ROBUST G-MUSIC

Romain Couillet1, Abla Kammoun2.

1Telecommunication Department, Supélec, Gif sur Yvette, France.
2King Abdullah University of Science and Technology, Saudi Arabia.

ABSTRACT
An improved MUSIC algorithm for direction-of-arrival estimation is introduced that accounts both for large array sizes \( N \) comparatively with the number of independent observations \( n \) and for the impulsiveness of the background environment (e.g., presence of outliers in the observations). This method derives from the spiked G-MUSIC algorithm proposed in [1] and from the recent works by one of the authors on the random matrix analysis of robust scatter matrix estimators [2]. The method is shown to be asymptotically consistent where classical approaches are not. This superiority is corroborated by simulations.

Index Terms—Random matrix theory, MUSIC, robust estimation, elliptical distribution.

I. INTRODUCTION

The recent advances in random matrix theory have allowed for several successive improvements of statistical inference methods when the number of observations \( n \) is not large compared to the population size \( N \). In particular in [3], a generalization of the well-known MUSIC algorithm [4] for direction-of-arrival (DoA) estimation, called G-MUSIC, was proposed which brings consistency to MUSIC in the regime where \( N, n \rightarrow \infty \) with \( N/n \rightarrow c \in (0, \infty) \).1 This work assumes that the signals are random and that each signal subspace has a rank growing with \( N \). Further extensions were in the direction of deterministic but unknown signals instead [5], and then in the direction of small signal subspace dimensions compared to \( N \) (the so-called spiked model) [1]. Finally, the generalization to non-white Gaussian noise was established in [6].

As a common denominator, all the works, which rely on subspace characterization of the sample covariance matrix of the observed samples, demand that the eigenvalues of the noise subspace have a limiting compact support. This allows one to isolate and identify the “signal eigenvalues” as those eigenvalues found away from this support. However, the eigenvalues of the noise subspace may not meet this condition when sample outliers occur, entailing false alarms, or when the background noise is very impulsive, leading then to a possibly unbounded eigenvalue support from which no signal eigenvalue can be isolated. In these scenarios, none of MUSIC or G-MUSIC (regular or spiked) are theoretically acceptable.

More recently, the second author proposed a new analytical framework of robust estimation of large dimensional scatter matrices in the series of works [7], [2], [8]. Robust estimation of scatter exactly aims at mitigating the effect of outliers and impulsiveness in the observed datasets. In particular, it is shown in [2] that the limiting eigenvalue distribution of impulsive samples has a bounded support with no eigenvalue escaping. The present article hinges on this observation by proposing a new MUSIC improvement, called robust G-MUSIC, which works under the spiked signal space assumption and accounts both for the large dimension of the antenna array and for the impulsive nature of the noise. The key idea is to trade the sample covariance matrix for a robust scatter estimate as studied in [2]. It appears that robust G-MUSIC manages to recover both the signal powers and their DoA in impulsive scenarios where MUSIC and G-MUSIC cannot. In particular, and maybe quite surprisingly at first sight, the low rank population signal subspace (if powerful enough) engenders isolated – hence detectable – eigenvalues in the robust scatter estimate found beyond a certain detectability threshold, while low rank sample outliers remain either contained in the noise subspace or found as isolated eigenvalues below the detectability threshold. Simulations are performed that corroborate the theoretical analysis.

The remainder of the article introduces the array processing system model in Section II and our main results in Section III.

II. MODEL

We consider the model

\[
y_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_l} w_i, \quad i \in \{1, \ldots, n\} \tag{1}
\]

with \( a_l = a(\theta_l) \in \mathbb{C}^N \) a unit norm directional vector parametrized by \( \theta_l \in [0, 2\pi) \), \( p_l \in \mathbb{R}^+ \) a power parameter, \( s_{li} \in \mathbb{C} \) random i.i.d. variables with zero mean and unit

\[1\]
variance, \( \tau_i \) random i.i.d. variables with distribution \( \nu \) such that \( \int t\nu(dt) = 1 \) and \( \int (1 + \varepsilon)\nu(dt) < \infty \) for some \( \varepsilon > 0 \), and \( w_j \in \mathbb{C}^N \) a unitarily invariant norm-\( \sqrt{N} \) vector. As for \( L \), it is assumed finite.

Our objective is to estimate the powers \( p_1, \ldots, p_L \) along with the angles \( \theta_1, \ldots, \theta_L \) from the independent samples \( y_1, \ldots, y_n \) in the regime \( N, n \to \infty \) with \( N/n \to c \in (0, 1) \).

Denoting \( \tilde{w}_i = [s_1, \ldots, s_L, w_1]^T \in \mathbb{C}^{N+L} \) (made of independent entries of zero mean and unit variance) and \( A_i = [\sqrt{n}a_1, \ldots, \sqrt{n}a_L, \sqrt{n}I_N] \), we have \( y_i = A_i\tilde{w}_i \).

We also denote
\[
A_i A_i^* = \sum_{l=1}^L p_l a_l a_l^* + \tau_l I_N \triangleq B + \tau_i I_N.
\]

We now introduce Maronna’s robust estimator of scatter \( \hat{C}_N \) as in [2]. The matrix \( \hat{C}_N \) is defined, when it exists, as the unique solution to the fixed-point matrix-valued equation in \( Z \):
\[
Z = \frac{1}{n} \sum_{i=1}^n u \left( \frac{1}{n} y_i^* Z^{-1} y_i \right) y_i y_i^* \tag{2}
\]
where \( u : \mathbb{R}^+ \to \mathbb{R}^+ \) is a nonnegative non-increasing function with \( u(0) = 0 \) and such that \( \phi(x) = xu(x) \) is increasing and bounded with \( \phi_\infty = \lim_{x \to \infty} \phi(x) \in (1, 1/e) \).

We recall from [2] that, if \( \hat{C}_N \) exists, then it can be written
\[
\hat{C}_N = \frac{1}{n} \sum_{i=1}^n v \left( \frac{1}{n} y_i^* \hat{C}_N^{-1} y_i \right) y_i y_i^* \tag{3}
\]
where \( v : x \mapsto u \circ g^{-1} \circ g : x \mapsto x/(1 - c\phi(x)) \), and \( \hat{C}_N = \hat{C}_N - \frac{1}{n} u \left( \frac{1}{n} y_i^* Z^{-1} y_i \right) y_i y_i^*. \) We shall also denote \( \psi(x) = x\phi(x) \). Recall that \( v \) is non-increasing and \( \psi \) is increasing with limit \( \psi_\infty = \phi_\infty/(1 - c\phi_\infty) \).

### III. MAIN RESULTS

#### III-A. Theoretical results

Our main results are twofold. Proposition 1 first generalizes [2, Theorem 2] on the asymptotic behavior of \( \hat{C}_N \) in the large \( N, n \) regime. Then Theorem 1 and Theorem 2 provide statistical inference methods that will be used to estimate the powers \( p_1, \ldots, p_L \) and angles \( \theta_1, \ldots, \theta_L \).

**Proposition 1 (Asymptotic Behavior):** Let \( \hat{C}_N \) be defined as the unique solution of (2) when it exists or arbitrarily otherwise. Then, as \( N, n \to \infty \) with \( N/n \to c \in (0, 1) \),
\[
\|\hat{C}_N - \hat{S}_N\| \xrightarrow{a.s.} 0 \tag{4}
\]
where
\[
\hat{S}_N \triangleq \frac{1}{n} \sum_{i=1}^n v(\gamma_i)A_i \hat{w}_i \hat{w}_i^* A_i^*.
\]
with \( \gamma \) the unique positive solution to
\[
1 = \int \frac{\psi(t\gamma)}{1 + c\psi(t\gamma)} \nu(dt). \tag{5}
\]

**Sketch of Proof:** We only provide an intuitive sketch of the proof. The complete proof is found in an extended version of this article. The asymptotic uniqueness of \( \hat{C}_N \) is an extension of [2, Theorem 1] which follows from the same ideas and is therefore not presented here. The interesting aspect is the convergence (3) which relates to [2, Theorem 2].

Note in particular that, for \( B = 0 \), the result matches [2, Theorem 2]. Following the intuitive approach delineated in [2, Section 2.2], denoting \( \hat{C}_{(i)} = \hat{C}_N - \frac{1}{n} u(\hat{y}_i^* \hat{C}_N^{-1} y_i) y_i y_i^* \), we expect that, for large \( N, n \)
\[
\frac{1}{N} y_i^* \hat{C}_{(i)}^{-1} y_i \approx \frac{1}{N} \text{tr} A_j A_j^* \hat{C}_N^{-1} = \frac{1}{N} \text{tr} B \hat{C}_N^{-1} + \tau_j \frac{1}{N} \text{tr} \hat{C}_N^{-1}.
\]

Denoting \( \hat{\alpha}_N = \frac{1}{N} \text{tr} B \hat{C}_N^{-1} \) and \( \hat{\tau}_N = \frac{1}{N} \text{tr} \hat{C}_N^{-1} \), we expect from random matrix intuition that \( \hat{\alpha}_N - \alpha_N \xrightarrow{a.s.} 0 \) and \( \hat{\tau}_N - \tau_N \xrightarrow{a.s.} 0 \) as \( N, n \to \infty \) for some deterministic equivalents \( \alpha_N, \tau_N > 0 \). If so, approximating \( \frac{1}{N} y_i^* \hat{C}_{(i)}^{-1} y_i \) by \( \alpha_N + \tau_j \gamma_N \) in the expression of \( \hat{C}_{(i)} \) and applying classical random matrix results (for instance from [9, Theorem 1]), we must have
\[
\gamma_N \approx \frac{N}{\text{tr} B} \left( \frac{n}{\text{tr} B} \right)^{-1}
\]
where the \( e_j \) are uniquely defined as the solution of
\[
e_j = v(\alpha_N + \tau_j \gamma_N) \times \frac{1}{N} \text{tr} A_j A_j^* \left( \frac{n}{\text{tr} B} \right)^{-1}
\]
The latter further simplifies as
\[
e_j = c\psi(\alpha_N + \tau_j \gamma_N) = c\psi(\alpha_N + \tau_j \gamma_N).
\]
This then gives the system
\[
\gamma_N = \frac{1}{N} \left( \frac{n}{\text{tr} B} \right)^{-1}
\]

It now follows from the small-rank property and norm boundedness of the vectors composing \( B \) that \( \alpha_N \to 0 \). Also, the small rank of \( B \) does not alter asymptotically the trace term in \( \gamma_N \) so that \( B \) can be discarded in this expression in the limit. Hence \( \gamma_N \to \gamma \) which, outside a zero measure set, satisfies (4) from the strong law of large numbers.

Proposition 1 states that, for many problems dealing with \( \hat{C}_N \), which is a complicated matrix defined as the solution of a fixed-point equation, one can instead use \( \hat{S}_N \), a much simpler object of the bi-correlated sample covariance matrix.
type, extensively analyzed in [10]. Indeed Proposition 1 ensures as a fundamental corollary that, denoting \( \lambda_1(X) \leq \ldots \leq \lambda_N(X) \) the ordered eigenvalues of a Hermitian matrix \( X \),

\[
\text{max}_{1 \leq k \leq N} |\lambda_k(\hat{C}_N) - \lambda_k(\hat{S}_N)| \xrightarrow{a.s.} 0
\]

so that the largest eigenvalues of \( \hat{C}_N \) coincide asymptotically with those of \( \hat{S}_N \) and, as far as (first order) G-estimation is concerned, we can equivalently work with \( \hat{C}_N \) or \( \hat{S}_N \).

A thorough analysis of the expression of \( \hat{S}_N \) now reveals that it has a limiting eigenvalue distribution with compact support and may exhibit isolated eigenvalues outside its limiting support. These isolated eigenvalues may be of two types. Some may first arise as sample outliers from large values of \( \tau_j \), which, as established by [2, Theorem 2], cannot exceed a limiting value. In particular, since \( \hat{S}_N \leq \frac{\phi}{\gamma(1-c\phi)} \sum_{i=1}^{N} w_i w_i^{*} \) when \( B = 0 \), from standard random matrix considerations, we find that, for arbitrary \( \varepsilon > 0 \), these eigenvalues cannot exceed \( S^* + \varepsilon \) infinitely often, where we defined

\[
S^* \triangleq \frac{\phi}{\gamma(1-c\phi)} (1 + \sqrt{c})^2.
\]

On the opposite, by the presence of the small rank \( B \neq 0 \) in each term \( A_i A_i^{*} \), a deeper analysis (similar to [11]) reveals that some other eigenvalues may be found beyond \( S^* + \varepsilon \) infinitely often. The value \( S^* \), although not optimal, will from now on constitute our signal detectability threshold for our improved signal classification algorithm based on \( \hat{C}_N \).

We are now in position to introduce our first main result.

**Theorem 1 (Estimation with known \( \nu \))**: Denote \( q_1 \geq \ldots \geq q_L > 0 \) the positive eigenvalues of \( B \) (they may depend of \( N \)), \( u_1, \ldots, u_L \) their respectively associated eigenvectors in \( B \), and \( \hat{u}_1, \ldots, \hat{u}_N \) the eigenvectors of \( \hat{C}_N \) respectively associated with the eigenvalues \( \lambda_1 \geq \ldots \geq \lambda_N \). For \( x > 0 \) sufficiently large, call \( \delta(x) \) the unique negative solution of

\[
\delta(x) = c \left( -x + \int \frac{tv(t\gamma)}{1 + \delta(x)tv(t\gamma)} \nu(dt) \right)^{-1}.
\]

Then, recalling the definition of \( \gamma \) as the solution of (4), we have the following two results:

1. **Eigenvalue estimation.** For each \( k \) with \( \lim \inf_N q_k \) sufficiently large,

\[
- \epsilon \left( \delta(\hat{\lambda}_k) \int \frac{tv(t\gamma)}{1 + \delta(\hat{\lambda}_k)tv(t\gamma)} \nu(dt) \right)^{-1} - q_k \xrightarrow{a.s.} 0.
\]

2. **Localization function estimation.** Assume that \( \lim \inf_N q_L \) is large enough. Denote \( \eta_N, \hat{\eta}_N \colon [0, 2\pi) \rightarrow \mathbb{R} \) the functions given by

\[
\eta_N(\theta) = 1 - \sum_{k=1}^{L} a(\theta)^* u_k \hat{u}_k a(\theta)
\]

\[
\hat{\eta}_N(\theta) = 1 - \sum_{k=1}^{L} w_k a(\theta)^* \hat{u}_k \hat{u}_k a(\theta)
\]

where \( w_j > 0 \) is given in (5). Then, for each \( \theta \in [0, 2\pi) \),

\[
\hat{\eta}_N(\theta) - \eta_N(\theta) \xrightarrow{a.s.} 0
\]

as \( N, n \rightarrow \infty \) and \( N/n \rightarrow c \).

**Proof**: The proof starts by transferring the difficult analysis of the behavior of \( \hat{C}_N \) into the simpler analysis of \( \hat{S}_N \), thanks to Theorem 1. This being done, as \( \hat{S}_N \) is a random matrix of a rather standard form, the proof then follows similar steps as in the proof of the first order results in [12] (see also [11]), although for a slightly different spiked model. For lack of space, we do not further detail these here. The complete derivation can be found in the extended version of this article [13].

Theorem 1 states that, based on the largest eigenvalues of \( \hat{C}_N \), one can consistently estimate the values of \( q_1, \ldots, q_L \) and the localization functions \( \eta_N(\theta) \) for each \( \theta \).

Note that the theorem mentions that \( q_k \) should be large enough. This sufficiently large value ensures that the number of eigenvalues found above the detectability threshold \( S^* \) must equal at least \( k \) (see [13] for more precise statements). Without this assumption, the above estimates may be invalid as some of the largest eigenvalues may correspond to sample noise outliers or eigenvalues from the noise bulk. This condition is classical and relates to the often named separability condition by which population eigenvalues \( q_k \) engender sample eigenvalues found away from a detectability threshold in the spectrum of \( \hat{C}_N \) (or equivalently of \( \hat{S}_N \)).

Note importantly that, if the sample covariance matrix were used in place of \( \hat{C}_N \) (therefore leading to a G-MUSIC
algorithm), the impulsive nature of noise may not allow for the eigen-spectrum of noise subspace to be of compact support, making it impossible to separate signal eigenvalues.

Theorem 1 assumes $\nu$ to be known. Using instead the fact that $\gamma - \hat{\gamma}_N \xrightarrow{a.s.} 0$ where

$$
\hat{\gamma}_N = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N} y^*_i \hat{C}_i^{-1} y_i
$$

and that $\tau_i$ can be well approximated by

$$
\hat{\tau}_i = \frac{1}{\gamma_N} \frac{1}{N} y^*_i \hat{C}_i^{-1} y_i
$$

(where we recall that $\hat{C}_N = \frac{1}{n} u^*_N (\frac{1}{N} y^*_N \hat{C}_N y_N) y_N u_N^*$), we deduce the following alternative empirical estimator.

**Theorem 2 (Empirical estimation):** Define $\delta(x)$ as the unique solution to

$$
\hat{\delta}(x) = \frac{N}{n} \left( -x + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \delta(x) \hat{\tau}_i v(\hat{\gamma}_N)} \right)^{-1}.
$$

Then we have the following results

1. **Eigenvalue estimation.** For each $k$ with $\lim \inf_N q_k$ large enough,

$$
- \left( \hat{\delta}(\hat{\lambda}_k) \frac{1}{N} \sum_{i=1}^{n} \frac{v(\hat{\tau}_i \hat{\gamma}_N)}{1 + \hat{\delta}(\hat{\lambda}_k) v(\hat{\tau}_i \hat{\gamma}_N)} \right)^{-1} - q_k \xrightarrow{a.s.} 0.
$$

2. **Localization function estimation.** Assume that $\lim \inf_N q_1$ is large enough. Denote $\eta_N^{\text{emp}} : [0, 2\pi) \rightarrow \mathbb{R}$ the function given by

$$
\eta_N^{\text{emp}}(\theta) = 1 - \sum_{k=1}^{L} \hat{w}_k a(\theta)^* \hat{u}_k \hat{u}_k^* a(\theta)
$$

where $\hat{w}_j$ is defined in (6). Then, for each $\theta \in [0, 2\pi)$,

$$
\eta_N^{\text{emp}}(\theta) - \eta_N(\theta) \xrightarrow{a.s.} 0.
$$

Note that Theorem 2 provides another set of consistent estimators for the same quantities $q_k$ and $\eta_N(\theta)$ as in Theorem 1 but that do not rely on the knowledge of $\nu$, as the $\tau_i$ are individually estimated. There is however no saying whether knowing $\nu$ or not improves the quality of the estimates.

**III-B. Application to array processing**

In this section, we use Theorem 1 and Theorem 2 to perform source detection, power estimation, and source localization in array processing.

We consider the model (1) in which $s_{lj}$, $l = 1, \ldots, L$ are $L \geq 0$ signal sources impinging at time instant $j$ an $N$-antenna array with angles $\theta_1, \ldots, \theta_L$. The signal $y_i$ received at time $i$ at the array is corrupted by the background noise $\sqrt{\tau_i} w_i$ with $w_i$ unitary and $\tau_i$ having unit-mean distribution $\nu$. In particular, if $\nu = \nu' / \int t dt' \nu'(t)$, then

$$
\nu' \sim \chi^2_{2N}
$$

(the chi-square distribution with $2N$ degrees of freedom), the noise is standard Gaussian. We however consider here scenarios where the noise may have a more impulsive behavior. In particular, one may take $\nu = \nu'/ \int t dt' \nu'(t)$ with $\nu' \sim (1-\varepsilon) \nu_G + \varepsilon \delta_\nu$ for some $\varepsilon$ small and a large $A$ value, introducing then outliers of high amplitude, which simulates strong but rare noise impulses. One may alternatively take $\nu$ continuous with heavier than Gaussian tails, leading in particular to scenarios where the limiting spectrum of the sample covariance matrix has unbounded support, translating here the large impulsive nature of the background noise.

Note now that $\eta_N(\theta)$ (in Theorem 1) can be written $\eta_N(\theta) = a(\theta)^* \Pi_W a(\theta)$ with $\Pi_W = I_N - \sum_{k=1}^{L} u_k u_k^*$ a projector on the population noise subspace. This is then the localization function found in the standard MUSIC algorithm [4]. Item 2) of Theorems 1 and 2 thus provides asymptotically consistent estimates for this localization function, under knowledge of $\nu$ or not. From Theorems 1 and 2, we may then define two novel DoA estimates $\hat{\theta}_1, \ldots, \hat{\theta}_L$, and $\hat{\theta}_1^{\text{emp}}, \ldots, \hat{\theta}_L^{\text{emp}}$ for the angles $\theta_1, \ldots, \theta_L$, which we hereafter call the robust G-MUSIC and empirical robust G-MUSIC estimates, respectively, and which we define as the deepest minima of the functions $\theta \mapsto \eta_N(\theta)$ and $\theta \mapsto \eta_N(\theta)^{\text{emp}}$, respectively. For comparison, we recall that the standard MUSIC algorithm assumes as an estimate for $\eta_N(\theta)$ the function $\theta \mapsto a(\theta)^* \Pi_W^{\text{SCM}} a(\theta)$ with $\Pi_W^{\text{SCM}}$ the noise subspace of the sample covariance matrix for the vectors $y_1, \ldots, y_n$. Robust (non-G) MUSIC is similar but with $\hat{C}_N$ in place of the sample covariance matrix. The (non-robust) G-MUSIC algorithm is similar to the robust G-MUSIC method but with the function $v$ set equal to one, which is somewhat similar to the G-MUSIC approach developed in [6]. In addition to G-MUSIC, we introduce the empirical version of G-MUSIC which, similar to the empirical robust G-MUSIC approach, uses estimates of $\tau_i$ instead of using the information of $\nu$.

As is standard in array processing, we may assume that $\alpha(\theta)$ is such that, for each $i \neq j$, $\alpha(\theta_j)^* \alpha(\theta_i) \rightarrow 0$. This is in particular the case if $[\alpha(\theta)]_k = N^{-\frac{1}{2}} \exp(2\pi i (k-1)\theta)$, corresponding to a uniform linear array. In this scenario, it is easily seen that, as $N \rightarrow \infty$, $q_i \rightarrow p_i$ for each $i \in \{1, \ldots, L\}$. The item 1) of both Theorems 1 and 2 can then be used here to estimate $p_1, \ldots, p_L$ under knowledge or not of $\nu$. Similar to the localization function estimate, for some $\nu$, classical sample covariance matrix-based techniques to estimate the $p_i$ such as [3], [14] cannot be adapted to account for the $\tau_i$.

Figure 1 compares the performance of robust G-MUSIC and empirical robust G-MUSIC against MUSIC and G-MUSIC for $L = 2$, $\theta_1 = 10^\circ$, $\theta_2 = 12^\circ$, and $\nu \sim$
Fig. 1. Mean square error (MSE) performance of the smallest angle estimation of the MUSIC estimators for $N = 20$, $n = 100$, two sources at $10^\circ$ and $12^\circ$, Gamma-$(0.8, 1.25)$ impulsions. Robust G-MUSIC (RG-MUSIC) and empirical robust G-MUSIC (Emp. RG-MUSIC) are compared against MUSIC, G-MUSIC, and empirical G-MUSIC (Emp. G-MUSIC). Gamma$(0.8, 1.25)$. This figure shows that, due to $N/n$ not being too small, MUSIC and robust MUSIC cannot separate the two close angles, which is a well-documented behavior. The G-MUSIC approaches accommodate for this, but suffer from the impulsive noise nature, and even more so if the shape parameter of the Gamma distribution gets reduced, as can be verified by further simulations that can be found, along with additional discussions, in [13].

IV. REFERENCES


