Robust covariance matrices estimation and applications in signal processing

F. Pascal
SONDRA/Supelec

GDR ISIS
Journée “Estimation et traitement statistique en grande dimension”

May 16^{th}, 2013

Motivations

Several SP applications require the covariance matrix estimation (sources localization, STAP, Polarimetric SAR classification, radar detection, MIMO...).

Classical radar applications consider the background to be Gaussian.

$\rightarrow$ The Sample Covariance Matrix (SCM)

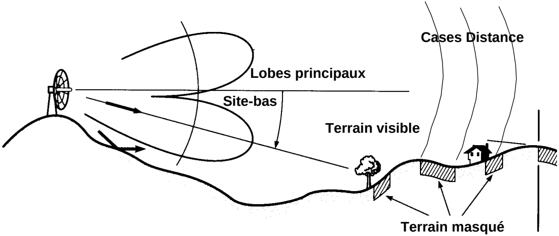
- a simple estimate
- well-known statistical properties

Robustness : what happens in non-Gaussian models ?

- High resolution techniques and/or low grazing angle radars
- Outliers and other parasites are not been taken into account with the Gaussian model.
- The SCM gives then poor results.

Why non Gaussian modeling (heterogeneous clutter) ?

- Grazing angle Radar



$\Rightarrow$ Impulsive Clutter

- High Resolution Radar

$\Rightarrow$ Small number of scatters in the Cell Under Test (CUT)
$\Rightarrow$ Central Limit Theorem (CLT) is not valid anymore

Failure of the OGD with non Gaussian background

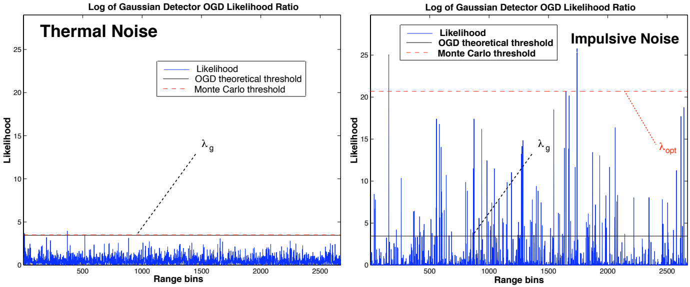


FIGURE : Failure of the OGD - Adjustment of the detection threshold - K-distributed clutter with same power as the Gaussian noise

- ⇒ Bad performance of the OGD in case of mismodeling
- ⇒ Introduction of elliptical distributions
- ⇒ Introduction of robust estimates

Results

- A more flexible and adjustable model
  $\rightsquigarrow$ *Elliptical distributions*
- A robust family of estimators
  $\rightsquigarrow$ *M-estimators*

To use the *M*-estimators for SP applications, we extend their statistical properties as well as

- the statistical property of the resulting ANMF (detection test)
- the statistical property of the MUSIC statistic (DoA estimation)

The relationship between its Probability of false alarm P_{fa} (Type-I error) and detection threshold is also derived.

Extension to the RMT

In many applications, the dimension of the observation m is large : Hyperspectral imaging, MIMO-STAP, ...

$\Rightarrow$ The required number N of observations for estimation purposes needs to be larger : $N \gg m$

$\Rightarrow$ BUT this is not the case in practice !

$\rightsquigarrow$ Random Matrix Theory

$\rightsquigarrow$ Main assumption : $N \to \infty, m \to \infty$ and $\frac{m}{N} \to c \in [0, 1]$

Preliminary results

Extension of the results on standard M-estimators :

- asymptotic distribution of the eigenvalues
- derivation of a robust G-MUSIC

Presentation outline

Modeling the background
Complex elliptical distributions

Let $\mathbf{z}$ be a complex circular random vector of length m. $\mathbf{z}$ has a complex elliptical distribution (CED) ($CE(\boldsymbol{\mu}, \boldsymbol{\Lambda}, g_{\mathbf{z}})$) if its PDF can be written

$$g_{\mathbf{z}}(\mathbf{z}) = |\boldsymbol{\Lambda}|^{-1} h_z((\mathbf{z}-\boldsymbol{\mu})^H \boldsymbol{\Lambda}^{-1}(\mathbf{z}-\boldsymbol{\mu})), \quad (1)$$

where $h_z : [0, \infty) \rightarrow [0, \infty)$ is the density generator and is such as (1) defines a pdf.

- $\boldsymbol{\mu}$ is the statistical mean
- $\boldsymbol{\Lambda}$ the scatter matrix

In general (finite second-order moment), $\mathbf{M} = \alpha\, \boldsymbol{\Lambda}$ where

- $\alpha = -2\phi'(0)$,
- ϕ is defined through the characteristic function $c_{\mathbf{x}}$of $\mathbf{x}$ by $c_{\mathbf{x}}(\mathbf{t}) = \exp(i\mathbf{t}^H\boldsymbol{\mu})\,\phi(\mathbf{t}^H\boldsymbol{\Lambda}\mathbf{t})$

Estimating the covariance matrix

M-estimators

PDF not specified
$\Rightarrow$ *MLE can not be derived*
$\Rightarrow$ *M-estimators are used instead*

Let $(\mathbf{z}_1, ..., \mathbf{z}_N)$ be a N-sample $\sim CE(\mathbf{0}, \mathbf{\Lambda}, g_{\mathbf{z}})$ of length m.

The complex M-estimator of $\mathbf{\Lambda}$ is defined as the solution of

$$\mathbf{V}_N = \frac{1}{N} \sum_{n=1}^{N} u\left(\mathbf{z}_n^H \mathbf{V}_N^{-1} \mathbf{z}_n^H\right) \mathbf{z}_n \mathbf{z}_n^H, \tag{2}$$

Maronna (1976), Kent and Tyler (1991)

- Existence
- Uniqueness
- Convergence of the recursive algorithm...

Examples of M-estimates

SCM :

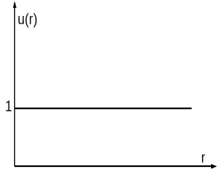
$u(r) = 1$

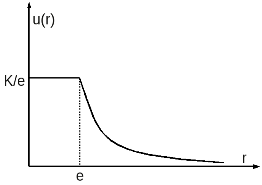
Huber's estimate (M-estimate) :

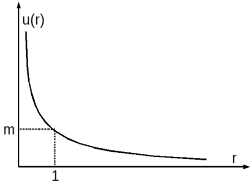
$$u(r) = \begin{cases} K/e \text{ if } r <= e \\ K/r \text{ if } r > e \end{cases}$$

FP Estimate :

$u(r) = \frac{m}{r}$







Remarks :

- Huber = mix between SCM and FP
- FP and SCM are "not" M-estimators
- FP estimator is the most robust.

FP Estimate (Tyler, 1987 ; Pascal, 2008) :

$$\mathbf{V}_N = \frac{m}{N} \sum_{n=1}^{N} \frac{\mathbf{z}_n \mathbf{z}_n^H}{\mathbf{z}_n^H \mathbf{V}_N^{-1} \mathbf{z}_n^H}$$

Context
M-estimators

Let us set

$$\mathbf{V} = E\left[u(\mathbf{z}'\mathbf{V}^{-1}\mathbf{z})\,\mathbf{z}\mathbf{z}'\right], \tag{3}$$

where $\mathbf{z} \sim CE(\mathbf{0}, \mathbf{\Lambda}, g_{\mathbf{z}})$.

- (3) admits a unique solution $\mathbf{V}$ and $\mathbf{V} = \sigma\mathbf{\Lambda} = \sigma/\alpha\,\mathbf{M}$ where σ is given by Tyler(1982),
- $\mathbf{V}_N$ is a consistent estimate of $\mathbf{V}$.

Asymptotic distribution of complex M-estimators

Using the results of Tyler (1982), we derived the following results (Ph.D of M. Mahot) :

Theorem 1 (Asymptotic distribution of $\mathbf{V}_N$)

$$\sqrt{N}\,\text{vec}(\mathbf{V}_N - \mathbf{V}) \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}\right), \quad (4)$$

where $\mathcal{CN}$ is the complex Gaussian distribution, $\boldsymbol{\Sigma}$ the CM and $\boldsymbol{\Omega}$ the pseudo CM :

$$\boldsymbol{\Sigma} = \sigma_1(\mathbf{V}^T \otimes \mathbf{V}) + \sigma_2 \text{vec}(\mathbf{V})\text{vec}(\mathbf{V})^H,$$
$$\boldsymbol{\Omega} = \sigma_1(\mathbf{V}^T \otimes \mathbf{V})\,\mathbf{K} + \sigma_2 \text{vec}(\mathbf{V})\text{vec}(\mathbf{V})^T,$$

where $\mathbf{K}$ is the commutation matrix.

The SCM is defined as $\mathbf{W}_N = \frac{1}{N}\sum_{n=1}^{N} \mathbf{z}_n \mathbf{z}_n^H$ where $\mathbf{z}_n$ are complex independent circular zero-mean Gaussian with CM $\mathbf{V}$. Then,

$$\sqrt{N}\,\text{vec}(\mathbf{W}_N - \mathbf{V}) \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Sigma}_W, \boldsymbol{\Omega}_W\right)$$

$$\boldsymbol{\Sigma}_W = (\mathbf{V}^T \otimes \mathbf{V})$$
$$\boldsymbol{\Omega}_W = (\mathbf{V}^T \otimes \mathbf{V})\,\mathbf{K}$$

An important property of complex M-estimators

- Let $\mathbf{V}_N$ an estimate of Hermitian positive-definite matrix $\mathbf{V}$ that satisfies

$$\sqrt{N}\left(\text{vec}(\mathbf{V}_N - \mathbf{V})\right) \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}, \mathbf{\Omega}\right), \tag{5}$$

with

$$\begin{cases} \mathbf{\Sigma} = \nu_1 \mathbf{V}^T \otimes \mathbf{V} + \nu_2 \text{vec}(\mathbf{V})\text{vec}(\mathbf{V})^H, \\ \mathbf{\Omega} = \nu_1 (\mathbf{V}^T \otimes \mathbf{V})\,\mathbf{K} + \nu_2 \text{vec}(\mathbf{V})\text{vec}(\mathbf{V})^T, \end{cases}$$

where ν_1 and ν_2 are any real numbers.

e.g.

	SCM	M-estimators	FP
ν_1	1	σ_1	$(m+1)/m$
ν_2	0	σ_2	$-(m+1)/m^2$
...	More accurate		More robust

- Let $H(\mathbf{V})$ be a r-multivariate function on the set of Hermitian positive-definite matrices, with continuous first partial derivatives and such as $H(\mathbf{V}) = H(\alpha\mathbf{V})$ for all $\alpha > 0$, e.g. the ANMF statistic, the MUSIC statistic.

An important property of complex M-estimators

Theorem 2 (Asymptotic distribution of $H(\mathbf{V}_N)$)

$$\sqrt{N}\left(H(\mathbf{V}_N) - H(\mathbf{V})\right) \xrightarrow{d} \mathcal{CN}\left(\mathbf{0}_{r,1}, \mathbf{\Sigma}_H, \mathbf{\Omega}_H\right) \quad (6)$$

where $\mathbf{\Sigma}_H$ and $\mathbf{\Omega}_H$ are defined as

$$\mathbf{\Sigma}_H = \nu_1 H'(\mathbf{V})(\mathbf{V}^T \otimes \mathbf{V}) H'(\mathbf{V})^H,$$
$$\mathbf{\Omega}_H = \nu_1 H'(\mathbf{V})(\mathbf{V}^T \otimes \mathbf{V})\, \mathbf{K} H'(\mathbf{V})^T,$$

where $H'(\mathbf{V}) = \left(\dfrac{\partial H(\mathbf{V})}{\partial \text{vec}(\mathbf{V})}\right)$.

$H(SCM)$ and $H(M$-estimators) share the same asymptotic distribution (differs from σ_1)

Application : Detection using the ANMF test

- In a m-vector $\mathbf{y}$, detecting a complex known signal $\mathbf{s} = A\,\mathbf{p}$ embedded in an additive noise $\mathbf{z}$ (with covariance matrix $\mathbf{V}$) , can be written as the following statistical test :

$$\begin{cases} \text{Hypothesis } H_0: & \mathbf{y} = \mathbf{z} & \mathbf{y}_n = \mathbf{z}_n & n = 1, \dots, N \\ \text{Hypothesis } H_1: & \mathbf{y} = \mathbf{s} + \mathbf{z} & \mathbf{y}_n = \mathbf{z}_n & n = 1, \dots, N \end{cases}$$

  where the $\mathbf{z}_n$'s are N "signal-free" independent observations (secondary data) used to estimate the noise parameters .
- Let $\mathbf{V}_N$ be an estimate of $\mathbf{V}$.

ANMF test

$$\Lambda(\mathbf{V}_N) = \frac{|\mathbf{p}^H \mathbf{V}_N^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \mathbf{V}_N^{-1} \mathbf{p})(\mathbf{y}^H \mathbf{V}_N^{-1} \mathbf{y})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$

One has $\Lambda(\mathbf{V}_N) = \Lambda(\alpha\,\mathbf{V}_N)$ for any $\alpha > 0$.

Probabilities of false alarm

P_{fa}-threshold relation in the Gaussian case of $\Lambda(SCM)$ (finite N)

$$P_{fa} = (1-\lambda)^{a-1} \, {}_2F_1(a, a-1; b-1; \lambda), \tag{7}$$

where $a = N - m + 2$, $b = N + 2$ and ${}_2F_1$ is the Hypergeometric function.

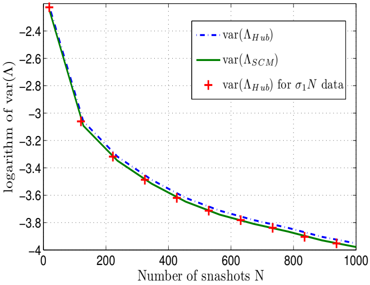
From theorem 2, one has

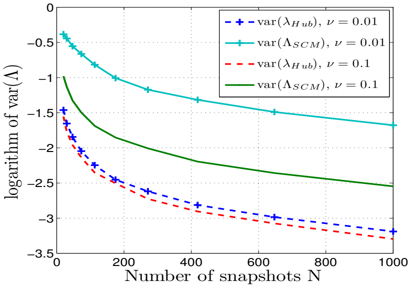
P_{fa}-threshold relation of $\Lambda(M\text{-estimators})$ for all elliptical distributions

For N large and any elliptically distributed noise, the PFA is still given by (7) if we replace N by N/σ_1.

Simulations

- Complex Huber's *M*-estimator.
- Figure 1 : Gaussian context, here $\sigma_1 = 1.066$.
- Figure 2 : K-distributed clutter (shape parameter : 0.1, and 0.01).



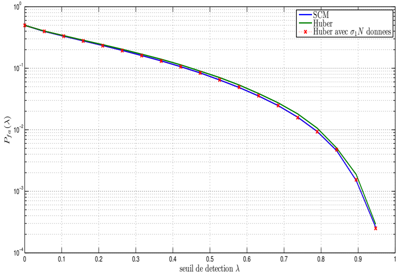


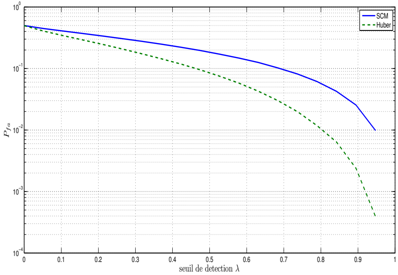
Validation of theorem (even for small *N*)

Interest of the *M*-estimators

Simulations : Probabilities of False Alarm

- Complex Huber's M-estimator.
- Figure 1 : Gaussian context, here $\sigma_1 = 1.066$.
- Figure 2 : K-distributed clutter (shape parameter : 0.1).



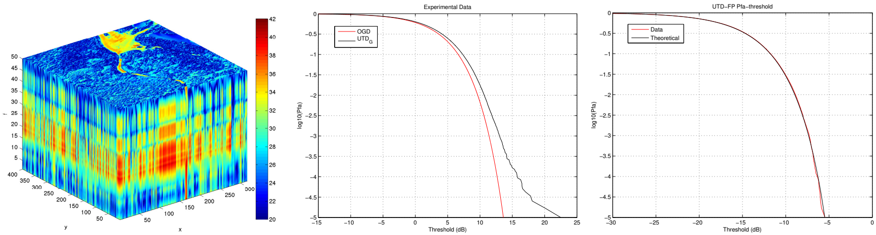


Validation of theorem (even for small N)

Interest of the M-estimators for False Alarm regulation

Hyperspectral Imaging - Ph.D of J. Frontera

Now, the statistical mean is non null



(a) Hyperspectral Data (b) AMF-H detector with the SCM (c) ANMF-H detector with the FP

FIGURE : Probability of false alarm versus the detection threshold for $m = 50$ and $N = 168$

Perspectives

- Open problem : joint M-estimators of the mean and the covariance matrix as solutions of fixed point equations
- Estimators performance
- Large dimensional problem : use of RMT

MUltiple Signal Classification (MUSIC) method for DoA estimation

- K direction of arrival θ_k on m antennas
- Gaussian stationary narrowband signal with DoA 20° with additive noise.
- the DoA is estimated from N snapshots, using the SCM and the Huber's M-estimator.

$$z_t = \sum_{k=1}^{K} \sqrt{p}_k s(\theta_k) y_{k,t} + \sigma w_t$$

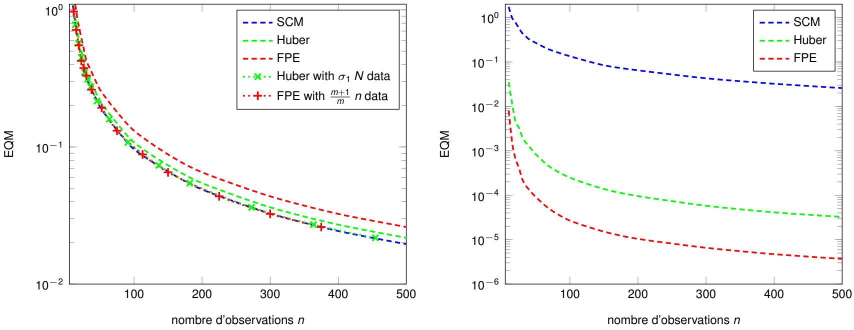
$$\begin{cases} H(\mathbf{V}) & = \gamma(\theta) = s(\theta)^H E_W E_W^H s(\theta), \qquad (\mathbf{V} \text{ known}) \\ H(\mathbf{V}_N) & = \hat{\gamma}(\theta) = \sum_{i=1}^{m-K} \lambda_i s(\theta)^H \hat{e}_i \hat{e}_i^H s(\theta) = H(\alpha \mathbf{V}_N), \qquad (\mathbf{V} \text{ unknown}) \end{cases}$$

where λ_i (resp. $\hat{e}_i$) are the eigenvalues (resp.eigenvectors) of $\mathbf{V}_N$.

The Mean Square Error (MSE) between the estimated angle $\widehat{\theta}$ and the real angle θ is then computed (case of one source).

Simulation using the MUltiple Signal Classification (MUSIC) method

- A $m = 3$ uniform linear array (ULA) with half wavelength sensors spacing is used,
- Gaussian stationary narrowband signal with DoA 20° with additive noise.
- the DoA is estimated from N snapshots, using the SCM, the Huber's M-estimator and the FP estimator.



(a) White Gaussian additive noise (b) K-distributed additive noise ($\nu = 0.1$)

FIGURE : MSE of $\hat{\theta}$ vs the number N of observations, with $m = 3$.

RMT - Classical results

Assumptions :

- $N, m \to \infty$ and $\dfrac{m}{N} \to c \in (0,1)$ and $\mathbf{W}_N = \dfrac{1}{N}\sum_{n=1}^{N} \mathbf{z}_n \mathbf{z}_n^H$ the SCM
- $(\mathbf{z}_1, ..., \mathbf{z}_N)$ be a N-sample, i.i.d with finite fourth-order moment

Thus one has :

1) $F^{\mathbf{W}_N} \Rightarrow F^{MP}$
where $F^{\mathbf{W}_N}$ (resp. F^{MP}) stands for the distribution of the eigenvalues of $\mathbf{W}_N$ (resp. the Marcenko-Pastur distribution) and $\Rightarrow$ stands for the weak convergence.

2) $\hat{\gamma}(\theta) = \sum_{i=1}^{m} \beta_i s(\theta)^H \hat{e}_i \hat{e}_i^H s(\theta)$ is the G-MUSIC statistic (Mestre, 2008)
where

$$\beta_i = \begin{cases} 1 + \sum_{k=N-K+1}^{N} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_i - \hat{\mu}_k} \right) & , i \leq N-K \\ -\sum_{k=1}^{N-K} \left(\frac{\hat{\lambda}_k}{\hat{\lambda}_i - \hat{\lambda}_k} - \frac{\hat{\mu}_k}{\hat{\lambda}_i - \hat{\mu}_k} \right) & , i > N-K \end{cases}$$

with $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_N$ the eigenvalues of $\mathbf{W}_N$ and $\hat{\mu}_1 \leq \ldots \leq \hat{\mu}_N$ the eigenvalues of

$\mathrm{diag}(\hat{\boldsymbol{\lambda}}) - \frac{1}{n}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^T$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^T$.

Robust RMT with R. Couillet and J.W. Silverstein

Assumptions :

- $N, m \to \infty$ and $\dfrac{m}{N} \to c \in (0,1)$ and $\mathbf{V}_N$ a M-estimor (with previous assumptions)
- $(\mathbf{z}_1, ..., \mathbf{z}_N)$ be a N-sample, i.i.d with finite fourth-order moment

Thus, one has shown in [1] :

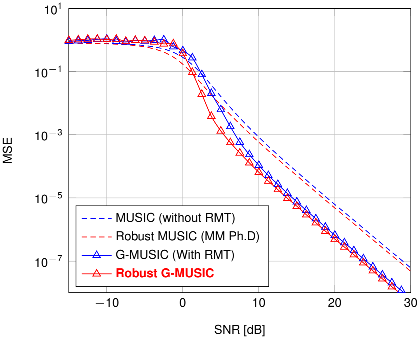
1) $\|\psi^{-1}(1)\,\mathbf{V}_N - \mathbf{W}_N\| \xrightarrow{a.s.} 0$ when $N, m \to \infty$ and $\dfrac{m}{N} \to c$
where $\|.\|$ stands for the spectral norm.

Classical results in RMT can be extended to the *M*-estimators

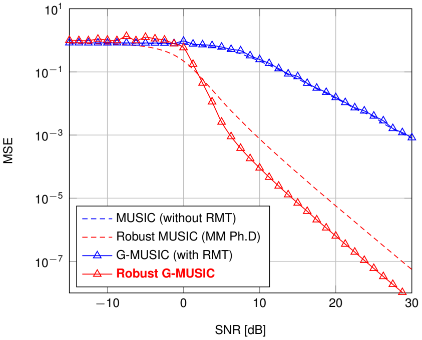
2) $\hat{\gamma}(\theta) = \sum_{i=1}^{m} \beta_i s(\theta)^H \hat{e}_i \hat{e}_i^H s(\theta)$ is STILL the G-MUSIC statistic for the M-estimators
(for the eigenvalues of $\mathbf{V}_N$)

[1] R. Couillet, F. Pascal et J. W. Silverstein, ”Robust M-Estimation for Array Processing : A Random Matrix Approach”, *Information Theory, IEEE Transactions on (submitted to)*, 2012. arXiv :1204.5320v1.

Application to DoA estimation with MUSIC for different additive clutter



(a) Homogeneous noise ($\simeq$ Gaussian), 50 data of size 10



(b) Heterogeneous clutter, 50 data of size 10

FIGURE : MSE performance of the various MUSIC estimators for $K = 1$ source

$\simeq$ 6dB gain for DoA estimation with MUSIC algorithm for various clutter scenarios

Resolution probability of 2 sources

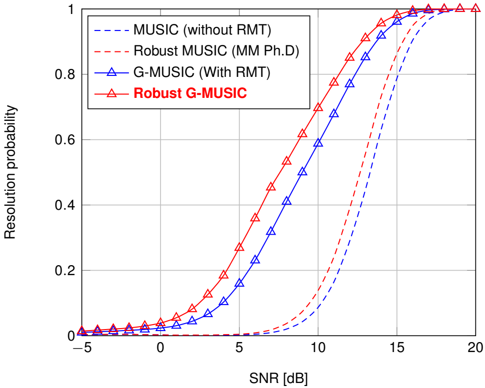


FIGURE : Resolution performance of the MUSIC estimators in homogeneous clutter for 50 data of size 10

$\simeq$ 7dB gain for resolution performance on classical MUSIC

Conclusions and Perspectives

- Conclusions
  - Derivation of the complex M-estimators asymptotic distribution, the robust ANMF and the MUSIC statistic asymptotic distributions.
  - In the Gaussian case, M-estimators built with $\sigma_1 N$ data behaves as SCM built with N data (i.e. slight loss of performance in Gaussian case).
  - Better estimation in non-Gaussian cases.
  - Extension to the Robust RMT and derivation of the Robust G-MUSIC method.
- Perspectives
  - Low Rank techniques for robust estimation
  - Robust estimation with a location parameter (non-zero-mean observation) : e.g. Hyperspectral imaging
  - Second-order moment in RMT
  - Eigenvalues distribution of the FP estimator (open-problem)

Thank you for your attention !

Questions ?