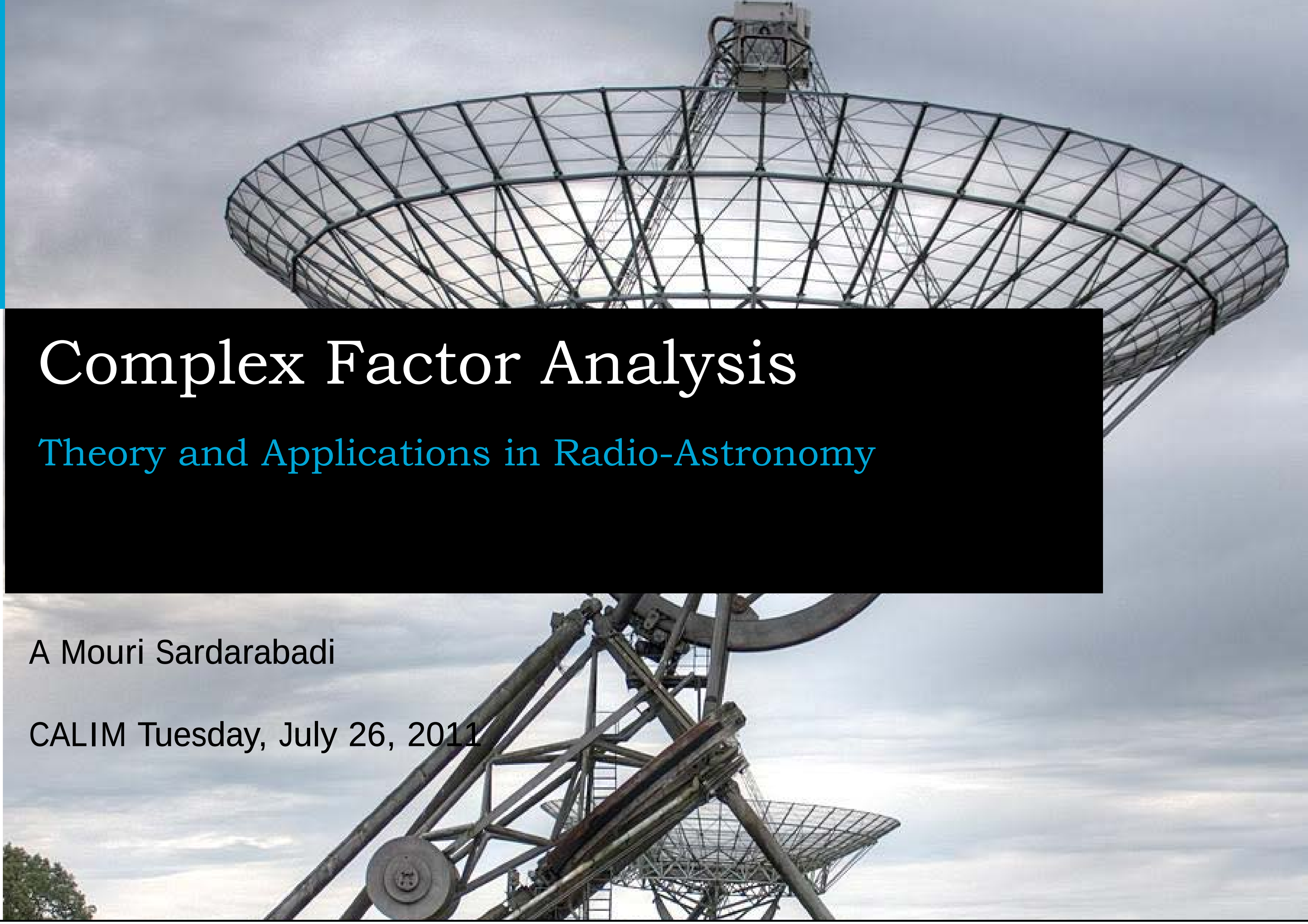


Complex Factor Analysis

Theory and Applications in Radio-Astronomy

A Mouri Sardarabadi

CALIM Tuesday, July 26, 2011

Content

- Data Model for Array Processing
- Factor Analysis
- Estimation
- Cramér–Rao Bound
- Validation/Detection
- Simulations
- Conclusion

Data Model

- Commonly used signal processing model
$$\mathbf{x}(t) = \widetilde{\mathbf{A}}\tilde{\mathbf{s}}(t) + \mathbf{n}(t)$$
- $\mathbf{x}$ is a $p \times 1$ vector of measurements
- $\widetilde{\mathbf{A}}$ is a $p \times m$ matrix containing the array response
- $\tilde{\mathbf{s}}$ is $m \times 1$ vector of sources
- $\mathbf{n}$ is a $p \times 1$ vector of noises
- Sampling the data at $t = kT_s$ where T_s is the sampling period, gives us
$$\mathbf{x}[k] = \widetilde{\mathbf{A}}\tilde{\mathbf{s}}[k] + \mathbf{n}[k]$$
- The covariance matrix is given by
$$\mathbf{\Sigma} = \widetilde{\mathbf{A}}\mathbf{\Sigma}_{\tilde{\mathbf{s}}}\widetilde{\mathbf{A}}^H + \mathbf{\Sigma}_{\mathbf{n}}$$

Factor Analysis

- The case that $\boldsymbol{\Sigma}_n = \sigma^2 \mathbf{I}$ is well studied in literature
- Noise is spatially uncorrelated

$$\boldsymbol{\Sigma}_{\mathbf{n}} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix

- Given any invertible matrix $\mathbf{Z}$ and any unitary matrix $\mathbf{Q}$

$$\mathbf{x} = \underbrace{\widetilde{\mathbf{A}}\mathbf{Z}\mathbf{Q}}_{\mathbf{A}}\,\underbrace{\mathbf{Q}^H\mathbf{Z}^{-1}\tilde{\mathbf{s}}}_{\mathbf{s}} + \mathbf{n}$$

$\mathbf{Z}$ could be chosen in such a way that we can rewrite

$$\boxed{\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^H + \mathbf{D}}$$

- $\mathbf{Q}$ is chosen in such a way that $\mathbf{A}^H\mathbf{D}^{-1}\mathbf{A}$ becomes a real diagonal matrix

Estimation (1)

- Given N samples from $\mathbf{x}$ the sample covariance matrix is given by

$$\mathbf{S} = \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{x}[k]\mathbf{x}^H[k]$$

- The model parameters in

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^H + \mathbf{D}$$

are estimated from $\boldsymbol{S}$

Various Algorithms for Estimation

- Various methods exist to find the model parameters
- Maximum likelihood estimation
  - Cost function is the likelihood $p(\mathbf{x}; \mathbf{A}, \mathbf{D}, \boldsymbol{S})$
  - Scoring method
  - Cost function Kullback–Leibler divergence
  - KLD Algorithm
- Least Squares
  - Cost function is the mean square error $\left\| \boldsymbol{S} - \left(\hat{\mathbf{A}}\hat{\mathbf{A}}^H + \hat{\mathbf{D}}\right) \right\|_F^2$
  - Alternating least squares

Maximum Likelihood Estimator

- Proper complex Gaussian distribution for noise and signals
- The likelihood that we want to maximize is

$$p(\mathbf{x};\mathbf{A},\mathbf{D},\mathbf{S})=\frac{1}{\pi^{pN}|\boldsymbol{\Sigma}|^{N}}\,\mathrm{e}^{-N\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})}$$

- By setting the Fisher score equal to zero the model parameters could be estimated

$$\begin{aligned}\mathbf{T}_{\mathbf{A}}&=-N\boldsymbol{\Sigma}^{-1}\mathbf{A}+N\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{A}\\ \mathbf{T}_{\mathbf{D}}&=N\mathrm{diag}(-\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1})\end{aligned}$$

- Too complex to solve analytically

Scoring Method

- Scoring Method is one way to find the MLE numerically
- It is an iterative method
- Let

$$\boldsymbol{\theta} = [\mathbf{a}_1^T, \dots, \mathbf{a}_m^T, \mathbf{d}^T]^T$$

- $\mathbf{a}_i$ is the ith column of $\mathbf{A}$ and $\mathbf{d}$ contains diagonal elements of $\mathbf{D}$

$$\widehat{\boldsymbol{\theta}}_{i+1} = \widehat{\boldsymbol{\theta}}_i + \mu \mathbf{F}^{\dagger}(\boldsymbol{\theta}) \mathbf{t}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_i}$$

where $\mathbf{t}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = [\mathrm{vect}(\mathbf{T}_{\mathbf{A}})^T, \mathrm{vect}(\mathbf{T}_{\mathbf{D}})^T]^T$ and $\mathbf{F}$ is the Fisher information matrix

Fisher Information Matrix

- If the estimated parameters are partitioned as $\boldsymbol{\theta} = [\mathbf{a}_1^T, \dots, \mathbf{a}_m^T, \mathbf{d}^T]^T$ then the Fisher information can be written as

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{\mathbf{a}_1\mathbf{a}_1} & \cdots & \mathbf{F}_{\mathbf{a}_1\mathbf{a}_m} & \mathbf{F}_{\mathbf{a}_1\mathbf{d}} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{F}_{\mathbf{a}_m\mathbf{a}_1} & \cdots & \mathbf{F}_{\mathbf{a}_m\mathbf{a}_m} & \mathbf{F}_{\mathbf{a}_m\mathbf{d}} \\ \mathbf{F}_{\mathbf{d}\mathbf{a}_1} & \cdots & \mathbf{F}_{\mathbf{d}\mathbf{a}_m} & \mathbf{F}_{\mathbf{d}\mathbf{d}} \end{pmatrix}$$

- For the FA model these sub-matrices are:

$$\mathbf{F}^*_{\mathbf{a}_k\mathbf{a}_n} = N\mathbf{a}_n^H\boldsymbol{\Sigma}^{-1}\mathbf{a}_k\boldsymbol{\Sigma}^{-1}$$
$$\mathbf{F}^*_{\mathbf{a}_k\mathbf{d}} = N\boldsymbol{\Sigma}^{-1}\mathrm{diag}(\boldsymbol{\Sigma}^{-1}\mathbf{a}_k)$$
$$\mathbf{F}_{\mathbf{d}\mathbf{d}} = N(\boldsymbol{\Sigma}^{-1}\odot\boldsymbol{\Sigma}^{-T})$$

Problems with Scoring Method

- Convergence is not guaranteed
- Sensitive to initial guess

$$\mathbf{D}_0 = \mathbf{diag}(\mathbf{S}^{-1})^{-1}$$

- Even if it converges, it might converge to a local maximum
- Size of the Fisher information could become very large

Kullback–Leibler divergence

- Tries to minimize the “distance” between two families of distributions
- The final iteration steps are

$$\widehat{\mathbf{A}}_{i+1} = \mathbf{S}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\mathbf{A}}_i\boldsymbol{\Phi}_i^{-1}$$
$$\widehat{\mathbf{D}}_{i+1} = \mathrm{diag}(\boldsymbol{S} - \widehat{\mathbf{A}}_{i+1}\widehat{\mathbf{A}}_i^H\widehat{\boldsymbol{\Sigma}}_i^{-1}\mathbf{S})$$
$$\boldsymbol{\Phi}_i = \mathbf{I} - \widehat{\mathbf{A}}_i^H\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\mathbf{A}}_i + \widehat{\mathbf{A}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1}\mathbf{S}\widehat{\boldsymbol{\Sigma}}_i^{-1}\widehat{\mathbf{A}}_i$$

- Shares the same convergence properties of the EM algorithm

Alternating Least Squares

- Minimize the MSE between $\mathbf{S}$ and $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H + \widehat{\mathbf{D}}$
- Two stage minimization
  - In the first stage $\widehat{\mathbf{A}}$ is held constant and $\widehat{\mathbf{D}}$ is found

$$\widehat{\mathbf{D}}_{i+1} = \mathrm{diag}(\mathbf{S} - \widehat{\mathbf{A}}_i\widehat{\mathbf{A}}_i^H)$$

  - In stage two the $\widehat{\mathbf{D}}$ is held constant and $\widehat{\mathbf{A}}$ is calculated

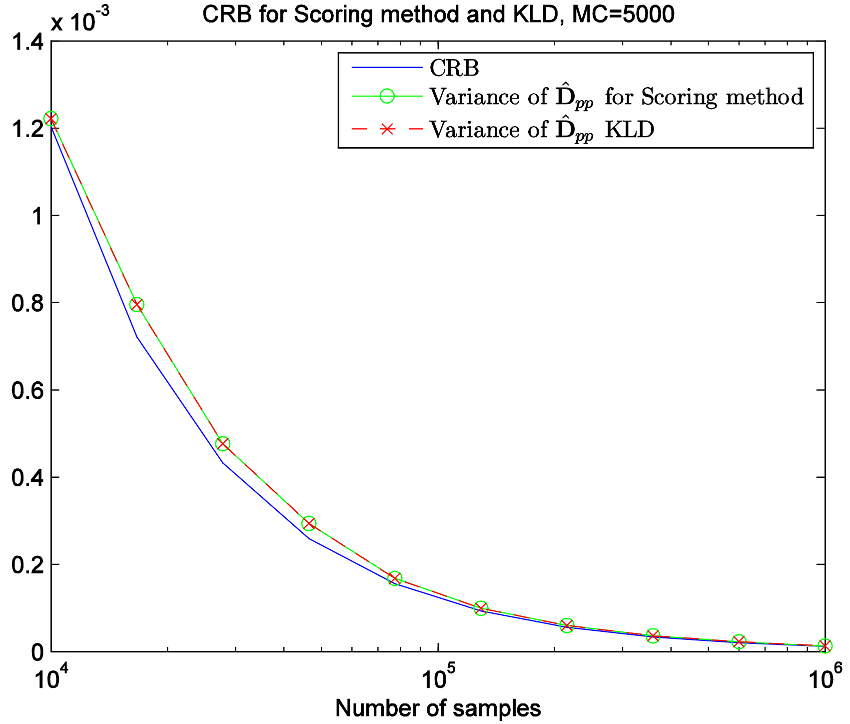
$$\widehat{\mathbf{A}}_{i+1} = \mathbf{U}_m\mathbf{L}_m^{\frac{1}{2}}$$

where $\mathbf{L}_m$ is a diagonal matrix containing m largest eigenvalues of $\mathbf{S} - \widehat{\mathbf{D}}_{i+1}$ and $\mathbf{U}_m$ is a matrix of size $\mathrm{p} \times m$ containing the corresponding eigenvectors

Cramér–Rao Bound

- For an unbiased estimator the CRB is the lowest bound on the covariance matrix, $\mathbf{C}$, of the estimated parameters, $\boldsymbol{\theta}$

$$\mathbf{C}(\boldsymbol{\theta}) \geq \mathbf{F}^{-1}$$



Validation/Detection

- After the parameters have been calculated the question remains if the FA model "explains" the data well enough
- General likelihood ratio test (GLRT)
- Two hypotheses:
  - H_1 is the case that no model is imposed on the data
  - H_0 is the case that the FA model "explains" the data

$$l = \frac{\log(L_1)}{\log(L_0)} < \gamma$$

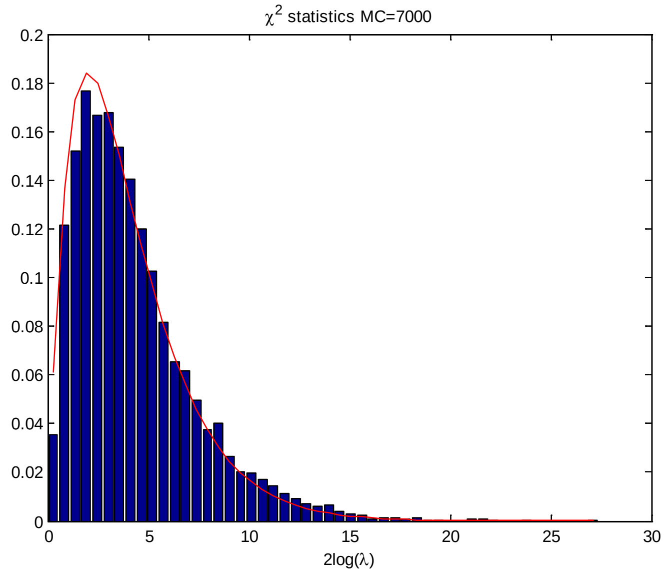
L_i is the maximum value of the likelihood under H_i

Test Statistics

- Under H_0 the GLRT, $2l$ has a central χ_s^2 distribution

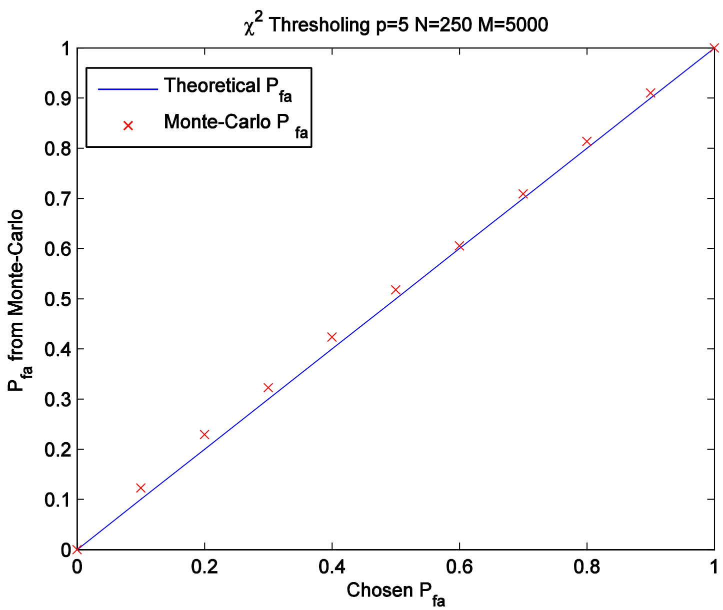
$$s = (p - m)^2 - p > 0$$

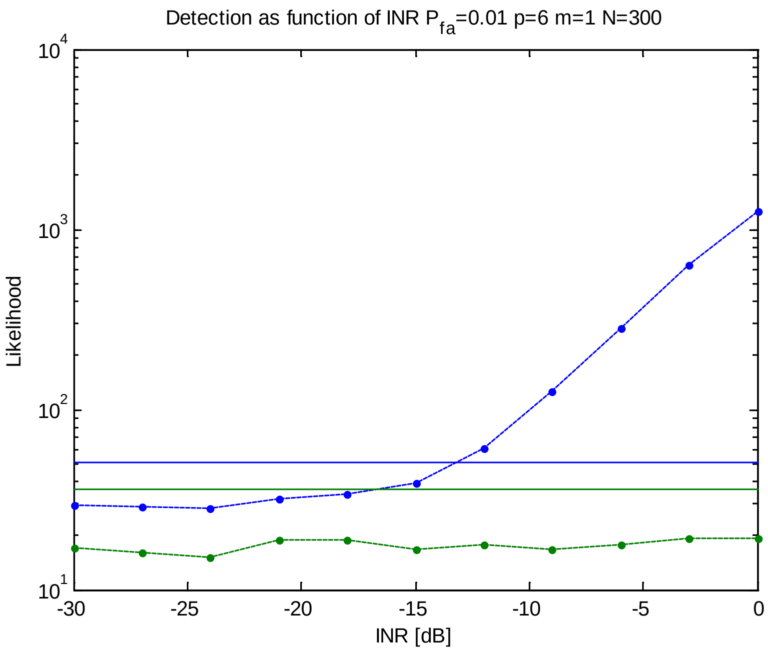
- Threshold could be found based on this distribution
- Simulation
  - $p = 5$
  - $m = 2$
  - $s = 4$



Constant False Alarm Detector

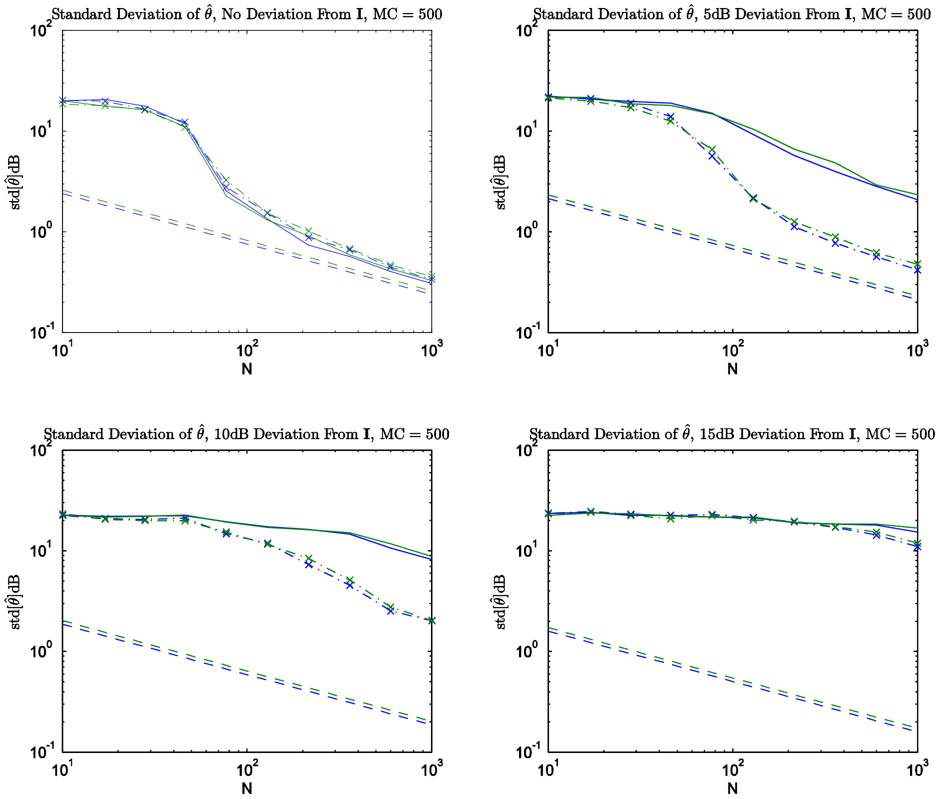
- Especial case is when $\hat{m} = 0$ then the GLRT becomes a constant false alarm detector

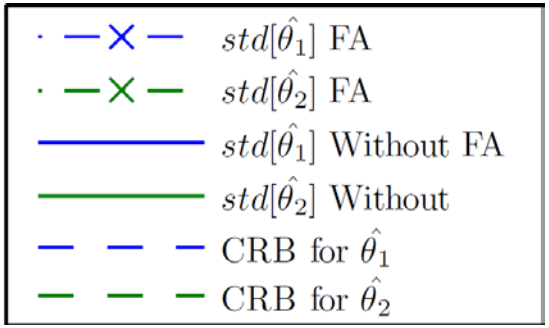


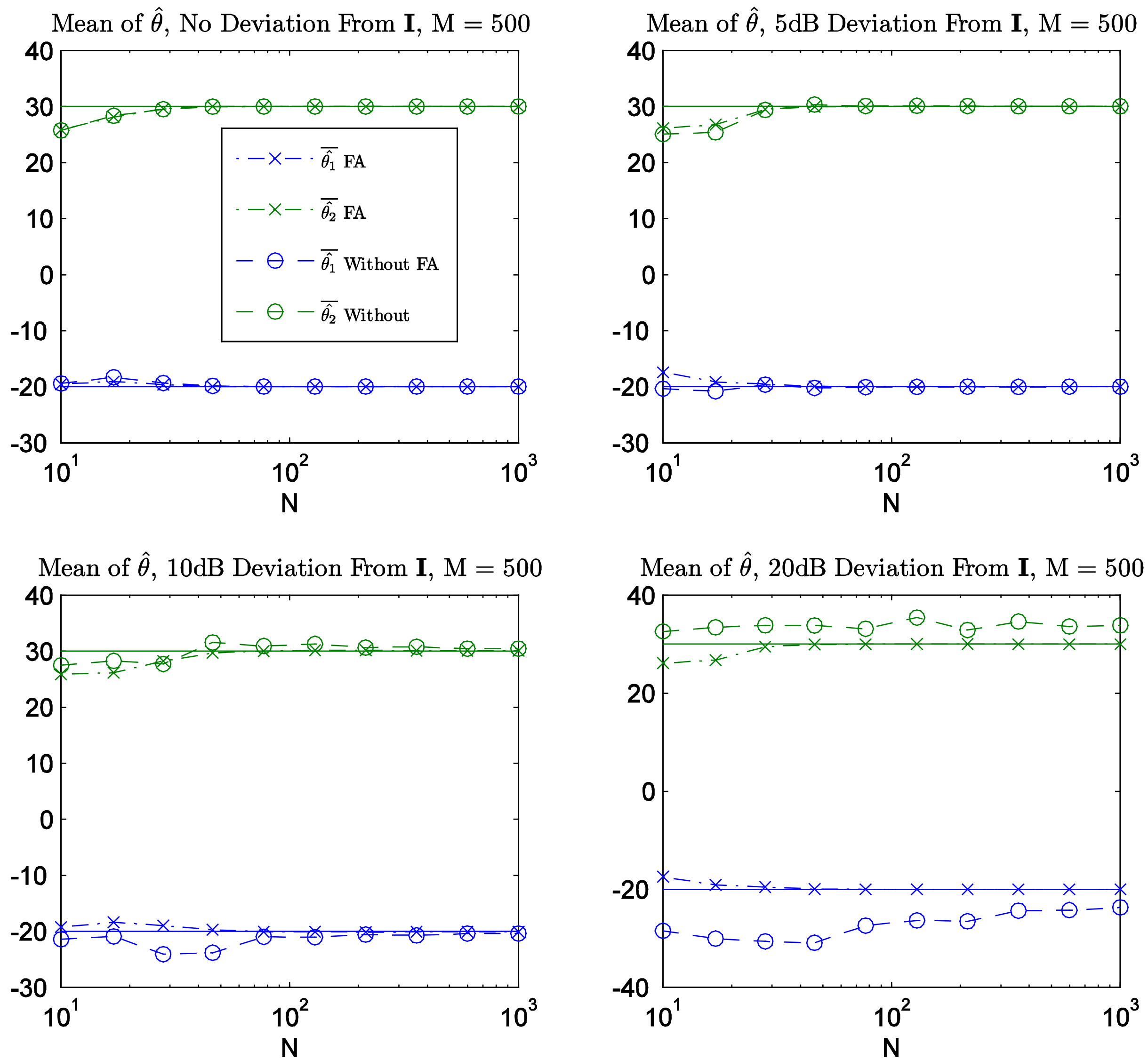


Simulations - DOA

- Direction of arrival (DOA) using ESPRIT
- 2 sources at -20 and 30 from broadside



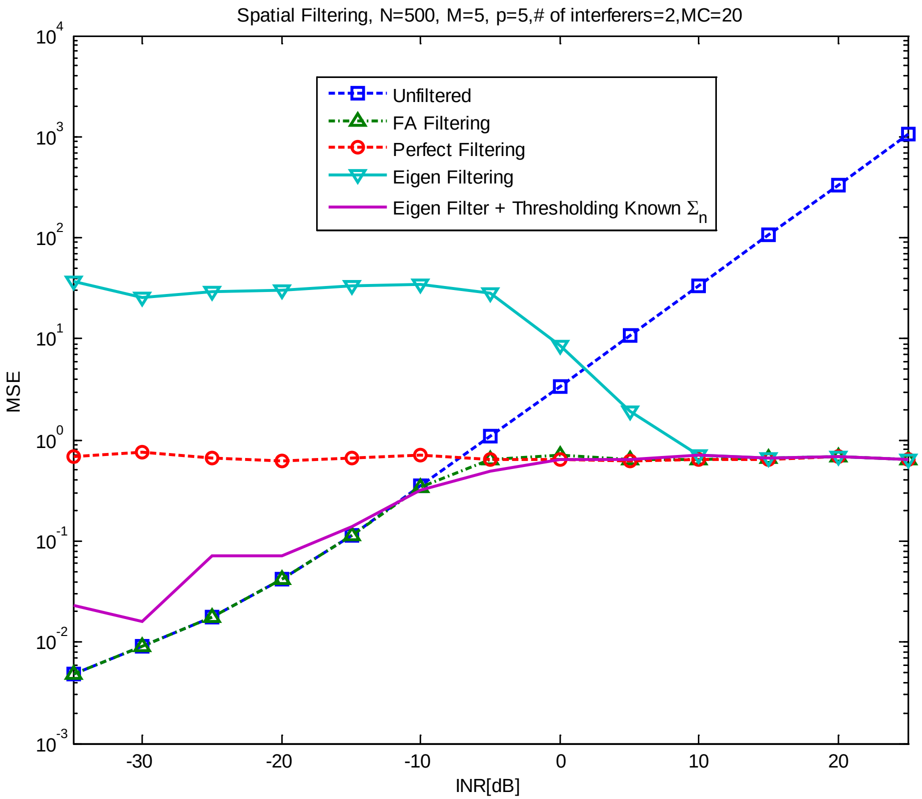




Mean of $\hat{\theta}$, No Deviation From **I**, M = 500
Mean of $\hat{\theta}$, 5dB Deviation From **I**, M = 500
Mean of $\hat{\theta}$, 10dB Deviation From **I**, M = 500
Mean of $\hat{\theta}$, 20dB Deviation From **I**, M = 500
$\overline{\hat{\theta}_1}$ FA
$\overline{\hat{\theta}_2}$ FA
$\overline{\hat{\theta}_1}$ Without FA
$\overline{\hat{\theta}_2}$ Without
N

Simulations – Spatial Filtering

- Celestial sources are very weak
- Spatial filtering on short-term correlations



Conclusions

- When the noise covariance is unknown FA can be used to model the data
- To make the model applicable for radio-astronomy it had to be extended to complex numbers
- Three different algorithms have been proposed for estimating the model and
- The validation of the model is shown with the help of a GLRT
- A constant false alarm detector and its statistics is shown
- With the help of simulations we showed that the algorithm has practical potential

Questions?