Supélec

Finite Dimensional Statistical Inference

Antonia Masucci, Øyvind Ryan, Sheng Yang, Mérouane Debbah

17 September 2009

Microsoft Research

Table of Contents

Introduction

Deconvolution

Given $\mathbf{A}$, $\mathbf{B}$ $n \times n$ independent hermitian random matrices:

1. Additive free deconvolution:
   To derive the eigenvalue distribution of $\mathbf{A}$ (denoted $\lambda(\mathbf{A})$) knowing $\lambda(\mathbf{A}+\mathbf{B})$ and $\lambda(\mathbf{B})$.
2. Multiplicative free deconvolution:
   To derive the eigenvalue distribution $\lambda(\mathbf{A})$ knowing $\lambda(\mathbf{AB})$ and $\lambda(\mathbf{B})$.

Previous Works

- $n = 1$ (Scalar Case - Classical Probability)
- $n \to \infty$ (Asymptotic Case - Free Probability)

Previous Works

- $n = 1$ (Scalar Case - Classical Probability)
- $n \to \infty$ (Asymptotic Case - Free Probability)

What happens between the Asymptotic Case and the Scalar Case?

Asymptotic Case

-Analytical Method

Stieltjes transform of a probability measure μ

$$s_\mu(z) = \int \frac{1}{\lambda - z} dF^\mu(\lambda)$$

-Moment Method

p-th moment of $\mathbf{A}$

$$t_{\mathbf{A}}^{n,p} = \mathbb{E}\left[\mathrm{tr}(\mathbf{A}^p)\right] = \int \lambda^p d\rho(\lambda)$$

Analytical Method

✓ Based on the Stieltjes transform
✓ Works for any measures
x Can not be implemented

Moment Method

✓ Based on the Moments
x Works only for measures with ALL its moments
✓ Can be implemented

Objective of the presentation

To propose algorithmic methods to compute deconvolution for finite size matrices.

In particular, we focus on Gaussian matrices.

Model 1: Multiplicative model

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

where

- where $\mathbf{X}$ is an $n \times N$ complex standard Gaussian matrix (i.e. with i.i.d., mean zero , variance 1, complex, Gaussian entries),
- $\mathbf{D}$ is an $n \times n$ deterministic matrix.

Multiplicative Deconvolution intends to express the moments of $\mathbf{D}\mathbf{D}^H$ based only on the moments of $\mathbf{Y}\mathbf{Y}^H$ (known as One side correleted zero mean Wishart matrix).

Model 2: Additive Model

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

where

- $\mathbf{X}$ is an $n \times N$ complex standard Gaussian matrix;
- $\mathbf{D}$ is an $n \times N$ deterministic matrix.

Additive Deconvolution intends to express the moments of $\mathbf{D}\mathbf{D}^H$ based only on the moments of $\mathbf{Y}\mathbf{Y}^H$ (known as Non-central Wishart matrix).

Main Results

Theorem 1:

Let n, N be positive integers, $\mathbf{X}$ be complex, standard, Gaussian and $\mathbf{D}$ a (deterministic) $n \times n$ matrix. For any positive integer p, we have

$$\mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{D}\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)^p\right)\right] = \sum_{\pi \in S_p} N^{k(\hat{\pi})-p} n^{l(\hat{\pi})-1} D_{\hat{\pi}|\mathrm{odd}},$$

where $\hat{\pi}|\mathrm{odd}$ is the partition consisting of the equivalence classes/blocks of $\hat{\pi}$ which are contained within the odd numbers; $D_\rho = \prod_{i=1}^{k} \mathrm{tr}\left(\mathbf{D}^{|\rho_i|}\right)$ whenever $\rho = \{\rho_1, \cdots, \rho_k\}$ is a partition with blocks ρ_i; and $|\rho_i|$ is the number of elements in ρ_i.

Denoted $M_p = \mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{D}\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)^p\right)\right]$, $D_p = \mathbb{E}\left[\mathrm{tr}\left(\mathbf{D}^p\right)\right]$ and $c = \frac{n}{N}$

$$\begin{aligned} M_1 &= D_1 \\ M_2 &= D_2 + cD_1^2 \\ M_3 &= \left(1 + \frac{1}{N^2}\right) D_3 + 3cD_2D_1 + c^2D_1^3 \\ &\vdots \end{aligned}$$

By a simple recursion, we can express D_p from M_p. For the first three moments these recursions become

$$\begin{aligned} D_1 &= M_1 \\ D_2 &= M_2 - cM_1^2 \\ D_3 &= \left[M_3 - 3c(M_2 - cM_1^2)M_1 + c^2 M_1^3\right] / \left(1 + \frac{1}{N^2}\right) \\ &\vdots \end{aligned}$$

Theorem 2:

Let $\mathbf{X}$ be an $n \times N$ complex, standard Gaussian matrix, $\mathbf{D}$ a deterministic $n \times N$ matrix, and set $D_p = \mathrm{tr}\left(\left(\frac{1}{N}\mathbf{D}\mathbf{D}^H\right)^p\right)$. Also, for $q \in S_{\rho_2}$, define π by the requirement that $\pi(\rho_1) = q(\rho_2)$. We have that

$$\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{D}+\mathbf{X})(\mathbf{D}+\mathbf{X})^H\right)^p\right)\right] = \sum_{\rho_1 \subset \{1,\ldots,p\}} \frac{1}{nN^{|\rho_1|}} \sum_{\rho_2 \subset \{1,\ldots,p\}} \sum_{q \in S_{\rho_2}} n^{k(\hat{\pi}) - kd(\hat{\pi})} \times N^{l(\hat{\pi}) - ld(\hat{\pi})} n^{|\sigma|} \prod_i D_{|\sigma_i|/2}$$

Defining $M_p = \mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{N}(\mathbf{D}+\mathbf{X})(\mathbf{D}+\mathbf{X})^H\right)^p\right)\right]$, $D_p = \mathbb{E}\left[\mathrm{tr}\left(\frac{1}{N}\left(\mathbf{D}\mathbf{D}^H\right)^p\right)\right]$ and $c = \frac{n}{N}$

$$\begin{aligned} M_1 &= D_1 + 1 \\ M_2 &= D_2 + (2+2c)\,D_1 + (1+c) \\ &\vdots \end{aligned}$$

- $\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{N}}(\mathbf{D}+\mathbf{X})\right)^p\right)\right]$, where $\mathbf{X}$ is selfadjoint standard Gaussian

- $\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{\sqrt{N}}\mathbf{D}\mathbf{X}\right)^p\right)\right]$, where $\mathbf{X}$ is selfadjoint standard Gaussian

- $$\mathbb{E}\left[\mathrm{tr}_n\left(\left(\mathbf{D}\frac{1}{N_1}\mathbf{X}_1\mathbf{X}_1^H\frac{1}{N_2}\mathbf{X}_2\mathbf{X}_2^H\cdots\frac{1}{N_k}\mathbf{X}_k\mathbf{X}_k^H\right)^p\right)\right],$$
  where $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_k$ are independent complex standard Gaussian $n\times N_1, n\times N_2, ..., n\times N_k$

- $$\mathbb{E}\left[\mathrm{tr}_n\left(\left(\mathbf{D}\frac{1}{\sqrt{N}}\mathbf{X}_1\frac{1}{\sqrt{N}}\mathbf{X}_2\cdots\frac{1}{N}\mathbf{X}_k\right)^p\right)\right],$$
  $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_k$ are independent selfadjoint standard Gaussian $N\times N$ matrices

Why do we study the Finite Case?

- Finite dimensional results are not strictly needed for some random matrix models
  $\Rightarrow$ Asymptotic results CAN also be applied

- Finite dimensional results are needed for some random matrix patterns
  $\Rightarrow$ Asymptotic results CAN NOT be applied
  $\Rightarrow$ We are not able to combine a set of observations into a larger compound matrix

How we can stacking the observations?

L observations of the matrix $\mathbf{R}$

- Horizontal

$$\mathbf{R} = [\mathbf{R}_{11}\mathbf{R}_{12}\cdots\mathbf{R}_{1L}]$$

- Vertical

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} \\ \mathbf{R}_{21} \\ \vdots \\ \mathbf{R}_{L1} \end{bmatrix}.$$

- Quadratic

$$\mathbf{R} = \mathbf{D} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{1L} \\ \mathbf{R}_{2L} & \mathbf{R}_{LL} \end{bmatrix}.$$

Example

$$\mathbf{Y} = \mathbf{D}\mathbf{X}_1 + \mathbf{X}_2$$

where

- $\mathbf{X}_1$ is a $m \times L$ matrix
- $\mathbf{X}_2$ is a $n \times N$ matrix
- $\mathbf{D}$ is $n \times m$ matrix

L independent observations $\mathbf{Y}_k = \mathbf{D}\mathbf{X}_{1,k} + \mathbf{X}_{2,k}$ $(1 \le k \le L)$
we can write

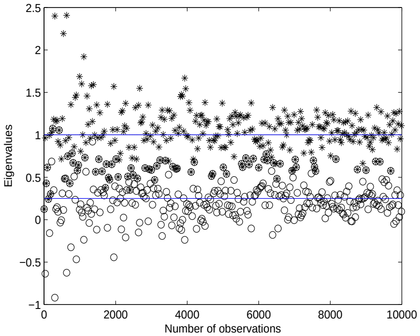
$$\mathbf{Y}_{1 \ldots L} = \mathbf{D}\mathbf{X}_{1,1 \ldots L} + \mathbf{X}_{2,1 \ldots L}$$

where

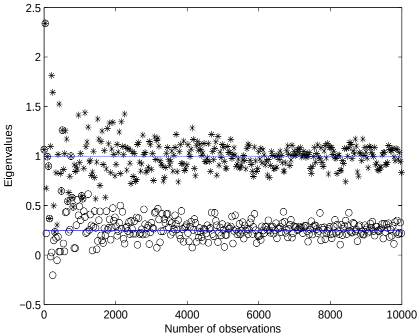
$\mathbf{Y}_{1 \ldots L} = [\mathbf{Y}_1 \mathbf{Y}_2 \cdots \mathbf{Y}_L]$
$\mathbf{X}_{1,1 \ldots L} = [\mathbf{X}_{1,1} \mathbf{X}_{1,2} \cdots \mathbf{X}_{1,L}]$
$\mathbf{X}_{2,1 \ldots L} = [\mathbf{X}_{2,1} \mathbf{X}_{2,2} \cdots \mathbf{X}_{2,L}]$

Taking

- $\mathbf{X}_1, \mathbf{X}_2$ are independent complex standard Gaussian matrices
- $$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

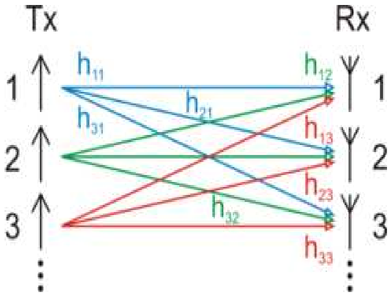


Averaging of moments



Horizontal Stacking

Rate Estimation

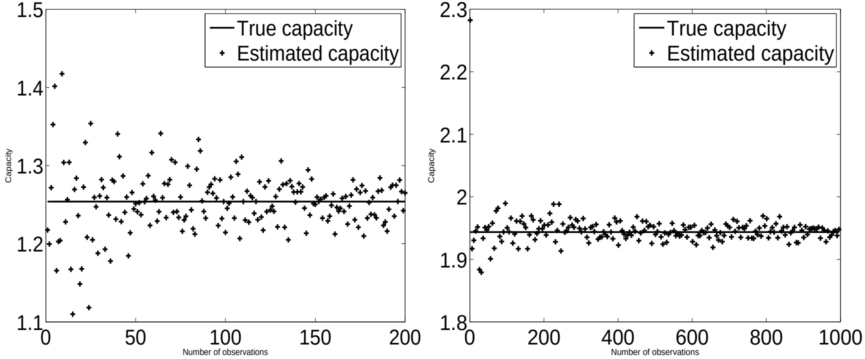


$$\mathbf{Y_i} = \mathbf{D} + \sigma \mathbf{N_i} \qquad (i = 1, \cdots, M)$$

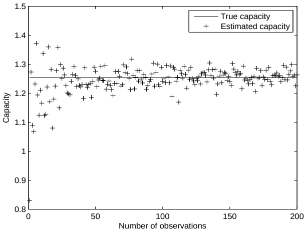
where

- $\mathbf{D}$ is an $n \times N$ deterministic channel matrix;
- $\mathbf{N_i}$ is an $n \times N$ standard Gaussian matrix representing the noise, with variance σ

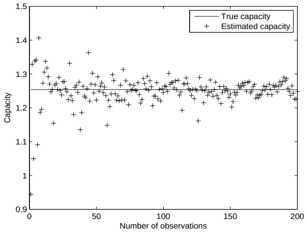
$$C = \frac{1}{n}\log_2 \det\left(\mathbf{I}_n + \frac{\rho}{N}\mathbf{D}\mathbf{D}^{\mathbf{H}}\right) = \frac{1}{n}\log_2\left(\prod_{i=1}^{n}(1+\rho\lambda_i)\right)$$



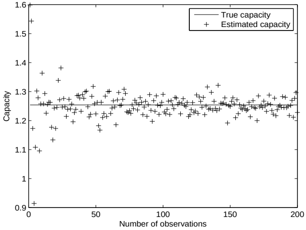
It is seen that less observations are needed for precise channel capacity estimation in the case of the 4×4-matrix.



Quadratic Stacking

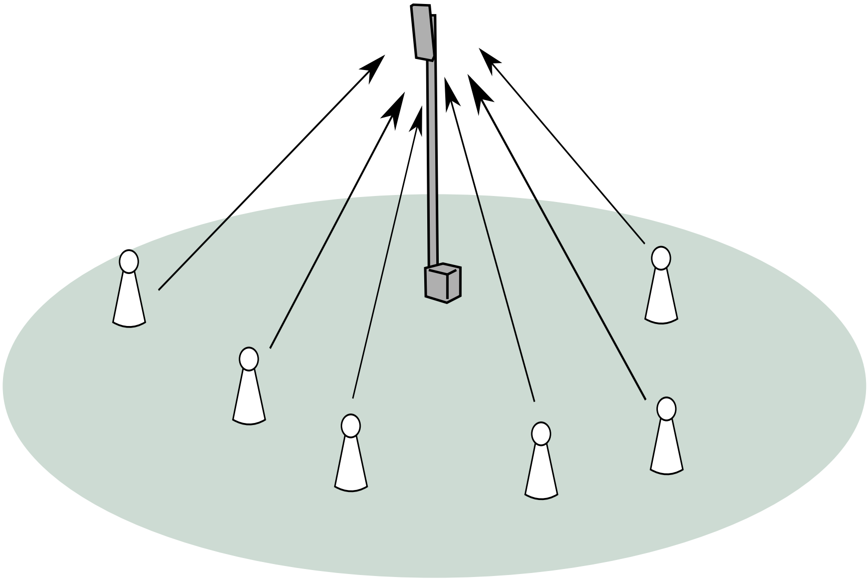


Horizontal Stacking



Averaging of Observations

Power Estimation

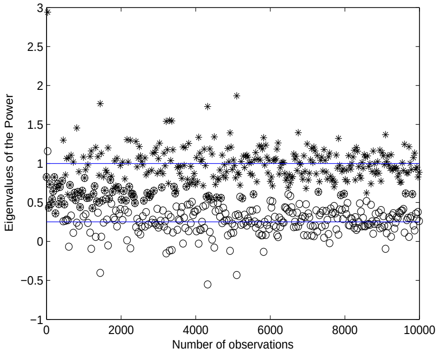


$$\mathbf{Y} = \mathbf{W}\mathbf{P}^{\frac{1}{2}}\mathbf{S} + \sigma\mathbf{N}$$

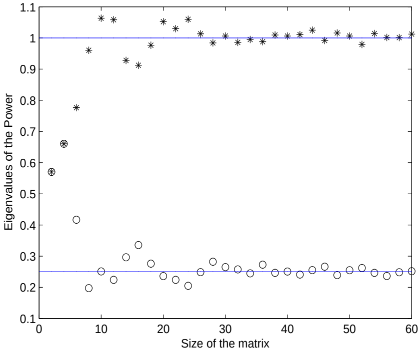
where

- **W** is $N \times K$ channel gain matrix,
- **P** is the $K \times K$ diagonal power matrix due to the different distances from which the users emit,
- **S** isthe $K \times M$ matrix of signals
- **N** is the $N \times M$ matrix representing the noise with variance σ.

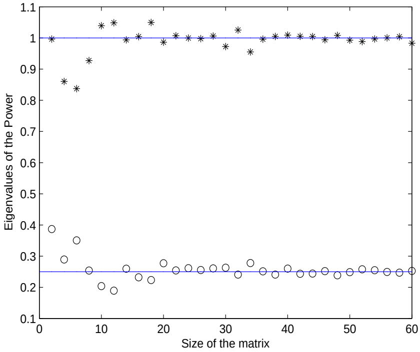
W, **S**, **N** are independent, complex, standard Gaussian matrices.



$K = M = N$ is increased



Observations $L = 15$



Observations $L = 50$

Conclusions

- We proposed a framework to compute the moments of many types of combinations of independent Gaussian and Wishart random matrices;
- We presented a Software Implementation
- We found the optimal way of stacking the observations for stackable models;
- We solved examples of non-stackable models where asymptotic results are not applicable.