

# A JOINT ROBUST ESTIMATION AND RANDOM MATRIX FRAMEWORK WITH APPLICATION TO ARRAY PROCESSING

Romain Couillet<sup>1</sup>, Frédéric Pascal<sup>2</sup>, and Jack W. Silverstein<sup>3</sup>

<sup>1</sup> Telecommunication department, Supélec, Gif sur Yvette, France.

<sup>2</sup> SONDRALaboratory, Supélec, Gif sur Yvette, France.

<sup>3</sup> Department of Mathematics, North Carolina State University, NC, USA.

## ABSTRACT

An original interface between robust estimation theory and random matrix theory for the estimation of population covariance matrices is proposed. Consider a random vector  $x = A_N y \in \mathbb{C}^N$  with  $y \in \mathbb{C}^M$  made of  $M \geq N$  independent entries,  $E[y] = 0$ , and  $E[yy^*] = I_N$ . It is shown that a class of robust estimators  $\hat{C}_N$  of  $C_N = A_N A_N^*$ , obtained from  $n$  independent copies of  $x$ , is  $(N, n)$ -consistent with the traditional sample covariance matrix  $\hat{S}_N$  in the sense that  $\|\hat{C}_N - \alpha \hat{S}_N\| \rightarrow 0$  in spectral norm for some  $\alpha > 0$ , almost surely, as  $N, n \rightarrow \infty$  with  $N/n$  and  $M/N$  bounded. This result, in general not valid in the fixed  $N$  regime, is used to propose improved subspace estimation techniques, among which an enhanced direction-of-arrival estimator called robust G-MUSIC.

**Index Terms**—random matrix theory, robust estimation.

## I. INTRODUCTION

Many multi-variate signal processing detection and estimation techniques are based on the empirical covariance matrix of a sequence of samples  $x_1, \dots, x_n$  from a random population vector  $x \in \mathbb{C}^N$ . Assuming  $E[x] = 0$  and  $E[xx^*] = C_N$ , the strong law of large numbers ensures that, for independent and identically distributed (i.i.d.) samples,  $\hat{S}_N = \frac{1}{n} \sum_{i=1}^n x_i x_i^* \rightarrow C_N$  almost surely (a.s.), as the number  $n$  of samples increases and  $N$  is fixed. Many subspace methods, such as the multiple signal classifier (MUSIC) algorithm and its derivatives [1], [2], heavily rely on this property by identifying  $C_N$  with  $\hat{S}_N$ , leading to appropriate approximations of functionals of  $C_N$  in the large  $n$  regime. However, this standard approach has two major limitations: the inherent inadequacy to small sample sizes (when  $n$  is not too large compared to  $N$ ) and the lack of robustness to outliers or heavy-tailed distribution of  $x$ .

The *sample covariance matrix* (SCM)  $\hat{S}_N$  is an object of primal interest since it is the maximum likelihood estimator of  $C_N$  for  $x$  Gaussian. When  $x$  is not Gaussian, the SCM as an approximation of  $C_N$  may however perform very poorly. This was particularly recognized in adaptive radar and sonar processing, where the signals under study are characterized by impulsive noise and outlying data. Robust estimates of  $C_N$  aim at tackling this problem [3], [4] and have imposed themselves as an appealing alternative to the SCM. These estimators, denoted  $\hat{C}_N$  here, are usually defined implicitly as a solution of an equation of the type

$$\hat{C}_N = \frac{1}{n} \sum_{i=1}^n u \left( \frac{1}{N} x_i^* \hat{C}_N^{-1} x_i \right) x_i x_i^* \quad (1)$$

Silverstein's work is supported by the U.S. Army Research Office, Grant W911NF-09-1-0266. Couillet's work is supported by the ERC Grant MORE EC-120133.

for  $u$  a nonnegative function with specific properties. These estimators are particularly appropriate as they are the maximum likelihood estimates of (a scaled version of)  $C_N$  for specific distributions of  $x$ , such as the family of elliptical distributions [5]. They are also used to cope with distributions of  $x$  with heavier-than-Gaussian tails, such as the K-distribution often met in the context of adaptive radar processing with impulsive clutter [6].

A second angle of improvement of subspace methods has recently emerged due to advances in random matrix theory. The latter aims at studying the statistical properties of matrices in the regime where both  $N$  and  $n$  grow large. It is known in particular that, if  $x = A_N y$  with  $y \in \mathbb{C}^M$ ,  $M \geq N$ , a vector of independent entries with zero mean and unit variance, then, under some conditions on  $C_N = A_N A_N^*$  and  $y$ , in the large  $N, n$  (and  $M$ ) regime, the eigenvalue distribution of (almost every)  $\hat{S}_N$  converges weakly to a limiting distribution described implicitly by its Stieltjes transform [7]. In the past ten years, this result and subsequent works have been applied to revisit classical signal processing techniques assuming  $n \gg N$  and turn them into improved methods assuming  $N$  and  $n$  of the same order of magnitude.

In this article, we study the asymptotic first order properties of the robust M-estimate  $\hat{C}_N$  of  $C_N$  as  $N, n$  (and  $M$ ) grow large simultaneously. Under the assumption that  $x$  is of the type  $x = A_N y$  with  $y$  having independent zero-mean entries, it is possible to prove that  $\hat{C}_N$  and  $\hat{S}_N$  have a close behaviour. Our main contribution consists in showing that, in the large  $N, n$  regime, and under some mild assumptions,  $\|\hat{C}_N - \alpha \hat{S}_N\| \rightarrow 0$ , a.s., for some  $\alpha > 0$  to be defined. A major consequence of our result is that the matrix  $\hat{S}_N$ , at the core of many random matrix-based estimators, can be straightforwardly replaced by  $\hat{C}_N$  without altering the first order properties of these estimators. We generically call the induced estimators *robust G-estimators*. As an application example, we provide a robust direction-of-arrival estimator, referred to as robust G-MUSIC, based on the G-MUSIC estimator from Mestre [8].

The remainder of the article is structured as follows. Section II provides our theoretical results. Section III introduces the robust G-MUSIC estimator. Section IV then concludes the article. The detailed proofs are available in the extended version of the present article [9].

*Notations:* The arrow ‘ $\xrightarrow{\text{a.s.}}$ ’ denotes almost sure convergence. The norm  $\|\cdot\|$  is the spectral norm for matrices and the Euclidean norm for vectors.  $A^T$  and  $A^*$  are the transpose and Hermitian transpose of  $A$ , respectively. For  $A \in \mathbb{C}^{N \times N}$  Hermitian,  $\lambda_1(A) \leq \dots \leq \lambda_N(A)$  are its ordered eigenvalues.

## II. MAIN RESULTS

Let  $X = [x_1, \dots, x_n] \in \mathbb{C}^{N \times n}$ , where  $x_i = A_N y_i \in \mathbb{C}^N$ , with  $y_i = [y_{i1}, \dots, y_{iM}]^\top \in \mathbb{C}^M$  having independent entries with zero mean and unit variance,  $A_N \in \mathbb{C}^{N \times M}$ , and  $C_N \triangleq A_N A_N^* \in \mathbb{C}^{N \times N}$  be a positive definite matrix. We denote  $c_N \triangleq N/n$ ,  $\bar{c}_N \triangleq M/N$ , and define the sample covariance matrix  $\hat{S}_N$  of the sequence  $x_1, \dots, x_n$  by

$$\hat{S}_N \triangleq \frac{1}{n} X X^* = \frac{1}{n} \sum_{i=1}^n x_i x_i^*.$$

Let  $u : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  ( $\mathbb{R}^+ = [0, \infty)$ ) be a function fulfilling the following conditions:

- (i)  $u$  is nonnegative, nonincreasing, and continuous on  $\mathbb{R}^+$ ;
- (ii) the function  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ ,  $s \mapsto su(s)$  is nondecreasing and bounded, with  $\sup_x \phi(x) = \phi_\infty > 1$ . Moreover,  $\phi$  is increasing in the interval where  $\phi(s) < \phi_\infty$ .

Classical M-estimators  $\hat{C}_N$  defined by (1) for such function  $u$  include the Huber estimator, with  $\phi(s) = s$  for  $s \in [0, \phi_\infty]$ ,  $\phi_\infty > 1$ , and  $\phi(s) = \phi_\infty$  for  $s \geq \phi_\infty$ . Since  $u(s) = 1$  for  $s \leq \phi_\infty$  and decreases for  $s \geq \phi_\infty$ , this estimator weights the majority of the samples  $x_1, \dots, x_n$  by a factor 1 and reduces the impact of outliers.

To pursue, we need the following statistical assumptions.

**A1.** The random variables  $y_{ij}$ ,  $i \leq n$ ,  $j \leq M$ , are independent either real or circularly symmetric complex (i.e.  $\mathbb{E}[y_{ij}^2] = 0$ ) with  $\mathbb{E}[y_{ij}] = 0$  and  $\mathbb{E}[|y_{ij}|^2] = 1$ . Also, there exists  $\eta > 0$  and  $\alpha > 0$ , such that, for all  $i, j$ ,  $\mathbb{E}[|y_{ij}|^{8+\eta}] < \alpha$ .

**A2.**  $\bar{c}_N \geq 1$  and, as  $n \rightarrow \infty$ ,

$$0 < \liminf_n c_N \leq \limsup_n c_N < 1, \quad \limsup_n \bar{c}_n < \infty.$$

**A3.** There exists  $C_-, C_+ > 0$  such that

$$C_- < \liminf_N \{\lambda_1(C_N)\} \leq \limsup_N \{\lambda_N(C_N)\} < C_+.$$

Note that the assumptions neither request the entries of  $y$  to be identically distributed nor impose the existence of a continuous density. The requirement of independence in the entries of  $y$  is nonetheless rather uncommon in robust estimation theory and excludes a number of practical applications. This assumption is however central in this article for the emergence of a concentration of the quadratic forms  $\frac{1}{N} x_i^* \hat{C}_N^{-1} x_i$ ,  $i = 1, \dots, n$ . Generalizations to e.g. elliptical distributions would break this effect and are therefore left to future work.

Technically, **A1–A3** mainly ensure that the eigenvalues of  $\hat{S}_N$  and  $\hat{C}_N$  lie within a compact set away from zero, a.s., for all  $N, n$  large, which is a consequence (although non immediate) of [10], [11]. Note also that **A2** demands  $\liminf_N c_N > 0$ , so that the following results *do not* contain the results from [4], [12], in which  $N$  is fixed and  $n \rightarrow \infty$ , as special cases. With these assumptions, we can now provide the main technical result of this article.

*Theorem 1:* Assume **A1–A3** and consider the following matrix-valued fixed-point equation in  $Z \in \mathbb{C}^{N \times N}$ ,

$$Z = \frac{1}{n} \sum_{i=1}^n u \left( \frac{1}{N} x_i^* Z^{-1} x_i \right) x_i x_i^*. \quad (2)$$

Then, we have the following results.

- (I) There exists a unique solution to (2) for all large  $N$  a.s. We denote  $\hat{C}_N$  this solution, given by

$$\hat{C}_N = \lim_{t \rightarrow \infty} Z^{(t)}$$

where  $Z^{(0)} = I_N$  and, for  $t \in \mathbb{N}$ ,

$$Z^{(t+1)} = \frac{1}{n} \sum_{i=1}^n u \left( \frac{1}{N} x_i^* (Z^{(t)})^{-1} x_i \right) x_i x_i^*.$$

- (II) Defining  $\hat{C}_N = I_N$  when (2) does not have a unique solution, we also have

$$\left\| \phi^{-1}(1) \hat{C}_N - \hat{S}_N \right\| \xrightarrow{\text{a.s.}} 0.$$

*Proof:* A sketch of the proof is provided in the appendix. The complete proof is available in the extended article [9]. ■

An immediate corollary of Theorem 1 is the asymptotic closeness of the ordered eigenvalues of  $\phi^{-1}(1) \hat{C}_N$  and  $\hat{S}_N$ .

*Corollary 1:* Under the assumptions of Theorem 1,

$$\max_{i \leq N} \left| \phi^{-1}(1) \lambda_i(\hat{C}_N) - \lambda_i(\hat{S}_N) \right| \xrightarrow{\text{a.s.}} 0.$$

Some comments are called for to understand Theorem 1 in the context of robust M-estimation.

Theorem 1–(I) can be first compared to the result from Maronna [4, Theorem 1] which states that a solution to (2) exists for each set  $\{x_1, \dots, x_n\}$  under certain conditions on the dimension of the space spanned by the  $n$  vectors, as well as on  $u(s)$ ,  $N$ , and  $n$  (in particular  $u(s)$  must satisfy  $\phi_\infty > n/(n - N)$  in [4]). Our result is more interesting in practice in the sense that  $N, n$  no longer condition  $\phi_\infty$  and therefore do not constrain the definition of  $u(s)$ , as long as  $N, n$  are taken large. Theorem 1–(I) can also be compared to the results on uniqueness [4], [12] which hold for all  $N, n$  under some further conditions on  $u(s)$ , such as  $\phi(s)$  is strictly increasing in [4]. The latter assumption is particularly demanding as it may reject some M-estimators such as the Huber M-estimator for which  $\phi(s)$  is constant for large  $s$ .

Theorem 1–(II), which is our main result, states that, as  $N$  and  $n$  grow large with a non trivial limiting ratio, the fixed-point solution  $\hat{C}_N$  is getting asymptotically close to the sample covariance matrix, up to a scaling factor. This implies in particular that, while  $\hat{C}_N$  is an  $n$ -consistent estimator of (a scaled version of)  $C_N$  for  $n \rightarrow \infty$  and  $N$  fixed, in the large  $N, n$  regime it has many of the same first order statistics as  $\hat{S}_N$ . This suggests that many results holding for  $\hat{S}_N$  in the large  $N, n$  regime should also hold for  $\hat{C}_N$ , at least concerning first order convergence.

In terms of applications to signal processing, recall first that the  $n$ -consistency results on robust estimation [4], [12] imply that many metrics based on functionals of  $C_N$  can be consistently estimated by replacing  $C_N$  by  $\phi^{-1}(1) \hat{C}_N$ . Theorem 1 suggests instead that this approach will lead in general to inconsistent estimators in the large  $N, n$  regime, and therefore to inaccurate estimates for moderate values of  $N, n, M$ . However, any metric based on  $C_N$ , and for which an  $(N, n)$ -consistent estimator involving  $\hat{S}_N$  exists, may still be  $(N, n)$ -consistently estimated by replacing  $\hat{S}_N$  by  $\phi^{-1}(1) \hat{C}_N$ . In the following section, we give a concrete example in the context of MUSIC-like estimation in array processing [8].

### III. APPLICATION: ROBUST G-MUSIC

Consider  $K$  signal sources impinging on a collection of  $N$  collocated sensors with angles of arrival  $\theta_1, \dots, \theta_K$ . The data  $x_t \in \mathbb{C}^N$  received at time  $t$  at the array is modeled as

$$x_t = \sum_{k=1}^K \sqrt{p_k} s(\theta_k) z_{k,t} + \sigma w_t$$

where  $s(\theta) \in \mathbb{C}^N$  is the deterministic unit norm steering vector for signals impinging the sensors at angle  $\theta$ ,  $z_{k,t} \in \mathbb{C}$  is the signal source modeled as a zero mean, unit variance, and finite  $8+\eta$  order moment random variable, i.i.d. across  $t$  and independent across  $k$ ,  $p_k > 0$  is the transmit power of source  $k$  ( $p_k < p_{\max}$  for some  $p_{\max} > 0$ ) and  $\sigma w_t \in \mathbb{C}^N$  is the received noise at time  $t$ , independent across  $t$ , with i.i.d. zero mean, variance  $\sigma^2 > 0$ , and finite  $8+\eta$  order moment entries.

We can write

$$x_t = A_N y_t, \quad A_N \triangleq \begin{bmatrix} S(\Theta) P^{\frac{1}{2}} & \sigma I_N \end{bmatrix}$$

where  $S(\Theta) = [s(\theta_1), \dots, s(\theta_K)]$ ,  $P = \text{diag}(p_1, \dots, p_K)$ , and  $y_t = (z_{1,t}, \dots, z_{K,t}, w_t^T)^T \in \mathbb{C}^{N+K}$ .

Taking  $n$  independent observations  $x_1, \dots, x_n$  of the process  $x_t$  and assuming  $n, N$ , and  $M = N + K$  large accordingly to Assumption **A2**, Assumptions **A1–A3** are met and Theorem 1 can be applied. This yields the following result.

*Theorem 2 (Robust G-MUSIC):* Under the current model, denote  $E_W \in \mathbb{C}^{N \times (N-K)}$  a matrix containing in columns the eigenvectors of  $C_N$  with eigenvalue  $\sigma^2$ . Also denote  $\hat{e}_k$  the eigenvector of  $\hat{C}_N$  with eigenvalue  $\hat{\lambda}_k \triangleq \lambda_k(\hat{C}_N)$  (recall that  $\hat{\lambda}_1 \leq \dots \leq \hat{\lambda}_N$ ), with  $\hat{C}_N$  defined as in Theorem 1 (with  $\hat{C}_N = I_N$  when (2) does not have a unique solution). Then, as  $N, n \rightarrow \infty$  in the regime of Assumption **A2**, and  $K$  fixed,

$$\gamma(\theta) - \hat{\gamma}(\theta) \xrightarrow{\text{a.s.}} 0$$

where

$$\begin{aligned} \gamma(\theta) &= s(\theta)^* E_W E_W^* s(\theta) \\ \hat{\gamma}(\theta) &= \sum_{i=1}^N \beta_i s(\theta)^* \hat{e}_i \hat{e}_i^* s(\theta) \end{aligned}$$

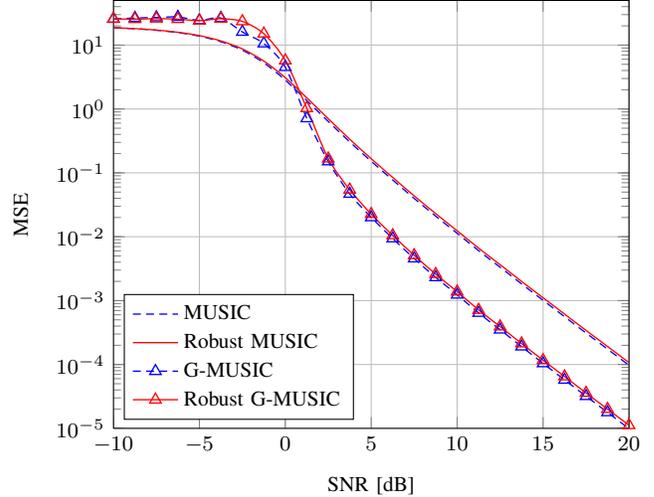
and

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

with  $\hat{\mu}_1 \leq \dots \leq \hat{\mu}_N$  the eigenvalues of  $\text{diag}(\hat{\lambda}) - \frac{1}{n} \sqrt{\hat{\lambda}} \sqrt{\hat{\lambda}}^T$ ,  $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_N)^T$ .

*Proof:* The proof is available in [9].  $\blacksquare$

The function  $\gamma(\theta)$  is the defining metric for the MUSIC algorithm [1], the zeros of which contain the  $\theta_i$ ,  $i \in \{1, \dots, K\}$ . Theorem 2 proves that the  $N, n$ -consistent G-MUSIC estimator of  $\gamma(\theta)$  proposed by Mestre in [13] can be extended into a *robust G-MUSIC* method. The latter consists in replacing the sample covariance matrix  $\hat{S}_N$  as in [13] by the robust estimator  $\hat{C}_N$ . The angles  $\theta_i$  are then estimated as the deepest minima of  $\hat{\gamma}(\theta)$ . This new technique is expected to perform better than either MUSIC or G-MUSIC in the finite  $(N, n)$  regime in the case of non-Gaussian noise, for an appropriate choice of the function  $u$ . Proving so



**Fig. 1.** MSE performance of the various MUSIC estimators for  $K = 1$ , Gaussian noise,  $N = 10$ , and  $n = 50$ .

requires the study of the second order statistics of  $\gamma(\theta)$ , which is left to future work.

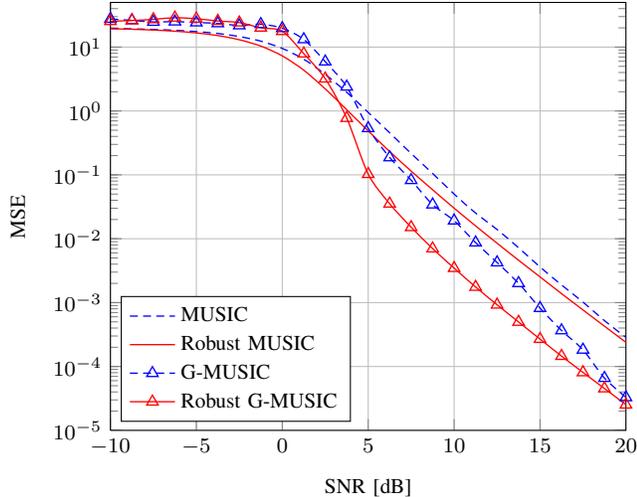
In the following, we provide comparative performance results between the classical MUSIC, the robust MUSIC, the G-MUSIC, and the robust G-MUSIC algorithms. We recall that the MUSIC algorithm consists in determining the deepest local minima of

$$\hat{\gamma}^\infty(\theta) = \sum_{i=1}^{N-K} s(\theta)^* \hat{e}_i^S \hat{e}_i^{S*} s(\theta)$$

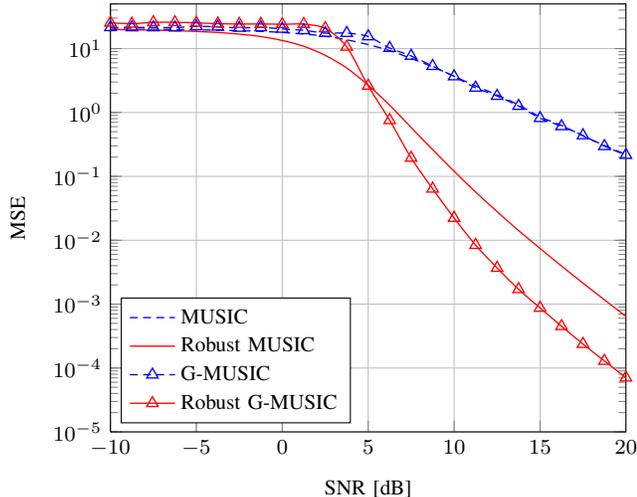
where  $\hat{e}_i^S$  is the eigenvector associated with the  $i$ -th smallest eigenvalue of  $\hat{S}_N$ . Robust MUSIC is equivalent to MUSIC but uses  $\hat{e}_i$  instead of  $\hat{e}_i^S$  in the expression of  $\hat{\gamma}^\infty(\theta)$ . G-MUSIC determines the local minima of  $\hat{\gamma}(\theta)$  but with  $\hat{e}_i^S$  instead of  $\hat{e}_i$ . Finally, robust G-MUSIC seeks the minima of  $\hat{\gamma}(\theta)$ , as described in Theorem 2.

We take  $z_{k,t}$  standard Gaussian, independent across  $k$  and  $t$ , and  $w_t$  a vector with independent zero-mean unit variance entries with either Gaussian or Student-t distribution with  $\nu > 2$  degrees of freedom. The case  $w_t$  Gaussian is used as a reference scenario. The choice of  $w_t$  with Student-t entries and  $\nu$  large is used to model the more realistic scenario of a sensor array with close-to-Gaussian noise. For small  $\nu$  (resulting into a noise distribution with heavier tails), the scenario can be either used to reflect independent antenna reading errors in a sensor array or to model a distributed sensor network in which each sensor faces independent impulsive noise (e.g. in a MIMO-STAP setting [14], [15]). We choose  $u(s) = (1 + \nu') / (\nu' + s)$ , for some  $\nu' > 0$  which controls the degree of robustness of the estimator ( $\nu' \rightarrow \infty$  brings  $u(s) = 1$ , hence reduced robustness). We set here  $\nu' = 0.5$  in all simulations. We model the steering vectors by  $[s(\theta)]_k = \exp(i\pi k \sin(\theta))$  as in a uniform linear array of  $N$  elements with half wavelength inter-element spacing. We take  $N = 10$ ,  $n = 50$ , and  $p_k = 1$  for all  $k$ . Under these conditions,  $\hat{C}_N$  satisfies [4, Assumption (E)], for  $\nu \geq 2.5$ , implying that  $\hat{C}_N$  is well defined for each  $x_1, \dots, x_n$  and not only for all large  $n$  as.

We first consider  $K = 1$  with  $\theta_1 = 18^\circ$ . Figure 1, Figure 2, and Figure 3 depict the mean-square error (MSE) performance

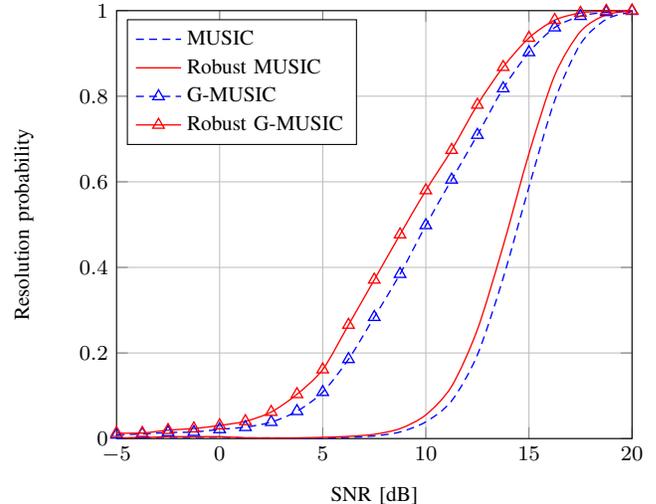


**Fig. 2.** MSE performance of the various MUSIC estimators for  $K = 1$ , Student-t noise with  $\nu = 5$ ,  $N = 10$ , and  $n = 50$ .



**Fig. 3.** MSE performance of the various MUSIC estimators for  $K = 1$ , Student-t noise with  $\nu = 2.5$ ,  $N = 10$ , and  $n = 50$ .

$E[|\gamma(\theta_1) - \hat{\gamma}(\theta_1)|^2]$  of the above estimators, as a function of the signal-to-noise ratio (SNR)  $\sigma^{-2}$ . In Figure 1, we take  $w_t$  Gaussian. In Figure 2,  $w_t$  has Student-t entries with  $\nu = 5$  degrees of freedom (close-to-Gaussian scenario). Finally, in Figure 3,  $w_t$  has Student-t entries with  $\nu = 2.5$  degrees of freedom (impulsive noise scenario). We naturally expect the robust techniques to bring larger performance gains in the latter scenario than in the close-to-Gaussian ones. The simulations are based on 50 000 Monte Carlo simulations per SNR value. We first observe that both robust methods perform almost identically to their non-robust counterparts in a Gaussian noise setting. In the close-to-Gaussian noise setting, the robust approaches then overcome the non-robust ones, especially in the low-to-medium SNR region where we see a significant performance advantage for the robust G-MUSIC method against G-MUSIC, while MUSIC and robust MUSIC perform similarly.



**Fig. 4.** Resolution performance of the various MUSIC estimators,  $\theta_1 = 10^\circ$ ,  $\theta_2 = 15^\circ$ , Student-t noise with  $\nu = 5$ ,  $N = 10$ , and  $n = 50$ .

In the far-from-Gaussian noise scenario, we then see both robust methods show a large gain compared to the non-robust ones. In this regime, the random matrix advantage of G-MUSIC versus MUSIC disappears completely, while being largely favorable to the robust scheme. The latter two results translate the fact that, if the noise non-Gaussianity and the small sample size are not both appropriately controlled, one of the two will overtake the other, making G-MUSIC or robust MUSIC inefficient. On the contrary, robust G-MUSIC, which controls both problems, always brings a significant performance advantage.

In Figure 4, we depict the performance of resolution of two close sources of the MUSIC estimators. For this, we take  $K = 2$ ,  $\theta_1 = 10^\circ$ ,  $\theta_2 = 15^\circ$ , and  $\nu = 5$ . The curves show the probability of detecting exactly two local minima of  $\hat{\gamma}$  (or  $\hat{\gamma}^\infty$ ) within  $[5^\circ, 20^\circ]$ , based on 50 000 Monte Carlo simulations for each SNR value. Note that again, in this close-to-Gaussian noise setting, the robust G-MUSIC algorithm has a much stronger resolution power than the G-MUSIC algorithm.

The robust G-MUSIC example is an illustrative application of Theorem 1 demonstrating the strong advantage brought by a joint robust and random matrix-based signal processing framework. The theoretical performance gains are however not easy to obtain as they would require the elaboration of central limit theorems (CLT), currently under study.

#### IV. CONCLUSION

We have proved that a certain family of robust M-estimates of population covariance matrices is consistent with the sample covariance matrix, in the regime of both large population  $N$  and sample  $n$  sizes. We applied this result to prove that a robust version of the G-MUSIC estimator of Mestre is still an  $(N, n)$ -consistent estimator of the direction of arrival in array processing. The simulation results then suggested that the induced robust G-estimator performs better than the MUSIC and G-MUSIC estimators under non-Gaussian noise and for  $N$  not small compared to  $n$ .

## V. SKETCH OF PROOF OF THEOREM 1

The almost sure existence of a solution to (2) for all large  $N, n$  unfolds from a modified application of Yates' standard interference function technique [16]. We do not detail this part of the proof here. Take now such a solution  $\hat{C}_N$  and denote  $d_i = \frac{1}{N} x_i^* \hat{C}_N^{-1} x_i$ , which we order as  $d_1 \leq \dots \leq d_n$  without loss of generality. Denote also  $D = \text{diag}(\{u(d_i)\}_{i=1}^n)$ . By definition

$$d_i = \frac{1}{N} x_i^* \left( \frac{1}{n} XDX^* \right)^{-1} x_i.$$

From the non increasing property of  $u$ , we have the inequality  $XDX^* \succeq u(d_n)XX^*$  (in the partial order of Hermitian matrices), which implies after inversion ([17, Corollary 7.7.4])

$$\frac{1}{u(d_n)} (XX^*)^{-1} \succeq (XDX^*)^{-1}$$

and therefore, recalling that  $n^{-1}XX^* = \hat{S}_N$ ,

$$d_n \leq \frac{1}{u(d_n)} \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n$$

or equivalently, since  $u(d_n) > 0$ ,

$$\phi(d_n) \leq \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n.$$

Similarly, we show

$$\phi(d_1) \geq \frac{1}{N} x_1^* \hat{S}_N^{-1} x_1.$$

Since  $\phi$  is non-decreasing, we also have  $\phi(d_1) \leq \phi(d_i) \leq \phi(d_n)$  for  $i \leq n$ , and we therefore obtain

$$\frac{1}{N} x_1^* \hat{S}_N^{-1} x_1 \leq \phi(d_i) \leq \frac{1}{N} x_n^* \hat{S}_N^{-1} x_n.$$

The remainder of the proof then consists in proving that

$$\max_{1 \leq i \leq n} \left| \frac{1}{N} x_i^* \hat{S}_N^{-1} x_i - 1 \right| \xrightarrow{\text{a.s.}} 0.$$

The main difficulty to prove this result lies in the joint control of the smallest eigenvalue of the matrices  $\{\hat{S}_N - x_i x_i^*\}_{i \in \{1, \dots, n\}}$  which is based on random matrix arguments. Details are provided in [9]. From this result and the fact that  $\phi$  is strictly increasing (then invertible with respect to composition) on a neighborhood of 1, we obtain

$$\max_{1 \leq i \leq n} |d_i - \phi^{-1}(1)| \xrightarrow{\text{a.s.}} 0$$

or equivalently

$$\max_{1 \leq i \leq n} \left| u(d_i) - \frac{1}{\phi^{-1}(1)} \right| \xrightarrow{\text{a.s.}} 0.$$

We conclude with the matrix inequalities

$$\begin{aligned} & \min_{i \leq n} \left\{ u(d_i) - \frac{1}{\phi^{-1}(1)} \right\} \frac{1}{n} XX^* \\ & \leq \frac{1}{n} \sum_{i=1}^n \left( u(d_i) - \frac{1}{\phi^{-1}(1)} \right) x_i x_i^* \\ & \leq \max_{i \leq n} \left\{ u(d_i) - \frac{1}{\phi^{-1}(1)} \right\} \frac{1}{n} XX^* \end{aligned}$$

and the fact that  $\|\frac{1}{n}XX^*\| < K$  for some  $K > 0$  and for all  $n$  almost surely, from which

$$\left\| \frac{1}{n} \sum_{i=1}^n \left( u(d_i) - \frac{1}{\phi^{-1}(1)} \right) x_i x_i^* \right\| = \left\| \hat{C}_N - \frac{\hat{S}_N}{\phi^{-1}(1)} \right\| \xrightarrow{\text{a.s.}} 0.$$

## VI. REFERENCES

- [1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [2] L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time-Series Analysis*. Boston, MA, USA: Addison-Wesley, 1991.
- [3] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
- [4] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The Annals of Statistics*, pp. 51–67, 1976.
- [5] D. Kelker, "Distribution theory of spherical distributions and a location-scale parameter generalization," *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 32, no. 4, pp. 419–430, 1970.
- [6] S. Watts, "Radar Detection Prediction in Sea Clutter Using the Compound K-Distribution model," *IEE Proceedings, Part F*, vol. 132, no. 7, pp. 613–620, December 1985.
- [7] J. W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 55, no. 2, pp. 331–339, 1995.
- [8] X. Mestre and M. Lagunas, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 598–614, Feb. 2008.
- [9] R. Couillet, F. Pascal, and J. W. Silverstein, "Robust M-Estimation for Array Processing: A Random Matrix Approach," *IEEE Transactions on Information Theory*, Submitted for publication. [Online]. Available: <http://arxiv.org/abs/1204.5320>
- [10] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *The Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.
- [11] R. Couillet, J. W. Silverstein, Z. D. Bai, and M. Debbah, "Eigen-inference for energy estimation of multiple sources," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2420–2439, 2011.
- [12] J. T. Kent and D. E. Tyler, "Redescending M-estimates of multivariate location and scatter," *The Annals of Statistics*, pp. 2102–2119, 1991.
- [13] X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.
- [14] A. M. Haimovich, R. S. Blum, and L. J. C. Jr, "MIMO Radar with Widely Separated Antennas," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 116–129, 2008.
- [15] J. Li and P. Stoica, *MIMO Radar Signal Processing*, 1st ed. Wiley, 2009.
- [16] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, 1995.
- [17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.