Robust Estimation, Random Matrix Theory and Applications to Signal Processing

Frédéric Pascal

CentraleSupelec, Laboratory of Signals and Systems (L2S), France
frederic.pascal@centralesupelec.fr
http://fredericpascal.blogspot.fr

IEEE/EURASIP Summer School on Robust Signal Processing
Sep. 21, 2016

Contents

- **Part A**
  Robust Covariance Matrix Estimation: CES distributions, ML-, M- and regularized M-estimators

- **Part B**
  M- and regularized M-estimators in the large dimensional regime

- **Part C**
  Application(s): DoA estimation, target detection, distribution fitting, Hyperspectral processing ...

Outline

Motivations ...

Signal Processing applications

- Application reality: only observations $\Rightarrow$ Unknown parameters
- Several SP applications require the covariance matrix estimation, e.g. sources localization, STAP, Polarimetric SAR classification, radar detection, MIMO, discriminant analysis, dimension reduction, PCA...
- The ultimate purpose is to characterize the system performance, not only the estimation performance $\Rightarrow$ ROC curves, PD vs SNR, PFA, MSE ...

Motivations ...

Robustness: what happens when models turn to be not Gaussian anymore?

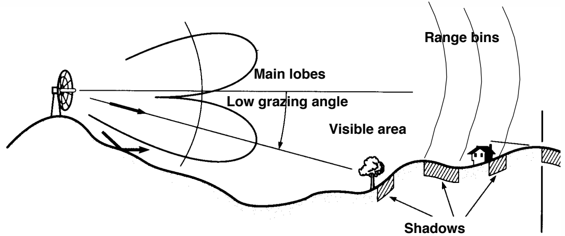
- Gaussian model $\Rightarrow$ Sample Covariance Matrix
- Outliers and other parasites
- Mismodelling
- Missing data

High dimensional problems

- Massive data
- Data size can be important...
- ... greater than the number of observations
- Link with robustness.

Why non Gaussian modeling? (e.g. 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

Radar signals amplitude - Gaussian or not?

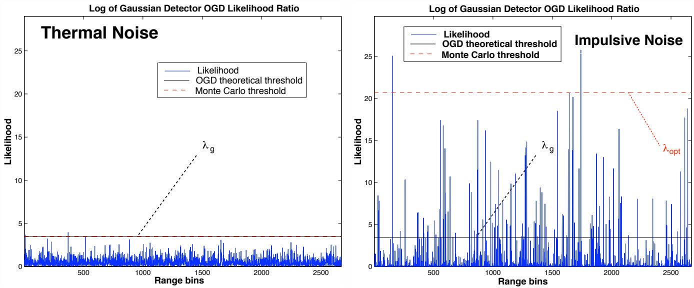


Figure: Failure of the OGD - Adjustment of the detection threshold - K-distributed clutter with same power as the Gaussian noise

- $\Rightarrow$ Bad performance of the OGD in case of mismodeling
- $\Rightarrow$ Introduction of elliptical distributions
- $\Rightarrow$ Introduction of robust estimates

Motivations

Why CES and M-estimation? Examples in Image processing

Polarimetric SAR image (RGB)
3-dimensional complex pixels

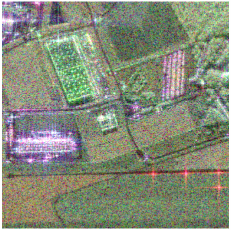


Figure: Brétigny area - RAMSES system (ONERA) - X-band - Resolution: 1.32m$\times$ 1.38m

Hyperspectral image
100-dimensional complex pixels

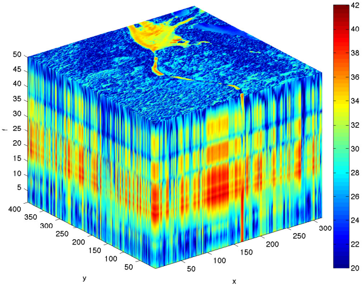


Figure: Indian Pines - $m = 100$ wavelengths

Some insights

* **Robust Estimation Theory**
  - More flexible and adjustable models ⇝ *CES distributions*
  - Robust family of estimators ⇝ M*-estimators*
  - Regularized estimators

- M-estimators statistical properties (complex case)
- Statistical properties of M-estimators functionals (e.g. MUSIC statistic for DoA estimation, ANMF detectors...)
- Regularized Tyler Estimator (RTE) derivation and asymptotics

Some insights

$\ast$ (Robust) Random Matrix Theory

In many applications, the dimension of the observation m is large (HSI...)

$\Rightarrow$ The required number N of observations for estimation purposes needs to be larger: $N \gg m$ BUT this is not the case in practice! Even $N < m$ is possible

$\rightsquigarrow$ New asymptotic regime: $N \to \infty, m \to \infty$ and $\dfrac{m}{N} \to c \in [0, 1]$

- Extension of "standards" for M-estimators for particular case and for general CES distribution.
- Asymptotic distribution of the eigenvalues
- Asymptotics for the RTE
- Application to DoA estimation: robust G-MUSIC

Connections between Robust Estimation Theory and RMT

Key references of this talk

Robust Estimation Theory

- E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *Signal Processing, IEEE Transactions on*, vol. 60, pp. 5597-5625, nov. 2012.

- F. Pascal, Y. Chitour and Y. Quek, "Generalized Robust Shrinkage Estimator and Its Application to STAP Detection Problem," *Signal Processing, IEEE Transactions on*, vol. 62, pp. 5640-5651, nov. 2014.

- A. Kammoun, R. Couillet, F. Pascal and M-S. Alouini, "Convergence and fluctuations of Regularized Tyler estimators," *Signal Processing, IEEE Transactions on*, vol. 64, pp. 1048-1060, feb. 2016.

Key references of this talk

(Robust) Random Matrix Theory

- R. Couillet, F. Pascal and J. W. Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples", *Journal of Multivariate Analysis*, vol. 139, pp. 56-78, 2015.
- (R. Couillet, F. Pascal and J. W. Silverstein, "Robust Estimates of Covariance Matrices in the Large Dimensional Regime," Information Theory, IEEE Transactions on, vol. 60, pp. 7269-7278, nov 2014.)
- R. Couillet, A. Kammoun, and F. Pascal, "Second order statistics of robust estimators of scatter. Application to GLRT detection for elliptical signals," *Journal of Multivariate Analysis*, vol.143, pp. 249-274 2016.
- A. Kammoun, R. Couillet, F. Pascal, and M.-S. Alouini, "Optimal Design of the Adaptive Normalized Matched Filter Detector," *submitted*, 2016. arXiv:1501.06027

Modeling the background

Complex elliptically symmetric (CES) distributions

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

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

where $h_z : [0,\infty) \to [0,\infty)$ 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

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

Characterizing property

- Unit complex m-sphere:

$$\boxed{\mathbb{C}S^m \triangleq \{\mathbf{z} \in \mathbb{C}^m \mid \|\mathbf{z}\| = 1\}}$$

- $\mathbf{u}$ (or $\mathbf{u}^{(m)}$) = r. v. with uniform distribution on $\mathbb{C}S^m$,

$$\boxed{\mathbf{u} \sim \mathcal{U}\left(\mathbb{C}S^m\right)}$$

Theorem (Stochastic representation theorem)

$\mathbf{z} \sim CE(\boldsymbol{\mu}, \boldsymbol{\Lambda}, h_{\mathbf{z}})$ *if and only if it admits the stochastic representation*

$$\mathbf{z} =_d \boldsymbol{\mu} + \mathcal{R}\mathbf{A}\mathbf{u}^{(k)}$$

where r. va. $\mathcal{R} \geq 0$, called the modular variate, is independent of $\mathbf{u}^{(k)}$ and $\boldsymbol{\Lambda} = \mathbf{A}\mathbf{A}^H$ is a factorization of $\boldsymbol{\Lambda}$, where $\mathbf{A} \in \mathbb{C}^{m \times k}$ with $k =$ rank(Λ).

Characterizing property

1. One-to-one relation with c.d.f $F_{\mathcal{R}}(.)$ of $\mathcal{R}$ and characteristic generator $\varphi(.)$

2. Ambiguity: both $(\mathcal{R}, \mathbf{A})$ and $(c^{-1}\mathcal{R}, c\mathbf{A}), c > 0$ are valid stochastic representations of $\mathbf{z}$ $\Rightarrow$ constraint for identifiability issues

3. Distribution of quadratic form: if rank$(\mathbf{\Lambda}) = m$, then

$$Q(\mathbf{z}) \triangleq (\mathbf{z} - \boldsymbol{\mu})^H \, \mathbf{\Lambda}^{-1} \, (\mathbf{z} - \boldsymbol{\mu}) =_d \mathcal{Q}$$

where $\mathcal{Q} \triangleq \mathcal{R}^2$ is called the 2^{nd}-order modular variate.

Characterizing property

4 Random number generation

- draw a random modular variate $\mathcal{R}$ from a distribution $F_{\mathcal{R}}(.)$ and $\mathbf{u}^{(k)}$ from $\mathcal{U}\left(\mathbb{C}S^m\right)$.
- set $\mathbf{z} =_d \boldsymbol{\mu} + \mathcal{R}\mathbf{A}\mathbf{u}^{(k)}$ for a given $\boldsymbol{\mu} \in \mathbb{C}^m$ and $\mathbf{A} \in \mathbb{C}^{m \times k}$.

5 Applications

- In practice, $\mathcal{R}$ accounts for the amplitude fluctuations from one observation to another (cf. example of low grazing angle).
- Can model heavy tailed-distribution (e.g. for image processing).

Examples

Compound Gaussian (CG) distributions

An important subclass of CES distributions, also called as

- Spherically Invariant Random Vectors (SIRV) [Yao, 1973]
- Scale mixture of normal distributions [Andrews and Mallows, 1974]

Compound Gaussian distributions

$\mathbf{z} \sim CG_m(\boldsymbol{\mu}, \boldsymbol{\Lambda}, F_\tau)$ if it admits a stochastic CG-representation

$$\mathbf{z} =_d \boldsymbol{\mu} + \sqrt{\tau}\mathbf{n}$$

where r. va. $\tau \geq 0$, with c.d.f. F_τ, called the texture, is independent of $\mathbf{n}$, with $\mathbf{n} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$, called the speckle.

Some remarks

Comments:

- Of course, it is a subclass of the CES!
- $\mathcal{R} = \sqrt{s}$ with $s \sim Gam(1,k)$, one has $\mathbf{n}_0 =_d \sqrt{s}\mathbf{u}^{(k)}$, with $\mathbf{n}_0 \sim \mathbb{CN}(\mathbf{0}, \mathbf{I})$.

1. Covariance matrix exists if $\mathbb{E}[\tau] < +\infty$,

$$\begin{aligned} \mathbf{M} &= \text{cov}\left(\sqrt{\tau}\mathbf{n}\right) = \sqrt{\tau}\text{cov}(\mathbf{n}) \\ &= \sigma_{\mathbf{M}}\mathbf{\Lambda} \text{ with } \sigma_{\mathbf{M}} = \mathbb{E}[\tau] \end{aligned}$$

2. Identifiability: Both $(\sqrt{\tau}, \mathbf{n})$ and $(a\sqrt{\tau}, \mathbf{n}/a), \forall a > 0$ leads to same CG dist. for $\mathbf{z}$

For proper identifiability, on has to impose a scale constraint:

- τ, e.g. $\mathbb{E}[\tau] = 1$
- $\mathbf{\Lambda}$, e.g. $\text{Tr}(\mathbf{\Lambda}) = m$

Examples

Student t-distribution and K-distribution

t-distribution with dof ν

- CG: $\tau^{-1} \sim Gam(\nu/2, 2/\nu)$, where $\nu > 0$
- CES: $(1/m)\,\mathcal{R}^2 \sim F_{2m,\nu}$ where $F_{a,b}$ is the F-dist with dof a and b

- $\nu = 1 \implies$ complex Cauchy dist.
- $\nu \to \infty \implies$ CN dist.
- finite 2nd-order moment for $\nu > 2$

K-distribution with shape parameter ν

- CG: $\tau \sim Gam(\nu, 1/\nu)$, where $\nu > 0$
- CES: closed-form PDF for $\mathcal{Q}$

- $\nu \downarrow \implies$ heavy-tailed dist.
- $\nu \to \infty \implies$ CN dist.
- $\mathbb{E}[\tau] = 1 \implies \mathbf{\Lambda} = \mathbf{M}$

Examples

Generalized Gaussian distribution

GG distribution with parameters s and η

- CES: $\mathcal{R}^2 =_d G^{1/s}$ where $G \sim Gam(m/s, \eta), s, \eta > 0$
- PDF: $f_{\mathbf{z}}(\mathbf{z}) = cte\,|\boldsymbol{\Lambda}|^{-1}\,\exp\left(-(\eta\,\mathbf{z}^H\boldsymbol{\Lambda}^{-1}\mathbf{z})^s\right)$

- Complex analog of the exponential power family, also called Box-Tiao distributions
- Subclass of multivariate symmetric Kotz-type distributions
- Case $s = 1 \Longrightarrow$ CN dist. Heavier tailed than normal for $s < 1$ and lighter tailed for $s > 1$
- $s = 1/2 \Longrightarrow$ generalization of Laplace dist.

Image analysis with GG distributions

The following results can be found in:

Image processing:

- F. Pascal , L. Bombrun, J.-Y. Tourneret and Y. Berthoumieu, "Parameter Estimation for Multivariate Generalized Gaussian Distributions" *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5960-5971, 2013.

Images are filtered by a stationary wavelet $\Rightarrow$ observed vector $\mathbf{z}$ contains the realizations of the wavelet coefficients for each channel of the RGB image.

Image analysis with GG distributions

(a)

(b)

Figure: Images from the VisTex database. (a) Bark.0000 and (b) Leaves.0008.

Image analysis with GG distributions

Table: Estimated MGGD parameters for the first subband of the Bark.0000 and Leaves.0008 images.

Image	$\hat{\eta}$	$\hat{s}$	$\hat{\boldsymbol{\Lambda}}$
Bark 0000	0.036	0.328	$\begin{bmatrix} 0.988 & 0.992 & 0.883 \\ 0.992 & 1.131 & 0.922 \\ 0.883 & 0.922 & 0.881 \end{bmatrix}$
Leaves 0008	0.054	0.265	$\begin{bmatrix} 0.935 & 0.966 & 0.871 \\ 0.966 & 1.074 & 0.976 \\ 0.871 & 0.976 & 0.991 \end{bmatrix}$

Image analysis with GG distributions

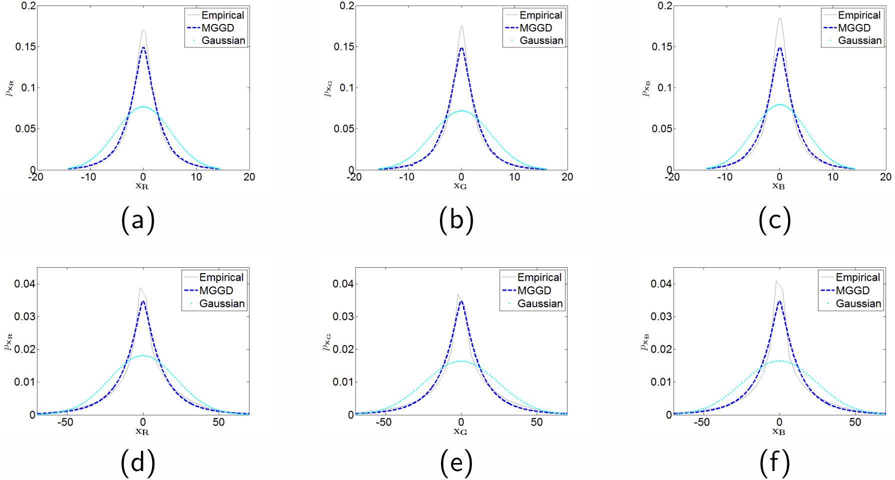


Figure: Marginal distributions of the wavelet coefficients with the estimated MGGD and Gaussian distributions of the first subband for the red, green and blue channels of the Bark.0000 (a,b,c) and Leaves.0008 images (d,e,f).

Estimating the covariance matrix

ML-estimators

CES PDF specified
$\Rightarrow$ *PDF of* $\mathcal{R}$
$\Rightarrow$ *PDF of* τ

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

- $\widehat{\mathbf{\Lambda}}$ that minimizes the negative log-likelihood function

$$L_n(\mathbf{\Lambda}) = N \ln |\mathbf{\Lambda}| - \sum_{n=1}^{N} \ln g_{\mathbf{z}}(\mathbf{z}_n^H \mathbf{\Lambda}^{-1} \mathbf{z}_n)$$

- Solution to the estimating equation

$$\widehat{\mathbf{\Lambda}} = \frac{1}{N} \sum_{n=1}^{N} \varphi(\mathbf{z}_n^H \widehat{\mathbf{\Lambda}}^{-1} \mathbf{z}_n) \mathbf{z}_n \mathbf{z}_n^H$$

with $\varphi(t) = -g'_{\mathbf{z}}(t)/g_{\mathbf{z}}(t)$.

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...

under a set of assumptions

From Maronna (1976) (real case),

Hypotheses: (case where $\boldsymbol{\mu} = \mathbf{0}$).

Let $\psi(s) = su(s)$ and $K = \sup_{s \geq 0} \psi(s)$.

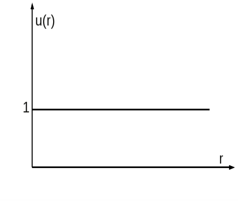
(H1) u is nonnegative, nonincreasing and continuous on $[0, \infty)$.

(H2) $m < K < \infty$ and ψ nondecreasing and strictly increasing in the interval where $\psi < K$.

(H3) Let $P_N(.)$ the empirical distribution of $(\mathbf{z}_1, ..., \mathbf{z}_N)$. It exists $a > 0$ s.t. for each hyperplan H, $\dim(H) \leq m - 1$, $P_N(H) \leq 1 - \frac{m}{K} - a$. (This assumption can be strongly relaxed.)

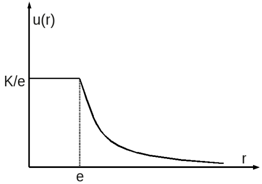
Examples of M-estimators

SCM:
$u(r) = 1$

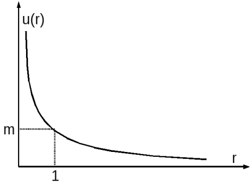


Huber's estimator (M-estimator):

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



Tyler Estimator (Tyler, 1987; Pascal, 2008):
$u(r) = \frac{m}{r}$



Remarks:

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

Tyler Estimator:

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

Estimating the covariance matrix

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 $\mathbf{V}$.

For any functional u (ensuring existence and uniqueness), $\mathbf{V}_N$ is a consistent estimate of both $\mathbf{M}$ and $\mathbf{\Lambda}$ (up to scale factors)

Asymptotic distribution of complex SCM

The SCM (MLE for normal dist.) is defined as

$$\widehat{\mathbf{S}}_N = \frac{1}{N} \sum_{n=1}^{N} \mathbf{z}_n \mathbf{z}_n^H$$

where $\mathbf{z}_n$ are complex independent circular zero-mean Gaussian with covariance matrix $\mathbf{V}$.

$$\sqrt{N}\, \text{vec}(\widehat{\mathbf{S}}_N - \mathbf{V}) \xrightarrow{d} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_W, \mathbf{\Omega}_W\right)$$

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

Remarks:

- Valid for Wishart distribution (up to scale $1/N$)!
- Wishart distribution $\Longrightarrow$ many interesting properties

Asymptotic distribution of complex M-estimators

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

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

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

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

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

with

$$\left\{\begin{array}{l}
\sigma_1 = a_1(m+1)^2(a_2+m)^{-2},\\
\sigma_2 = a_2^{-2}\left\{(a_1-1) - 2a_1(a_2-1)\left[2m+(2m+4)a_2\right](2a_2+2m)^{-2}\right\},
\end{array}\right.$$

and

$$\left\{\begin{array}{l}
a_1 = \left[m(m+1)\right]^{-1} E\left[\psi^2(\sigma|\mathbf{t}|^2)\right],\\
a_2 = m^{-1}\, E[\sigma|\mathbf{t}|^2\psi'(\sigma|\mathbf{t}|^2)],
\end{array}\right.$$

where σ is the solution of $E[\psi(\sigma|\mathbf{t}|^2)] = m$, where $\mathbf{t} \sim CE(\mathbf{0}, \mathbf{I}, g_{\mathbf{z}})$ and $\psi(x) = xu(x)$.

Asymptotic distribution of Tyler estimator

Tyler estimator is obtained for $u(t) = m/t$

$$\widehat{\mathbf{M}}_{FP} = \frac{m}{N} \sum_{n=1}^{N} \frac{\mathbf{z}_n \mathbf{z}_n^H}{\mathbf{z}_n^H \widehat{\mathbf{M}}_{FP}^{-1} \mathbf{z}_n}$$

Theorem 1 (Asymptotic distribution of $\widehat{\mathbf{M}}_{FP}$)

$$\sqrt{N}\ \mathrm{vec}(\widehat{\mathbf{M}}_{FP} - \mathbf{V}) \xrightarrow{d} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{FP}, \boldsymbol{\Omega}_{FP}\right),$$

where $\mathbb{C}\mathcal{N}$ is the complex Gaussian distribution, $\boldsymbol{\Sigma}_{FP}$ the CM and $\boldsymbol{\Omega}_{FP}$ the pseudo CM:

$$\boldsymbol{\Sigma}_{FP} = \tfrac{m+1}{m}(\mathbf{V}^T \otimes \mathbf{V}) - \tfrac{m+1}{m^2}\mathrm{vec}(\mathbf{V})\mathrm{vec}(\mathbf{V})^H,$$

$$\boldsymbol{\Omega}_{FP} = \tfrac{m+1}{m}(\mathbf{V}^T \otimes \mathbf{V})\,\mathbf{K} - \tfrac{m+1}{m^2}\mathrm{vec}(\mathbf{V})\mathrm{vec}(\mathbf{V})^T,$$

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

Reminders and comments

- General family of distributions $\Longrightarrow$ Model many real phenomenon
- General family of CM estimators related OR NOT to the underlying distribution
- Common asymptotic distributions: differs from scale factors $\Longrightarrow$ highlight the tradeoff efficiency-robustness!
    - mis-modelling
    - non-Gaussian model
    - smaller scale factor (1 and 0) for the SCM

How to use this framework in practice ?

An important property of complex M-estimators

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

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

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	Tyler
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 CM estimators

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

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

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

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

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

Some comments:

Perfect (but asymptotic) characterization of several objects properties, such as detectors, classifiers, estimators...

$H(SCM)$, $H(M$-estimators) and $H(FP)$ share the same asymptotic distribution (differs from σ_1 or $\frac{m+1}{m}$)

$$\Downarrow$$

- Link to the classical Gaussian case
- Quantification of the loss involved by robust estimator

Application: target detection

Problem statement

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

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

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

  $\Rightarrow$ Neyman-Pearson criterion

Detection: generalities

- Detection test: comparison between the Likelihood Ratio $\Lambda(\mathbf{y})$ and a detection threshold λ :

$$\Lambda(\mathbf{y}) = \frac{p_{\mathbf{y}}(\mathbf{y}/H_1)}{p_{\mathbf{y}}(\mathbf{y}/H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \,,$$

λ is obtained for a given PFA (set by the user):

- Probability of False Alarm (type-I error):

$$P_{fa} = \mathbb{P}(\Lambda(\mathbf{y}) > \lambda/H_0)$$

- Probability of Detection (to evaluate the performance):

$$P_D = \mathbb{P}(\Lambda(\mathbf{y}) > \lambda/H_1)$$

for different Signal-to-Noise Ration (SNR).

Detection under Gaussian/non-Gaussian assumption

- Gaussian case (OGD): if $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{M})$ then

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

with $\lambda_g = \sqrt{-\ln(PFA)}$ and $p_{\mathbf{z}}(\mathbf{z}) = \frac{1}{(\pi)^m |\mathbf{M}|} \exp\left(-\mathbf{z}^H \mathbf{M}^{-1} \mathbf{z}\right)$.

- Heterogeneous case (NMF):

$$\Lambda(\mathbf{y}) = \frac{|\mathbf{p}^H \mathbf{M}^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \mathbf{M}^{-1} \mathbf{p})(\mathbf{y}^H \mathbf{M}^{-1} \mathbf{y})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{NMF}$$

The False Alarm regulation can be theoretically done thanks to

$$\lambda_{NMF} = 1 - PFA^{\frac{1}{m-1}}.$$

This comes from a Beta distribution of the test.

Generalities: $\mathbb{M}$ unknown $\Rightarrow$ Adaptive detection

Gaussian model $\Rightarrow \widehat{\mathbf{S}}_N = \frac{1}{N} \sum_{n=1}^{N} \mathbf{z}_n \mathbf{z}_n^H$

- AMF test [1]

$$\Lambda_{AMF}(\mathbf{y}) = \frac{\left| \mathbf{p}^H \, \widehat{\mathbf{S}}_N^{-1} \, \mathbf{y} \right|^2}{\left(\mathbf{p}^H \, \widehat{\mathbf{S}}_N^{-1} \, \mathbf{p} \right)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{AMF} \,. \tag{5}$$

[1] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector", *Aerospace and Electronic Systems, IEEE Transactions on,* vol. 28, no. 1, pp. 208-216, 1992.

- Kelly test [2]

$$\Lambda_{Kelly}(\mathbf{y}) = \frac{\left| \mathbf{p}^H \, \widehat{\mathbf{S}}_N^{-1} \, \mathbf{y} \right|^2}{\left(\mathbf{p}^H \, \widehat{\mathbf{S}}_N^{-1} \, \mathbf{p} \right) \, \left(N + \mathbf{y}^H \, \widehat{\mathbf{S}}_N^{-1} \, \mathbf{y} \right)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{Kelly} \,. \tag{6}$$

[2] E. J. Kelly, "An adaptive detection algorithm", *Aerospace and Electronic Systems, IEEE Transactions on,* pp. 115-127, November 1986.

CES distribution $\Rightarrow$ ANMF

ANMF test (ACE, GLRT-LQ) [3,4]

$$\Lambda_{ANMF}(\mathbf{y}, \widehat{\mathbf{M}}) = \frac{|\mathbf{p}^H \widehat{\mathbf{M}}^{-1} \mathbf{y}|^2}{(\mathbf{p}^H \widehat{\mathbf{M}}^{-1} \mathbf{p})(\mathbf{y}^H \widehat{\mathbf{M}}^{-1} \mathbf{y})} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{ANMF} \tag{7}$$

where $\widehat{\mathbf{M}}$ stands for any estimators presented before: SCM, M-estimators, Tyler estimator.

One has, **conditionally to $\mathbf{y}$!!**, $\Lambda_{ANMF}(\widehat{\mathbf{M}}) = \Lambda_{ANMF}(\alpha \, \widehat{\mathbf{M}})$ for any $\alpha > 0$.

[3] E. Conte, M. Lops, and G. Ricci, "Asymptotically Optimum Radar Detection in Compound-Gaussian Clutter", *Aerospace and Electronic Systems, IEEE Transactions on,*, vol. 31, pp. 617-625, April 1995.

[4] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT", *Signal Processing, IEEE Transactions on,* vol. 47, no. 9, pp. 2538-2541, 1999.

Properties

- The ANMF is scale-invariant, i.e.
  $\forall \alpha, \beta \in \mathbb{R}\,,\ \Lambda_{ANMF}(\alpha\,\mathbf{y}, \beta\,\widehat{\mathbf{M}}) = \Lambda_{ANMF}(\mathbf{y}, \widehat{\mathbf{M}})$
- Its asymptotic distribution (**conditionally to $\mathbf{y}$**) is known (thanks to theorem 2)

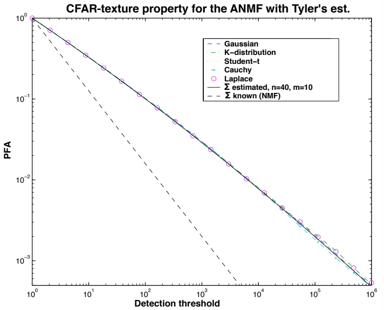
Considering $\Lambda_{ANMF}(\mathbf{y}, \widehat{\mathbf{M}})$ **conditionally to $\mathbf{y}$**, i.e. $\Lambda_{ANMF}(\widehat{\mathbf{M}})$, allows to directly apply theorem 2. Else see next slide!

- It is CFAR w.r.t the covariance/scatter matrix, i.e. its distribution does not depend on the covariance/scatter matrix
- It is CFAR w.r.t the texture (if considering Compound-Gaussian model)

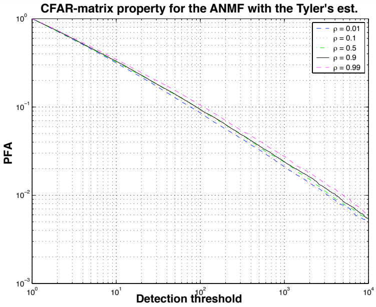
CFAR = Constant False Alarm Rate

Illustration of the CFAR properties

False Alarm regulation



(a) CFAR-texture



(b) CFAR-matrix

Figure: Illustration of the CFAR properties of the ANMF built with the Tyler's estimator, for a Toeplitz CM whose (i, j)-entries are $\rho^{|i-j|}$

Probability of false alarm

PFA-threshold relation of $\Lambda_{ANMF}(\widehat{\mathbf{S}}_N)$ (Gaussian case, finite N)

Let $(\mathbf{z}_1, ..., \mathbf{z}_N)$ be a N-sample $\sim \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{M})$ with dimension m

$$P_{fa} = \mathbb{P}(\Lambda_{ANMF}(\mathbf{y}, \widehat{\mathbf{S}}_N) > \lambda / H_0) = (1-\lambda)^{a-1} \, {}_2F_1(a, a-1; b-1; \lambda) \tag{8}$$

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

$${}_2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)}{\Gamma(c+k)} \frac{x^k}{k!}$$

[5] F. Pascal, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Constant false alarm rate detection in spherically invariant random processes," in *Proc. of the European Signal Processing Conf., EUSIPCO-04*, (Vienna), pp. 2143-2146, Sept. 2004.

Comments

Three possible approaches to characterize the performance:

1. Use the (very) poor approximation of the FA regulation of the NMF
2. Use the asymptotics of theorem 2 (but it is conditionally to the dist. of $\mathbf{y}$!) $\Rightarrow$ a slight loss of performance
3. Combine the asymptotics of theorem 2 and the finite-distance result on PFA-threshold...

From theorem 2, one has

PFA-threshold relation of Λ_{ANMF}(M-est.) for CES distributions

For N large enough and for any CES distributed noise, the PFA is still given by (8) if we replace N by N/ν_1.

The third one seems to provide more accurate results...

Probabilities of false alarm for robust detection

From theorem 2, one has both results (Λ_{ANMF} denoted now Λ)

P_{fa}-threshold relation of $\Lambda(M$-est$)$ for CES distributions

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

where $a = N/\nu_1 - m + 2$, $b = N/\nu_1 + 2$ and ${}_2F_1$ is the Hypergeometric function.

Here, something is missing...

P_{fa}-threshold relation of $\Lambda(M$-est$)$ for CES distributions (conditionally to the dist. of $\mathbf{y}$)

$$\sqrt{N}\left(\Lambda(\mathbf{V}_N) - \Lambda(\mathbf{V})\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, 2\,\nu_1 \Lambda(\mathbf{V})\left(\Lambda(\mathbf{V}) - 1\right)^2\right)$$

Then, integrate (numerically) over the dist. of $\mathbf{y}$, see [6] for details

[6] F. Pascal and J.-P. Ovarlez, "Asymptotic Properties of the Robust ANMF," in *Proc. of ICASSP-15*, (Brisbane, Australia), Apr. 2015.

What is missing...

Theorem (Asymptotic distribution of $\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{S}}_N$)

$$\sqrt{N}\mathrm{vec}(\widehat{\mathbf{M}}_{FP} - \widehat{\mathbf{S}}_N) \xrightarrow{d} \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}, \Omega)$$

where $\mathbf{\Sigma}$ and Ω are defined by

$$\mathbf{\Sigma} = \frac{1}{m}\mathbf{M}^T \otimes \mathbf{M} + \frac{m-1}{m^2}\mathrm{vec}(\mathbf{M})\mathrm{vec}(\mathbf{M})^H,$$
$$\Omega = \frac{1}{m}(\mathbf{M}^T \otimes \mathbf{M})\mathbf{K} + \frac{m-1}{m^2}\mathrm{vec}(\mathbf{M})\mathrm{vec}(\mathbf{M})^T.$$

Remark:

- $\mathbf{\Sigma} < \mathbf{\Sigma}_{FP}$ and $\Omega < \Omega_{FP}$
- $m \to \infty \implies 1/m \to 0 \implies \mathbf{\Sigma} \ll \mathbf{\Sigma}_{FP}$ and $\Omega \ll \Omega_{FP}$

Result for the ANMF test

Theorem (Asymptotic distribution of $\Lambda(\widehat{\mathbf{M}}_{FP}) - \Lambda(\widehat{\mathbf{S}}_N)$, conditionally to the dist. of $\mathbf{y}$

$$\sqrt{N}(\Lambda(\widehat{\mathbf{M}}_{FP}) - \Lambda(\widehat{\mathbf{S}}_N))_{\mathbf{y}} \xrightarrow{d} \mathcal{N}(0, \mathbf{\Sigma}_T)$$

where $\mathbf{\Sigma}_T$ is defined by

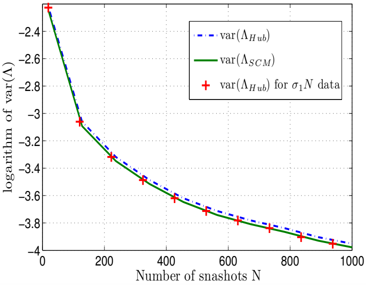
$$\mathbf{\Sigma}_T = \frac{2}{m}\Lambda(\mathbf{M})(\Lambda(\mathbf{M}) - 1)^2.$$

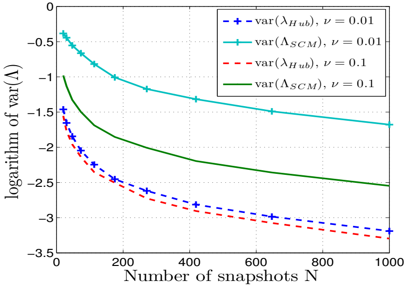
Remark: $\mathbf{\Sigma}_T < \mathbf{\Sigma}_H$ and $m \to \infty \implies 1/m \to 0 \implies \mathbf{\Sigma}_T \ll \mathbf{\Sigma}_H$

This theoretically justifies the use of relation (8)

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).



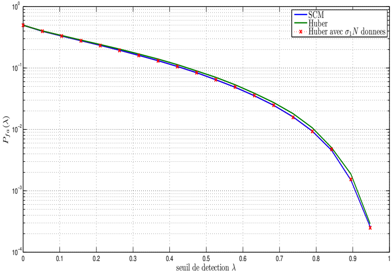


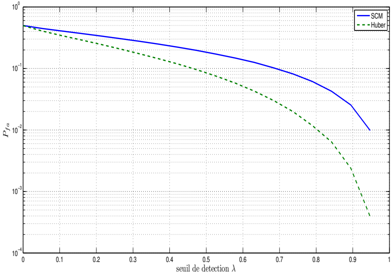
Thm validation (even for small N)

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).





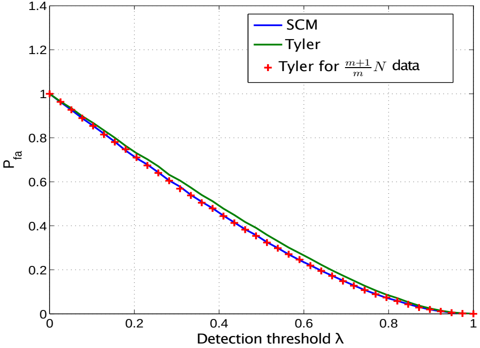
Validation of theorem (even for small N)

Interest of the M-estimators for False Alarm regulation

Tyler's estimator: Gaussian context, $n = 10$, $m = 3$

PFA-threshold relation of Λ_{ANMF}(Tyler's est.) for CES distributions

For n large and any elliptically distributed noise, the PFA is still given by (8) if we replace N by $N/\frac{m+1}{m}$.



MUltiple Signal Classification (MUSIC) method for DoA estimation

- K (known) direction of arrival θ_k on m antennas
- Gaussian stationary narrowband signal with additive noise.
- the DoA is estimated from N snapshots, (SCM, M- and Tyler estimator).

$$\mathbf{z}_t = \sum_{k=1}^{K} \sqrt{p}_k \mathbf{s}(\theta_k) y_{k,t} + \sigma \mathbf{w}_t = \mathbf{A}(\boldsymbol{\theta}) \mathbf{y}_t + \sigma \mathbf{w}_t$$

- $\boldsymbol{\theta} = (\theta_1, \theta_2, ...\theta_K)^T$,
- $\mathbf{A}(\boldsymbol{\theta}) = (\sqrt{p}_1 \mathbf{s}(\theta_1), \sqrt{p}_2 \mathbf{s}(\theta_2), ..., \sqrt{p}_K \mathbf{s}(\theta_K))$ is the steering matrix
- $\mathbf{y}_t = \begin{pmatrix} y_{1,t} & y_{2,t} & \dots & y_{K,t} \end{pmatrix}^T$ the signal vector,
- $\mathbf{w}_t$ stationary additive noise.

$$\mathbf{M} = \mathbb{E}[\mathbf{z}\mathbf{z}^H] = \mathbf{A}(\boldsymbol{\theta})\mathbb{E}[\mathbf{y}\mathbf{y}^H]\mathbf{A}(\boldsymbol{\theta})^H + \sigma^2\mathbf{I}$$

which can be rewritten

$$\mathbf{M} = \mathbb{E}[\mathbf{z}\mathbf{z}^H] = \mathbf{E}_S\mathbf{D}_S\mathbf{E}_S^H + \sigma^2\mathbf{E}_W\mathbf{E}_W^H.$$

where $\mathbf{E}_S$ (resp. $\mathbf{E}_W$) are the signal (resp. noise) subspace eigenvectors. The MUSIC statistic is

$$\begin{cases} H(\mathbf{M}) & = \gamma(\theta) = \mathbf{s}(\theta)^H\mathbf{E}_W\mathbf{E}_W^H\mathbf{s}(\theta), & (\mathbf{M}\ \text{known}) \\ H(\widehat{\mathbf{M}}) & = \hat{\gamma}(\theta) = \displaystyle\sum_{i=1}^{m-K} \lambda_i\mathbf{s}(\theta)^H\hat{\mathbf{e}}_i\hat{\mathbf{e}}_i^H\mathbf{s}(\theta) = H(\alpha\,\widehat{\mathbf{M}}), & (\mathbf{M}\ \text{unknown}) \end{cases}$$

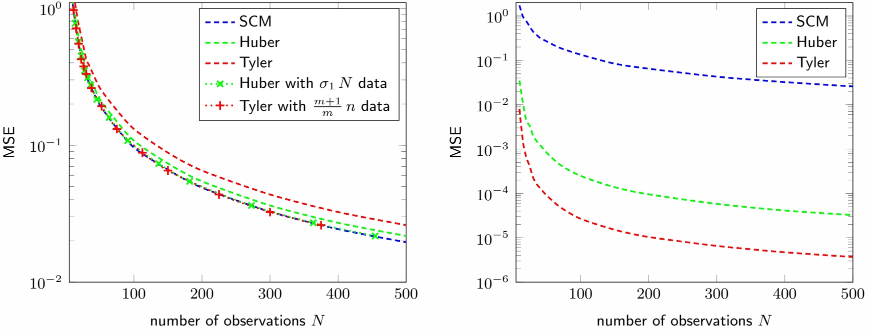
where λ_i (resp. $\hat{\mathbf{e}}_i$) are the eigenvalues (resp. eigenvectors) of $\widehat{\mathbf{M}}$.

This function respects assumptions of theorem 2

Simulation using the MUltiple Signal Classification (MUSIC) method

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

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

Similar conclusions as for detection can be drawn...

Interest of RMT: A very simple example...

Problem: Estimation of 1 DoA embedded in white Gaussian noise

$$\mathbf{z}_t = \sqrt{p}\mathbf{s}(\theta)\mathbf{y}_t + \mathbf{w}_t$$

where the $\mathbf{w}_t$'s are N independent realizations of circular white Gaussian noise, i.e. $\mathbf{w}_t \sim \mathbb{C}\mathcal{N}(0, \mathbf{I})$.

Classical approach

- $\widehat{\mathbf{S}}_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} \mathbf{w}_t \mathbf{w}_t^H \underset{N\to\infty}{\longrightarrow} \mathbf{I}$
- Then, MUSIC algorithm allows to estimate the DoA...

What happens when the dimension m is large?

- $\widehat{\mathbf{S}}_N = \dfrac{1}{N}\displaystyle\sum_{t=1}^{N} \mathbf{w}_t \mathbf{w}_t^H \underset{m,N\to\infty}{\not\longrightarrow} \mathbf{I}$
- Then, MUSIC algorithm IS NOT the best way to estimate the DoA...

Classical approach: $N \gg m$

e.g. STAP context, 4 sensors and 64 pulses, $m = 256$ and $N = 10^4$

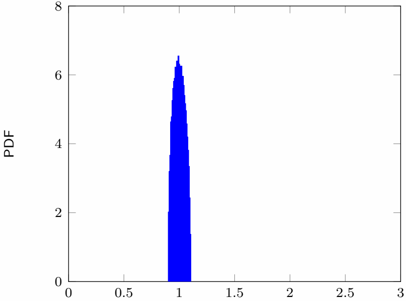


Figure: Empirical distribution for the eigenvalues of the SCM in the case of a white Gaussian noise of dimension $m = 256$ for $N = 10^4$ secondary data

What happens when the dimension m is large? (compared to N)

STAP context, 4 sensors and 64 pulses, $m = 256$ and $N = 10^3$

Marcenko-Pastur Law...

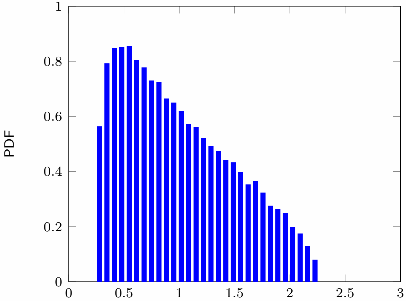


Figure: Empirical distribution for the eigenvalues of the SCM in the case of a white Gaussian noise of dimension $m = 256$ for $N = 10^3$ secondary data

What happens when the dimension m is large? (compared to N)

STAP context, 4 sensors and 64 pulses, $m = 256$ and $N = 500$

Marcenko-Pastur Law...

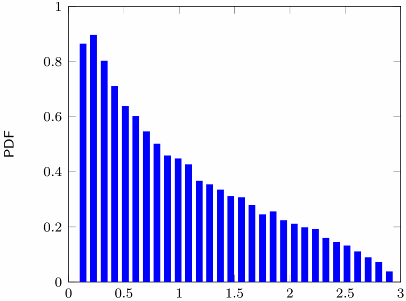
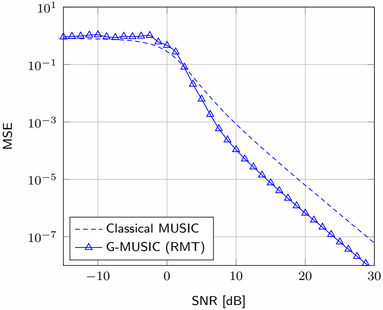


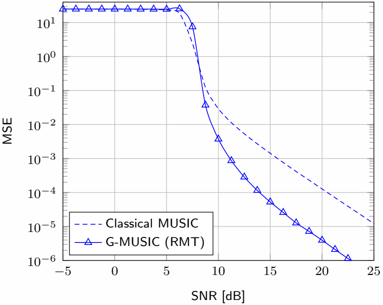
Figure: Empirical distribution for the eigenvalues of the SCM in the case of a white Gaussian noise of dimension $m = 256$ for $N = 500$ secondary data

Consequences

Bad assumptions $\Longrightarrow$ Bad performance



(a) 50 data of size 10



(b) 250 data of size 50

Figure: MSE on the different DoA estimators for $K = 1$ source embedded in an additive white Gaussian noise

RMT - Classical results

Assumptions:

- $N, m \to \infty$ and $\dfrac{m}{N} \to c \in (0,1)$ and $\widehat{\mathbf{S}}_N = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i^H$ the SCM
- $(\mathbf{z}_1, ..., \mathbf{z}_N)$ be a N-sample, i.i.d (i.e. $E[\mathbf{z}_i^{(j)}\,\mathbf{z}_k^{(l)}] = 0$) with finite fourth-order moment.

Remark: CES dist. do not respect this assumptions!

Thus one has:

1) $F^{\widehat{\mathbf{S}}_N} \Rightarrow F^{MP}$

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

The MP PDF is defined by

$$\mu(x) = \begin{cases} \left(1 - \frac{1}{c}\right) \mathbf{1}_{x=0} + f(x) & \text{if } c > 1 \\ f(x) & \text{if } c \in (0,1] \end{cases}$$

with $f(x) = \dfrac{1}{2\pi\sigma^2} \dfrac{\sqrt{(c_+ - x)(x - c_-)}}{cx} \mathbf{1}_{x \in [c_-, c_+]}$ and $c_\pm = \sigma^2\,(1 \pm \sqrt{c})^2$.

RMT - Classical results

Exploiting the MP dist for the SCM eigenvalues leads to a new MUSIC statistic:

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

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

with $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_m$ (resp. $\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_m$ the eigenvalues (resp. the eigenvectors) of $\widehat{\mathbf{S}}_N$ and $\hat{\mu}_1 \leq \ldots \leq \hat{\mu}_m$ the eigenvalues of $\mathrm{diag}(\hat{\boldsymbol{\lambda}}) - \frac{1}{m}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^T$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m)^T$.

Remark: Contrary to MUSIC or Robust-MUSIC, all the eigenvectors are used to compute G-MUSIC.

Robust RMT

Assumptions:

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

Thus, it is shown that:

1) There exists a unique solution to the M-estimator fixed-point equation for all large m a.s. The recursive algorithm associated converges to this solution.

2) $\|\phi^{-1}(1)\,\mathbf{V}_N - \widehat{\mathbf{S}}_N\| \xrightarrow{a.s.} 0$ when $N, m \to \infty$ and $\dfrac{m}{N} \to c$
where $\|.\|$ stands for the spectral norm and ϕ such that $\phi(t) = t.u(t)$.

Remark: This result is similar to those presented in the classical asymptotic regime (m fixed and $N \to +\infty$).

Robust RMT

2) is the key result! Notably, it implies that

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

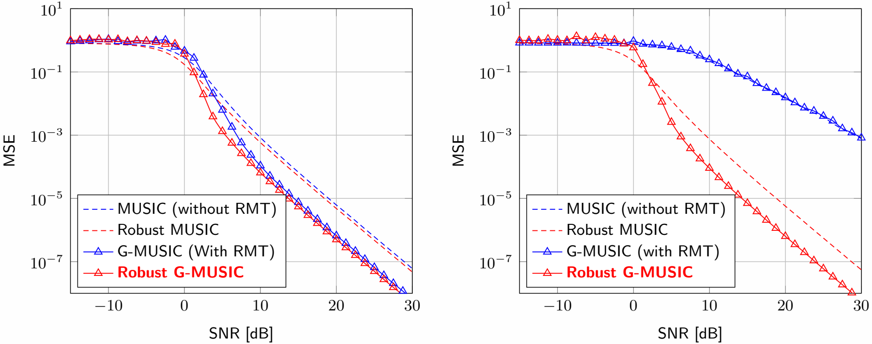
3) $\hat{\gamma}(\theta) = \sum_{i=1}^{m} \beta_i \mathbf{s}(\theta)^H \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^H \mathbf{s}(\theta)$ is STILL the G-MUSIC statistic for the M-estimators

where

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

with $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_m$ (resp. $\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_m$ the eigenvalues (resp. the eigenvectors) of $\mathbf{V}_N$ and $\hat{\mu}_1 \leq \ldots \leq \hat{\mu}_m$ the eigenvalues of $\mathrm{diag}(\hat{\boldsymbol{\lambda}}) - \frac{1}{m}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^T$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_m)^T$.

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

Resolution probability of 2 sources

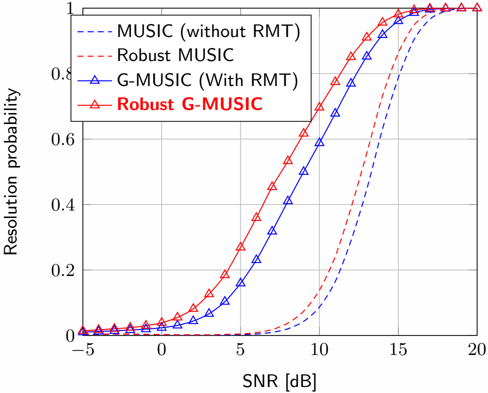


Figure: Resolution performance of the MUSIC estimators in homogeneous clutter for 50 data of size 10

Pros and Cons of these results

- Advantages
  - Original results on robust RMT
  - Now, possibility of using robust estimators in a RMT context: extension of classical RMT results such DoA estimation (done), sources power estimation, number of sources estimation (challenging problem), detection...
  - Great improvement: sources resolution, MUSIC statistic est.
- Limitations
  - Assumption of independence, i.e. not CES dist:

$$\mathbf{z}_i = \begin{pmatrix} \tau_1 \, x_i^{(1)} \\ \vdots \\ \tau_m \, x_i^{(m)} \end{pmatrix} \text{ instead of } \mathbf{z}_i = \tau_i \begin{pmatrix} x_i^{(1)} \\ \vdots \\ x_i^{(m)} \end{pmatrix}$$

    where all the quantity are independent (means $\neq$ random amplitude on the different sensors).
  - Improvement on MSE is valid for the MUSIC statistic estimate and NOT for the DoA estimate.

Robust RMT under CES distributions

- Previous results remain valid under CES distributions, i.e. where τ_i are r.va. with unknown PDF (M-estimators, (Couillet, 2015)).

  Technical condition: For each $a > b > 0$, one has $\lim_{t\to\infty} \frac{\limsup_N \nu_N([t,\infty))}{\phi(at)-\phi(bt)} \to 0$. where $\nu_N = \frac{1}{N}\sum_{i=1}^N \delta_{\tau_i}$ and $\phi(t) = t.u(t)$.

  Meaning: one has to control the queue of the dist. of τ_i.

- Also valid for Tyler's estimator (Zhang, 2016): $\phi(t) = m, \forall t > 0$. More tight condition but same idea for the proof.

Robust RMT under CES distributions

Results on the eigenvalues distributions of the M-estimators for CES

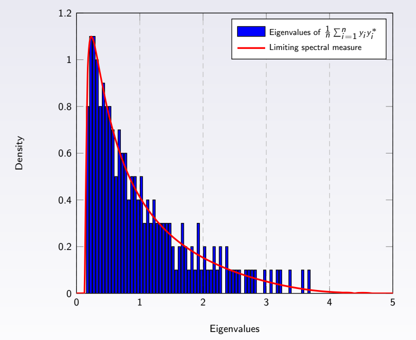
R. Couillet, F. Pascal, and J. W. Silverstein, "The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples", *JMVA*, vol. 139, 2015.

Ideas of the proofs? Break and discusssions.

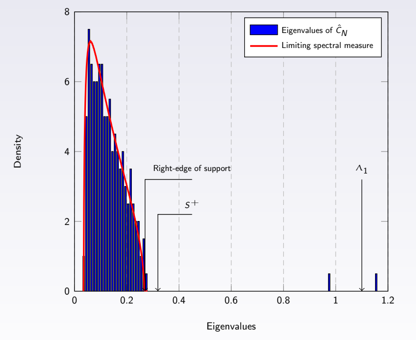
Results on the eigenvalues distributions of the Tyler's estimator for CES

T. Zhang, X. Cheng, and A. Singer, "Marchenko-Pastur Law for Tyler's and Maronna's M-estimators", arXiv preprint arXiv:1401.3424, 2016.

Robust RMT under CES distributions



(a) SCM

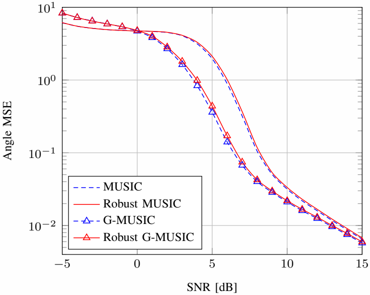


(b) Student M-estimator

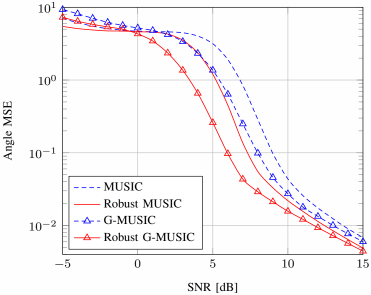
Histogram of the eigenvalues of the SCM and a M-estimator against the limiting spectral measure, with 2 sources, $p_1 = p_2 = 1$, $m = 200, N = 1000$, Student-t distributions

Robust RMT under CES distributions

MSE on the DoA estimation



(c) Gaussian

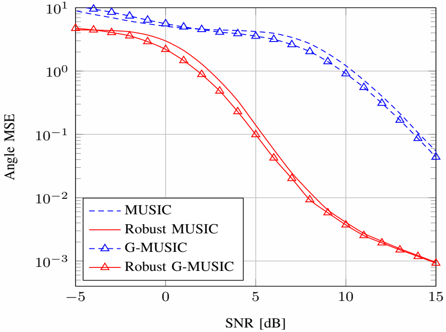


(d) K-dist ($\nu = 1$, homogeneous)

MSE vs SNR of the DoA estimation in the case of 2 sources ($\theta_1 = 14^\circ$ and $\theta_2 = 18^\circ$), for Gaussian noise and K-distributed noise, where $N = 100$ and $m = 20$.

Robust RMT under CES distributions

MSE on the DoA estimation

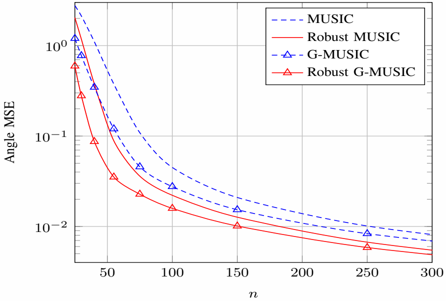


K-dist ($\nu = 0.11$, heterogeneous)

MSE vs SNR of the DoA estimation in the case of 2 sources ($\theta_1 = 14^\circ$ and $\theta_2 = 18^\circ$), for Gaussian noise and K-distributed noise, where $N = 100$ and $m = 20$.

Interest on sources resolution

Robust RMT under CES distributions



MSE vs the ration m/N of the DoA estimation in the case of 2 sources ($\theta_1 = 14^\circ$ and $\theta_2 = 18^\circ$), for homogeneous K-distributed noise, where $SNR = 10dB$ and $m = 20$.

I. Introduction

II. Estimation, background and applications

III. Random Matrix Theory

IV. Regularized M-estimators and link to RMT

- Motivations and definitions
- Optimization and detection

V. Conclusions and perspectives

Motivations

Some advantages

- Regularized problem, with norm penalties (e.g. for sparsity)
- Combined with M-estimators $\Rightarrow$ robustness to outliers
- May allow to include *a priori* informations
- Case of small number of observations or under-sampling $N < m$: matrix is not invertible $\Rightarrow$ Problem when using M-estimators or Tyler's estimator!

It is an active research on this topic:
see the works of Yuri Abramovich, Olivier Besson, Romain Couillet, Mathew McKay, Ami Wiesel...

Regularized Tyler's estimators (RTE)

Chen estimator

$$\widehat{\boldsymbol{\Sigma}}_C(\rho) = (1-\rho)\,\frac{m}{N}\sum_{i=1}^{N}\frac{\mathbf{z}_i\mathbf{z}_i^H}{\mathbf{z}_i^H\widehat{\boldsymbol{\Sigma}}_C^{-1}(\rho)\mathbf{z}_i} + \rho\mathbf{I}$$

subject to the constraint $\text{Tr}(\widehat{\boldsymbol{\Sigma}}_C(\rho)) = m$ and for $\rho \in (0,1]$.

- Originally introduced in (Abramovich, 2007)
- Existence, uniqueness and algorithm convergence proved in (Chen, 2011)

  Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4097-4107, 2011.

Remark: Constraint $\text{Tr}(\widehat{\boldsymbol{\Sigma}}_C(\rho)) = m$ has two interests:

- Allowing ρ to live in $[0,1]$
- Making the prove easier

Regularized Tyler's estimators

Pascal estimator

$$\widehat{\boldsymbol{\Sigma}}_P(\rho) = (1-\rho)\,\frac{m}{N}\sum_{i=1}^{N} \frac{\mathbf{z}_i \mathbf{z}_i^H}{\mathbf{z}_i^H \widehat{\boldsymbol{\Sigma}}_P^{-1}(\rho)\mathbf{z}_i} + \rho \mathbf{I}$$

subject to the **no** trace constraint but for $\rho \in (\bar{\rho}, 1]$, where $\bar{\rho} := \max(0, 1 - N/m)$.

- Existence, uniqueness and algorithm convergence proved in (Pascal, 2013)

  F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to STAP detection problem," *Signal Processing, IEEE Transactions on,* vol. 62, pp. 5640-5651, Nov. 2014.

- $\widehat{\boldsymbol{\Sigma}}_P(\rho)$ **(naturally) verifies** **Tr**$(\widehat{\boldsymbol{\Sigma}}_P^{-1}(\rho)) = m$ **for all** $\rho \in (0, 1]$

Regularized Tyler's estimators

The main challenge is to find the optimal ρ!

According to the applications...**MSE, detection performances...**

One (theoretical) answer is given thanks to RMT in ...

R. Couillet and M. R. McKay, "Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators," Journal of Multivariate Analysis, vol. 131, pp. 99-120, 2014.

where it is also proved that

- Both estimators have asymptotically (RMT regime) the same performance (achieved for a different value of beta)
- They asymptotically perform as a normalized version of the Ledoit-Wolf estimator (similar to previous results).

O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of multivariate analysis* , vol. 88, no. 2, pp. 365-411, 2004.

Regularized Tyler's estimators

Objective: Robust estimate of $\mathbf{M} = E[\mathbf{z}_i \mathbf{z}_i^H]$, for $\mathbf{z}_1, \ldots, \mathbf{z}_N \in \mathbb{C}^m$ i.i.d. with

- $\mathbf{z}_i = \sqrt{\tau_i}\, \mathbf{M}^{1/2} \mathbf{x}_i$, $\mathbf{x}_i$ has i.i.d. entries, $E[\mathbf{x}_i] = \mathbf{0}$, $E[\mathbf{x}_i \mathbf{x}_i^H] = \mathbf{I}$
- $\tau_i > 0$ random impulsions with $E[\tau_i] = 1$.
- m fixed and $N \to \infty$ (Classical asymptotics!)

OR

- $\mathbf{z}_i = \sqrt{\tau_i}\, \mathbf{M}^{1/2} \mathbf{x}_i$, $\mathbf{x}_i$ has i.i.d. entries, $E[\mathbf{x}_i] = \mathbf{0}$, $E[\mathbf{x}_i \mathbf{x}_i^H] = \mathbf{I}$
- $\tau_i > 0$ random impulsions with $E[\tau_i] = 1$.
- $c_m \triangleq \frac{m}{N} \to c$ as $m, N \to \infty$
- few data: $m \sim N$.

Find "optimal" regularized parameter!

RTE Asymptotics

Assumptions: m fixed and $N \to +\infty$

Let us set

$$\boldsymbol{\Sigma}_0(\rho) = m\,(1-\rho) E\left[\frac{\mathbf{z}\mathbf{z}^H}{\mathbf{z}^H \boldsymbol{\Sigma}_0^{-1}(\rho)\mathbf{z}}\right] + \rho \mathbf{I}$$

for $\rho \in (\bar{\rho}, 1]$, where $\bar{\rho} := \max(0, 1 - N/m)$.

Then, for any $\kappa > 0$, one has

$$\sup_{\rho \in [\kappa, 1]} \left\| \widehat{\boldsymbol{\Sigma}}_P(\rho) - \boldsymbol{\Sigma}_0(\rho) \right\| \xrightarrow[m \text{ fixed } , N \to \infty]{a.s} 0$$

Remark: Of course, $\boldsymbol{\Sigma}_0(\rho) \neq \mathbf{M}$!!! What is $\boldsymbol{\Sigma}_0(\rho)$? ... it can be shown that they share the same eigenvectors space.

RTE Asymptotics

Characterization of $\Sigma_0(\rho)$

Let us first denote $\mathbf{\Sigma}_0 = \mathbf{\Sigma}_0(\rho)$.

- Multiplying by $\mathbf{M}^{-1/2}$, one obtains:

$$\mathbf{M}^{-1/2}\,\mathbf{\Sigma}_0\,\mathbf{M}^{-1/2} = m\,(1-\rho)E\left[\frac{\mathbf{x}\mathbf{x}^H}{\mathbf{x}^H\mathbf{M}^{1/2}\,\mathbf{\Sigma}_0^{-1}\,\mathbf{M}^{1/2}\mathbf{x}}\right] + \rho\mathbf{M}^{-1}$$

- Let the eigenvalue decomposition of $\mathbf{M}^{-1/2}\,\mathbf{\Sigma}_0\,\mathbf{M}^{-1/2} = \mathbf{V}\mathbf{D}\mathbf{V}^H$.
- Then, $m\,(1-\rho)E\left[\frac{\mathbf{x}\mathbf{x}^H}{\mathbf{x}^H\mathbf{D}\mathbf{x}}\right] + \rho\mathbf{V}^H\mathbf{M}^{-1}\mathbf{V} = \mathbf{D}^{-1}$

  $\Longrightarrow E\left[\frac{\mathbf{x}\mathbf{x}^H}{\mathbf{x}^H\mathbf{D}\mathbf{x}}\right] = \mathrm{diag}(\alpha_1,\ldots,\alpha_m)$ is diagonal implying $\mathbf{\Sigma}_0$ and $\mathbf{M}$ share the same eigenvector space.

Lemma If $\mathbf{D} = \mathrm{diag}(d_1,\ldots,d_m)$, then α_i are given by

$$\alpha_i = \frac{1}{2^m\,m}\frac{1}{\prod_{j=1}^m d_j}F_D^{(m)}\left(m,1,\ldots,2,1,\ldots,1,m+1,\frac{d_1-1/2}{d_1},\ldots,\frac{d_m-1/2}{d_m}\right)$$

where $F_D^{(m)}$ is the Lauricella's type D hypergeometric function.

RTE Asymptotics

Characterization of $\Sigma_0(\rho)$

- Denote by $\alpha_i(\{d_j\}_{j=1}^m) = E\left[\frac{|x_i|^2}{\mathbf{x}^H \mathbf{D}\mathbf{x}}\right]$. Then

$$m(m-1)\alpha_i(\{d_i\}_{i=1}^m) + \frac{\rho}{\lambda_i} = \frac{1}{d_i},$$

where λ_i are the eigenvalues of $\mathbf{M}$: $\mathbf{M} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^H$ with $\boldsymbol{\Delta} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ and $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_m$.

- Start from $d_1^{(0)}, \ldots, d_m^{(0)}$ and compute iteratively

$$d_i^{(t+1)} = \frac{1}{\frac{\rho}{\lambda_i} + m(1-m)\alpha_i(\mathrm{diag}(\mathbf{d}^{(t)}))}$$

until convergence. If $d_{1,\infty}, \ldots, d_{m,\infty}$ are the obtained values, then...

- Set $s_{i,\infty} = \lambda_i \, d_{i,\infty}$, Then,

$$\boldsymbol{\Sigma}_0 = \mathbf{V}\,\mathrm{diag}(s_{1,\infty}, \ldots, s_{m,\infty})\mathbf{V}^H.$$

RTE Asymptotics

Assumptions: m fixed and $N \to +\infty$

Similarly to M-estimators, one can establish a CLT:

Theorem 1 (Asymptotic distribution of $\widehat{\boldsymbol{\Sigma}}_P(\rho)$)

$$\sqrt{N}\ \mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_P(\rho) - \boldsymbol{\Sigma}_0(\rho)) \xrightarrow{d} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{M}_1, \mathbf{M}_2\right),$$

where $\mathbb{C}\mathcal{N}$ is the complex Gaussian distribution, $\mathbf{M}_1$ the CM and $\mathbf{M}_2$ the pseudo CM.

RMT Asymptotic Behavior

Theorem (Asymptotic Behavior (Couillet-McKay, 2014))

For $\varepsilon \in (0, \min\{1, c^{-1}\})$, *define* $\hat{\mathcal{R}}_\varepsilon = [\varepsilon + \max\{0, 1 - c^{-1}\}, 1]$.

Then, as $m, N \to \infty$, $m/N \to c \in (0, \infty)$,

$$\sup_{\rho \in \hat{\mathcal{R}}_\varepsilon} \left\| \widehat{\boldsymbol{\Sigma}}_P(\rho) - \widetilde{\mathbf{S}}_m(\rho) \right\| \xrightarrow{\text{a.s.}} 0$$

with

$$\widetilde{\mathbf{S}}_m(\rho) = \frac{1}{\underline{\gamma}(\rho)} \frac{1-\rho}{1-(1-\rho)c} \frac{1}{N} \sum_{i=1}^{N} \mathbf{M}^{\frac{1}{2}} \mathbf{x}_i \mathbf{x}_i^H \mathbf{M}^{\frac{1}{2}} + \rho \mathbf{I}$$

and $\underline{\gamma}(\rho)$ *unique positive solution to equation*

$$1 = \frac{1}{m} \operatorname{Tr}\left(\mathbf{M} \left(\rho \underline{\gamma}(\rho) \mathbf{I} + (1-\rho)\mathbf{M} \right)^{-1} \right).$$

Moreover, $\rho \mapsto \underline{\gamma}(\rho)$ *continuous on* $(0, 1]$.

Asymptotic Model Equivalence

Theorem (Model Equivalence (Couillet-McKay, 2014))

For each $\rho \in (0,1]$, *there exist unique* $\underline{\rho} \in (\max\{0, 1-c^{-1}\}, 1]$ *such that*

$$\frac{\widetilde{\mathbf{S}}_m(\rho)}{\frac{1}{\underline{\gamma}(\underline{\rho})}\frac{1-\underline{\rho}}{1-(1-\underline{\rho})c} + \underline{\rho}} = (1-\underline{\rho})\frac{1}{N}\sum_{i=1}^{N} \mathbf{M}^{\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^*\mathbf{M}^{\frac{1}{2}} + \underline{\rho}\mathbf{I}.$$

Besides, $(0,1] \to (\max\{0, 1-c^{-1}\}, 1]$, $\rho \mapsto \underline{\rho}$ *is increasing and onto.*

- Estimator behaves similar to impulsion-free Ledoit-Wolf estimator
- **About uniformity:** Uniformity over ρ essential to find optimal values of ρ.
- $\widetilde{\mathbf{S}}_m$ is unobservable!

Context

Hypothesis testing: Two sets of data

- Initial pure-noise data: $\mathbf{z}_1, \ldots, \mathbf{z}_N$, $\mathbf{z}_n = \sqrt{\tau_n}\,\mathbf{M}^{1/2}\mathbf{x}_n$ as before.
- $$\begin{cases} \text{Hypothesis } H_0\text{:} & \mathbf{y} = \mathbf{z} & \mathbf{y}_n = \mathbf{z}_n & n = 1, \ldots, N \\ \text{Hypothesis } H_1\text{:} & \mathbf{y} = \mathbf{s} + \mathbf{z} & \mathbf{y}_n = \mathbf{z}_n & n = 1, \ldots, N \end{cases}$$

  with $\mathbf{z} = \sqrt{\tau}\,\mathbf{M}^{1/2}\mathbf{x}$, $\mathbf{s} = \alpha\,\mathbf{p}$, $\mathbf{p} \in \mathbb{C}^m$ deterministic known, α unknown.

GLRT detection test:

$$T_m(\rho) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \Gamma$$

for some detection threshold Γ where

$$T_m(\rho) \triangleq \frac{\left|\mathbf{y}^H \widehat{\boldsymbol{\Sigma}}_P^{-1}(\rho)\mathbf{p}\right|}{\sqrt{\mathbf{y}^H \widehat{\boldsymbol{\Sigma}}_P^{-1}(\rho)\mathbf{y}}\sqrt{\mathbf{p}^H \widehat{\boldsymbol{\Sigma}}_P^{-1}(\rho)\mathbf{p}}}.$$

Context

Originally found to be $\widehat{\boldsymbol{\Sigma}}_P(0)$ *but*

- *only valid for* $m < N$
- $\rho \geq 0$ *can only bring improvements.*

Basic comments:

- For $\Gamma > 0$, as $m, N \to \infty$, $m/N \to c > 0$, under H_0,

$$T_m(\rho) \xrightarrow{\text{a.s.}} 0.$$

$\Rightarrow$ Zero false alarm, trivial result.

- Non-trivial solutions for $\Gamma = \gamma/\sqrt{m}$, $\gamma > 0$ fixed.

Objectives

Objective: For finite but large m, N, solve

$$\rho^\star = \operatorname{argmin}_\rho \left\{ P\left(\sqrt{m} T_m(\rho) > \gamma\right) \right\}.$$

Several steps:

- for each ρ, central limit theorem to evaluate

$$\lim_{\substack{m,N\to\infty \\ m/N\to c}} P\left(\sqrt{m} T_m(\rho) > \gamma\right)$$

(very involved due to intricate structure of $\widehat{\boldsymbol{\Sigma}}_P$*)*

- find minimizing ρ
- estimate minimizing ρ

Main results

Theorem (Asymptotic detector performance (Couillet-Pascal, 2015))

As $m, N \to \infty$ with $m/N \to c \in (0, \infty)$,

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| P\left(T_m(\rho) > \frac{\gamma}{\sqrt{m}} \right) - \exp\left(-\frac{\gamma^2}{2\sigma_m^2(\underline{\rho})} \right) \right| \to 0$$

with $\rho \mapsto \underline{\rho}$ aforementioned mapping and $\sigma_m^2(\underline{\rho}) \triangleq$

$$\frac{1}{2} \frac{\mathbf{p}^H \mathbf{M} Q_m^2(\underline{\rho}) \mathbf{p}}{\mathbf{p}^H Q_m(\underline{\rho}) \mathbf{p} \cdot \frac{1}{m} \operatorname{Tr}\left(\mathbf{M} Q_m(\underline{\rho}) \right) \cdot \left(1 - c(1-\underline{\rho})^2 f(-\underline{\rho})^2 \frac{1}{N} \operatorname{Tr}\left(\mathbf{M}^2 Q_m^2(\underline{\rho}) \right) \right)}$$

with $Q_m(\underline{\rho}) \triangleq (\mathbf{I} + (1-\underline{\rho}) f(-\underline{\rho}) \mathbf{M})^{-1}$.

- Limiting Rayleigh distribution (weak convergence to Rayleigh $R_m(\underline{\rho})$)
- **Remark:** σ_m and $\underline{\rho}$ not function of γ

$\Rightarrow$ There exists uniformly optimal ρ.

Simulation

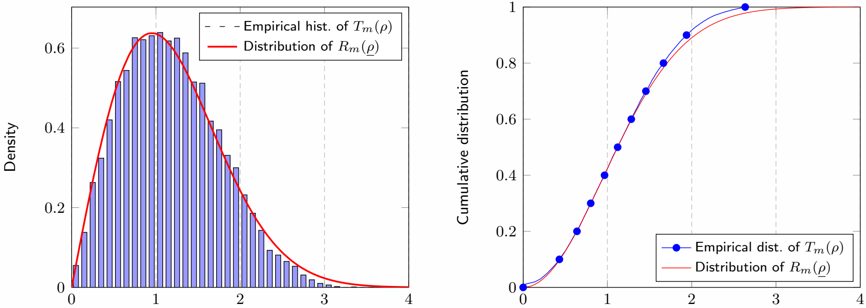


Figure: Histogram distribution function of the $\sqrt{m}T_m(\rho)$ versus $R_m(\underline{\rho})$, $m = 20$, $N = 40$ $\mathbf{p} = m^{-\frac{1}{2}}[1, \ldots, 1]^T$, $\mathbf{M}$ Toeplitz from AR of order 0.7, $\rho = 0.2$.

Simulation

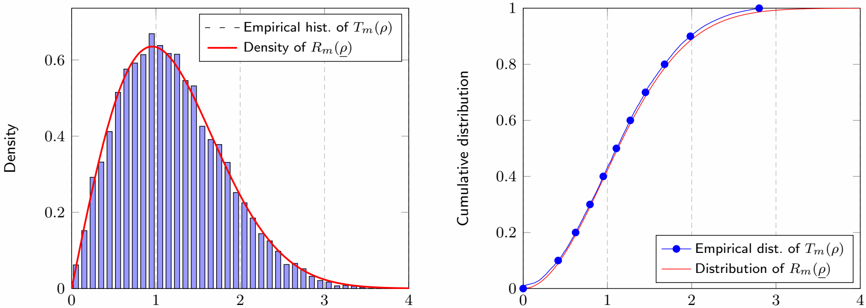


Figure: Histogram distribution function of the $\sqrt{N}T_m(\rho)$ versus $R_m(\underline{\rho})$, $m = 100$, $N = 200$ $\mathbf{p} = m^{-\frac{1}{2}}[1, \ldots, 1]^T$, $\mathbf{M}$ Toeplitz from AR of order 0.7, $\rho = 0.2$.

Empirical estimation of optimal ρ

Optimal ρ depends on unknown $\mathbf{M}$. We need:

- empirical estimate $\sigma_m(\underline{\rho})$
- minimize the estimate
- prove asymptotic optimality of estimate.

Theorem (Empirical performance estimation (Couillet-Pascal, 2015))

For $\rho \in (\max\{0, 1 - c_m^{-1}\}, 1)$, *let*

$$\hat{\sigma}_m^2(\underline{\rho}) \triangleq \frac{1}{2} \frac{1 - \rho \cdot \frac{\mathbf{p}^H \widehat{\boldsymbol{\Sigma}}_P^{-2}(\rho)\mathbf{p}}{\mathbf{p}^H \widehat{\boldsymbol{\Sigma}}_P^{-1}(\rho)\mathbf{p}}}{(1 - c_m + c_m \rho)(1 - \rho)}.$$

Also let $\hat{\sigma}_m^2(1) \triangleq \lim_{\underline{\rho} \uparrow 1} \hat{\sigma}_m^2(\underline{\rho})$.
Then

$$\sup_{\rho \in \mathcal{R}_\kappa} \left| \sigma_m^2(\underline{\rho}) - \hat{\sigma}_m^2(\underline{\rho}) \right| \xrightarrow{\text{a.s.}} 0.$$

Final result

Theorem (Optimality of empirical estimator (Couillet-Pascal, 2015))

Define

$$\underline{\rho}_m^* = \operatorname{argmin}_{\{\rho \in \mathcal{R}'_\kappa\}} \left\{ \hat{\sigma}_m^2(\underline{\rho}) \right\}.$$

Then, for every $\gamma > 0$*,*

$$P\left(\sqrt{m} T_m(\underline{\rho}_m^*) > \gamma\right) - \inf_{\rho \in \mathcal{R}_\kappa} \left\{ P\left(\sqrt{m} T_m(\rho) > \gamma\right) \right\} \to 0.$$

Simulations

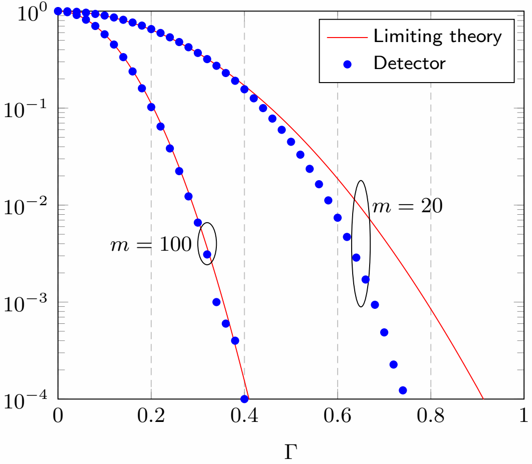


Figure: False alarm rate $P(T_m(\rho) > \Gamma)$ for $m = 20$ and $m = 100$, $\mathbf{p} = m^{-\frac{1}{2}}[1, \ldots, 1]^T$, $M_{ij} = 0.7^{|i-j|}$, $c_m = 1/2$.

Analogous results can be obtained under H_1 (more useful!).

A. Kammoun, R. Couillet, F. Pascal, and M.-S. Alouini, "Optimal Design of the Adaptive Normalized Matched Filter Detector," *Information Theory, IEEE Transactions on (submitted to)*, 2016. arXiv:1501.06027

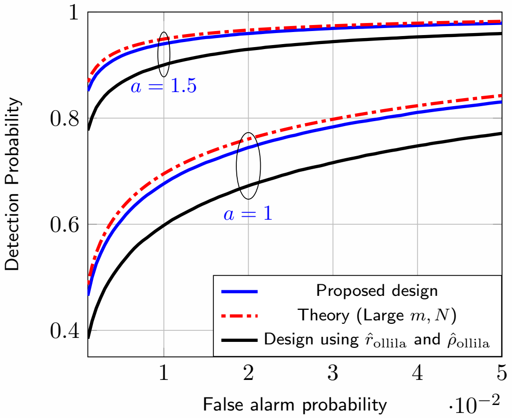


Figure: ROC curves for non-Gaussian clutters when $m = 250$ (STAP application $N_a = 10$, $N_p = 25$), $N = 250$, $f_d = 0.6$

Conclusions and Perspectives

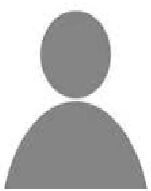
- Conclusions
  - Derivation of the complex M-estimators asymptotic distribution, the robust ANMF and the MUSIC statistic asymptotic distributions.
  - In the Gaussian case, M-estimators built with $\sigma_1 N$ data behaves as SCM built with N data (i.e. slight loss of performance in Gaussian case).
  - Better estimation in non-Gaussian cases.
  - Extension to the Robust RMT and derivation of the Robust G-MUSIC method.
  - Shrinkage M-estimators: one more degree of freedom (for Big data problems, robust methods...)

Conclusions and Perspectives

- 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
  - Asymptotics for regularized robust estimators
  - RMT analysis for regularized robust estimators

References and thanks to...

- my co-authors:



Yacine Chitour Abla Kammoun Romain Couillet Jack Silverstein

- and many inspiring people working in this field
  Yuri Abramovich, Olivier Besson, Ernesto Conte, Antonio De Maio, Alfonso Farina, Fulvio Gini, Maria Greco, Shawn Kraut, Jean-Philippe Ovarlez, Louis Scharf, Ami Wiesel . . .

List of references

Y.I. Abramovich and O. Besson.
Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 1: The over-sampled case.
Signal Processing, IEEE Transactions on, 61(23):5807–5818, 2013.

YI Abramovich and Nicholas K Spencer.
Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering.
In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 3, pages 1105–1108. IEEE, 2007.

O. Besson and Y.I. Abramovich.
Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach-part 2: The under-sampled case.
Signal Processing, IEEE Transactions on, 61(23):5819–5829, 2013.

Olivier Besson and Yuri I Abramovich.

Invariance properties of the likelihood ratio for covariance matrix estimation in some complex elliptically contoured distributions.
Journal of Multivariate Analysis, 124:237–246, 2014.

Yilun Chen, Ami Wiesel, and Alfred O Hero.
Robust shrinkage estimation of high-dimensional covariance matrices.
Signal Processing, IEEE Transactions on, 59(9):4097–4107, 2011.

R. Couillet and F. Pascal.
Robust M-estimator of scatter for large elliptical samples.
In *IEEE Workshop on Statistical Signal Processing, SSP-14*, Gold Coast, Australia, June 2014.

R Couillet, F Pascal, and J W Silverstein.
A Joint Robust Estimation and Random Matrix Framework with Application to Array Processing.
In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-13*, Vancouver, Canada, May 2013.

R Couillet, F Pascal, and J W Silverstein.

Robust M-Estimation for Array Processing: A Random Matrix Approach.
Information Theory, IEEE Transactions on (submitted to), 2014. arXiv:1204.5320v1.

R Couillet, F Pascal, and J W Silverstein.
The Random Matrix Regime of Maronna's M-estimator with elliptically distributed samples.
Journal of Multivariate Analysis (submitted to), 2014. arXiv:1311.7034.

Romain Couillet and Matthew R McKay.
Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators.
arXiv preprint arXiv:1401.4083, 2014.

Olivier Ledoit and Michael Wolf.
A well-conditioned estimator for large-dimensional covariance matrices.
Journal of multivariate analysis, 88(2):365–411, 2004.

F. Pascal and Y. Chitour.
Shrinkage covariance matrix estimator applied to STAP detection.
In *IEEE Workshop on Statistical Signal Processing, SSP-14*, Gold Coast, Australia, June 2014.

F. Pascal, Y. Chitour, and Y. Quek.
Generalized robust shrinkage estimator and its application to STAP detection problem.
Signal Processing, IEEE Transactions on (submitted to), 2014 arXiv:1311.6567.

Ilya Soloveychik and Ami Wiesel.
Non-asymptotic Error Analysis of Tyler's Scatter Estimator.
arXiv preprint arXiv:1401.6926, 2014.

Ami Wiesel.
Unified framework to regularized covariance estimation in scaled Gaussian models.
Signal Processing, IEEE Transactions on, 60(1):29–38, 2012.

Teng Zhang, Xiuyuan Cheng, and Amit Singer.

Marchenko-Pastur Law for Tyler's and Maronna's M-estimators.
arXiv preprint arXiv:1401.3424, 2014.

Thank you for your attention!

Questions?