Covariance Matrix Estimation and Applications in Radar

Frédéric Pascal

1L2S
Supélec, France

Crash course/Tutorial
January 2015
Contact info:

- Frédéric Pascal, frederic.pascal@supelec.fr
  http://www.supelec.fr/426_p_16381/frederic-pascal.html

This crash course is an extension of the tutorial ”CES distributions with applications” given at ICASSP’14 with

- Esa Ollila, esa.ollila@aalto.fi
  http://signal.hut.fi/~esollila/

- David E. Tyler, dtyler@rci.rutgers.edu
  http://www.rci.rutgers.edu/~dtyler/
Contents

■ Part A
  CES distributions

■ Part B
  ML and $M$-estimators of scatter matrix

■ Part C
  Applications

■ Part D
  $M$-estimators of scatter in the large dimensional regime
Motivation and applications

In many applications, the complex multinormality is a poor approximation of the underlying physics

- radar clutter,
- noise and interference in indoor and outdoor mobile communication channels
- noise in imaging problems, etc.

Complex elliptically symmetric (CES) distributions form a natural and flexible extension of the complex normal (CN) distribution:

- Allowing heavier/lighter tails than the CN distribution while maintaining the elliptical geometry of the equidensity contours.
- Many results/properties for the CN distribution carry over in this broader class.
Motivation and applications

- CES distributions can also be useful
  - to assess the robustness of estimators/detectors to non-Gaussianity
  - derive alternative (robust) estimators of the parameters of the CN distribution, the mean vector and the covariance matrix.

- ML and $M$-estimators of scatter matrix parameter
  - Widely used estimators of covariance matrix in many applications (e.g., array and radar)
  - Can be tuned to have desirable properties: high efficiency over neighborhood of target model(s) and smooth and bounded influence function
Part A

CES distributions
Part A: Contents

1 Preliminaries
   - Complex vectors and distributions
   - Circular symmetry
   - Complex normal distribution

2 CES distributions
   - Definition
   - The absolutely continuous case
   - Compound Gaussian distributions

3 Examples
   - $t$-distribution
   - $K$-distribution
   - Generalized Gaussian distribution
   - Inverse Gaussian distribution

4 Angular central Gaussian distribution

5 Radar clutter analysis
Part A and B are based on

- Esa, Dave, V. Koivunen and H.V. Poor
  Complex elliptically symmetric distributions: survey, new results and applications

- Referred in the text as [CES,2012]
Outline

1 Preliminaries
   ■ Complex vectors and distributions
   ■ Circular symmetry
   ■ Complex normal distribution

2 CES distributions
   ■ Definition
   ■ The absolutely continuous case
   ■ Compound Gaussian distributions

3 Examples
   ■ $t$-distribution
   ■ $K$-distribution
   ■ Generalized Gaussian distribution
   ■ Inverse Gaussian distribution

4 Angular central Gaussian distribution

5 Radar clutter analysis
A complex vector in $\mathbb{C}^m$: $z = x + jy \in \mathbb{C}^m$, $x, y \in \mathbb{R}^m$.

Complex-to-real vector mapping (Isomorphism) $[,]_R : \mathbb{C}^m \mapsto \mathbb{R}^{2m}$

$$[z]_R = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Complex-to-real matrix mapping $\{\cdot\}_R : \mathbb{C}^{n \times m} \mapsto \mathbb{R}^{2n \times 2m}$

$$\{C\}_R = \begin{pmatrix} A & -B \\ B & A \end{pmatrix} \quad \text{for} \quad C = A + jB$$

The (real) matrix on RHS is said to be of complex form.

$\mathbb{C}$-linear transforms

$$[Cz]_R = \{C\}_R [z]_R$$
A complex random vector (r.v.) $z \in \mathbb{C}^m$ is comprised of a pair of real r.v.’s $x = \text{Re}(z)$ and $y = \text{Im}(z)$ in $\mathbb{R}^m$.

The distribution of $z$ on $\mathbb{C}^m$ determines the joint real $2m$-variate distribution of $x$ and $y$ on $\mathbb{R}^{2m}$ and conversely:

$$F_z(c) \triangleq \mathbb{P}([z]_\mathbb{R} \leq [c]_\mathbb{R}).$$

The probability density function (p.d.f.) of $z = x + jy$

$$f(z) \triangleq f([z]_\mathbb{R}) \equiv f(x, y).$$

The mean (expectation)

$$\mathbb{E}[z] = \mathbb{E}[x] + j\mathbb{E}[y].$$
- **Characteristic function (c.f.)**

\[
\Phi_z(c) \triangleq \mathbb{E}\left[ \exp \left\{ j \left[ c \right]_{\mathbb{R}} \left[ z \right]_{\mathbb{R}} \right\} \right] \\
= \mathbb{E}\left[ \exp \left\{ j \text{Re} (c^H z) \right\} \right] \\
= \mathbb{E}\left[ \exp \left\{ \frac{1}{2} (c^H z + c^T z^*) \right\} \right]
\]

- **Covariance matrix**

\[
\mathbf{C} \equiv \text{cov}(z) \triangleq \mathbb{E}\left[ (z - \mathbb{E}[z])(z - \mathbb{E}[z])^H \right] \\
= \text{cov}(x) + \text{cov}(y) + j\{\text{cov}(y, x) - \text{cov}(x, y)\}
\]

is positive semidefinite Hermitian ($\mathbf{C}^H = \mathbf{C}$), denoted $\mathbf{C} \in \mathcal{H}_m^0$.

- $\mathcal{H}_m$ denotes the class of positive definite Hermitian $m \times m$ matrices.
The class of circularly symmetric distributions

Circular symmetry

- R.v. \( z \) is said to have a **circularly symmetric** distribution about \( \mu \), called the **symmetry center**, if

\[
(z - \mu) \overset{d}{=} e^{j\theta}(z - \mu), \quad \forall \theta \in \mathbb{R},
\]

where the notation \( \overset{d}{=} \) should be read “has same distribution as”.

- R.v. \( z \) is said to be **circular** if it is circularly symmetric about the origin (\( \mu = 0 \)).

- Symmetry center equals \( \mu = \mathbb{E}[z] \) under finite 1st order moments.

- All marginals \( z_i \) of \( z \) have spherical distributions.
If a circular r.v. \( z \) possess a density, then its p.d.f \( f(z) \) satisfies

\[
f(e^{j\theta}z) = f(z), \quad \forall \theta \in \mathbb{R}.
\]

In the univariate case \((m = 1)\),

\[
f(z) = C \cdot g(|z|^2)
\]

for some non-neg. fnc \( g(\cdot) \) and normalizing constant \( C \).

The kurtosis of a circularly symmetric random variable (r.va.) \( z \in \mathbb{C} \):

\[
kurt(z) \triangleq \frac{\mathbb{E}[|z - \mu|^4]}{(\mathbb{E}[|z - \mu|^2])^2} - 2
\]

measures peakedness and heavy-tailedness as in the real case.
2nd-order circularity:

(i) \( \text{cov}(\mathbf{x}) = \text{cov}(\mathbf{y}) \) and \( \text{cov}(\mathbf{x}, \mathbf{y}) = -\text{cov}(\mathbf{y}, \mathbf{x}) \)

(ii) \( \mathbb{E}[(\mathbf{z} - \mathbb{E}[\mathbf{z}]) (\mathbf{z} - \mathbb{E}[\mathbf{z}])^\top] = \mathbf{0} \).

(iii) \( \text{cov}([\mathbf{z}]_\mathbb{R}) = (1/2)\{\mathbf{C}\}_\mathbb{R} \).

The above conditions are pairwise equivalent.

If \( \mathbf{z} \) circularly symmetric \( \rightarrow \) \( \mathbf{z} \) is 2nd-order circular (given the second-order moments exist), but the converse is naturally not necessarily true.
Complex (circular) normal distribution

\[ z \sim \mathbb{C} \mathcal{N}_m(\mu, \Sigma) \] with parameters \( \mu \in \mathbb{C}^m \) and \( \Sigma \in \mathcal{H}_m^0 \) if its c.f. is

\[ \Phi_z(c) = \exp\{j \text{Re}(c^H \mu)\} \exp\left\{ -\frac{1}{4} c^H \Sigma c \right\} \]

- This definition admits singular normal distributions.
- Equivalent to stating that

\[ [z]_R \sim \mathcal{N}_{2m}([\mu]_R, \frac{1}{2} \{\Sigma\}_R) \]

i.e., \([z]_R\) has a \(2m\)-variate real normal distribution with covariance matrix of complex form
CN distribution: Important properties

1. Param’s $\mu$ and $\Sigma$ are equal to $\mathbb{E}[z]$ and the covariance matrix $C$, respectively.

2. All marginals have a CN distribution, e.g., $z_i \sim \mathcal{CN}(\mu_i, \sigma_i^2)$, where $\sigma_i^2 = [\Sigma]_{ii}$.

3. The class is closed under affine transformations:
\[ Bz + b \sim \mathcal{CN}_k(B\mu + b, B\Sigma B^H) \quad \forall \ B \in \mathbb{C}^{k \times m}, b \in \mathbb{C}^k \]

4. If $\text{rank}(\Sigma) = m$, the quadratic form
\[ Q(z) = (z - \mu)^H \Sigma^{-1} (z - \mu) \sim \text{Gam}(m, 1). \]

5. If $\text{rank}(\Sigma) = m$, the p.d.f. exists and is of the form
\[ f(z) = \pi^{-m} |\Sigma|^{-1} \exp\{-(z - \mu)^H \Sigma^{-1} (z - \mu)\}, \]
where $|\Sigma|$ denotes the determinant of $\Sigma$. 
Outline

1 Preliminaries
   - Complex vectors and distributions
   - Circular symmetry
   - Complex normal distribution

2 CES distributions
   - Definition
   - The absolutely continuous case
   - Compound Gaussian distributions

3 Examples
   - $t$-distribution
   - $K$-distribution
   - Generalized Gaussian distribution
   - Inverse Gaussian distribution

4 Angular central Gaussian distribution

5 Radar clutter analysis
Definition

**Complex elliptically symmetric distribution**

\[ z \sim \text{CE}_m(\mu, \Sigma, \phi) \] if its characteristic function is

\[
\Phi_z(c) = \exp\{j\text{Re}(c^H \mu)\} \phi(c^H \Sigma c)
\]

for some fnc \( \phi : \mathbb{R}^+ \to \mathbb{R} \), called the characteristic generator, parameters \( \Sigma \in \mathcal{H}_m \), called the scatter matrix, and symmetry center \( \mu \in \mathbb{C}^m \).

- The case \( \mu = 0 \) is called as the centered CES distribution.
- Definition allows singular CES distribution
- Equivalent to stating that

\[
[z]_\mathbb{R} \sim \text{RE}_{2m}\left([\mu]_\mathbb{R}, \frac{1}{2}\{\Sigma\}_\mathbb{R}, \phi\right)
\]
Some remarks

1. **CN distribution**: $\phi(t) = \exp(-\frac{1}{4}t)$

2. $\phi$ may depend on $m$. (Sometimes we write $\phi_m$ when there is possibility for confusion).

3. $\Sigma$ and $\phi(\cdot)$ do not *uniquely* identify the $m$-variate CES distribution.

Write $\Sigma_0 = c^2 \Sigma$ and $\phi_0(t) = \phi(t/c^2)$, $c > 0$

$$\Rightarrow \text{CE}_m(\mu, \Sigma, \phi) = \text{CE}_m(\mu, \Sigma_0, \phi_0).$$

This ambiguity is easily avoided by imposing a scale constraint on $\phi(\cdot)$ or on $\Sigma$. 
The class is closed under affine transformations

If \( z \sim \text{CE}_m(\mu, \Sigma, \phi) \), then

\[
Bz + b \sim \text{CE}_k(B\mu + b, B\Sigma B^H, \phi) \quad \forall B \in \mathbb{C}^{k \times m}, \ b \in \mathbb{C}^k.
\]

Parameters are transformed as \((\mu, \Sigma) \mapsto (B\mu + b, B\Sigma B^H)\).

All subvectors of \( z \sim \text{CE}_m(\mu, \Sigma, \phi) \) have CES distributions, where the characteristic generator \( \phi \) remains unchanged.

\[
\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim \text{CE}_m\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \phi\right)
\]

\[\Rightarrow z_1 \sim \text{CE}_d(\mu_1, \Sigma_{11}, \phi) \text{ and } z_2 \sim \text{CE}_{m-d}(\mu_2, \Sigma_{22}, \phi)\]
Characterizing property

- **Unit complex $m$-sphere**: 
  \[ \mathbb{C}S^m \triangleq \{ z \in \mathbb{C}^m : \| z \| = 1 \} \]

- **$u$ (or $u^{(m)}$) = r.v. with a uniform distr. on $\mathbb{C}S^m$**, \( u \sim U(\mathbb{C}S^m) \).

**Stochastic representation theorem**

\( z \sim \text{CE}_m(\mu, \Sigma, \phi) \) if and only if it admits the stochastic representation

\[ z =_d \mu + \mathcal{R} A u^{(k)}, \]

where r.v.a. \( \mathcal{R} \geq 0 \), called the **modular variate**, is independent of \( u^{(k)} \) and \( \Sigma = A A^H \) is a factorization of \( \Sigma \), where \( A \in \mathbb{C}^{m \times k} \) with \( k = \text{rank}(\Sigma) \).
Some remarks

1. **One-to-one relation** with c.d.f. \( F_R(\cdot) \) of \( R \) and characteristic generator \( \phi(\cdot) \).

2. **Ambiguity**: both \((R, A)\) and \((c^{-1}R, cA)\), \(c > 0\) are valid stochastic representations of \( z \).
   - a reformulation of the earlier identifiability issue which can be solved by restricting the scale of \( R \) or \( A \)

3. **Distribution of quadratic form**: If \( \text{rank}(\Sigma) = m \), then

   \[
   Q(z) \triangleq (z - \mu)\Sigma^{-1}(z - \mu) =_d Q.
   \]

   where \( Q \triangleq R^2 \) is called as the **2nd-order modular variate**.
4 Random number generation:
- draw a random deviate $R$ from a distribution $F_R$ and $u^{(k)}$ from $U(CS^k)$
- set $z = \mu + RAu^{(k)}$ for a given $\mu \in \mathbb{C}^m$ and $A \in \mathbb{C}^{m \times k}$.

5 Marginal r.v.a. $z_i \sim CE_1(\mu_i, \sigma_i^2, \phi)$ of $z$ possess a circular distribution and it admits the following stochastic representation

$$z_i = \mu_i + \sigma_i R_1 u,$$

where
- $u \sim$ uniform on the unit circle, $u \perp R_1$.
- $R_1 = \gamma R$, where $\gamma^2 \sim Beta(1, m - 1)$, $\gamma \perp R$.
Moments

1. **Existence:** $p$th-order moments exists if and only if $E[R^p] < \infty$

2. **Mean:** If $E[R] < \infty$, then $E[z] = \mu + E[R] A E[u] = \mu$.

3. **Covariance matrix:** If $E[R^2] < \infty$, then

   \[ C = \sigma_C \Sigma, \quad \text{(i.e., } C \propto \Sigma) \]

   where \( \sigma_C \triangleq \frac{E[R^2]}{\text{rank}(\Sigma)} = -4\phi'(0). \)

   For proper identifiability of $(R, \Sigma)$, the scale constraint:

   \[ E[R^2] = \text{rank}(\Sigma) \text{ is convenient as it implies } C = \Sigma. \]

4. **Mixed moments:** All odd-order central moments vanish. Also even-order moments vanish if the multiplicity of $z_i$ and $z_i^*$ differ.
The 6th-order moment $\mathbb{E}[(z_1 - \mu_1)^2(z_2 - \mu_2)^2(z_1 - \mu_1)^*2] = 0.$

Kurtosis: If $\mathbb{E}[R^4] < \infty$ and $\text{rank}(\Sigma) = m$, then

$$\text{kurt}(z_i) = 2\kappa,$$

where

$$\kappa = \frac{\phi''(0)}{\phi'(0)^2} - 1$$

$$= \frac{m\mathbb{E}[R^4]}{(m + 1)(\mathbb{E}[R^2])^2} - 1$$

$$= \frac{2\text{kurt}(x_i)}{3}$$

with $x_i$ being the real (or imaginary) part of $z_i$. 
Existence of a density

- R.v. $z \sim \text{CE}_m(\mu, \Sigma, \phi)$, where $\text{rank}(\Sigma) = m$ and absolutely continuous modular variate $\mathcal{R}$.

**P.d.f. of CES distribution**

\[ f_z(z) = C_{m,g} |\Sigma|^{-1} g((z - \mu)^H \Sigma^{-1} (z - \mu)), \]

where $g : \mathbb{R}_0^+ \rightarrow \mathbb{R}^+$, called the density generator, satisfies

\[ \delta_{m,g} \triangleq \int_0^\infty t^{m-1} g(t) dt < \infty, \]

to ensure the integrability of $f_z(\cdot)$, and $C_{m,g}$ is a normalizing constant, which ensures that $f_z(z)$ integrates to 1. We write $z \sim \text{CE}_m(\mu, \Sigma, g)$. 

Some remarks

1. **CN distribution**: \( g(t) = \exp(-t) \).

2. **Normalizing constant** has the form:
   \[
   C_{m,g} \triangleq 2(s_m \delta_{m,g})^{-1},
   \]
   where \( s_m \triangleq 2\pi^m / \Gamma(m) \) is the surface area of \( \mathbb{C}S^m \).

3. **Quadratic form** \( Q(z) = (z - \mu)^H \Sigma^{-1} (z - \mu) \) and \( Q = \mathcal{R}^2 \) have p.d.f.
   \[
   f_Q(t) = t^{m-1} g(t) \delta_{m,g}^{-1}.
   \]

4. **Level sets** of the density \( f_z(z) \) are ellipsoids in the complex Euclidean \( m \)-space (since p.d.f. depends on \( z \) only via \( Q(z) \)).
Marginal r.v.a. $z_i$ has p.d.f.

$$f_i(z) = C_{1|m,g} \frac{1}{\sigma_i^2} g_1|m \left( \frac{|z - \mu_i|^2}{\sigma_i^2} \right)$$

where the normalizing constant is

$$C_{1|m,g} = C_{m,g} \frac{\pi^{m-1}}{\Gamma(m-1)}$$

and the density generator $g_1|m : \mathbb{R}_0^+ \rightarrow \mathbb{R}^+$ is of the form

$$g_1|m(t) = \int_t^\infty (y - t)^{m-2} g(y) dy.$$
Note:

- The integral in marginal density generator may not have closed-form solution.
- The “standardized” \((\mu_i = 0, \sigma^2_i = 1)\) r.v.a. \(z_i\) has amplitude \(|z_i| = \mathcal{R}_1\ \forall i\) with p.d.f.

\[
f_{\mathcal{R}_1}(r) = 2\pi r \ C_{1|m,g} \ g_{1|m}(r^2).
\]

6. **Covariance matrix exists** if \(\mathbb{E}[\mathcal{R}^2] = \mathbb{E}[Q] = \delta_{m+1,g} < \infty\).

- \(\Sigma\) is a well defined parameter (even if \(\mathcal{C}\) does not exist) that determines the shape and orientation of the elliptical equidensity contours.
Compound Gaussian (CG) distributions

An important subclass of CES distributions, also called as

- Spherically invariant random vectors
- Scale mixture of normal distributions [Andrews and Mallows, 1974]

**Compound Gaussian distributions**

\[ z \sim CG_m(\mu, \Sigma, F_\tau) \text{ if it admits a stochastic CG-representation} \]

\[ z =_d \mu + \sqrt{\tau} n \]

for r.v. \( \tau > 0 \) with c.d.f. \( F_\tau \), called the *texture*, independent of r.v. \( n \sim \mathcal{CN}_m(0, \Sigma) \), called the *speckle*. 
Some remarks

1. **P.d.f:** If \( \text{rank}(\Sigma) = m \), the density always exist and is:

\[
f_z(z) = \pi^{-m} |\Sigma|^{-1} \int_0^\infty \tau^{-m} \exp(-Q(z)/\tau) dF_\tau(\tau) = \pi^{-m} |\Sigma|^{-1} \int_0^\infty \tau^{-m} \exp(-Q(z)/\tau) f_\tau(\tau) d\tau
\]

where the latter eq. holds when \( \tau \) possesses a density \( f_\tau(\tau) = F'_\tau(\tau) \).

2. **Density generator:** \( z \) has a \( \text{CE}_m(\mu, \Sigma, g) \) distribution with

\[
g(t) \propto \int_0^\infty \tau^{-m} \exp(-t/\tau) f_\tau(\tau) d\tau.
\]
3 Consistency property: Marginals belong to the same class with the same mixing distribution $F_\tau(\cdot)$.

Ex $z_1 \sim CG_d(\mu_1, \Sigma_{11}, F_\tau)$ where $z_1$, $\mu_1$ and $\Sigma_{11}$ are as earlier.

4 Covariance matrix exists if $E[\tau] < \infty$,

$$
C = \text{cov}(\sqrt{\tau} n) = E[\tau]\text{cov}(n) \\
= \sigma_C \Sigma \quad \text{with} \quad \sigma_C = E[\tau].
$$

5 Identifiability: Both $(\sqrt{\tau}, n)$ and $(\sqrt{\tau}a, n/a)$ $\forall a > 0$ are proper CG-representations of $z$.

For proper identifiability, one can impose a scale constraint on:

- $\tau$, e.g., (unit mean assumption) $E[\tau] = 1$
- $\Sigma$, e.g., $\text{Tr}(\Sigma) = m$.

Note: If $\tau$ does not have finite mean, then a sensible constraint is $\text{Med}(\tau) = 1$. 

Outline

1 Preliminaries
   - Complex vectors and distributions
   - Circular symmetry
   - Complex normal distribution
2 CES distributions
   - Definition
   - The absolutely continuous case
   - Compound Gaussian distributions
3 Examples
   - t-distribution
   - K-distribution
   - Generalized Gaussian distribution
   - Inverse Gaussian distribution
4 Angular central Gaussian distribution
5 Radar clutter analysis
Examples of CES distributions

- For simplicity of notation, we let $\mu = 0$.
- Stochastic representation
  \[
  z = d \mathcal{R} Au \quad u \sim U(\mathbb{C}S^m), \quad \Sigma = AA^H, \quad u \perp \mathcal{R}, \quad \mathcal{R} > 0
  \]
- Stochastic CG-representation:
  \[
  z = d \sqrt{\tau} n \quad n \sim \mathcal{CN}_m(0, \Sigma), \quad \tau \perp n, \quad \tau > 0
  \]

The distribution class of $z \sim \text{CE}_m(0, \Sigma, g)$ can be defined by stating:

1. distribution of modular variate $\mathcal{R} > 0$ or 1-to-1 fnc of it ($Q = \mathcal{R}^2$)
2. distribution of texture variate $\tau > 0$ or 1-to-1 fnc of it
3. density generator $g(t)$
The $m$-variate $t_\nu$-distribution $z \sim C_{t_{m,\nu}}(\mu, \Sigma)$

**CG characterization**

$\tau^{-1}$ has unit mean Gamma distribution with shape $\nu/2$, i.e.,

$$\tau^{-1} \sim \text{Gam}(\nu/2, 2/\nu)$$

where $\nu > 0$.

**CES characterization**

The second order modular variate $Q = \mathcal{R}^2$ verifies

$$(1/m)Q \sim F_{2m,\nu},$$

where $F_{l,q}$ denotes the $F$-distribution with $l$ and $q$ d.o.f.

**P.d.f**

$$f_z(z) = C_{m,\nu} \left| \Sigma \right|^{-1} \left( 1 + 2z^H \Sigma^{-1} z / \nu \right)^{-(2m+\nu)/2}, \quad \nu > 0$$
Some remarks

1. $\nu > 0$ is the degrees of freedom parameter:
   - $\nu = 1$ is called the complex Cauchy distribution
   - $\nu \to \infty$ yields the CN distribution
   - finite 2nd-order moments for $\nu > 2$

2. The density generator and the normalizing constant are

   $$g(t) = (1 + 2t/\nu)^{-(2m+\nu)/2}$$

   $$C_{m,g} = \frac{2^m \Gamma(\frac{2m+\nu}{2})}{(\pi \nu)^{m/2} \Gamma(\frac{\nu}{2})}$$

3. Covariance matrix: $\mathbf{C} = \sigma_C \mathbf{\Sigma}$ with

   $$\sigma_C = \mathbb{E}[^C] = \frac{\nu}{\nu - 2} \text{ for } \nu > 2.$$
4. Belongs to the class of CG distributions $\Leftrightarrow$ marginal distributions belong to the same class (consistency property). For example,

$$z_i \sim \mathbb{C}t_{1,\nu}(\mu_i, \sigma_i^2), \quad i = 1, \ldots, m$$

5. Kurtosis exists for $\nu > 4$:

$$\text{kurt}(z_i) = \frac{4}{\nu - 4}, \quad i = 1, \ldots, m.$$ 

6. Amplitude p.d.f. of $|z_i|$ in the standardized ($\mu_i = 0, \sigma_i^2 = 1$) case:

$$f_{R_1}(r) = 2r \left(1 + \frac{2r^2}{\nu}\right)^{-(\nu+2)/2}$$

which is the square-root of an r.v.a. with the $F_{2,\nu}$ distribution.

NOTE: recall stochastic decomposition $z_i = d \mu_i + \sigma_i R_1 u$ for $i = 1, \ldots, m$
Alternative parametrization

- We presumed scale constraint: $\mathbb{E}[\tau^{-1}] = 1$. Alternatively, we can restrict the scale of $\Sigma$ (i.e., model $\tau^{-1}$ with a scale parameter.)

- Write

  \[ V = \eta \Sigma, \quad \text{where } \eta = \frac{m}{\text{Tr}(\Sigma)} \]

- Reparametrize the d.o.f. parameter: $\lambda = \nu/2$

  \[ (\tau')^{-1} = \eta \tau^{-1} \sim \text{Gam}\left(\lambda, \frac{\eta}{\lambda}\right) \]
Alternative parametrization (cont’d)

- We denote this case as \( \mathbf{z} \sim Ct_{\lambda}(\mathbf{0}, \mathbf{V}, \eta) \).
- The p.d.f. then rewrites as

\[
f_{\mathbf{z}}(\mathbf{z}) = \frac{\Gamma(m + \lambda)}{\pi^m \Gamma(\lambda) |\mathbf{V}|^{-1}} \left( \frac{\lambda}{\eta} + \mathbf{z}^H \mathbf{V}^{-1} \mathbf{z} \right)^{-(m+\lambda)} , \quad \lambda, \eta > 0
\]

This parametrization is useful if the scale \( \sigma^2(\mathbf{\Sigma}) = \eta^{-1} \) of \( \mathbf{\Sigma} \),

\[
\sigma^2 = \frac{\text{Tr}(\mathbf{\Sigma})}{m} = \frac{1}{m} \sum_{i=1}^{m} \sigma_i^2
\]

= average (measurement/sensor/clutter) power

can be assumed to be known (e.g., when \( z_i \)'s are standardized).
The \( m \)-variate \( K \)-distribution \( \mathbf{z} \sim \mathbb{C}K_{m,\nu}(\mathbf{\mu}, \mathbf{\Sigma}) \)

**CG characterization**

\( \tau \) has unit mean Gamma distribution with shape parameter \( \nu > 0 \), i.e.,

\[
\tau \sim \text{Gam}(\nu, 1/\nu),
\]

with p.d.f.

\[
f_\tau(\tau) = \frac{\nu^\nu}{\Gamma(\nu)} \tau^{\nu-1} e^{-\nu \tau}.
\]

**Density**

\[
f_{\mathbf{z}}(\mathbf{z}) = \frac{2\nu^{(\nu+m)/2}}{\Gamma(\nu)\pi^m} |\mathbf{\Sigma}|^{-1} \left( \mathbf{z}^\mathsf{H} \mathbf{\Sigma}^{-1} \mathbf{z} \right)^{(\nu-m)/2} K_{\nu-m}(2\sqrt{\nu} \mathbf{z}^\mathsf{H} \mathbf{\Sigma}^{-1} \mathbf{z})
\]

where \( K_{\ell}(\cdot) \) denotes the modified Bessel fnc of 2nd kind of order \( \ell \).
Some remarks

1. All moments exist and hence the constraint $E[\tau] = 1 \ (\Rightarrow \Sigma = C)$.

2. Consistency property: All marginals belong to the same class.

3. Kurtosis is

$$kurt(z_i) = 2 \text{Var}(\tau) = \frac{2}{\nu} \text{ for } i = 1, \ldots, m.$$ 

$\nu \downarrow$ the heavier-tailed is the distribution

$\nu \to \infty$ yields the CN distribution (and $kurt(z_i) \to 0$).

4. Amplitude p.d.f. of $|z_i|$ in the standardized ($\mu_i = 0, \sigma_i^2 = 1$) case:

$$f_{R_1}(r) = \frac{2\sqrt{\nu}}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}r)^\nu K_{\nu-1}(2\sqrt{\nu}r)$$

equals the real univariate $K$-distribution with shape $\nu$ and unit power.
The \textit{m}-variate Generalized Gaussian (GG) distribution

\textbf{Characterization}

The 2nd-order modular variate \( Q = R^2 \) verifies

\[
Q =_d G^{1/s} \quad \text{where } G \sim \text{Gam} \left( \frac{m}{s}, \eta^{-s} \right), \quad s, \eta > 0
\]

\textbf{Density}

\[
f_{\mathbf{z}}(\mathbf{z}) = \frac{s \Gamma(m) \eta^m}{\pi^m \Gamma(m/s)} \left| \Sigma \right|^{-1} \exp \left( - \left( \eta \mathbf{z}^H \Sigma^{-1} \mathbf{z} \right)^s \right)
\]

\[
g(t) = \exp(- (\eta t)^s)
\]

\[
\text{norm. cst. } C_{m,g}
\]

- Is a multivariate complex analog of the (real) exponential power family also called as \textit{Box-Tiao distributions}.
- A subclass of multivariate symmetric Kotz type distributions.
For proper identifiability we can either:

1. Restrict the scale of $R$ (and hence the scale parameter $\eta$)
   - We set $\mathbb{E}[R^2] = m \implies \mathcal{C} = \Sigma$ which is equivalent to
     \[
     \eta^{-1} = \frac{m \Gamma\left(\frac{m}{s}\right)}{\Gamma\left(\frac{m+1}{s}\right)}
     \]
   - We denote this case by $z \sim \mathcal{CGG}_{m,s}(\mu, \Sigma)$.

2. Restrict the scale of $\Sigma$ and set $\Sigma = V \ (\text{Tr}(V) = m)$.
   - In this case $\eta^{-1} = \text{Tr}(\Sigma)/m$ is the scale of $\Sigma$.
   - We denote this case by $z \sim \mathcal{CGG}_{m,s}(\mu, V, \eta)$.

We assume the former parametrization. The latter is useful when the scale of $\Sigma$ is assumed to be known.
Some remarks

1. Case $s = 1$ equals the CN distribution. Heavier tailed than normal for $s < 1$ and lighter tailed for $s > 1$.

2. Case $s = 1/2$ is a generalization of Laplace distribution.
   - In this case $g(t) = \exp(-\sqrt{\eta t})$ and the marginal density generator is
     \[ g_1|m(t) = \int_t^\infty (y - t)^{m-2} \exp(-\sqrt{\eta y}) dy. \]
   - Kurtosis becomes $\text{kurt}(z_i) = 2\kappa$
     \[ \kappa = \frac{2}{2m + 1}. \]

3. NOTE: Also for general $s \neq 1$, the kurtosis depends on dimension $m$.
   \[ \Rightarrow \] GG distribution is not consistent and hence NOT a member of CG distributions.
Inverse Gaussian distribution $\tau \sim \text{CIG}_{m,\lambda}(\mu, \Sigma)$

**CG characterization**

Texture $\tau$ follows an inverse Gaussian (IG) distribution with shape $\lambda > 0$ and unit mean ($\mathbb{E}[\tau] = 1$), the p.d.f. being

$$f_\tau(\tau) = \sqrt{\frac{\lambda}{2\pi}} \tau^{-3/2} \exp\left(-\frac{\lambda(\tau - 1)^2}{2\tau}\right).$$

**Density**

$$f_z(z) = C_{m,g} |\Sigma|^{-1} \left(1 + \frac{2z^H \Sigma^{-1} z}{\lambda}\right)^{-\left(m+\frac{1}{2}\right)/2} K_{m+\frac{1}{2}}\left(\lambda \sqrt{1 + \frac{2z^H \Sigma^{-1} z}{\lambda}}\right)$$

where the normalizing cnst. is $C_{m,g} = 2^{\nu(m+\nu)/2} / [\Gamma(\nu) \pi^m]$. 
Some remarks

1. All moments exists: hence the scale constraint $\mathbb{E}[\tau] = 1 \Rightarrow \Sigma = \mathcal{C}$

2. Density generator is

$$g(t) = \left(1 + \frac{2t}{\lambda}\right)^{-(m+\frac{1}{2})/2} K_{m+\frac{1}{2}}\left(\lambda \sqrt{1 + \frac{2t}{\lambda}}\right)$$

3. Consistency property: All marginals belong to the same class.

4. Kurtosis is $\text{kurt}(z_i) = 2 \text{Var}(\tau) = 2/\lambda$ or $i = 1, \ldots, m$.

   $\lambda \downarrow$ the heavier-tailed (or spikier) is the distribution

   $\lambda \uparrow$ the distribution gets closer to multinormal ($\text{kurt}(z_i) \to 0$)
Some remarks

4 Amplitude p.d.f. of $|z_i|$ in the standardized ($\mu_i = 0, \sigma_i^2 = 1$) case:

$$f_{R_1}(r) = 2^{3/2} \sqrt{\frac{\lambda}{\pi}} e^{\lambda r} \left( \sqrt{1 + \frac{2r^2}{\lambda}} \right)^{-3/2} K_{3/2} \left( \lambda \sqrt{1 + \frac{2r^2}{\lambda}} \right)$$

**NOTE:** recall stochastic decomposition $z_i = d \mu_i + \sigma_i R_1 u$ for $i = 1, \ldots, m$

5 Amplitude c.d.f.:

$$F_{R_1}(r) = 1 - e^{\lambda} \exp \left( -\lambda \sqrt{1 + \frac{2r^2}{\lambda}} \right) \left( 1 + \frac{2r^2}{\lambda} \right)^{-1/2}.$$
Outline

1 Preliminaries
   - Complex vectors and distributions
   - Circular symmetry
   - Complex normal distribution

2 CES distributions
   - Definition
   - The absolutely continuous case
   - Compound Gaussian distributions

3 Examples
   - $t$-distribution
   - $K$-distribution
   - Generalized Gaussian distribution
   - Inverse Gaussian distribution

4 Angular central Gaussian distribution

5 Radar clutter analysis
Angular Gaussian distribution $z_a \sim CAG_m(0, \Sigma)$

**Definition**

R.v $z_a$ has complex angular central Gaussian (ACG) distribution if

$$z_a = d \frac{z}{\|z\|}, \quad \text{where} \quad z \sim \mathbb{C}N_m(0, \Sigma)$$

**Density (for nonsingular $\Sigma$)**

$$f_{z_a}(z) = s_m^{-1} |\Sigma|^{-1} (z^H \Sigma^{-1} z)^{-m}$$

where $s_m = 2\pi^m/\Gamma(m)$ is the surface area of $\mathbb{C}S^m$.

- The parameter $\Sigma$ can only be identified up to a scale.
- Widely used in statistical shape analysis.
Some remarks

1. The term “angular central Gaussian” is a slight misnomer. In its definition, $CN_m(0, \Sigma)$ can be replaced by any central CES distribution $CE_m(0, \Sigma, \phi)$.

2. ACG distribution does not belong to CES distributions. Note: CES density is w.r.t. Lebesgue measure on $\mathbb{R}^{2m}$ and ACG density is w.r.t. the uniform measure on $\mathbb{C}S^m$.

3. If $u \sim U(\mathbb{C}S^m)$, then $Au/\|Au\| \sim CAG_m(0, \Sigma)$ for $\Sigma = AA^H$ and nonsingular $A \in \mathbb{C}^{m \times m}$.

4. The class is closed under standardized linear transformations:

$$z_a \sim CAG_m(0, \Sigma) \Rightarrow Bz_a/\|Bz_a\| \sim CAG_m(0, B\Sigma B^H)$$

for any nonsingular $B \in \mathbb{C}^{m \times m}$. 
Outline

1 Preliminaries
   - Complex vectors and distributions
   - Circular symmetry
   - Complex normal distribution

2 CES distributions
   - Definition
   - The absolutely continuous case
   - Compound Gaussian distributions

3 Examples
   - $t$-distribution
   - $K$-distribution
   - Generalized Gaussian distribution
   - Inverse Gaussian distribution

4 Angular central Gaussian distribution

5 Radar clutter analysis
Radar clutter analysis

- McMaster IPIX radar lake-clutter data [Bakker and Currie, ]
- Polarimetric measurements (HH, HV, VH, and VV) of the backscattering from the lake surface, taken at three different resolutions.
- Following [Conte et al., 2004] we focus our analysis on the datasets 84, 85 and 86 taken at three different range resolutions;
- We use 6 different CES models. See [CES,2012].
Figure: Clutter amplitudes (first row) and a.p.d.f. (second row) for HH and VV polarisations, 19th range cell, 30m, dataset 84
Empirical quantiles of the data
Theoretical quantiles of the ML-fit

(a) complex $t$-distribution
(b) IG-CG distribution

Figure: QQ plots for the dataset 84 (HH polarization, 19th range cell, 30m). The horizontal (resp. vertical) lines indicate the values of the theoretical (resp. empirical) 0.95th and 0.99th quantiles.
List of references

For a more complete list and appropriate pointers to results, see [CES, 2012]
Scale mixtures of normal distributions.

Bakker, R. and Currie, B.
The McMaster IPIX Radar Sea Clutter Database.
http://soma.crl.mcmaster.ca/ipix/.

Statistical analysis of real clutter at different range resolutions.

Characterisation of radar clutter as a spherically invariant random process.

Spherically invariant random processes for modeling non-Gaussian radar clutter.

Part B

ML and $M$-estimators of scatter matrix
Part B : Contents

1  ML-estimators of scatter
   ▪ Definition
   ▪ Existence, Uniqueness, Convergence

2  $M$-estimators of scatter
   ▪ Complex Huber’s $M$-estimator
   ▪ Complex Tyler’s $M$-estimator

3  Robustness and asymptotics
   ▪ The influence function
   ▪ Asymptotics
Problem statement

Given \( m \)-variate i.i.d. data set \( z_1, \ldots, z_n, n > m \) from an elliptical population \( CE_m(0, \Sigma, g) \), we wish to find the MLE \( \hat{\Sigma} \).

Classical estimator is the sample covariance matrix (SCM)

\[
\hat{C} = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^H
\]

- We study conditions for existence, uniqueness and computation of the M(L)-solution using an iterative fixed point algorithm.
- Also a general expression for the influence function (IF) is derived as well asymptotic normality and covariances.
Outline

1. **ML-estimators of scatter**
   - Definition
   - Existence, Uniqueness, Convergence

2. **M-estimators of scatter**
   - Complex Huber’s $M$-estimator
   - Complex Tyler’s $M$-estimator

3. **Robustness and asymptotics**
   - The influence function
   - Asymptotics
The maximum likelihood estimator (MLE)

**negative log-likelihood function (divided by \( n \))**

\[
\mathcal{L}_n(\Sigma) = \frac{1}{n} \sum_{i=1}^{n} \rho(z_i^H \Sigma^{-1} z_i) - \ln |\Sigma^{-1}| \quad \text{where} \quad \rho(t) = -\ln g(t)
\]

**MLE**

\[
\hat{\Sigma} = \arg \min_{\Sigma \in \mathcal{H}_m} \mathcal{L}_n(\Sigma)
\]

Critical points are solutions to **estimating equation**

\[
\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \varphi(z_i^H \hat{\Sigma}^{-1} z_i) z_i z_i^H, \quad \text{where} \quad \varphi(t) = -g'(t)/g(t) = \rho'(t)
\]
The Gaussian MLE

- \( z_1, \ldots, z_n \) i.i.d. sample from \( \mathcal{CN}_m(0, \Sigma) \).
- Density generator is \( g(t) = \exp(-t) \Rightarrow \rho(t) = -\ln g(t) = t \) and \( \varphi(t) = \rho'(t) \equiv 1 \)

The neg. log-likelihood function is

\[
\mathcal{L}_n(\Sigma) = \frac{1}{n} \sum_{i=1}^{n} \rho(z_i^H \Sigma^{-1} z_i) - \ln |\Sigma^{-1}|
= \text{Tr}(\hat{C} \Sigma^{-1}) - \ln |\Sigma^{-1}|
\]

and the est. eq. is explicit with (local) solution given by the SCM \( \hat{\Sigma} \).
- SCM is the unique minimizer (i.e., the MLE \( \hat{\Sigma} \) of the problem).
General Case

\[ \mathcal{L}_n(\Sigma) = \frac{1}{n} \sum_{i=1}^{n} \rho(z_i^H \Sigma^{-1} z_i) - \ln |\Sigma^{-1}| \]

where \[ \rho(t) = -\ln g(t) \]

- Existence
- Uniqueness
- Implicit estimating equation
  - A fast algorithm to find the solution is needed.
Some notations

- \( z_i = x_i + jy_i \in \mathbb{C}^m \) maps to a pair of orthogonal vec’s in \( \mathbb{R}^{2m} \):

\[
\mathbf{v}_i = [z_i]_{\mathbb{R}} = \begin{pmatrix} x_i \\ y_i \end{pmatrix},
\]

\[
\mathbf{v}_{n+i} = [jz_i]_{\mathbb{R}} = \begin{pmatrix} -y_i \\ x_i \end{pmatrix} \quad \text{for } i = 1, \ldots, n.
\]

Above \( \mathbf{v}_i \) and \( \mathbf{v}_{n+i} \) have identical \( \text{RE}_{2m}(\mathbf{0}, (1/2)[\Sigma]_{\mathbb{R}}, g) \) distributions, but are naturally not independent.

- Let \( \mathbb{P}_{2n}(\cdot) \) denotes the empirical distribution of the sample \( \mathbf{v}_1, \ldots, \mathbf{v}_{2n} \).
Assumptions

\[ \mathcal{L}_n(\Sigma) = \frac{1}{n} \sum_{i=1}^{n} \rho(z_i^H \Sigma^{-1} z_i) - \ln |\Sigma^{-1}| \]

where \( \rho(t) = -\ln g(t) \)

Assumptions

**M1** \( \varphi(t) = \rho'(t) \) is non-negative, continuous and nonincreasing

**M2** \( \psi(t) = t \varphi(t) \) is strictly increasing

**M3** For all linear subspaces \( V \subset \mathbb{R}^{2m} \), with \( \dim(V) \leq 2m - 1 \), it holds that

\[ \mathbb{P}_{2n}(V) < 1 - \frac{2m - \dim(V)}{2K}. \]

where \( K \triangleq \sup\{\psi(t) | t \geq 0\} \).

Note: M3 implies that \( K > m \). If \( K = \infty \), then M3 states that all the data cannot lie in some lower dimensional subspace.
Existence, Uniqueness, Convergence

Theorem 6 [CES, 2012] Suppose M1, M2 and M3 hold. Then,

- The MLE $\hat{\Sigma}$ exists and corresponds to the unique solution of the estimating equation.
- Given any initial estimate $\Sigma_0 \in \mathcal{H}_m$, the iterations
  \[
  \Sigma_{k+1} = \frac{1}{n} \sum_{i=1}^{n} \varphi(z_i^H \Sigma_k^{-1} z_i) z_i z_i^H
  \]
  converge to $\hat{\Sigma}$.

- The proof relies upon the results in [Kent and Tyler, 1991] and assumption M2 above can be further relaxed:
  - M2** $\psi(t) = t \varphi(t)$ is nondecreasing, but strictly increasing for $\psi(t) < m + \epsilon$ for some $\epsilon > 0$.
- This result has also been established in [Maronna, 1976] and in [Tyler, 1988] (real case) but under more stringent conditions.
Existence, Uniqueness, Convergence

Theorem 7(b) [CES, 2012] Suppose M1, M2** and M3 hold. Then,

- The MLE $\hat{\Sigma}$ exists and corresponds to the unique solution of the estimating equation.
- Given any initial estimate $\Sigma_0 \in \mathcal{H}_m$, the iterations

$$
\Sigma_{k+1} = \frac{1}{n} \sum_{i=1}^{n} \varphi(z_i^H \Sigma_k^{-1} z_i) z_i z_i^H
$$

converge to $\hat{\Sigma}$.

- The proof relies upon the results in [Kent and Tyler, 1991] and assumption M2 above can be further relaxed:
- M2** $\psi(t) = t\varphi(t)$ is nondecreasing, but strictly increasing for $\psi(t) < m + \epsilon$ for some $\epsilon > 0$.
- This result has also been established in [Maronna, 1976] and in [Tyler, 1988] (real case) but under more stringent conditions.
Some remarks

1. Key result (using the isomorphism $\{\cdot\}_{\mathbb{R}}$) is the equivalence with the real $M$-estimating equation

$$\tilde{\Sigma} = \frac{1}{2n} \sum_{i=1}^{2n} \varphi \left( \frac{1}{2} v_i^\top \tilde{\Sigma}^{-1} v_i \right) v_i v_i^\top,$$

where $\tilde{\Sigma}$ is positive definite symmetric matrix of complex form.

2. Condition M2 admits most popular $M(L)$-estimators, but not all.

- M2** is needed for Huber’s $M$-estimator.
- Some estimators such as Tyler’s $M$-estimator do not satisfy M2**.
Existence and uniqueness revisited

\[ \mathcal{L}_n(\Sigma) = \frac{1}{n} \sum_{i=1}^{n} \rho(z_i^H \Sigma^{-1} z_i) - \ln |\Sigma^{-1}| \]

where \( \rho(t) = -\ln g(t) \)

**M2**\* \( \psi(t) = t \varphi(t) \) is nondecreasing

Suppose M1, M2\* and M3 hold. Then,

- \( \hat{\Sigma} \) exists and corresponds to a solution of the estimating equation.
- Furthermore, if both \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) satisfy the estimating equation then
  \[ \hat{\Sigma}_1 \propto \hat{\Sigma}_2 \]

(uniqueness property under proportional scatter matrices)
Outline

1. ML-estimators of scatter
   - Definition
   - Existence, Uniqueness, Convergence

2. $M$-estimators of scatter
   - Complex Huber’s $M$-estimator
   - Complex Tyler’s $M$-estimator

3. Robustness and asymptotics
   - The influence function
   - Asymptotics
Generalization of ML-estimators of the scatter matrix parameter.

Real-valued case by [Maronna, 1976]
- Among the first proposals for robust covariance matrix estimators
- Extensively studied in the statistics literature; see e.g. [Huber, 1981, Tyler, 1987, Tyler, 1988, Kent and Tyler, 1991]

Complex $M$-estimators of scatter studied mainly in the engineering literature; See bibliography in [CES, 2012].

Examples:
- Complex Tyler’s $M$-estimator
- Complex Huber’s $M$-estimator
- For more examples, see [CES, 2012] and references therein.
**M-estimators of scatter**

A positive definite Hermitian $m \times m$ matrix that solves the $M$-estimating equation

$$
\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \varphi(z_i^H \hat{\Sigma}^{-1} z_i) z_i z_i^H,
$$

where $\varphi$ is any real-valued *weight function* on $[0, \infty)$.

- $\varphi$ need not be related to the density of a CES distribution $\Rightarrow$ the class include the MLE’s of the scatter parameter $\Sigma$.
- Existence, uniqueness and computation of the $M$-estimates of scatter follows from previous result (note: M2 can be relaxed by M2*).
Complex Huber’s $M$-estimator

- Introduced in [Ollila and Koivunen, 2003].
- $\hat{\Sigma} \in \mathcal{H}_m$ for chosen $0 < q < 1$ is a solution to

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \varphi(z_i^H \hat{\Sigma}^{-1} z_i) z_i z_i^H$$

with

$$\varphi(t) = \begin{cases} 
1/b, & \text{for } t \leq c^2 \\
\frac{c^2}{tb}, & \text{for } t > c^2 
\end{cases}$$

where $c$ is a tuning constant s.t. $q = F_{\chi^2_{2m}}(2c^2)$.

- The scaling factor $b$ is chosen s.t. $\hat{\Sigma}$ is consistent to the covariance matrix for Gaussian data,

- $q \to 1 \implies$ SCM. If $q \to 0$ then $\hat{\Sigma}$ approaches Tyler’s $M$-estimator.
- Existence, Uniqueness and computation: Theorem 7(b) applies.
Fig: Weight function $\varphi(t)$ for different choices of $q$
Complex Tyler’s $M$-estimator

- A popular distribution-free $M$-estimator introduced in [Tyler, 1987] (real case)

Complex Tyler’s $M$-estimator

Weight function $\varphi(t) = p/t$: $\hat{\Sigma} \in \mathcal{H}_m$ is a solution to estimating eq.:

$$
\hat{\Sigma} = \frac{m}{n} \sum_{i=1}^{n} \frac{z_i z_i^H}{z_i^H \hat{\Sigma}^{-1} z_i},
$$

- Not related to any elliptical density.
- Defined up to a scale: if $\hat{\Sigma}$ is a solution, then so is $c\hat{\Sigma}$ for $c > 0$.
  - For uniqueness, we consider a solution verifying $\text{Tr}(\hat{\Sigma}) = m$
  - Essentially, $\hat{\Sigma}$ is an estimator of the shape matrix $\mathbf{V} = m\Sigma / \text{Tr}(\Sigma)$.
ML interpretations of Tyler’s $M$-estimator

- Many authors have arrived to Tyler’s $M$-estimator from different likelihood perspectives.
- We can think (at least) 3 ML-interpretations.

**Case A:** MLE for complex angular Gaussian data $\{z_i\}_{i=1}^{n}$ iid $\sim \text{CAG}(0, \Sigma)$

The MLE $\hat{\Sigma}$ minimizing the negative log-likelihood fnc (divided by $n$)

$$\mathcal{L}_n^*(\Sigma) = \frac{1}{n} \sum_{i=1}^{n} \rho(t) = p \ln(t)$$

$$= p \ln(t) = p \ln t$$

$$= p \ln t$$

is Tyler’s $M$-estimator.  

[Kent, 1997]
Case B. MLE for independent observations having
- with Gaussian distribution
- and proportional scatter matrices

Formally, Tyler’s $M$-estimator $\hat{\Sigma}$ is the MLE of the shape matrix $\Sigma$ (verifying $\text{Tr}(\Sigma) = m$) of an independent sample:

$$z_i \sim \mathbb{C}N_m(0, \eta_i \Sigma)$$

possessing the same unknown shape matrix $\Sigma$ but different unknown scales $\eta_i > 0$

[Gini and Greco, 2002, Conte et al., 2002]

- This result can be extended.
- Provides a formal explanation why the Tyler’s $M$-estimator often offers good performance in practical applications.
- EX: in radar applications, the secondary data is often inhomogeneous in nature.
Case C. MLE for independent observations having
- possibly different elliptical distributions
- and proportional scatter matrices

Formally, Tyler’s $M$-estimator $\hat{\Sigma}$ is the MLE of the shape matrix $\Sigma$ (verifying $\text{Tr}(\Sigma) = m$) of an independent sample:

$$z_i \sim \text{CE}_m(0, \eta_i \Sigma, g_i)$$

possessing the same unknown shape matrix $\Sigma$ but different unknown scales $\eta_i > 0$ and possibly different density generator $g_i(t)$ for $i = 1, \ldots, n$ [Ollila and Tyler, 2012]

- This result can be extended.
- Provides a formal explanation why the Tyler’s $M$-estimator often offers good performance in practical applications.
- EX: in radar applications, the secondary data is often inhomogeneous in nature.
Algorithm

\[ \Sigma_{k+1} = \frac{m}{n} \sum_{i=1}^{n} \frac{z_i z_i^H}{z_i^H \Sigma_k^{-1} z_i} \]

\[ \Sigma_{k+1} \leftarrow m \Sigma_{k+1} / \text{Tr}(\Sigma_{k+1}) \]

converges to \( \hat{\Sigma} \) which exist and is unique up to a positive scalar under the assumption [CES, 2012]:

\( \mathbb{P}_{2n}(\{0\}) = 0 \) and for all linear subspaces \( V \subset \mathbb{R}^{2m} \), with

\[ 0 < \text{dim}(V) \leq 2m - 1, \quad \mathbb{P}_{2n}(V) < \text{dim}(V)/(2m). \]

Result follows from [Kent and Tyler, 1988]; See also [Kent, 1997, Pascal et al., 2008].

Note: normalization not needed: simply normalize after convergence.
Algorithm

\[ \Sigma_{k+1} = \frac{m}{n} \sum_{i=1}^{n} \frac{z_i z_i^H}{z_i^H \Sigma_k^{-1} z_i} \]

\[ \Sigma_{k+1} \leftarrow m \Sigma_{k+1} / \text{Tr}(\Sigma_{k+1}) \]

converges to \( \hat{\Sigma} \) which exist and is unique up to a positive scalar under the assumption [CES, 2012]:

**M3'** \( \mathbb{P}_{2n}(\{0\}) = 0 \) and for all linear subspaces \( V \subset \mathbb{R}^{2m} \), with \( 0 < \dim(V) \leq 2m - 1 \), \( \mathbb{P}_{2n}(V) < \dim(V)/(2m) \).

Result follows from [Kent and Tyler, 1988]; See also [Kent, 1997, Pascal et al., 2008].

**Note:** normalization not needed: simply normalize after convergence.
Algorithm

$$\Sigma_{k+1} = \frac{m}{n} \sum_{i=1}^{n} \frac{z_i z_i^H}{z_i^H \Sigma_k^{-1} z_i}$$

converges to $\hat{\Sigma}$ which exist and is unique up to a positive scalar under the assumption [CES, 2012]:

**M3’** $\mathbb{P}_{2n}(\{0\}) = 0$ and for all linear subspaces $V \subset \mathbb{R}^{2m}$, with $0 < \dim(V) \leq 2m - 1$, $\mathbb{P}_{2n}(V) < \dim(V)/(2m)$.

Result follows from [Kent and Tyler, 1988]; See also [Kent, 1997, Pascal et al., 2008].

**Note:** normalization not needed: simply normalize after convergence.
Outline

1. ML-estimators of scatter
   - Definition
   - Existence, Uniqueness, Convergence

2. $M$-estimators of scatter
   - Complex Huber’s $M$-estimator
   - Complex Tyler’s $M$-estimator

3. Robustness and asymptotics
   - The influence function
   - Asymptotics
The \( M \)-functional of scatter matrix

\[
\Sigma_\varphi(F) \in \mathcal{H}_m \text{ is defined as a solution of }
\]

\[
\Sigma_\varphi(F) = \mathbb{E}[\varphi(z^H \Sigma^{-1}_\varphi(F) z)zz^H]
\]

Fisher Consistency:

At \( F = CE_m(0, \Sigma, g) \) distribution:

\[
\Sigma_\varphi(F) = \sigma_\varphi \Sigma,
\]

where the scalar factor \( \sigma_\varphi > 0 \) solves

\[
\mathbb{E}[\varphi(Q/\sigma_\varphi)Q/\sigma_\varphi] = m
\]

Often \( \sigma_\varphi \) needs to be solved numerically, although in some cases (e.g. Huber’s \( M \)-estimator) an analytical expression can be derived.
The influence function

The **influence function** of $M$-functional $\Sigma_\varphi$ at $F = CE_m(0, \Sigma, g)$

\[
\text{IF}(\xi; \Sigma_\varphi, F) = \lim_{\varepsilon \downarrow 0} \frac{\Sigma_\varphi \left( (1 - \varepsilon)F(z) + \varepsilon \Delta_\xi(z) \right) - \Sigma_\varphi (F)}{\varepsilon} = \frac{\partial}{\partial \varepsilon} \Sigma_\varphi \left( (1 - \varepsilon)F(z) + \varepsilon \Delta_\xi(z) \right) \bigg|_{\varepsilon = 0}
\]

- The IF describes the effect of an *infinitesimal* contamination at a point $\xi$ on the estimator, standardized by the mass of the contamination.
- A robust estimator should have a bounded and continuous IF
  - **boundedness** implies that a small amount of contamination at any point $\xi$, does not have an arbitrarily large influence on the estimator
  - **continuity** implies that the small changes in the data set cause only small changes in the estimator.
The influence function

The influence function of $M$-functional $\Sigma_\varphi$ at $F = CE_m(0, \Sigma, g)$

$$\text{IF}(\xi; \Sigma_\varphi, F) = \lim_{\varepsilon \downarrow 0} \frac{\Sigma_\varphi \left( (1 - \varepsilon) F(z) + \varepsilon \Delta_\xi(z) \right) - \Sigma_\varphi (F)}{\varepsilon}$$

$$= \frac{\partial}{\partial \varepsilon} \Sigma_\varphi \left( (1 - \varepsilon) F(z) + \varepsilon \Delta_\xi(z) \right) \bigg|_{\varepsilon = 0}$$

- The IF describes the effect of an infinitesimal contamination at a point $\xi$ on the estimator, standardized by the mass of the contamination.
- A robust estimator should have a bounded and continuous IF
  - boundedness implies that a small amount of contamination at any point $\xi$ does not have an arbitrarily large influence on the estimator
  - continuity implies that the small changes in the data set cause only small changes in the estimator.
The influence function

The **influence function** of $M$-functional $\Sigma_\varphi$ at $F = CE_m(0, \Sigma, g)$

\[
\text{IF}(\xi; \Sigma_\varphi, F) = \lim_{\varepsilon \downarrow 0} \frac{\Sigma_\varphi((1 - \varepsilon)F(z) + \varepsilon \Delta_\xi(z)) - \Sigma_\varphi(F)}{\varepsilon} = \frac{\partial}{\partial \varepsilon} \Sigma_\varphi((1 - \varepsilon)F(z) + \varepsilon \Delta_\xi(z))\bigg|_{\varepsilon=0}
\]

- The IF describes the effect of an *infinitesimal* contamination at a point $\xi$ on the estimator, standardized by the mass of the contamination.
- A robust estimator should have a bounded and continuous IF
  - boundedness implies that a small amount of contamination at any point $\xi$ does not have an arbitrarily large influence on the estimator
  - continuity implies that the small changes in the data set cause only small changes in the estimator.
Assumptions

We need the condition

\( N \) The function \( t\psi'(t) \) is bounded

- Condition \( N \) implies that the expectation

\[
c_\phi \triangleq \mathbb{E} \left[ \varphi' \left( \frac{Q}{\sigma_\phi} \right) \frac{Q^2}{\sigma_\phi^2} \right]
\]

exists.

- Recall that \( Q = \mathcal{R}^2 \) has p.d.f. \( f_Q(t) = C t^{m-1} g(t) \) (\( C \) a norm. cnst.)

- Next theorem generalizes the results in [Maronna, 1976] and [Huber, 1981] to the complex case.
Main result

**Theorem 8** [CES, 2012] Suppose that the condition N holds. Then, the IF of $\Sigma_\varphi(\cdot)$ at $F = CE_m(0, \Sigma, g)$ exists and is given by

$$\text{IF}(\xi; \Sigma_\varphi, F) = w_1(t)\Sigma^{1/2}(uu^H - m^{-1}I)\Sigma^{1/2} + w_2(t)\Sigma,$$

where $t = \xi^H\Sigma^{-1}\xi$, $u = \Sigma^{-1/2}\xi/\sqrt{t}$ and

$$w_1(t) = \frac{\varphi(t/\sigma_\varphi)t}{1 + [m(m + 1)]^{-1}c_\varphi}, \quad w_2(t) = \frac{\varphi(t/\sigma_\varphi)t - m\sigma_\varphi}{m + c_\varphi}.$$

⇒ The IF is continuous and bounded iff $\psi(t) = \varphi(t)t$ is continuous and bounded.

EX SCM ($\varphi \equiv 1$) has unbounded IF, i.e., it is not robust. This is also true for MLE of Generalized Gaussian distribution; See [CES, 2012].
Main result

Theorem 8 [CES, 2012] Suppose that the condition N holds. Then, the IF of \( \Sigma \varphi(\cdot) \) at \( F = \text{CE}_m(0, \Sigma, g) \) exists and is given by

\[
\text{IF}(\xi; \Sigma \varphi, F) = w_1(t)\Sigma^{1/2}(uu^\text{H} - m^{-1}I)\Sigma^{1/2} + w_2(t)\Sigma,
\]

where \( t = \xi^\text{H} \Sigma^{-1} \xi, \ u = \Sigma^{-1/2} \xi / \sqrt{t} \) and

\[
w_1(t) = \frac{\varphi(t/\sigma_\varphi)t}{1 + [m(m+1)]^{-1}c_\varphi}, \quad w_2(t) = \frac{\varphi(t/\sigma_\varphi)t - m\sigma_\varphi}{m + c_\varphi}.
\]

⇒ The IF is continuous and bounded iff \( \psi(t) = \varphi(t)t \) is continuous and bounded.

EX SCM (\( \varphi \equiv 1 \)) has unbounded IF, i.e., it is not robust. This is also true for MLE of Generalized Gaussian distribution; See [CES, 2012].
Asymptotic normality of scatter matrices

**Theorem 9** [CES, 2012]. If conditions M1, M2* and N holds. Then

\[ \sqrt{n} \text{vec}(\hat{\Sigma} - \Sigma_\varphi) \xrightarrow{d} \mathcal{CN}_{m^2}(0, C, P), \]

where the asymptotic covariance and pseudo-covariance matrices are

\[ C = \vartheta_1 (\Sigma^* \otimes \Sigma) + \vartheta_2 \text{vec}(\Sigma)\text{vec}(\Sigma)^H, \]

\[ P = \vartheta_1 (\Sigma^* \otimes \Sigma)K_{m,m} + \vartheta_2 \text{vec}(\Sigma)\text{vec}(\Sigma)^\top, \]

where \( K_{m,m} \) denotes the commutation matrix.

The constants \( \vartheta_1 > 0 \) and \( \vartheta_2 \geq -\vartheta_1/m \) are given by

\[ \vartheta_1 = \frac{\mathbb{E}[\varphi^2(Q/\sigma_\varphi)Q^2]}{m(m+1)(1 + [m(m+1)]^{-1}c_\varphi)^2} \]

and

\[ \vartheta_2 = \frac{\mathbb{E}\left[(\varphi(Q/\sigma_\varphi)Q - m\sigma_\varphi)^2\right]}{(m + c_\varphi)^2} - \vartheta_1 \frac{1}{m} \]
Asymptotic normality of shape matrices

\[ \hat{V} = \frac{m \hat{\Sigma}}{\text{Tr}(\hat{\Sigma})} \quad \text{and} \quad V = \frac{m \Sigma}{\text{Tr}(\Sigma)} \]

[CES, 2012]. If conditions M1, M2* and N holds. Then

\[ \sqrt{n} \text{vec}(\hat{V} - V) \xrightarrow{d} \mathcal{CN}_{m^2}(0, C_V, P_V), \]

where the asymptotic covariance and pseudo-covariance matrices are

\[ C_V = \frac{\vartheta_1}{\sigma^2_\phi} \Gamma_V (V^* \otimes V) \Gamma^H_V \]

\[ P_V = \frac{\vartheta_1}{\sigma^2_\phi} \Gamma_V (V^* \otimes V) K_{m,m} \Gamma^\top_V \]

where \( K_{m,m} \) and \( \vartheta_1 > 0 \) are as in Theorem 9.
Image analysis with GG distributions

The following results can be found in:

Image processing:


Images are filtered by a stationary wavelet ⇒ observed vector $z$ contains the realizations of the wavelet coefficients for each channel of the RGB image.
Image analysis with GG distributions

Figure: Images from the VisTex database. (a) Bark.0000 and (b) Leaves.0008.
Table: Estimated MGGD parameters for the first subband of the Bark.0000 and Leaves.0008 images.

<table>
<thead>
<tr>
<th>Image</th>
<th>$\hat{\nu}$</th>
<th>$\hat{s}$</th>
<th>$\hat{\Sigma}$</th>
</tr>
</thead>
</table>
| 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     | \[
\begin{bmatrix}
0.935 & 0.966 & 0.871 \\
0.966 & 1.074 & 0.976 \\
0.871 & 0.976 & 0.991 \\
\end{bmatrix}
\] |
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).
List of references

Please see a more complete list in [CES, 2012]
**Conte, E., De Maio, A., and Ricci, G. (2002).**
Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection.


**Gini, F. and Greco, M. (2002).**
Covariance matrix estimation for CFAR detection in correlated heavy-tailed clutter.


**Huber, P. J. (1981).**

*R robust Statistics.*

Wiley, New York.

**Kent, J. T. (1997).**

Data analysis for shapes and images.


**Kent, J. T. and Tyler, D. E. (1988).**

Maximum likelihood estimation for the wrapped Cauchy distribution.

Redescending M-estimates of multivariate location and scatter.

Robust M-estimators of multivariate location and scatter.

Influence functions for array covariance matrix estimators.
In *Proc. IEEE Workshop on Statistical Signal Processing (SSP’03)*, pages 445–448, St. Louis, MO, USA.

Distribution-free detection under complex elliptically symmetric clutter distribution.

Covariance structure maximum-likelihood estimates in compound Gaussian noise: existence and algorithm analysis.


A distribution-free M-estimator of multivariate scatter.


Some results on the existence, uniqueness, and computation of the M-estimates of multivariate location and scatter.

Contents

- **Part A**
  CES distributions

- **Part B**
  ML and $M$-estimators of scatter matrix

- **Part C**
  Applications

- **Part D**
  $M$-estimators of scatter in the large dimensional regime
Part C

Applications
Part C: Contents

1 Preliminaries
   ■ Motivations
   ■ Reminders
   ■ An important property

2 Detection
   ■ Problem statement and definitions
   ■ The ANMF and its properties
   ■ Simulations

3 Radar applications: Doppler detection/estimation, STAP
   ■ Experimentation on real data - Doppler
   ■ STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
Key references of Part C


Outline

1 Preliminaries
   ■ Motivations
   ■ Reminders
   ■ An important property

2 Detection
   ■ Problem statement and definitions
   ■ The ANMF and its properties
   ■ Simulations

3 Radar applications: Doppler detection/estimation, STAP
   ■ Experimentation on real data - Doppler
   ■ STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
Motivations

- 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...
- The ultimate purpose is to characterize the system performance, not only the estimation performance $\Rightarrow$ ROC curves, probability of detection vs SNR, false alarm regulation, MSE characterization...
Motivations
Why CES and $M$-estimation? Examples in Radar processing

Classical radar applications consider the background to be Gaussian.

- 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 may give poor results.
Grazing angle Radar

- Impulsive Clutter
- Spatial heterogeneity (e.g. in SAR or HS images)

High Resolution Radar

- Small number of scatters in the Cell Under Test (CUT)
- Central Limit Theorem (CLT) is not valid anymore
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
⇒ Need/Use of CES distributions
⇒ Need/Use 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
Outline

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
Reminder: Asymptotic distribution of complex $M$-estimators

Using the results of Tyler (1982), we derived the following results (Mahot, 2013):

\[ \sqrt{n} \text{vec}(\hat{\Sigma} - \Sigma) \xrightarrow{d} \mathcal{CN}_{m^2}(0, C, P), \]  

(1)

where $\mathcal{CN}$ is the complex Gaussian distribution, $C$ the CM and $P$ the pseudo CM:

\[
C = \sigma_1 (\Sigma^* \otimes \Sigma) + \sigma_2 \text{vec}(\Sigma)\text{vec}(\Sigma)^H, \\
P = \sigma_1 (\Sigma^* \otimes \Sigma) K + \sigma_2 \text{vec}(\Sigma)\text{vec}(\Sigma)^T,
\]

where $K$ is the commutation matrix and where the constant $\sigma_1$ and $\sigma_1$ have been defined in Part B.
Reminder: asymptotic distribution of the SCM under Gaussian assumptions

The SCM is defined as

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^H$$

where $z_i$ are complex independent circular zero-mean Gaussian with covariance matrix $\Sigma$.

Then,

$$\sqrt{n} \text{vec}(\hat{S}_n - \Sigma) \xrightarrow{d} \mathcal{CN}(0, C, P)$$

where

$$C = (\Sigma^* \otimes \Sigma)$$

$$P = (\Sigma^* \otimes \Sigma) K_{m,m}$$
Outline

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
An important property of complex $M$-estimators

Let $\hat{\Sigma}$ an estimate of Hermitian positive-definite matrix $\Sigma$ that satisfies

$$\sqrt{n} \left( \text{vec}(\hat{\Sigma} - \Sigma) \right) \xrightarrow{d} \mathcal{CN} (0, C, P), \quad (2)$$

with

$$\left\{ \begin{array}{l}
C = \nu_1 \Sigma^* \otimes \Sigma + \nu_2 \text{vec}(\Sigma) \text{vec}(\Sigma)^H, \\
P = \nu_1 (\Sigma^* \otimes \Sigma) K_{m,m} + \nu_2 \text{vec}(\Sigma) \text{vec}(\Sigma)^T,
\end{array} \right.$$  

where $\nu_1$ and $\nu_2$ are any real numbers.
<table>
<thead>
<tr>
<th></th>
<th>SCM</th>
<th>$M$-estimators</th>
<th>FP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_1$</td>
<td>1</td>
<td>$\sigma_1$</td>
<td>$(m + 1)/m$</td>
</tr>
<tr>
<td>$\nu_2$</td>
<td>0</td>
<td>$\sigma_2$</td>
<td>$-(m + 1)/m^2$</td>
</tr>
<tr>
<td>...</td>
<td>More accurate</td>
<td></td>
<td>More robust</td>
</tr>
</tbody>
</table>

Let $H(V)$ be a $r$-multivariate function on the set of Hermitian positive-definite matrices, with continuous first partial derivatives and such as $H(V) = H(\alpha V)$ for all $\alpha > 0$, e.g. the ANMF statistic, the MUSIC statistic.
Theorem 1 (Asymptotic distribution of $H(\hat{\Sigma})$)

$$\sqrt{n} \left( H(\hat{\Sigma}) - H(\Sigma) \right) \xrightarrow{d} \mathcal{CN}(0, 1, C_H, P_H)$$

(3)

where $C_H$ and $P_H$ are defined as

$$C_H = \nu_1 H'(\Sigma)(\Sigma^T \otimes \Sigma) H'(\Sigma)^H,$$
$$P_H = \nu_1 H'(\Sigma)(\Sigma^T \otimes \Sigma) K_{m,m} H'(\Sigma)^T,$$

where $H'(\Sigma) = \left( \begin{array}{c} \frac{\partial H(\Sigma)}{\partial \text{vec}(\Sigma)} \end{array} \right)$. 

Frederic Pascal  
Covariance Matrix Estimation and Applications in Radar
Some comments:

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

\( H(\text{SCM}) \) and \( H(\text{M-estimators}) \) share the same asymptotic distribution (differs from \( \sigma_1 \))

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

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
Problem Statement


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

$$\begin{align*}
\text{Hypothesis } H_0: \quad & y = z \quad y_i = z_i \quad i = 1, \ldots, n \\
\text{Hypothesis } H_1: \quad & y = s + z \quad y_i = z_i \quad i = 1, \ldots, n
\end{align*}$$

where the $z_i$'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 $\Lambda(y)$ and a detection threshold $\lambda$:

$$\Lambda(y) = \frac{p_y(y/H_1)}{p_y(y/H_0)} \frac{H_1}{H_0} \lambda,$$

$\lambda$ is obtained for a given $PFA$ (set by the user):

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

$$PFA = \mathbb{P}(\Lambda(y) > \lambda/H_0)$$

- **Probability of Detection** (to evaluate the performance):

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

for different Signal-to-Noise Ratio (SNR).
Detection under Gaussian/non-Gaussian assumption

- **Gaussian case (OGD):** if \( z \sim \mathcal{CN}(0, \Sigma) \) then

\[
\Lambda(y) = \frac{|p^H\Sigma^{-1}y|^2}{p^H\Sigma^{-1}p} \begin{cases} H_1 \\ \geq \lambda_g \\ H_0 \end{cases}
\]

with \( \lambda_g = \sqrt{-\ln(PFA)} \) et \( p_z(z) = \frac{1}{(\pi^m|\Sigma|} \exp(-z^H\Sigma^{-1}z). \)

- **Heterogeneous case (NMF):**

\[
\Lambda(y) = \frac{|p^H\Sigma^{-1}y|^2}{(p^H\Sigma^{-1}p)(y^H\Sigma^{-1}y)} \begin{cases} H_1 \\ \geq \lambda_{NMF} \\ H_0 \end{cases}
\]

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: $\Sigma$ unknown $\Rightarrow$ Adaptive detection

Gaussian model $\Rightarrow$ $\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^H$

- AMF test [1]

$$\Lambda_{AMF}(y) = \frac{\left| p^H \hat{S}_n^{-1} y \right|^2}{\left( p^H \hat{S}_n^{-1} p \right)} \begin{cases} \frac{H_1}{H_0} & \lambda_{AMF} \end{cases}$$


- Kelly test [2]

$$\Lambda_{Kelly}(y) = \frac{\left| p^H \hat{S}_n^{-1} y \right|^2}{\left( p^H \hat{S}_n^{-1} p \right) \left( N + y^H \hat{S}_n^{-1} y \right)} \begin{cases} \frac{H_1}{H_0} & \lambda_{Kelly} \end{cases}$$

Outline

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
CES distribution $\Rightarrow$ ANMF

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

$$\Lambda_{ANMF}(y, \hat{\Sigma}) = \frac{|p^H \hat{\Sigma}^{-1} y|^2}{(p^H \hat{\Sigma}^{-1} p)(y^H \hat{\Sigma}^{-1} y)}$$

$$\begin{cases} H_1 & \Lambda_{ANMF} \geq \lambda_{ANMF} \\ H_0 & \Lambda_{ANMF} < \lambda_{ANMF} \end{cases}$$

(6)

where $\hat{\Sigma}$ stands for any estimators presented before: SCM, $M$-estimators, Tyler’s estimator...

One has, conditionally to $y$, $\Lambda(\hat{\Sigma}) = \Lambda(\alpha \hat{\Sigma})$ for any $\alpha > 0$.


The ANMF is **scale-invariant**, i.e.
\[ \forall \alpha, \beta \in \mathbb{R}, \Lambda_{ANMF}(\alpha y, \beta \hat{\Sigma}) = \Lambda_{ANMF}(y, \hat{\Sigma}) \]

- Its **asymptotic distribution** (conditionally to \( y \)) is known (tks to theorem 1)

Considering \( \Lambda_{ANMF}(y, \hat{\Sigma}) \) conditionally to \( y \), i.e. \( \Lambda_{ANMF}(\hat{\Sigma}) \), allows to directly apply theorem 1. 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)
Illustration of the CFAR properties

False Alarm regulation

**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** $\Lambda_{ANMF}(\hat{S}_n)$ **(Gaussian case, finite n)**

\[
P_{fa} = (1 - \lambda)^{a-1} 2F_1(a, a - 1; b - 1; \lambda),
\]

where $a = n - m + 2$, $b = n + 2$ and $2F_1$ is the Hypergeometric function defined as

\[
2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{\infty} \frac{\Gamma(a + k)\Gamma(b + k)}{\Gamma(c + k) k!} x^k
\]

Comments

Three possible approaches to characterize the performance:

- Use the (very) poor approximation of the FA regulation of the NMF
- Use the asymptotics of theorem 2 (but it is conditionally to the dist. of $y$) ⇒ a slight loss of performance
- Combine the asymptotics of theorem 9 of Part B and the finite-distance result on PFA-threshold...

From theorem 9 (Part B), one has

**PFA-threshold relation of $\Lambda_{ANMF}(M\text{-est.})$ for CES distributions**

For $n$ large enough and for any elliptically distributed noise, the PFA is still given by (7) if we replace $n$ by $n/\sigma_1$.

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

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
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
Tyler’s estimator: Gaussian context, $n = 10$, $m = 3$

**PFA-threshold relation of $\Lambda_{ANMF}(Tyler’s est.)$ for CES distributions**

For $n$ large and any elliptically distributed noise, the PFA is still given by (7) if we replace $n$ by $n/\frac{m+1}{m}$.

![Graph showing the PFA-threshold relation for Tyler's estimator](image-url)
Comments

Conclusions on the detection part:

Accurate approximation of the (theoretical) FA regulation

Cost: having a little bit more data: \( \sigma_1 n \) instead of \( n \).

This \( \sigma_1 \) can be interpreted as the loss brought by robust estimators compared to optimal Gaussian estimator **BUT** performance stability of the robust estimators in various distributions contexts.
Outline

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
"Range-azimuth" map from ground clutter data collected by a radar from THALES Air Defence, placed 13 meters above ground and illuminating area at low grazing angle.

Ground clutter complex echoes collected in $n = 868$ range bins for 70 different azimuth angles and for $m = 8$ pulses.
Data processing

Figure: Ground clutter data level (in dB) corresponding to the 1\textsuperscript{st} pulse (previous map in 3 dimensions)
Data processing

- Rectangular CFAR mask $5 \times 5$ for different steering vectors $p$.

For each $y$, computation of associated detector $\Lambda_{ANMF}(\hat{\Sigma}_{Tyler})$.

- Mask moving all over the map.

$$p = \begin{pmatrix}
1 \\
\exp \left( \frac{2i\pi(k-1)}{m} \right) \\
\exp \left( \frac{2i\pi(k-1)2}{m} \right) \\
\vdots \\
\exp \left( \frac{2i\pi(k-1)(m-1)}{m} \right)
\end{pmatrix}$$
Figure: False alarm regulation for $\mathbf{p} = (1 \ldots 1)^T$
Outline

1 Preliminaries
   ■ Motivations
   ■ Reminders
   ■ An important property

2 Detection
   ■ Problem statement and definitions
   ■ The ANMF and its properties
   ■ Simulations

3 Radar applications: Doppler detection/estimation, STAP
   ■ Experimentation on real data - Doppler
   ■ STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
Space Time Adaptive Processing: Principles

(a) STAP principles

\[ p(\theta, f_d) = \begin{pmatrix} 1 \\ \exp(-2i\pi d \sin(\theta)/\lambda) \\ \vdots \\ \exp(-2i\pi (K - 1) d \sin(\theta)/\lambda) \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \exp(-2i\pi f_d T_r) \\ \vdots \\ \exp(-2i\pi f_d (M - 1) T_r) \end{pmatrix} \]

(b) STAP datacube
Data

Problem: estimate the position (angle) and the Doppler frequency (speed) of the target ⇒ use of the ANMF with a particular steering vector

Data parameters: real clutter with synthetic target

X-Band ≃ 10e9, wavelength 0.03, flight speed 100m/s, distance to the scene 30km, 5 of incidence, PRF (Pulse Repetition Frequency) of 1 kHz, inter-sensor distance 0.3m, 12 trials with 410 range bins, 64 pulses and 4 sensors.

This means observations of size $m = 256$ while $n \leq 410$!

Clutter more or less homogeneous BUT targets are present in the secondary data
No target is present in the secondary data - homogeneous noise

Figure: Doppler-angle map for the range bin 255 with $n = 404$ secondary data (targets and guard cells are removed) and $m = 256$

(c) AMF detector with the SCM

(d) ANMF detector with Huber’s est. (parameter $q = 0.6$)
Preliminaries
Detection
Radar applications: Doppler detection/estimation, STAP
DoA estimation
Hyperspectral Imaging (ongoing work)

No target is present in the secondary data - homogeneous noise

(a) AMF detector with the Student est. (parameter $\nu = 2$)
(b) ANMF detector with Tyler’s est.

Figure: Doppler-angle map for the range bin 255 with $n = 404$ secondary data (targets and guard cells are removed) and $m = 256$
Two targets are present in the secondary data - homogeneous noise

(a) AMF detector with the SCM (parameter $\nu = 2$)

(b) ANMF detector with Huber’s est. (parameter $q = 0.6$)

**Figure:** Doppler-angle map for the range bin 255 with $n = 404$ secondary data (guard cells are removed) and $m = 256$
Two targets are present in the secondary data - homogeneous noise

(a) AMF detector with the Student est. (parameter $\nu = 2$)
(b) ANMF detector with Tyler’s est.

Figure: Doppler-angle map for the range bin 255 with $n = 404$ secondary data (guard cells are removed) and $m = 256$
Comments

$M$-estimators and Tyler’s estimators continue to ”well” perform when dimension is high (even if it requires a matrix inversion)

[already said] $M$-estimators and Tyler’s estimators provide good performance in homogeneous contexts (close to Gaussian distribution)

$M$-estimators and Tyler’s estimators are robust to outliers (link to the robustness criteria: breakdown point and influence function)
Outline

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
MUltiple Signal Classification (MUSIC) method for DoA estimation

- $K$ (known) direction of arrival $\theta_k$ on $m$ antennas
- Gaussian stationary narrowband signal with additive noise.
- the DoA is estimated from $n$ snapshots, using the SCM, the Huber’s $M$-estimator and the Tyler’s estimator.

$$y(t) = A(\theta_0)s(t) + w(t)$$

- $\theta_0 = (\theta_1 \ \theta_2 \ \ldots \ \theta_K)^T$
- the steering matrix $A(\theta_0) = (a(\theta_1) \ a(\theta_2) \ \ldots \ a(\theta_K))$
- $s(t) = (s_1(t) \ s_2(t) \ \ldots \ s_K(t))^T$ signal vector,
- $w(t)$ stationary additive noise.
\[ \Sigma = \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{A}(\theta_0)\mathbb{E}[\mathbf{s}\mathbf{s}^H]\mathbf{A}^H(\theta_0) + \sigma^2\mathbf{I} \]

which can be rewritten

\[ \Sigma = \mathbb{E}[\mathbf{y}\mathbf{y}^H] = E_S D_S E_S^H + \sigma^2 E_W E_W^H. \]

where \( E_S \) (resp. \( E_W \)) are the signal (resp. noise) subspace eigenvectors. The MUSIC statistic is

\[
\begin{aligned}
    H(\Sigma) &= \gamma(\theta) = s(\theta)^H E_W E_W^H s(\theta), & (\Sigma \text{ known}) \\
    H(\hat{\Sigma}) &= \hat{\gamma}(\theta) = \sum_{i=1}^{m-K} \lambda_i s(\theta)^H \hat{e}_i \hat{e}_i^H s(\theta) = H(\alpha \hat{\Sigma}), & (\Sigma \text{ unknown})
\end{aligned}
\]

where \( \lambda_i \) (resp. \( \hat{e}_i \)) are the eigenvalues (resp. eigenvectors) of \( \hat{\Sigma} \).

This function respects assumptions of theorem 1!
Simulation using the MUltiple Signal Classification (MUSIC) method

The Mean Square Error (MSE) between the estimated angle $\hat{\theta}$ and the real angle $\theta$ 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 $20^\circ$ plus additive noise.
- the DoA is estimated from $n$ snapshots, using the SCM, the Huber’s $M$-estimator and the Tyler’s estimator.
Preliminaries
Detection
Radar applications: Doppler detection/estimation, STAP
DoA estimation
Hyperspectral Imaging (ongoing work)

(a) White additive Gaussian 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...
Outline

1 Preliminaries
   - Motivations
   - Reminders
   - An important property

2 Detection
   - Problem statement and definitions
   - The ANMF and its properties
   - Simulations

3 Radar applications: Doppler detection/estimation, STAP
   - Experimentation on real data - Doppler
   - STAP

4 DoA estimation

5 Hyperspectral Imaging (on going work)
Hyperspectral Imaging: motivations

Figure: Hyperspectral Data - Indian Pines data set - $m = 100$ wavelengths

Figure: Empirical distributions of different physical elements

Heavy tailed distributions $\Rightarrow$ CES framework
Hyperspectral Imaging: context and difficulties

**Problem**

Now, the statistical mean is non null ⇒ \( M \)-estimator of the mean is required

\[
\hat{\mu} = \frac{\sum_{i=1}^{n} u_1(t_i) z_i}{\sum_{i=1}^{n} u_1(t_i)} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} u_2\left(t_i^2\right) (z_i - \hat{\mu}) (z_i - \hat{\mu})^H,
\]

where \( t_i = \left((z_i - \hat{\mu})^H \hat{\Sigma}^{-1} (z_i - \hat{\mu})\right)^{1/2} \) and \( u_1(.) \), \( u_2(.) \) denote any real-valued weight functions (following the conditions of Maronna).

⚠️ No proofs of existence, uniqueness, consistency and convergence of the recursive algorithm!
Methodology

- Rectangular CFAR mask $k \times k$ for different steering vectors $p$.
- For each $y$, computation of the detector $\Lambda_{ANMF}(\hat{\Sigma})$.
- Mask moving all over the hyperspectral image.

Assumptions

- Pixels of the mask are statistically independent, i.e. spatially independence.
- Pixels of the mask are identically distributed.

FA regulation proved in non-zero mean Gaussian case

Hyperspectral Imaging - False Alarm regulation

Figure: Probability of false alarm versus the detection threshold for $m = 50$ and $n = 168$
Hyperspectral Imaging - Detection performance

Figure: Detection probability versus SNR for a $P_{fa} = 10^{-2}$

Improvement of $\approx 10$ dB in detection due to the detection test and due to the more appropriate covariance matrix estimator
Conclusions

- General statistical modeling for HS images: CES distributions
- Theoretical performance of covariance matrix $M$-estimators as well as of detectors (more general with functional $H(.)$)
- Good agreement for false alarm regulation (although based on asymptotic results)

Perspectives

- Open problem: joint $M$-estimators of the mean and the covariance matrix as solutions of fixed point equations
- Performance of the mean (location) $M$-estimators
- Large dimensional problem (more than 200 spectral bands): use of RMT - Part D
There have been other applications for CES distributions and robust estimators...
One can cite:

- **Multivariate radar imaging**

- **Image processing**

**List of references**
E. Conte, M. Lops, and G. Ricci. 
Asymptotically Optimum Radar Detection in Compound-Gaussian Clutter. 

A. Farina, A. Russo, and F. Scannapieco. 
Radar detection in coherent Weibull clutter. 

J. Frontera-Pons, F. Pascal, and J-P. Ovarlez. 
False-alarm regulation for target detection in hyperspectral imaging. 

F. Gini, A. Farina, and M. V. Greco.
Selected list of references on radar signal processing.


Performance analysis of two adaptive radar detectors against non-gaussian real sea clutter data.

E. J. Kelly.
An adaptive detection algorithm.

S. Kraut and L. L. Scharf.
The CFAR adaptive subspace detector is a scale-invariant GLRT.
*Signal Processing, IEEE Transactions on, 47(9):2538–2541, 1999*
J. Liu, Z-J. Zhang, Y. Yang, and H. Liu. 
A CFAR Adaptive Subspace Detector for First-Order or Second-Order Gaussian Signals Based on a Single Observation. 

M. Mahot, F. Pascal, P. Forster, and J-P. Ovarlez. 
Asymptotic properties of robust covariance matrix estimates. 

R. A. Maronna, D. R. Martin, and J. V. Yohai. 

E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. 
Complex elliptically symmetric distributions: Survey, new results and applications.


On a SIRV-CFAR Detector with radar experimentations in impulsive noise.

F. Pascal and A. Renaux.
Statistical analysis of the covariance matrix MLE in K-distributed clutter.

Non-Gaussian random vector identification using spherically invariant random processes.


Covariance Matrix Estimation and Applications in Radar

Frédéric Pascal

1L2S
Supélec, France

Crash course/Tutorial
January 2015
Part D

$M$-estimators of scatter in the large dimensional regime
Part D: Contents

1 Motivations

2 Random Matrix Theory
   - Interest of RMT: A very simple example
   - Classical Results
   - Robust RMT
   - Applications to DoA estimation

3 Shrinkage $M$-estimator and link to RMT
   - Motivations
Key references of Part D


+ references on Regularized estimators (see the List of References)
1 Motivations

2 Random Matrix Theory
   - Interest of RMT: A very simple example
   - Classical Results
   - Robust RMT
   - Applications to DoA estimation

3 Shrinkage $M$-estimator and link to RMT
   - Motivations
Motivations for the RMT and the robust RMT

- **Grazing angle Radar (Impulsive Clutter)**

- **High Resolution Radar**
  - Small number of scatters in the Cell Under Test (CUT)
  - Central Limit Theorem (CLT) is not valid anymore

- **Big data: Many dimensions in Radar applications...**
  - STAP (Doppler, Space), typically, 64 pulses $\times$ 4 sensors
  - MIMO (Space, Space), typically, 4 emitters $\times$ 4 receivers
  - Polarimetry, typically 2 or 3 channels
  - Hyperspectral imaging
  - High dimensional problem ($\geq 250!$)
Extension to the RMT

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

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

⇒ BUT this is not the case in practice!

⇝ Random Matrix Theory

⇝ 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 robust covariance matrix estimation:

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

Random Matrix Theory

Shrinkage $M$-estimator and link to RMT

Outline

1 Motivations

2 Random Matrix Theory
   - Interest of RMT: A very simple example
   - Classical Results
   - Robust RMT
   - Applications to DoA estimation

3 Shrinkage $M$-estimator and link to RMT
   - Motivations
Interest of RMT: A very simple example...

Problem: Estimation of 1 DoA embedded in white Gaussian noise

\[ y(t) = \sqrt{p} A(\theta) s(t) + w(t), \]

where the \( w_t \)’s are \( n \) independent realizations of circular white Gaussian noise, i.e. \( w_t \sim \mathcal{CN}(0, I) \).

Classical approach

\[ \hat{S}_n = \frac{1}{n} \sum_{t=1}^{n} w_t w_t^H \xrightarrow{n \to \infty} I \]

Then, MUSIC algorithm allows to estimate the DoA...

What happens when the dimension \( m \) is large?

\[ \hat{S}_n = \frac{1}{m,n \to \infty} \sum_{t=1}^{n} w_t w_t^H \xrightarrow{m,n \to \infty} I \]

Then, MUSIC algorithm IS NOT the best way to estimate the DoA...
Classical approach: $n \gg m$

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 $\implies$ Bad performance

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

1 Motivations

2 Random Matrix Theory
   - Interest of RMT: A very simple example
   - Classical Results
   - Robust RMT
   - Applications to DoA estimation

3 Shrinkage $M$-estimator and link to RMT
   - Motivations
RMT - Classical results

Assumptions:

- $n, m \to \infty$ and $\frac{m}{n} \to c \in (0, 1)$ and $\hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^H$ the SCM
- $(z_1, \ldots, z_n)$ be a $n$-sample, i.i.d with finite fourth-order moment.

Thus one has:

1) $F_{\hat{S}_n} \Rightarrow F_{MP}$

where $F_{\hat{S}_n}$ (resp. $F_{MP}$) stands for the distribution of the eigenvalues of $\hat{S}_n$ (resp. the Marcenko-Pastur distribution) and $\Rightarrow$ stands for the weak convergence.
RMT - Classical results

Assumptions:

- \( n, m \to \infty \) and \( \frac{m}{n} \to c \in (0, 1) \) and \( \hat{S}_n = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^H \) the SCM
- \((z_1, \ldots, z_n)\) be a \( n \)-sample, i.i.d with finite fourth-order moment.

Thus one has:

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

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

with \( \hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_n \) the eigenvalues of \( \hat{S}_n \) and \( \hat{\mu}_1 \leq \ldots \leq \hat{\mu}_n \) the eigenvalues of

\[
\text{diag}(\hat{\lambda}) - \frac{1}{n} \sqrt{\hat{\lambda}} \sqrt{\hat{\lambda}}^T, \quad \hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_n)^T.
\]
Outline

1 Motivations

2 Random Matrix Theory
   - Interest of RMT: A very simple example
   - Classical Results
   - Robust RMT
   - Applications to DoA estimation

3 Shrinkage $M$-estimator and link to RMT
   - Motivations
Robust RMT

Assumptions:

- $n, m \to \infty$ and $\frac{m}{n} \to c \in (0, 1)$ and $\hat{\Sigma}$ a $M$-estimator (with previous assumptions)
- $(z_1, ..., z_n)$ be a $n$-sample, i.i.d ($\Delta !!!!!$) with finite fourth-order moment

Thus, one has shown in [1]:

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.
Robust RMT

2) \( \|\psi^{-1}(1) \hat{\Sigma} - \hat{S}_n\| \xrightarrow{a.s.} 0 \) when \( n, m \to \infty \) and \( \frac{m}{n} \to c \)

where \( \| . \| \) stands for the spectral norm and \( \psi(x) = x \varphi(x) \) (\( \varphi \) is the weight function of the \( M \)-estimator)

Classical results in RMT can be extended to the \( M \)-estimators!

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

(for the eigenvalues of \( \hat{\Sigma} \))

Outline

1 Motivations

2 Random Matrix Theory
   - Interest of RMT: A very simple example
   - Classical Results
   - Robust RMT
   - Applications to DoA estimation

3 Shrinkage $M$-estimator and link to RMT
   - Motivations
Application to DoA estimation with MUSIC for different additive clutter

(a) Homogeneous noise ($\sim$ 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

\[ \sim 7\text{dB gain for resolution performance on classical MUSIC} \]
Pros and Cons of these results

- Advantages
  - Original results on robust RMT
  - Now, possibility of using robust estimate in a RMT context: means possible extension of classical RMT results such DoA estimation (done), sources power estimation, number of sources estimation (challenging problem), detection...
  - Great improvement on sources resolution
  - Great improvement on MUSIC statistic estimation
Pros and Cons of these results

- **Limitations**
  - Assumption of independence, i.e. NOT exactly CES data:
    \[
    z_i = \begin{pmatrix}
    \tau_1 x_i^{(1)} \\
    \vdots \\
    \tau_m x_i^{(m)}
    \end{pmatrix}
    \]
    instead of
    \[
    z_i = \tau_i 
    \begin{pmatrix}
    x_i^{(1)} \\
    \vdots \\
    x_i^{(m)}
    \end{pmatrix}
    \]
    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.
Motivations
Random Matrix Theory
Shrinkage $M$-estimator and link to RMT

Interest of RMT: A very simple example
Classical Results
Robust RMT
Applications to DoA estimation

Update on this work

MSE on the DoA estimation

(a) Gaussian

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

(c) 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 $m = 20$. 
Update on this work

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 $m = 20$. 
Robust RMT under CES distributions

Results on the eigenvalues distributions of the $M$-estimators have now been proved for CES distribution in...


Results on the eigenvalues distributions of the Tyler’s estimator (!) have now been proved for CES distribution in...

Outline

1. Motivations

2. Random Matrix Theory
   - Interest of RMT: A very simple example
   - Classical Results
   - Robust RMT
   - Applications to DoA estimation

3. Shrinkage $M$-estimator and link to RMT
   - Motivations
Some advantages

- 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...
Shrinkage Tyler’s estimators

Chen estimator

\[ \hat{\Sigma}_C = (1 - \beta) \frac{m}{n} \sum_{i=1}^{n} \frac{z_i z_i^H}{z_i^H \hat{\Sigma}^{-1} C z_i} + \beta I \]

subject to the constraint \( \text{Tr}(\hat{\Sigma}) = m \) and for \( \beta \in (0, 1] \).

- Originally introduced in

- Existence, uniqueness and algorithm convergence proved in
Shrinkage Tyler’s estimators

**Pascal estimator**

\[
\hat{\Sigma}_P = (1 - \beta) \frac{m}{n} \sum_{i=1}^{n} \frac{z_i z_i^H}{z_i^H \hat{\Sigma}_P^{-1} z_i} + \beta I
\]

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


\( \hat{\Sigma}_P \) (naturally) verifies \( \text{Tr}(\hat{\Sigma}_P^{-1}) = m \) for all \( \beta \in (0, 1] \)
Shrinkage Tyler’s estimators

The main challenge is to find the optimal $\beta$!

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


where it is also proved that

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

Applications to STAP data for $\neq$ values of $\beta$, $m = 256$ and $N = 400$

(Trial 10, beta= 0.5, 400 secondary data) (Trial 10, beta= 0.6, 400 secondary data)

(Trial 10, beta= 0.7, 400 secondary data) (Trial 10, beta= 0.8, 400 secondary data)

(Trial 10, beta= 0.9, 400 secondary data) (Trial 10, beta= 1, 400 secondary data)

(d) SCM (e) Shrinkage FPE
Applications to STAP data for $\neq$ values of $\beta$, $m = 256$ and $N = 200$

(f) SCM

(g) Shrinkage FPE
Motivations

Random Matrix Theory

Shrinkage $M$-estimator and link to RMT

References and thanks to...

- my co-authors:

Yacine Chitour  Jean-Philippe Ovarlez  Romain Couillet  Jack Silverstein

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

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.

YI Abramovich and Nicholas K Spencer.
Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering.

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.
Olivier Besson and Yuri I Abramovich. 
Invariance properties of the likelihood ratio for covariance matrix estimation in some complex elliptically contoured distributions. 

Yilun Chen, Ami Wiesel, and Alfred O Hero. 
Robust shrinkage estimation of high-dimensional covariance matrices. 

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.
Motivations
Random Matrix Theory
Shrinkage $M$-estimator and link to RMT


R Couillet, F Pascal, and J W Silverstein.
Robust $M$-Estimation for Array Processing: A Random Matrix Approach.

R Couillet, F Pascal, and J W Silverstein.
The Random Matrix Regime of Maronna’s $M$-estimator with elliptically distributed samples.

Romain Couillet and Matthew R McKay.
Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators.
Olivier Ledoit and Michael Wolf.
A well-conditioned estimator for large-dimensional covariance matrices.

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

Ilya Soloveychik and Ami Wiesel.
Non-asymptotic Error Analysis of Tyler’s Scatter Estimator.

Ami Wiesel.
Unified framework to regularized covariance estimation in scaled Gaussian models.

Teng Zhang, Xiuyuan Cheng, and Amit Singer.
Marchenko-Pastur Law for Tyler’s and Maronna’s $M$-estimators.