Crash Course on Random Matrix Theory
Part II: Advanced notions and applications to signal processing
Afternoon Session: Signal Processing

Romain Couillet and **Mérouane Debbah**

SUPELEC

27-01-2012

Outline

Outline

Outline

Problem formulation

- We consider the model

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

Problem formulation

- We consider the model

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

- We wish to confront the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ given the data matrix $\mathbf{Y} \triangleq [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(M)}] \in \mathbb{C}^{N \times M}$.

Problem formulation

- We consider the model

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

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

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

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

A Bayesian framework for cognitive radios

- We assume prior statistical and deterministic knowledge I on $\mathbf{H}$, σ, P

A Bayesian framework for cognitive radios

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

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

A Bayesian framework for cognitive radios

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

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

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

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

- Using maximum entropy arguments, **X** and **W** are standard Gaussian matrix with $X_{ij}, W_{ij} \sim \mathcal{CN}(0,1)$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

where
- $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathrm{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathrm{H}}$;

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

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

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

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

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

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

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

where
  - $P_{\mathbf{Y}|\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}},\mathcal{H}_1}$ is Gaussian with zero mean and variance $\mathbf{U}\mathbf{G}\mathbf{U}^{\mathsf{H}}$;
  - $P_{\mathbf{U}}$ is a constant ($d\mathbf{U}$ is a Haar measure);

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

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

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

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

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

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

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

where

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

Result in the Gaussian case, $n = 1$

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

Theorem (Neyman-Pearson test)

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

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

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

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

Result in the Gaussian case, $n = 1$

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

Theorem (Neyman-Pearson test)

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

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

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

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

- non trivial dependency on $\lambda_1, \ldots, \lambda_N$
- contrary to energy detector, $\sum_i \lambda_i$ is not a sufficient statistic;

Result in the Gaussian case, $n = 1$

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

Theorem (Neyman-Pearson test)

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

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

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

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

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

Comparison to energy detector

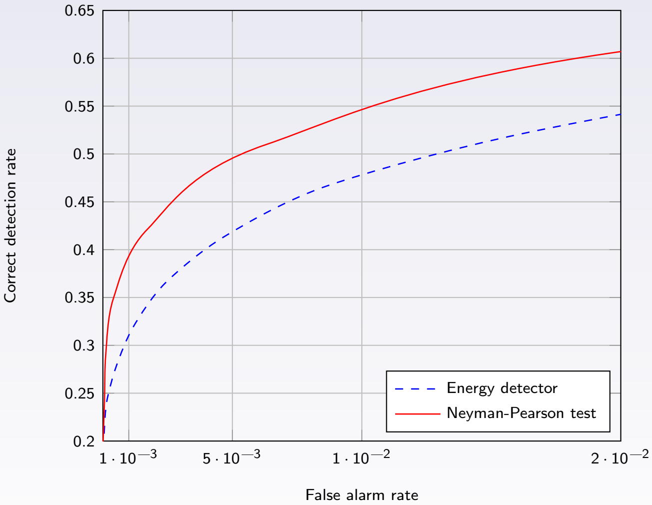


Figure: ROC curve for single-source detection, $K = 1$, $N = 4$, $M = 8$, $\mathrm{SNR} = -3$ dB, FAR range of practical interest, with signal power $E = 0$ dBm, either known or unknown at the receiver.

Comparison to energy detector

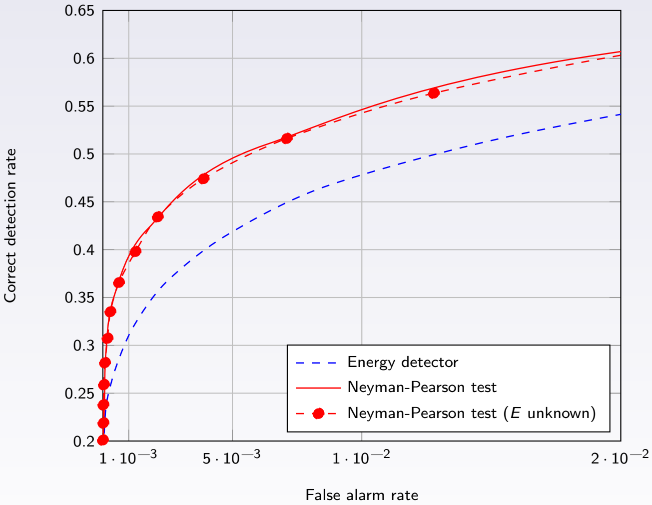


Figure: ROC curve for single-source detection, $K = 1$, $N = 4$, $M = 8$, $\mathrm{SNR} = -3$ dB, FAR range of practical interest, with signal power $E = 0$ dBm, either known or unknown at the receiver.

Unknown power and noise variances

- Bayesian approaches:

$$P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y}) = \int_{\mathbb{R}_+^2} P_{\mathbf{Y}|\mathcal{H}_i,\sigma,P}(\mathbf{Y}) P_{(\sigma,P)}(\sigma,P) d(\sigma,P)$$

  - limited by computational complexity (two-dimension numerical integration);
  - inconsistence in MaxEnt uninformative priors on σ, P.

Unknown power and noise variances

- Bayesian approaches:

$$P_{\mathbf{Y}|\mathcal{H}_i,I}(\mathbf{Y}) = \int_{\mathbb{R}_+^2} P_{\mathbf{Y}|\mathcal{H}_i,\sigma,P}(\mathbf{Y}) P_{(\sigma,P)}(\sigma,P) d(\sigma,P)$$

  - limited by computational complexity (two-dimension numerical integration);
  - inconsistence in MaxEnt uninformative priors on σ, P.
- instead, we will explore nonparametric methods based on large dimensional RMT.

Outline

Reminder of the hypothesis testing problem

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

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

Reminder of the hypothesis testing problem

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

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

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

Exploiting the conditioning number

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

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

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

  - if $\mathcal{H}_1$,

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

  with $t_1 = \sum_{k=1}^{N} |h_k|^2 + \sigma^2$

Exploiting the conditioning number

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

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

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

  - if $\mathcal{H}_1$,

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

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

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

for some security margin ε.

Comments on the method

- Advantages:
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known

Comments on the method

- Advantages:
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known
- Drawbacks:
  - only stands for very large N (dimension N for which asymptotic results arise function of σ!)

Comments on the method

- Advantages:
  - much simpler than finite size analysis
  - ratio independent of σ, so σ needs not be known
- Drawbacks:
  - only stands for very large N (dimension N for which asymptotic results arise function of σ!)
  - *ad-hoc* method, does not rely on performance criterion.

Generalized likelihood ratio test

P. Bianchi, M. Debbah, M. Maida, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, 2011.

- Alternative generalized likelihood ratio test (GLRT) decision criterion, i.e.

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

- Denote

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

To guarantee a maximum false alarm ratio of α,

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

for some threshold ξ_N that can be explicitly given as a function of α.

Generalized likelihood ratio test

P. Bianchi, M. Debbah, M. Maida, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, 2011.

- Alternative generalized likelihood ratio test (GLRT) decision criterion, i.e.

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

- Denote

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

  To guarantee a maximum false alarm ratio of α,

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

  for some threshold ξ_N that can be explicitly given as a function of α.
- Optimal test with respect to GLR.
- Performs better than conditioning number test.

Performance comparison for unknown σ^2, P

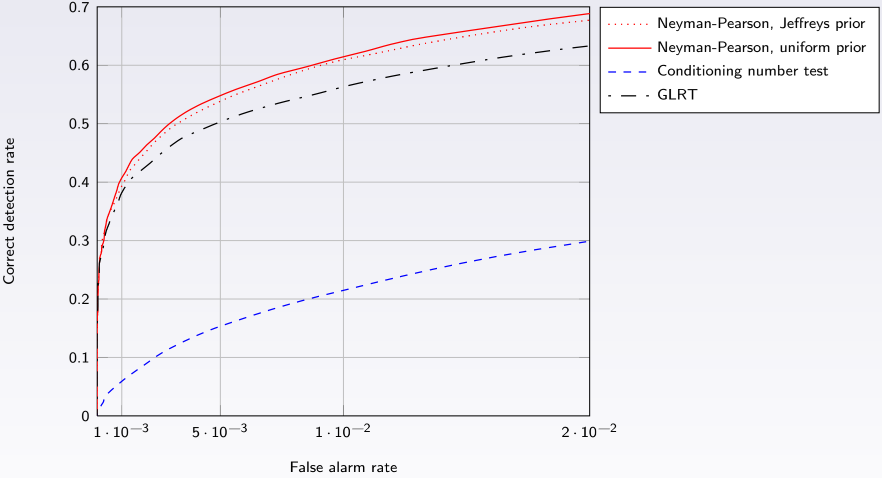


Figure: ROC curve for *a priori* unknown σ^2 of the Neyman-Pearson test, conditioning number method and GLRT, $K = 1$, $N = 4$, $M = 8$, $\text{SNR} = 0$ dB. For the Neyman-Pearson test, both uniform and Jeffreys prior, with exponent $\beta = 1$, are provided.

Related biography

- R. Couillet, M. Debbah, "A Bayesian Framework for Collaborative Multi-Source Signal Sensing", IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5186-5195, 2010.
- T. Ratnarajah, R. Vaillancourt, M. Alvo, "Eigenvalues and condition numbers of complex random matrices," SIAM Journal on Matrix Analysis and Applications, vol. 26, no. 2, pp. 441-456, 2005.
- M. Matthaiou, M. R. McKay, P. J. Smith, J. A. Mossek, "On the condition number distribution of complex Wishart matrices," IEEE Transactions on Communications, vol. 58, no. 6, pp. 1705-1717, 2010.
- C. Zhong, M. R. McKay, T. Ratnarajah, K. Wong, "Distribution of the Demmel condition number of Wishart matrices," IEEE Trans. on Communications, vol. 59, no. 5, pp. 1309-1320, 2011.
- L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative spectrum sensing using random matrix theory," International Symposium on Wireless Pervasive Computing, pp. 334-338 , 2008.
- P. Bianchi, M. Debbah, M. Maida, J. Najim, "Performance of Statistical Tests for Source Detection using Random Matrix Theory," IEEE Trans. on Information Theory, vol. 57, no. 4, pp. 2400-2419, 2011.

Outline

Generic inference scenario

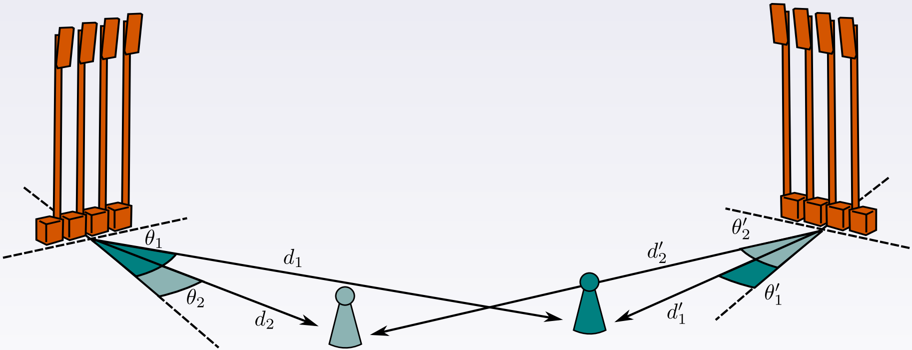


Figure: Signal sensing and angle of arrival detection

Outline

Problem Statement

- Consider the model

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

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

Problem Statement

- Consider the model

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

  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).

Problem Statement

- Consider the model

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

  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

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

Problem Statement

- Consider the model

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

  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

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

Problem Statement

- Consider the model
$$\mathbf{y}^{(m)} = \sum_{k=1}^{K} \sqrt{P_k}\mathbf{H}_k\mathbf{x}_k^{(m)} + \sigma\mathbf{w}^{(m)}$$
and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

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

Problem Statement

- Consider the model

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

  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

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

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

Problem Statement

- Consider the model

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

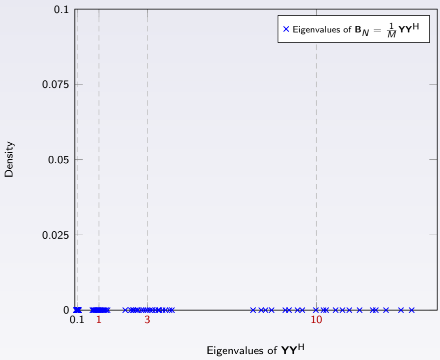
  and wish to infer $P_1, \ldots, P_K$.
- This gives information on transmit power / source distance.
- Applications in localization (radar, sensor network).
- With $\mathbf{Y} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, this can be rewritten

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

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

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

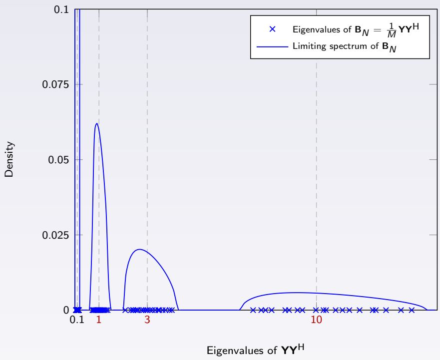
From small to large system analysis



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

- assuming Gaussian parameters, this is similar to previous calculus
- leads to a sum of two-dimensional integrals
- prohibitively expensive to evaluate even for small N, n_k, M

From small to large system analysis



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

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

Free deconvolution approach

- one can infer the moment of $F^{\mathbf{P}}$ from those of $F^{\mathbf{YY}^{\mathsf{H}}}$.

Free deconvolution approach

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

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

(the "deterministic" matrix can be taken random as long as it has a l.s.d.)

Free deconvolution approach

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

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

    (the "deterministic" matrix can be taken random as long as it has a l.s.d.)
  - 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 $\mathbf{X}$,

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

Free deconvolution approach

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

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

    (the "deterministic" matrix can be taken random as long as it has a l.s.d.)
  - 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 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 μ_c the Marčenko-Pastur law and $c = N/M$.

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 μ_c the Marčenko-Pastur law and $c = N/M$.

- signal deconvolution

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

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 μ_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 $\mu_{\mathbf{P}}$ can be computed from those of $\mu_{\mathbf{YY}^{\mathrm{H}}}$.

Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mu_{\mathbf{P}}$ can be computed from those of $\mu_{\mathbf{YY}^{\mathrm{H}}}$.
- this process can be automatized by combinatorics softwares

Free deconvolution: moments

- from the three previous steps (plus addition of null eigenvalues), the moments of $\mu_{\mathbf{P}}$ can be computed from those of $\mu_{\mathbf{YY}^{\mathsf{H}}}$.
- this process can be automatized by combinatorics softwares
- finite size formulas are also available

Free deconvolution: moments

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

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

Free deconvolution: inferring powers

- For practical finite size applications, the deconvolved moments will exhibit errors. Different strategies are available,

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 Π_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 $S_k = \sum_{i=1}^{k} P_i^k$.

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 Π_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

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.

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.

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 $\mathbf{p} = (P_1, \ldots, P_K)$,

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

  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 $\mathbf{p}$.

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

Remarks on free deconvolution approach

- convenient approach, computationally not expensive
- necessarily suboptimal when finitely many moments are considered

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.

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

Stieltjes transform method

- **Reminder:** Method consists in:

Stieltjes transform method

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.

Stieltjes transform method

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.

Stieltjes transform method

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.
  - **Step 3:** Using $m_{\mathbf{B}_N}$ as an approximation of m_F, residue calculus provides estimator.

Stieltjes transform method

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.
  - **Step 3:** Using $m_{\mathbf{B}_N}$ as an approximation of m_F, residue calculus provides estimator.
- Extending $\mathbf{Y}$ with zeros, our model is a "double sample covariance matrix"

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

Stieltjes transform method

- **Reminder:** Method consists in:
  - **Step 1:** link between Stieltjes transform $m_{\mathbf{P}}$ of $\mathbf{P}$ and *limiting* Stieltjes transform m_F of $\mathbf{B}_N$.
  - **Step 2:** Cauchy integral of the parameter to estimate.
  - **Step 3:** Using $m_{\mathbf{B}_N}$ as an approximation of m_F, residue calculus provides estimator.
- Extending $\mathbf{Y}$ with zeros, our model is a "double sample covariance matrix"

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

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

Theorem (Spectral analysis of $\mathbf{B}_N$)

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

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

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

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

Stieltjes transform method (2)

R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," *to appear in* IEEE Trans. on Inf. Theory, 2010.

- estimator calculus

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

Let $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ be defined as in Theorem 2, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, $\lambda_1 < \ldots < \lambda_N$. Assume that asymptotic cluster separability condition is fulfilled for some k. Then, as N, n, $M \to \infty$,

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

where

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

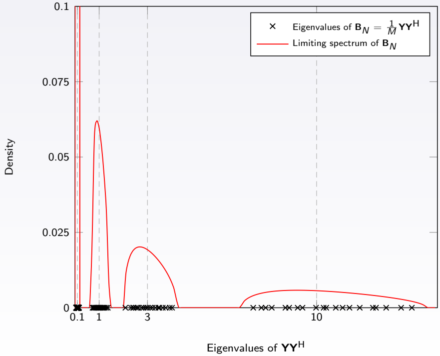
with $\mathcal{N}_k$ the set indexing the eigenvalues in cluster k of F, $\eta_1 < \ldots < \eta_N$ the eigenvalues of $\text{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^{\mathsf{T}}$ and $\mu_1 < \ldots < \mu_N$ the eigenvalues of $\text{diag}(\boldsymbol{\lambda})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

Remarks

- solution is computationally simple, explicit, and the final formula compact.

Remarks

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



Simulations

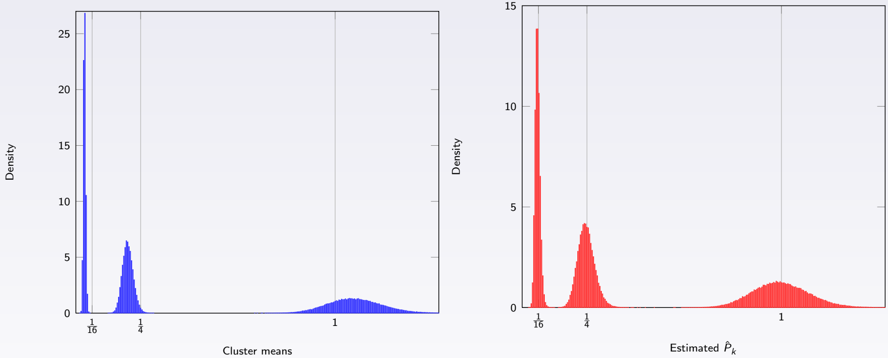


Figure: Histogram of the cluster-mean approach and of $\hat{P}_k$ for $k \in \{1, 2, 3\}$, $P_1 = 1/16$, $P_2 = 1/4$, $P_3 = 1$, $n_1 = n_2 = n_3 = 4$ antennas per user, $N = 24$ sensors, $M = 128$ samples and $\mathrm{SNR} = 20$ dB.

Performance comparison

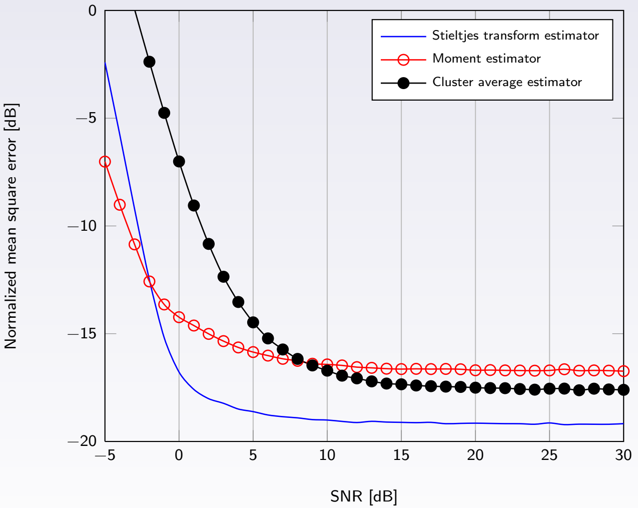


Figure: Normalized mean square error of largest estimated power $\hat{P}_3$, $P_1 = 1/16, P_2 = 1/4, P_3 = 1$, $n_1 = n_2 = n_3 = 4$,$N = 24$, $M = 128$. Comparison between classical, moment and Stieltjes transform approaches.

Outline

Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.

Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.
- Received signal $\mathbf{y}^{(t)} \in \mathbb{C}^N$ at time t

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{s}(\theta_k) x_k^{(t)} + \sigma \mathbf{w}^{(t)}$$

with $E[s_k] = 0$, $E[|x_k|^2] = P_k$.

Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.
- Received signal $\mathbf{y}^{(t)} \in \mathbb{C}^N$ at time t

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{s}(\theta_k) x_k^{(t)} + \sigma \mathbf{w}^{(t)}$$

with $E[s_k] = 0$, $E[|x_k|^2] = P_k$.

- Therefore

$$E[\mathbf{y}^{(t)} \mathbf{y}^{(y)\mathsf{H}}] \triangleq \mathbf{R} = \mathbf{S}(\Theta)\mathbf{P}\mathbf{S}(\Theta)^{\mathsf{H}} + \sigma^2 \mathbf{I}_N$$

where $\mathbf{S}(\Theta) = [\mathbf{s}(\theta_1), \ldots, \mathbf{s}(\theta_K)] \in \mathbb{C}^{N \times K}$, $\mathbf{P} = \operatorname{diag}(P_1, \ldots, P_K)$.

Position of the problem

- We consider the sensor network scenario with:
  - K signal sources
  - an array of N receive antennas, $N > K$
  - line-of-sight signal sensing from angles $\theta_1, \ldots, \theta_K$.
- Received signal $\mathbf{y}^{(t)} \in \mathbb{C}^N$ at time t

$$\mathbf{y}^{(t)} = \sum_{k=1}^{K} \mathbf{s}(\theta_k) x_k^{(t)} + \sigma \mathbf{w}^{(t)}$$

with $E[s_k] = 0$, $E[|x_k|^2] = P_k$.
- Therefore

$$E[\mathbf{y}^{(t)} \mathbf{y}^{(y)\mathrm{H}}] \triangleq \mathbf{R} = \mathbf{S}(\Theta)\mathbf{P}\mathbf{S}(\Theta)^{\mathrm{H}} + \sigma^2 \mathbf{I}_N$$

where $\mathbf{S}(\Theta) = [\mathbf{s}(\theta_1), \ldots, \mathbf{s}(\theta_K)] \in \mathbb{C}^{N \times K}$, $\mathbf{P} = \mathrm{diag}(P_1, \ldots, P_K)$.
- **Objective:** Based on $\mathbf{Y} \triangleq [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(M)}]$, estimate $\theta_1, \ldots, \theta_K$,

MUSIC method

- Write

$$\mathbf{R} = \begin{pmatrix} \mathbf{E}_W & \mathbf{E}_S \end{pmatrix} \begin{pmatrix} \sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S \end{pmatrix} \begin{pmatrix} \mathbf{E}_W^{\mathsf{H}} \\ \mathbf{E}_S^{\mathsf{H}} \end{pmatrix}$$

with $\mathbf{L}_S = \mathrm{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.

MUSIC method

- Write

$$\mathbf{R} = \begin{pmatrix} \mathbf{E}_W & \mathbf{E}_S \end{pmatrix} \begin{pmatrix} \sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S \end{pmatrix} \begin{pmatrix} \mathbf{E}_W^H \\ \mathbf{E}_S^H \end{pmatrix}$$

  with $\mathbf{L}_S = \text{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.

- By definition,

$$\eta(\theta_k) \triangleq \mathbf{s}(\theta_k)^H \mathbf{E}_W \mathbf{E}_W^H \mathbf{s}(\theta_k) = 0$$

MUSIC method

- Write
$$\mathbf{R} = \begin{pmatrix}\mathbf{E}_W & \mathbf{E}_S\end{pmatrix} \begin{pmatrix}\sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S\end{pmatrix} \begin{pmatrix}\mathbf{E}_W^H \\ \mathbf{E}_S^H\end{pmatrix}$$
with $\mathbf{L}_S = \mathrm{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.
- By definition,
$$\eta(\theta_k) \triangleq \mathbf{s}(\theta_k)^H \mathbf{E}_W \mathbf{E}_W^H \mathbf{s}(\theta_k) = 0$$
- MUSIC algorithm consists in finding θ such that
$$\hat{\eta}(\theta) \triangleq \mathbf{s}(\theta)^H \hat{\mathbf{E}}_W \hat{\mathbf{E}}_W^H \mathbf{s}(\theta).$$
reaches a local minimum, with $\hat{\mathbf{E}}_W = [\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_{N-K}] \in \mathbb{C}^{N \times (N-K)}$ the subspace spanned by the $N - K$ smallest eigenvalues of
$$\mathbf{R}_N = \frac{1}{M} \sum_{t=1}^{M} \mathbf{y}^{(t)} \mathbf{y}^{(t)H}.$$

MUSIC method

- Write

$$\mathbf{R} = \begin{pmatrix}\mathbf{E}_W & \mathbf{E}_S\end{pmatrix} \begin{pmatrix}\sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_S\end{pmatrix} \begin{pmatrix}\mathbf{E}_W^{\mathrm{H}} \\ \mathbf{E}_S^{\mathrm{H}}\end{pmatrix}$$

  with $\mathbf{L}_S = \operatorname{diag}(\lambda_{N-K+1}, \ldots, \lambda_N)$, $\mathbf{E}_S = [\mathbf{e}_{N-K+1}, \ldots, \mathbf{e}_N]$ the *signal subspace* and $\mathbf{E}_W = [\mathbf{e}_1, \ldots, \mathbf{e}_{N-K}]$ the *noise subspace*.

- By definition,

$$\eta(\theta_k) \triangleq \mathbf{s}(\theta_k)^{\mathrm{H}} \mathbf{E}_W \mathbf{E}_W^{\mathrm{H}} \mathbf{s}(\theta_k) = 0$$

- MUSIC algorithm consists in finding θ such that

$$\hat{\eta}(\theta) \triangleq \mathbf{s}(\theta)^{\mathrm{H}} \hat{\mathbf{E}}_W \hat{\mathbf{E}}_W^{\mathrm{H}} \mathbf{s}(\theta).$$

  reaches a local minimum, with $\hat{\mathbf{E}}_W = [\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_{N-K}] \in \mathbb{C}^{N \times (N-K)}$ the subspace spanned by the $N - K$ smallest eigenvalues of

$$\mathbf{R}_N = \frac{1}{M} \sum_{t=1}^{M} \mathbf{y}^{(t)} \mathbf{y}^{(t)\mathrm{H}}.$$

Only M-consistent!

RMT will provide an (N, M)-consistent procedure.

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.
- Starting point: Cauchy integration formula

$$\mathbf{s}(\theta_k)^{\mathsf{H}}\mathbf{E}_W\mathbf{E}_W^{\mathsf{H}}\mathbf{s}(\theta_k) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathbf{s}(\theta_k)\left(\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{s}(\theta_k)dz$$

with $\mathcal{C}$ circling around σ^2 only (only one pole in $z = \sigma^2$).

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.
- Starting point: Cauchy integration formula

$$\mathbf{s}(\theta_k)^{\mathsf{H}}\mathbf{E}_W\mathbf{E}_W^{\mathsf{H}}\mathbf{s}(\theta_k) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathbf{s}(\theta_k)\left(\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{s}(\theta_k)dz$$

with $\mathcal{C}$ circling around σ^2 only (only one pole in $z = \sigma^2$).

- We then use the result:

Lemma

For $\mathbf{a} \in \mathbb{C}^N$ deterministic bounded, independent of $\mathbf{R}_N$,

$$\mathbf{a}^{\mathsf{H}}\left(\mathbf{R}_N - z\mathbf{I}_N\right)^{-1}\mathbf{a} - \mathbf{a}^{\mathsf{H}}\left(\frac{1}{1+ce_N(z)}\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{a} \xrightarrow{\text{a.s.}} 0$$

with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{t}{1+ce} - z} dF^{\mathbf{R}}(t).$$

Result on quadratic forms

- Contrary to power inference, we need here results on quadratic forms.
- Starting point: Cauchy integration formula

$$\mathbf{s}(\theta_k)^{\mathsf{H}}\mathbf{E}_W\mathbf{E}_W^{\mathsf{H}}\mathbf{s}(\theta_k) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathbf{s}(\theta_k)\left(\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{s}(\theta_k)dz$$

with $\mathcal{C}$ circling around σ^2 only (only one pole in $z = \sigma^2$).
- We then use the result:

Lemma

For $\mathbf{a} \in \mathbb{C}^N$ deterministic bounded, independent of $\mathbf{R}_N$,

$$\mathbf{a}^{\mathsf{H}}\left(\mathbf{R}_N - z\mathbf{I}_N\right)^{-1}\mathbf{a} - \mathbf{a}^{\mathsf{H}}\left(\frac{1}{1 + ce_N(z)}\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{a} \xrightarrow{\text{a.s.}} 0$$

with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{t}{1+ce} - z} dF^{\mathbf{R}}(t).$$

- By change of variable, dominated convergence arguments, and residue calculus, we conclude.

G-MUSIC

X. Mestre, M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," IEEE Trans. on Signal Processing, vol. 54, no. 1, pp. 69-82, 2006.

Theorem

Under the above conditions,

$$\eta(\theta) - \bar{\eta}(\theta) \xrightarrow{\text{a.s.}} 0$$

as $N, M \to \infty$ *with* $0 < \lim N/M < \infty$, *where*

$$\bar{\eta}(\theta) = \mathbf{s}(\theta)^{\mathsf{H}} \left(\sum_{n=1}^{N} \phi(n) \hat{\mathbf{e}}_n \hat{\mathbf{e}}_n^{\mathsf{H}} \right) \mathbf{s}(\theta)$$

with $\phi(n)$ *defined as*

$$\phi(n) = \begin{cases} 1 + \sum_{k=N-K+1}^{N} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_n - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_n - \hat{\mu}_k} \right) & , \ n \leqslant N-K \\ -\sum_{k=1}^{N-K} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_n - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_n - \hat{\mu}_k} \right) & , \ n > N-K \end{cases}$$

and with $\mu_1 \leqslant \ldots \leqslant \mu_N$ *the eigenvalues of diag*$(\hat{\boldsymbol{\lambda}}) - \frac{1}{M}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^{\mathsf{T}}$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^{\mathsf{T}}$.

Simulation results

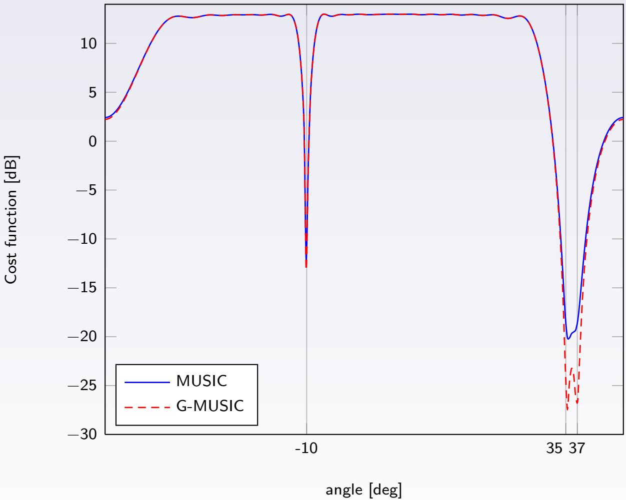


Figure: MUSIC against G-MUSIC for DoA detection of $K = 3$ signal sources, $N = 20$ sensors, $M = 150$ samples, SNR of 10 dB. Angles of arrival of 10°, 35°, and 37°.

Simulation results (2)

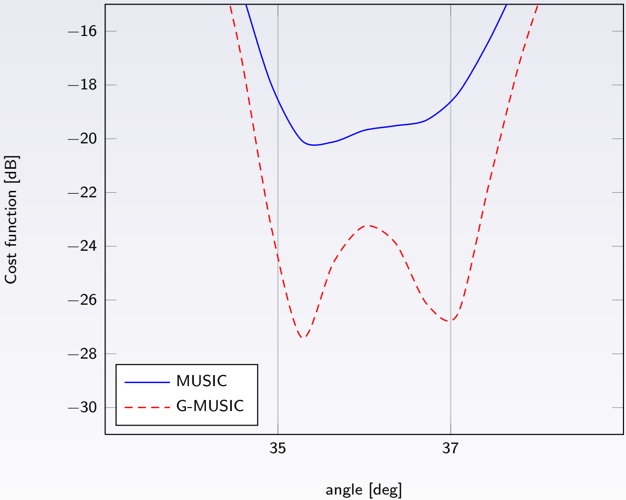


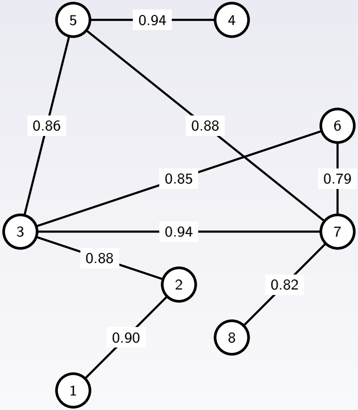
Figure: MUSIC against G-MUSIC for DoA detection of $K = 3$ signal sources, $N = 20$ sensors, $M = 150$ samples, SNR of 10 dB. Angles of arrival of $10°$, $35°$, and $37°$.

Related bibliography

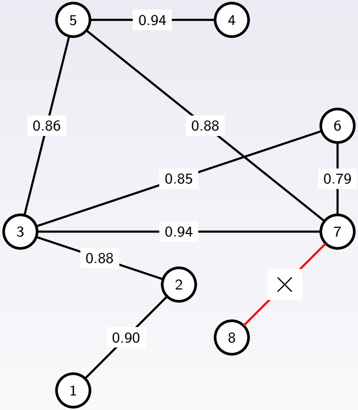
- Ø. Ryan, M. Debbah, "Free deconvolution for signal processing applications," ISIT, pp. 1846-1850, 2007.
- R. Rao, J. A. Mingo, R. Speicher, A. Edelman, "Statistical eigen-inference from large Wishart matrices," Annals of Statistics, vol. 36, no. 6, pp. 2850-2885, 2008.
- A. Masucci, Ø. Ryan, S. Yang, M. Debbah, "Finite dimensional statistical inference," vol. 57, no. 4, pp. 2457-2473, 2011.
- J. W. Silverstein, P. L. Combettes, "Large dimensional random matrix theory for signal detection and estimation in array processing," Workshop on Statistical Signal and Array Processing, pp. 276-279, 1992.
- R. Couillet, J. W. Silverstein, Z. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources," *to appear in* IEEE Trans. on Inf. Theory, 2010.
- J. Yao, R. Couillet, J. Najim, M. Debbah, "Fluctuations of an Improved Population Eigenvalue Estimator in Sample Covariance Matrix Models", (submitted to) IEEE Transactions on Information Theory.
- X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices", IEEE Transactions on Signal Processing, vol. 56, no. 11, 2008.
- X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates", IEEE Transactions on Information Theory, vol. 54, no. 11, Nov. 2008.
- A. Johnson, Y. Abramovich, X. Mestre, "MUSIC, G-MUSIC and Maximum Likelihood Performance Breakdown", IEEE Transactions on Signal Processing, vol. 56, no. 8, pp. 3944-3958, 2008.
- X. Mestre, M. A. Lagunas, "Modified Subspace Algorithms for DoA Estimation in the Small Sample Size Regime" IEEE Transactions on Signal Processing, Vol. 56, pp. 598-614, Feb. 2008.
- P. Vallet, P. Loubaton, X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case," arxiv preprint 1002.3234, 2010.
- W. Hachem, P. Loubaton, X. Mestre, P. Vallet, "A subspace estimator of finite rank perturbations of large random matrices," arXiv preprint, 2011.

Outline

Problem statement



Problem statement



- Localize local failures based on observations from a sensor network.
- Focus on failures modeled as small rank perturbations of large random matrices.

Target

- Systems with failures modeled by small rank perturbations
- Observation matrix $\mathbf{\Sigma} = [\mathbf{s}_1, \ldots, \mathbf{s}_n] \in \mathbb{C}^{N \times n}$ modeled by

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

with $\mathbf{P}_k \in \mathbb{C}^{N \times N}$ of rank $r_k \ll N$, $\mathbf{X}$ with independent $\mathcal{CN}(0, 1/n)$ entries.

Target

- Systems with failures modeled by small rank perturbations
- Observation matrix $\mathbf{\Sigma} = [\mathbf{s}_1, \ldots, \mathbf{s}_n] \in \mathbb{C}^{N \times n}$ modeled by

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

  with $\mathbf{P}_k \in \mathbb{C}^{N \times N}$ of rank $r_k \ll N$, $\mathbf{X}$ with independent $\mathcal{CN}(0, 1/n)$ entries.
- Failure scenarios:
  - $(\mathcal{H}_0)$: no failure, $E[\mathbf{ss}^{\mathsf{H}}] = \mathbf{I}_N$.
  - $(\mathcal{H}_k)$: $1 \leqslant k \leqslant K$, failure of type k, $E[\mathbf{ss}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$.

Target

- Systems with failures modeled by small rank perturbations
- Observation matrix $\mathbf{\Sigma} = [\mathbf{s}_1, \ldots, \mathbf{s}_n] \in \mathbb{C}^{N \times n}$ modeled by

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

  with $\mathbf{P}_k \in \mathbb{C}^{N \times N}$ of rank $r_k \ll N$, $\mathbf{X}$ with independent $\mathcal{CN}(0, 1/n)$ entries.
- Failure scenarios:
  - $(\mathcal{H}_0)$: no failure, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N$.
  - $(\mathcal{H}_k)$: $1 \leqslant k \leqslant K$, failure of type k, $E[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N + \mathbf{P}_k$.
- Subspace approach for:
  - detecting a failure: decide between $\mathcal{H}_0$ and $\bar{\mathcal{H}}_0$
  - diagnosing a failure: upon failure detection, decide on the most probable $\mathcal{H}_k$.

Example 1

Node failure in sensor networks

- Consider the model

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

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

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

Example 1

Node failure in sensor networks

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

Example 2

Sudden parameter change detection in sensor networks

- Upon sudden change of parameter θ_k,

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

- Then

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

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

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

with

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

of rank-1.

Eigenvalue and eigenvectors statistics: Method

- Consider the model
$$\mathbf{\Sigma} = (\mathbf{I}_N + \mathbf{P})^{\frac{1}{2}}\mathbf{X}$$
with, for simplicity
  - $\mathbf{X}$ standard Gaussian
  - $\mathbf{P} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}$, $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{C}^{N\times r}$, $\mathbf{\Omega} = \mathrm{diag}(\omega_1, \ldots, \omega_r)$, $\omega_1 > \ldots > \omega_r > 0$.
- Convergence properties of
  - $\lambda_1 > \ldots > \lambda_r$, the r largest eigenvalues of $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathsf{H}}$
  - $\mathbf{u}_i^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_i$, with $\hat{\mathbf{u}}_i$ the eigenvector associated to λ_i.

Eigenvalue and eigenvectors statistics: Method

- Consider the model

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

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

First order limits: Reminder

- The *limiting* ρ_k are given by:

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

- Consider ω_i and its corresponding eigenvector $\mathbf{u}_i$, then

$$\mathbf{u}_i^{\mathrm{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathrm{H}}\mathbf{u}_i \xrightarrow{\text{a.s.}} \zeta_i \triangleq \frac{1 - c\omega_i^{-2}}{1 + c\omega_i^{-1}}.$$

Fluctuations

Second order behaviour for the joint variable

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

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

Theorem

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

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

where

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

Reminder: Fluctuations at the edge of the bulk

- The previous theorem holds for $\omega_i > \sqrt{c}$, i.e. "strong perturbations"

Reminder: Fluctuations at the edge of the bulk

- The previous theorem holds for $\omega_i > \sqrt{c}$, i.e. "strong perturbations"
- For $\omega_i < \sqrt{c}$, the eigenvalue fluctuations are:

Theorem

If $0 \leqslant \omega_i < \sqrt{c}$,

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

where T_2 *is the complex Tracy-Widom distribution function.*

Failure detection and localization

- The proposed subspace procedure is a two-step approach:
  - **Failure detection procedure**, $\mathcal{H}_0$ vs. $\bar{\mathcal{H}}_0$: We evaluate the statistics of λ_1 against the Tracy-Widom law for a false alarm rate η,

$$\lambda_1' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta)$$

where $\lambda_1' \triangleq N^{\frac{2}{3}}(1+\sqrt{c_N})^{-\frac{4}{3}} c_N^{-\frac{1}{2}}(\lambda_1 - (1+\sqrt{c_N})^2)$.

Failure detection and localization

- The proposed subspace procedure is a two-step approach:
  - **Failure detection procedure**, $\mathcal{H}_0$ vs. $\bar{\mathcal{H}}_0$: We evaluate the statistics of λ_1 against the Tracy-Widom law for a false alarm rate η,

$$\lambda_1' \underset{\bar{\mathcal{H}}_0}{\overset{\mathcal{H}_0}{\lessgtr}} (T_2)^{-1}(1-\eta)$$

    where $\lambda_1' \triangleq N^{\frac{2}{3}}(1+\sqrt{c_N})^{-\frac{4}{3}} c_N^{-\frac{1}{2}}(\lambda_1 - (1+\sqrt{c_N})^2)$.
  - **Failure diagnosis**, selection of $\mathcal{H}_k$: We evaluate the joint statistics of λ_i, $\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_{k,i}$ for each $k \in \{1,\ldots,K\}$, and obtain the maximum-likelihood test,

$$\hat{k} = \arg\max_{1 \leqslant k \leqslant K} \prod_{i=1}^{r} f\left(\left(\left(\sqrt{N}(\lambda_i - \rho_{k,i})\right)_{i=1}^{r}, \left(\sqrt{N}(\mathbf{u}_{k,i}^{\mathsf{H}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{u}_{k,i} - \zeta_{k,i})\right)_{i=1}^{r}\right); C(\rho_{k,i})\right)$$

    with $f(x;\mathbf{R})$ the Gaussian density with zero mean and variance $\mathbf{R}$, and indices k corresponding to hypothesis $\mathcal{H}_k$.

Results

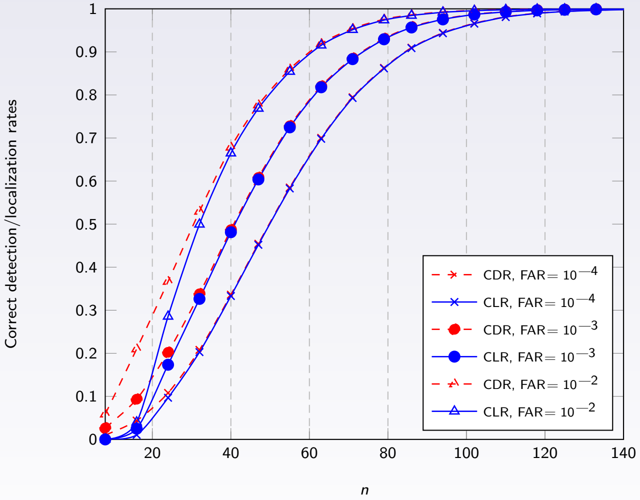


Figure: Simulation of sensor failure in an $N = 10$ node network. Correct detection (CDR) and localization (CLR) rates for different false alarm rates (FAR) and different n.

Outline

Hypothesis testing in large data sets
- Finite dimensional approach
- Large dimensional considerations

Statistical inference: improved subspace estimators
- Sensor networks: distance estimation
  - Free probability approach
  - Stieltjes transform approach
- Sensor networks: angle-of-arrival estimation

Local failure detection schemes

Research today: Robust estimation and RMT

Problem statement

R. A. Maronna, "Robust M-estimators of multivariate location and scatter", The annals of statistics, pp. 51-67, 1976.

- Observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ of a random vector $\mathbf{x}$ with zero mean, variance $\mathbf{C}$.

Problem statement

R. A. Maronna, "Robust M-estimators of multivariate location and scatter", The annals of statistics, pp. 51-67, 1976.

- Observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ of a random vector $\mathbf{x}$ with zero mean, variance $\mathbf{C}$.
- Asymptotic behaviour slow to arise for heavy-tailed distributions.
- Stability issues with outliers.
- Statistical inference methods using sample covariance matrix $\hat{\mathbf{S}}_n$ (SCM) not appropriate,

$$\hat{\mathbf{S}}_N = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathsf{H}}.$$

Problem statement

R. A. Maronna, "Robust M-estimators of multivariate location and scatter", The annals of statistics, pp. 51-67, 1976.

- Observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ of a random vector $\mathbf{x}$ with zero mean, variance $\mathbf{C}$.
- Asymptotic behaviour slow to arise for heavy-tailed distributions.
- Stability issues with outliers.
- Statistical inference methods using sample covariance matrix $\hat{\mathbf{S}}_n$ (SCM) not appropriate,

$$\hat{\mathbf{S}}_N = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}.$$

- Instead, one uses robust M-estimators, such as fixed-point SCM $\hat{\mathbf{C}}_N$, solution to

$$\hat{\mathbf{C}}_N = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}\mathbf{x}_i^{\mathrm{H}}\hat{\mathbf{C}}_N^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}$$

for some well-chosen $u(x)$.

Problem statement

R. A. Maronna, "Robust M-estimators of multivariate location and scatter", The annals of statistics, pp. 51-67, 1976.

- Observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ of a random vector $\mathbf{x}$ with zero mean, variance $\mathbf{C}$.
- Asymptotic behaviour slow to arise for heavy-tailed distributions.
- Stability issues with outliers.
- Statistical inference methods using sample covariance matrix $\hat{\mathbf{S}}_n$ (SCM) not appropriate,

$$\hat{\mathbf{S}}_N = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}.$$

- Instead, one uses robust M-estimators, such as fixed-point SCM $\hat{\mathbf{C}}_N$, solution to

$$\hat{\mathbf{C}}_N = \frac{1}{n}\sum_{i=1}^{n} u\left(\frac{1}{N}\mathbf{x}_i^{\mathrm{H}}\hat{\mathbf{C}}_N^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}$$

  for some well-chosen $u(x)$.
- Typically,
  - **Tyler:** $u(x) = 1/x$ (but $\hat{\mathbf{C}}_N$ non-unique)
  - **Maronna:** $u(x)$ continuous on $[0, \infty)$ nonincreasing, $\phi(x) = xu(x)$ nondecreasing bounded.

Problem statement

R. A. Maronna, "Robust M-estimators of multivariate location and scatter", The annals of statistics, pp. 51-67, 1976.

- Observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ of a random vector $\mathbf{x}$ with zero mean, variance $\mathbf{C}$.
- Asymptotic behaviour slow to arise for heavy-tailed distributions.
- Stability issues with outliers.
- Statistical inference methods using sample covariance matrix $\hat{\mathbf{S}}_n$ (SCM) not appropriate,

$$\hat{\mathbf{S}}_N = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}.$$

- Instead, one uses robust M-estimators, such as fixed-point SCM $\hat{\mathbf{C}}_N$, solution to

$$\hat{\mathbf{C}}_N = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} \mathbf{x}_i^{\mathrm{H}} \hat{\mathbf{C}}_N^{-1} \mathbf{x}_i\right) \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}$$

  for some well-chosen $u(x)$.
- Typically,
  - **Tyler:** $u(x) = 1/x$ (but $\hat{\mathbf{C}}_N$ non-unique)
  - **Maronna:** $u(x)$ continuous on $[0, \infty)$ nonincreasing, $\phi(x) = xu(x)$ nondecreasing bounded.
- For elliptical distributions with density

$$f(\mathbf{x}) = Kg\left((\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{H}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right)$$

  $\hat{\mathbf{C}}_n$ is an n-consistent estimator of the scatter matrix $\boldsymbol{\Sigma}$.

Problem statement

R. A. Maronna, "Robust M-estimators of multivariate location and scatter", The annals of statistics, pp. 51-67, 1976.

- Observations $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$ of a random vector $\mathbf{x}$ with zero mean, variance $\mathbf{C}$.
- Asymptotic behaviour slow to arise for heavy-tailed distributions.
- Stability issues with outliers.
- Statistical inference methods using sample covariance matrix $\hat{\mathbf{S}}_n$ (SCM) not appropriate,

$$\hat{\mathbf{S}}_N = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}.$$

- Instead, one uses robust M-estimators, such as fixed-point SCM $\hat{\mathbf{C}}_N$, solution to

$$\hat{\mathbf{C}}_N = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N} \mathbf{x}_i^{\mathrm{H}} \hat{\mathbf{C}}_N^{-1} \mathbf{x}_i\right) \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}}$$

  for some well-chosen $u(x)$.
- Typically,
  - **Tyler:** $u(x) = 1/x$ (but $\hat{\mathbf{C}}_N$ non-unique)
  - **Maronna:** $u(x)$ continuous on $[0, \infty)$ nonincreasing, $\phi(x) = xu(x)$ nondecreasing bounded.
- For elliptical distributions with density

$$f(\mathbf{x}) = Kg\left((\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{H}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right)$$

  $\hat{\mathbf{C}}_n$ is an n-consistent estimator of the scatter matrix $\boldsymbol{\Sigma}$.
- Objective is to study $\hat{\mathbf{C}}_N$ in large dimensional RMT setting.

Stieltjes transform approach

R. Couillet, F. Pascal, *(on-going work)*.

Theorem

Assume $u(x)$ of Maronna-type. As $N, n \to \infty$ with $N/n \to c$, for almost every sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$,

$$F^{\hat{\mathbf{C}}_N} - F^{u(1)\hat{\mathbf{S}}_N} \Rightarrow 0.$$

Stieltjes transform approach

R. Couillet, F. Pascal, *(on-going work)*.

Theorem

Assume $u(x)$ of Maronna-type. As $N, n \to \infty$ with $N/n \to c$, for almost every sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$,

$$F^{\hat{\mathbf{C}}_N} - F^{u(1)\hat{\mathbf{S}}_N} \Rightarrow 0.$$

- Proof based on relaxation of $\hat{\mathbf{C}}_n$ into

$$\hat{\mathbf{C}}_n(z) = \frac{1}{u\left(\frac{e_N(z)}{1+ce_N(z)}\right)} \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N}\mathbf{x}_i^{\mathsf{H}}(\hat{\mathbf{C}}_n(z) - z\mathbf{I}_N)^{-1}\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^{\mathsf{H}}$$

  with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{t}{1+ce} - z} dF^{\mathbf{C}_N}(t).$$

- In particular, $\hat{\mathbf{C}}_N(0) = \frac{1}{u(1)}\hat{\mathbf{C}}_N$.

Stieltjes transform approach

R. Couillet, F. Pascal, *(on-going work)*.

Theorem

Assume $u(x)$ of Maronna-type. As $N, n \to \infty$ with $N/n \to c$, for almost every sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^N$,

$$F^{\hat{\mathbf{C}}_N} - F^{u(1)\hat{\mathbf{S}}_N} \Rightarrow 0.$$

- Proof based on relaxation of $\hat{\mathbf{C}}_n$ into

$$\hat{\mathbf{C}}_n(z) = \frac{1}{u\left(\frac{e_N(z)}{1+ce_N(z)}\right)} \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N}\mathbf{x}_i^{\mathrm{H}}(\hat{\mathbf{C}}_n(z) - z\mathbf{I}_N)^{-1}\mathbf{x}_i\right) \mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}$$

  with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{t}{1+ce} - z} dF^{\mathbf{C}_N}(t).$$

- In particular, $\hat{\mathbf{C}}_N(0) = \frac{1}{u(1)}\hat{\mathbf{C}}_N$.
- We show that:
  - $\hat{\mathbf{C}}_n(z)$ exists and is unique for $z \leqslant 0$
  - $\frac{1}{N}\mathrm{tr}\,(\hat{\mathbf{S}}_N - z'\mathbf{I}_N)^{-1} - \frac{1}{N}\mathrm{tr}\,(\hat{\mathbf{C}}_N(z) - z'\mathbf{I}_N)^{-1} \to 0$ for all z, z'
  - Extension to $\hat{\mathbf{C}}_N$ done by analytic continuation arguments.

Sketch of proof

Some reminders:

- Limiting spectrum of $\hat{\mathbf{S}}_N$

$$\frac{1}{N}\operatorname{tr}\left(\hat{\mathbf{S}}_N - z\mathbf{I}_N\right)^{-1} - m_n(z) \xrightarrow{\text{a.s.}} 0$$

where

$$m_n(z) = \int \frac{1}{\frac{1}{1+ce_N(z)}t - z} dF^{\mathbf{C}_N}(t)$$

with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{1}{1+ce}t - z} dF^{\mathbf{C}_N}(t).$$

Sketch of proof

Some reminders:

- Limiting spectrum of $\hat{\mathbf{S}}_N$

$$\frac{1}{N}\operatorname{tr}\left(\hat{\mathbf{S}}_N - z\mathbf{I}_N\right)^{-1} - m_n(z) \xrightarrow{\text{a.s.}} 0$$

  where

$$m_n(z) = \int \frac{1}{\frac{1}{1+ce_N(z)}t - z} dF^{\mathbf{C}_N}(t)$$

  with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{1}{1+ce}t - z} dF^{\mathbf{C}_N}(t).$$

- Restriction of $e_N(z)$ to $\mathbb{R}^-$ is an increasing function, so $\frac{e_N(z)}{1+ce_N(z)}$ is increasing and has image on $(0, 1]$.

Sketch of proof

Some reminders:

- Limiting spectrum of $\hat{\mathbf{S}}_N$

$$\frac{1}{N}\mathrm{tr}\,(\hat{\mathbf{S}}_N - z\mathbf{I}_N)^{-1} - m_n(z) \xrightarrow{\mathrm{a.s.}} 0$$

  where

$$m_n(z) = \int \frac{1}{\frac{1}{1+ce_N(z)}t - z} dF^{\mathbf{C}_N}(t)$$

  with $e_N(z)$ solution to

$$e = \int \frac{t}{\frac{1}{1+ce}t - z} dF^{\mathbf{C}_N}(t).$$

- Restriction of $e_N(z)$ to $\mathbb{R}^-$ is an increasing function, so $\frac{e_N(z)}{1+ce_N(z)}$ is increasing and has image on $(0,1]$.
- For all i,

$$\mathbf{x}_i^{\mathrm{H}}(\hat{\mathbf{S}}_N - z\mathbf{I}_N)^{-1}\mathbf{x}_i - \frac{e_N(z)}{1+ce_N(z)} \xrightarrow{\mathrm{a.s.}} 0.$$

Sketch of proof (2)

- Identification of asymptotic equivalence between $\hat{\mathbf{S}}_N$ and $\hat{\mathbf{C}}_N(z)$.

Lemma

Denote

$$\mathbf{Q}_S^z = (\hat{\mathbf{S}}_N - z\mathbf{I}_N)^{-1}.$$

For $z \leqslant 0$ *and* $z' \in \mathbb{C} \setminus \mathbb{R}^+$,

$$(I)\ \frac{1}{N} tr \mathbf{Q}_S^{z'} = \frac{1}{N} tr \left(\frac{1}{u\left(\frac{e_N(z)}{1+ce_N(z)}\right)} \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N}\mathbf{x}_i^{\mathrm{H}} \mathbf{Q}_S^z \mathbf{x}_i\right) \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}} - z'\mathbf{I}_N \right)^{-1} + \varepsilon_n(z)$$

$$(II)\ \frac{1}{N} \mathbf{x}_j^{\mathrm{H}} \mathbf{Q}_S^z \mathbf{x}_j = \frac{1}{N} \mathbf{x}_j^{\mathrm{H}} \left(\frac{1}{u\left(\frac{e_N(z)}{1+ce_N(z)}\right)} \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{1}{N}\mathbf{x}_i^{\mathrm{H}} \mathbf{Q}_S^z \mathbf{x}_i\right) \mathbf{x}_i \mathbf{x}_i^{\mathrm{H}} - z\mathbf{I}_N \right)^{-1} \mathbf{x}_j + \varepsilon_n^j(z)$$

where $\varepsilon_n(z) \xrightarrow{\text{a.s.}} 0$ *and* $\sup_j |\varepsilon_n^j(z)| \xrightarrow{\text{a.s.}} 0$ *as* N, n *grow large.*

Sketch of proof (3)

- Securing the spectrum of $\hat{\mathbf{C}}_N(z)$ for $z \to 0$.

Lemma

There exists $\hat{c}_-, \hat{c}_+ > 0$ *such that, with probability one, for all large* N*,*

$$\hat{c}_- \leqslant \lambda_{\min}(\hat{\mathbf{C}}_N(z)) < \lambda_{\max}(\hat{\mathbf{C}}_N(z)) \leqslant \hat{c}_+.$$

Sketch of proof (3)

- Securing the spectrum of $\hat{\mathbf{C}}_N(z)$ for $z \to 0$.

Lemma

There exists $\hat{c}_-, \hat{c}_+ > 0$ *such that, with probability one, for all large* N*,*

$$\hat{c}_- \leqslant \lambda_{\min}(\hat{\mathbf{C}}_N(z)) < \lambda_{\max}(\hat{\mathbf{C}}_N(z)) \leqslant \hat{c}_+.$$

- In order for this to hold, continuity of $u(x)$ in $x = 0$ is fundamental!
- $u(x) = 1/x$ does not work here... This is a major limitation to generalization to Tyler-type!

Sketch of proof (4)

- Under the above conditions, one can then show

For all $z \leqslant 0$, $z' \in \mathbb{C} \setminus \mathbb{R}^+$,

$$\frac{1}{N}\mathbf{x}_i^{\mathrm{H}}\left(\hat{\mathbf{S}}_N - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i - \frac{1}{N}\mathbf{x}_i^{\mathrm{H}}\left(\hat{\mathbf{C}}_N(z) - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i \xrightarrow{\text{a.s.}} 0$$

$$\frac{1}{N}\operatorname{tr}\left(\hat{\mathbf{S}}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\operatorname{tr}\left(\hat{\mathbf{C}}_N(z) - z'\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0.$$

which gives the final result.

Sketch of proof (4)

- Under the above conditions, one can then show

For all $z \leqslant 0$, $z' \in \mathbb{C} \setminus \mathbb{R}^+$,

$$\frac{1}{N}\mathbf{x}_i^{\mathrm{H}}\left(\hat{\mathbf{S}}_N - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i - \frac{1}{N}\mathbf{x}_i^{\mathrm{H}}\left(\hat{\mathbf{C}}_N(z) - z\mathbf{I}_N\right)^{-1}\mathbf{x}_i \xrightarrow{\text{a.s.}} 0$$

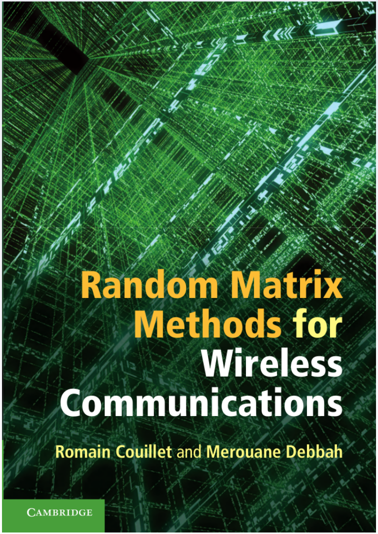
$$\frac{1}{N}\mathrm{tr}\left(\hat{\mathbf{S}}_N - z\mathbf{I}_N\right)^{-1} - \frac{1}{N}\mathrm{tr}\left(\hat{\mathbf{C}}_N(z) - z'\mathbf{I}_N\right)^{-1} \xrightarrow{\text{a.s.}} 0.$$

which gives the final result.

Conclusions:

- Most methods of statistical inference for SCM carry over to FP-SCM!
- CLT results should provide efficiency of these statistical tests.

To know more about all this



The end

Thank you