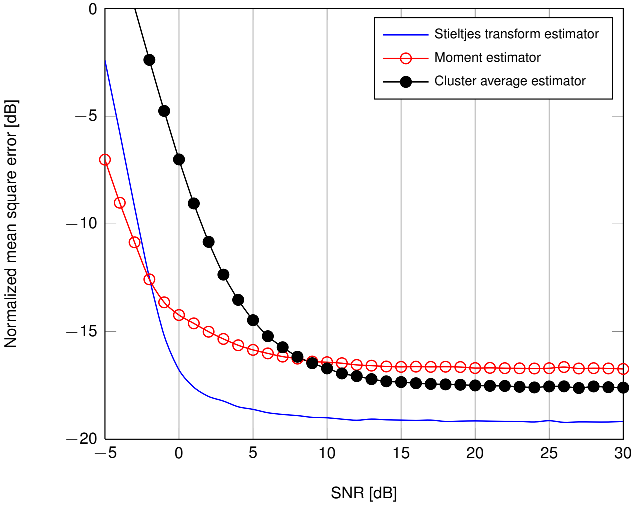


Figure: Normalized mean square error of largest estimated power $\hat{P}_3$, $P_1 = 1/16$, $P_2 = 1/4$, $P_3 = 1$, $n_1 = n_2 = n_3 = 4$,$N = 24$, $M = 128$. Comparison between classical, moment and Stieltjes transform approaches.

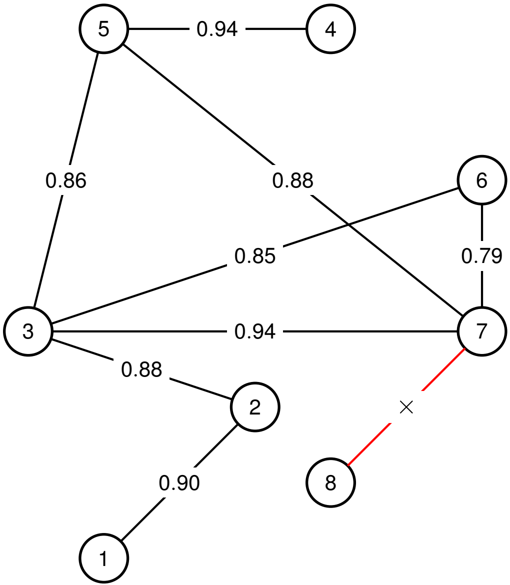
Related bibliography

- N. El Karoui, “Spectrum estimation for large dimensional covariance matrices using random matrix theory,” Annals of Statistics, vol. 36, no. 6, pp. 2757-2790, 2008.
- 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.
- R. Couillet, M. Debbah, “Free deconvolution for OFDM multicell SNR detection”, PIMRC 2008, Cannes, France.
- 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, M. Debbah, “Eigen-inference for multi-source power estimation”, *submitted to ISIT 2010*.
- 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, “CLT of linear spectral statistics of large dimensional sample covariance matrices,” Annals of Probability, vol. 32, no. 1A, pp. 553-605, 2004.
- J. Silverstein, Z. Bai, “Exact separation of eigenvalues of large dimensional sample covariance matrices” Annals of Probability, vol. 27, no. 3, pp. 1536-1555, 1999.
- Ø. Ryan, M. Debbah, “Free Deconvolution for Signal Processing Applications,” IEEE International Symposium on Information Theory, pp. 1846-1850, 2007.

Outline

Outline

Failure detection



Node failure detection in sensor networks

- Consider the model
$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \sigma\mathbf{w}$$
with $\mathbf{H} \in \mathbb{C}^{N \times p}$ deterministic, $\boldsymbol{\theta} \sim \mathcal{CN}(0, \mathbf{I}_p)$, $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}_N)$.
- In particular $\mathrm{E}[\mathbf{y}] = 0$ and $\mathrm{E}[\mathbf{y}\mathbf{y}^\mathsf{H}] = \mathbf{R} \triangleq \mathbf{H}\mathbf{H}^\mathsf{H} + \sigma^2\mathbf{I}_N$
- With $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}$,
$$\mathrm{E}[\mathbf{s}\mathbf{s}^\mathsf{H}] = \mathbf{I}_N.$$
- Upon failure of sensor k, $\mathbf{y}$ becomes
$$\mathbf{y}' = (\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^\mathsf{H})\mathbf{H}\boldsymbol{\theta} + \sigma_k\mathbf{e}_k\mathbf{e}_k^*\boldsymbol{\theta}' + \sigma\mathbf{w}$$
for some noise variance σ_k^2.
- Now $\mathrm{E}[\mathbf{y}'] = 0$ and
$$\mathrm{E}[\mathbf{y}'\mathbf{y}'^\mathsf{H}] = (\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^\mathsf{H})\mathbf{H}\mathbf{H}^\mathsf{H}(\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^\mathsf{H}) + \sigma_k^2\mathbf{e}_k\mathbf{e}_k^\mathsf{H} + \sigma^2\mathbf{I}_N.$$
- With now $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}'$,
$$\mathrm{E}[\mathbf{s}\mathbf{s}^\mathsf{H}] = \mathbf{I}_N + \mathbf{P}_k$$
with
$$\mathbf{P}_k = -\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^\mathsf{H}\mathbf{e}_k\mathbf{e}_k^\mathsf{H}\mathbf{R}^{-\frac{1}{2}} + \mathbf{R}^{-\frac{1}{2}}\mathbf{e}_k\left[(\mathbf{e}_k^\mathsf{H}\mathbf{H}\mathbf{H}^\mathsf{H}\mathbf{e}_k + \sigma_k^2)\mathbf{e}_k^\mathsf{H}\mathbf{R}^{-\frac{1}{2}} - \mathbf{e}_k^\mathsf{H}\mathbf{H}\mathbf{H}^\mathsf{H}\mathbf{R}^{-\frac{1}{2}}\right]$$
of rank-2 (image of $\mathbf{P}_k$ in $\mathrm{Span}(\mathbf{R}^{-\frac{1}{2}}\mathbf{e}_k, \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^\mathsf{H}\mathbf{e}_k)$)

Node failure detection in sensor networks

- Consider the model

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \sigma \mathbf{w}$$

with $\mathbf{H} \in \mathbb{C}^{N \times p}$ deterministic, $\boldsymbol{\theta} \sim \mathcal{CN}(0, \mathbf{I}_p)$, $\mathbf{w} \sim \mathcal{CN}(0, \mathbf{I}_N)$.

- In particular $\mathrm{E}[\mathbf{y}] = 0$ and $\mathrm{E}[\mathbf{y}\mathbf{y}^{\mathsf{H}}] = \mathbf{R} \triangleq \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N$
- With $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}$,

$$\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N.$$

- Upon failure of sensor k, $\mathbf{y}$ becomes

$$\mathbf{y}' = (\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^{\mathsf{H}})\mathbf{H}\boldsymbol{\theta} + \sigma_k \mathbf{e}_k \mathbf{e}_k^{*} \boldsymbol{\theta}' + \sigma \mathbf{w}$$

for some noise variance σ_k^2.

- Now $\mathrm{E}[\mathbf{y}'] = 0$ and

$$\mathrm{E}[\mathbf{y}'\mathbf{y}'^{\mathsf{H}}] = (\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^{\mathsf{H}})\mathbf{H}\mathbf{H}^{\mathsf{H}}(\mathbf{I}_N - \mathbf{e}_k\mathbf{e}_k^{\mathsf{H}}) + \sigma_k^2 \mathbf{e}_k\mathbf{e}_k^{\mathsf{H}} + \sigma^2 \mathbf{I}_N.$$

- With now $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}'$,

$$\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$$

with

$$\mathbf{P}_k = -\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}} + \mathbf{R}^{-\frac{1}{2}}\mathbf{e}_k\left[(\mathbf{e}_k^{\mathsf{H}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{e}_k + \sigma_k^2)\mathbf{e}_k^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}} - \mathbf{e}_k^{\mathsf{H}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}}\right]$$

of rank-2 (image of $\mathbf{P}_k$ in $\mathrm{Span}(\mathbf{R}^{-\frac{1}{2}}\mathbf{e}_k, \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{H}^{\mathsf{H}}\mathbf{e}_k)$)

Sudden parameter change detection in sensor networks

- Upon sudden change of parameter θ_k,

$$\mathbf{y}' = \mathbf{H}(\mathbf{I}_p + \alpha_k \mathbf{e}_k \mathbf{e}_k^*)\boldsymbol{\theta} + \mu_k \mathbf{H}\mathbf{e}_k + \sigma \mathbf{w}$$

- Then

$$\mathrm{E}[\mathbf{y}'\mathbf{y}'^{\mathsf{H}}] = \mathbf{H}(\mathbf{I}_p + [\mu_k^2 + (1+\alpha_k)^2 - 1]\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}})\mathbf{H}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N.$$

- With $\mathbf{R} = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2 \mathbf{I}_N$ and $\mathbf{s} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}'$,

$$\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$$

with

$$\mathbf{P}_k = [\mu_k^2 + (1+\alpha_k)^2 - 1]\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{e}_k\mathbf{e}_k^{\mathsf{H}}\mathbf{H}^{\mathsf{H}}\mathbf{R}^{-\frac{1}{2}}.$$

Outline

Classical approach

- With K the number of failure scenarios, hypothesis test between:
  - no failure
  - failure of type 1
  - ...
  - failure of type K
- Maximum-likelihood approach computationally constraining!

$$\text{calculus cost } \simeq O(N^3 K)$$

which is

$$\text{calculus cost } \simeq O(N^{3+m})$$

for m simultaneous node failures detection.

- Ad-hoc approaches/PCA can reduce this amount
- We propose here a “maximum-likelihood-type” method in

$$\text{one SVD } + O(K)$$

Classical approach

- With K the number of failure scenarios, hypothesis test between:
  - no failure
  - failure of type 1
  - ...
  - failure of type K
- Maximum-likelihood approach computationally constraining!

$$\text{calculus cost} \simeq O(N^3 K)$$

which is

$$\text{calculus cost} \simeq O(N^{3+m})$$

for m simultaneous node failures detection.

- Ad-hoc approaches/PCA can reduce this amount
- We propose here a "maximum-likelihood-type" method in

$$\text{one SVD} + O(K)$$

Classical approach

- With K the number of failure scenarios, hypothesis test between:
  - no failure
  - failure of type 1
  - ...
  - failure of type K
- Maximum-likelihood approach computationally constraining!

$$\text{calculus cost } \simeq O(N^3 K)$$

  which is

$$\text{calculus cost } \simeq O(N^{3+m})$$

  for m simultaneous node failures detection.
- Ad-hoc approaches/PCA can reduce this amount
- We propose here a "maximum-likelihood-type" method in

$$\text{one SVD } + O(K)$$

Classical approach

- With K the number of failure scenarios, hypothesis test between:
  - no failure
  - failure of type 1
  - ...
  - failure of type K
- Maximum-likelihood approach computationally constraining!

$$\text{calculus cost } \simeq O(N^3 K)$$

  which is

$$\text{calculus cost } \simeq O(N^{3+m})$$

  for m simultaneous node failures detection.
- Ad-hoc approaches/PCA can reduce this amount
- We propose here a "maximum-likelihood-type" method in

$$\text{one SVD } + O(K)$$

Failure detection and identification

R. Couillet and W. Hachem, “Local failure detection and identification in large sensor networks,” *submitted to* IEEE Transaction on Information Theory, 2011.

- Upon reception of $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]$,
  - Failure detection based on hypothesis test
    - $\mathcal{H}_0$: no failure
    - $\bar{\mathcal{H}}_0$: failure
  - If $\bar{\mathcal{H}}_0$ is decided, multi-hypothesis test

$$\mathcal{H}_k = \text{“failure of type } k\text{”}$$

- Detection test on largest eigenvalue $\hat{\lambda}_1$ of $\frac{1}{n}\mathbf{S}\mathbf{S}^{\mathrm{H}}$: for a false alarm rate η,

$$\hat{\lambda}_1' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta)$$

with

$$\hat{\lambda}_1' = N^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c_N})^2}{(1+\sqrt{c_N})^{\frac{4}{3}} c_N^{\frac{1}{2}}}$$

and T_2 the complex Tracy-Widom distribution.

Failure detection and identification

R. Couillet and W. Hachem, "Local failure detection and identification in large sensor networks," *submitted to* IEEE Transaction on Information Theory, 2011.

- Upon reception of $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]$,
  - Failure detection based on hypothesis test
    - $\mathcal{H}_0$: no failure
    - $\bar{\mathcal{H}}_0$: failure
  - If $\bar{\mathcal{H}}_0$ is decided, multi-hypothesis test

$$\mathcal{H}_k = \text{“failure of type } k\text{”}$$

- Detection test on largest eigenvalue $\hat{\lambda}_1$ of $\frac{1}{n}\mathbf{S}\mathbf{S}^{\mathsf{H}}$: for a false alarm rate η,

$$\hat{\lambda}_1' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta)$$

with

$$\hat{\lambda}_1' = N^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1+\sqrt{c_N})^2}{(1+\sqrt{c_N})^{\frac{4}{3}} c_N^{\frac{1}{2}}}$$

and T_2 the complex Tracy-Widom distribution.

Failure localization

- For localization, eigenvalues are poor statistics
- Denote, in case of failure of type k

$$\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = \mathbf{I}_N + \omega_k \mathbf{u}_{k,1}\mathbf{u}_{k,1}^{\mathrm{H}}$$

(rank-1 perturbation for simplicity)

- We use the eigenvector $\hat{\mathbf{u}}_1$ corresponding to λ_1, and

$$|\hat{\mathbf{u}}_1^{\mathrm{H}}\mathbf{u}_{k,1}|^2 \xrightarrow{\text{a.s.}} \xi(\omega_k) > 0$$

for k the failure index.

- With the CLT on $|\hat{\mathbf{u}}_1^{\mathrm{H}}\mathbf{u}_{k,1}|^2 - \xi(\omega_k)$, we have the estimator

$$k^\star = \arg\max_{1\leq k\leq K} f\left(\sqrt{N}(|\hat{\mathbf{u}}_1^{\mathrm{H}}\mathbf{u}_{k,1}|^2 - \xi(\omega_k)); \sigma_k^2\right)$$

with f the Gaussian density.

- Test can be reinforced by including
  - projection statistics on other vectors
  - statistics of eigenvalues
  - take the joint probability over multiple spikes.
- Further generalizations are possible assuming unknown failure amplitude.

Failure localization

- For localization, eigenvalues are poor statistics
- Denote, in case of failure of type k

$$\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \omega_k \mathbf{u}_{k,1} \mathbf{u}_{k,1}^{\mathsf{H}}$$

  (rank-1 perturbation for simplicity)
- We use the eigenvector $\hat{\mathbf{u}}_1$ corresponding to λ_1, and

$$|\hat{\mathbf{u}}_1^{\mathsf{H}} \mathbf{u}_{k,1}|^2 \xrightarrow{\text{a.s.}} \xi(\omega_k) > 0$$

  for k the failure index.
- With the CLT on $|\hat{\mathbf{u}}_1^{\mathsf{H}} \mathbf{u}_{k,1}|^2 - \xi(\omega_k)$, we have the estimator

$$k^\star = \arg \max_{1 \leq k \leq K} f\left(\sqrt{N}(|\hat{\mathbf{u}}_1^{\mathsf{H}} \mathbf{u}_{k,1}|^2 - \xi(\omega_k)); \sigma_k^2\right)$$

  with f the Gaussian density.
- Test can be reinforced by including
  - projection statistics on other vectors
  - statistics of eigenvalues
  - take the joint probability over multiple spikes.
- Further generalizations are possible assuming unknown failure amplitude.

Failure localization

- For localization, eigenvalues are poor statistics
- Denote, in case of failure of type k

$$\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \omega_k \mathbf{u}_{k,1}\mathbf{u}_{k,1}^{\mathsf{H}}$$

  (rank-1 perturbation for simplicity)
- We use the eigenvector $\hat{\mathbf{u}}_1$ corresponding to λ_1, and

$$|\hat{\mathbf{u}}_1^{\mathsf{H}}\mathbf{u}_{k,1}|^2 \xrightarrow{\text{a.s.}} \xi(\omega_k) > 0$$

  for k the failure index.
- With the CLT on $|\hat{\mathbf{u}}_1^{\mathsf{H}}\mathbf{u}_{k,1}|^2 - \xi(\omega_k)$, we have the estimator

$$k^\star = \arg\max_{1\leq k\leq K} f\left(\sqrt{N}(|\hat{\mathbf{u}}_1^{\mathsf{H}}\mathbf{u}_{k,1}|^2 - \xi(\omega_k)); \sigma_k^2\right)$$

  with f the Gaussian density.
- Test can be reinforced by including
  - projection statistics on other vectors
  - statistics of eigenvalues
  - take the joint probability over multiple spikes.
- Further generalizations are possible assuming unknown failure amplitude.

Failure localization

- For localization, eigenvalues are poor statistics
- Denote, in case of failure of type k

$$\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathrm{H}}] = \mathbf{I}_N + \omega_k \mathbf{u}_{k,1}\mathbf{u}_{k,1}^{\mathrm{H}}$$

(rank-1 perturbation for simplicity)

- We use the eigenvector $\hat{\mathbf{u}}_1$ corresponding to λ_1, and

$$|\hat{\mathbf{u}}_1^{\mathrm{H}}\mathbf{u}_{k,1}|^2 \xrightarrow{\text{a.s.}} \xi(\omega_k) > 0$$

for k the failure index.

- With the CLT on $|\hat{\mathbf{u}}_1^{\mathrm{H}}\mathbf{u}_{k,1}|^2 - \xi(\omega_k)$, we have the estimator

$$k^\star = \arg\max_{1\leq k\leq K} f\left(\sqrt{N}(|\hat{\mathbf{u}}_1^{\mathrm{H}}\mathbf{u}_{k,1}|^2 - \xi(\omega_k)); \sigma_k^2\right)$$

with f the Gaussian density.

- Test can be reinforced by including
  - projection statistics on other vectors
  - statistics of eigenvalues
  - take the joint probability over multiple spikes.
- Further generalizations are possible assuming unknown failure amplitude.

Performance results

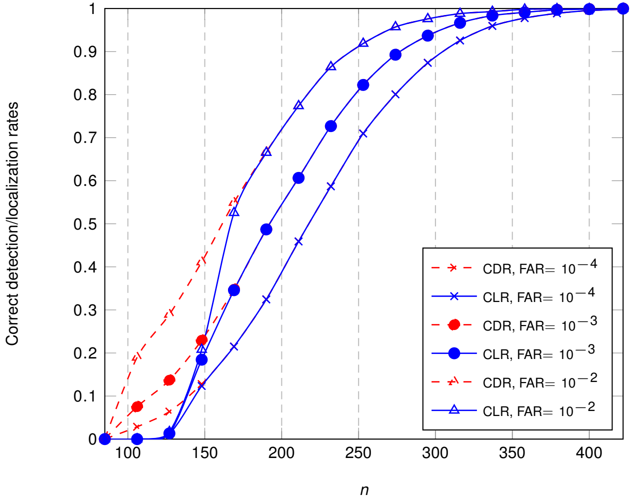


Figure: Correct detection (CDR) and localization (CLR) rates for different false alarm rates (FAR) and different n, worst case node failure in a 100-node network.

Selected authors' recent bibliography

- Articles in Journals,

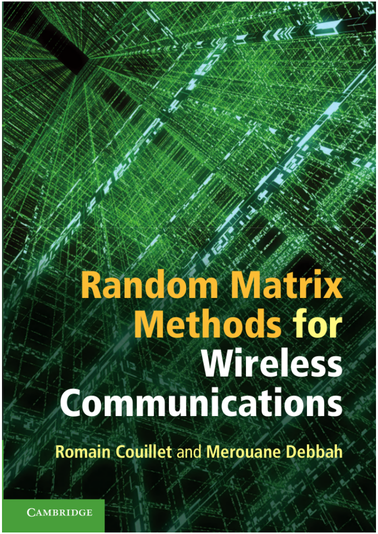
  - R. Couillet, W. Hachem, "Local failure detection and identification in large sensor networks," IEEE Transactions on Information Theory, *submitted*.
  - R. Couillet, J. Hoydis, M. Debbah, "Random Unitary Beamforming over Correlated Fading Channels," IEEE Transactions on Information Theory, *submitted*.
  - R. Couillet, J. Hoydis, M. Debbah, "A deterministic equivalent approach to the performance analysis of isometric random precoded systems," IEEE Transactions on Information Theory, *submitted*.
  - R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," IEEE Trans. on Information Theory, 2010, *to be published*.
  - R. Couillet, J. W. Silverstein, M. Debbah, "A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-User MIMO Channels," IEEE Trans. on Information Theory, *to be published*.
  - P. Bianchi, J. Najim, M. Maida, M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," IEEE Trans. on Information Theory, *to be published*.
  - R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Sensing," IEEE Trans. on Signal Processing, vol. 58, no. 10, pp. 5186-5195 ,2010.
  - S. Wagner, R. Couillet, M. Debbah, D. Slock, "Large System Analysis of Linear Precoding in MISO Broadcast Channels with Limited Feedback," IEEE Trans. on Information Theory, 2010, *submitted*.
  - A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Gaussian Finite Dimensional Statistical Inference," IEEE Trans. on Information Theory, 2009, *submitted*.
  - Ø. Ryan, M. Debbah, "Asymptotic Behaviour of Random Vandermonde Matrices with Entries on the Unit Circle," IEEE Trans. on Information Theory, vol. 55, no. 7 pp. 3115-3148, July 2009.
  - M. Debbah, R. Müller, "MIMO channel modeling and the principle of maximum entropy," IEEE Trans. on Information Theory, vol. 51, no. 5, pp. 1667-1690, 2005.

Selected authors' recent bibliography

- Articles in International Conferences
  - A. Kammoun, R. Couillet, J. Najim, M. Debbah, "A G-estimator for rate adaption in cognitive radios," *submitted to* IEEE International Symposium on Information Theory, St Petersburg, Russia, 2011.
  - J. Yao, R. Couillet, J. Najim, E. Moulines, M. Debbah, "CLT for eigen-inference methods in cognitive radios," IEEE International Conf. on Acoustics, Speech and Signal Proc., Prague, Czech Rep., 2011.
  - J. Hoydis, R. Couillet, M. Debbah, "Deterministic Equivalents for the Performance Analysis of Isometric Random Precoded Systems," IEEE International Conference on Communications, Kyoto, Japan, 2011.
  - J. Hoydis, J. Najim, R. Couillet, M. Debbah, "Fluctuations of the Mutual Information in Large Distributed Antenna Systems with Colored Noise," Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing, Allerton, IL, USA, 2010.
  - R. Couillet, S. Wagner, M. Debbah, A. Silva, "The Space Frontier: Physical Limits of Multiple Antenna Information Transfer", Inter-Perf 2008, Athens, Greece. **BEST STUDENT PAPER AWARD.**
  - R. Couillet, M. Debbah, V. Poor, "Self-organized spectrum sharing in large MIMO multiple access channels", submitted to ISIT 2010.
  - L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, "Cooperative spectrum sensing using random matrix theory," 3rd International Symposium on Wireless Pervasive Computing (ISWPC), 2008.
  - R. Couillet, M. Debbah, "Uplink capacity of self-organizing clustered orthogonal CDMA networks in flat fading channels", ITW 2009 Fall, Taormina, Sicily.

- Book Chapters
  - **Mathematical Foundations for Signal Processing, Communications and Networking**
    - *Editors:* T. Chen, D. Rajan and E. Serpedin
    - *Chapter title:* "Random matrix theory"
    - *Chapter authors:* R. Couillet and M. Debbah
    - *Publisher:* CRC Press, Taylor & and Francis Group
    - *Year:* 2011 (to appear)

Coming up soon...



Coming up soon...

Romain Couillet, Mérouane Debbah, *Random Matrix Methods for Wireless Communications*.

1. Theoretical aspects
   1. Random matrices
   2. The Stieltjes transform method
   3. Free probability theory
   4. Combinatoric approaches
   5. Deterministic equivalents
   6. Spectrum analysis
   7. Eigen-inference
   8. Extreme eigenvalues
   9. Summary and partial conclusions
2. Applications to wireless communications
   1. Introduction to applications in telecommunications
   2. System performance of CDMA technologies
   3. Performance of multiple antennas systems
   4. Rate performance in multiple access and broadcast channels
   5. Performance of multi-cellular and relay networks
   6. Detection
   7. Estimation
   8. System modelling
   9. Perspectives
   10. Conclusion