Theoretical Analysis of an Improved Covariance Matrix Estimator in non-Gaussian Noise
Frédéric Pascal, Philippe Forster, Jean-Philippe Ovarlez, Pascal Larzabal

To cite this version:

HAL Id: hal-02495012
https://hal-centralesupelec.archives-ouvertes.fr/hal-02495012
Submitted on 29 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
THEORETICAL ANALYSIS OF AN IMPROVED COVARIANCE MATRIX ESTIMATOR IN NON-GAUSSIAN NOISE

F. Pascal123, P. Forster2, J-P. Ovarlez1, P. Larzabal3.

1 ONERA, Chemin de la Hunière, F-91761 Palaiseau Cedex, France
2 GEA, 1 Chemin Desvalières, F-92410 Ville d’Avray, France
3 ENS Cachan, 61 Avenue du Président Wilson, F-94235 Cachan Cedex, France
Email : pascal@onera.fr, ovarlez@onera.fr, philippe.forster@cva.u-paris10.fr, pascal.larzabal@satie.ens-cachan.fr

ABSTRACT

This paper presents a detailed theoretical analysis of a recently introduced covariance matrix estimator, called the Fixed Point Estimate (FPE). It plays a significant role in radar detection applications. This estimate is provided by the Maximum Likelihood Estimation (MLE) theory when the non-Gaussian noise is modelled as a Spherically Invariant Random Process (SIRP). We study in details its properties: existence, uniqueness, unbiasedness, consistency and asymptotic distribution. We propose also an algorithm for its computation and prove the convergence of this numerical procedure. These results will allow to study the performance analysis of the adaptive CFAR radar detectors (GLRT-LQ, BORD,...).

1. PROBLEM STATEMENT AND BACKGROUND

Non-Gaussian noise characterization has gained many interests since experimental radar clutter measurements showed that these data are correctly described by non-Gaussian statistical models. One of the most tractable and elegant non-Gaussian model comes from the so-called Spherically Invariant Random Process (SIRP) theory. A SIRP is the product of a Gaussian random process - called speckle - with a non-negative random variable - called texture. This model leads to many results [1, 2, 3, 4].

The basic problem of detecting a complex signal corrupted by an additive SIRP noise \( e \) in a \( m \)-dimensional complex vector \( y \) allowed to build several Generalized Likelihood Ratio Tests like the GLRT-Linear Quadratic (GLRT-LQ) in [1, 2] or the Bayesian Optimum Radar Detector (BORD) in [3, 4].

Let us recall some SIRP theory results. A noise modelled as a SIRP is a non-homogeneous Gaussian process with random power. More precisely, a SIRP [5] is the product of a positive random variable \( \tau \) (texture) and a \( m \)-dimensional independent complex Gaussian vector \( x \) (speckle) with zero mean covariance matrix \( \mathbf{M} = \mathbb{E}(xx^\dagger) \) with normalization \( \text{tr}(\mathbf{M}) = m \), where \( \dagger \) denotes the conjugate transpose operator:

\[
\mathbf{e} = \sqrt{\tau} \mathbf{x}.
\]

The SIRP PDF expression is:

\[
p_m(e) = \int_0^{+\infty} g_m(e, \tau) p(\tau) d\tau,
\]

where

\[
g_m(e, \tau) = \frac{1}{(\pi \tau)^{m/2}} \exp\left(\frac{-e^\dagger \mathbf{M}^{-1} e}{\tau}\right).
\]

In many problems, non-Gaussian noise can be characterized by SIRPs but the covariance matrix \( \mathbf{M} \) is generally not known and an estimate \( \hat{\mathbf{M}} \) is required. Obviously, it has to satisfy the \( \mathbf{M} \)-normalization: \( \text{tr}(\hat{\mathbf{M}}) = m \).

In the literature [6, 7], the Normalized Sample Covariance Matrix Estimate (NSCME) defined as follows is usually used:

\[
\hat{\mathbf{M}}_{\text{NSCME}} = \frac{m}{N} \sum_{i=1}^{N} \left( \mathbf{e}_i \mathbf{e}_i^\dagger \right) = \frac{m}{N} \sum_{i=1}^{N} \left( \mathbf{x}_i \mathbf{x}_i^\dagger \right),
\]

\[
\hat{\mathbf{M}}_{\text{NSCME}} = \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{e}_i \mathbf{e}_i^\dagger}{\mathbf{e}_i \mathbf{M}^{-1} \mathbf{e}_i} \right),
\]

where \( 1 \leq i \leq N \). \( \mathbf{e}_i = \sqrt{\tau_i} \mathbf{x}_i \) are \( m \)-dimensional independent complex SIRP vectors and \( \mathbf{x}_i \) are \( m \)-dimensional independent complex Gaussian vectors with zero mean and normalized covariance matrix \( \mathbf{M} \).

This estimate is remarkably independent of the texture statistics. However, despite of this interesting property, the NSCME (3) suffers the following drawbacks:

- it is a biased estimate;
- the resulting adaptive Generalized Likelihood Ratio (GLR) is not independent of matrix \( \mathbf{M} \) characteristics.

In the next sections, we propose to introduce and analyze an improved estimator of \( \mathbf{M} \): the FPE estimator.

2. THE FIXED POINT ESTIMATOR \( \hat{\mathbf{M}}_{\text{FP}} \)

Conte and Gini in [8, 9] have shown that the Maximum Likelihood estimator \( \hat{\mathbf{M}} \) of \( \mathbf{M} \) is a solution of the following equation:

\[
\hat{\mathbf{M}} = \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{e}_i \mathbf{e}_i^\dagger}{\mathbf{e}_i \mathbf{M}^{-1} \mathbf{e}_i} \right).
\]

Existence and uniqueness of the above equation solution have been investigated in a previous paper [10] and we briefly recall here the main results in Theorem 1.
Let function \( f \) be defined as:
\[
f(M) = \frac{m}{N} \sum_{i=1}^{N} \left( \frac{c_i x_i^T}{c_i M^{-1} c_i} \right),
\]
and notice that \( f \) can be rewritten as follows, which shows that \( \hat{M} \) is also statistically independent:
\[
f(\hat{M}) = \frac{m}{N} \sum_{i=1}^{N} \left( \frac{x_i x_i^T}{x_i M^{-1} x_i} \right).
\]

**Theorem 1**

1. The function \( f \) admits a single fixed point, called \( \hat{M}_{fp} \), which verifies
   \[
   f(\hat{M}_{fp}) = \hat{M}_{fp}.
   \]
2. Let us consider the recurrence relation
   \[
   \hat{M}_{t+1} = \frac{m}{tr(f(\hat{M}_t))} f(\hat{M}_t)
   \]
   Then
   \[
   \hat{M}_t \xrightarrow{t \to \infty} \hat{M}_{fp}.
   \]

The point one of the theorem shows that the Maximum Likelihood estimate \( \hat{M}_{fp} \) exists and is unique.

**Fig. 1.** Illustration of the convergence to the fixed point.

Figure 1 illustrates the second point of the above theorem. On this figure, the relative error \( \|f(M_t) - \hat{M}_t\|/\|\hat{M}_t\| \) has been plotted versus \( t \) with initial value \( \hat{M}_0 = \hat{M}_{NSCME} \) (3).

Other simulations have been performed with different initial values \( \hat{M}_0 \) (ex: sample Gaussian covariance estimate, matrix with uniform PDF, deterministic Toeplitz matrix), each of them conducting to the same value with an extremely fast convergence:
\[
\|f(\hat{M}_t) - \hat{M}_t\| \xrightarrow{t \to \infty} 0
\]

But \( \hat{M}_{fp} \) is only given in (7) as an implicit function of the data: there is no closed form expression for \( \hat{M}_{fp} \). So, it is a significant feature to characterize \( \hat{M}_{fp} \) by its properties. The statistical properties have never been investigated. The purpose of this paper is to fill these gaps by establishing the bias, the consistency and the asymptotic distribution of this random matrix: this is the aim of the next section.

### 3. \( \hat{M}_{fp} \) Properties

In order to simplify the notations in the proofs, let us define the following quantities:
- \( \hat{M} = \hat{M}_{fp} \);
- \( \Delta = M^{-1/2} \hat{M} M^{-1/2} - I_m \) where \( I_m \) is the \( m \times m \) identity matrix;
- \( X = \text{vec}(\Delta) \) where \( X \) is the vector containing all the elements of \( \Delta \) and \( \text{vec} \) denotes the operator which reshapes the \( m \times n \) matrix elements into a \( mn \) column vector;
- \( \top \) denotes the transpose operator.

#### 3.1. Bias and Consistency

**Proposition 1** \( \hat{M}_{fp} \) is unbiased and it is a consistent estimator.

**Proof 1** This proof will appear in a forthcoming journal paper.

#### 3.2. \( \hat{M}_{fp} \) Covariance Matrix

**Proposition 2** We have the following original results with the above notations:

1. \[\sqrt{N} \begin{bmatrix} \text{Re}(X) \\ \text{Im}(X) \end{bmatrix} \xrightarrow{law} N(0,2m^2, C) \text{, where } \xrightarrow{law} \text{ represents the convergence in distribution};\]
2. \[N E(XX^\top) = \left( \frac{m+1}{m} \right)^2 C_1 \text{ where}\]
   \[C_1 = \frac{m}{m+1} \left( p - \frac{1}{m} \text{vec}(I_m)\text{vec}(I_m)^\top \right); \quad (9)\]
3. \[N E(XX^\dagger) = \left( \frac{m+1}{m} \right)^2 C_2 \text{ where}\]
   \[C_2 = \frac{m}{m+1} \left( I_{m^2} - \frac{1}{m} \text{vec}(I_m)\text{vec}(I_m)^\top \right). \quad (10)\]

Notice that \( C \), which is the covariance matrix of \( \begin{bmatrix} \text{Re}(X) \\ \text{Im}(X) \end{bmatrix} \), is fully characterized by the two quantities \( E(XX^\top) \) and \( E(XX^\dagger) \).

**Proof 2** First, let us define \( \hat{M} = M + \Delta_1 \). Notice that
\[
\Delta_1 = M^{1/2} \Delta M^{1/2}.
\]
For large \( N \), \( \Delta_1 \simeq 0_{m \times m} \) because of the \( \hat{M} \) consistency.

We suppose \( N \) to be large enough to ensure the validity of the first order expressions.

We can write
\[
\hat{M}^{-1} = (M(I_m + M^{-1} \Delta_1))^{-1} = (I_m + M^{-1} \Delta_1)^{-1} M^{-1} \approx (I_m - M^{-1} \Delta_1) M^{-1} \approx (M^{-1} - M^{-1} \Delta_1 M^{-1}).
\]
For $N$ large enough, this implies that:

$$
\hat{\mathbf{M}} \simeq \frac{m}{N} \sum_{i=1}^{N} \mathbf{x}_i (\mathbf{M}^{-1} - \mathbf{M}^{-1} \Delta_i \mathbf{M}^{-1}) \mathbf{x}_i .
$$

(12)

Hence,

$$
\Delta_i \simeq \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{x}_i (\mathbf{M}^{-1} - \mathbf{M}^{-1} \Delta_i \mathbf{M}^{-1}) \mathbf{x}_i}{\mathbf{x}_i (\mathbf{M}^{-1} - \mathbf{M}^{-1} \Delta_i \mathbf{M}^{-1}) \mathbf{x}_i} - \mathbf{M} \right). 
$$

(13)

Let us define $\mathbf{y}_i = \mathbf{M}^{-1/2} \mathbf{x}_i$, where $\mathbf{y}_i$ are Gaussian iid complex vectors with identity covariance matrix. Then,

$$
\mathbf{M}^{-1/2} \mathbf{y}_i \mathbf{M}^{-1/2} \simeq \frac{m}{N} \sum_{i=1}^{N} \mathbf{y}_i (\mathbf{I}_m - \mathbf{M}^{-1/2} \Delta_i \mathbf{M}^{-1/2}) \mathbf{y}_i - \mathbf{I}_m ,
$$

or equivalently using expression (11),

$$
\Delta \simeq \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{y}_i \Delta_i \mathbf{y}_i}{\mathbf{y}_i \mathbf{y}_i} \right) - \mathbf{I}_m .
$$

(14)

We obtain at the first order, for large $N$,

$$
\Delta \simeq \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{y}_i \Delta_i \mathbf{y}_i}{\mathbf{y}_i \mathbf{y}_i} \right) - \mathbf{I}_m .
$$

(15)

To find the explicit expression of $\Delta$ in terms of data, the above expression can be reorganized as:

$$
\Delta = \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{y}_i \Delta_i \mathbf{y}_i}{\mathbf{y}_i \mathbf{y}_i} \right) - \mathbf{I}_m .
$$

(16)

To solve this $m^2$-system, the above equation has to be rewritten as:

$$
\mathbf{B} \mathbf{X} \simeq \text{vec} \left( \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{y}_i \Delta_i \mathbf{y}_i}{\mathbf{y}_i \mathbf{y}_i} \right) - \mathbf{I}_m \right) ,
$$

(18)

where

- $\mathbf{X} = \text{vec}(\Delta)$,
- $\mathbf{B} = \mathbf{I}_{m^2} - \frac{m}{N} \sum_{i=1}^{N} \mathbf{D}_i / \sqrt{\mathbf{y}_i \mathbf{y}_i}$,
- $\mathbf{D}_i$ is the $m^2 \times m^2$ matrix defined by $\mathbf{D}_i = (d_{kl})_{1 \leq k,l \leq m^2}$ with
  
  $$d_{kl} = \mathbf{y}_p \mathbf{y}_q \mathbf{y}_{p'} \mathbf{y}_{q'}$$
  
  and
  
  $$\begin{cases} k + p + m(q - 1) & \text{with } 1 \leq p, q \leq m, \\
  l + p' + m(q' - 1) & \text{with } 1 \leq p', q' \leq m .
  \end{cases}
  $$

Notice that the vec operator reorganizes the matrix elements as follows: if $\mathbf{H} = (h_{ij})_{1 \leq i,j \leq m}$ and $\text{vec}(\mathbf{H}) = (v_k)_{1 \leq k \leq m^2}$, then $h_{ij} = v_k$ for $k = (j - 1)m + i$.

Let us set in equation (18):

$$
\mathbf{A} = \text{vec} \left( \frac{m}{N} \sum_{i=1}^{N} \left( \frac{\mathbf{y}_i \Delta_i \mathbf{y}_i}{\mathbf{y}_i \mathbf{y}_i} \right) - \mathbf{I}_m \right) .
$$

(19)

Using the Central Limit Theorem (CLT), the right hand side of equation (18) satisfies:

$$
\sqrt{N} \left[ \frac{\text{Re}(\mathbf{A})}{\text{Im}(\mathbf{A})} \right] \xrightarrow{\text{law}} N(0,2m^2,G),
$$

(20)

where $G$ is the covariance matrix of $\text{Re}(\mathbf{A}) / \text{Im}(\mathbf{A})$.

While $\mathbf{B}$ at the left hand side, by the Strong Law of Large Numbers (SLLN), has the following property:

$$
\mathbf{B} \xrightarrow{N \to +\infty} \mathbf{C}_2 = \mathbf{I}_{m^2} - \mathbb{E} \left( \frac{\mathbf{D}_i}{(\mathbf{y}_i \mathbf{y}_i)^2} \right) .
$$

(21)

Thus, from standard probability convergence considerations, we obtain the first point of proposition 2. Moreover, for large $N$, we have the following equations:

$$
\begin{align*}
\mathbb{E} \left( \mathbf{X} \mathbf{X}^\top \right) &= \mathbf{C}_2^{-1} \mathbb{E} \left( \mathbf{A} \mathbf{A}^\top \right) \mathbf{C}_2^{-1} , \\
\mathbb{E} \left( \mathbf{X} \mathbf{X} \right) &= \mathbf{C}_2^{-1} \mathbb{E} \left( \mathbf{A} \mathbf{A} \right) \mathbf{C}_2^{-1} .
\end{align*}
$$

(22)

Let us now turn to the closed form expression of $\mathbf{C}_2$.

$$
\mathbf{C}_2 = \mathbf{I}_{m^2} - \mathbb{E} \left( \frac{\mathbf{D}_i}{(\mathbf{y}_i \mathbf{y}_i)^2} \right) ,
$$

(23)

where $\mathbf{D}$ is the $m^2 \times m^2$ matrix defined by $\mathbf{D} = (d_{kl})_{1 \leq k,l \leq m^2}$ with $d_{kl} = \mathbf{y}_p \mathbf{y}_q \mathbf{y}_{p'} \mathbf{y}_{q'}$.

Now, as we have $\mathbf{y} = (y_1, \ldots, y_m)^\top \sim \mathcal{N}(\mathbf{0}_m, \mathbf{I}_m)$, we can rewrite $y_i = \sqrt{r_i}/2 \exp(i \theta_i)$ where $\forall j \in \{1, \ldots, m\}$, $r_j$ and $\theta_j$ are independent variables, with $r_j \sim \chi^2(2)$ and $\theta_j \sim U([0,2\pi])$.

Then, $\forall 1 \leq p, q, p', q' \leq m$, let us set for the elements of $\mathbf{E} = \mathbb{E} \left( \frac{\mathbf{D}_i}{(\mathbf{y}_i \mathbf{y}_i)^2} \right)$ in (23):

$$
E_{kl} = \mathbb{E} \left( \frac{\mathbf{y}_p \mathbf{y}_q \mathbf{y}_{p'} \mathbf{y}_{q'}}{(\mathbf{y}_i \mathbf{y}_i)^2} \right) .
$$

(24)

Thus, with the above notations, we have:

$$
E_{kl} = \mathbb{E} \left( \frac{\exp(i(\theta_p - \theta_q + \theta_{p'} - \theta_{q'}))}{(\sum_{j=1}^{m} r_j)^2} \right) .
$$

(24)

We can notice that:

$$
\mathbb{E} \left( \exp(i(\theta_p - \theta_q + \theta_{p'} - \theta_{q'})) \right) \neq 0
$$

if and only if

1. $p = q = p' = q'$,
2. $p = q, p' = q'$ et $p \neq p'$,
3. $p = p', q = q'$ et $p \neq q$.

which is equivalent to:

1. $k = l = p + m(p - 1)$,
2. $k = p + m(p - 1), l = p' + m(p' - 1)$ et $p \neq p'$, i.e. $k \neq l$,
3. $k = p + m(q - 1), l = p + m(q - 1)$ et $p \neq q$.

Finally, the non zero elements of the matrix $\mathbf{E}$ are:
by (10). In a similar way, we have derived the following results:

1. \( E_{p+m(p-1),p+m(p-1)} = \mathbb{E}\left( \frac{p^2}{\sum_{j=1}^{m} r_j^2} \right) = \frac{2}{m(m+1)} \)

2. \( E_{p+m(p-1),p'+m(p'-1)} = \mathbb{E}\left( \frac{r_p r_{p'}}{\sum_{j=1}^{m} r_j^2} \right) = \frac{1}{m(m+1)} \)

3. \( E_{p+m(p-1),p+m(q-1)} = \mathbb{E}\left( \frac{r_p r_q}{\sum_{j=1}^{m} r_j^2} \right) = \frac{1}{m(m+1)} \)

In summary, we obtain the closed form expression of \( C_2 \) given by (10). In a similar way, we have derived the following results:

- \( C_1 = \frac{m}{m+1} \left( \mathbf{P} - \frac{1}{m} \text{vec}(\mathbf{I}_m) \text{vec}(\mathbf{I}_m)^\top \right) \) where \( \mathbf{P} \) is the \( m^2 \times m^2 \) nilpotent matrix defined as follows:
  - \( P_{ij} = 0 \) and for \( 1 \leq p, p' \leq m : \)
    - \( P_{p+m(p-1),p+m(p-1)} = 1 \),
    - \( P_{p+m(p-1),p'+m(p-1)} = 1 \),
- \( \mathbb{E}(\mathbf{A}\mathbf{A}^\top) = \mathbf{C}_1 \),
- \( \mathbb{E}(\mathbf{A}\mathbf{A}^\top) = \mathbf{C}_2 \).

4. APPLICATIONS

![Fig. 2. Relation “PFA-threshold” for different SIRP noises](image)

In radar detection, the following test is used to detect, in an \( m \)-vector observation \( \mathbf{z} \), a complex known signal \( s \) whose Doppler characteristics are represented here by its steering vector \( \mathbf{p} \):

\[
\hat{\Lambda}(\mathbf{M}) = \frac{p^\top \mathbf{M}^{-1} p^\top}{(p^\top \mathbf{M}^{-1} p)(z^\top \mathbf{M}^{-1} z)} \geq \lambda.
\]

In a previous paper [10], we have derived the closed form expression of the relation between the Probability of False Alarm (PFA) and the detection threshold \( \lambda \), under Gaussian noise assumption and for \( \mathbf{M} \) equal to the sample covariance matrix (Wishart distributed).

As illustrated on the Figure 2, results obtained in this paper allowed us to show that the previous relation is surprisingly still valid under any SIRP noise and for \( \mathbf{M} \) equal to the FPE. Indeed, the covariance matrix of the Wishart matrix (not done in this paper) is the same near to a multiplicative scalar factor as the FPE covariance matrix established in this paper. Then, the asymptotic behavior of the previous estimates is the same near to a multiplicative scalar factor.

5. CONCLUSIONS AND OUTLOOK

We have established in this paper the theoretical statistical properties (unbiasedness, consistency, asymptotic distribution) of the MLE SIRP kernel covariance matrix estimate. Simulation results have confirmed the validity of this work.

These properties will be of practical use in radar detection for establishing the statistics of the GLRT-LQ or BORD radar detectors. These adaptive detectors built with the FPE will have the following significant feature: they will be SIRP-CFAR (independent of the texture statistics as well as the structure of the covariance matrix \( \mathbf{M} \) of the SIRP gaussian kernel).

6. REFERENCES


