Robust Burg Estimation of stationary autoregressive mixtures covariance

Alexis Decurninge∗ and Frédéric Barbaresco†

∗alexis.decurninge@upmc.fr
†frederic.barbaresco@thalesgroup.com

Abstract. Burg estimators are classically used for the estimation of the autocovariance of a stationary autoregressive process. We propose to consider scale mixtures of stationary autoregressive processes, a non-Gaussian extension of the latter. The traces of such processes are Spherically Invariant Random Vectors (SIRV) with a constraint on the scatter matrix due to the autoregressive model. We propose adaptations of the Burg estimators to the considered models and their associated robust versions based on geometrical considerations.

Keywords: Burg technique, autoregressive process, elliptical distributions, SIRV

PACS: 02

INTRODUCTION

Motivation and notations

Non-Gaussian models of strong clutters such as ground or sea clutters are used in the field of radar processing. The family of complex elliptically symmetric distributions [1] is a useful generalization of Gaussian random vectors, inheriting of similar shape and location parameters. We consider the Spherically Invariant Random Vectors (SIRV), that compose a large subfamily of complex elliptical distributions. A centered SIRV $X = (X_1, \ldots, X_d)^T \in \mathbb{C}^d$ is characterized by the existence of a positive random variable $\tau$ (the amplitude) and a Gaussian random vector $Y$ of covariance $\Sigma$ such that $X = \tau Y$. Within this framework, we consider two kinds of robustness concepts for the estimation of the scatter matrix $\Sigma$:

• (R1) a robustness with respect to the distribution of the amplitude which is often heavy-tailed
• (R2) a robustness with respect to contamination in the observed sample

We consider stationary samples where stationarity is defined as the second order one. This assumption adds a Toeplitz structure constraint for the scatter matrix $\Sigma$. The Toeplitz structure allows us to split the estimation of the matrix $\Sigma$ of size $d \times d$ into $d$ estimations of Toeplitz matrices of size $2 \times 2$. This splitting corresponds to the so-called "Burg technique" [3]. Indeed, instead of estimating the covariance of the raw sample $x_1, \ldots, x_N \in \mathbb{C}^d$, we iteratively define second-order samples in $\mathbb{C}^2$ whose theoretical covariance can be expressed in function of $\Sigma$.

This technique was originally proposed in the context of stationary Gaussian autore-
gressive time series. The sample \( x_1, ..., x_N \) can be viewed as the collection of \( N \) traces of such a time series. The parallel between a time series and its trace is often implicit in the signal processing literature. For this reason, we will refer at this trace as autoregressive vector.

Moreover, if we consider \( X \) as the trace of an autoregressive process of order \( M < d - 1 \), we add more structure on the matrix \( \Sigma \) than the Toeplitz one. Actually, given the auto-covariance \( \mathbb{E}[X_i X_k] \) for \( k = 1 \ldots M \) with \( M \leq d - 1 \), it is well known that the maximum entropy model pertaining to the vector \( X = (X_1, ..., X_d)^T \) in \( \mathbb{C}^d \) results as the complex Gaussian distribution in \( \mathbb{C}^d \), whose covariance coincides with the autoregressive autocovariance of size \( d \times d \) (see [3][4]). We propose here to adapt these techniques for non-Gaussian scale mixtures of autoregressive vectors, a subfamily of the class of SIRV.

Moreover, in order to deal with the robustness (R2), we propose a geometrical method consisting in computing the median of autoregressive models estimated for subsamples of \( x_1, ..., x_N \). The known robustness of the median with respect to outliers (see [5][6][7]) will be illustrated.

### STATIONARY SIRV MODELS : SCALE MIXTURE OF AUTOREGRESSIVE VECTORS

#### Presentation of the model

Let \( X \in \mathbb{C}^d \) be a random variable sampled from a scale mixture of stationary Gaussian autoregressive random vectors. Then, similarly to SIRV distributions, \( X \) is characterized by the existence of a scalar random variable \( \tau \in \mathbb{R}_+ \) and a scatter matrix \( \Sigma \) such that:

\[
X \overset{d}{=} \tau Y
\]

where \( Y \sim \mathcal{N}_d(0, \Sigma) \) is the trace of a stationary Gaussian autoregressive process (i.e. a Gaussian vector, called speckle, of Toeplitz covariance \( \Sigma \)) independent of \( \tau \) (called texture). As \( Y \) is the trace of a stationary Gaussian autoregressive process of order \( M \leq d - 1 \) of parameters \( a^{(M)}_1, ..., a^{(M)}_M \in \mathbb{C} \), it holds for \( 1 \leq n \leq d \):

\[
Y_n + \sum_{i=1}^{M} a^{(M)}_i Y_{n-i} = b_n
\]

where \( b_n \) is a complex standard Gaussian random variable independent of \( Y_{n-1}, ..., Y_{n-M} \) with the convention \( Y_{-i} = 0 \) for all \( i \geq 0 \).

We note that \( X \) is also the trace of an autoregressive process with dependent non-Gaussian noise:

\[
X_n + \sum_{i=1}^{M} a^{(M)}_i X_{n-i} = \tau b_n
\]
Burg method applied to Gaussian autoregressive vectors

We first present the Burg method for Gaussian autoregressive vectors. All the definitions which we introduce for the process underlying \( Y \) remain valid for the process associated to \( X \). Let us define the autocovariance function \( \gamma \) of the underlying Gaussian autoregressive process. For \( t \geq 0 \), we have \( \gamma(t) = \mathbb{E}[Y_{n+t}Y_n] \) for any \( n \). The Levinson algorithm inverts the stationarity equations by introducing the successive autoregressive parameters \( (a_k^{(m)})_{1 \leq k \leq m} \) of order \( 1 \leq m \leq M \):

- Initialization : let us define \( P_0 = \gamma(0) \) and

\[
\begin{align*}
\mu_1 &:= a_1^{(1)} = -\frac{\gamma(1)}{P_0} \\
P_1 &:= P_0(1 - |\mu_1|^2)
\end{align*}
\]

(4)

- for \( 1 \leq m \leq M - 1 \)

\[
\begin{align*}
\mu_{m+1} &:= a_{m+1}^{(m+1)} = -\frac{\gamma(m+1)+\sum_{k=1}^{m} a_k^{(m)} \gamma(m+1-k)}{P_m} \\
P_{m+1} &:= P_m(1 - |\mu_m|^2) \\
\begin{pmatrix}
a_1^{(m+1)} \\
a_2^{(m+1)} \\
\vdots \\
a_{m}^{(m+1)} \\
a_{m+1}^{(m+1)}
\end{pmatrix} &:= \begin{pmatrix}
a_1^{(m)} \\
a_2^{(m)} \\
\vdots \\
a_{m}^{(m)} \\
a_{m+1}^{(m)}
\end{pmatrix} + \mu_{m+1} \begin{pmatrix}
\bar{a}_m^{(m)} \\
\bar{a}_m^{(m)} \\
\vdots \\
\bar{a}_m^{(m)} \\
\bar{a}_m^{(m)}
\end{pmatrix}
\end{align*}
\]

(5)

This algorithm enhances the role of the parameters \( (\mu_m)_{1 \leq m \leq M} \), called reflection (or Verblunsky) parameters, that are sufficient, together with \( P_0 \), in order to describe the autoregressive vector in \( \mathbb{C}^d \). Instead of estimating the covariance matrix directly from the samples which does not guarantee the Toeplitz constraint, we estimate these reflection parameters satisfying the Toeplitz structure (we will then use the bijection between \( (P_0, \mu_1, ..., \mu_M) \) and \( \Sigma \) given by equations (4) and (5) to recover an estimated covariance).

For this purpose, Burg proposes in the Gaussian framework to minimize an error at each step \( 1 \leq m \leq M \):

\[
U^{(m)} = \sum_{n=m+1}^{d} |f_m(n)|^2 + |b_m(n)|^2
\]

(6)

with \( f_m \) and \( b_m \) respectively the "forward" and "backward" errors defined for \( m+1 \leq n \leq d \):

\[
\begin{align*}
f_m(n) &:= Y_n + \sum_{k=1}^{m} a_k^{(m)} Y_{n-k} \\
b_m(n) &:= Y_{n-m} + \sum_{k=1}^{m} \bar{a}_k^{(m)} Y_{n-m+k}
\end{align*}
\]

(7)

Thanks to the equation (5), we can state for \( m+2 \leq n \leq d \):

\[
\begin{align*}
f_{m+1}(n) &= f_m(n) + \mu_{m+1} b_m(n-1) \\
b_{m+1}(n) &= b_m(n-1) + \bar{\mu}_{m+1} f_m(n)
\end{align*}
\]

(8)

Note that the errors are random variables and that \( f_m(n) \) and \( b_m(n) \), both depending on \( (a_1^{(m)}, ..., a_{m}^{(m)}) \), are not directly observable from \( Y \) for \( m > 0 \). Although Burg introduces
the criterion 6 as an iterative least square method for autoregressive process of order \( m \) (for \( m \) going from 1 to \( M \)), it can be justified as an approximation of the likelihood of the errors \( e_m(n) := \left( \begin{array}{c} f_m(n) \\ b_m(n-1) \end{array} \right) \) for \( n \in \{m+1, \ldots, d\} \). Let us first give the moments of \( e_m \) by the following lemma due to Brockwell and Dalhaus [10]

**Lemma 1.** If \( m \geq 0 \) and \( m+1 \leq n \leq d \)

\[
\begin{align*}
\mathbb{E}[|f_m(n)|^2] &= \mathbb{E}[|b_m(n-1)|^2] = P_m \\
\mathbb{E}[f_m(n)b_m(n-1)] &= -P_m \mu_{m+1} 
\end{align*}
\]

(9)

Burg’s technique lies in the iterative estimation of the correlation \(-\mu_{m+1}\) of the coordinates \( f_m(n) \) and \( b_m(n-1) \). For each \( i \), \( f_{i,m} \) and \( b_{i,m} \) denote the observed forward and backward errors for the sample \( Y_i \). Knowing \( \mu_1, \ldots, \mu_m \), we estimate \( \mu_{m+1} \) by the minimization of the empirical version of the criterion 6

\[
\hat{\mu}_{m+1}^{(gauss)} = -2 \frac{\sum_{i=1}^N \sum_{n=m+2}^d f_{i,m}(n)b_{i,m}(n-1)}{\sum_{i=1}^N \sum_{n=m+2}^d |f_{i,m}(n)|^2 + |b_{i,m}(n-1)|^2}
\]

(10)

**Burg method for non-Gaussian vectors**

We now consider the autoregressive vector \( X \). The estimator defined by equation (10) applied on the sample \((x_1, \ldots, x_N)\) will suffer from the disparity of the realizations of the scalar part \((\tau_1, \ldots, \tau_N)\). This lead us to consider two different criterion from equation 6.

**Normalized Energy**

The first idea could be to consider an error which is normalized with respect to \( \tau \):

\[
U^{(m+1)} = \sum_{n=m+2}^d \frac{|f_{m+1}(n)|^2 + |b_{m+1}(n)|^2}{|f_{m}(n)|^2 + |b_{m}(n-1)|^2}.
\]

(11)

The minimum of the empirical version of the previous error is then:

\[
\hat{\mu}_{m+1} = -\frac{2}{N(d-m-1)} \sum_{i=1}^N \sum_{n=m+2}^d \frac{\hat{b}_{i,m}(n-1)f_{i,m}(n)}{|f_{i,m}(n)|^2 + |b_{i,m}(n-1)|^2}.
\]

(12)

The drawback is that \( \hat{\mu}_{m+1} \) is not consistent. Indeed, from Lemma 9 , it holds

\[
\hat{\mu}_{m} \xrightarrow{a.e.} B_1(|\mu_m|) \frac{\mu_m}{|\mu_m|}
\]

with \( B_1(x) = \frac{1-x^2}{x} \left( \frac{\log(1-x) - \log(1+x)}{2x} + \frac{1}{1-x^2} \right) \). The consistent version of (12) is then obtained through:

\[
\mu_m = B_1^{-1}(|\mu_m|) \frac{\hat{\mu}_m}{|\hat{\mu}_m|},
\]

(13)
Elliptical Energy

In this section, we introduce natural estimators for a specific dependent sample in the context of the elliptical distribution. We also prove that this estimator solves a minimum energy problem for an energy functional which we call elliptic energy.

As the forward and backward errors defined by the equation 7, $e_m(n)$, are 2-dimensional elliptical random vectors with known covariance (up to a multiplicative constant) given by equation 9. We can apply the elliptical approach given in [11] in order to estimate the covariance of this vector. This leads to the following consistent estimator

$$\hat{\mu}_{m+1}^{(ell)} = \arg\min_{\mu_{m+1} \in \mathbb{C}, |\mu_{m+1}| < 1} \sum_{n=m+2}^{d} \sum_{i=1}^{N} 2 \log \left( e_{im}(n) + \left( \frac{1}{\mu_{m+1}} \mu_{m+1} + 1 \right) e_{im}(n) \right) - N(d - m - 1) \log(1 - |\mu_{m+1}|^2).$$

GEODESIC BURG ESTIMATORS

Median of Burg estimators

The idea of these estimators is to cut the sample $x_1, \ldots, x_N$ into $S$ subsamples and perform on each subsample an estimation process such as a classical Burg estimator or Normalized Burg estimators (preferably fast but not necessarily robust). Then, we will compute a "representative" for the $S$ estimates $\hat{\mu}_1, \ldots, \hat{\mu}_S$.

As representative, we will consider the median of $\hat{\mu}_1, \ldots, \hat{\mu}_S$ defined by the following minimization problems

$$\hat{\mu}_{\text{median}} = \arg\min_{\mu} \sum_{i=1}^{S} d(\mu, \hat{\mu}_i)$$

where $d$ is a distance defined on the Poincaré disk $D = \{ z \in \mathbb{C} \text{ s.t. } |z| < 1 \}$. It is natural to consider a Riemannian distance that reflects the geometry of the model. A natural metric to consider in the Poincaré disk is the Poincaré metric. It corresponds to the geometry linked to the entropy of an autoregressive vector. These median computations for this metric were the object of Arnaudon et al. [5] and especially the median computation in its generality in Riemannian spaces was studied by Yang [13].
Geodesic Median Burg estimator

The limit case $S = N$ is intuitively the most robust estimator of this geodesic class. However, it corresponds to an estimation of the scatter matrix of the autoregressive vector with a single sample. It can however otherwise be explained in the framework of autoregressive process. If the order of the underlying autoregressive $M$ is small with respect to $d$, then the regularization process we presented can be interpreted as an implicit estimation of the true order of the autoregressive process [6]. The regularization is then essential in this context.

APPLICATIONS FOR RADAR DETECTION

As an illustration of the performances of the algorithms, we present a simulation of the model with a Weibull texture for the clutter. We recall the expression of the density for a Weibull distribution of scale parameter $\sigma$ (representing the clutter power level set to 20 dB) and a shape parameter $\nu$ (representing the disparity of the distribution). We will take $N = 64$ samples and a speckle built from an autoregressive vector of order 1 and of dimension $d = 8$.

Quality of the estimation without outliers

The measure of the error of estimation for the covariance matrix has to be normalized because the scatter matrices are defined up to a multiplicative constant. The Riemannian mean error for $N$ estimations is the Riemannian normalized distance which is a natural distance in the space of positive definite matrices:

$$MCE = \frac{1}{N} \sum_{i=1}^{N} \left\| \log \left( \left( \frac{\hat{\Sigma}_i}{\| \hat{\Sigma}_i \|} \right)^{-1/2} \Sigma_0 \left( \frac{\hat{\Sigma}_i}{\| \hat{\Sigma}_i \|} \right)^{-1/2} \right) \right\|_F$$

(15)

where $\|M\|_F = Tr(MM^+)$. We will compare four proposed estimators (namely Normalized Burg, Elliptical Burg, Geodesic Median Burg and Median of Normalized Burg) with of the scatter matrix with those classically used in the literature : the empirical Burg estimator and the Fixed Point (FP) proposed by Tyler [11] in the elliptical context. The Burg estimators of the scatter matrix are used for a maximal order $M = d - 1$. Figure 2 illustrates the fact that the robustness of the different estimators with respect to the shape parameter of the Weibull texture holds except for the empirical covariance as expected.

Quality of the estimation for a clutter transition

We consider a scenario of a clutter transition between a clutter of frequency 0 (that could represent a ground clutter) and a clutter of frequency 0.3 (that could represent a sea clutter). The power of the clutters is set to $40dB$. Figure 3 and 4 show the estimated
spectra for 64 cases around each range case. We see that only "Geodesic Median Burg" and "Median of Normalized Burg" correctly estimate the transition. These estimators are said to be robust to a large contamination of the data.

**CONCLUSION**

We presented estimators for the scatter matrix of scale mixtures of stationary Gaussian autoregressive random vectors. Our estimators are inspired from Burg technique and are independent of the texture. Robust versions of these estimators through median consideration were presented showing an improvement of the estimation process in a contamination context.

**ACKNOWLEDGMENTS**

The authors would like to thank the DGA/MRIS and Thales for their support.

**REFERENCES**

Figure 2. Riemannian Mean Error for an AR(1) ($\mu_1 = 0.9$) with respect to the shape parameter of the Weibull texture.

Figure 3. Simulated spectra and estimated spectra for robust estimators:
(a) Simulated spectra representing a clutter transition
(b) Geodesic Median Burg estimator
(c) Median of Normalized Burg estimator

Figure 4. Estimated spectra for non robust estimators:
(a) Normalized Burg estimator
(b) Elliptical Burg estimator
(c) Fixed Point estimator
(d) Empirical estimator