# Eigen-Inference Statistical Methods for Cognitive Radio

**Romain Couillet**[1,2], **Mérouane Debbah**[1]
[1]Alcatel-Lucent Chair on Flexible Radio, Supélec, Gif sur Yvette, FRANCE
[2]ST-Ericsson, Sophia-Antipolis, FRANCE
{*romain.couillet,merouane.debbah*}*@supelec.fr*

European Wireless

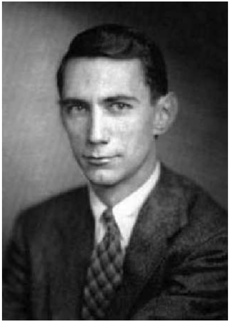
# Outline

# Outline

# 1948: Cybernetics and Theory of Communications

C. E. Shannon, “A Mathematical Theory of Communication,” Bell System Technical Journal, 1948.

N. Wiener, “Cybernetics, or Control and Communication in the Animal and the Machine,” Herman et Cie, The Technology Press, 1948.



Claude Shannon, 1916-2001

Norbert Wiener, 1894-1964

# Information and Noise *against* Black Box and Feedback

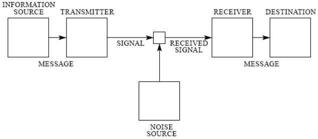



Fig. 1—Schematic diagram of a general communication system.

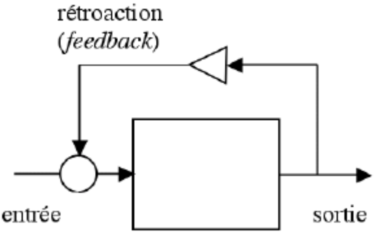
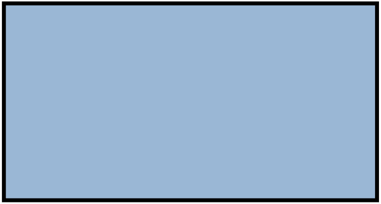
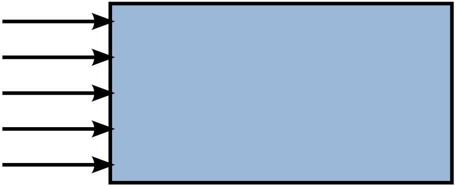



Fig. 1 – Boucle de rétroaction

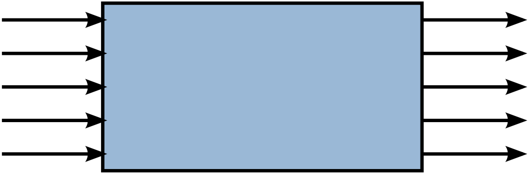
# 2008: 60 years later... MIMO Random Networks



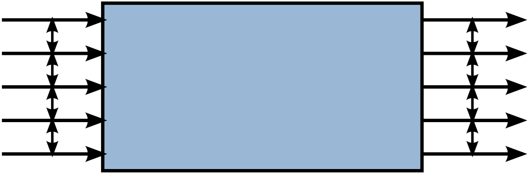
# 2008: 60 years later... MIMO Random Networks



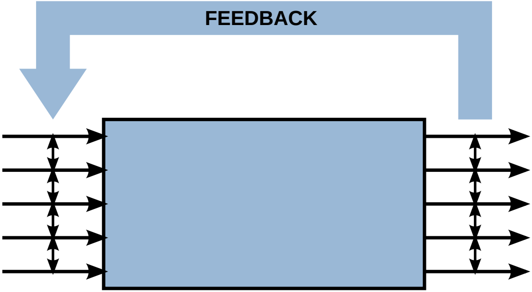
# 2008: 60 years later... MIMO Random Networks



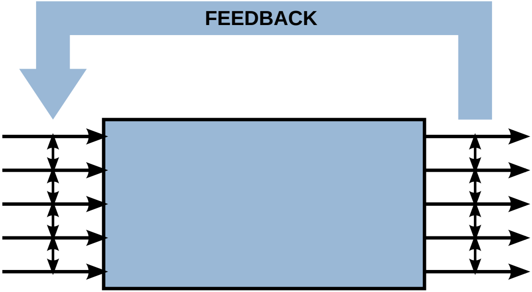
# 2008: 60 years later... Flexible MIMO Random Networks



# 2008: 60 years later... Flexible MIMO Random Networks



# 2008: 60 years later... Mobile Flexible MIMO Random Networks

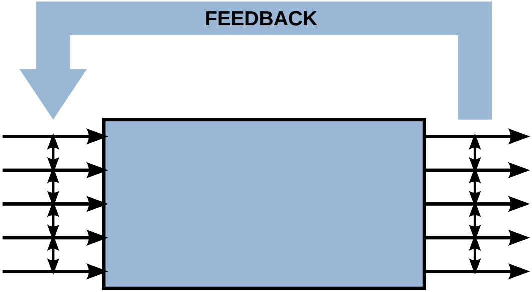



We must learn and control the black box

- within a fraction of time
- with finite energy.

In many cases, the number of inputs/outputs (the dimensionality of the system) is of the same order as the time scale changes of the box.

# 2008: 60 years later... Mobile Flexible MIMO Random Networks

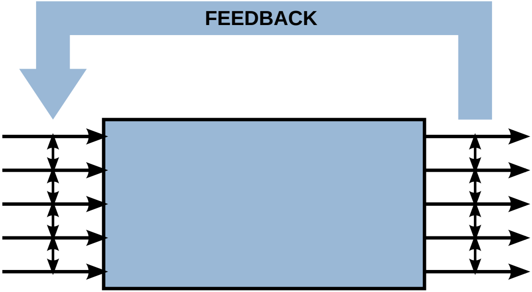



We must learn and control the black box

- within a fraction of time
- with finite energy.

In many cases, the number of inputs/outputs (the dimensionality of the system) is of the same order as the time scale changes of the box.

# 2008: 60 years later... Mobile Flexible MIMO Random Networks

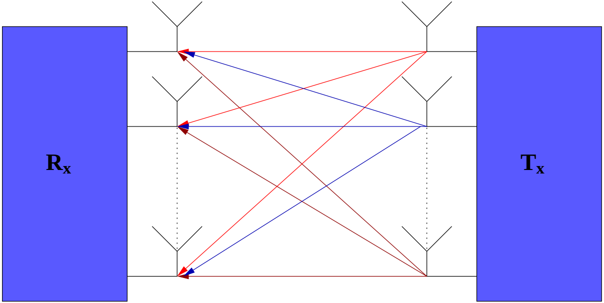



We must learn and control the black box

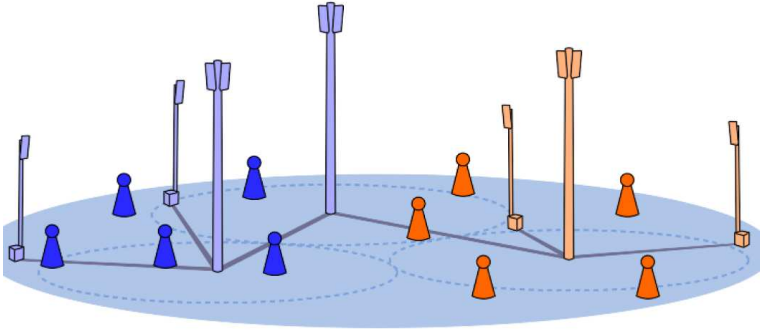
- within a fraction of time
- with finite energy.

In many cases, the number of inputs/outputs (the dimensionality of the system) is of the same order as the time scale changes of the box.

# Example: Multi-antenna systems

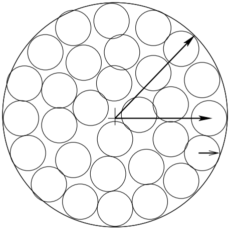


# Example: Cognitive Network MIMO



# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

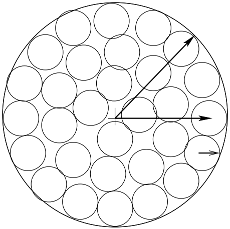


$$C = H(\mathbf{y}) - H(\mathbf{y} \mid \mathbf{x})$$
$$= \log\det(\pi e \mathbf{R}_y) - \log\det(\pi e \mathbf{R}_n)$$

$$C = \log\left(\frac{\det(\mathbf{R}_y)}{\det(\mathbf{R}_n)}\right)$$

# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

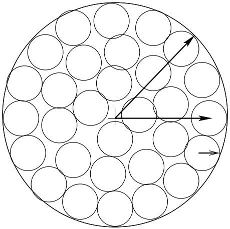


$$C = H(\mathbf{y}) - H(\mathbf{y} \mid \mathbf{x})$$
$$= \log\det\left(\pi e \mathbf{R}_y\right) - \log\det\left(\pi e \mathbf{R}_n\right)$$

$$C = \log\left(\frac{\det\left(\mathbf{R}_y\right)}{\det\left(\mathbf{R}_n\right)}\right)$$

# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

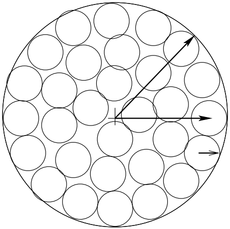


$$C = H(\mathbf{y}) - H(\mathbf{y} \mid \mathbf{x})$$
$$= \log\det(\pi e \mathbf{R}_y) - \log\det(\pi e \mathbf{R}_n)$$

$$C = \log\left(\frac{\det(\mathbf{R}_y)}{\det(\mathbf{R}_n)}\right)$$

# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$



$$C = H(\mathbf{y}) - H(\mathbf{y} \mid \mathbf{x})$$
$$= \log\det(\pi e \mathbf{R}_y) - \log\det(\pi e \mathbf{R}_n)$$

$$C = \log\left(\frac{\det(\mathbf{R}_y)}{\det(\mathbf{R}_n)}\right)$$

# Understanding the network in a finite time

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

- $$\text{Rate} = \log \frac{\det(\mathbf{R_y})}{\det(\mathbf{R_n})}$$
- In the Gaussian case, one can write
$$\mathbf{y}_i = \mathbf{R_y}^{\frac{1}{2}}\mathbf{u_i}$$
where $\mathbf{u_i}$ is zero mean i.i.d Gaussian.
- One has only $n$ samples:
$$\hat{\mathbf{R}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^H = \mathbf{R_y}^{\frac{1}{2}}\left(\frac{1}{L}\mathbf{U}\mathbf{U}^H\right)\mathbf{R_y}^{\frac{1}{2}} \to \frac{1}{L}\mathbf{U}\mathbf{U}^H\mathbf{R_y}$$
- The non-zero eigenvalues of $\hat{\mathbf{R}}$ are the same as the eigenvalues of $\frac{1}{L}\mathbf{U}\mathbf{U}^H\mathbf{R_y}$.
- We know the eigenvalues of $\frac{1}{L}\mathbf{U}\mathbf{U}^H$ and $\hat{\mathbf{R}}$. Can we determine the eigenvalues of $\mathbf{R_y}$?

# Understanding the network in a finite time

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

- $$\text{Rate} = \log \frac{\det(\mathbf{R_y})}{\det(\mathbf{R_n})}$$
- In the Gaussian case, one can write
$$\mathbf{y}_i = \mathbf{R_y}^{\frac{1}{2}} \mathbf{u_i}$$
where $\mathbf{u_i}$ is zero mean i.i.d Gaussian.
- One has only $n$ samples:
$$\hat{\mathbf{R}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i^H = \mathbf{R_y}^{\frac{1}{2}} \left(\frac{1}{L}\mathbf{U}\mathbf{U}^H\right)\mathbf{R_y}^{\frac{1}{2}} \to \frac{1}{L}\mathbf{U}\mathbf{U}^H\mathbf{R_y}$$
- The non-zero eigenvalues of $\hat{\mathbf{R}}$ are the same as the eigenvalues of $\frac{1}{L}\mathbf{U}\mathbf{U}^H\mathbf{R_y}$.
- We know the eigenvalues of $\frac{1}{L}\mathbf{U}\mathbf{U}^H$ and $\hat{\mathbf{R}}$. Can we determine the eigenvalues of $\mathbf{R_y}$?

# Understanding the network in a finite time

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

- $$\text{Rate} = \log \frac{\det(\mathbf{R_y})}{\det(\mathbf{R_n})}$$
- In the Gaussian case, one can write
$$\mathbf{y}_i = \mathbf{R_y}^{\frac{1}{2}}\mathbf{u_i}$$
where $\mathbf{u_i}$ is zero mean i.i.d Gaussian.
- One has only $n$ samples:
$$\hat{\mathbf{R}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^H = \mathbf{R_y}^{\frac{1}{2}}\left(\frac{1}{L}\mathbf{U}\mathbf{U}^H\right)\mathbf{R_y}^{\frac{1}{2}} \to \frac{1}{L}\mathbf{U}\mathbf{U}^H\mathbf{R_y}$$
- The non-zero eigenvalues of $\hat{\mathbf{R}}$ are the same as the eigenvalues of $\frac{1}{L}\mathbf{U}\mathbf{U}^H\mathbf{R_y}$.
- We know the eigenvalues of $\frac{1}{L}\mathbf{U}\mathbf{U}^H$ and $\hat{\mathbf{R}}$. Can we determine the eigenvalues of $\mathbf{R_y}$?

# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

The capacity per dimension is given by:

$$C = \frac{1}{N} \log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{W}\mathbf{W}^H\right) = \frac{1}{N}\sum_{i=1}^{N} \log(1 + \frac{1}{\sigma^2}\lambda_i) = \int \log(1 + \frac{1}{\sigma^2}\lambda) f^N(\lambda) d\lambda$$

with

$$f^N(\lambda) = \frac{1}{N}\sum_{i=1}^{N} \delta(\lambda - \lambda_i)$$

All we need to know is how the empirical eigenvalue distribution behaves.
It is often sufficient to determine the moments $M_1^N, M_2^N, \ldots$

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

Sometimes, we need more involved tools, such as Fourier transform, or Stieltjes transform...

# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

The capacity per dimension is given by:

$$C = \frac{1}{N}\log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{W}\mathbf{W}^H\right) = \frac{1}{N}\sum_{i=1}^{N}\log(1 + \frac{1}{\sigma^2}\lambda_i) = \int \log(1 + \frac{1}{\sigma^2}\lambda)f^N(\lambda)d\lambda$$

with

$$f^N(\lambda) = \frac{1}{N}\sum_{i=1}^{N}\delta(\lambda - \lambda_i)$$

All we need to know is how the empirical eigenvalue distribution behaves.

It is often sufficient to determine the moments $M_1^N, M_2^N, \ldots$

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

Sometimes, we need more involved tools, such as Fourier transform, or Stieltjes transform...

# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

The capacity per dimension is given by:

$$C = \frac{1}{N} \log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{W}\mathbf{W}^H\right) = \frac{1}{N}\sum_{i=1}^{N} \log(1 + \frac{1}{\sigma^2}\lambda_i) = \int \log(1 + \frac{1}{\sigma^2}\lambda) f^N(\lambda) d\lambda$$

with

$$f^N(\lambda) = \frac{1}{N}\sum_{i=1}^{N} \delta(\lambda - \lambda_i)$$

All we need to know is how the empirical eigenvalue distribution behaves.
It is often sufficient to determine the moments $M_1^N, M_2^N, \ldots$

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

Sometimes, we need more involved tools, such as Fourier transform, or Stieltjes transform...

# Information transfer in MIMO flexible networks

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{n}$$

The capacity per dimension is given by:

$$C = \frac{1}{N} \log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{W}\mathbf{W}^H\right) = \frac{1}{N}\sum_{i=1}^{N} \log(1 + \frac{1}{\sigma^2}\lambda_i) = \int \log(1 + \frac{1}{\sigma^2}\lambda) f^N(\lambda) d\lambda$$

with

$$f^N(\lambda) = \frac{1}{N}\sum_{i=1}^{N} \delta(\lambda - \lambda_i)$$

All we need to know is how the empirical eigenvalue distribution behaves.
It is often sufficient to determine the moments $M_1^N, M_2^N, \ldots$

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

Sometimes, we need more involved tools, such as Fourier transform, or Stieltjes transform...

# Outline

# 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 $n \to \infty$,

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

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 $(i, j)$,

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

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 $\mathbf{R} = \mathrm{E}[\mathbf{w}_1 \mathbf{w}_1^{\mathsf{H}}] \in \mathbb{C}^{N \times N}$.

## Law of large numbers

As $n \to \infty$,

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

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

- if $n \gg N$,

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

  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 $\mathbf{R} = \mathrm{E}[\mathbf{w}_1 \mathbf{w}_1^{\mathsf{H}}] \in \mathbb{C}^{N \times N}$.

## Law of large numbers

As $n \to \infty$,

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

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

- if $n \gg N$,

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

  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

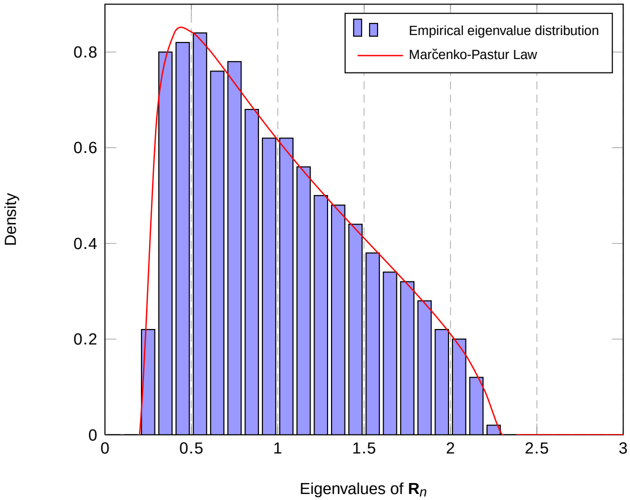



Figure: Histogram of the eigenvalues of $\mathbf{R}_n$ for $n = 2000$, $N = 500$, $\mathbf{R} = \mathbf{I}_N$

# The Marčenko-Pastur Law

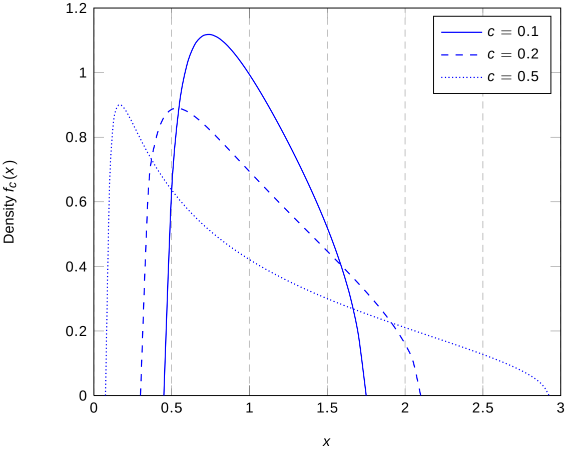



Figure: Marčenko-Pastur law for different limit ratios $c = \lim N/n$.

# Deriving the Marčenko-Pastur law

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

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

- Denoting $\alpha = N/n$, the moments of this distribution are given by

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

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

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

# Deriving the Marčenko-Pastur law

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

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

- Denoting $\alpha = N/n$, the moments of this distribution are given by

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

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

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

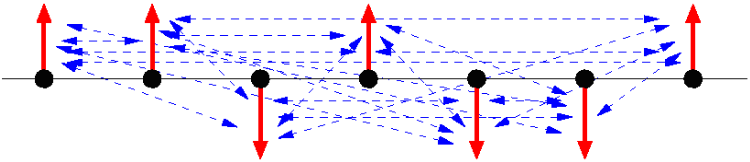
# Outline

# Wigner and semi-circle law

Schrödinger's equation

$$H \mathbf{\Phi}_i = E_i \mathbf{\Phi}_i$$

where $\mathbf{\Phi}_i$ 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

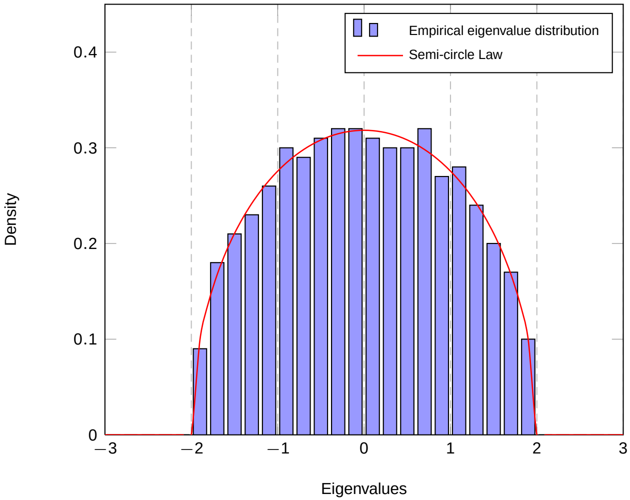



Figure: Histogram of the eigenvalues of Wigner matrices and the semi-circle law, for $N = 500$

# Circular law

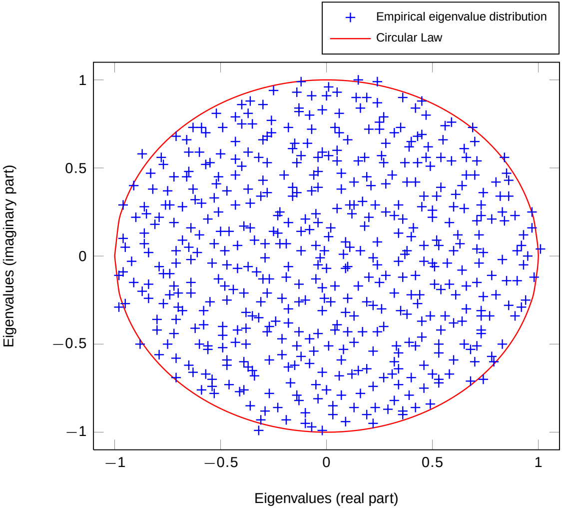



Figure: Eigenvalues of $\mathbf{X}_N$ with i.i.d. standard Gaussian entries, for $N = 500$.

# 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 $C_k$,

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

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,

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

# Non-crossing partitions

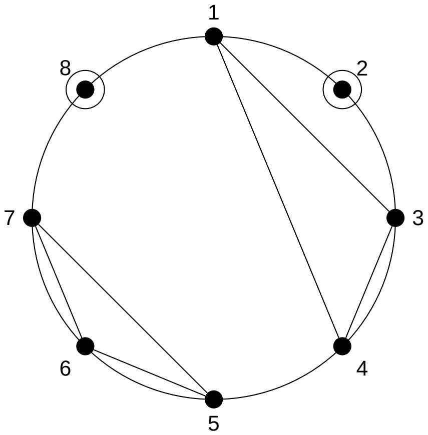



Figure: Non-crossing partition $\pi = \{\{1, 3, 4\}, \{2\}, \{5, 6, 7\}, \{8\}\}$ of $NC(8)$.

# Moments of sums and products of random matrices

- Combinatorial calculus of all moments

**Theorem**

*For free random matrices* **A** *and* **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, it is impossible to retrieve the distribution function from moments. 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, it is impossible to retrieve the distribution function from moments. This is in particular the case of Vandermonde matrices.

# Free convolution

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

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

- 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 $\mu_{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}} \Rightarrow \mu_A$, as $n/N \to c$. Then the eigenvalue distribution of

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

converges weakly and almost surely to $\mu_B$ such that

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

with $\mu_c$ the Marčenko-Pastur law with ratio $c$.

# Free convolution

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

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

- 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* $\mu_{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}} \Rightarrow \mu_A$*, as* $n/N \to c$*. Then the eigenvalue distribution of*

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

*converges weakly and almost surely to* $\mu_B$ *such that*

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

*with* $\mu_c$ *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 | $m_n = \sum_{\pi \in \mathcal{P}(n)} \prod_{V \in \pi} c_{\|V\|}$ | $M_n = \sum_{\pi \in \mathcal{NC}(n)} \prod_{V \in \pi} C_{\|V\|}$ |
| 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 $m_{\mathbf{X}} \stackrel{\Delta}{=} m_F$, 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.

# 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.

# 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.

# 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*

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

*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 $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N\mathbf{T}_N^{\frac{1}{2}}$ with asymptotic distribution $F$,

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

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

# 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*

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

*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 $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N^{\mathsf{H}}\mathbf{X}_N\mathbf{T}_N^{\frac{1}{2}}$ with asymptotic distribution $F$,

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

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

# Getting $F'$ from $m_F$

- Remember that, for $a < b$ real,

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

- 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 $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^H \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ 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, for $a < b$ real,

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

- 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 $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^H \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$ 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, for $a < b$ real,

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

- 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 $K = 4$.

# Spectrum of the sample covariance matrix

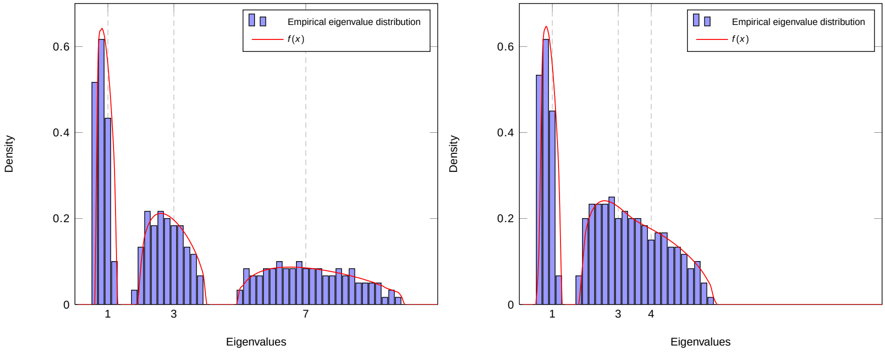



Figure: Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N^{\mathsf{H}} \mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}}$, $N = 3000$, $n = 300$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 on top, (ii) 1, 3 and 4 at bottom.

# 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 $\mathbf{X}\mathbf{X}^{\mathrm{H}} \in \mathbb{C}^{N \times N}$,

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

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

# 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 $\sigma_{i,j}^2$.
  - 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 $\mathbf{H}_k = \sum_{l=1}^{L} \mathbf{R}'^{\frac{1}{2}}_{kl}\mathbf{X}_{kl}\mathbf{T}'_{kl}\mathbf{X}_{kl}^{\mathsf{H}}\mathbf{R}'^{\frac{1}{2}}_{kl}$.
- *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, $\mathbf{XTX}^{\mathsf{H}}$ 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 $\sigma_{i,j}^2$.
  - 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 moments/free probability

- asymptotic results
  - most of the above models with Gaussian $\mathbf{X}$.
  - products $\mathbf{V}_1\mathbf{V}_1^H\mathbf{T}_1\mathbf{V}_2\mathbf{V}_2^H\mathbf{T}_2\ldots$ 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 $\mathbf{V}_1\mathbf{V}_1^H\mathbf{T}_1\mathbf{V}_2\mathbf{V}_2^H\mathbf{T}_2\ldots$ 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 $\mathrm{E}[m_{\mathbf{X}_N}]$.

# Open problems, to be explored

- Stieltjes transform methods for more structured matrices: e.g. Vandermonde matrices
- clean framework for band matrix models
- finite dimensional methods for Ricean matrices
- other ?

# Related bibliography

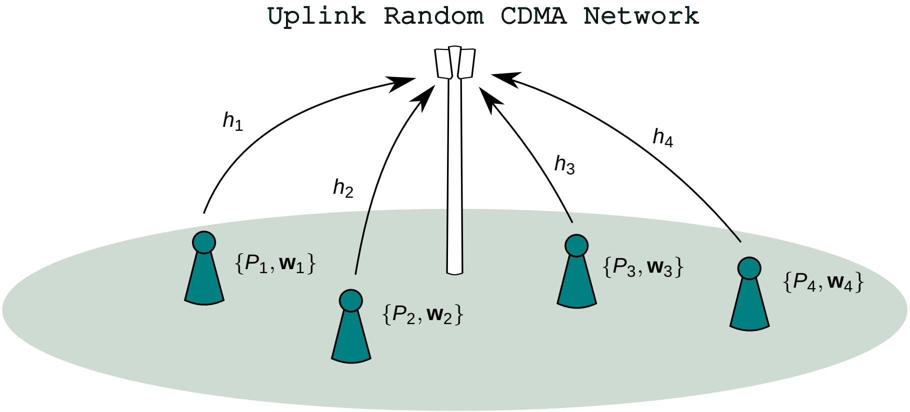

- R. B. Dozier, J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices," Journal of Multivariate Analysis, vol. 98, no. 4, pp. 678-694, 2007.
- J. W. Silverstein, Z. D. Bai, "On the empirical distribution of eigenvalues of a class of 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.*

# Technical Bibliography


- W. Rudin, “Real and complex analysis,” New York, 1966.
- P. Billingsley, “Probability and measure,” Wiley New York, 2008.
- P. Billingsley, “Convergence of probability measures,” Wiley New York, 1968.

# Outline

# Example of use: uplink random CDMA



# Capacity of uplink random CDMA

- System model conditions,
  - uplink random CDMA
  - $K$ mobile users, 1 base station
  - $N$ chips per CDMA spreading code.
  - User $k$, $k \in \{1, \dots, K\}$ has code $\mathbf{w}_k \sim \mathcal{CN}(0, \mathbf{I}_N)$
  - User $k$ transmits the symbol $s_k$.
  - User $k$'s channel is $h_k \sqrt{P_k}$, with $P_k$ the power of user $k$
- The base station receives

$$\mathbf{y} = \sum_{k=1}^{K} h_k \mathbf{w}_k \sqrt{P_k} s_k + \mathbf{n}$$

- This can be written in the more compact form

$$\mathbf{y} = \mathbf{WHP}^{\frac{1}{2}} \mathbf{s} + \mathbf{n}$$

with

- $\mathbf{s} = [s_1, \dots, s_K]^{\mathsf{T}} \in \mathbb{C}^K$,
- $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_K] \in \mathbb{C}^{N \times K}$,
- $\mathbf{P} = \mathrm{diag}(P_1, \dots, P_K) \in \mathbb{C}^{K \times K}$,
- $\mathbf{H} = \mathrm{diag}(h_1, \dots, h_K) \in \mathbb{C}^{K \times K}$.

# Outline

# MMSE decoder

- Consists into taking

$$r_k = \mathbf{w}_k^{\mathrm{H}} \left( \mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{y}$$

  as symbol for user $k$.
- The SINR for user's $k$ signal is

$$\gamma_k^{(\mathrm{MMSE})} = P_k|h_k|^2 \mathbf{w}_k^{\mathrm{H}} \Big( \sum_{\substack{1 \leq i \leq K \\ i \neq k}} P_i|h_i|^2 \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \Big)^{-1} \mathbf{w}_k \quad (1)$$

$$= P_k|h_k|^2 \mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k. \quad (2)$$

- Now we have the following result

**Theorem (Trace Lemma)**

*If $\mathbf{x} \in \mathbb{C}^N$ is i.i.d. with entries of zero mean, variance $1/N$, and $\mathbf{A} \in \mathbb{C}^{N \times N}$ is independent of $\mathbf{x}$, then*

$$\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} = \sum_{i,j} x_i^* x_j A_{ij} \xrightarrow{\mathrm{a.s.}} \frac{1}{N} \operatorname{tr} \mathbf{A}.$$

- Applying this result, for $N$ large,

$$\mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k - \frac{1}{N} \operatorname{tr}(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \xrightarrow{\mathrm{a.s.}} 0.$$

# MMSE decoder

- Consists into taking

$$r_k = \mathbf{w}_k^{\mathrm{H}} \left( \mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{y}$$

as symbol for user $k$.

- The SINR for user's $k$ signal is

$$\gamma_k^{(\mathrm{MMSE})} = P_k |h_k|^2 \mathbf{w}_k^{\mathrm{H}} \Big( \sum_{\substack{1 \le i \le K \\ i \ne k}} P_i |h_i|^2 \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \Big)^{-1} \mathbf{w}_k \quad (1)$$

$$= P_k |h_k|^2 \mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k . \quad (2)$$

- Now we have the following result

**Theorem (Trace Lemma)**

*If $\mathbf{x} \in \mathbb{C}^N$ is i.i.d. with entries of zero mean, variance $1/N$, and $\mathbf{A} \in \mathbb{C}^{N \times N}$ is independent of $\mathbf{x}$, then*

$$\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} = \sum_{i,j} x_i^* x_j A_{ij} \xrightarrow{\text{a.s.}} \frac{1}{N} \operatorname{tr} \mathbf{A}.$$

- Applying this result, for $N$ large,

$$\mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k - \frac{1}{N} \operatorname{tr} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

# MMSE decoder

- Consists into taking

$$r_k = \mathbf{w}_k^{\mathrm{H}} \left( \mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{y}$$

  as symbol for user $k$.
- The SINR for user's $k$ signal is

$$\gamma_k^{(\mathrm{MMSE})} = P_k |h_k|^2 \mathbf{w}_k^{\mathrm{H}} \Big( \sum_{\substack{1 \leq i \leq K \\ i \neq k}} P_i |h_i|^2 \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \Big)^{-1} \mathbf{w}_k \quad (1)$$

$$= P_k |h_k|^2 \mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k . \quad (2)$$

- Now we have the following result

**Theorem (Trace Lemma)**

*If $\mathbf{x} \in \mathbb{C}^N$ is i.i.d. with entries of zero mean, variance $1/N$, and $\mathbf{A} \in \mathbb{C}^{N \times N}$ is independent of $\mathbf{x}$, then*

$$\mathbf{x}^{\mathrm{H}} \mathbf{A} \mathbf{x} = \sum_{i,j} x_i^* x_j A_{ij} \xrightarrow{\text{a.s.}} \frac{1}{N} \operatorname{tr} \mathbf{A}.$$

- Applying this result, for $N$ large,

$$\mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k - \frac{1}{N} \operatorname{tr} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

# MMSE decoder

- Consists into taking

$$r_k = \mathbf{w}_k^{\mathrm{H}} \left( \mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{y}$$

as symbol for user $k$.

- The SINR for user's $k$ signal is

$$\gamma_k^{(\mathrm{MMSE})} = P_k |h_k|^2 \mathbf{w}_k^{\mathrm{H}} \Big( \sum_{\substack{1 \leq i \leq K \\ i \neq k}} P_i |h_i|^2 \mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} + \sigma^2 \mathbf{I}_N \Big)^{-1} \mathbf{w}_k \tag{1}$$

$$= P_k |h_k|^2 \mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k. \tag{2}$$

- Now we have the following result

**Theorem (Trace Lemma)**

*If $\mathbf{x} \in \mathbb{C}^N$ is i.i.d. with entries of zero mean, variance $1/N$, and $\mathbf{A} \in \mathbb{C}^{N \times N}$ is independent of $\mathbf{x}$, then*

$$\mathbf{x}^{\mathrm{H}} \mathbf{A} \mathbf{x} = \sum_{i,j} x_i^* x_j A_{ij} \xrightarrow{\text{a.s.}} \frac{1}{N} \operatorname{tr} \mathbf{A}.$$

- Applying this result, for $N$ large,

$$\mathbf{w}_k^{\mathrm{H}} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{w}_k - \frac{1}{N} \operatorname{tr} (\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k |h_k|^2 \mathbf{w}_k \mathbf{w}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

# MMSE decoder

$$\mathbf{w}_k^{\mathsf{H}}(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N)^{-1}\mathbf{w}_k - \frac{1}{N}\operatorname{tr}(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

- Second important result,

**Theorem (Rank 1 perturbation Lemma)**

*Let* $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{x} \in \mathbb{C}^N$, $t > 0$, *then*

$$\left|\frac{1}{N}\operatorname{tr}(\mathbf{A} + t\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}(\mathbf{A} + \mathbf{x}\mathbf{x}^{\mathsf{H}} + t\mathbf{I}_N)^{-1}\right| \leq \frac{1}{tN}$$

- As $N$ grows large,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1} \to 0,$$

- The RHS is the Stieltjes transform of $\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}$ in $z = -\sigma^2$!

$$m_{\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}}(-\sigma^2)$$

# MMSE decoder

$$\mathbf{w}_k^{\mathsf{H}}(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N)^{-1}\mathbf{w}_k - \frac{1}{N}\operatorname{tr}(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

- Second important result,

**Theorem (Rank 1 perturbation Lemma)**

*Let* $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{x} \in \mathbb{C}^N$, $t > 0$, *then*

$$\left|\frac{1}{N}\operatorname{tr}(\mathbf{A} + t\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}(\mathbf{A} + \mathbf{x}\mathbf{x}^{\mathsf{H}} + t\mathbf{I}_N)^{-1}\right| \leq \frac{1}{tN}$$

- As $N$ grows large,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}} + \sigma^2\mathbf{I}_N\right)^{-1} \to 0,$$

- The RHS is the Stieltjes transform of $\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}$ in $z = -\sigma^2$!

$$m_{\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}}(-\sigma^2)$$

# MMSE decoder

$$\mathbf{w}_k^{\mathrm{H}}(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathrm{H}} + \sigma^2\mathbf{I}_N)^{-1}\mathbf{w}_k - \frac{1}{N}\operatorname{tr}(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathrm{H}} + \sigma^2\mathbf{I}_N)^{-1} \xrightarrow{\text{a.s.}} 0.$$

- Second important result,

**Theorem (Rank 1 perturbation Lemma)**

*Let* $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{x} \in \mathbb{C}^N$, $t > 0$, *then*

$$\left|\frac{1}{N}\operatorname{tr}(\mathbf{A} + t\mathbf{I}_N)^{-1} - \frac{1}{N}\operatorname{tr}(\mathbf{A} + \mathbf{x}\mathbf{x}^{\mathrm{H}} + t\mathbf{I}_N)^{-1}\right| \leq \frac{1}{tN}$$

- As $N$ grows large,

$$\frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} - P_k|h_k|^2\mathbf{w}_k\mathbf{w}_k^{\mathrm{H}} + \sigma^2\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\left(\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}} + \sigma^2\mathbf{I}_N\right)^{-1} \to 0,$$

- The RHS is the Stieltjes transform of $\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}$ in $z = -\sigma^2$!

$$m_{\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}}(-\sigma^2)$$

# MMSE decoder

- From previous result,

$$m_{\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}}(-\sigma^2) - m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

with $m_N(-\sigma^2)$ the unique positive solution of

$$m = \left[\frac{1}{N}\operatorname{tr}\mathbf{HPH}^{\mathrm{H}}\left(m\mathbf{HPH}^{\mathrm{H}} + \mathbf{I}_K\right)^{-1} + \sigma^2\right]^{-1}$$

independent of $k$!

- This is also

$$m = \left[\sigma^2 + \frac{1}{N}\sum_{1\leq i\leq K}\frac{P_i|h_i|^2}{1+mP_i|h_i|^2}\right]^{-1}$$

- Finally,

$$\gamma_k^{(\mathrm{MMSE})} - m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

and the capacity reads

$$C^{(\mathrm{MMSE})}(\sigma^2) - \log_2(1 + m_N(-\sigma^2)) \xrightarrow{\text{a.s.}} 0.$$

# MMSE decoder

- From previous result,

$$m_{\mathbf{WHPH}^{\mathrm{H}}\mathbf{W}^{\mathrm{H}}}(-\sigma^2) - m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

  with $m_N(-\sigma^2)$ the unique positive solution of

$$m = \left[ \frac{1}{N} \operatorname{tr} \mathbf{HPH}^{\mathrm{H}} \left( m\mathbf{HPH}^{\mathrm{H}} + \mathbf{I}_K \right)^{-1} + \sigma^2 \right]^{-1}$$

  independent of $k$!

- This is also

$$m = \left[ \sigma^2 + \frac{1}{N} \sum_{1 \le i \le K} \frac{P_i |h_i|^2}{1 + mP_i |h_i|^2} \right]^{-1}$$

- Finally,

$$\gamma_k^{(\mathrm{MMSE})} - m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

  and the capacity reads

$$C^{(\mathrm{MMSE})}(\sigma^2) - \log_2(1 + m_N(-\sigma^2)) \xrightarrow{\text{a.s.}} 0.$$

# MMSE decoder

- From previous result,

$$m_{\mathbf{WHPH}^{\mathsf{H}}\mathbf{W}^{\mathsf{H}}}(-\sigma^2) - m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

with $m_N(-\sigma^2)$ the unique positive solution of

$$m = \left[ \frac{1}{N} \operatorname{tr} \mathbf{HPH}^{\mathsf{H}} \left( m\mathbf{HPH}^{\mathsf{H}} + \mathbf{I}_K \right)^{-1} + \sigma^2 \right]^{-1}$$

independent of $k$!

- This is also

$$m = \left[ \sigma^2 + \frac{1}{N} \sum_{1 \leq i \leq K} \frac{P_i |h_i|^2}{1 + mP_i |h_i|^2} \right]^{-1}$$

- Finally,

$$\gamma_k^{(\mathrm{MMSE})} - m_N(-\sigma^2) \xrightarrow{\text{a.s.}} 0$$

and the capacity reads

$$C^{(\mathrm{MMSE})}(\sigma^2) - \log_2(1 + m_N(-\sigma^2)) \xrightarrow{\text{a.s.}} 0.$$

# MMSE decoder

$$C^{(\mathrm{MMSE})}(\sigma^2) - \log_2(1 + m_N(-\sigma^2)) \xrightarrow{\text{a.s.}} 0.$$

- AWGN channel, $P_k = P$, $h_k = 1$,

$$C^{(\mathrm{MMSE})}(\sigma^2) \xrightarrow{\text{a.s.}} c \log_2\left(1 + \frac{-(\sigma^2 + (c-1)P) + \sqrt{(\sigma^2 + (c-1)P)^2 + 4P\sigma^2}}{2\sigma^2}\right)$$

- Rayleigh channel, $P_k = P$, $|h_k|$ Rayleigh,

$$m = \left[\sigma^2 + c\int \frac{Pt}{1 + Ptm} e^{-t} dt\right]^{-1}$$

and

$$C_{\mathrm{MMSE}}(\sigma^2) \xrightarrow{\text{a.s.}} c \int \log_2\left(1 + Ptm(-\sigma^2)\right) e^{-t} dt.$$

# Outline

# Matched-Filter, Optimal decoder ...

R. Couillet, M. Debbah, J. W. Silverstein, “A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-User MIMO Channels,” IEEE Trans. on Information Theory, *accepted*, on arXiv.

- Similarly, we can compute deterministic equivalents for the matched-filter performance,

$$C_{\rm MF}(\sigma^2) - \frac{1}{N}\sum_{k=1}^{K}\log_2\left(1 + \frac{P_k|h_k|^2}{\frac{1}{N}\sum_{i=1}^{K}P_i|h_i|^2+\sigma^2}\right) \xrightarrow{\rm a.s.} 0$$

- AWGN case,

$$C_{\rm MF}(\sigma^2) \xrightarrow{\rm a.s.} c\log_2\left(1+\frac{P}{Pc+\sigma^2}\right)$$

- Rayleigh case,

$$C_{\rm MF}(\sigma^2) \xrightarrow{\rm a.s.} -c\log_2(e)e^{\frac{Pc+\sigma^2}{P}}\,{\rm Ei}\left(-\frac{Pc+\sigma^2}{P}\right)$$

- ... and the optimal joint-decoder performance

$$C_{\rm opt}(\sigma^2) - \log_2\left(1+\frac{1}{\sigma^2 N}\sum_{k=1}^{K}\frac{P_k|h_k|^2}{1+cP_k|h_k|^2 m_N(-\sigma^2)}\right) - \frac{1}{N}\sum_{k=1}^{K}\log_2\left(1+cP_k|h_k|^2 m_N(-\sigma^2)\right)$$
$$-\log_2(e)\left(\sigma^2 m_N(-\sigma^2)-1\right) \xrightarrow{\rm a.s.} 0.$$

  with $m_N(-\sigma^2)$ defined as previously.
- Similar expressions are obtained for the AWGN and Rayleigh cases.

# Matched-Filter, Optimal decoder ...

R. Couillet, M. Debbah, J. W. Silverstein, "A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-User MIMO Channels," IEEE Trans. on Information Theory, *accepted*, on arXiv.

- Similarly, we can compute deterministic equivalents for the matched-filter performance,

$$C_{\rm MF}(\sigma^2) - \frac{1}{N}\sum_{k=1}^{K}\log_2\left(1 + \frac{P_k|h_k|^2}{\frac{1}{N}\sum_{i=1}^{K}P_i|h_i|^2 + \sigma^2}\right) \xrightarrow{\rm a.s.} 0$$

- AWGN case,

$$C_{\rm MF}(\sigma^2) \xrightarrow{\rm a.s.} c\log_2\left(1 + \frac{P}{Pc + \sigma^2}\right)$$

- Rayleigh case,

$$C_{\rm MF}(\sigma^2) \xrightarrow{\rm a.s.} -c\log_2(e)e^{\frac{Pc+\sigma^2}{P}}\,{\rm Ei}\left(-\frac{Pc+\sigma^2}{P}\right)$$

- ... and the optimal joint-decoder performance

$$C_{\rm opt}(\sigma^2) - \log_2\left(1 + \frac{1}{\sigma^2 N}\sum_{k=1}^{K}\frac{P_k|h_k|^2}{1 + cP_k|h_k|^2 m_N(-\sigma^2)}\right) - \frac{1}{N}\sum_{k=1}^{K}\log_2\left(1 + cP_k|h_k|^2 m_N(-\sigma^2)\right)$$
$$- \log_2(e)\left(\sigma^2 m_N(-\sigma^2) - 1\right) \xrightarrow{\rm a.s.} 0.$$

  with $m_N(-\sigma^2)$ defined as previously.

- Similar expressions are obtained for the AWGN and Rayleigh cases.

# Simulation results: AWGN case

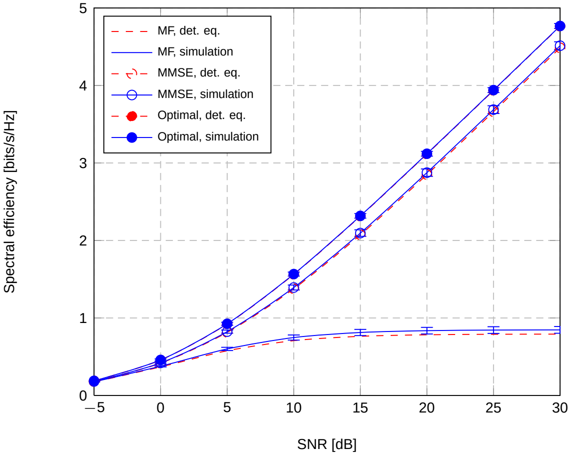



Figure: Spectral efficiency of random CDMA decoders, AWGN channels. Comparison between simulations and deterministic equivalents (det. eq.), for the matched-filter, the MMSE decoder and the optimal decoder, $K = 16$ users, $N = 32$ chips per code. Rayleigh channels. Error bars indicate two standard deviations.

# Simulation results: Rayleigh case

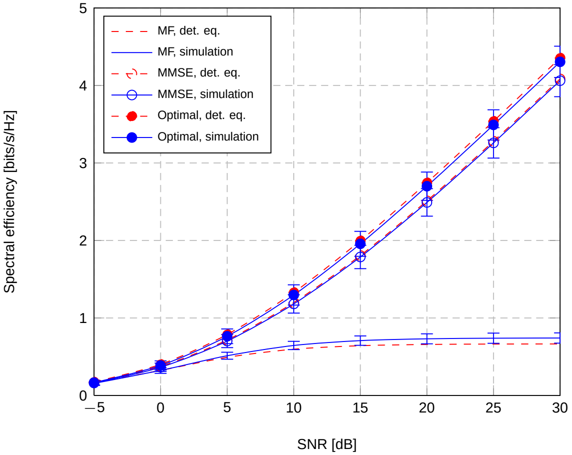



Figure: Spectral efficiency of random CDMA decoders, Rayleigh fading channels. Comparison between simulations and deterministic equivalents (det. eq.), for the matched-filter, the MMSE decoder and the optimal decoder, $K = 16$ users, $N = 32$ chips per code. Rayleigh channels. Error bars indicate two standard deviations.

# Simulation results: Performance as a function of $K/N$

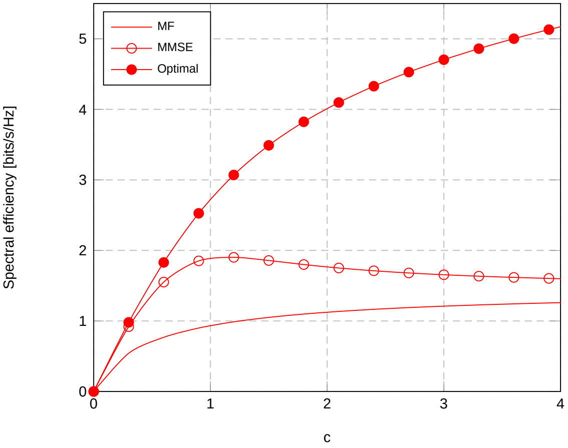
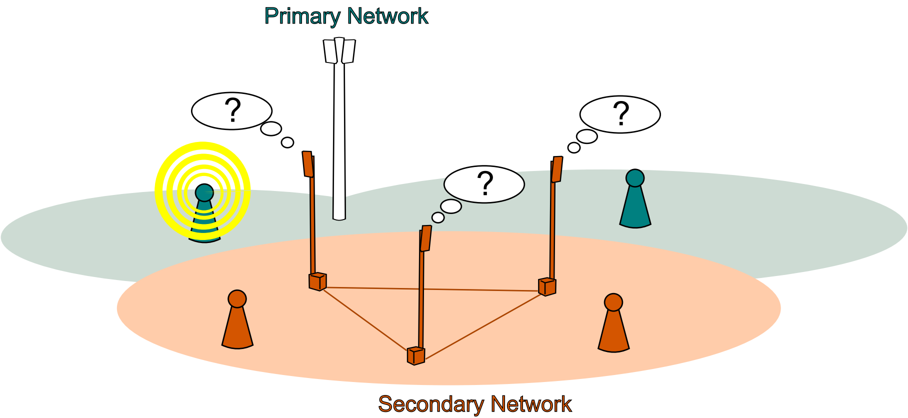



Figure: Spectral efficiency of random CDMA decoders, for different asymptotic ratios $c = K/N$, SNR=10 dB, AWGN channels. Deterministic equivalents for the matched-filter, the MMSE decoder and the optimal decoder. Rayleigh channels.

# Outline

# Signal Sensing in Cognitive Radios



# Position of the Problem

Decide on presence of *informative signal* or *pure noise*.

## Limited *a priori* Knowledge

- Known parameters: the prior information $I$
  - $N$ sensors
  - $L$ sampling periods
  - unit transmit power
  - unit channel variance
- Possibly unknown parameters
  - $M$ signal sources
  - noise power equals $\sigma^2$

## One situation, one solution

For a given prior information $I$, there must be a unique solution to the detection problem.

# Position of the Problem

Decide on presence of *informative signal* or *pure noise*.

## Limited *a priori* Knowledge

- Known parameters: the prior information $I$
  - $N$ sensors
  - $L$ sampling periods
  - unit transmit power
  - unit channel variance
- Possibly unknown parameters
  - $M$ signal sources
  - noise power equals $\sigma^2$

## One situation, one solution

For a given prior information $I$, there must be a unique solution to the detection problem.

# Position of the Problem

Decide on presence of *informative signal* or *pure noise*.

## Limited *a priori* Knowledge

- Known parameters: the prior information $I$
  - $N$ sensors
  - $L$ sampling periods
  - unit transmit power
  - unit channel variance
- Possibly unknown parameters
  - $M$ signal sources
  - noise power equals $\sigma^2$

## One situation, one solution

For a given prior information $I$, there must be a unique solution to the detection problem.

# Problem statement

Signal detection is a typical hypothesis testing problem.

- $\mathcal{H}_0$: only background noise.

$$\mathbf{Y} = \sigma\mathbf{\Theta} = \sigma \begin{pmatrix} \theta_{11} & \cdots & \theta_{1L} \\ \vdots & \ddots & \vdots \\ \theta_{N1} & \cdots & \theta_{NL} \end{pmatrix}$$

- $\mathcal{H}_1$: informative signal plus noise.

$$\mathbf{Y} = \begin{pmatrix} h_{11} & \dots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \dots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix} \begin{pmatrix} s_1^{(1)} & \cdots & \cdots & s_1^{(L)} \\ \vdots & \vdots & \vdots & \vdots \\ s_M^{(1)} & \cdots & \cdots & s_M^{(L)} \\ \theta_{11} & \cdots & \cdots & \theta_{1L} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{N1} & \cdots & \cdots & \theta_{NL} \end{pmatrix}$$

# Problem statement

Signal detection is a typical hypothesis testing problem.

- $\mathcal{H}_0$: only background noise.

$$\mathbf{Y} = \sigma\mathbf{\Theta} = \sigma \begin{pmatrix} \theta_{11} & \cdots & \theta_{1L} \\ \vdots & \ddots & \vdots \\ \theta_{N1} & \cdots & \theta_{NL} \end{pmatrix}$$

- $\mathcal{H}_1$: informative signal plus noise.

$$\mathbf{Y} = \begin{pmatrix} h_{11} & \ldots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \ldots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix} \begin{pmatrix} s_1^{(1)} & \cdots & \cdots & s_1^{(L)} \\ \vdots & \vdots & \vdots & \vdots \\ s_M^{(1)} & \cdots & \cdots & s_M^{(L)} \\ \theta_{11} & \cdots & \cdots & \theta_{1L} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{N1} & \cdots & \cdots & \theta_{NL} \end{pmatrix}$$

# Problem statement

Signal detection is a typical hypothesis testing problem.

- $\mathcal{H}_0$: only background noise.

$$\mathbf{Y} = \sigma\boldsymbol{\Theta} = \sigma\begin{pmatrix} \theta_{11} & \cdots & \theta_{1L} \\ \vdots & \ddots & \vdots \\ \theta_{N1} & \cdots & \theta_{NL} \end{pmatrix}$$

- $\mathcal{H}_1$: informative signal plus noise.

$$\mathbf{Y} = \begin{pmatrix} h_{11} & \ldots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \ldots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix} \begin{pmatrix} s_1^{(1)} & \cdots & \cdots & s_1^{(L)} \\ \vdots & \vdots & \vdots & \vdots \\ s_M^{(1)} & \cdots & \cdots & s_M^{(L)} \\ \theta_{11} & \cdots & \cdots & \theta_{1L} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{N1} & \cdots & \cdots & \theta_{NL} \end{pmatrix}$$

# Outline

# Solution

Solution of hypothesis testing is the function

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

If the receiver does not know if $\mathcal{H}_1$ is more likely than $\mathcal{H}_0$,

$$P_{\mathcal{H}_1} = P_{\mathcal{H}_0} = \frac{1}{2}$$

Therefore,

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

# Solution

Solution of hypothesis testing is the function

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

If the receiver does not know if $\mathcal{H}_1$ is more likely than $\mathcal{H}_0$,

$$P_{\mathcal{H}_1} = P_{\mathcal{H}_0} = \frac{1}{2}$$

Therefore,

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

# Solution

Solution of hypothesis testing is the function

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

If the receiver does not know if $\mathcal{H}_1$ is more likely than $\mathcal{H}_0$,

$$P_{\mathcal{H}_1} = P_{\mathcal{H}_0} = \frac{1}{2}$$

Therefore,

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

# Odds for hypothesis $\mathcal{H}_0$

If the SNR is known then the maximum Entropy Principle leads to

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

# Odds for hypothesis $\mathcal{H}_1$

If known $N$, $M$, $\mathrm{SNR}$ only then

$$
\begin{aligned}
P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) &= \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma})P_{\mathbf{\Sigma}}(\mathbf{\Sigma})d\mathbf{\Sigma} \\
&= \int_{\mathcal{U}(N)\times\mathbb{R}^{+N}} P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y},\mathbf{U},L\mathbf{\Lambda})P_{\mathbf{\Lambda}}(\mathbf{\Lambda})d\mathbf{U}d\mathbf{\Lambda}
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{\Sigma} &= L\begin{pmatrix} h_{11} & \dots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \dots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix}\begin{pmatrix} h_{11} & \dots & h_{1M} & \sigma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{N1} & \dots & h_{NM} & 0 & \cdots & \sigma \end{pmatrix}^H \\
&= \mathbf{U}\left(L\mathbf{\Lambda}\right)\mathbf{U}^H
\end{aligned}
$$

# Odds for hypothesis $\mathcal{H}_1$ (2)

Case $M = 1$.
Maximum Entropy distribution for **H** is Gaussian i.i.d channel. *Unordered* eigenvalue distribution for $\mathbf{\Sigma}$,

$$P_{\mathbf{\Lambda}}(\mathbf{\Lambda})d\mathbf{\Lambda} = \mathbf{1}_{(\lambda_1>\sigma^2)}\frac{1}{N}(\lambda_1-\sigma^2)^{N-1}\frac{e^{-(\lambda_1-\sigma^2)}}{(N-1)!}\prod_{i=2}^{N}\delta(\lambda_i-\sigma^2)d\lambda_1\dots d\lambda_N$$

Maximum Entropy distribution for $\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1$ is correlated Gaussian,

$$P_{\mathbf{Y}|\mathbf{\Sigma}I_1}(\mathbf{Y},\mathbf{U},L\mathbf{\Lambda}) = \frac{1}{\pi^{LN}\det(\mathbf{\Lambda})^L}e^{-\operatorname{tr}\left(\mathbf{Y}\mathbf{Y}^H\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^H\right)}$$

# Odds for hypothesis $\mathcal{H}_1$ (2)

Case $M = 1$.
Maximum Entropy distribution for **H** is Gaussian i.i.d channel. *Unordered* eigenvalue distribution for $\boldsymbol{\Sigma}$,

$$P_{\boldsymbol{\Lambda}}(\boldsymbol{\Lambda})d\boldsymbol{\Lambda} = \mathbf{1}_{(\lambda_1>\sigma^2)}\frac{1}{N}(\lambda_1-\sigma^2)^{N-1}\frac{e^{-(\lambda_1-\sigma^2)}}{(N-1)!}\prod_{i=2}^{N}\delta(\lambda_i-\sigma^2)d\lambda_1\ldots d\lambda_N$$

Maximum Entropy distribution for $\mathbf{Y}|\boldsymbol{\Sigma}\mathcal{H}_1$ is correlated Gaussian,

$$P_{\mathbf{Y}|\boldsymbol{\Sigma}I_1}(\mathbf{Y},\mathbf{U},L\boldsymbol{\Lambda}) = \frac{1}{\pi^{LN}\det(\boldsymbol{\Lambda})^L}e^{-\operatorname{tr}\left(\mathbf{Y}\mathbf{Y}^H\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^H\right)}$$

# Neyman-Pearson Test

- $M = 1$,

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

with $(\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{YY}^{\mathrm{H}})$ and

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

- From which we have the Neyman-Pearson test

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

Neyman-Pearson test only depends on the eigenvalues! But in an involved way!

# Neyman-Pearson Test

- $M = 1$,

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

with $(\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathrm{H}})$ and

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

- From which we have the Neyman-Pearson test

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

Neyman-Pearson test only depends on the eigenvalues! But in an involved way!

# Neyman-Pearson Test

- $M = 1$,

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

with $(\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathrm{H}})$ and

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

- From which we have the Neyman-Pearson test

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

Neyman-Pearson test only depends on the eigenvalues! But in an involved way!

# Neyman-Pearson Test

- $M = 1$,

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

with $(\lambda_1, \ldots, \lambda_N) = \mathrm{eig}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}})$ and

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

- From which we have the Neyman-Pearson test

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

Neyman-Pearson test only depends on the eigenvalues! But in an involved way!

# Neyman-Pearson Test against energy detector, SNR known

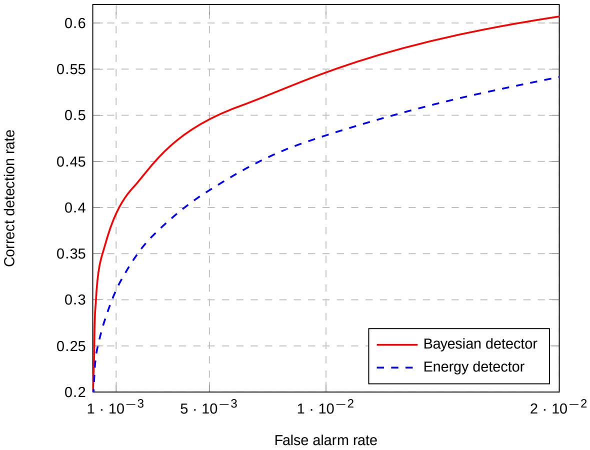



Figure: ROC curve for SIMO transmission, $M = 1$, $N = 4$, $L = 8$, SNR $= -3$ dB, FAR range of practical interest.

# What if $N_t$ is unknown?

Need to integrate out prior for $M$

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

$$= \frac{1}{M_{\max}} \sum_{i=1}^{M_{\max}} P(\mathbf{Y}|\text{``}M = i\text{''}, I_0)$$

# Neyman-Pearson Test, Unknown SNR

- We need to integrate out the prior for $\sigma^2$.
- This leads to

$$C(\mathbf{Y}) = \frac{\int P_{\mathbf{Y}|\sigma^2, I'_M}(\mathbf{Y}, \sigma^2) P_{\sigma^2}(\sigma^2) d\sigma^2}{\int P_{\mathbf{Y}|\sigma^2, \mathcal{H}_0}(\mathbf{Y}, \sigma^2) P_{\sigma^2}(\sigma^2) d\sigma^2}$$

- prior $P_{\sigma^2}(\sigma^2)$ is chosen to be
  - uniform over $[\sigma_-^2, \sigma_+^2]$
  - Jeffrey over $(0, \infty)$

# Neyman-Pearson Test, Unknown SNR

- We need to integrate out the prior for $\sigma^2$.
- This leads to

$$C(\mathbf{Y}) = \frac{\int P_{\mathbf{Y}|\sigma^2, I'_M}(\mathbf{Y}, \sigma^2) P_{\sigma^2}(\sigma^2) d\sigma^2}{\int P_{\mathbf{Y}|\sigma^2, \mathcal{H}_0}(\mathbf{Y}, \sigma^2) P_{\sigma^2}(\sigma^2) d\sigma^2}$$

- prior $P_{\sigma^2}(\sigma^2)$ is chosen to be
  - uniform over $[\sigma_-^2, \sigma_+^2]$
  - Jeffrey over $(0, \infty)$

# Outline

# Reminder: the Marčenko-Pastur Law

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

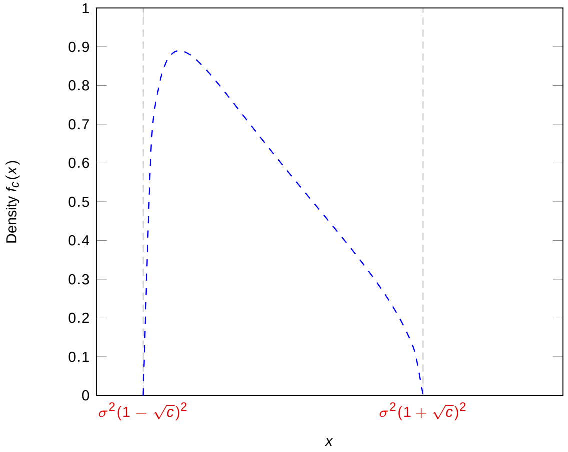



Figure: Marčenko-Pastur law with $c = \lim N/L$.

# Alternative Tests in Large Random Matrix Theory

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

**Theorem**

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

- 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

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

**Theorem**

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

- 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

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

**Theorem**

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

- 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

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

- 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?

# Alternative Tests in Large Random Matrix Theory (2)

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,

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

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

- GLRT test

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

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

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

# Alternative Tests in Large Random Matrix Theory (2)

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,

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

- 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

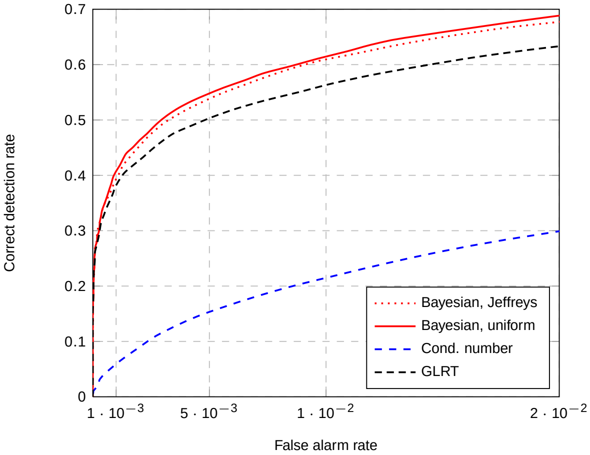
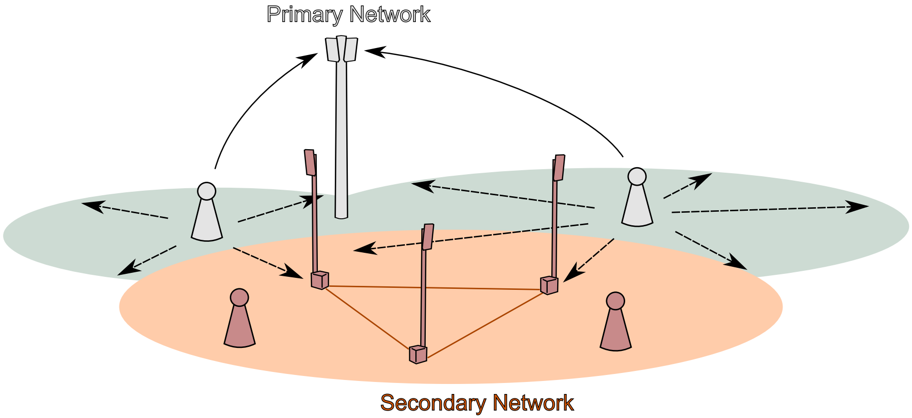



Figure: ROC curve for *a priori* unknown $\sigma^2$ of the Bayesian method, conditioning number method and GLRT method, $M = 1$, $N = 4$, $L = 8$, $\mathrm{SNR} = 0$ dB. For the Bayesian method, both uniform and Jeffreys prior, with exponent $\alpha = 1$, are provided.

# Outline

# Application Context: Coverage range in Femtocells



# Problem statement

- a device embedded with $N$ antennas receives a signal
  - originating from multiple sources
  - number of sources $K$ is not necessarily known
  - source $k$ is equipped with $n_k$ antennas (ideally $n_k >> 1$)
  - signal $k$ goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance $\sigma^2$ of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K$, $n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing $K$
  - $K$, $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $\mathbf{YY}^{\mathsf{H}}$, infer those of $\mathbf{P}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. from analytical formulas on the asymptotic eigenvalue distribution of $\mathbf{YY}^{\mathsf{H}}$, we derive consistent estimates of each $P_k$.

# Problem statement

- a device embedded with $N$ antennas receives a signal
  - originating from multiple sources
  - number of sources $K$ is not necessarily known
  - source $k$ is equipped with $n_k$ antennas (ideally $n_k >> 1$)
  - signal $k$ goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance $\sigma^2$ of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing $K$
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $\mathbf{YY}^{\mathrm{H}}$, infer those of $\mathbf{P}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. from analytical formulas on the asymptotic eigenvalue distribution of $\mathbf{YY}^{\mathrm{H}}$, we derive consistent estimates of each $P_k$.

# Problem statement

- a device embedded with $N$ antennas receives a signal
  - originating from multiple sources
  - number of sources $K$ is not necessarily known
  - source $k$ is equipped with $n_k$ antennas (ideally $n_k >> 1$)
  - signal $k$ goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance $\sigma^2$ of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing $K$
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, infer those of $\mathbf{P}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. from analytical formulas on the asymptotic eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, we derive consistent estimates of each $P_k$.

# Problem statement

- a device embedded with $N$ antennas receives a signal
  - originating from multiple sources
  - number of sources $K$ is not necessarily known
  - source $k$ is equipped with $n_k$ antennas (ideally $n_k >> 1$)
  - signal $k$ goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance $\sigma^2$ of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing $K$
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, infer those of $\mathbf{P}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. from analytical formulas on the asymptotic eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, we derive consistent estimates of each $P_k$.

# Problem statement

- a device embedded with $N$ antennas receives a signal
  - originating from multiple sources
  - number of sources $K$ is not necessarily known
  - source $k$ is equipped with $n_k$ antennas (ideally $n_k >> 1$)
  - signal $k$ goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance $\sigma^2$ of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing $K$
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $\mathbf{YY}^{\mathrm{H}}$, infer those of $\mathbf{P}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. from analytical formulas on the asymptotic eigenvalue distribution of $\mathbf{YY}^{\mathrm{H}}$, we derive consistent estimates of each $P_k$.

# Problem statement

- a device embedded with $N$ antennas receives a signal
  - originating from multiple sources
  - number of sources $K$ is not necessarily known
  - source $k$ is equipped with $n_k$ antennas (ideally $n_k >> 1$)
  - signal $k$ goes through unknown MIMO channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$
  - the variance $\sigma^2$ of the additive noise is not necessarily known
- the problem is to infer
  - $P_1, \ldots, P_K$ knowing $K, n_1, \ldots, n_K$
  - $P_1, \ldots, P_K$ and $n_1, \ldots, n_K$ knowing $K$
  - $K, P_1, \ldots, P_K$ and $n_1, \ldots, n_K$
- we will regard the problem under the angle of
  - free deconvolution: i.e. from the moments of the receive $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, infer those of $\mathbf{P}$, and infer on $\mathbf{P}$
  - Stieltjes transform: i.e. from analytical formulas on the asymptotic eigenvalue distribution of $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, we derive consistent estimates of each $P_k$.

# System model

- at time $t$, source $k$ transmit signal $\mathbf{x}_k^{(t)} \in \mathbb{C}^{n_k}$ with i.i.d. entries of zero mean and variance 1.
- we denote $P_k$ the power emitted by user $k$
- the channel $\mathbf{H}_k \in \mathbb{C}^{N\times n_k}$ from user $k$ to the receiver has i.i.d. entries of zero mean and variance $1/N$.
- at time $t$, the additive noise is denoted $\sigma\mathbf{w}^{(t)}$, with $\mathbf{w}^{(t)} \in \mathbb{C}^N$ with i.i.d. entries of zero mean and variance 1.
- hence the receive signal $\mathbf{y}^{(t)}$ at time $t$,

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

Gathering $M$ time instant into $\mathbf{Y} = [\mathbf{y}^{(1)} \dots \mathbf{y}^{(M)}] \in \mathbb{C}^{N\times M}$, this can be written

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{H}_k \sqrt{P_k}\mathbf{X}_k + \sigma\mathbf{W} = \mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X} + \sigma\mathbf{W}$$

with $\mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_K] \in \mathbb{C}^{N\times n}$, $n = \sum_{k=1}^{K} n_k$,
$\mathbf{P} = \mathrm{diag}(P_1, \dots, P_1, P_2, \dots, P_2, \dots, P_K, \dots, P_K)$ where $P_k$ has multiplicity $n_k$ on the diagonal, $\mathbf{X}^{\mathrm{H}} = [\mathbf{X}_1^{\mathrm{H}} \dots \mathbf{X}_K^{\mathrm{H}}]^{\mathrm{H}} \in \mathbb{C}^{n\times M}$, $\mathbf{X}_k = [\mathbf{x}_k^{(1)} \dots \mathbf{x}_k^{(M)}] \in \mathbb{C}^{n_k\times M}$, $\mathbf{W}$ defined similarly.

# System model

- at time $t$, source $k$ transmit signal $\mathbf{x}_k^{(t)} \in \mathbb{C}^{n_k}$ with i.i.d. entries of zero mean and variance 1.
- we denote $P_k$ the power emitted by user $k$
- the channel $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ from user $k$ to the receiver has i.i.d. entries of zero mean and variance $1/N$.
- at time $t$, the additive noise is denoted $\sigma \mathbf{w}^{(t)}$, with $\mathbf{w}^{(t)} \in \mathbb{C}^N$ with i.i.d. entries of zero mean and variance 1.
- hence the receive signal $\mathbf{y}^{(t)}$ at time $t$,

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

Gathering $M$ time instant into $\mathbf{Y} = [\mathbf{y}^{(1)} \dots \mathbf{y}^{(M)}] \in \mathbb{C}^{N \times M}$, this can be written

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{H}_k \sqrt{P_k} \mathbf{X}_k + \sigma \mathbf{W} = \mathbf{H}\mathbf{P}^{\frac{1}{2}} \mathbf{X} + \sigma \mathbf{W}$$

with $\mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_K] \in \mathbb{C}^{N \times n}$, $n = \sum_{k=1}^{K} n_k$,
$\mathbf{P} = \mathrm{diag}(P_1, \dots, P_1, P_2, \dots, P_2, \dots, P_K, \dots, P_K)$ where $P_k$ has multiplicity $n_k$ on the diagonal, $\mathbf{X}^{\mathsf{H}} = [\mathbf{X}_1^{\mathsf{H}} \dots \mathbf{X}_K^{\mathsf{H}}]^{\mathsf{H}} \in \mathbb{C}^{n \times M}$, $\mathbf{X}_k = [\mathbf{x}_k^{(1)} \dots \mathbf{x}_k^{(M)}] \in \mathbb{C}^{n_k \times M}$, $\mathbf{W}$ defined similarly.

# Outline

# Reminder on free deconvolution

- Free probability provides tools to compute

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

as a function of

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

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

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

with $\mu_c$ the Marčenko-Pastur law with ratio $c$.

# Reminder on free deconvolution

- Free probability provides tools to compute

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

as a function of

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

- 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 $\mathbf{B}_N = \frac{1}{n}(\mathbf{A}_N + \sigma\mathbf{W}_N)(\mathbf{A}_N + \sigma\mathbf{W}_N)^{\mathsf{H}}$,

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

with $\mu_c$ the Marčenko-Pastur law with ratio $c$.

# Reminder on free deconvolution

- Free probability provides tools to compute

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

as a function of

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

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

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

  with $\mu_c$ the Marčenko-Pastur law with ratio $c$.

# 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 $\mathbf{YY}^\mathsf{H}$ 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 operations

In terms of free probability operations, this is

- noise deconvolution

$$\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathrm{H}}} = \left(\left(\mu_{\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}} \boxslash \mu_c\right) \boxminus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

with $\mu_c$ the Marčenko-Pastur law and $c = N/M$.

- signal deconvolution

$$\mu_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}} = \frac{N}{n}\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathrm{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathrm{H}}} + \left(1 - \frac{N}{n}\right)\delta_0$$

- channel deconvolution

$$\mu_{\mathbf{P}} = \mu_{\mathbf{P}\frac{1}{n}\mathbf{H}^{\mathrm{H}}\mathbf{H}} \boxslash \mu_{\eta_{c_1}}$$

with $c_1 = n/N$

# Free deconvolution operations

In terms of free probability operations, this is

- noise deconvolution

$$\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} = \left( \left( \mu_{\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} \boxslash \mu_c \right) \boxminus \delta_{\sigma^2} \right) \boxtimes \mu_c$$

  with $\mu_c$ the Marčenko-Pastur law and $c = N/M$.

- signal deconvolution

$$\mu_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \frac{N}{n}\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} + \left(1 - \frac{N}{n}\right)\delta_0$$

- channel deconvolution

$$\mu_{\mathbf{P}} = \mu_{\mathbf{P}\frac{1}{n}\mathbf{H}^{\mathsf{H}}\mathbf{H}} \boxslash \mu_{\eta_{c_1}}$$

  with $c_1 = n/N$

# Free deconvolution operations

In terms of free probability operations, this is

- noise deconvolution

$$\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} = \left(\left(\mu_{\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}} \boxslash \mu_c\right) \boxminus \delta_{\sigma^2}\right) \boxtimes \mu_c$$

  with $\mu_c$ the Marčenko-Pastur law and $c = N/M$.

- signal deconvolution

$$\mu_{\frac{1}{M}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}} = \frac{N}{n}\mu_{\frac{1}{M}\mathbf{H}\mathbf{P}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{P}^{\frac{1}{2}}\mathbf{H}^{\mathsf{H}}} + \left(1 - \frac{N}{n}\right)\delta_0$$

- channel deconvolution

$$\mu_{\mathbf{P}} = \mu_{\mathbf{P}\frac{1}{n}\mathbf{H}^{\mathsf{H}}\mathbf{H}} \boxslash \mu_{\eta_{c_1}}$$

  with $c_1 = n/N$

# Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of **P** can be computed from those of $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments $m_k$ of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ as a function of the first moments $d_k$ of **P** read

$$
\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right)d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right)d_1^2 \\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right)d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right)d_3 \\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right)d_2d_1 \\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right)d_1^3 \\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right)d_2 \\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right)d_1^2 \\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right)d_1
\end{aligned}
$$

where

$$
m_k = \frac{1}{N}\sum_{i=1}^{N}\lambda\left(\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k \text{ and } d_k = \frac{1}{K}\sum_{i=1}^{K}\lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K}P_i^k
$$

# Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of **P** can be computed from those of $\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments $m_k$ of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ as a function of the first moments $d_k$ of **P** read

$$\begin{aligned}
m_1 &= N^{-1}nd_1 + 1\\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right)d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right)d_1^2\\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right)d_1 + \left(1 + NM^{-1}\right)\\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right)d_3\\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right)d_2d_1\\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right)d_1^3\\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right)d_2\\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right)d_1^2\\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right)d_1
\end{aligned}$$

where

$$m_k = \frac{1}{N}\sum_{i=1}^{N}\lambda\left(\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\right)^k \text{ and } d_k = \frac{1}{K}\sum_{i=1}^{K}\lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K}P_i^k$$

# Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mathbf{P}$ can be computed from those of $\mathbf{YY}^{\mathsf{H}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments $m_k$ of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$ as a function of the first moments $d_k$ of $\mathbf{P}$ read

$$
\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right)d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right)d_1^2 \\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right)d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right)d_3 \\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right)d_2d_1 \\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right)d_1^3 \\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right)d_2 \\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right)d_1^2 \\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right)d_1
\end{aligned}
$$

where

$$
m_k = \frac{1}{N}\sum_{i=1}^{N}\lambda\left(\frac{1}{M}\mathbf{YY}^{\mathsf{H}}\right)^k \text{ and } d_k = \frac{1}{K}\sum_{i=1}^{K}\lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K}P_i^k
$$

# Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mathbf{P}$ can be computed from those of $\mathbf{YY}^{\mathsf{H}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available
- the first moments $m_k$ of $\frac{1}{M}\mathbf{YY}^{\mathsf{H}}$ as a function of the first moments $d_k$ of $\mathbf{P}$ read

$$
\begin{aligned}
m_1 &= N^{-1}nd_1 + 1 \\
m_2 &= \left(N^{-2}M^{-1}n + N^{-1}n\right) d_2 + \left(N^{-2}n^2 + N^{-1}M^{-1}n^2\right) d_1^2 \\
&\quad + \left(2N^{-1}n + 2M^{-1}n\right) d_1 + \left(1 + NM^{-1}\right) \\
m_3 &= \left(3N^{-3}M^{-2}n + N^{-3}n + 6N^{-2}M^{-1}n + N^{-1}M^{-2}n + N^{-1}n\right) d_3 \\
&\quad + \left(6N^{-3}M^{-1}n^2 + 6N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 3N^{-1}M^{-1}n^2\right) d_2 d_1 \\
&\quad + \left(N^{-3}M^{-2}n^3 + N^{-3}n^3 + 3N^{-2}M^{-1}n^3 + N^{-1}M^{-2}n^3\right) d_1^3 \\
&\quad + \left(6N^{-2}M^{-1}n + 6N^{-1}M^{-2}n + 3N^{-1}n + 3M^{-1}n\right) d_2 \\
&\quad + \left(3N^{-2}M^{-2}n^2 + 3N^{-2}n^2 + 9N^{-1}M^{-1}n^2 + 3M^{-2}n^2\right) d_1^2 \\
&\quad + \left(3N^{-1}M^{-2}n + 3N^{-1}n + 9M^{-1}n + 3NM^{-2}n\right) d_1
\end{aligned}
$$

where

$$
m_k = \frac{1}{N}\sum_{i=1}^{N} \lambda\left(\frac{1}{M}\mathbf{YY}^{\mathsf{H}}\right)^k \text{ and } d_k = \frac{1}{K}\sum_{i=1}^{K} \lambda(\mathbf{P})^k = \frac{1}{K}\sum_{i=1}^{K} P_i^k
$$

# Free deconvolution: inferring powers

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,
- direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1, \ldots, P_K$ are given by the $K$ solutions of the polynomial

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

where the $\Pi_m$'s (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k \Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.
  - may lead to non-real solutions!
  - does not minimize any conventional error criterion
  - convenient for one-shot power inference
  - when multiple realizations are available, statistical solutions are preferable

# Free deconvolution: inferring powers

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,
- direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1,\ldots,P_K$ are given by the $K$ solutions of the polynomial

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

where the $\Pi_m$'s (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k\Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.

  - may lead to non-real solutions!
  - does not minimize any conventional error criterion
  - convenient for one-shot power inference
  - when multiple realizations are available, statistical solutions are preferable

# Free deconvolution: inferring powers

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,
- direct inversion with Newton-Girard formulas. Assuming perfect evaluation of $\frac{1}{K}\sum_{k=1}^{K} P_k^m$, $P_1, \ldots, P_K$ are given by the $K$ solutions of the polynomial

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

where the $\Pi_m$'s (known as the *elementary symmetric polynomials*) are iteratively defined as

$$(-1)^k k \Pi_k + \sum_{i=1}^{k} (-1)^{k+i} S_i \Pi_{k-i} = 0$$

where $S_k = \sum_{i=1}^{k} P_i^k$.

  - may lead to non-real solutions!
  - does not minimize any conventional error criterion
  - convenient for one-shot power inference
  - when multiple realizations are available, statistical solutions are preferable

# Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

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.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order $k$ empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

where $X$ is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))$$

with, for some $p$, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^\circ(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}$$

# Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

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.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order $k$ empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where $X$ is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))^{\mathrm{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))$$

  with, for some $p$, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^\circ(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathrm{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathrm{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}$$

# Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

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.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order $k$ empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where $X$ is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^\circ(\mathbf{p}))$$

  with, for some $p$, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^\circ(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^\circ(\mathbf{p}))}d\mathbf{p}}$$

# Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

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.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order $k$ empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where $X$ is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))$$

  with, for some $p$, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^{\circ}(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p} \det(\mathbf{C}^{-1}(\mathbf{p})) e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))} d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p})) e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))} d\mathbf{p}}$$

# Free deconvolution: inferring powers

- alternative approach: estimators that minimize conventional error metrics

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.

- for the model $\mathbf{Y} = \mathbf{T}^{\frac{1}{2}}\mathbf{X}$, an asymptotic central limit result is known for the moments, i.e. for $m_k^{(N)}$ the order $k$ empirical moment of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and $m_k^{\circ(N)}$ its deterministic equivalent, as $N \to \infty$,

$$N\left(m_k^{(N)} - m_k^{\circ(N)}\right) \Rightarrow X$$

  where $X$ is a central Gaussian random variable.
- for the model under consideration, no such result is known.
- if a given model turns out to be Gaussian, then maximum-likelihood or MMSE estimators are of order. Denoting $\mathbf{p} = (P_1, \ldots, P_K)$,

$$\hat{\mathbf{p}}_{\mathrm{ML}} = \arg\min_{\mathbf{p}} \log\det(\mathbf{C}(\mathbf{p})) + (\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m} - \mathbf{m}^{\circ}(\mathbf{p}))$$

  with, for some $p$, $\mathbf{m} = (m_1^{(N)}, \ldots, m_p^{(N)})$, $\mathbf{m}^{\circ}(\mathbf{p}) = (m_1^{\circ(N)}, \ldots, m_p^{\circ(N)})$, and $\mathbf{C}(\mathbf{p})$ the covariance matrix of the Gaussian moment vector assuming powers $\mathbf{p}$.
- and for the MMSE,

$$\hat{\mathbf{p}}_{\mathrm{MMSE}} = \frac{\int \mathbf{p}\det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))}d\mathbf{p}}{\int \det(\mathbf{C}^{-1}(\mathbf{p}))e^{-(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))^{\mathsf{T}}\mathbf{C}(\mathbf{p})^{-1}(\mathbf{m}-\mathbf{m}^{\circ}(\mathbf{p}))}d\mathbf{p}}$$

# Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

# Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

# Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

# Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered
- problem to move from moments to estimates: Newton-Girard method may lead to non real solutions.
- more elaborate methods, e.g. ML, MMSE, are prohibitively expensive

# Outline

# Stieltjes transform approach

- remember the matrix model

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

with $\mathbf{W}, \mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{H} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{n\times M}$, and $\mathbf{P} \in \mathbb{C}^{n\times n}$ diagonal.

- this can be written in the following way

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{N\times M}$$

and extend it into the matrix

$$\mathbf{Y}_{\rm ext} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{(N+n)\times M}$$

which is a sample covariance matrix model.

- the population covariance matrix is

$$\begin{pmatrix} \mathbf{H}\mathbf{P}\mathbf{H}^{\rm H} + \sigma^2\mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$$

itself a sample covariance matrix.

# Stieltjes transform approach

- remember the matrix model

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

with $\mathbf{W}, \mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{H} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{n\times M}$, and $\mathbf{P} \in \mathbb{C}^{n\times n}$ diagonal.

- this can be written in the following way

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{N\times M}$$

and extend it into the matrix

$$\mathbf{Y}_{\rm ext} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{(N+n)\times M}$$

which is a sample covariance matrix model.

- the population covariance matrix is

$$\begin{pmatrix} \mathbf{H}\mathbf{P}\mathbf{H}^{\rm H} + \sigma^2\mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$$

itself a sample covariance matrix.

# Stieltjes transform approach

- remember the matrix model

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

  with $\mathbf{W}, \mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{H} \in \mathbb{C}^{N\times n}$, $\mathbf{X} \in \mathbb{C}^{n\times M}$, and $\mathbf{P} \in \mathbb{C}^{n\times n}$ diagonal.
- this can be written in the following way

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{N\times M}$$

  and extend it into the matrix

$$\mathbf{Y}_{\rm ext} = \begin{bmatrix} \mathbf{H}\mathbf{P}^{\frac{1}{2}} & \sigma\mathbf{I} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \in \mathbb{C}^{(N+n)\times M}$$

  which is a sample covariance matrix model.
- the population covariance matrix is

$$\begin{pmatrix} \mathbf{H}\mathbf{P}\mathbf{H}^{\rm H} + \sigma^2\mathbf{I}_N & 0 \\ 0 & 0 \end{pmatrix}$$

  itself a sample covariance matrix.

# Asymptotic spectrum

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-inference Energy Estimation of Multiple Sources", IEEE Trans. on Information Theory, 2010, *submitted*.

- the asymptotic spectrum of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ has Stietljes transform $m(z)$, $z \in \mathbb{C}^+$, such that

$$m(z) = \frac{M}{N}\underline{m}_N(z) + \frac{M-N}{N}\frac{1}{z}$$

where $\underline{m}_N(z)$ is the unique solution in $\mathbb{C}^+$ of

$$\frac{1}{\underline{m}_N(z)} = -\sigma^2 + \frac{1}{f(z)} - \frac{1}{N}\sum_{k=1}^{K}\frac{n_k P_k}{1+P_k f(z)}$$

where $f(z)$ is given by

$$f(z) = \frac{M-N}{N}\underline{m}_N(z) - \frac{M}{N}z\underline{m}_N(z)^2$$

# Asymptotic spectrum of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$

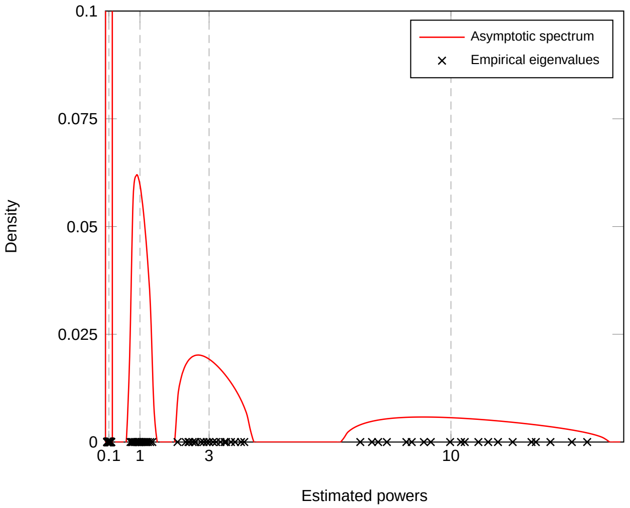



Figure: Empirical and asymptotic eigenvalue distribution of $\frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ when $\mathbf{P}$ has three distinct entries $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1 = n_2 = n_3$, $N/n = 10$, $M/N = 10$, $\sigma^2 = 0.1$. Empirical test: $n = 60$.

# Complex Integration

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.

- Cauchy integration formula

**Theorem**

*Let $f$ be holomorphic on $\mathbb{C}$ and $\gamma \subset \mathbb{C}$ be a continuous contour. Then, for $a$ inside $\gamma$ and $b$ outside $\gamma$,*

$$f(a) = \frac{1}{2\pi i}\oint_{\gamma} \frac{f(\omega)}{\omega - a} d\omega \quad \text{and} \quad 0 = \frac{1}{2\pi i}\oint_{\gamma} \frac{f(\omega)}{\omega - b} d\omega$$

- to estimate $P_k$, notice that, for some contour $\mathcal{C}_k$ enclosing $P_k$,

$$P_k = -\frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{\omega}{P_k - \omega} d\omega = -\frac{n}{n_k}\frac{1}{2\pi i}\oint_{\mathcal{C}_k} \frac{1}{N}\sum_{r=1}^{K} n_r \frac{\omega}{P_r - \omega} d\omega$$

- We recognize a Stieltjes transform of the distribution with masses in $(P_1, \ldots, P_N)$

# Complex Integration

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.

- Cauchy integration formula

**Theorem**

*Let $f$ be holomorphic on $\mathbb{C}$ and $\gamma \subset \mathbb{C}$ be a continuous contour. Then, for $a$ inside $\gamma$ and $b$ outside $\gamma$,*

$$f(a) = \frac{1}{2\pi i} \oint_\gamma \frac{f(\omega)}{\omega - a} d\omega \quad \text{and} \quad 0 = \frac{1}{2\pi i} \oint_\gamma \frac{f(\omega)}{\omega - b} d\omega$$

- to estimate $P_k$, notice that, for some contour $\mathcal{C}_k$ enclosing $P_k$,

$$P_k = -\frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{\omega}{P_k - \omega} d\omega = -\frac{n}{n_k} \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{1}{N} \sum_{r=1}^{K} n_r \frac{\omega}{P_r - \omega} d\omega$$

- We recognize a Stieltjes transform of the distribution with masses in $(P_1, \ldots, P_N)$

# Complex analysis: application

The strategy is the following,

- **variable change.** Write $\frac{1}{N}\sum_{r=1}^{K} n_r \frac{\omega}{P_r - \omega}$ as a function of $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$,

$$\text{remember that } \left(\frac{1}{\underline{m}_N(z)} = -\sigma^2 + \frac{1}{f(z)} - \frac{1}{N}\sum_{r=1}^{K} n_r \frac{P_r}{1 + P_r f(z)}\right)$$

- if clusters are separated, the contour image encircles cluster $k$!
- **approximation.** For large $N$, $m(z) \simeq \hat{m}(z) = \frac{1}{N}\operatorname{tr}(\mathbf{Y}\mathbf{Y}^{\mathrm{H}} - z\mathbf{I}_N)^{-1}$, the accessible data!
- **calculus.** We replace $m(z)$ by $\hat{m}(z)$ and do the (residue) calculus.

# Complex analysis: application

The strategy is the following,

- **variable change.** Write $\frac{1}{N}\sum_{r=1}^{K} n_r \frac{\omega}{P_r-\omega}$ as a function of $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$,

$$\text{remember that } \left(\frac{1}{\underline{m}_N(z)} = -\sigma^2 + \frac{1}{f(z)} - \frac{1}{N}\sum_{r=1}^{K} n_r \frac{P_r}{1+P_r f(z)}\right)$$

- if clusters are separated, the contour image encircles cluster $k$!
- **approximation.** For large $N$, $m(z) \simeq \hat{m}(z) = \frac{1}{N}\operatorname{tr}(\mathbf{Y}\mathbf{Y}^{\mathrm{H}} - z\mathbf{I}_N)^{-1}$, the accessible data!
- **calculus.** We replace $m(z)$ by $\hat{m}(z)$ and do the (residue) calculus.

# Complex analysis: application

The strategy is the following,

- **variable change.** Write $\frac{1}{N}\sum_{r=1}^{K} n_r \frac{\omega}{P_r-\omega}$ as a function of $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$,

$$\text{remember that } \left(\frac{1}{\underline{m}_N(z)} = -\sigma^2 + \frac{1}{f(z)} - \frac{1}{N}\sum_{r=1}^{K} n_r \frac{P_r}{1+P_r f(z)}\right)$$

- if clusters are separated, the contour image encircles cluster $k$!
- **approximation.** For large $N$, $m(z) \simeq \hat{m}(z) = \frac{1}{N}\operatorname{tr}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}$, the accessible data!
- **calculus.** We replace $m(z)$ by $\hat{m}(z)$ and do the (residue) calculus.

# Complex analysis: application

The strategy is the following,

- **variable change.** Write $\frac{1}{N}\sum_{r=1}^{K} n_r \frac{\omega}{P_r-\omega}$ as a function of $m(z)$, the asymptotic Stieltjes transform of $\frac{1}{N}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$,

$$\text{remember that } \left(\frac{1}{\underline{m}_N(z)} = -\sigma^2 + \frac{1}{f(z)} - \frac{1}{N}\sum_{r=1}^{K} n_r \frac{P_r}{1+P_r f(z)}\right)$$

- if clusters are separated, the contour image encircles cluster $k$!
- **approximation.** For large $N$, $m(z) \simeq \hat{m}(z) = \frac{1}{N}\operatorname{tr}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}$, the accessible data!
- **calculus.** We replace $m(z)$ by $\hat{m}(z)$ and do the (residue) calculus.

# Stieltjes transform approach: final result

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-inference Energy Estimation of Multiple Sources", IEEE Trans. on Information Theory, 2010, *submitted*.

## Theorem

*Let* $\mathbf{B}_N = \frac{1}{M}\mathbf{Y}\mathbf{Y}^{\mathsf{H}} \in \mathbb{C}^{N\times N}$*, with* $\mathbf{Y}$ *defined as previously. Denote its ordered eigenvalues vector* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$*,* $\lambda_1 < \ldots, \lambda_N$*. Further assume asymptotic spectrum separability. Then, for* $k \in \{1, \ldots, K\}$*, as N, n, M grow large, we have*

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

*where the estimate* $\hat{P}_k$ *is given by*

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

*with* $\mathcal{N}_k = \{N - \sum_{i=k}^{K} n_i + 1, \ldots, N - \sum_{i=k+1}^{K} n_i\}$ *the set of indexes matching the cluster corresponding to* $P_k$*,* $(\eta_1, \ldots, \eta_N)$ *the ordered eigenvalues of* $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ *and* $(\mu_1, \ldots, \mu_N)$ *the ordered eigenvalues of* $\mathrm{diag}(\boldsymbol{\lambda}) - \frac{1}{M}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$*.*

# Comments on the result

- very compact formula
- low computational complexity
- assuming cluster separation, it allows also to infer the number of eigenvalues, as well as the multiplicity of each eigenvalue.
- however, strong requirement on cluster separation
- if separation is not true, the mean of the eigenvalues instead of the eigenvalues themselves is computed.
- it is possible to infer $K$, all $n_k$ and all $P_k$ using the Stieltjes transform method.

# Comments on the result

- very compact formula
- low computational complexity
- assuming cluster separation, it allows also to infer the number of eigenvalues, as well as the multiplicity of each eigenvalue.
- however, strong requirement on cluster separation
- if separation is not true, the mean of the eigenvalues instead of the eigenvalues themselves is computed.
- it is possible to infer $K$, all $n_k$ and all $P_k$ using the Stieltjes transform method.

# Comments on the result

- very compact formula
- low computational complexity
- assuming cluster separation, it allows also to infer the number of eigenvalues, as well as the multiplicity of each eigenvalue.
- however, strong requirement on cluster separation
- if separation is not true, the mean of the eigenvalues instead of the eigenvalues themselves is computed.
- it is possible to infer $K$, all $n_k$ and all $P_k$ using the Stieltjes transform method.

# Multi-Source Power Estimation: Performance Comparison

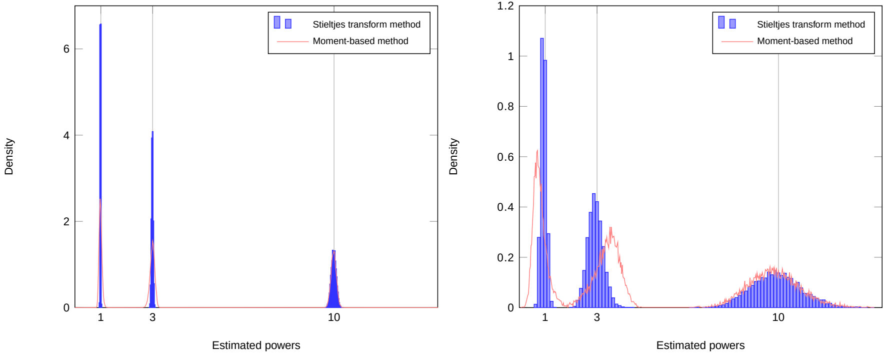



Figure: Multi-source power estimation, for $K = 3$, $P_1 = 1$, $P_2 = 3$, $P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$ ,$n/N = N/M = 1/10$, $\mathrm{SNR} = 10$ dB, for $10,000$ simulation runs; Top $n = 60$, bottom $n = 6$.

# Multi-Source Power Estimation: Performance Comparison

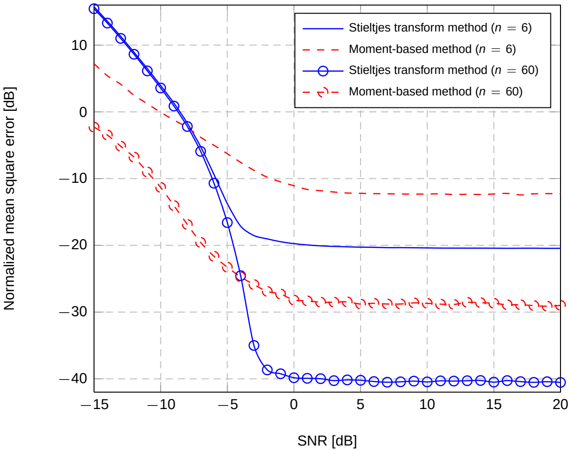



Figure: Normalized mean square error of the vector $(\hat{P}_1, \hat{P}_2, \hat{P}_3)$, $P_1 = 1, P_2 = 3, P_3 = 10$, $n_1/n = n_2/n = n_3/n = 1/3$ ,$n/N = N/M = 1/10$, for $10,000$ simulation runs.

# General comments and steps left to fulfill

- up to this day
  - the moment approach is much simpler to derive
  - it does not require any cluster separation
  - the finite size case is treated in the mean, which the Stieltjes transform approach cannot do.
  - however, the Stieltjes transform approach makes full use of the spectral knowledge, when the moment approach is limited to a few moments.
  - the results are more natural, and more "telling"
- in the future, it is expected that the cluster separation requirement can be overtaken.
- a natural general framework attached to the Stieltjes transform method could arise
- central limit results on the estimates is expected

# General comments and steps left to fulfill

- up to this day
  - the moment approach is much simpler to derive
  - it does not require any cluster separation
  - the finite size case is treated in the mean, which the Stieltjes transform approach cannot do.
  - however, the Stieltjes transform approach makes full use of the spectral knowledge, when the moment approach is limited to a few moments.
  - the results are more natural, and more “telling”
- in the future, it is expected that the cluster separation requirement can be overtaken.
- a natural general framework attached to the Stieltjes transform method could arise
- central limit results on the estimates is expected

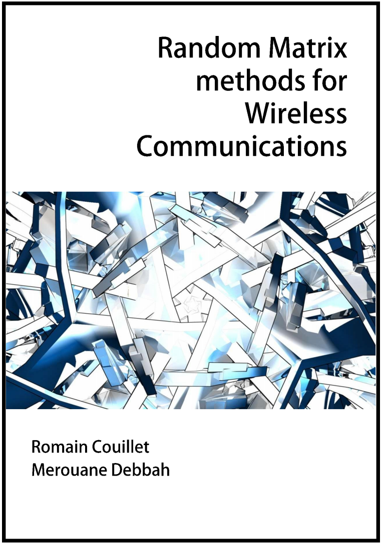
# Related bibliography


- 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.

# Selected authors' recent bibliography


- Articles in Journals,
  - R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," IEEE Trans. on Information Theory, 2010, *submitted*.
  - 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, 2010, *accepted for publication*.
  - P. Bianchi, J. Najim, M. Maida, M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," IEEE Trans. on Information Theory, 2010, *accepted for publication*.
  - R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Sensing," IEEE Trans. on Signal Processing, 2010, *accepted for publication*.
  - S. Wagner, R. Couillet, M. Debbah, D. T. M. 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.
- Articles in International Conferences
  - **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.
  - R. Couillet, M. Debbah, "Uplink capacity of self-organizing clustered orthogonal CDMA networks in flat fading channels", ITW 2009 Fall, Taormina, Sicily.

# Coming up soon...



# Coming up soon...

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

1. Theoretical aspects
    1. Preliminary
    2. Tools for random matrix theory
    3. Deterministic equivalents
    4. Central limit theorems
    5. Spectrum analysis
    6. Eigen-inference
    7. Extreme eigenvalues
2. Applications to wireless communications
    1. Introduction
    2. System performance: capacity and rate-regions
        1. Introduction
        2. Performance of CDMA technologies
        3. Performance of multiple antenna systems
        4. Multi-user communications, rate regions and sum-rate
        5. Design of multi-user receivers
        6. Analysis of multi-cellular networks
        7. Communications in ad-hoc networks
    3. Detection
    4. Estimation
    5. Modelling
    6. Random matrix theory and self-organizing networks
    7. Perspectives
    8. Conclusion