Robust M-Estimation for Array Processing: A Random Matrix Approach

Romain Couillet[1], Frédéric Pascal[2], and Jack W. Silverstein[3]
[1] *Telecommunication department, Supélec, Gif sur Yvette, France.*
[2] *SONDRA Laboratory, Supélec, Gif sur Yvette, France.*
[3] *Department of Mathematics, North Carolina State University, NC, USA.*

***Abstract*—This article studies the limiting behavior of a robust M-estimator of population covariance matrices as both the number of available samples and the population size are large. Using tools from random matrix theory, we prove that the difference between the sample covariance matrix and (a scaled version of) the robust M-estimator tends to zero in spectral norm, almost surely. This result is applied to prove that recent subspace methods arising from random matrix theory can be made robust without altering their first order behavior.**

I. Introduction

Many multi-variate signal processing detection and estimation techniques are based on the empirical covariance matrix of a sequence of samples $x_1, \ldots, x_n$ from a random population vector $x \in \mathbb{C}^N$. Assuming $\mathrm{E}[x] = 0$ and $\mathrm{E}[xx^*] = C_N$, the strong law of large numbers ensures that, for independent and identically distributed (i.i.d.) samples,

$$\hat{S}_N = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^* \to C_N$$

almost surely (a.s.), as the number n of samples increases. Many subspace methods, such as the 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 the entries of x. Although the former issue was probably the first historically recognized, it is only recently that significant advances have been made using random matrix theory [3]. As for the latter, it has spurred a strong wave of interest in the seventies, starting with the works from Huber [4] on robust M-estimation. The objective of this article is to provide a first bridge between the two disciplines by introducing new fundamental results on robust M-estimates in the random matrix regime where both N and n grow large at the same rate.

Aside from its obvious simplicity of analysis, 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 estimation theory aims at tackling this problem [5]. Among other solutions, the so-called robust M-estimators of the population covariance matrix, originally introduced by Huber [4] and investigated in the seminal work of Maronna [6], have imposed themselves as an appealing alternative to the SCM. This estimator, which we denote $\hat{C}_N$, is defined implicitly as a solution of[1]

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

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 [7]. 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 [8].

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 of 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 $\hat{S}_N$ converges a.s. to a limiting distribution described implicitly by its Stieltjes transform [9]. When C_N is the identity matrix for all N, this distribution takes an explicit form known as the Marčenko-Pastur law [10]. Under some additional moment conditions on the entries of y, it has also been shown that the eigenvalues of $\hat{S}_N$ cannot lie infinitely often away from the support of the limiting distribution [11]. In the past years, these two results and subsequent works have been applied to revisit classical signal processing techniques such as signal detection schemes [12] or subspace methods [13], [14]. In these works, traditional *n-consistent* detection and estimation methods were improved into (N, n)*-consistent* approaches, i.e. they provide estimates that are consistent in the large N, n

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

[1] Our expression differs from the standard convention where $x_i^* \hat{C}_N^{-1} x_i$ is traditionally not scaled by $1/N$. The current form is however more convenient for analysis in the large N, n regime.

regime rather than in the fixed N and large n regime. These improved estimators are often referred to as G-estimators.

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. Although the study of the SCM $\hat{S}_N$ for elliptically distributed vectors x in this regime was recently done in [15], the equivalent analysis for $\hat{C}_N$ is much more challenging. Nonetheless, under the assumption that x 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. One important technical challenge brought by $\hat{C}_N$, usually not met in random matrix theory, lies in the dependence structure between the columns of $\hat{C}_N$ (as opposed to $\hat{S}_N$). We fundamentally rely on the set of assumptions on the function u taken by Maronna in [6] to overcome this difficulty. Our main contribution consists in showing that, in the large N, n regime, and under some mild assumptions, $\|\hat{C}_N - \alpha\hat{S}_N\| \to 0$, a.s., for some $\alpha > 0$ to be defined. Note that this result is in line with the conjecture made in [16] according to which $\|\hat{C}_N - \alpha\hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ for the function $u(s) = 1/s$ studied extensively by Tyler [17], [18]; however, this function does not enter our present scheme and creates some additional difficulties which leave the conjecture open. 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 [19].

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. All technical proofs are detailed in the appendices.

Notations: The arrow '$\xrightarrow{\text{a.s.}}$' denotes almost sure convergence. For $A \in \mathbb{C}^{N\times N}$ Hermitian, $\lambda_1(A) \leq \ldots \leq \lambda_N(A)$ are its ordered eigenvalues. The norm $\|\cdot\|$ is the spectral norm for matrices and the Euclidean norm for vectors. For A, B Hermitian, $A \succeq B$ means that $A - B$ is nonnegative definite. The notation A^* denotes the Hermitian transpose of A. We also write $\imath = \sqrt{-1}$.

II. Main results

Let $X = [x_1, \ldots, x_n] \in \mathbb{C}^{N\times n}$, where $x_i = A_N y_i \in \mathbb{C}^N$, with $y_i = [y_{i1}, \ldots, y_{iM}]^\mathsf{T} \in \mathbb{C}^M$ 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 $x_1, \ldots, x_n$ by

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

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

(i) u is nonnegative, nonincreasing, and continuous on $\mathbb{R}^+$;

(ii) the function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$, $s \mapsto su(s)$ is nondecreasing and bounded, with $\sup_x \phi(x) = \phi_\infty > 1$. Moreover, ϕ is increasing on $[0, \phi_\infty)$.

Classical M-estimators within this framework include the Huber estimator, defined by $\phi(s) = s$ 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 $x_1, \ldots, x_n$ by a factor 1 and reduces the impact of the outliers.

To pursue, we need the following statistical assumptions on the large dimensional random matrices under study.

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

A2. $M \geq N$ and, as $n \to \infty$,

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

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 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 $\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i$, $i = 1, \ldots, n$. Further generalizations, e.g. to elliptical distributions, would break this effect and would certainly entail a much different asymptotic behavior of $\hat{C}_N$. These important considerations are 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 [11], [14]. Note also that **A2** demands $\liminf_n c_N > 0$, so that the following results *do not* contain the results from [6], [18], in which N is fixed and $n \to \infty$, as special cases. With these assumptions, we are now in position to provide the main technical result of this article.

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

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

Then, we have the following results.

(I) There exists a unique solution to (1) for all large N a.s. We denote $\hat{C}_N$ this solution, defined as

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

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 (1) 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: The proof is provided in Appendix A. ■

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,

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

Proof: The proof is provided in Appendix A. ■

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

Theorem 1–(I) extends first the result from Maronna [6, Theorem 1] which states that a solution to (1) exists for each set $\{x_1, \ldots, x_n\}$ 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 [6]). Theorem 1–(I) also extends the results on uniqueness [6], [18] which hold for all N, n under some further conditions on $u(s)$, such as $\phi(s)$ is strictly increasing in [6]. These assumptions are particularly demanding as they may reject some M-estimators such as the Huber M-estimator for which $\phi(s)$ is constant for large s. Theorem 1–(I) trades these assumptions against a requirement for N and n to be "sufficiently large" and for $\{x_1, \ldots, x_n\}$ to belong to a probability one sequence. Precisely, we demand that there exists an integer n_0 depending on the random sequence $\{(x_1, \ldots, x_n)\}_{n=1}^{\infty}$, such that for all $n \geq n_0$, existence and uniqueness are established under no further condition than the definition (i)–(ii) of $u(s)$ and **A1**–**A3**.

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 $\hat{C}_N$ (either always defined under the assumptions of [6], [18] or defined a.s. for large enough 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 $n \to \infty$ 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 [6], [18] 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 [19].

III. Application: Robust G-MUSIC

Consider K signal sources impinging on a collection of N collocated sensors with angles of arrival $\theta_1, \ldots, \theta_K$. 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 θ, $z_{k,t} \in \mathbb{C}$ 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

$$x_t = A_N y_t, \quad A_N \triangleq \left[S(\Theta) P^{\frac{1}{2}} \quad \sigma I_N \right]$$

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

Taking n independent observations $x_1, \ldots, x_n$ 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 σ^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 $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_N$), with $\hat{C}_N$ defined as in Theorem 1 (with $\hat{C}_N = I_N$ when (1) does not have a unique solution). Then, as $N, n \to \infty$ in the regime of Assumption **A2**, and K fixed,

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

where

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

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 $\hat{\mu}_1 \leq \ldots \leq \hat{\mu}_N$ the eigenvalues of $\operatorname{diag}(\hat{\boldsymbol{\lambda}}) - \frac{1}{n}\sqrt{\hat{\boldsymbol{\lambda}}}\sqrt{\hat{\boldsymbol{\lambda}}}^{\mathsf{T}}$, $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)^{\mathsf{T}}$.

Proof: The proof can be found in Appendix E. ■

The function $\gamma(\theta)$ is the defining metric for the MUSIC algorithm [1], the zeros of which contain the θ_i, $i \in \{1, \ldots, K\}$. 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 θ_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 requires the study of the second order statistics of $\gamma(\theta)$, which is left to future work. Note also that our result does *not* prove the N, n-consistency in the estimates of $\theta_1, \ldots, \theta_K$, which would demand to show

$$\sup_{\theta \in [-\pi,\pi]} |\gamma(\theta) - \hat{\gamma}(\theta)| \xrightarrow{\text{a.s.}} 0.$$

Proving this convergence requires more advanced techniques; see [20, Section 4.3.2] for a discussion on this topic.

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 the function

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

where $\hat{e}_i^S$ is the eigenvector associated with the i-th smallest eigenvalue of $\hat{S}_N$ (the notation ∞ recalls the fact that $n \to \infty$ for N fixed in this setting). 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 ν large is used to model the more realistic scenario of a sensor array with close-to-Gaussian noise. For small ν (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 [21], [22]). We choose $u(s) = (1 + \nu')/(\nu' + s)$, for some $\nu' > 0$ which controls the degree of robustness of the estimator ($\nu' \to \infty$ brings $u(s) = 1$, hence reduced robustness). We set here $\nu' = 0.5$ in all simulations. We model the steering vectors by $[s(\theta)]_k = \exp(\imath \pi k \sin(\theta))$ 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 [6, Assumption (E)], for $\nu \geq 2.5$, implying that $\hat{C}_N$ is well defined for each $x_1, \ldots, x_n$ and not only for all large n a.s.

We first consider $K = 1$ with $\theta_1 = 18°$. Figure 1, Figure 2, and Figure 3 depict the mean-square error (MSE) performance $\mathrm{E}[|\gamma(\theta_1) - \hat{\gamma}(\theta_1)|^2]$ of the above estimators, as a function of the signal-to-noise ratio (SNR) σ^{-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

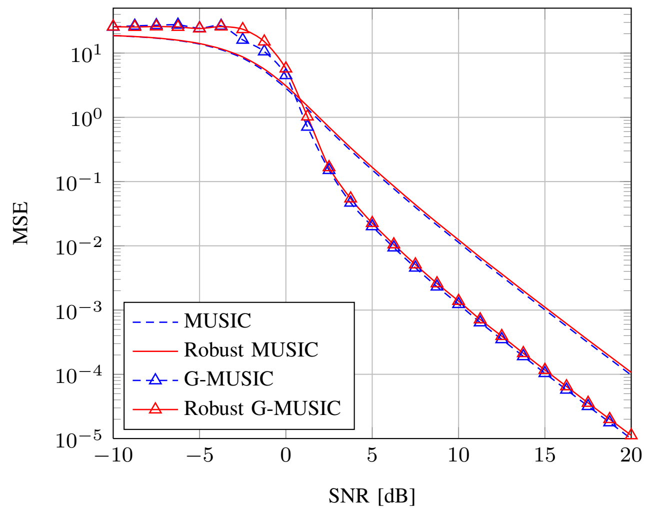

Fig. 1. MSE performance of the various MUSIC estimators for $K = 1$, Gaussian noise, $N = 10$, and $n = 50$.

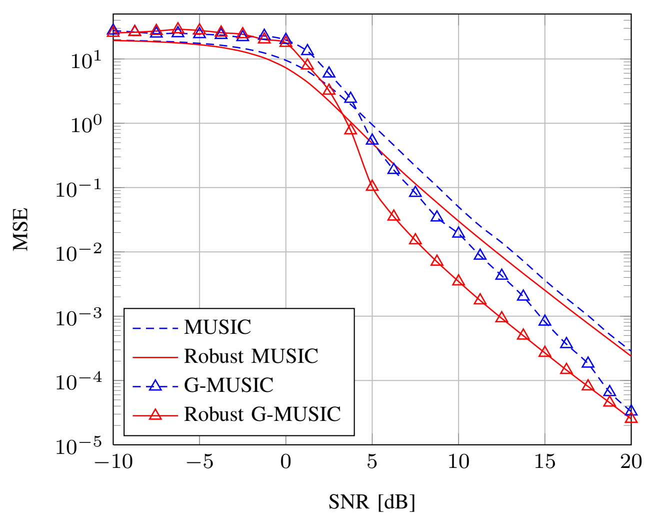

Fig. 2. MSE performance of the various MUSIC estimators for $K = 1$, Student-t noise with $\nu = 5$, $N = 10$, and $n = 50$.

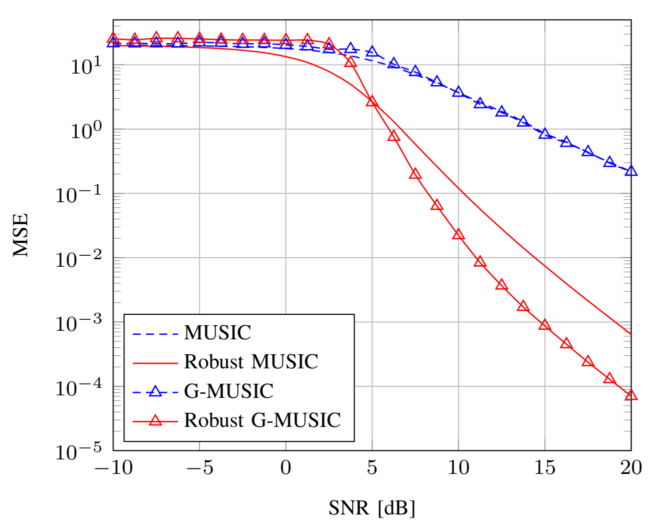

Fig. 3. MSE performance of the various MUSIC estimators for $K = 1$, Student-t noise with $\nu = 2.5$, $N = 10$, and $n = 50$.

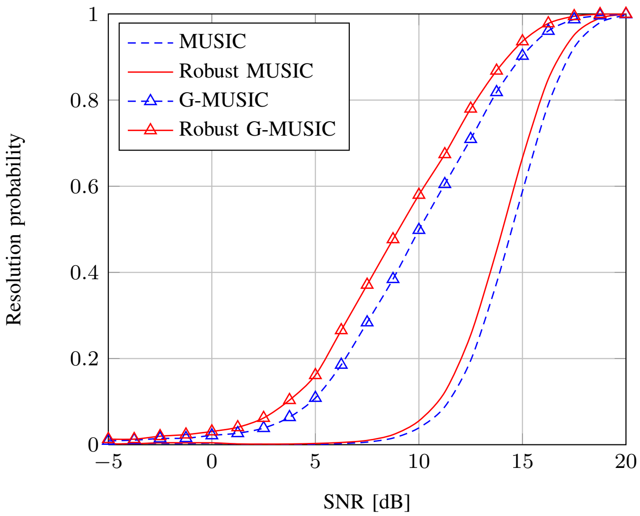

Fig. 4. Resolution performance of the various MUSIC estimators, $\theta_1 = 10°$, $\theta_2 = 15°$, Student-t noise with $\nu = 5$, $N = 10$, and $n = 50$.

perform slightly worse than 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. 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 $K = 2$, $\theta_1 = 10°$, $\theta_2 = 15°$, and $\nu = 5$. The curves show the probability of detecting exactly two local minima of $\hat{\gamma}$ (or $\hat{\gamma}^{\infty}$) within $[5°, 20°]$, based on 50 000 Monte Carlo simulations for each SNR value. Note that, in this close-to-Gaussian noise setting, the robust G-MUSIC algorithm has a much stronger resolution power than the G-MUSIC, although both operate at close MSE for single source detection (from Figure 2).

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). In the robust G-MUSIC example, this demands a CLT for the quantity $n(\hat{\theta} - \theta)$, which requires more advanced tools than these presented in this article.

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.

Appendix A
Proof of Theorem 1 and Corollary 1

Proof of Theorem 1: In order to prove the existence and uniqueness of a solution to (1) for all large n, we use the framework of standard interference functions from [23].

Definition 1: A function $h = (h_1, \dots, h_n) : \mathbb{R}_+^n \to \mathbb{R}_+^n$ is said to be a standard interference function if it fulfills the following conditions:

1) *Positivity:* if $q_1, \dots, q_n \geq 0$, then $h_j(q_1, \dots, q_n) > 0$, for all j.
2) *Monotonicity:* if $q_1 \geq q_1', \dots, q_n \geq q_n'$, then for all j, $h_j(q_1, \dots, q_n) \geq h_j(q_1', \dots, q_n')$.
3) *Scalability:* for all $\alpha > 1$ and for all j, $\alpha h_j(q_1, \dots, q_n) \geq h_j(\alpha q_1, \dots, \alpha q_n)$.

Theorem 3: If an n-variate function $h(q_1, \dots, q_n)$ is a standard interference function and there exists $(q_1, \dots, q_n)$ such that for all j, $q_j \geq h_j(q_1, \dots, q_n)$, then the system of equations

$$q_j = h_j(q_1, \dots, q_n) \tag{2}$$

for $j = 1, \dots, n$, has at least one solution, given by $\lim_{t\to\infty}(q_1^{(t)}, \dots, q_n^{(t)})$, where

$$q_j^{(t+1)} = h_j(q_1^{(t)}, \dots, q_n^{(t)})$$

for $t \geq 1$ and any initial values $q_1^{(0)}, \dots, q_n^{(0)} \geq 0$.

Proof: The proof is provided in Appendix D. ∎

Remark 1: Note that our definition of a standard interference function differs from that of [23] in which the scalability requirement reads: for all j, $\alpha h_j(q_1, \dots, q_n) > h_j(\alpha q_1, \dots, \alpha q_n)$. Changing the strict inequality to a loose one alters the consequences for the theorem above, where only existence is ensured. However, for our present purposes with $\phi(s)$ possibly possessing a flat region, requesting a strict inequality would be too demanding.

Since $\{x_1, \dots, x_n\}$ spans $\mathbb{C}^N$ for all large n a.s. (as a consequence of Proposition 2 in Appendix F), we can define for these n the functions h_j, $j = 1, \dots, n$,

$$h_j(q_1, \dots, q_n) \triangleq \frac{1}{N} x_j^* \left(\frac{1}{n} \sum_{i=1}^{n} u(q_i) x_i x_i^* \right)^{-1} x_j. \tag{3}$$

We first show that $h = (h_1, \dots, h_n)$ meets the conditions of Theorem 3 for all large n a.s. Due to **A1**, from standard arguments using the Markov inequality and the Borel Cantelli lemma, we have that $\min_{i\leq n}\|x_i\|\neq 0$ for all large n a.s. (this is also a corollary of Lemma 2 below). Therefore, we clearly have $h_j>0$ for all j, for all large n a.s. Also, since u is non-increasing, taking $q_1,\ldots,q_n$ and $q_1',\ldots,q_n'$ such that $q_i'\geq q_i\geq 0$ for all i, $u(q_i')\leq u(q_i)$ and then

$$\frac{1}{n}\sum_{i=1}^n u(q_i)x_ix_i^*\succeq \frac{1}{n}\sum_{i=1}^n u(q_i')x_ix_i^*$$

From [24, Corollary 7.7.4], this implies

$$\left(\frac{1}{n}\sum_{i=1}^n u(q_i')x_ix_i^*\right)^{-1}\succeq\left(\frac{1}{n}\sum_{i=1}^n u(q_i)x_ix_i^*\right)^{-1}$$

from which $h_j(q_1',\ldots,q_n')\geq h_j(q_1,\ldots,q_n)$, proving the monotonicity of h.

For $\alpha>1$, $\phi(\alpha q_i)\geq\phi(q_i)$, so that $u(\alpha q_i)\geq\frac{u(q_i)}{\alpha}$. Therefore

$$\frac{1}{n}\sum_{i=1}^n u(\alpha q_i)x_ix_i^*\succeq\frac{1}{\alpha}\frac{1}{n}\sum_{i=1}^n u(q_i)x_ix_i^*$$

From [24, Corollary 7.7.4] again, we then have

$$\alpha\left(\frac{1}{n}\sum_{i=1}^n u(q_i)x_ix_i^*\right)^{-1}\succeq\left(\frac{1}{n}\sum_{i=1}^n u(\alpha q_i)x_ix_i^*\right)^{-1}$$

so that $\alpha h_j(q_1,\ldots,q_n)\geq h_j(\alpha q_1,\ldots,\alpha q_n)$. Therefore h is a standard interference function. In order to prove that (3) admits a solution, from Theorem 3, we now need to prove that there exists $(q_1,\ldots,q_n)$ such that for all j, $q_j\geq h_j(q_1,\ldots,q_n)$. Note that this may not hold for all fixed N,n as discussed in [6, pp. 54]. We will prove instead that a solution exists for all large n a.s.

To pursue, we need random matrix results and additional notations. Take c_-,c_+ such that $0<c_-<\liminf_N c_N$ and $\limsup_N c_N<c_+<1$, and denote $X_{(i)}=[x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n]\in\mathbb{C}^{N\times(n-1)}$. We start with the following fundamental lemmas, which allow for a control of the joint convergence of the quadratic forms $\frac{1}{N}x_i^*\hat{S}_N^{-1}x_i-1$.

Lemma 1: Assume **A1–A3**. There exists $\varepsilon>0$ such that

$$\min_{i\leq n}\left\{\lambda_1\left(\frac{1}{n}X_{(i)}X_{(i)}^*\right)\right\}>\varepsilon$$

for all large n a.s.

Proof: The proof is provided in Appendix B. ■

Lemma 2: Assume **A1–A3**. Then, a.s.,

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

Proof: The proof is provided in Appendix C. ■

Let $q_1=\ldots=q_n\triangleq q>0$. Then,

$$h_i(q_1,\ldots,q_n)=\frac{1}{u(q)}\frac{1}{N}x_i^*\hat{S}_N^{-1}x_i=\frac{q}{\phi(q)}\frac{1}{N}x_i^*\hat{S}_N^{-1}x_i.$$

Take $\varepsilon>0$ such that $(1+\varepsilon)/(\phi_\infty-\varepsilon)<1$. This is always possible since $\phi_\infty>1$. Choose now q such that $\phi(q)=\phi_\infty-\varepsilon$, which also exists since ϕ is increasing on $[0,\phi^{-1}(\phi_\infty-))$ with image $[0,\phi_\infty)$. From Lemma 2, for all large n a.s.,

$$\sup_i\left|\frac{1}{q}h_i(q_1,\ldots,q_n)(\phi_\infty-\varepsilon)-1\right|<\varepsilon.$$

Therefore,

$$\frac{1}{q}h_i(q_1,\ldots,q_n)<\frac{1+\varepsilon}{\phi_\infty-\varepsilon}<1$$

from which $h_i(q,\ldots,q)<q$ for all i. From Theorem 3, we therefore prove the existence of a solution to (2) with h_j given in (3). Since these quadratic forms define the solutions of the fixed-point equation (1), this proves the existence of a solution $\hat{C}_N$ for all large n a.s. Note that Lemma 2 is crucial here and that, for ϕ_∞ close to one, there is little hope to prove existence for all fixed N,n, consistently with the results [6], [18].

We now prove uniqueness. Take a solution $\hat{C}_N$ and denote $d_i=\frac{1}{N}x_i^*\hat{C}_N^{-1}x_i$, which we order as $d_1\leq\ldots\leq d_n$ without loss of generality. Denote also $D=\operatorname{diag}(\{u(d_i)\}_{i=1}^n)$. By definition

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

From the nonincreasing property of u, we have the inequality

$$XDX^*\succeq u(d_n)XX^*$$

which implies after inversion

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

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,

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

from which we also have

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

Since ϕ 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

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

Take $0<\varepsilon<\min\{1,(\phi_\infty-1)\}$. From Lemma 2, for all large n a.s.,

$$0<1-\varepsilon<\phi(d_i)<1+\varepsilon<\phi_\infty.$$

Since ϕ is continuous and increasing on $(0,\phi^{-1}(\phi_\infty-))$ with image contained in $(0,\phi_\infty)$, ϕ is invertible there and we obtain that for all large n a.s.,

$$\phi^{-1}\left(1-\varepsilon\right)<d_i<\phi^{-1}\left(1+\varepsilon\right).\qquad(4)$$

We can now prove the almost sure uniqueness of $\hat{C}_N$ for all large n. Take ε in (4) to satisfy the previous conditions and to be such that $(\phi^{-1}(1+\varepsilon))^2/\phi^{-1}(1-\varepsilon) < \phi^{-1}(\phi_\infty-)$, which is always possible as the left-hand side expression is continuous in ε with limit $\phi^{-1}(1) < \phi^{-1}(\phi_\infty-)$ as $\varepsilon \to 0$.

We now follow the arguments of [23, Theorem 1]. Assume $(d_1^{(1)}, \ldots, d_n^{(1)})$ and $(d_1^{(2)}, \ldots, d_n^{(2)})$ are two distinct solutions of the fixed-point equation $d_j = h_j(d_1, \ldots, d_n)$ for $j = 1, \ldots, n$, where h_j is defined by (3). Then (up to a change in the indices 1 and 2), there exists k such that, for some $\alpha > 1$, $\alpha d_k^{(1)} = d_k^{(2)}$ and $\alpha d_i^{(1)} \geq d_i^{(2)}$ for $i \neq k$. From (4), for sufficiently large n a.s. the ratio $\alpha = d_k^{(1)}/d_k^{(2)}$ is also constrained to satisfy $\alpha < \phi^{-1}(1+\varepsilon)/\phi^{-1}(1-\varepsilon)$. Using this inequality and the upper bound in (4), we have for all j

$$0 < \alpha d_j^{(1)} < \frac{(\phi^{-1}(1+\varepsilon))^2}{\phi^{-1}(1-\varepsilon)} < \phi^{-1}(\phi_\infty-).$$

Since ϕ is increasing on $(0, \phi^{-1}(\phi_\infty-))$, we have in particular $\phi(\alpha d_j^{(1)}) > \phi(d_j^{(1)})$ from which $\alpha u(\alpha d_j^{(1)}) > u(d_j^{(1)})$, for all j and then, with similar arguments as previously, $\alpha h_j(d_1^{(1)}, \ldots, d_n^{(1)}) > h_j(\alpha d_1^{(1)}, \ldots, \alpha d_n^{(1)})$ for all j. Using the monotonicity of h, we conclude in particular

$$\begin{aligned} d_k^{(2)} = h_k(d_1^{(2)}, \ldots, d_n^{(2)}) &\leq h_k(\alpha d_1^{(1)}, \ldots, \alpha d_n^{(1)}) \\ &< \alpha h_k(d_1^{(1)}, \ldots, d_n^{(1)}) = \alpha d_k^{(1)} \end{aligned}$$

which contradicts $\alpha d_k^{(1)} = d_k^{(2)}$ and proves the uniqueness of $\hat{C}_N$ and Part (I) of Theorem 1.

We now prove Part (II) of the theorem. In order to proceed, we start again from (4). Since ε is arbitrary, we conclude that

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

Applying the continuous mapping theorem, we then have

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

Noticing that $\phi^{-1}(1) u(\phi^{-1}(1)) = \phi(\phi^{-1}(1)) = 1$, and therefore that $u(\phi^{-1}(1)) = 1/\phi^{-1}(1)$, this can be rewritten

$$\max_{i \leq n} \left| u(d_i) - \frac{1}{\phi^{-1}(1)} \right| \xrightarrow{\text{a.s.}} 0. \tag{5}$$

Now, we also have the matrix inequalities

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

From Proposition 2 in Appendix F, $\|\frac{1}{n}XX^*\| < K$ for some $K > 0$ and for all n a.s. From (5), we then conclude that

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

which completes the proof of Theorem 1. ∎

Proof of Corollary 1: The identity follows from [24, Theorem 4.3.7], according to which, for $1 \leq i \leq N$,

$$\begin{aligned} \lambda_i\left(\hat{S}_N\right) &\leq \lambda_i\left(\phi^{-1}(1)\hat{C}_N\right) + \lambda_N\left(\hat{S}_N - \phi^{-1}(1)\hat{C}_N\right) \\ \lambda_i\left(\hat{S}_N\right) &\geq \lambda_i\left(\phi^{-1}(1)\hat{C}_N\right) - \lambda_N\left(\hat{S}_N - \phi^{-1}(1)\hat{C}_N\right). \end{aligned}$$

The result follows by noticing that the second term in both right-hand sides tends to zero a.s. according to Theorem 1. ∎

Appendix B
Proof of Lemma 1

If the set of the eigenvalues of $\frac{1}{n}X_{(i)}X_{(i)}^*$ is contained within the set of the eigenvalues of $\frac{1}{n}XX^*$, then the result is immediate from Proposition 2 in Appendix F. We can therefore assume the existence of eigenvalues of $\frac{1}{n}X_{(i)}X_{(i)}^*$ which are not eigenvalues of $\frac{1}{n}XX^*$. By definition, the eigenvalues of $\frac{1}{n}X_{(i)}X_{(i)}^*$ solve the equation in λ

$$\det\left(\frac{1}{n}X_{(i)}X_{(i)}^* - \lambda I_N\right) = 0.$$

Take λ not to be also an eigenvalue of $\frac{1}{n}XX^*$. Then, developing the above expression, we get

$$\begin{aligned} &\det\left(\frac{1}{n}X_{(i)}X_{(i)}^* - \lambda I_N\right) \\ &= \det\left(\frac{1}{n}XX^* - \frac{1}{n}x_i x_i^* - \lambda I_N\right) \\ &= \det Q(\lambda) \det\left(I_N - Q(\lambda)^{-\frac{1}{2}} \frac{1}{n} x_i x_i^* Q(\lambda)^{-\frac{1}{2}}\right) \\ &= \det Q(\lambda) \left(1 - \frac{1}{n} x_i^* Q(\lambda)^{-1} x_i\right) \end{aligned}$$

with the notation $Q(\lambda) \triangleq \frac{1}{n}XX^* - \lambda I_N$, where we used $\det(I_N + AB) = \det(I_p + BA)$ in the last line, for $A \in \mathbb{C}^{N \times p}$ and $B \in \mathbb{C}^{p \times N}$, with $p = 1$ here.

Therefore, since λ cannot cancel the first determinant,

$$\frac{1}{n} x_i^* Q(\lambda)^{-1} x_i = \frac{1}{n} x_i^* \left(\frac{1}{n}XX^* - \lambda I_N\right)^{-1} x_i = 1.$$

Let us study the function

$$x \mapsto f_{n,i}(x) \triangleq \frac{1}{n} x_i^* \left(\frac{1}{n}XX^* - xI_N\right)^{-1} x_i.$$

First note, from a basic study of the asymptotes and limits of $f_{n,i}(x)$, that the eigenvalues of $\frac{1}{n}X_{(i)}X_{(i)}^*$ are interleaved with those of $\frac{1}{n}XX^*$ (a property known as Weyl's interlacing lemma) and in particular that

$$\lambda_1\left(\frac{1}{n}X_{(i)}X_{(i)}^*\right) \leq \lambda_1\left(\frac{1}{n}XX^*\right) \leq \lambda_2\left(\frac{1}{n}X_{(i)}X_{(i)}^*\right). \tag{6}$$

Since $\lambda_1(\frac{1}{n}XX^*)$ is a.s. away from zero for all large N (Proposition 2), only $\lambda_1(\frac{1}{n}X_{(i)}X_{(i)}^*)$ may remain in the neighborhood of zero for at least one $i \leq n$, for all large n.

We will show that this is impossible. Precisely, for all large n a.s., we will show that $f_{n,i}(x) < 1$ for any $i \leq n$ and for all x in some interval $[0, \xi)$, $\xi > 0$, confirming that no eigenvalue of $\frac{1}{n}X_{(i)}X_{(i)}^*$ can be found there. For this, we first use the fact that the $f_{n,i}(x)$ can be uniformly well estimated for all $x<0$ through Proposition 1 in Appendix F by a quantity strictly less than one. We then show that the growth of the $f_{n,i}(x)$ for x in a neighborhood of zero can be controlled, so to ensure that none of them reaches 1 for all $x<\xi$. This will conclude the proof.

We start with the study of $f_{n,i}(x)$ on $\mathbb{R}^-$. From Lemma 3,

$$f_{n,i}(x)=\frac{\frac{1}{n}x_i^*\left(\frac{1}{n}X_{(i)}X_{(i)}^*-xI_N\right)^{-1}x_i}{1+\frac{1}{n}x_i^*\left(\frac{1}{n}X_{(i)}X_{(i)}^*-xI_N\right)^{-1}x_i}.$$

Define

$$\bar{f}_n(x)\triangleq\frac{c_Ne_N(x)}{1+c_Ne_N(x)}$$

with $e_N(x)$ the unique positive solution of (see Proposition 1)

$$e_N(z)=\int\frac{t}{(1+c_Ne_N(z))^{-1}t-z}dF^{C_N}(t). \tag{7}$$

Then, with $Q(x)\triangleq\frac{1}{n}XX^*-xI_N$, $Q_i(x)\triangleq\frac{1}{n}X_{(i)}X_{(i)}^*-xI_N$,

$$\begin{aligned}\left|f_{n,i}(x)-\bar{f}_n(x)\right|&=\left|\frac{\frac{1}{n}x_i^*Q_i(x)^{-1}x_i}{1+\frac{1}{n}x_i^*Q_i(x)^{-1}x_i}-\frac{c_Ne_N(x)}{1+c_Ne_N(x)}\right|\\&\leq\left|\frac{1}{n}x_i^*Q_i(x)^{-1}x_i-c_Ne_N(x)\right|\\&\leq\left|\frac{1}{n}x_i^*Q_i(x)^{-1}x_i-\frac{1}{n}\operatorname{tr}C_NQ_i(x)^{-1}\right|\\&+\left|\frac{1}{n}\operatorname{tr}C_NQ_i(x)^{-1}-\frac{1}{n}\operatorname{tr}C_NQ(x)^{-1}\right|\\&+\left|\frac{1}{n}\operatorname{tr}C_NQ(x)^{-1}-c_Ne_N(x)\right|\end{aligned} \tag{8}$$

Using $(a+b+c)^p\leq 3^p(a^p+b^p+c^p)$ for $a,b,c>0$, and $p\geq 1$ (Hölder's inequality), and applying Lemma 5, Lemma 4, and Proposition 1 to the right-hand side terms of (8), respectively, with $p=4+\eta/2$, we obtain

$$\mathrm{E}\left[\left|f_{n,i}(x)-\bar{f}_n(x)\right|^{4+\frac{\eta}{2}}\right]\leq\frac{K}{n^{2+\frac{\eta}{4}}}$$

for some constant K independent of i, where we implicitly used **A1**. Therefore, using Boole's inequality on the above event for $i\leq n$, and the Markov inequality, for all $\zeta>0$,

$$\begin{aligned}&P\left(\max_{i\leq n}\left|f_{n,i}(x)-\bar{f}_n(x)\right|>\zeta\right)\\&\leq\sum_{i=1}^nP\left(\left|f_{n,i}(x)-\bar{f}_n(x)\right|>\zeta\right)<\frac{K}{\zeta^{4+\frac{\eta}{2}}n^{1+\frac{\eta}{4}}}.\end{aligned}$$

The Borel Cantelli lemma therefore ensures, for all $x<0$,

$$\max_{i\leq n}\left|f_{n,i}(x)-\bar{f}_n(x)\right|\xrightarrow{\text{a.s.}}0. \tag{9}$$

We now extend the study of $f_{n,i}(x)$ to x in a neighborhood of zero. From Proposition 2, $\lambda_1(\frac{1}{n}XX^*)>C_-(1-\sqrt{c_+})^2$ for all large n a.s. (recall that $\limsup_N c_N<c_+<1$) so that $f_{n,i}(x)$ is well-defined and continuously differentiable on $U=(-\varepsilon,\varepsilon)$ for $0<\varepsilon<C_-(1-\sqrt{c_+})^2$, for all large n a.s. Take $x\in U$. Since the smallest eigenvalue of $\frac{1}{n}XX^*-xI_N$ is lower bounded by $C_-(1-\sqrt{c_+})^2-\varepsilon$ for all large n, and that

$$\max_{i\leq n}\left|\frac{1}{n}\|x_i\|^2-\frac{1}{n}\operatorname{tr}C_N\right|\xrightarrow{\text{a.s.}}0$$

(using similar arguments based on the Boole and Markov inequality reasoning as above), we also have that for all large n a.s.

$$0<f'_{n,i}(x)<\frac{c_+C_+}{(C_-(1-\sqrt{c_+})^2-\varepsilon)^2}\triangleq K'$$

where we used $\limsup_N\frac{1}{n}\operatorname{tr}C_N<c_+C_+$.

From this result, along with the continuity of $f_{n,i}$, for $x\in U$ and for all large n a.s.,

$$f_{n,i}(x)<f_{n,i}(-x)+2xK'.$$

In particular, for $\xi=\min\{\varepsilon/2,(1-c_+)/(2K')\}$,

$$f_{n,i}(\xi)<f_{n,i}(-\xi)+(1-c_+). \tag{10}$$

Since $e_N(0)=1+c_Ne_N(0)$ by definition (14),

$$\bar{f}_n(0)=c_N<c_+$$

and $\bar{f}_n(x)$ is continuous and increasing on U, so that

$$\bar{f}_n(-\xi)<c_+.$$

Recalling (9), we then conclude that, for all large n a.s.

$$\max_{i\leq n}f_{n,i}(-\xi)<c_+$$

which, along with (10), gives, for all large n a.s.

$$\max_{i\leq n}f_{n,i}(\xi)<1.$$

Since $f_{n,i}(x)$ is continuous and increasing on $[0,\xi)$, the equation $f_{n,i}(x)=1$ has no solution on this interval for any $i\leq n$, for all large n a.s., which concludes the proof.

Appendix C
Proof of Lemma 2

Define $\hat{S}_{N,(i)}=\hat{S}_N-\frac{1}{n}x_ix_i^*$ and denote $\hat{S}_{N,(i)}^{-1}$ its inverse when it exists or the identity matrix otherwise. Take $2\leq p\leq 4+\eta/2$ (see **A1**) and $\varepsilon>0$ as in Lemma 1. Denoting E_{x_i} the expectation with respect to x_i and $\phi_i=1_{\{\lambda_1(\hat{S}_{N,(i)})>\varepsilon\}}$,

$$\begin{aligned}&\mathrm{E}_{x_i}\left[\phi_i\left|\frac{\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i}{1+\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i}-\frac{\frac{1}{n}\operatorname{tr}C_N\hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr}C_N\hat{S}_{N,(i)}^{-1}}\right|^p\right]\\&=\mathrm{E}_{x_i}\left[\phi_i\left|\frac{\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i-\frac{1}{n}\operatorname{tr}C_N\hat{S}_{N,(i)}^{-1}}{\left(1+\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i\right)\left(1+\frac{1}{n}\operatorname{tr}C_N\hat{S}_{N,(i)}^{-1}\right)}\right|^p\right]\\&\leq\mathrm{E}_{x_i}\left[\phi_i\left|\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i-\frac{1}{n}\operatorname{tr}C_N\hat{S}_{N,(i)}^{-1}\right|^p\right].\end{aligned}$$

Recalling that $x_i=A_Ny_i$ with y_i having independent zero mean and unit variance entries, from Lemma 5, we have

$$\begin{aligned}&\mathrm{E}_{x_i}\left[\phi_i\left|\frac{\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i}{1+\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i}-\frac{\frac{1}{n}\operatorname{tr}C_N\hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr}C_N\hat{S}_{N,(i)}^{-1}}\right|^p\right]\\&\leq\frac{\phi_iK_p}{n^{\frac{p}{2}}}\left[\left(\frac{\nu_4}{n}\operatorname{tr}(C_N\hat{S}_{N,(i)}^{-1})^2\right)^{\frac{p}{2}}+\frac{\nu_{2p}}{n^{\frac{p}{2}}}\operatorname{tr}\left((C_N\hat{S}_{N,(i)}^{-1})^2\right)^{\frac{p}{2}}\right]\end{aligned}$$

for some constant K_p depending only on p, with ν_ℓ any value such that $\mathrm{E}[|y_{ij}|^\ell] \le \nu_\ell$ (well defined from **A1**). Using $\frac{1}{n^k}\operatorname{tr} A^k \le (\frac{1}{n}\operatorname{tr} A)^k$ for $A \in \mathbb{C}^{N\times N}$ nonnegative definite and $k \ge 1$, with here $A = (C_N \hat{S}_{N,(i)}^{-1})^2$, $k = p/2$, this gives

$$\begin{aligned}
&\mathrm{E}_{x_i}\left[\phi_i \left| \frac{\frac{1}{n}x_i^* \hat{S}_{N,(i)}^{-1} x_i}{1+\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i} - \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}\right|^p\right] \\
&\le \frac{\phi_i K_p}{n^{\frac{p}{2}}}\left(\nu_4^{\frac{p}{2}} + \nu_{2p}\right)\left(\frac{1}{n}\operatorname{tr}(C_N \hat{S}_{N,(i)}^{-1})^2\right)^{\frac{p}{2}} \\
&\le \frac{K_p}{n^{\frac{p}{2}}}\left(\nu_4^{\frac{p}{2}} + \nu_{2p}\right)(c_+ C_+^2 \varepsilon^{-2})^{\frac{p}{2}} \triangleq \frac{K_p'}{n^{\frac{p}{2}}} \qquad (11)
\end{aligned}$$

where, in (11), we used $\operatorname{tr} AB \le \|A\| \operatorname{tr} B$ for $A, B \succeq 0$, $\phi_i \le 1$, $\|\hat{S}_{N,(i)}^{-1}\| \le \varepsilon^{-1}$ when $\phi_i = 1$, and $\frac{1}{n}\operatorname{tr} C_N^2 \le c_+ C_+^2$.

This being valid irrespective of $X_{(i)}$, we can take the expectation of the above expression over $X_{(i)}$ to obtain

$$\mathrm{E}\left[\phi_i \left| \frac{\frac{1}{n}x_i^* \hat{S}_{N,(i)}^{-1} x_i}{1+\frac{1}{n}x_i^*\hat{S}_{N,(i)}^{-1}x_i} - \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}\right|^p\right] \le \frac{K_p'}{n^{\frac{p}{2}}}.$$

Therefore, from Lemma 3,

$$\mathrm{E}\left[\phi_i \left| \frac{1}{n}x_i^* \hat{S}_N^{-1} x_i - \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}\right|^p\right] \le \frac{K_p'}{n^{\frac{p}{2}}}.$$

Using Boole's inequality on the n events above with $i = 1,\ldots,n$, and Markov inequality, for $\zeta > 0$,

$$\begin{aligned}
&P\left(\max_{i\le n}\left\{\phi_i \left| \frac{1}{n}x_i^* \hat{S}_N^{-1} x_i - \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}\right|\right\} > \zeta\right) \\
&\le \frac{K_p' \zeta^{-p}}{n^{\frac{p}{2}-1}}.
\end{aligned}$$

Choosing $4 < p \le 4 + \eta/2$, the right-hand side is summable. The Borel-Cantelli lemma then ensures that

$$\max_{i\le n}\left\{\phi_i \left| \frac{1}{n}x_i^* \hat{S}_N^{-1} x_i - \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}\right|\right\} \xrightarrow{\text{a.s.}} 0.$$

But, from Lemma 1, $\min_i\{\phi_i\} = 1$ for all large n a.s. Therefore, we conclude

$$\max_{i\le n}\left\{\left| \frac{1}{n}x_i^* \hat{S}_N^{-1} x_i - \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}\right|\right\} \xrightarrow{\text{a.s.}} 0. \qquad (12)$$

Since $\hat{S}_{N,(i)} - \varepsilon I_N \succ 0$ for these large n, we also have

$$\begin{aligned}
&\max_{i\le n}\left| \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}} - \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_N^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_N^{-1}}\right| \\
&= \max_{i\le n}\left| \frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_N^{-1} - \frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}}{\left(1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_{N,(i)}^{-1}\right)\left(1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_N^{-1}\right)}\right| \le \frac{1}{n}\frac{C_+}{\varepsilon}
\end{aligned}$$

where, in the last inequality, we used Lemma 4 with $B = C_N$, $A = \hat{S}_{N,(i)} - \varepsilon I_N$ and $x = \varepsilon$, along with the fact that $(1+x)^{-1} \le 1$ for $x \ge 0$.

From Proposition 1, since $\lambda_1(\hat{S}_N) \ge \lambda_1(\hat{S}_{N,(i)}) > \varepsilon$ for these large n (see (6)), we also have

$$\left|\frac{1}{n}\operatorname{tr} C_N \hat{S}_N^{-1} - \frac{c_N}{1-c_N}\right| \xrightarrow{\text{a.s.}} 0$$

and thus, from $c_N(1-c_N)^{-1}/(1+c_N(1-c_N)^{-1}) = c_N$,

$$\left|\frac{\frac{1}{n}\operatorname{tr} C_N \hat{S}_N^{-1}}{1+\frac{1}{n}\operatorname{tr} C_N \hat{S}_N^{-1}} - c_N\right| \xrightarrow{\text{a.s.}} 0.$$

Putting things together, this finally gives

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

an expression which, since $c_N > c_- > 0$, can be divided by c_N, concluding the proof.

Appendix D
Proof of Theorem 3

The proof immediately follows from the arguments of [23]. When the scalability assumption is satisfied with strict inequality, the result is exactly [23, Theorem 2]. When the scalability assumption is reduced to a loose inequality, [23, Theorem 1] does not hold, and therefore uniqueness cannot be satisfied. Nonetheless, the existence of a solution follows from the proof of [23, Lemma 1] which does not call for the scalability assumption. Indeed, since there exists $(q_1,\ldots,q_n)$ such that $q_i \ge h(q_1,\ldots,q_n)$ for all i, the algorithm

$$q_j^{(t+1)} = h_j(q_1^{(t)},\ldots,q_n^{(t)})$$

with $q_j^{(0)} = q_j$, satisfies $q_j^{(1)} \le q_j^{(0)}$ for all j. Assuming $q_j^{(t+1)} \le q_j^{(t)}$ for all j, the monotonicity assumption ensures that $q_j^{(t+2)} \le q_j^{(t+1)}$ which, by recursion, means that $q_j^{(t)}$ is a non-increasing sequence. Now, since $q_j^{(t)}$ is in the image of h_j, $q_j^{(t)} > 0$ by positivity, and therefore $q_j^{(t)}$ converges to a fixed-point (not necessarily unique). Such a fixed-point therefore exists. Note that [23, Lemma 2] provides an algorithm for reaching this fixed-point, starting with $q_j^{(0)} = 0$ for all j.

Appendix E
Proof of Theorem 2

If $\hat{C}_N$ is replaced by $\hat{S}_N$ in the statement of the result, then Theorem 2 is exactly [19, Theorem 1], which is a direct consequence of [13, Theorem 3] with some updated remarks on the $\hat{\mu}_i$ found in the discussion around [25, Theorem 17.1]. In order to prove Theorem 2, we need to justify the substitution of $\hat{S}_N$ by $\hat{C}_N$. First observe that the result is independent of a scaling of $\hat{S}_N$, and therefore we can freely substitute $\hat{S}_N$ by $\phi^{-1}(1)\hat{C}_N$ instead of $\hat{C}_N$. Using the notations of Mestre in [13], we first need to extend [13, Proposition 4]. Call $\hat{g}_M^C(z)$ the equivalent of $\hat{g}_M(z)$ designed from the eigenvectors of $\phi^{-1}(1)\hat{C}_N$ instead of those of $\hat{S}_N$ (referred to as $\hat{R}_M$ in [13] with M in place of N, and N in place of n). Then, on the chosen rectangular contour $\partial\mathbb{R}_y^-(m)$, both $\hat{g}_M^C(z)$ and $\hat{g}_M(z)$ are a.s. bounded holomorphic functions for all large N; this is due to the exact separation [14, Theorem 3] of the eigenvalues of $\hat{S}_N$ and the fact that Corollary 1 ensures the convergence between the eigenvalues of $\phi^{-1}(1)\hat{C}_N$ and of $\hat{S}_N$.

From [13, Equation (29)], $\hat{g}_M(z)$ consists of the functions $\hat{b}_M(z)$ and $\hat{m}_M(z)$ for which we also call $\hat{b}_M^C(z)$ and $\hat{m}_M^C(z)$ their equivalents for $\phi^{-1}(1)\hat{C}_N$. We need to show that the respective differences of these functions go to zero. From the definition [13, Equation (4)] of $\hat{b}_M(z)$, Theorem 1 and the fact that $\left|\frac{1}{N}\operatorname{tr}(A^{-1}-B^{-1})\right| \leq \|A^{-1}\|\|B^{-1}\|\|A-B\|$ for invertible $A,B\in\mathbb{C}^{N\times N}$, we have immediately that

$$\sup_{z\in\partial\mathbb{R}_y^-(m)}\left|\hat{b}_M(z)-\hat{b}_M^C(z)\right| \xrightarrow{\text{a.s.}} 0.$$

Similarly, using [13, Equation (6)], and $\left|a^*(A^{-1}-B^{-1})b\right| \leq |a^*b|\|A^{-1}\|\|B^{-1}\|\|A-B\|$ for $a,b\in\mathbb{C}^N$, we find

$$\sup_{z\in\partial\mathbb{R}_y^-(m)}\left|\hat{m}_M(z)-\hat{m}_M^C(z)\right| \xrightarrow{\text{a.s.}} 0.$$

By the dominated convergence theorem, this gives

$$\oint_{\partial\mathbb{R}_y^-(m)}\left(\hat{g}_M^C(z)-\hat{g}_M(z)\right)dz \xrightarrow{\text{a.s.}} 0$$

which then immediately extends [13, Proposition 4] to the present scenario. The second step to be proved is that the residue calculus performed in [13, Equations (32)–(33)] carries over to the present scenario. The poles within the contour $\partial\mathbb{R}_y^-(m)$ are the $\hat{\lambda}_k$ and the $\hat{\mu}_k$ found in the contour. The indices k such that the $\hat{\lambda}_k$ and $\hat{\mu}_k$ are within $\partial\mathbb{R}_y^-(m)$ are the same for $\hat{S}_N$ and $\phi^{-1}(1)\hat{C}_N$ for all large N, due to the exact separation property and Corollary 1. This completes the proof.

Appendix F
Useful lemmas and results

Lemma 3 (A matrix-inversion lemma): Let $x\in\mathbb{C}^N$, $A\in\mathbb{C}^{N\times N}$, and $t\in\mathbb{R}$. Then, whenever the inverses exist

$$x^*\left(A+txx^*\right)^{-1}x = x^*A^{-1}x(1+tx^*A^{-1}x)^{-1}.$$

Lemma 4 (Rank-one perturbation): Let $v\in\mathbb{C}^N$, $A,B\in\mathbb{C}^{N\times N}$ nonnegative definite, and $x>0$. Then

$$\operatorname{tr}B\left(A+vv^*+xI_N\right)^{-1}-\operatorname{tr}B\left(A+xI_N\right)^{-1}\leq x^{-1}\|B\|.$$

Lemma 5 (Trace lemma): [26, Lemma B.26] Let $A\in\mathbb{C}^{N\times N}$ be non-random and $y=[y_1,\ldots,y_N]^\mathsf{T}\in\mathbb{C}^N$ be a vector of independent entries with $\mathrm{E}[y_i]=0$, $\mathrm{E}[|y_i|^2]=1$, and $\mathrm{E}[|y_i|^\ell]\leq\nu_\ell$ for all $\ell\leq 2p$, with $p\geq 2$. Then,

$$\mathrm{E}\left[|y^*Ay-\operatorname{tr}A|^p\right]\leq C_p\left(\left(\nu_4\operatorname{tr}AA^*\right)^{\frac{p}{2}}+\nu_{2p}\operatorname{tr}(AA^*)^{\frac{p}{2}}\right)$$

for C_p a constant depending on p only.

Proposition 1 (A random matrix result): Let $X=[x_1,\ldots,x_n]\in\mathbb{C}^{N\times n}$ with $x_i=A_Ny_i$, $A_N\in\mathbb{C}^{N\times M}$, $M\geq N$, where $y_i=[y_{i1},\ldots,y_{iM}]\in\mathbb{C}^M$ has independent entries satisfying $\mathrm{E}[y_{ij}]=0$, $\mathrm{E}[|y_{ij}|^2]=1$, $\mathrm{E}[|y_{ij}|^\ell]<\nu_\ell$ for all $\ell\leq 2p$ and $C_N\triangleq A_NA_N^*$ is nonnegative definite with $\|C_N\|<C_+<\infty$. Assume $c_N=N/n$ and $\bar{c}_N=M/N\geq 1$ satisfy $\limsup_N c_N<\infty$ and $\limsup_N \bar{c}_N<\infty$, as $N,n,M\to\infty$. Then, for $z<0$, and $p>2$,

$$\mathrm{E}\left[\left|\frac{1}{N}\operatorname{tr}C_N\left(\frac{1}{n}XX^*-zI_N\right)^{-1}-e_N(z)\right|^p\right]\leq\frac{K_p}{N^{\frac{p}{2}}} \quad (13)$$

for K_p a constant depending only on p, ν_ℓ for $\ell\leq 2p$, and z, while $e_N(z)$ is the unique positive solution of

$$e_N(z)=\int\frac{t}{(1+c_Ne_N(z))^{-1}t-z}dF^{C_N}(t) \qquad (14)$$

where F^{C_N} is the eigenvalue distribution of C_N. The function $\mathbb{R}^-\to\mathbb{R}^+,\ z\mapsto e_N(z)$ is increasing.

Moreover, for any N_0, as $N,n\to\infty$ with $\limsup_N c_N<\infty$, for $z\in\mathbb{R}\setminus\mathcal{S}_{N_0}$, where $\mathcal{S}_{N_0}$ is the union of the supports of the eigenvalue distributions of $\frac{1}{n}XX^*$ for all $N\geq N_0$,

$$\frac{1}{N}\operatorname{tr}C_N\left(\frac{1}{n}XX^*-zI_N\right)^{-1}-e_N(z)\xrightarrow{\text{a.s.}}0. \qquad (15)$$

Proof: To prove the first part of Proposition 1, we follow the steps of the proof of [27]. Note first that we can append A_N into an $M\times M$ matrix by adding rows of zeros, without altering the left-hand side of (13). Using the notations of [27], we consider the simple case where $A_n=0$ and $\sigma_{ij}^n=C_i^n$, where C_i^n denotes the i-th eigenvalue of C_N. Although this updated proof of [27] would impose C_N to be diagonal, it is rather easy to generalize to non-diagonal C_N (see e.g. [28], [29]). The proof then extends to the non i.i.d. case when using Lemma 5 instead of [27, (B.1)]. The second part follows from the first part immediately for $z<0$. In order to extend the result to $z\in\mathbb{R}\setminus\mathcal{S}_{N_0}$, note that both left-hand side terms in (15) are uniformly bounded in any compact $\mathcal{D}$ away from $\mathcal{S}_{N_0}$ and including part of $\mathbb{R}^-$, and are holomorphic on $\mathcal{D}$. From Vitali's convergence theorem [30], their difference therefore tends to zero on $\mathcal{D}$, which is what we need. ■

Proposition 2 (No eigenvalue outside the support): Let $X=[x_1,\ldots,x_n]\in\mathbb{C}^{N\times n}$ with $x_i=A_Ny_i$, $A_N\in\mathbb{C}^{N\times M}$, where $y_i=[y_{i1},\ldots,y_{iM}]\in\mathbb{C}^M$ has independent entries satisfying $\mathrm{E}[y_{ij}]=0$, $\mathrm{E}[|y_{ij}|^2]=1$ and $\mathrm{E}[|y_{ij}|^{8+\eta}]<\alpha$ for some $\eta,\alpha>0$, $C_N\triangleq A_NA_N^*$ has bounded spectral norm, and $N,n,M\to\infty$ with $\limsup_N N/n<1$, and $1\leq\limsup_N M/N<\infty$. Let N_0 be an integer and $[a,b]\subset\mathbb{R}\cup\{\pm\infty\}$, $b>a$, a segment outside the closure of the union of the supports $F^{N/n,C_N}$, $N\geq N_0$, with $F^{t,A}$ the limiting support of the eigenvalues of $\frac{1}{n}XX^*$ when C_N has the same spectrum as A for all N and $N/n\to t$. Then, for all large n a.s., no eigenvalue of $\frac{1}{n}XX^*$ is found in $[a,b]$.

Proof: Appending A_N into an $M\times M$ matrix filled with zeros, this unfolds from [14, Theorem 3], with the supports $F^{N/n,C_N}$ appended with the singleton $\{0\}$. Now, for $A_N\in\mathbb{C}^{N\times M}$, such that $A_NA_N^*$ is positive definite, zero is not an eigenvalue of $\frac{1}{n}XX^*$ for all N, a.s., which gives the result. ■

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] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5353–5368, Nov. 2008.

[4] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[5] R. A. Maronna, D. R. Martin, and J. V. Yohai, *Robust Statistics: Theory and Methods*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 2006.

[6] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The annals of statistics*, pp. 51–67, 1976.

[7] 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.

[8] S. Watts, "Radar Detection Prediction in Sea Clutter Using the Compound K-Distribution model," *IEE Proceeding, Part. F*, vol. 132, no. 7, pp. 613–620, December 1985.

[9] 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.

[10] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.

[11] 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.

[12] P. Bianchi, J. Najim, M. Maida, and M. Debbah, "Performance of some eigen-based hypothesis tests for collaborative sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2400–2419, 2011.

[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] 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.

[15] N. E. Karoui, "Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond," *The Annals of Applied Probability*, vol. 19, no. 6, pp. 2362–2405, 2009.

[16] U. J. G. Frahm, "Tylers m-estimator, random matrix theory, and generalized elliptical distributions with applications to finance," *Discussion papers in statistics and econometrics*, vol. 2, no. 7, 2008.

[17] D. E. Tyler, "Some results on the existence, uniqueness, and computation of the m-estimates of multivariate location and scatter," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, p. 354, 1988.

[18] J. T. Kent and D. E. Tyler, "Redescending M-estimates of multivariate location and scatter," *The Annals of Statistics*, pp. 2102–2119, 1991.

[19] 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.

[20] P. Vallet, "Random matrix theory and applications to statistical signal processing," Ph.D. dissertation, Université de Paris-Est, Marne-la-Vallée, 2011.

[21] 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.

[22] J. Li and P. Stoica, *MIMO Radar Signal Processing*, 1st ed. Wiley, 2009.

[23] 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.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[25] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.

[26] Z. D. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, 2nd ed. New York, NY, USA: Springer Series in Statistics, 2009.

[27] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.

[28] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011.

[29] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012. [Online]. Available: http://arxiv.org/abs/0906.3682

[30] E. C. Titchmarsh, *The Theory of Functions*. New York, NY, USA: Oxford University Press, 1939.