Robust spiked random matrices and a robust G-MUSIC estimator

Romain Couillet

Robust spiked random matrices and a robust G-MUSIC estimator[☆]

Romain Couillet[a]

[a] *Telecommunication department, Supélec, Gif sur Yvette, France*

Abstract

A class of robust estimators of scatter applied to information-plus-impulsive noise samples is studied, where the sample information matrix is assumed of low rank; this generalizes the study (Couillet et al., 2013b) to spiked random matrix models. It is precisely shown that, as opposed to sample covariance matrices which may have asymptotically unbounded (eigen-)spectrum due to the sample impulsiveness, the robust estimator of scatter has bounded spectrum and may contain isolated eigenvalues which we fully characterize. We show that, if found beyond a certain detectability threshold, these eigenvalues allow one to perform statistical inference on the eigenvalues and eigenvectors of the information matrix. We use this result to derive new eigenvalue and eigenvector estimation procedures, which we apply in practice to the popular array processing problem of angle of arrival estimation. This gives birth to an improved algorithm based on the MUSIC method, which we refer to as robust G-MUSIC.

Keywords: random matrix theory, robust estimation, spiked models, MUSIC.

1. Introduction

The mathematical advances in the field of random matrix theory have recently allowed for the improvement of sometimes old statistical estimation methods when the data have population size N is commensurable with the sample size n, therefore disrupting the traditional assumption $n \gg N$. One of the recent contributions of random matrix theory lies in the introduction of methods to retrieve information contained in low rank perturbations of large matrices with independent entries, which are referred to as spiked models. The initial study of such models (Baik and Silverstein, 2006) for matrices of the type $\hat{S}_N = \frac{1}{n}(I_N + A)XX^*(I_N + A^*)$, where $X \in \mathbb{C}^{N\times n}$ has independent and identically distributed (i.i.d.) zero mean, unit variance, and finite fourth moment entries and A has fixed rank L, has shown that, as $N, n \to \infty$ with

[☆]This work is jointly supported by the French ANR DIONISOS project (ANR-12-MONU-OOO3) and the GDR ISIS–GRETSI "Jeunes Chercheurs" Project.

Email address: romain.couillet@supelec.fr (Romain Couillet)

$N/n \to c \in (0,\infty)$, $\hat{S}_N$ may exhibit up to L isolated eigenvalues strictly away from the *bounded* support of the limiting empirical distribution μ of $\hat{S}_N^\circ = \frac{1}{n}XX^*$, while the other eigenvalues of $\hat{S}_N$ get densely compacted in the support of μ. This result has triggered multiple works on various low rank perturbation models for Gram, Wigner, or general square random matrices (Benaych-Georges and Rao, 2011; Paul, 2007; Benaych-Georges et al., 2010) with similar conclusions. Of particular interest to us here is the information-plus-noise model $\hat{S}_N = \frac{1}{n}(X+A)(X+A)^*$ introduced in (Benaych-Georges and Rao, 2011) which is closer to our present model. Other generalizations explored the direction of turning X into the more general $XT^{\frac{1}{2}}$ model for $T = \mathrm{diag}(\tau_1, \ldots, \tau_n) \succeq 0$, such that $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\tau_i} \to \nu$ weakly, where ν has bounded support $\mathrm{Supp}(\nu)$ and $\max_i\{\mathrm{dist}(\tau_i, \mathrm{Supp}(\nu))\} \to 0$ (Chapon et al., 2012). In this scenario again, thanks to the fundamental assumption that no τ_i can escape $\mathrm{Supp}(\nu)$ asymptotically, only finitely many eigenvalues of $\hat{S}_N$ can be found away from the support of the limiting spectral distribution of $\frac{1}{n}XTX^*$, and these eigenvalues are intimately linked to A.

The major interest of the spiked models in practice is twofold. First, if the (non observable) perturbation matrix A constitutes the relevant information to the system observer, then the observable isolated eigenvalues and associated eigenvectors of $\hat{S}_N$ contain information about A. These isolated eigenvalues and eigenvectors are therefore important objects to characterize. Moreover, since $\hat{S}_N$ has the same limiting spectrum as that of simple random matrix models, this characterization is usually quite easy and leads to tractable expressions and computationally efficient algorithms. This led to notable contributions to statistical inference and in particular to detection and estimation techniques for signal processing (Mestre, 2008a; Nadler, 2010; Hachem et al., 2013; Couillet and Hachem, 2012).

However, from the discussion of the first paragraph, these works have a few severe practical limitations in that: (i) the support of the limiting spectral distribution of $\hat{S}_N$ must be bounded for isolated eigenvalues to be detectable and exploitable and (ii) no eigenvalue of $\hat{S}_N^\circ$ (the unperturbed model) can be isolated, to avoid risking a confusion between isolated eigenvalues of $\hat{S}_N$ arising from A and isolated eigenvalues of $\hat{S}_N$ intrinsically linked to $\hat{S}_N^\circ$. This therefore rules out the possibility to straightforwardly extend these techniques in practice to impulsive noise models $XT^{\frac{1}{2}}$ where $T = \mathrm{diag}(\tau_1, \ldots, \tau_n)$ with either τ_i i.i.d. arising from a distribution with unbounded support or $\tau_i = 1$ for all but a few indices i. In the former case, the support of the limiting spectrum of $\hat{S}_N^\circ$ is unbounded (Couillet and Hachem, 2013, Proposition 3.4), therefore precluding information detection, while in the latter spurious eigenvalues in the spectrum of $\hat{S}_N$ may arise that are also found in $\hat{S}_N^\circ$ and therefore constitute false information (note that this case can be seen as one where low rank perturbations are present *both in the population and in the sample directions* which cannot be discriminated). Such impulsive models are nonetheless fundamental in many applications such as statistical finance or radar array processing, where impulsive samples are classically met.

Traditional statistical techniques to accommodate for impulsive samples fall in the realm of robust estimation (Maronna et al., 2006), the study of which has long remained limited to the assumption $n \gg N$. Recently though, in a series of articles (Couillet et al., 2013a,b; Couillet and McKay, 2013), the author of the present article and his coauthors provided a random matrix analysis of robust estimation, i.e., assuming $N, n \to \infty$ and $N/n \to c \in (0, 1)$, which revealed that robust sample estimates $\hat{C}_N^\circ$ of scatter (or covariance) matrices can be fairly easily analyzed through simpler equivalent random matrix models. In (Couillet et al., 2013b), a noise-only setting of the present article is considered, i.e., with $A = 0$, for which it is precisely shown that robust estimators of scatter can be assimilated as special models of the type of $\hat{S}_N^\circ$.[1] Besides, it importantly appears that the limiting spectrum distribution of $\hat{C}_N^\circ$ always has bounded support, irrespective of the impulsiveness of the samples. Also, it is proved (although not mentioned explicitly) that, asymptotically, isolated eigenvalues of $\hat{C}_N^\circ$ (arising from isolated τ_i) can be found but that none of the eigenvalues can exceed a fixed finite value.

In the present work, we extend the model studied in (Couillet et al., 2013b) by introducing a finite rank perturbation A to the robust estimator of scale $\hat{C}_N^\circ$, the resulting matrix being denoted $\hat{C}_N$. As opposed to non-robust models, it shall appear (quite surprisingly on the onset) that $\hat{C}_N$ now allows for finitely many isolated eigenvalues to appear beyond the aforementioned fixed finite value (referred from now on to as the detection threshold), these eigenvalues being related to A. This holds even if $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\tau_i}$ has unbounded support in the large n regime. As such, any isolated eigenvalue of $\hat{C}_N$ found below the detection threshold may carry information about A or may merely be an outlier due to an isolated τ_i (as in the non-robust context) but any eigenvalue found beyond the detection threshold necessarily carries information about A. This has important consequences in practice as now low rank perturbations *in the sample direction* are appropriately harnessed by the robust estimator while the (more relevant) low rank perturbations *in the population direction* can be properly estimated. We shall introduce an application of these results to array processing by providing two novel estimators for the power and steering direction of signals sources captured by a large sensor array under impulsive noise.

Our contribution thus lies on both theoretical and practical grounds. We first introduce in Theorem 1 the generalization of (Couillet et al., 2013b) to the perturbed model $\hat{C}_N$ which we precisely define in Section 2. The main results are then contained in Section 3. In this section, Theorem 2 provides the localization of the eigenvalues of $\hat{C}_N$ in the large system regime along with associated population eigenvalue and eigenvector estimators when the limiting distribution for $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\tau_i}$ is known. This result is then extended in Theorem 3 thanks to a two-step estimator where the τ_i are directly estimated. A practical

[1] These models are special in that XTX^* becomes now $XTVX^*$ for a diagonal matrix V which makes VT bounded in norm. However, V contains non-observable information about T, which makes $\hat{S}_N^\circ$ only observable through its approximation by $\hat{C}_N^\circ$.

application of these novel methods to the context of steering angle estimation for array processing is then provided, leading to an improved algorithm referred to as robust G-MUSIC. Simulation results in this context are then displayed that confirm the improved performance of using robust schemes versus traditional sample covariance matrix-based techniques. We finally close the article with concluding remarks in Section 4.

Notations: Vectors and matrices are represented in lower- and upper-case characters, respectively. Transpose and Hermitian transpose of X are denoted respectively by X^T and X^*. The norm $\|\cdot\|$ is the spectral norm for matrices and the Euclidean norm for vectors. The matrix $T^{\frac{1}{2}}$ is the nonnegative definite square root of the Hermitian nonnegative definite matrix T. The eigenvalues of a Hermitian matrix $X \in \mathbb{C}^{N\times N}$ are denoted in order as $\lambda_1(X) \geq \ldots \geq \lambda_N(X)$. Hermitian matrix ordering is denoted $X \succeq Y$, i.e., $X - Y$ is nonnegative definite. The support of a measure μ is denoted $\mathrm{Supp}(\mu)$. Almost sure convergence will be sometimes denoted "$\xrightarrow{\text{a.s.}}$". The Dirac measure at x is denoted $\boldsymbol{\delta}_x$.

2. Model and Motivation

Let $n \in \mathbb{N}$. For $i \in \{1, \ldots, n\}$, we consider the following statistical model

$$y_i = \sum_{l=1}^{L} \sqrt{p_l} a_l s_{li} + \sqrt{\tau_i} w_i \tag{1}$$

with $y_i \in \mathbb{C}^N$ satisfying the following hypotheses.

Assumption 1. *The vectors $y_1, \ldots, y_n \in \mathbb{C}^N$ satisfy the following conditions:*

1. *$\tau_1, \ldots, \tau_n \in (0, \infty)$ are random scalars such that $\nu_n \triangleq \frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\tau_i} \to \nu$ weakly, almost surely, where $\int t\nu(dt) = 1$;*
2. *$w_1, \ldots, w_n \in \mathbb{C}^N$ are random independent unitarily invariant $\sqrt{N}$-norm vectors, independent of $\tau_1, \ldots, \tau_n$;*
3. *$L \in \mathbb{N}$, $p_1 \geq \ldots \geq p_L \geq 0$ are deterministic and independent of N*
4. *$a_1, \ldots, a_L \in \mathbb{C}^N$ are deterministic or random and such that*

$$A^*A \xrightarrow{\text{a.s.}} \mathrm{diag}(p_1, \ldots, p_L)$$

   as $N \to \infty$, with $A \triangleq [\sqrt{p_1}a_1, \ldots, \sqrt{p_L}a_L] \in \mathbb{C}^{N\times L}$
5. *$s_{1,1}, \ldots, s_{Ln} \in \mathbb{C}$ are independent with zero mean, unit variance, and uniformly bounded moments of all orders.*

For further use, we shall define

$$A_i \triangleq \begin{bmatrix} \sqrt{p_1}a_1 & \ldots & \sqrt{p_L}a_L & \sqrt{\tau_i} I_N \end{bmatrix} \in \mathbb{C}^{N\times(N+L)}.$$

In particular, $A_i A_i^* = AA^* + \tau_i I_N$.

Remark 1 (Application contexts). *The system* (1) *can be adapted to multiple scenarios in which the s_{li} model scalar signals or data originated from L sources of respective powers $p_1, \ldots, p_L$ carried by the vectors $a_1, \ldots, a_L$, while the $\sqrt{\tau_i} w_i$ model additive impulsive noise. Two examples are:*

- *wireless communication channels in which signals s_{li} originating from L transmitters are captured by an N-antenna receiver. The vectors a_l are here random independent channels for which it is natural to assume that $a_l^* a_{l'} \to \boldsymbol{\delta}_{l-l'}$ (e.g., for independent $a_l \sim \mathcal{CN}(0, I_N/N)$);*
- *array processing in which L sources emit signals s_{li} captured by an antenna array through steering vectors $a_l = a(\theta_l)$ for a given $a(\theta)$ function and angles of arrival $\theta_1, \ldots, \theta_L \in [0, 2\pi)$. In the case of uniform linear arrays with inter-antenna distance d, $[a(\theta)]_j = N^{-\frac{1}{2}} \exp(2\pi \imath d j \sin(\theta))$.*

The noise impulsiveness is translated by the τ_i coefficients. The vectors $\sqrt{\tau_i} w_i$ are for instance i.i.d. elliptic random vectors if the τ_i are i.i.d. with absolutely continuous measure $\tilde{\nu}_n$ having a limit $\tilde{\nu}$ (in which case, we easily verify that $\nu_n \to \nu = \tilde{\nu}$ almost surely (a.s.)). This particularizes to additive white Gaussian noise if $2N\tau_i$ is chi-square with $2N$ degrees of freedom (in this case, $\nu = \boldsymbol{\delta}_1$). Of interest in this article is however the scenarios where ν has unbounded support, e.g., when the τ_i are either random i.i.d. and heavy-tailed or contain a few arbitrarily large outliers, which both correspond to impulsive noise scenarios.

Remark 2 (Technical comments). *From a purely technical perspective, it is easily seen from the proofs of our main results in Section 5 that some of the items of Assumption 1 could have been relaxed. In particular, Item (4) could have been relaxed into "all accumulation points of A^*A are similar to* $\mathrm{diag}(q_1, \ldots, q_L)$ *for given $q_1 \geq \ldots \geq q_L$" as in e.g., (Chapon et al., 2012). Also, similar to (Couillet et al., 2013b), the convergence of ν_n in Item (1) could be relaxed to the cost of introducing a tightness condition on the sequence $\{\nu_n\}_{n=1}^{\infty}$ and to loose the convergence of measure in the discussion following Theorem 1. For readability and since Assumption 1 gathers most of the scenarios of interest, we restrict ourselves to those (already quite general) hypotheses.*

We now define the robust estimate of scatter $\hat{C}_N$. We start by denoting $u : [0, \infty) \to (0, \infty)$ any function satisfying the following hypotheses.

Assumption 2. *The function u is characterized by*

1. *u is continuous, nonnegative, and non-increasing from $[0, \infty)$ onto $(0, u(0)] \subset (0, \infty)$;*
2. *for $x \geq 0$, $\phi(x) \triangleq xu(x)$ is increasing and bounded with*

$$\phi_\infty \triangleq \lim_{x \to \infty} \phi(x) > 1$$

3. *there exists $m > 0$ such that $\nu([0, m)) < 1 - \phi_\infty^{-1}$;*

4. *for all* $a > b > 0$,

$$\limsup_{t\to\infty} \frac{\nu((t,\infty))}{\phi(at)-\phi(bt)} = 0.$$

These assumptions are the same as in Couillet et al. (2013b) which are therefore not altered by the updated model (1).

The function u being given, we now define $\hat{C}_N$, when it exists, as the unique solution to the fixed-point matrix-valued equation in Z:

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

For $i \in \{1,\ldots,N\}$, we shall denote $\hat{\lambda}_i \triangleq \lambda_i(\hat{C}_N)$ and $\hat{u}_i \in \mathbb{C}^N$ the i-th largest eigenvalue of $\hat{C}_N$ and its associated eigenvector.

Due to its implicit formulation, the study of $\hat{C}_N$ for every fixed N, n couple is quite involved in general. As such, similar to (Couillet et al., 2013b), we shall place ourselves in the regime where both N and n are large but with non trivial ratio. Hence, we shall assume the following system growth regime.

Assumption 3. *The integer* $N = N(n)$ *is such that* $c_n \triangleq N/n$ *satisfies*

$$\lim_{n\to\infty} c_n = c \in (0, \phi_\infty^{-1}).$$

Meanwhile, L *remains constant independently of* N, n.

Up to differences in the hypotheses of Assumption 2 and Assumption 3, and a slight difference in notations, $\hat{C}_N$ is exactly the robust estimator of scatter proposed by Maronna in (Maronna, 1976). As a direct application of (Chitour et al., 2014), under Assumption 1 and Assumption 2, $\hat{C}_N$ is almost surely well defined for each couple N, n with $N < n$. Also, from (Couillet et al., 2013b), $\hat{C}_N$ can be written (at least for all large n) in the technically more convenient form (see discussions in (Couillet et al., 2013b))

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

where $v : x \mapsto u \circ g^{-1}$, $g : x \mapsto x/(1-c_n\phi(x))$, and $\hat{C}_{(i)} = \hat{C}_N - u\left(\frac{1}{N} y_i^* \hat{C}_N^{-1} y_i\right) y_i y_i^*$. We shall further denote $\psi(x) = xv(x)$. It is easy to see that v is non-increasing while ψ is increasing with limit $\psi_\infty = \phi_\infty/(1 - c_n\phi_\infty)$.

With these definitions in place, we are now in position to present our main results.

3. Main Results

The first objective of the article is to study the spectrum of $\hat{C}_N$ and in particular its largest eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_L$ and associated eigenvectors $\hat{u}_1, \ldots, \hat{u}_L$, in the large N, n regime. This study will in turn allow us to retrieve information on $p_1, \ldots, p_L$ and $a_1, \ldots, a_L$. As an application, a novel improved angle estimator for array processing will then be provided.

3.1. Localisation and estimation

Our first result is an extension of (Couillet et al., 2013b, Theorem 2) which states that $\hat{C}_N$, the implicit structure of which makes it complicated to analyze, can be appropriately replaced by a more practical random matrix $\hat{S}_N$, which is much easier to study.

Theorem 1 (Asymptotic model equivalence). *Let Assumptions 1, 2, and 3 hold. Then*

$$\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{S}_N \triangleq \frac{1}{n}\sum_{i=1}^{n} v_c(\tau_i\gamma) A_i \bar{w}_i \bar{w}_i^* A_i^*$$

with γ the unique solution to

$$1 = \int \frac{\psi_c(t\gamma)}{1 + c\psi_c(t\gamma)}\nu(dt)$$

v_c and ψ_c the limits of v and ψ as $c_n \to c$, and $\bar{w}_i = [s_{1i}, \ldots, s_{Li}, w_i r_i/\sqrt{N}]^\mathsf{T}$, with $r_i \geq 0$ such that $2Nr_i^2$ is a chi-square random variable with $2N$ degrees of freedom, independent of w_i.[2]

Remark 3 (From robust estimator to sample covariance matrix). *Note that, if the function v_c in the expression of $\hat{S}_N$ were replaced by the constant 1 (and $r_i/\sqrt{N}$ set to one), $\hat{S}_N$ would be the classical sample covariance matrix of $y_1, \ldots, y_n$. Although it is here highly non rigorous to let v_c tend to 1 uniformly in Theorem 1, this remark somewhat reveals the classical robust estimation intuition according to which the larger ϕ_∞ (as a consequence of u and v_c being close to 1) the less robust $\hat{C}_N$.*

As a corollary of Theorem 1, we have

$$\max_{1\leq i\leq N} \left|\hat{\lambda}_i - \lambda_i(\hat{S}_N)\right| \xrightarrow{\text{a.s.}} 0 \tag{2}$$

(which unfolds from applying (Horn and Johnson, 1985, Theorem 4.3.7)) and therefore all eigenvalues of $\hat{C}_N$ can be accurately controlled through the eigenvalues of $\hat{S}_N$.

[2]Note that $w_i r_i/\sqrt{N}$ as defined above is a standard Gaussian vector and therefore $\bar{w}_i$ has independent entries of zero mean and unit variance. In fact, the result can be equivalently formulated with $\bar{w}_i$ replaced by $\bar{\bar{w}}_i \triangleq [s_{1i}, \ldots, s_{Li}, w_i]^\mathsf{T}$, but the former vector, having independent entries, is of more interest statistically.

Let us assume for a moment that $p_1 = \ldots = p_L = 0$. Then, from Theorem 1, Assumption 1, and (Silverstein and Choi, 1995), $\mu_n \triangleq \frac{1}{N}\sum_{i=1}^N \boldsymbol{\delta}_{\hat{\lambda}_i} \to \mu$ weakly, a.s., where μ has a density on $\mathbb{R}$ with bounded support $\mathrm{Supp}(\mu) \subset \mathbb{R}^+$. Denote

$$\begin{aligned} S_\mu^- &\triangleq \inf(\mathrm{Supp}(\mu)) \\ S_\mu^+ &\triangleq \sup(\mathrm{Supp}(\mu)) \\ S^+ &\triangleq \frac{\phi_\infty(1+\sqrt{c})^2}{\gamma(1-c\phi_\infty)}. \end{aligned}$$

Since $\tau_i v_c(\tau_i\gamma) = \gamma^{-1}\psi_c(\tau_i\gamma) < \gamma^{-1}\psi_{c,\infty}$ with $\psi_{c,\infty} = \phi_\infty/(1-c\phi_\infty)$, we have

$$\hat{S}_N \preceq \frac{\phi_\infty}{\gamma(1-c\phi_\infty)}\frac{1}{n}\sum_{i=1}^n w_i w_i^*$$

so that, according to (Marčenko and Pastur, 1967; Bai and Silverstein, 1998) and (2), for each $\varepsilon > 0$, $\hat{\lambda}_1 < S^+ + \varepsilon$ for all large n a.s. Of course, $S^+ \geq S_\mu^+$. If in addition $\max_{1\leq i\leq n}\{\mathrm{dist}(\tau_i, \mathrm{Supp}(\nu))\} \xrightarrow{\text{a.s.}} 0$, then from (Bai and Silverstein, 1998), we even have $\hat{\lambda}_1 \xrightarrow{\text{a.s.}} S_\mu^+$; but this constraint is of little practical interest so that in general one may have $S_\mu^+ < \hat{\lambda}_1 < S^+$ infinitely often.

Coming back to generic values for $p_1, \ldots, p_L$, the idea of the results below is that, for sufficiently large $p_1, \ldots, p_L$, the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_L$ may exceed $S^+ + \varepsilon$ and contain information to estimate $p_1, \ldots, p_L$ as well as bilinear forms involving $a_1, \ldots, a_L$. The exact location of the eigenvalues and the value of these estimates shall be expressed as a function of the fundamental object $\delta(x)$, defined for $x \in \mathbb{R}^* \setminus [S_\mu^-, S_\mu^+]$ as the unique real solution to

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

The function $\delta(x)$ is the restriction to $\mathbb{R}^* \setminus [S_\mu^-, S_\mu^+]$ of the Stieltjes transform of $c\mu + (1-c)\boldsymbol{\delta}_0$ and is, as such, increasing on $(S^+, \infty) \subset (S_\mu^+, \infty)$; see (Silverstein and Choi, 1995; Couillet and Hachem, 2013) and Section 5 for details. Therefore, the following definition of p_-, which will be referred to as the detectability threshold, is licit

$$p_- \triangleq \lim_{x \downarrow S^+} -c\left(\int \frac{\delta(x) v_c(t\gamma)}{1+\delta(x) t v_c(t\gamma)}\nu(dt)\right)^{-1}.$$

We shall further denote $\mathcal{L} \triangleq \{j, p_j > p_-\}$.

We are now in position to provide our main results.

Theorem 2 (Robust estimation under known ν). *Let Assumptions 1, 2, and 3 hold. Denote u_k the eigenvector associated with the k-th largest eigenvalue*

of AA^* (in case of multiplicity, take any vector in the eigenspace with $u_1, \ldots, u_L$ orthogonal) and $\hat{u}_1, \ldots, \hat{u}_N$ the eigenvectors of $\hat{C}_N$ respectively associated with the eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_N$. Then, we have the following three results.

0. Extreme eigenvalues. *For each $j \in \mathcal{L}$,*

$$\hat{\lambda}_j \xrightarrow{\text{a.s.}} \Lambda_j > S^+$$

while $\limsup_n \hat{\lambda}_{|\mathcal{L}|+1} \leq S^+$ a.s., where Λ_j is the unique positive solution to

$$-c\left(\delta(\Lambda_j)\int \frac{v_c(\tau\gamma)}{1+\delta(\Lambda_j)\tau v_c(\tau\gamma)}\nu(d\tau)\right)^{-1} = p_j.$$

1. Power estimation. *For each $j \in \mathcal{L}$,*

$$-c\left(\delta(\hat{\lambda}_j)\int \frac{v_c(\tau\gamma)}{1+\delta(\hat{\lambda}_j)\tau v_c(\tau\gamma)}\nu(d\tau)\right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

2. Bilinear form estimation. *For each $a, b \in \mathbb{C}^N$ with $\|a\| = \|b\| = 1$, and $j \in \mathcal{L}$*

$$\sum_{k, p_k = p_j} a^* u_k u_k^* b - \sum_{k, p_k = p_j} w_k a^* \hat{u}_k \hat{u}_k^* b \xrightarrow{\text{a.s.}} 0$$

where

$$w_k = \frac{\displaystyle\int \frac{v_c(t\gamma)}{\left(1+\delta(\hat{\lambda}_k)tv_c(t\gamma)\right)^2}\nu(dt)}{\displaystyle\int \frac{v_c(t\gamma)}{1+\delta(\hat{\lambda}_k)tv_c(t\gamma)}\nu(dt)\left(1-\frac{1}{c}\int \frac{\delta(\hat{\lambda}_k)^2 t^2 v_c(t\gamma)^2}{\left(1+\delta(\hat{\lambda}_k)tv_c(t\gamma)\right)^2}\nu(dt)\right)}.$$

Item 0. in Theorem 2 provides a necessary and sufficient condition, i.e., $p_j > p_-$, for the existence of outlying eigenvalues in the spectrum of $\hat{C}_N$. In turn, this provides a means to estimate each p_j, $j \in \mathcal{L}$, along with bilinear forms involving a_j, from $\hat{\lambda}_j$ and $\hat{u}_j$. It is important here to note that, although the right-edge of the spectrum of μ is S_μ^+, due to the little control on τ_i in practice (in particular some of the τ_i may freely be arbitrarily large), isolated eigenvalues may be found infinitely often beyond S_μ^+ which do not carry information. This is why the (possibly pessimistic) choice of S^+ as an eigenvalue discrimination threshold was made. The major potency of the robust estimator $\hat{C}_N$ is indeed to be able to maintain these non informative eigenvalues below the known value S^+. As such, eigenvalues found above S^+ must contain information about A (at least with high probability) and this information can be retrieved, while isolated eigenvalues found below S^+ may arise from spurious values of τ_i, therefore

containing no relevant information, or may contain relevant information but that cannot be trusted.

Figure 1 and Figure 2 provide the histogram and limiting spectral distribution of $\hat{C}_N$ and $\frac{1}{n}YY^*$, $Y = [y_1, \ldots, y_n]$, respectively, for $u(x) = (1+\alpha)/(\alpha+x)$, $\alpha = 0.2$, $N = 200$, $n = 1000$, τ_i i.i.d. equal in distribution to $t^2(\beta-2)\beta^{-1}$ with t a Student-t random scalar of parameter $\beta = 100$, and $L = 2$ with $p_1 = p_2 = 1$, $a_1 = a(\theta_1)$, $a_2 = a(\theta_2)$, $\theta_1 = 10°$, $\theta_2 = 12°$, $a(\theta)$ being defined in Remark 1 (as well as in Assumption 4 below). These curves confirm that, while the limiting spectral measure of $\frac{1}{n}YY^*$ is unbounded, that of $\hat{C}_N$ is bounded. The numerically evaluated values of S_μ^+ and S^+ are reported in Figure 1. They reveal a rather close proximity between both values. In terms of empirical eigenvalues, note the particularly large gap between the isolated eigenvalues of $\hat{C}_N$ and the $N-2$ smallest ones, which may seem at first somewhat surprising for $p_1 = p_2 = 1$ since this setting induces a ratio 1 between the power carried by information versus noise (indeed, $A^*A \simeq I_2$ while $\mathrm{E}[\tau_i w_i w_i^*] = I_N$); this in fact results from the function u which, in attenuating the rare samples of large amplitudes, significantly reduces the noise power but only weakly affects the information part which has roughly constant amplitude across the samples. Also observe from Figure 2 that, as predicted, the largest two eigenvalues of $\frac{1}{n}YY^*$ do not isolate from the majority of the eigenvalues.

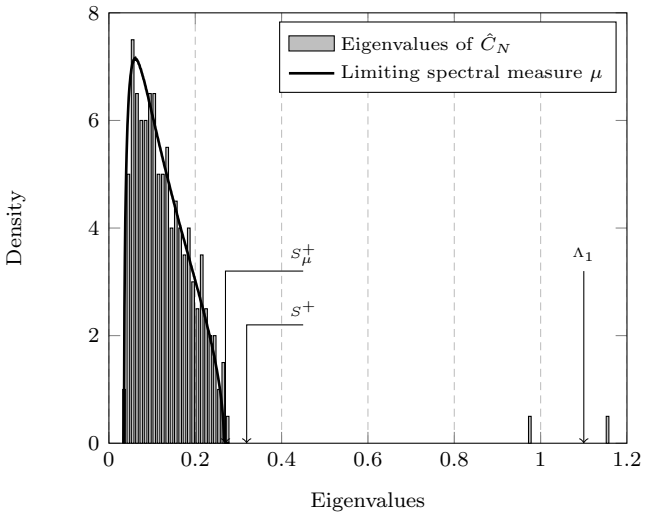

Figure 1: Histogram of the eigenvalues of $\hat{C}_N$ against the limiting spectral measure, for $u(x) = (1+\alpha)/(\alpha+x)$ with $\alpha = 0.2$, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Student-t impulsions.

Items 1. and 2. in Theorem 2 then provide a means to estimate $p_1, \ldots, p_{|\mathcal{L}|}$ and bilinear forms involving the eigenvectors of AA^*. In particular, if p_k has multiplicity one in $\mathrm{diag}(p_1, \ldots, p_L)$, the summations in Item 2. are irrelevant and

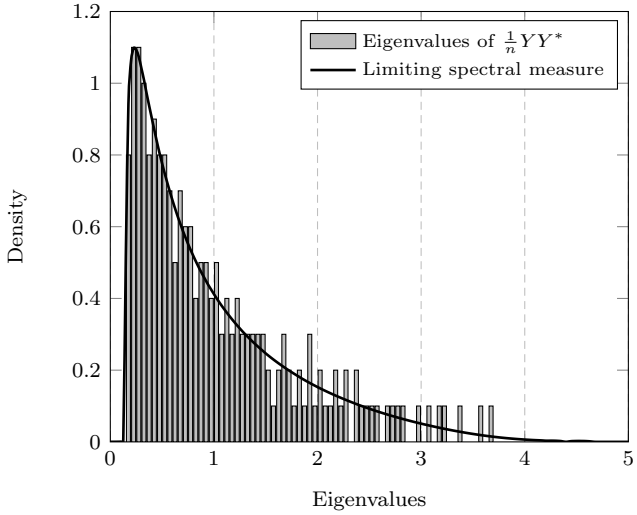

Figure 2: Histogram of the eigenvalues of $\frac{1}{n}YY^*$ against the limiting spectral measure, $L = 2$, $p_1 = p_2 = 1$, $N = 200$, $n = 1000$, Sudent-t impulsions.

we obtain an estimator for $a^* u_k u_k^* b$. These however explicitly rely on ν which, for practical purposes, might be of limited interest if the τ_i are statistically unknown. It turns out, from a careful understanding of γ, that

$$\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{\gamma}_n \triangleq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{N} y_i^* \hat{C}_{(i)}^{-1} y_i \tag{3}$$

and $\hat{C}_{(i)} = \hat{C}_N - \frac{1}{n} u (\frac{1}{N} y_i^* \hat{C}_N^{-1} y_i) y_i y_i^*$. Also, for any $M > 0$,

$$\max_{\substack{1 \le j \le n \\ \tau_j \le M}} |\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0, \quad \max_{\substack{1 \le j \le n \\ \tau_j > M}} \left|1 - \tau_j^{-1} \hat{\tau}_j\right| \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{\tau}_i \triangleq \frac{1}{\hat{\gamma}_n} \frac{1}{N} y_i^* \hat{C}_{(i)}^{-1} y_i. \tag{4}$$

Details of these results are provided in Section 5. Letting $\varepsilon > 0$ small, for $x \in (S^+ + \varepsilon, \infty)$ and for all large n a.s., we then denote $\hat{\delta}(x)$ the unique negative

solution to[3]

$$\hat{\delta}(x) = c_n \left(-x + \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\tau}_i v_c(\hat{\tau}_i \hat{\gamma}_n)}{1 + \hat{\delta}(x) \hat{\tau}_i v_c(\hat{\tau}_i \hat{\gamma}_n)} \right)^{-1}. \tag{5}$$

From this, we then deduce the following alternative set of power and bilinear form estimators.

Theorem 3 (Robust estimation for unknown ν). *With the same notations as in Theorem 2, and with $\hat{\gamma}_n$, $\hat{\tau}_i$, and $\hat{\delta}$ defined in (3)–(5), we have the following results.*

1. Purely empirical power estimation. *For each $j \in \mathcal{L}$,*

$$-\left(\hat{\delta}(\hat{\lambda}_j) \frac{1}{N} \sum_{i=1}^{n} \frac{v(\hat{\tau}_i \hat{\gamma}_n)}{1 + \hat{\delta}(\hat{\lambda}_j) \hat{\tau}_i v(\hat{\tau}_i \hat{\gamma}_n)} \right)^{-1} \xrightarrow{\text{a.s.}} p_j.$$

2. Purely empirical bilinear form estimation. *For each $a, b \in \mathbb{C}^N$ with $\|a\| = \|b\| = 1$, and each $j \in \mathcal{L}$,*

$$\sum_{k, p_k = p_j} a^* u_k u_k^* b - \sum_{k, p_k = p_j} \hat{w}_k a^* \hat{u}_k \hat{u}_k^* b \xrightarrow{\text{a.s.}} 0$$

where

$$\hat{w}_k = \frac{\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \frac{v(\hat{\tau}_i \hat{\gamma}_n)}{\left(1 + \hat{\delta}(\hat{\lambda}_k) \hat{\tau}_i v(\hat{\tau}_i \hat{\gamma}_n)\right)^2}}{\dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \frac{v(\hat{\tau}_i \hat{\gamma}_n)}{1 + \hat{\delta}(\hat{\lambda}_k) \hat{\tau}_i v(\hat{\tau}_i \hat{\gamma}_n)} \left(1 - \frac{1}{N} \displaystyle\sum_{i=1}^{n} \frac{\hat{\delta}(\hat{\lambda}_k)^2 \tau_i^2 v(\hat{\tau}_i \hat{\gamma}_n)^2}{\left(1 + \hat{\delta}(\hat{\lambda}_k) \hat{\tau}_i v(\hat{\tau}_i \hat{\gamma}_n)\right)^2} \right)}.$$

Theorem 3 provides a means to estimate powers and bilinear forms without any statistical knowledge on the τ_i, which are individually estimated. It is interesting to note that, since ν is only a limiting distribution, for practical systems, there is a priori no advantage in using the knowledge of ν or not. In particular, if n is not too large in practice or if ν has heavy tails, it is highly probable that ν_n be quite distinct from ν, leading the estimators in Theorem 1 to be likely less accurate than the estimators in Theorem 2. Conversely, if N is not too large, $\hat{\tau}_i$ may be a weak estimate for τ_i so that, if ν has much lighter tails, the estimators of Theorem 1 may have a better advantage. Theoretical performance comparison between both schemes would require to exhibit central limit theorems for these quantities, which we discuss in Section 4 but goes here beyond the scope of the present work.

[3]Remark here that, since $\hat{\tau}_i$, similar to τ_i, may be found away from $\mathrm{Supp}(\nu)$, $\hat{\delta}(x)$ may not be defined everywhere in (S_μ^+, S^+) but is defined beyond $S^+ + \varepsilon$ for n large a.s.

3.2. Application to angle estimation

An important application of Theorem 1 and Theorem 2 is found in the context of array processing, briefly evoked in the second item of Remark 1, in which $a_i = a(\theta_i)$ for some $\theta_i \in [0, 2\pi)$. For theoretical convenience, we use the classical linear array representation for a_i as follows.

Assumption 4. *For $i \in \{1, \ldots, L\}$, $a_i = a(\theta_i)$ with $\theta_1, \ldots, \theta_L$ distinct and, for $d > 0$ and $\theta \in [0, 2\pi)$,*

$$a(\theta) = N^{-\frac{1}{2}} [\exp(2\pi \imath d j \sin(\theta))]_{j=0}^{N-1}.$$

The objective in this specific model is to estimate $\theta_1, \ldots, \theta_L$ from the observations $y_1, \ldots, y_n$. In the regime $n \gg N$ with non-impulsive noise, this is efficiently performed by the traditional multiple signal classification (MUSIC) algorithm from (Schmidt, 1986). Using the fact that the vectors $a(\theta_i)$, $i \in \{1, \ldots, L\}$, are orthogonal to the subspace spanned by the eigenvectors with eigenvalue 1 of $\mathrm{E}[y_1 y_1^*] = AA^* + I_N$, the algorithm consists in retrieving the deepest minima of the nonnegative localization function $\hat{\eta}$ defined for $\theta \in [0, 2\pi)$ by

$$\hat{\eta}(\theta) = a(\theta)^* \Pi_{\frac{1}{n} YY^*} a(\theta)$$

where $\Pi_{\frac{1}{n} YY^*}$ is a projection matrix on the subspace associated with the $N - L$ smallest eigenvalues of $\frac{1}{n} YY^*$. Indeed, as $\frac{1}{n} YY^*$ is an almost surely consistent estimate for $\mathrm{E}[y_1 y_1^*]$ in the large n regime, $\hat{\eta}(\theta) \xrightarrow{\text{a.s.}} \eta(\theta)$ where

$$\eta(\theta) = a(\theta)^* \Pi_{\mathrm{E}[y_1 y_1^*]} a(\theta)$$

with here $\Pi_{\mathrm{E}[y_1 y_1^*]}$ a projection matrix on the subspace associated with the eigenvalue 1 in $\mathrm{E}[y_1 y_1^*]$; as such, $\hat{\eta}(\theta) \xrightarrow{\text{a.s.}} 0$ for $\theta \in \{\theta_1, \ldots, \theta_L\}$ and to a positive quantity otherwise. In (Mestre, 2008b), Mestre proved that this algorithm is however inconsistent in the regime of Assumption 3. This led to (Mestre, 2008a) in which an improved estimator (the G-MUSIC estimator) for $\theta_1, \ldots, \theta_L$ was designed, however for a more involved model than the spiked model (i.e., L is assumed commensurable with N). In (Loubaton and Vallet, 2010), a spiked model hypothesis was then assumed (i.e., with L small compared to N, n) which unfolded into a more practical and more theoretically tractable spiked G-MUSIC estimator. Similar to MUSIC, the latter consists in determining the deepest minima of an alternative localization function $\hat{\eta}_{\mathrm{G}}(\theta)$, which we shall define in a moment.

Although improved with respect to MUSIC, both algorithms still rely on exploiting the largest isolated eigenvalues of $\frac{1}{n} YY^*$ and the asymptotic boundedness of the noise spectrum. From the discussions in Section 1 and after Theorem 2, under the generic Assumption 1 with τ_i allowed to grow unbounded, these methods are now unreliable and in fact inefficient. From Item 2. in both Theorem 2 and Theorem 3, it is now possible to provide a consistent estimation method based on two novel localization functions $\hat{\eta}_{\mathrm{RG}}$ and $\hat{\eta}_{\mathrm{RG}}^{\mathrm{emp}}$. The resulting algorithms are from now on referred to as robust G-MUSIC and empirical robust G-MUSIC, respectively.

Corollary 1 (Robust G-MUSIC). *Let Assumptions 1–4 hold. Let* $0 < \kappa < \min_{i,j} |\theta_i - \theta_j|$ *and denote* $\mathcal{R}_i^\kappa = [\theta_i - \kappa/2, \theta_i + \kappa/2]$. *Also define* $\hat{\eta}_{\mathrm{RG}}(\theta)$ *and* $\hat{\eta}_{\mathrm{RG}}^{\mathrm{emp}}(\theta)$ *as*

$$\hat{\eta}_{\mathrm{RG}}(\theta) = 1 - \sum_{k=1}^{|\mathcal{L}|} w_k a(\theta)^* \hat{u}_k \hat{u}_k a(\theta)$$
$$\hat{\eta}_{\mathrm{RG}}^{\mathrm{emp}}(\theta) = 1 - \sum_{k=1}^{|\mathcal{L}|} \hat{w}_k a(\theta)^* \hat{u}_k \hat{u}_k a(\theta)$$

where we used the notations from Theorems 2 and 3. Then, for each $j \in \mathcal{L}$,

$$\hat{\theta}_j \xrightarrow{\text{a.s.}} \theta_j$$
$$\hat{\theta}_j^{\mathrm{emp}} \xrightarrow{\text{a.s.}} \theta_j$$

where

$$\hat{\theta}_j \triangleq \mathrm{argmin}_{\theta \in \mathcal{R}_j^\kappa} \{\hat{\eta}_{\mathrm{RG}}(\theta)\}$$
$$\hat{\theta}_j^{\mathrm{emp}} \triangleq \mathrm{argmin}_{\theta \in \mathcal{R}_j^\kappa} \{\hat{\eta}_{\mathrm{RG}}^{\mathrm{emp}}(\theta)\}.$$

With the same reasoning as in Remark 3, it is now easy to check that, letting the v_c or v functions be replaced by the constant 1 in the expressions of w_k and $\hat{w}_k$, respectively, we fall back on G-MUSIC schemes devised in e.g., (Loubaton and Vallet, 2010). In what follows, we then define $\hat{\eta}_{\mathrm{G}}(\theta)$ and $\hat{\eta}_{\mathrm{G}}^{\mathrm{emp}}(\theta)$ similarly to $\hat{\eta}_{\mathrm{RG}}(\theta)$ and $\hat{\eta}_{\mathrm{RG}}^{\mathrm{emp}}(\theta)$ but with the functions v_c and v replaced by the constant 1 and with the couples $(\hat{\lambda}_k, \hat{u}_k)$ replaced by the k-th largest eigenvalue and associated eigenvectors of $\frac{1}{n} YY^*$. For a further comparison of the various methods, we also denote by $\hat{\eta}_{\mathrm{R}}(\theta)$ the robust counterpart to $\hat{\eta}(\theta)$ defined by $\hat{\eta}_{\mathrm{R}}(\theta) = a(\theta)^* \Pi_{\hat{C}_N} a(\theta)$ with $\Pi_{\hat{C}_N}$ a projection matrix on the subspace associated with the $N - L$ smallest eigenvalues of $\hat{C}_N$.

Simulation curves are provided below which compare the performance of the various improved MUSIC techniques. Since the methods based on the extraction of $\delta(\hat{\lambda}_i)$ may be void when this value does not exist, we blindly proceed by solving the fixed-point equation defining $\delta(\hat{\lambda}_i)$ thanks to the standard fixed-point algorithm until convergence or until a maximum number of iterations is reached. This effect is in fact marginal as it is theoretically highly probable that eigenvalues be found beyond S_μ^+ for each finite N, n. We also assume $\mathcal{L} = \{1, \ldots, L\}$ even if this does not hold, which in practice one cannot anticipate. Voluntarily disrupting from the theoretical claims of Theorems 1–3 will allow for an observation of problems arising when the assumptions are not fully satisfied. In all simulation figures, we consider $u(x) = (1 + \alpha)(\alpha + x)^{-1}$ with $\alpha = 0.2$, $N = 20$, $n = 100$, $L = 2$, $\theta_1 = 10°$, $\theta_2 = 12°$. The noise impulsions are of two types: (i) single outlier impulsion for which $\tau_i = 1$, $i \in \{1, \ldots, n-1\}$ and $\tau_n = 100$, or (ii) Student impulsions for which $\tau_i = t^2(\beta - 2)\beta^{-1}$ with t a

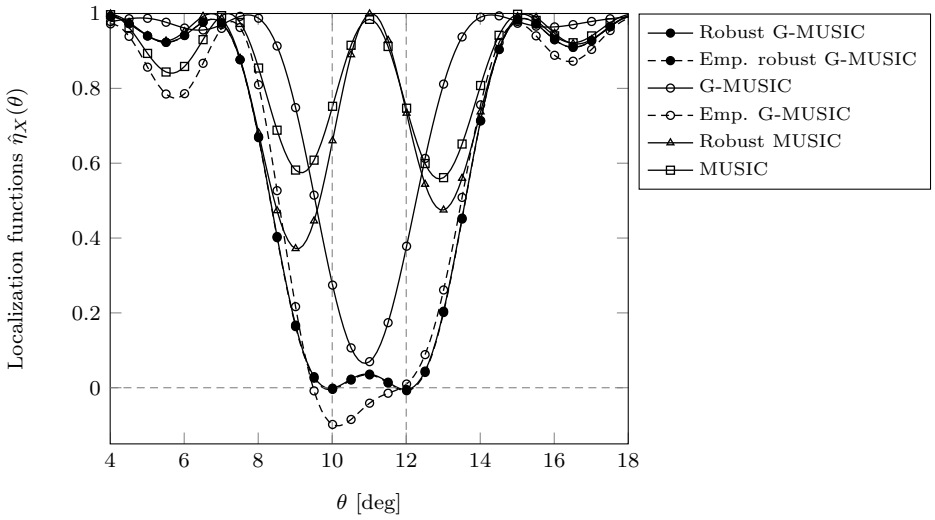

Figure 3: Random realization of the localization functions for the various MUSIC estimators, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 100$, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$. Powers $p_1 = p_2 = 10^{0.5} = 5$ dB.

Student-t random variable with parameter $\beta = 100$ (the normalization ensures $\mathrm{E}[\tau_1] = 1$).

Figure 3 provides a single realization (but representative of the multiple realizations we simulated) of the various localization functions $\hat{\eta}_X$ and $\hat{\eta}_X^{\mathrm{emp}}$ for θ in the vicinity of θ_1, θ_2, X being void, R, or RG. The scenario considered is that of a Student-t noise and $p_1 = p_2 = 1$. The figure confirms the advantage of the methods based on $\hat{C}_N$ over $\frac{1}{n}YY^*$ which unfolds from the proper extreme eigenvalue isolation observed under the same setting in Figure 1 against Figure 2. Due to N/n being non trivial, while the robust G-MUSIC methods accurately discriminate both angles at their precise locations and with appropriate localization function amplitude, the robust MUSIC approach discriminates the two angles at erroneous locations and erroneous localization function amplitude. Benefiting from the random matrix advantage, G-MUSIC in turn behaves better in amplitude than MUSIC but cannot discriminate angles. Observe also here that both empirical and non-empirical robust G-MUSIC approaches behave extremely similar (both curves are visually superimposed), suggesting that with $\beta = 100$ the samples from the Student-t distribution represent sufficiently well the actual distribution of $\tau_1 v(\tau_1 \gamma)$. This no longer holds for G-MUSIC versus empirical G-MUSIC, in which case the approximation of ν_n by the distribution ν of τ_1 is not appropriate.

Figure 4 and Figure 5 provide the mean square error performance for the first angle estimation $\mathrm{E}[|\hat{\theta}_1 - \theta_1|^2]$ as a function of the source powers $p_1 = p_2$; the estimates are based for each estimator on retrieving the local minima of $\hat{\eta}_X$. For fair comparison, the two deepest minima of the localization functions are

extracted and $\hat{\theta}_1$ is declared to be the estimated angle closest to θ_1 (in particular, if a unique minimum is found close to any θ_i, $\hat{\theta}_1$ is attached to this minimum). Figure 4 assumes the Student-t impulsion scenario of Figure 3, while Figure 5 is concerned with the outlier impulsion model previously described. Both figures further confirm the advantage brought by the robust G-MUSIC scheme with asymptotic equivalence between empirical or non-empirical in the large source power regime. We observe in particular the outstanding advantage of (robust or not) G-MUSIC methods which perform well at high source power, while standard methods saturate. Interestingly, from Figure 4, the G-MUSIC schemes perform well in the high source power regime, which corresponds to scenarios in which the noise impulsion amplitudes are often small enough compared to source power to be assumed bounded and G-MUSIC is then consistent. Nonetheless, G-MUSIC never closes the gap with robust G-MUSIC which is likely explained by the much larger spacing between noise and information eigenvalues in the spectrum of $\hat{C}_N$. The situation is different in Figure 5 where G-MUSIC almost meets the performance of robust G-MUSIC at very high power, while performing poorly below 20 dB. This is explained by the presence of a single additional eigenvalue of amplitude around 100 (i.e., 20 dB) in the spectrum of $\frac{1}{n}YY^*$ which corrupts the G-MUSIC algorithm as long as this amplitude is larger than these of the two informative eigenvalues due to the steering vectors (about p_1).

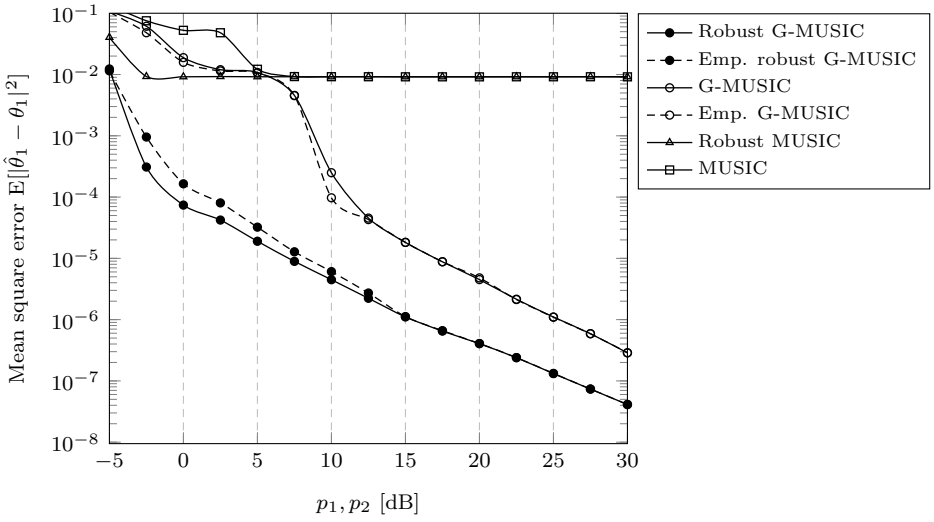

Figure 4: Means square error performance of the estimation of $\theta_1 = 10°$, with $N = 20$, $n = 100$, two sources at 10° and 12°, Student-t impulsions with parameter $\beta = 10$, $u(x) = (1+\alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

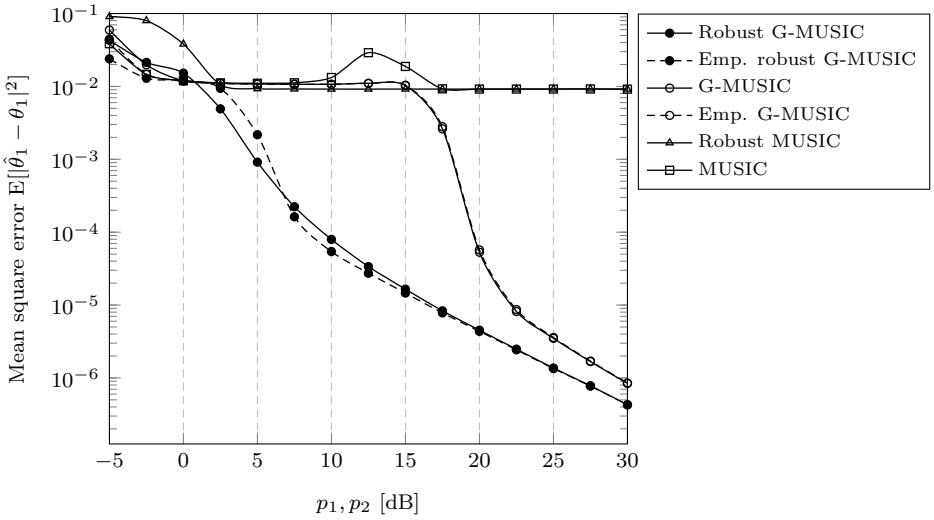

Figure 5: Means square error performance of the estimation of $\theta_1 = 10°$, with $N = 20$, $n = 100$, two sources at 10° and 12°, sample outlier scenario $\tau_i = 1$, $i < n$, $\tau_n = 100$, $u(x) = (1 + \alpha)/(\alpha + x)$ with $\alpha = 0.2$, $p_1 = p_2$.

4. Concluding Remarks

Robust estimators of scatter were originally designed to provide improved covariance (or scatter) matrix estimates of non-Gaussian zero mean random vectors, consistent in the regime $n \to \infty$, which are particularly suited to elliptical samples (Maronna, 1976; Tyler, 1987) or to accommodate for outliers (Huber, 1964). Similar to the more classical sample covariance matrix, the large n consistency however falls short when the population size N is large as well. Random matrix methods allows one to restore consistency in this regime by providing alternative estimation methods of spectral properties of the population covariance or scatter matrices. This is the result of a two-step method: (i) the analysis of the limiting spectrum of the covariance estimators ((Marčenko and Pastur, 1967) for sample covariance matrices and (Couillet et al., 2013b) for robust estimates of scatter) and (ii) the introduction of improved statistical inference methods. For sample covariance matrices, Point (ii) is the result of the works of Girko (Girko, 1987) and more recently Mestre (Mestre, 2008a). The present article provides a first instance of Point (ii) for robust estimators of scatter. The need here for a restriction to a spiked model (while (Girko, 1987; Mestre, 2008a) treat more generic models) is intimately related to the structure of the approximation $\hat{S}_N$ of $\hat{C}_N$ which heavily depends on a non-observable variable γ which may in general be itself an involved function of the parameters to be estimated.

The interest of robust methods is to harness the effect of rare sample outliers, the concatenation of which can be seen as a small rank perturbation matrix of

the data sample matrix. A non obvious outcome of the present study is that, while sample covariance matrices equally treat small rank sample and population perturbations by creating non distinguishable spikes in the spectrum, robust estimates of scatter isolate sample versus population perturbations. This makes it possible to specifically estimate information carried by population perturbations, which is one important consequence of Theorem 2. The practical purpose of this discriminative advantage is obvious and was exemplified by the introduction in Corollary 1 of an improved angle of arrival estimation method which is resilient to sample outliers.

However, since robust estimators of scatter are non unique (Maronna's estimators are defined through u and other estimators such as Tyler's exist), this naturally raises the question of an optimal estimator choice. These questions demand more advanced studies on second order statistics for given performance metrics. Initial investigations are optimistic as they suggest that, on top of $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$, differences of linear spectrum functionals of the type $\int f d\mu_{\hat{C}_N} - \int f d\mu_{\hat{S}_N}$, with μ_X the empirical spectral distribution of X and f a continuous and bounded function, have much weaker fluctuations than each integral around its mean; this indicates that fluctuations of functionals of $\hat{C}_N$ can be studied equivalently through the much more tractable fluctuations of functionals of $\hat{S}_N$.

5. Proof of the main results

5.1. Notations

Throughout the proof, we shall use the following shortcut notations:

$$\begin{aligned} T &= \text{diag}(\{\tau_i\}_{i=1}^n) \in \mathbb{C}^{n\times n} \\ V &= \text{diag}(\{v_c(\tau_i\gamma\}_{i=1}^n) \in \mathbb{C}^{n\times n} \\ S &= [\{s_{ij}\}_{1\le i\le L, 1\le j\le n}] \in \mathbb{C}^{L\times n} \\ W &= [w_1, \ldots, w_n] \in \mathbb{C}^{N\times n} \\ \tilde{W} &= [\tilde{w}_1, \ldots, \tilde{w}_n] \in \mathbb{C}^{N\times n} \end{aligned}$$

with $\tilde{w}_i = w_i r_i/\sqrt{N}$ as in the statement of Theorem 1. We shall expand A as the singular value decomposition $A = U\Omega\bar{U}^*$ with $U \in \mathbb{C}^{N\times L}$ isometric, $\Omega = \text{diag}(\sigma_1, \ldots, \sigma_L)$, $\sigma_1 \ge \ldots \ge \lambda_L \ge 0$, and $\bar{U} \in \mathbb{C}^{L\times L}$ unitary.

We also define

$$\hat{S}_N^\circ = \frac{1}{n}\sum_{i=1}^n \tau_i v_c(\tau_i\gamma)\tilde{w}_i\tilde{w}_i^* = \frac{1}{n}\tilde{W}TV\tilde{W}^*$$

which corresponds to $\hat{S}_N$ with $p_1 = \ldots = p_L = 0$, i.e., with no perturbation, and

$$Q_z^\circ = (\hat{S}_N^\circ - zI_N)^{-1} = \left(\frac{1}{n}\tilde{W}TV\tilde{W}^* - zI_N\right)^{-1}$$

the resolvent of $\hat{S}_N^\circ$.

For couples (η, M_η), $\eta < 1$, such that $\nu((0, M_\eta)) > 0$ and $\nu((M_\eta, \infty)) < \eta$, it will be necessary to define T_η the matrix T in which all values of τ_i greater or equal to M_η are replaced by zeros, and similarly for V_η. Denote also γ^η the unique solution to

$$1 = \int_{\tau < M_\eta} \frac{\psi_c(\tau\gamma^\eta)\nu(d\tau)}{1 + c\psi_c(\tau\gamma^\eta)}. \tag{6}$$

and $\hat{S}_{N,\eta}$ the resulting $\hat{S}_N$ matrix with all τ_i greater than M_η discarded and γ replaced by γ^η.

Finally, we further define $T_{(j)} = \mathrm{diag}(\{\tau_i\}_{i \neq j})$ and similarly for $V_{(j)}$, $S_{(j)}$, $\tilde{W}_{(j)}$, $\hat{S}_{(j)} = \hat{S}_{N,(j)}$ the matrices with column or component j discarded, as well as $T_{(j),\eta}$ the matrix T_η with row-and-column j discarded, and similarly $V_{(j),\eta}$, $S_{(j),\eta}$, $\tilde{W}_{(j),\eta}$, $\hat{S}_{(j),\eta}$ the corresponding matrices with column or component j discarded.

5.2. Overall proof strategy

The existence and uniqueness of $\hat{C}_N$ as defined in the statement of Theorem 1 follows immediately from (**?**, Theorem 2.2) (which is more general than (Couillet et al., 2013b, Theorem 1) established by the author in a previous article). One of the key elements of the proof of convergence in Theorem 1 is to ensure that there exists $\varepsilon > 0$ such that, for all large n a.s., all eigenvalues of $\{\hat{S}_{(j)}, 1 \leq j \leq n\}$ (and also of $\{\hat{S}_{(j),\eta}, 1 \leq j \leq n\}$ for given η small) are greater than ε. This is an important condition to ensure that the quadratic forms $\frac{1}{N}\tilde{w}_j^* \hat{S}_{(j)}^{-1} \tilde{w}_j$, which play a central role in the proof, are jointly controllable. In (Couillet et al., 2013b), where the convergence $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ is obtained for $p_1 = \ldots = p_L = 0$, this unfolded readily from (Couillet et al., 2013a, Lemma 2) (i.e., (Couillet et al., 2013a, Lemma 2) states that the matrices $\frac{1}{n}\tilde{W}_{(j)}\tilde{W}_{(j)}^*$ have their smallest eigenvalue uniformly away from zero). Here, due to the existence of a small rank matrix A, the approach from (Couillet et al., 2013a, Lemma 2) no longer holds as $\hat{S}_{(j)}$ may a priori exhibit finitely many isolated eigenvalues getting close to zero as $n \to \infty$. We shall show that this is not possible. Precisely, we shall prove that the large n spectrum of $\hat{S}_N$ is similar to that of $\hat{S}_N^\circ$ but possibly for finitely many isolated eigenvalues, none of which can be asymptotically found close to zero. We shall however characterize those eigenvalues of $\hat{S}_N$ found beyond the right-edge of the limiting spectrum of $\hat{S}_N^\circ$. Once this result is obtained, to complete the proof of Theorem 1, it will then suffice to check that most spectral statistics involved in the proof of (Couillet et al., 2013b, Theorem 2) are not affected by the presence of the additional small rank matrix AS in the model. Since most results need be proved jointly for the matrix sets $\{\hat{S}_{(j)}, 1 \leq j \leq n\}$ (or $\{\hat{S}_{(j),\eta}, 1 \leq j \leq n\}$), high order moment bounds will be required to then apply union bound along with Markov inequality techniques. As the proof in (Couillet et al., 2013b) is rather long and technical and since the main contribution of

the present article lies in Theorem 2, we only discuss in what follows the main new technical elements that differ from (Couillet et al., 2013b).

When Theorem 1 is obtained, the proofs of Theorems 2 and 3 unfolds from classical techniques for spiked random matrix models, using the approximation $\hat{S}_N$ for $\hat{C}_N$. The model $\hat{S}_N$ considered here is closely related to the scenario of (Chapon et al., 2012), but for the random non-Gaussian structure of the matrix S; also, (Chapon et al., 2012) imposes $\max_i \mathrm{dist}(\tau_i, \mathrm{Supp}(\nu)) \to 0$ which we do not enforce here.

5.3. Localization of the eigenvalues of $\hat{S}_N$ and $\hat{S}_{(i)}$

We first study the localization of the eigenvalues of $\hat{S}_N$ and $\{\hat{S}_{(j),\eta}, 1 \leq j \leq n\}$. The strategy being the same, we concentrate mostly on the study of $\hat{S}_N$ and then briefly generalize the approach to $\{\hat{S}_{(j),\eta}, 1 \leq j \leq n\}$.

By isolating the small rank perturbation terms, we first develop $\hat{S}_N$ as

$$\hat{S}_N = \frac{1}{n}\sum_{i=1}^{n} v_c(\tau_i\gamma) A_i \bar{w}_i \bar{w}_i^* A_i^* \tag{7}$$

$$= \hat{S}_N^\circ + \frac{1}{n} ASVS^*A^* + \frac{1}{n} AST^{\frac{1}{2}} V\tilde{W}^* + \frac{1}{n}\tilde{W}T^{\frac{1}{2}}VS^*A^*. \tag{8}$$

Let $\lambda \in \mathbb{R} \setminus [\varepsilon, S^+ + \varepsilon]$ for some $\varepsilon > 0$ small be an eigenvalue of $\hat{S}_N$. Note that such a λ may not exist. However, from (Bai and Silverstein, 1998) and since in particular $\limsup_n \|AA^*\| < \infty$ and $\limsup_n \|T^{\frac{1}{2}}V\| < \infty$, the spectral norm of each matrix above is asymptotically bounded almost surely and thus $\limsup_n \lambda < \infty$ a.s. Also, from (Couillet et al., 2013b) and from the discussion prior to the statement of Theorem 1, for all large n a.s., λ is not an eigenvalue of $\hat{S}_N^\circ$ (for ε chosen small enough). Thus, by definition, λ is a solution of $\det(\hat{S}_N - \lambda I_N) = 0$ while $\|(\hat{S}_N^\circ - \lambda I_N)^{-1}\| < M$ for some $M > 0$ independent of n but increasing as $\varepsilon \to 0$. As such, from the development above, for all large n a.s.,

$$0 = \det\left(\hat{S}_N^\circ - \lambda I_N + \Gamma\right) = \det\left(Q_\lambda^\circ\right)^{-1} \det\left(I_N + (Q_\lambda^\circ)^{\frac{1}{2}} \Gamma (Q_\lambda^\circ)^{\frac{1}{2}}\right)$$

where $\Gamma = \frac{1}{n} ASVS^*A^* + \frac{1}{n} AST^{\frac{1}{2}} V\tilde{W}^* + \frac{1}{n}\tilde{W}T^{\frac{1}{2}}VS^*A^*$ can be further written

$$\Gamma = \begin{bmatrix} U\Omega^{\frac{1}{2}} & \frac{1}{n}\tilde{W}T^{\frac{1}{2}}VS^*\bar{U}\Omega^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Omega^{\frac{1}{2}}\bar{U}^*\frac{1}{n}\tilde{W}V\tilde{W}^*\bar{U}\Omega^{\frac{1}{2}} & I_L \\ I_L & 0 \end{bmatrix} \begin{bmatrix} \Omega^{\frac{1}{2}}U^* \\ \Omega^{\frac{1}{2}}\frac{1}{n}\bar{U}^*ST^{\frac{1}{2}}V\tilde{W} \end{bmatrix}. \tag{9}$$

Exploiting the small rank of S and A, and the formula $\det(I+AB) = \det(I+BA)$ for properly sized A, B matrices, this induces

$$0 = \det\left(I_{2L} + \Gamma_L(\lambda)\right)$$

where

$$\Gamma_L(\lambda) \triangleq \begin{bmatrix} \Omega^{\frac{1}{2}} \bar{U}^* \frac{1}{n} \tilde{W} V \tilde{W}^* \bar{U} \Omega^{\frac{1}{2}} & I_L \\ I_L & 0 \end{bmatrix} \begin{bmatrix} \Omega^{\frac{1}{2}} U^* \\ \Omega^{\frac{1}{2}} \frac{1}{n} \bar{U}^* S T^{\frac{1}{2}} V \tilde{W} \end{bmatrix} Q_\lambda^\circ \begin{bmatrix} U \Omega^{\frac{1}{2}} & \frac{1}{n} \tilde{W} T^{\frac{1}{2}} V S^* \bar{U} \Omega^{\frac{1}{2}} \end{bmatrix}.$$

We now need the following central lemmas.

Lemma 1. *Let $\varepsilon > 0$ and $\mathcal{A}_\varepsilon$ be the event $\varepsilon < \lambda_N(\hat{S}_N^\circ) < \lambda_1(\hat{S}_N^\circ) < S^+ + \varepsilon$. Let also $a, b \in \mathbb{C}^N$ be two vectors of unit norm. Then, for every $z \in \mathcal{C} \subset \mathbb{C} \backslash [\varepsilon, S^+ + \varepsilon]$ with $\mathcal{C}$ compact,*

$$\mathrm{E}\left[\left| \frac{1}{n} S^* V S - \frac{1}{n} \operatorname{tr} V \right|^p \right] \leq K N^{-\frac{p}{2}}$$

$$\mathrm{E}\left[1_{\mathcal{A}_\varepsilon} \left| \frac{1}{n} S T^{\frac{1}{2}} V \frac{1}{n} \tilde{W}^* Q_z^\circ \tilde{W} V T^{\frac{1}{2}} S^* - \left[\frac{1}{n} \operatorname{tr} V + z \frac{1}{n} \operatorname{tr} V \tilde{Q}_z^\circ \right] \right|^p \right] \leq K N^{-\frac{p}{2}}$$

$$\mathrm{E}\left[1_{\mathcal{A}_\varepsilon} \left| a^* Q_z^\circ b - a^* b \frac{1}{N} \operatorname{tr} Q_z^\circ \right|^p \right] \leq K N^{-\frac{p}{2}}$$

$$\mathrm{E}\left[1_{\mathcal{A}_\varepsilon} \left\| \frac{1}{n} a^* Q_z^\circ \tilde{W} T^{\frac{1}{2}} V S^* \right\|^p \right] \leq K N^{-\frac{p}{2}}$$

where $\tilde{Q}_z^\circ = (\frac{1}{n} T^{\frac{1}{2}} V^{\frac{1}{2}} \tilde{W}^* \tilde{W} V^{\frac{1}{2}} T^{\frac{1}{2}} - z I_N)^{-1}$ and $K > 0$ does not depend on z.

Proof. The first convergence is a mere application of (Bai and Silverstein, 2009, Lemma B.26). Similarly, noticing that

$$\begin{aligned} \frac{1}{n} S T^{\frac{1}{2}} V \frac{1}{n} \tilde{W}^* Q_z^\circ \tilde{W} V T^{\frac{1}{2}} S^* &= \frac{1}{n} S V^{\frac{1}{2}} \left[T^{\frac{1}{2}} V^{\frac{1}{2}} \frac{1}{n} \tilde{W}^* \tilde{W} V^{\frac{1}{2}} T^{\frac{1}{2}} \tilde{Q}_z^\circ \right] V^{\frac{1}{2}} S^* \\ &= \frac{1}{n} S V S^* + z \frac{1}{n} S \tilde{Q}_z^\circ V^{\frac{1}{2}} S^* \end{aligned}$$

the second result follows again by (Bai and Silverstein, 2009, Lemma B.26) and the fact that $\limsup_n \|\tilde{Q}_z^\circ\| < 1/\operatorname{dist}(\mathcal{C}, [\varepsilon, S^+ + \varepsilon])$. Using the fact that $\tilde{W}$ is Gaussian, the third result follows from the same proof as in (Loubaton and Vallet, 2010, Lemma 3) using additionally $[VT]_{ii} < \psi_\infty$. Similarly, conditioning first on S, which is independent of $\tilde{W}$, we obtain by the same proof as in (Loubaton and Vallet, 2010, Lemma 4) that

$$\mathrm{E}_{\tilde{W}}\left[1_{\mathcal{A}_\varepsilon} \left| \frac{1}{n} a^* Q_z^\circ \tilde{W} T^{\frac{1}{2}} V s_i \right|^p \right] \leq K \|n^{-\frac{1}{2}} s_i\|^p N^{-\frac{p}{2}}$$

where we denoted $S^* = [s_1, \ldots, s_L]$ (the proof follows from exploiting the left-unitary invariance of $\tilde{W}$ and applying the integration by parts and Poincaré–Nash inequality method for unitary Haar matrices described in (Pastur and Ŝerbina, 2011, Chapter 8)). Now, $\mathrm{E}[\|n^{-\frac{1}{2}} s_i\|^p] = O(1)$ by Hölder's inequality, and we obtain the last inequality.

Lemma 2. *For $z \in \mathbb{C} \setminus [S_\mu^-, S_\mu^+]$, let $\delta(z)$ be the unique solution to the equation*

$$\delta(z) = c\left(-z + \int \frac{tv_c(t\gamma)}{1+\delta(z)tv_c(t\gamma)} d\nu(t)\right)^{-1}$$

where we recall that γ is the unique positive solution to

$$1 = \int \frac{\psi_c(t\gamma)}{1+c\psi_c(t\gamma)} d\nu(t).$$

Let now $z \in \mathcal{C}$, with $\mathcal{C}$ a compact set of $\mathbb{C} \setminus [\varepsilon, S_\mu^+ + \varepsilon]$ for some $\varepsilon > 0$ small enough. Then, denoting $\Psi_z^\circ = (I_n + \delta(z)VT)^{-1}$,

$$\sup_{z\in\mathcal{C}} \left| \frac{1}{N} \operatorname{tr} Q_z^\circ - \frac{\delta(z)}{c} \right| \xrightarrow{\text{a.s.}} 0$$

$$\sup_{z\in\mathcal{C}} \left| \frac{1}{n} \operatorname{tr} V + z\frac{1}{n} \operatorname{tr} V\tilde{Q}_z^\circ - \delta(z)\frac{1}{n} \operatorname{tr} V^2 T\Psi_z^\circ \right| \xrightarrow{\text{a.s.}} 0.$$

Proof. The almost sure convergences to zero of the terms inside the norms (i.e., for each $z \in \mathcal{C}$) are classical, see e.g., (Silverstein and Bai, 1995). Considering a countable sequence $z_1, z_2, \ldots$ of such $z \in \mathcal{C}$ having an accumulation point, by the union bound, there exists a probability one set on which the convergence is valid for each point of the sequence. Now, by (Couillet et al., 2013b), for all large n a.s., Q_z° and $\tilde{Q}_z^\circ$ are analytic on $\mathcal{C}$. Since $\delta(z)$ is also analytic on $\mathcal{C}$, by Vitali's convergence theorem (Titchmarsh, 1939), the convergences are uniform on $\mathcal{C}$.

From (Couillet et al., 2013b) again, for $\varepsilon > 0$ small enough, the set $\mathcal{A}_\varepsilon$ introduced in Lemma 1 satisfies $1_{\mathcal{A}_\varepsilon} \xrightarrow{\text{a.s.}} 1$. As such, using the Markov inequality and the Borel Cantelli lemma, Lemma 1 for $p > 2$ ensures that all quantities in absolute values in the statement of Lemma 1 converge to zero almost surely as $n \to \infty$. Since the quantities involved are analytic on compact $\mathcal{C} \subset \mathbb{C}\setminus[\varepsilon, S^+ + \varepsilon]$, considering a countable sequence of $z \in \mathcal{C}$ having a limit point, it is clear by Vitali's convergence theorem (Titchmarsh, 1939) that these convergences are uniform on $\mathcal{C}$. Applying successively Lemma 1 for $p > 2$ and Lemma 2, we then obtain, for $\mathcal{C} \subset \mathbb{C} \setminus [\varepsilon, S^+ + \varepsilon]$,

$$\sup_{z\in\mathcal{C}} \left\{ \left\| \Gamma_L(z) - \begin{bmatrix} \Omega\frac{1}{n}\operatorname{tr} V & I_L \\ I_L & 0 \end{bmatrix} \begin{bmatrix} \Omega\frac{\delta(z)}{c} & 0 \\ 0 & \Omega\delta(z)\frac{1}{n}\operatorname{tr} V^2 T\Psi_z^\circ \end{bmatrix} \right\| \right\} \xrightarrow{\text{a.s.}} 0$$

or equivalently

$$\sup_{z\in\mathcal{C}} \left\{ \left\| \Gamma_L(z) - \begin{bmatrix} \Omega^2\frac{\delta(z)}{c}\frac{1}{n}\operatorname{tr} V & \Omega\delta(z)\frac{1}{n}\operatorname{tr} V^2 T\Psi_z^\circ \\ \Omega\frac{\delta(z)}{c} & 0 \end{bmatrix} \right\| \right\} \xrightarrow{\text{a.s.}} 0. \tag{10}$$

We may then particularize this result to $z = \lambda$ which, for ε sufficiently small, remains bounded away from $[\varepsilon, S^+ + \varepsilon]$ as n grows (but of course depends on

n) to obtain

$$\left\| \Gamma_L(\lambda) - \begin{bmatrix} \Omega^2 \frac{\delta(\lambda)}{c} \frac{1}{n} \operatorname{tr} V & \Omega \delta(\lambda) \frac{1}{n} \operatorname{tr} V^2 T \Psi_\lambda^\circ \\ \Omega \frac{\delta(\lambda)}{c} & 0 \end{bmatrix} \right\| \xrightarrow{\text{a.s.}} 0. \tag{11}$$

For $\bar{\lambda} \in \mathbb{R} \setminus [\varepsilon, S^+ + \varepsilon]$, let us now study the equation

$$\det \left(I_{2L} + \begin{bmatrix} \Omega^2 \frac{\delta(\bar{\lambda})}{c} \frac{1}{n} \operatorname{tr} V & \Omega \delta(\bar{\lambda}) \frac{1}{n} \operatorname{tr} V^2 T \Psi_{\bar{\lambda}}^\circ \\ \Omega \frac{\delta(\bar{\lambda})}{c} & 0 \end{bmatrix} \right) = 0. \tag{12}$$

After development of the determinant, this equation is equivalent to

$$\sigma_\ell^2 \frac{\delta(\bar{\lambda})}{c} \left(\frac{1}{n} \operatorname{tr} V - \delta(\bar{\lambda}) \frac{1}{n} \operatorname{tr} V^2 T \Psi_{\bar{\lambda}}^\circ \right) + 1 = 0$$

for some $\ell \in \{1, \ldots, L\}$, or equivalently, using $V - \delta(\bar{\lambda}) V^2 T \Psi_{\bar{\lambda}}^\circ = V \Psi_{\bar{\lambda}}^\circ$

$$\sigma_\ell^2 \delta(\bar{\lambda}) \frac{1}{N} \sum_{i=1}^n \frac{v_c(\tau_i \gamma)}{1 + \tau_i v_c(\tau_i \gamma) \delta(\bar{\lambda})} = 0.$$

In the limit $n \to \infty$, using $A^* A \xrightarrow{\text{a.s.}} \operatorname{diag}(p_1, \ldots, p_L)$ and $\frac{1}{n} \sum_{i=1}^n \boldsymbol{\delta}_{\tau_i} \to \nu$ a.s., any accumulation point $\bar{\Lambda} \in (\mathbb{R} \setminus (\varepsilon, S^+ + \varepsilon)) \cup \{\infty\}$ of $\bar{\lambda}$ must satisfy

$$1 + p_\ell \frac{1}{c} \int \frac{\delta(\bar{\Lambda}) v_c(\tau\gamma)}{1 + \delta(\bar{\Lambda}) \tau v_c(\tau\gamma)} \nu(dt) = 0. \tag{13}$$

This unfolds from dominated convergence, using $\delta((S^+, \infty)) \subset (-(\tau_+ v(\tau_+ \gamma))^{-1}, 0)$ with $\tau_+ \in (0, \infty]$ the right-edge of the support of ν; in particular, if $\operatorname{Supp}(\nu)$ is unbounded, $\delta((S^+, \infty)) \subset (-\gamma/\psi_\infty, 0)$ (Couillet and Hachem, 2013). Let us then consider the equation in the variable $\Lambda \in (S^+, \infty)$

$$-\left(\frac{1}{c} \int \frac{\delta(\Lambda) v_c(\tau\gamma)}{1 + \delta(\Lambda) \tau v_c(\tau\gamma)} \nu(d\tau) \right)^{-1} = p_\ell. \tag{14}$$

We know from (Couillet et al., 2013b) that, since $\nu([0, m)) < 1 - \phi_\infty^{-1}$ for some $m > 0$ (by Assumption 2), $S_\mu^- > 0$. Also, as the Stieltjes transform of a measure with support included in $[S_\mu^-, S_\mu^+] \subset [S_\mu^-, S^+]$, δ is increasing on both $[0, S_\mu^-)$ and (S^+, ∞). Moreover, $\delta([0, S_\mu^-)) \subset (0, \infty)$ and $\delta((S^+, \infty)) \subset (-(\tau_+ v(\tau_+ \gamma))^{-1}, 0)$. Therefore, the left-hand side of (14) is negative for $\Lambda \in [0, S_\mu^-)$ and the equation has no solution in this set. It is now easily seen that the left-hand side of (14) is increasing with Λ with limits infinity as $\Lambda \to \infty$ and $p_- > 0$ as $\Lambda \downarrow S^+$. Therefore, if $p_- < p_\ell$, the above equation has a unique solution $\Lambda_\ell \in (S^+, \infty)$, distinct for each distinct p_ℓ. Hence, $\bar{\lambda} \to \bar{\Lambda} = \Lambda_\ell$.

By the argument principal, for all n large a.s., the number of eigenvalues of $\hat{S}_N$, i.e., the number of zeros of $\det(I_{2L} + \Gamma_L(\lambda))$, in any open set $\mathcal{V} \subset \mathbb{R} \setminus [\varepsilon, S^+ + \varepsilon]$ is

$$\frac{1}{2\pi\imath} \oint_\jmath \frac{[\det(I_{2L} + \Gamma_L(z))]'}{\det(I_{2L} + \Gamma_L(z))} dz$$

with $\mathcal{I}$ a contour enclosing $\mathcal{V}$. By the uniform convergence of (10) on $\mathcal{V}$, the analyticity of the quantities involved, and the fact that the involved determinant is a polynomial of order at most $2L$ of its entries, this value asymptotically corresponds to the number of solutions to (12) in $\mathcal{V}$ counted with multiplicity, which in the limit are the $\Omega_k \in \mathcal{V}$. Particularizing $\mathcal{V}$ to $(-1, 2\varepsilon)$ for $\varepsilon > 0$ small enough and then to any small open ball around Λ_ℓ for each ℓ such that $p_\ell > p_-$, we then conclude that $\hat{S}_N$ has asymptotically no eigenvalue in $[0, \varepsilon]$ but that $\lambda_\ell(\hat{S}_N) \xrightarrow{\text{a.s.}} \Lambda_\ell$ for all $\ell \in \mathcal{L}$, which is the expected result.

The precise localization of the eigenvalues of $\hat{S}_N$ will be fundamental for the proof of Theorems 2 and 3. To prove Theorem 1 though, we need to generalize part of this result to the matrices $\hat{S}_{(j)}$ and $\hat{S}_{(j),\eta}$ defined at the beginning of the section. Precisely, we need to show that there exists $\varepsilon > 0$ such that $\min_{1 \le j \le N}\{\lambda_N(\hat{S}_{(j)})\} > \varepsilon$ for all large n a.s., and similarly for $\hat{S}_{(j),\eta}$.

Take $j \in \{1, \ldots, n\}$. Replacing $\hat{S}_N$ by $\hat{S}_{(j)}$ in the proof above leads to the same conclusions. Indeed, by a rank-one perturbation argument (Silverstein and Bai, 1995, Lemma 2.6), for each $\varepsilon > 0$, for all large n a.s.

$$\frac{1}{n}\operatorname{tr}\tilde{Q}_z^\circ - \frac{1}{n}\operatorname{tr}\left(\frac{1}{n}\tilde{W}_{(j)}T_{(j)}V_{(j)}\tilde{W}_{(j)} - zI_N\right)^{-1} \le \frac{1}{n}\frac{1}{\operatorname{dist}(z, [\varepsilon, S^+ + \varepsilon])}$$

and therefore, up to replacing all matrices X by $X_{(j)}$ in their statements, Lemmas 1 and 2 hold identically (with $\delta(z)$ unchanged). Exploiting $\frac{1}{n-1}\sum_{i \neq j} \boldsymbol{\delta}_{\tau_i} \to \nu$ a.s., the remainder of the proof unfolds all the same and we have in particular that for all large n a.s. $\hat{S}_{(j)}$ has no eigenvalue below some $\varepsilon > 0$.

We now prove that this result can be made uniform across j. Denote $\Gamma_{L,(j)}(z)$ the matrix $\Gamma_L(z)$ with all matrices X replaced by $X_{(j)}$. Also rename Lemmas 1 and 2 respectively Lemma 1-(j) and Lemma 2-(j), and rename $\mathcal{A}_\varepsilon$ by $\mathcal{A}_{\varepsilon,(j)}$ in the statement of Lemma 1-(j). Then, taking $p > 4$ in Lemma 1-(j), by the union bound and the Markov inequality, for $e > 0$,

$$\begin{aligned}
&P\left(\max_{1 \le j \le n} 1_{A_{\varepsilon,(j)}} \left\| \Gamma_{L,(j)}(z) - \begin{bmatrix} \Omega^2 \frac{\delta(z)}{c}\frac{1}{n}\operatorname{tr} V & \Omega\delta(z)\frac{1}{n}\operatorname{tr} V^2 T \Psi_z^\circ \\ \Omega\frac{\delta(z)}{c} & 0 \end{bmatrix} \right\| > e \right) \\
&\le \frac{1}{e^p}\sum_{j=1}^n \mathrm{E}\left[1_{A_{\varepsilon,(j)}} \left\| \Gamma_{L,(j)}(z) - \begin{bmatrix} \Omega^2 \frac{\delta(z)}{c}\frac{1}{n}\operatorname{tr} V & \Omega\delta(z)\frac{1}{n}\operatorname{tr} V^2 T \Psi_z^\circ \\ \Omega\frac{\delta(z)}{c} & 0 \end{bmatrix} \right\|^p \right] \\
&= O(N^{1-\frac{p}{2}})
\end{aligned}$$

which is summable. By the Borel Cantelli lemma, the event in the probability parentheses then converges a.s. to zero. Finally, from (Couillet et al., 2013a), there exists $\varepsilon > 0$ such that $1_{\cap_{j=1}^n A_{\varepsilon,(j)}} \xrightarrow{\text{a.s.}} 1$. We then conclude that, for each $z \in \mathcal{C} \subset \mathbb{C} \setminus [\varepsilon, S_\mu^+ + \varepsilon]$ for some $\varepsilon > 0$,

$$\sup_{1 \le j \le n} \left\| \Gamma_{L,(j)}(z) - \begin{bmatrix} \Omega^2 \frac{\delta(z)}{c}\frac{1}{n}\operatorname{tr} V & \Omega\delta(z)\frac{1}{n}\operatorname{tr} V^2 T \Psi_z^\circ \\ \Omega\frac{\delta(z)}{c} & 0 \end{bmatrix} \right\| \xrightarrow{\text{a.s.}} 0$$

Let now $\mathcal{V} \subset \mathbb{C} \setminus [\varepsilon, S_\mu^+ + \varepsilon]$ be a bounded open set containing $[0, \varepsilon/2]$ and $\mathcal{I}$ be its smooth boundary. Taking the determinant of each matrix inside the norm and using again the analyticity of the functions involved, we now get that the quantity

$$\frac{1}{2\pi\imath} \oint_{\mathcal{I}} \frac{[\det(I_{2L} + \Gamma_{L,(j)}(z))]'}{\det(I_{2L} + \Gamma_{L,(j)}(z))} dz$$

converges almost surely uniformly across $j \in \{1, \ldots, n\}$ to the number of eigenvalues of any of the $\hat{S}_{(j)}$ within $[0, \varepsilon/2]$. But by the previous proof, this must be zero. Hence, for all large n a.s., none of the $\hat{S}_{N,(j)}$ has eigenvalues smaller than $\varepsilon/2$, which is what we wanted.

Let now (η, M_η) be such that $\nu((0, M_\eta)) > 0$ and $\nu((M_\eta, \infty)) < \eta$. We have now $\frac{1}{n}\sum_{i=1}^n 1_{\tau_i \leq M_\eta} \boldsymbol{\delta}_{\tau_i} \xrightarrow{\text{a.s.}} \nu\eta \triangleq c_\eta \nu + (1 - c_\eta)\boldsymbol{\delta}_0$ with $c_\eta = \lim_n n^{-1}|\{\tau_i \leq M_\eta\}| = 1 - \eta$ (which almost surely exists by the law of large numbers), so that $\nu_\eta([0, m)) < \eta + (1 - \eta)(1 - \phi_\infty^{-1})$ for some $m > 0$ (Assumption 2). Taking η small enough so that $\nu_\eta([0, m)) < 1 - \phi_\infty^{-1}$, we are still under the assumptions of (Couillet et al., 2013b, Theorem 2) and therefore we again have that for all large n a.s. none of the matrices $\hat{S}_{(j),\eta}$ has eigenvalues below a certain positive value $\varepsilon_\eta > 0$.

These elements are sufficient to now turn to the proof of the main theorems.

5.4. Proof of Theorem 1

When $p_1 = \ldots = p_L = 0$, Theorem 1 unfolds directly from (Couillet et al., 2013b, Theorem 2). Indeed, in this scenario, the latter result states

$$\left\| \hat{C}_N - \frac{1}{n}\sum_{i=1}^n v(\tau_i \gamma_N) w_i w_i^* \right\| \xrightarrow{\text{a.s.}} 0 \tag{15}$$

with γ_N the unique positive solution to

$$1 = \frac{1}{n}\sum_{i=1}^n \frac{\psi(\tau_i \gamma_N)}{1 + c_n \psi(\tau_i \gamma_N)}.$$

Using $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\tau_i} \xrightarrow{\text{a.s.}} \nu$, $c_n \to c$, along with the boundedness of ψ, we have that any accumulation point $\gamma \in [0, \infty]$ of γ_N as $n \to \infty$ must satisfy

$$1 = \int \frac{\psi_c(\tau\gamma)\nu(d\tau)}{1 + c\psi_c(\tau\gamma)}$$

the solution of which is easily shown to be unique in $(0, \infty)$ as the right-hand side term is increasing in γ with limits zero as $\gamma \to 0$ and $\psi_\infty > 1$ as $\gamma \to \infty$ (unless $\nu = \boldsymbol{\delta}_0$ which is excluded). Using the continuity and boundedness of v, it then comes $\max_i |v(\tau_i \gamma_N) - v_c(\tau_i \gamma)| \xrightarrow{\text{a.s.}} 0$. Now, $w_i w_i^* = (w_i w_i^* r_i^2/N)/(r_i^2/N)$

where in the numerator $w_i r_i/\sqrt{N}$ is Gaussian and where the denominator satisfies $\max_i |r_i^2/N - 1| \xrightarrow{\text{a.s.}} 0$ (using classical probability bounds on the chi-square distribution). With these results, along with (Bai and Silverstein, 1998) which ensures that $\frac{1}{nN}\sum_i w_i w_i^* r_i^2$ has bounded spectral norm for all large n a.s., (Couillet et al., 2013b, Theorem 2) implies

$$\left\| \hat{C}_N - \frac{1}{nN} \sum_{i=1}^{n} v_c(\tau_i \gamma) w_i w_i^* r_i^2 \right\| \xrightarrow{\text{a.s.}} 0$$

which is the desired result for $p_1 = \ldots = p_L = 0$.

The generalization to generic $p_1, \ldots, p_L$ follows from a careful control of the elements of proof of (Couillet et al., 2013b, Theorem 2). We see that (Couillet et al., 2013b, Lemma 1) and (Couillet et al., 2013b, Remark 1) are not affected by $p_1, \ldots, p_L$ as these results only depend on $\tau_1, \ldots, \tau_n$. The fundamental lemma (Couillet et al., 2013b, Lemma 2) (and its extension remark (Couillet et al., 2013b, Remark 2)) as well as the lemma (Couillet et al., 2013b, Lemma 3) however need be updated.

We shall not go into the details of every generalization which is painstaking and in fact similar for each lemma. Instead, we detail the generalization of the important remark (Couillet et al., 2013b, Remark 2) and merely give elements for the other results. The remark (Couillet et al., 2013b, Remark 2) is now updated as follows.

Lemma 3. *Let (η, M_η) be couples indexed by $\eta \in (0,1)$ such that $\nu((0, M_\eta)) > 0$ and $\nu((M_\eta, \infty)) < \eta$ and define γ^η as the unique solution to (6). Also let $M > 0$ be arbitrary. Then, for all η small enough,*

$$\max_{\substack{1 \le j \le n \\ \tau_j \le M}} \left| \frac{1}{N} y_j^* \left(\frac{1}{n} \sum_{\tau_i \le M_\eta, i \ne j} v\left(\tau_i \gamma^\eta\right) y_i y_i^* \right)^{-1} y_j - \tau_j \gamma^\eta \right| \xrightarrow{\text{a.s.}} 0$$

$$\max_{\substack{1 \le j \le n \\ \tau_j > M}} \left| \frac{1}{\tau_j} \frac{1}{N} y_j^* \left(\frac{1}{n} \sum_{\tau_i \le M_\eta, i \ne j} v\left(\tau_i \gamma^\eta\right) y_i y_i^* \right)^{-1} y_j - \gamma^\eta \right| \xrightarrow{\text{a.s.}} 0. \tag{16}$$

Proof. Note that, replacing the terms y_i by $\tau_i w_i$ in (16) gives exactly (Couillet et al., 2013b, Remark 2). To ensure that the result holds, we then only need verify that the terms involving AS become negligible.

For η sufficiently small, define

$$\check{S}_{(j),\eta} = \frac{1}{n} \sum_{\tau_i \le M_\eta, i \ne j} v\left(\tau_i \gamma^\eta\right) y_i y_i^* = \frac{1}{n} (AS_{(j)} + W_{(j)}) V_{(j),\eta} (AS_{(j)} + W_{(j)})^*.$$

Using the fact that $\max_{1 \le i \le n} \{|r_i/\sqrt{N} - 1|\} \xrightarrow{\text{a.s.}} 0$ and that all matrices in the equality above have bounded norm almost surely by (Bai and Silverstein, 1998), we then have $\sup_{1 \le j \le n} \|\check{S}_{(j),\eta} - \hat{S}_{(j),\eta}\| \xrightarrow{\text{a.s.}} 0$. From the results in the previous

section, we then conclude that there exists $\varepsilon > 0$ such that the eigenvalues of $\check{S}_{(j),\eta}$ for all j are all greater than ε for all large n almost surely. Now, recalling that $S = [s_1, \ldots, s_n]$,

$$\frac{1}{N} y_j^* \check{S}_{(j),\eta}^{-1} y_j = \frac{1}{N} s_j^* A^* \check{S}_{(j),\eta}^{-1} A s_j + 2\Re\left[\sqrt{\tau_j} \frac{1}{N} s_j^* A^* \check{S}_{(j),\eta}^{-1} w_j\right] + \tau_j \frac{1}{N} w_j^* \check{S}_{(j),\eta}^{-1} w_j.$$

By the trace lemma (Bai and Silverstein, 2009, Lemma B.26), denoting $\mathcal{A}$ the probability set over which the eigenvalues of $\check{S}_{(j),\eta}$ for all j are greater than ε, for each $p > 2$,

$$\mathrm{E}\left[1_{\mathcal{A}} \left| \frac{1}{N} \tilde{w}_j^* \check{S}_{(j),\eta}^{-1} \tilde{w}_j - \frac{1}{N} \operatorname{tr} \check{S}_{(j),\eta}^{-1} \right|^p \right] \leq K N^{-\frac{p}{2}}$$

where K only depends on ε (which is obtained by first conditioning on $\tilde{W}_{(j)}$ then averaging over it). Taking $p > 3$ and using the union bound on n events, the Markov inequality and the Borel Cantelli lemma, along with $1_{\mathcal{A}} \xrightarrow{\text{a.s.}} 1$ and $\max_j \{|r_j^2/N - 1|\} \xrightarrow{\text{a.s.}} 0$, leads to

$$\max_{1 \leq j \leq n} \left| \frac{1}{N} w_j^* \check{S}_{(j),\eta}^{-1} w_j - \frac{1}{N} \operatorname{tr} \check{S}_{(j),\eta}^{-1} \right| \xrightarrow{\text{a.s.}} 0.$$

Using the same result and the fact that $\frac{1}{N} \operatorname{tr} A^* \check{S}_{(j),\eta}^{-1} A \leq K\varepsilon^{-1}/N$ for all large n a.s., we also have

$$\max_{1 \leq j \leq n} \left| \frac{1}{N} s_j^* A^* \check{S}_{(j),\eta}^{-1} A s_j \right| \xrightarrow{\text{a.s.}} 0.$$

Using both results and $|\frac{1}{N} s_j^* A^* \check{S}_{(j),\eta}^{-1} w_j|^2 \leq \frac{1}{N} s_j^* A^* \check{S}_{(j),\eta}^{-1} A s_j \frac{1}{N} w_j^* \check{S}_{(j),\eta}^{-1} w_j$ (Cauchy-Schwarz inequality), we finally get

$$\max_{1 \leq j \leq n} \left| \frac{1}{N} s_j^* A^* \check{S}_{(j),\eta}^{-1} w_j \right| \xrightarrow{\text{a.s.}} 0.$$

All this then ensures that

$$\max_{1 \leq j \leq n, \tau_j \leq M} \left| \frac{1}{N} y_j^* \check{S}_{(j),\eta}^{-1} y_j - \tau_j \frac{1}{N} \operatorname{tr} \check{S}_{(j),\eta}^{-1} \right| \xrightarrow{\text{a.s.}} 0$$

$$\max_{1 \leq j \leq n, \tau_j > M} \left| \frac{1}{\tau_j} \frac{1}{N} y_j^* \check{S}_{(j),\eta}^{-1} y_j - \frac{1}{N} \operatorname{tr} \check{S}_{(j),\eta}^{-1} \right| \xrightarrow{\text{a.s.}} 0.$$

Since A has rank at most L, $\check{S}_{(j),\eta}$ is an at most rank-$2L+1$ perturbation of $\frac{1}{n} W T_\eta V_\eta W^*$, i.e., the matrix obtained for $p_1 = \ldots = p_L = 0$, by an additive symmetric matrix. A $(2L+1)$-fold application of the rank-one perturbation lemma (Silverstein and Bai, 1995, Lemma 2.6) along with the facts that $\|W - \tilde{W}\| \xrightarrow{\text{a.s.}} 0$ and that all eigenvalues of the matrices involved are uniformly away from zero almost surely then ensures that

$$\max_{1 \leq j \leq n} \left| \frac{1}{N} \operatorname{tr} \check{S}_{(j),\eta}^{-1} - \frac{1}{N} \operatorname{tr} \left(\frac{1}{n} W T_\eta V_\eta W^* \right)^{-1} \right| \xrightarrow{\text{a.s.}} 0.$$

But now, recalling (Couillet et al., 2013b, Remark 2), $\frac{1}{N}\operatorname{tr}\left(\frac{1}{n}WT_\eta V_\eta W^*\right)^{-1} \xrightarrow{\text{a.s.}} \gamma^\eta$. Putting these results together finally leads to the requested result

$$\max_{1\le j\le n,\tau_j\le M}\left|\frac{1}{N}y_j^*\check{S}_{(j),\eta}^{-1}y_j-\tau_j\gamma^\eta\right|\xrightarrow{\text{a.s.}}0$$
$$\max_{1\le j\le n,\tau_j> M}\left|\frac{1}{\tau_j}\frac{1}{N}y_j^*\check{S}_{(j),\eta}^{-1}y_j-\gamma^\eta\right|\xrightarrow{\text{a.s.}}0.$$

Note that the proof only exploits the boundedness away from zero of the various matrices involved and not their bounded spectral norm. Therefore, with the same derivations, we also generalize (Couillet et al., 2013b, Lemma 2) as follows.

Lemma 4. *For every $M>0$, we have*

$$\max_{\substack{1\le j\le n\\ \tau_j\le M}}\left|\frac{1}{N}y_j^*\left(\frac{1}{n}\sum_{i\neq j}v\left(\tau_i\gamma^\eta\right)y_iy_i^*\right)^{-1}y_j-\tau_j\gamma\right|\xrightarrow{\text{a.s.}}0$$
$$\max_{\substack{1\le j\le n\\ \tau_j> M}}\left|\frac{1}{\tau_j}\frac{1}{N}y_j^*\left(\frac{1}{n}\sum_{i\neq j}v\left(\tau_i\gamma^\eta\right)y_iy_i^*\right)^{-1}y_j-\gamma\right|\xrightarrow{\text{a.s.}}0.$$

Define now $d_i=\frac{1}{\tau_i}\frac{1}{N}y_i^*\hat{C}_{(i)}^{-1}y_i$ with $\hat{C}_{(i)}=\hat{C}_N-\frac{1}{n}u(\frac{1}{N}y_i^*\hat{C}_N^{-1}y_i)y_iy_i^*$, from which in passing we can write

$$\hat{C}_N=\frac{1}{n}\sum_{i=1}^n v(\tau_i d_i)y_iy_i^*=\frac{1}{n}\sum_{i=1}^n v(\tau_i d_i)A_i\bar{\bar{w}}_i\bar{\bar{w}}_i^*A_i^* \tag{17}$$

with $\bar{\bar{w}}_i=[s_{1i},\ldots,s_{Li},w_i]^\mathsf{T}$. Then (Couillet et al., 2013b, Lemma 3) remains valid and reads

Lemma 5. *There exists $d_+>d_->0$ such that, for all large n a.s.*

$$d_i<\liminf_n\min_{1\le i\le n}d_i\le\limsup_n\max_{1\le i\le n}d_i<d_+.$$

Proof. Taking $m>0$ small enough and denoting $d_{\max}=\max_j d_j$, Equation (Couillet et al., 2013b, (14)) becomes here

$$\hat{C}_{(j)}\succeq mv(md_{\max})\frac{1}{n}\sum_{\substack{i\neq j\\ \tau_i\ge m}}\frac{1}{\tau_i}\left(As_is_i^*A^*+2\sqrt{\tau_i}\Re\left[w_i^*As_i\right]+\tau_iw_iw_i^*\right)$$

so that, taking j such that $d_j=d_{\max}$,

$$d_{\max}\le\frac{1}{mv(md_{\max})}\frac{1}{\tau_j}\frac{1}{N}y_j^*\left(\frac{1}{n}\sum_{\substack{i\neq j\\ \tau_i\ge m}}\frac{As_is_i^*A^*+2\sqrt{\tau_i}\Re\left[w_i^*As_i\right]+\tau_iw_iw_i^*}{\tau_i}\right)^{-1}y_j.$$

If $\liminf_n \tau_j > 0$ (with j always defined to be such that $d_j = d_{\max}$), with the same arguments as in the proof of Lemma 3 (here the boundedness from above of the τ_i is irrelevant) and recalling (Couillet et al., 2013b, Lemma 6), the right-hand side term can be bounded by $(mv_c(md_{\max})(1-c))^{-1}(1+\varepsilon)$ for arbitrarily small $\varepsilon > 0$ by taking m small enough and n large enough. From there the proof of (Couillet et al., 2013b, Lemma 3) for the boundedness of $d_{\max}$ remains valid. If instead $\liminf_n \tau_j = 0$, we restrict ourselves to a subsequence over which $\tau_j \to 0$. Multiplying both sides of the equation above by τ_j, we get by a similar result as Lemma 3 that $\tau_j d_{\max}$ can be bounded by $\tau_j(mv_c(md_{\max})(1-c))^{-1}(1+\varepsilon)$ for arbitrarily small $\varepsilon > 0$ (again taking m small and n large), and the result unfolds again.

To obtain the lower bound, in the proof of (Couillet et al., 2013b, Lemma 3), denoting $d_{\min} = \min_j d_j$, one needs now write

$$\hat{C}_{(j)} \preceq Mv(Md_{\min})\frac{1}{n} \sum_{\substack{i\neq j \\ m\leq\tau_i\leq M}} \frac{1}{\tau_i}\left(As_is_i^*A^* + 2\sqrt{\tau_i}\Re\left[w_i^*As_i\right] + \tau_i w_i w_i^*\right)$$
$$+ v(0)\frac{1}{n} \sum_{\substack{i\neq j \\ \tau_i\in\mathbb{R}\setminus[m,M]}} \left(As_is_i^*A^* + 2\sqrt{\tau_i}\Re\left[w_i^*As_i\right] + \tau_i w_i w_i^*\right).$$

The controls established for the upper bound on $d_{\max}$ can be similarly used here for $d_{\min}$ and the proof of (Couillet et al., 2013b, Lemma 3) for $d_{\min}$ unfolds then similarly.

Equipped with these lemmas, the proof of Theorem 1 unfolds similar to the proof of (Couillet et al., 2013b, Theorem 2) but for a particular care to be taken for terms involving $\tau_j^{-1}y_j$ which need to be controlled if $\liminf_n \tau_j = 0$. This is easily performed as previously by either using approximations of d_j or of $\tau_j d_j$ depending on whether $\liminf_n \tau_j > 0$ or $\liminf_n \tau_j = 0$, respectively. Assumption 2, which reproduces the assumptions of (Couillet et al., 2013b), is precisely used here. In particular, by the end of the proof, we obtain similar to (Couillet et al., 2013b) the important convergence

$$\max_{1\leq j\leq n, \tau_j<M} |\tau_j d_j - \tau_j\gamma| \xrightarrow{\text{a.s.}} 0$$
$$\max_{1\leq j\leq n, \tau_j\geq M} |d_j - \gamma| \xrightarrow{\text{a.s.}} 0 \tag{18}$$

from which, expanding both $\hat{S}_N$ and $\hat{C}_N$ as in (8) (noting the similarity between (7) and (17)) and exploiting the almost sure asymptotic boundedness in norm of the various matrices then involved, we obtain $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ as desired.

5.5. Eigenvalues of $\hat{C}_N$ and power estimation

From Theorem 1, $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ so that in particular $\max_{1\leq i\leq n} |\lambda_i(\hat{C}_N) - \lambda_i(\hat{S}_N)| \xrightarrow{\text{a.s.}} 0$. This means that it suffices to study the individual eigenvalues of $\hat{S}_N$ in order to study the individual eigenvalues of $\hat{C}_N$. In particular, from the

results of Section 5.3, we have that, for any small $\varepsilon > 0$, $\hat{C}_N$ has asymptotically no eigenvalue in $[0, \varepsilon]$ almost surely, that $\hat{\lambda}_{|\mathcal{L}|+i} < S^+ + \varepsilon$ for all large n a.s. for each $i \in \{1, \ldots, N - |\mathcal{L}|\}$ and that $\hat{\lambda}_i \xrightarrow{\text{a.s.}} \Lambda_i > S^+$ for each $i \in \mathcal{L}$, where Λ_i is as in the statement of Theorem 2, Item 0. Along with the continuity of δ and $\delta((S^+, \infty)) \subset (-(\tau_+ v_c(\tau_+ \gamma)), 0)$, we then get Theorem 2, Item 1.

5.6. Localization function estimation

Let $a, b \in \mathbb{C}^N$ be two vectors of unit norm. Then, from the first part of Theorem 2 and from Cauchy's integral formula, for any $k \in \mathcal{L}$ and for all large N a.s.,

$$\sum_{\substack{1 \le i \le L \\ p_i = p_\ell}} a^* \hat{u}_i \hat{u}_i^* b = -\frac{1}{2\pi\imath} \oint_{\mathcal{J}_\ell} a^* \left(\hat{C}_N - z I_N \right)^{-1} b \, dz \tag{19}$$

for $\mathcal{J}_\ell$ defined as above as a positively oriented contour around a sufficiently small neighborhood of Λ_ℓ, where Λ_ℓ is the unique positive solution of the equation in Λ (14) when $p_k = p_\ell$. Using $\|\hat{C}_N - \hat{S}_N\| \xrightarrow{\text{a.s.}} 0$ along with the uniform boundedness of $\|(\hat{S}_N - z I_N)^{-1}\|$ and $\|(\hat{C}_N - z I_N)^{-1}\|$ on $\mathcal{J}_\ell$ (for all n large), we then have

$$\sum_{\substack{1 \le i \le L \\ p_i = p_\ell}} a^* \hat{u}_i \hat{u}_i^* b + \frac{1}{2\pi\imath} \oint_{\mathcal{J}_\ell} a^* \left(\hat{S}_N - z I_N \right)^{-1} b \, dz \xrightarrow{\text{a.s.}} 0$$

so that it suffices to determine the second left-hand side expression.

Let us develop the term $a^*(\hat{S}_N - z I_N)^{-1} b$. Proceeding similar to Section 5.3, we find

$$a^* \left(\hat{S}_N - z I_N \right)^{-1} b = a^* \left(\hat{S}_N^\circ - z I_N + \Gamma \right)^{-1} b$$

with Γ defined in (9). Using Woodbury's identity $(A + BCB^*)^{-1} = A^{-1} - A^{-1}B(C^{-1} + B^* A^{-1} B)^{-1} B^* A^{-1}$ for invertible A, B, this becomes, with the same notations as in the previous paragraph,

$$a^* \left(\hat{S}_N - z I_N \right)^{-1} b = a^* Q_z^\circ b - a^* Q_z^\circ G \left(H^{-1} + G^* Q_z^\circ G \right)^{-1} G^* Q_z^\circ b \tag{20}$$

where

$$G = \begin{bmatrix} U \Omega^{\frac{1}{2}} & \frac{1}{n} \tilde{W} T^{\frac{1}{2}} V S^* \bar{U} \Omega^{\frac{1}{2}} \end{bmatrix}$$
$$H = \begin{bmatrix} \Omega^{\frac{1}{2}} \bar{U}^* \frac{1}{n} \tilde{W} V \tilde{W}^* \bar{U} \Omega^{\frac{1}{2}} & I_L \\ I_L & 0 \end{bmatrix}.$$

The matrix H is clearly invertible and we then find, using Lemma 1 and Lemma 2 that, uniformly on z in a small neighborhood of Λ_ℓ,

$$\left\| H^{-1} - \begin{bmatrix} 0 & I_L \\ I_L & -\Omega \frac{1}{n} \operatorname{tr} V \end{bmatrix} \right\| \xrightarrow{\text{a.s.}} 0$$

so that, again by Lemma 1 and Lemma 2,

$$\left\| H^{-1} + G^* Q_z^\circ G - \begin{bmatrix} \Omega \frac{\delta(z)}{c} & I_L \\ I_L & -\Omega \frac{1}{n} \operatorname{tr} V \Psi_z^\circ \end{bmatrix} \right\| \xrightarrow{\text{a.s.}} 0. \tag{21}$$

To ensure that $H^{-1}+G^*Q_z^\circ G$ is invertible for $z \in \mathcal{I}_\ell$, let us study the determinant of the rightmost matrix. We have easily

$$\det \left(\begin{bmatrix} \Omega \frac{\delta(z)}{c} & I_L \\ I_L & -\Omega \frac{1}{n} \operatorname{tr} V \Psi_z^\circ \end{bmatrix} \right) = \det \left(-\Omega^2 \frac{\delta(z)}{c} \frac{1}{n} \operatorname{tr} V \Psi_z^\circ - I_L \right).$$

From the discussion around (14), the right-hand side term cancels exactly once in a neighborhood of $z = \Lambda_k$ for each $k \in \mathcal{L}$. Now, for $z \in \mathbb{C} \setminus \mathbb{R}$, it is easily seen that it has non-zero imaginary part. Therefore, since the convergence (21) is uniform on a small neighborhood of Λ_ℓ, for all large n a.s., the determinant of $H^{-1} + G^* Q_z^\circ G$ is uniformly away from zero on $\mathcal{I}_\ell$ (up to taking n larger). We can then freely take inverses in (21) and have, uniformly on $\mathcal{I}_\ell$,

$$\left\| (H^{-1} + G^* Q_z^\circ G)^{-1} - \begin{bmatrix} \Omega \frac{\delta(z)}{c} & I_L \\ I_L & -\Omega \frac{1}{n} \operatorname{tr} V \Psi_z^\circ \end{bmatrix}^{-1} \right\| \xrightarrow{\text{a.s.}} 0.$$

To compute the inverse of the rightmost matrix, it is convenient to write

$$\begin{bmatrix} \Omega \frac{\delta(z)}{c} & I_L \\ I_L & \Omega \frac{1}{n} \operatorname{tr} V \Psi_z^\circ \end{bmatrix} = P \left\{ \begin{bmatrix} \sigma_k \frac{\delta(z)}{c} & 1 \\ 1 & -\sigma_k \frac{1}{n} \operatorname{tr} V \Psi_z^\circ \end{bmatrix} \right\}_{k=1}^{L} P^*$$

where $\{A_k\}_{k=1}^L$ is a block-diagonal matrix with diagonal blocks $A_1, \ldots, A_L$ in this order, and where $P \in \mathbb{C}^{2L \times 2L}$ is the symmetric permutation matrix with $[P]_{ij} = \boldsymbol{\delta}_{j-(L+i/2)}$ for even $i \le L$ and $[P]_{ij} = \boldsymbol{\delta}_{j-(i+1)/2}$ for odd $i \le L$. With this notation, we have

$$\begin{bmatrix} \Omega \frac{\delta(z)}{c} & I_L \\ I_L & \Omega \frac{1}{n} \operatorname{tr} V \Psi_z^\circ \end{bmatrix}^{-1} = P \left\{ \frac{-1}{\frac{\delta(z)}{c} \sigma_k^2 \frac{1}{n} \operatorname{tr} V \Psi_z^\circ + 1} \begin{bmatrix} -\sigma_k \frac{1}{n} \operatorname{tr} V \Psi_z^\circ & -1 \\ -1 & \sigma_k \frac{\delta(z)}{c} \end{bmatrix} \right\}_{k=1}^{L} P^*.$$

Denoting $U = [u_1, \ldots, u_L] \in \mathbb{C}^{N \times L}$ and $\bar{U} = [\bar{u}_1, \ldots, \bar{u_L}] \in \mathbb{C}^{L \times L}$, we have

$$GP = \begin{bmatrix} \sqrt{\sigma_1} u_1 & \sqrt{\sigma_1} \frac{1}{n} \tilde{W} T^{\frac{1}{2}} V S^* \bar{u}_1 & \cdots & \sqrt{\sigma_L} u_L & \sqrt{\sigma_L} \frac{1}{n} \tilde{W} T^{\frac{1}{2}} V S^* \bar{u}_L \end{bmatrix}.$$

From this remark, using again Lemma 1 and Lemma 2, we finally have

$$\sup_{z \in \mathcal{I}_\ell} \left| a^* Q_z^\circ G (H^{-1} + G^* Q_z^\circ G)^{-1} G^* Q_z^\circ b - \sum_{k=1}^{L} a^* u_k u_k^* b \frac{\frac{\delta(z)^2}{c^2} \sigma_k^2 \frac{1}{n} \operatorname{tr} V \Psi_z^\circ}{\frac{\delta(z)}{c} \sigma_k^2 \frac{1}{n} \operatorname{tr} V \Psi_z^\circ + 1} \right| \xrightarrow{\text{a.s.}} 0.$$

Putting things together, using the results above which we recall are uniform on $\mathcal{I}_\ell$, and also using the fact that Q_z° has no pole in $\mathcal{I}_\ell$, we finally have

$$\sum_{\substack{1 \le i \le L \\ p_i = p_\ell}} a^* \hat{u}_i \hat{u}_i^* b - \sum_{k=1}^{L} \frac{1}{2\pi \imath} \oint_{\mathcal{I}_\ell} a^* u_k u_k^* b \frac{\frac{\delta(z)^2}{c^2} \sigma_k^2 \frac{1}{n} \operatorname{tr} V \Psi_z^\circ}{\frac{\delta(z)}{c} \sigma_k^2 \frac{1}{n} \operatorname{tr} V \Psi_z^\circ + 1} dz \xrightarrow{\text{a.s.}} 0$$

which, after taking the limits on the fraction in the integrand, gives

$$\sum_{\substack{1\le i\le L\\ p_i=p_\ell}} a^*\hat{u}_i\hat{u}_i^*b - \sum_{k=1}^{L}\frac{1}{2\pi\imath}\oint_{\mathcal{I}_\ell} a^*u_ku_k^*b\frac{\frac{\delta(z)^2}{c^2}p_k\int\frac{v(\tau\gamma)\nu(d\tau)}{1+\tau v(\tau\gamma)\delta(z)}}{\frac{\delta(z)}{c}p_k\int\frac{v(\tau\gamma)\nu(d\tau)}{1+\tau v(\tau\gamma)\delta(z)}+1}dz \xrightarrow{\text{a.s.}} 0$$

For $z\in(S^+,\infty)$, we already saw that $\delta(z)$ is negative while $\int\frac{v(\tau\gamma)\nu(d\tau)}{1+\tau v(\tau\gamma)\delta(z)}$ is positive. For z non real, both quantities are non real, and therefore do no have poles in $\mathcal{I}_\ell$. The only pole is then obtained for $\frac{\delta(z)}{c}p_k\int\frac{v(\tau\gamma)\nu(d\tau)}{1+\tau v(\tau\gamma)\delta(z)}+1=0$, that is for $z=\Lambda_\ell$ as defined in the previous section. Using l'Hospital rule, the residue of the right complex integral is then evaluated to be

$$\begin{aligned}
\text{Res}(\Lambda_\ell) &= \lim_{z\to\Lambda_\ell}(z-\Lambda_\ell)a^*\Pi_\ell b\frac{\delta(z)}{c}\frac{\int\frac{p_\ell v(t\gamma)}{1+tv(t\gamma)\delta(z)}\nu(dt)}{\int\frac{p_\ell v(t\gamma)}{1+tv(t\gamma)\delta(z)}\nu(dt)+\frac{c}{\delta(z)}}\\
&= \lim_{z\to\Lambda_\ell} a^*\Pi_\ell b\frac{\frac{\delta(z)}{c}\int\frac{v(t\gamma)p_\ell}{1+tv(t\gamma)\delta(z)}\nu(dt)+(z-\Lambda_\ell)\frac{d}{dz}\left(\frac{\delta(z)}{c}\int\frac{p_\ell v(t\gamma)}{1+tv(t\gamma)\delta(z)}\nu(dt)\right)}{-c\frac{\delta'(z)}{\delta(z)^2}-\int\frac{tv(t\gamma)^2p_\ell\delta'(z)}{(1+tv(t\gamma)\delta(z))^2}\nu(dt)}\\
&= a^*\Pi_\ell b\left(c\frac{\delta'(\Lambda_\ell)}{\delta(\Lambda_\ell)^2}+p_\ell\delta'(\Lambda_\ell)\int\frac{\tau v_c(\tau\gamma)^2\nu(d\tau)}{(1+\tau v_c(\tau\gamma)\delta(\Lambda_\ell))^2}\right)^{-1} \qquad (22)
\end{aligned}$$

where $\Pi_\ell\triangleq\sum_{i,p_i=p_\ell}u_iu_i^*$ and the last equality uses $\frac{\delta(\Lambda_\ell)}{c}p_\ell\int\frac{v(\tau\gamma)\nu(d\tau)}{1+\tau v(\tau\gamma)\delta(\Lambda_\ell)}=-1$. Recall now that

$$\delta(\Lambda_\ell)=c\left(-\Lambda_\ell+\int\frac{\tau v_c(\tau\gamma)}{1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma)}\nu(d\tau)\right)^{-1}$$

from which

$$\delta'(\Lambda_\ell)=\frac{\delta(\Lambda_\ell)^2}{c}\left(1-\frac{\delta(\Lambda_\ell)^2}{c}\int\frac{\tau^2v_c(\tau\gamma)^2}{(1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma))^2}\nu(d\tau)\right)^{-1}>0.$$

From the expression of p_ℓ in the previous paragraph and these values, we then further find

$$\begin{aligned}
\text{Res}(\Lambda_\ell) &= a^*\Pi_\ell b\left(1-\frac{\int\frac{\delta(\Lambda_\ell)\tau v_c(\tau\gamma)^2\nu(d\tau)}{(1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma))^2}}{\int\frac{v_c(\tau\gamma)\nu(d\tau)}{1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma)}}\right)^{-1}\left(1-\frac{\delta(\Lambda_\ell)^2}{c}\int\frac{t^2v_c(\tau\gamma)^2\nu(d\tau)}{(1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma))^2}\right)\\
&= a^*\Pi_\ell b\frac{\int\frac{v_c(\tau\gamma)\nu(d\tau)}{1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma)}\left(1-\frac{\delta(\Lambda_\ell)^2}{c}\int\frac{t^2v_c(\tau\gamma)^2\nu(d\tau)}{(1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma))^2}\right)}{\int\frac{v_c(\tau\gamma)\nu(d\tau)}{(1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma))^2}}.
\end{aligned}$$

Inverting the relation

$$\sum_{\substack{1\le i\le L\\ p_i=p_\ell}} a^*\hat{u}_i\hat{u}_i^*b - a^*\Pi_\ell b\frac{\int\frac{v_c(\tau\gamma)\nu(d\tau)}{1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma)}\left(1-\frac{\delta(\Lambda_\ell)^2}{c}\int\frac{t^2v_c(\tau\gamma)^2\nu(d\tau)}{(1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma))^2}\right)}{\int\frac{v_c(\tau\gamma)\nu(d\tau)}{(1+\delta(\Lambda_\ell)\tau v_c(\tau\gamma))^2}}\xrightarrow{\text{a.s.}}0$$

and using $\hat{\lambda}_\ell\xrightarrow{\text{a.s.}}\Lambda_\ell$ for all $\ell\in\mathcal{L}$ then completes the proof.

5.7. Empirical estimators

To prove Theorem 3, one needs to ensure that the empirical estimators introduced in the statement of the theorem are consistent with the estimators introduced in Theorem 2.

Note first that $\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0$ is a consequence of (18). Indeed, letting $M > 0$, from (18),

$$\frac{1}{n}\sum_{\tau_j<M} \tau_j d_j - \gamma \frac{1}{n}\sum_{\tau_j<M} \tau_j \xrightarrow{\text{a.s.}} 0.$$

Still from (18), we also have, a.s.

$$\frac{1}{n}\sum_{\tau_j\geq M} \tau_j d_j - \gamma \frac{1}{n}\sum_{\tau_j\geq M} \tau_j = o\left(\frac{1}{n}\sum_{\tau_j\geq M} \tau_j\right).$$

But $\frac{1}{n}\sum_{\tau_j\geq M}\tau_j \xrightarrow{\text{a.s.}} \int_{(M,\infty)} t\nu(dt) \leq 1$ (say M is a continuity point of ν). Also, $\frac{1}{n}\sum_j \tau_j \xrightarrow{\text{a.s.}} 1$. Putting the results together then gives $\gamma - \hat{\gamma}_n \xrightarrow{\text{a.s.}} 0$. From this, we now get, again with (18),

$$\max_{1\leq j\leq n, \tau_j\leq M} \left|\frac{\tau_j d_j}{\hat{\gamma}_n} - \tau_j\right| \xrightarrow{\text{a.s.}} 0$$
$$\max_{1\leq j\leq n, \tau_j> M} \left|\frac{d_j}{\hat{\gamma}_n} - 1\right| \xrightarrow{\text{a.s.}} 0$$

which is $\max_{\tau_j\leq M}|\tau_j - \hat{\tau}_j| \xrightarrow{\text{a.s.}} 0$ and $\max_{\tau_j>M}|\tau_j^{-1}\hat{\tau}_j - 1| \xrightarrow{\text{a.s.}} 0$, as desired.

We now need to prove that $\hat{\delta}(x) - \delta(x) \xrightarrow{\text{a.s.}} 0$ uniformly on any bounded set of $(S^+ + \varepsilon, \infty)$. For this, recall first that both $\hat{\delta}$ and δ are Stieltjes transforms of distributions with support contained in $[0, S^+]$ and, as such, are analytic in $(S^+ + \varepsilon, \infty)$ and uniformly bounded in any compact of $(S^+ + \varepsilon, \infty)$. Taking the difference and denoting $\hat{\nu}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_{\hat{\tau}_i}$, we have

$$\begin{aligned}
&\hat{\delta}(x) - \delta(x)\\
&= \left(1 - \frac{c}{c_n}\right)\hat{\delta}(x) + \frac{\hat{\delta}(x)\delta(x)}{c_n}\left(\int \frac{tv_c(t\gamma)\nu(dt)}{1+\delta(x)tv_c(t\gamma)} - \int \frac{tv_c(t\hat{\gamma}_n)\hat{\nu}_n(dt)}{1+\hat{\delta}(x)tv_c(t\hat{\gamma}_n)}\right)\\
&= \left(1 - \frac{c}{c_n}\right)\hat{\delta}(x) + \frac{\hat{\delta}(x)\delta(x)}{c_n}\left((\hat{\delta}(x) - \delta(x))\int \frac{t^2 v_c(t\gamma)v_c(t\hat{\gamma}_n)\nu(dt)}{(1+\delta(x)tv_c(t\gamma))(1+\hat{\delta}(x)tv_c(t\hat{\gamma}_n))}\right.\\
&\left. + \int \frac{t(v_c(t\gamma) - v_c(t\hat{\gamma}_n))\nu(dt)}{(1+\delta(x)tv_c(t\gamma))(1+\hat{\delta}(x)tv_c(t\hat{\gamma}_n))} + \int \frac{tv_c(t\hat{\gamma}_n)(\hat{\nu}_n(dt) - \nu(dt))}{1+\hat{\delta}(x)tv_c(t\hat{\gamma}_n)}\right).
\end{aligned}$$

From uniform boundedness of $tv_c(t\hat{\gamma}_n)$ and $tv_c(t\gamma)$, and $\hat{\nu}_n((t,M)) \xrightarrow{\text{a.s.}} \nu((t,M))$ weakly and $\hat{\gamma}_n \xrightarrow{\text{a.s.}} \gamma$, it is easily seen that the last two integrals on the right-hand side can be made arbitrarily small (e.g., by isolating $\tau_i \leq M$ and $\tau_i > M$

and letting M large enough in the previous convergence). Also, the first integral on the right hand side is clearly bounded. Gathering the terms $\hat{\delta}(x) - \delta(x)$ on the left-hand side and taking x large enough so to ensure $\hat{\delta}(x)\delta(x)$ is uniformly smaller than one (recall that their limit is zero as $x \to \infty$), we finally get that $\hat{\delta}(x) - \delta(x)$ can be made arbitrarily small. This is valid for any given large x and therefore on some sequence $\{x^{(i)}\}$ of $(S^+ + \varepsilon, \infty)$ having an accumulation point, $\hat{\delta}(x^{(i)})\delta(x^{(i)}) \xrightarrow{\text{a.s.}} 0$. Since $\hat{\delta}(x) - \delta(x)$ is complex analytic in $(S^+ + \varepsilon, \infty)$, by Vitali's convergence theorem, we therefore get that the convergence is uniform over any bounded set of $(S^+ + \varepsilon, \infty)$, which is what we wanted.

Since, for $i \in \mathcal{L}$ and for some $\varepsilon, M > 0$, $\hat{\lambda}_i \in [S^+ + \varepsilon, M]$ for all large n a.s., we therefore have that $\hat{\delta}(\hat{\lambda}_i) - \delta(\lambda_i) \xrightarrow{\text{a.s.}} 0$ for each $i \in \mathcal{L}$. Using all these convergence results, we then obtain, with the same line of arguments the asymptotic consistence between the estimates in Item 1. and Item 2. of both Theorems 2 and 3. This concludes the proof of Theorem 3.

5.8. Proof of Corollary 1

We are here in the same setting as (Hachem et al., 2013, Theorem 3), only for our improved model. The proof is the same as in (Hachem et al., 2013) and relies on showing the uniform convergence of $\hat{\eta}_{\mathrm{RG}}(\theta) - \eta(\theta)$ across θ, from which the result unfolds. In our setting, the point-wise convergence easily follows from Items 3. in both Theorem 2 and Theorem 3. Uniform convergence then hinges on a regular discretization of the set $[0, 2\pi)$ into N^2 subsets and on (i) a Lipschitz control of the differences $\hat{\eta}_{\mathrm{RG}}(\theta) - \hat{\eta}_{\mathrm{RG}}(\theta')$ for $|\theta - \theta'| = O(N^{-2})$ and (ii) a joint convergence of $\hat{\eta}_{\mathrm{RG}}(\theta) - \eta(\theta)$ over the $N^2 + 1$ edges of the subsets. Point (i) uses the defining properties of $a(\theta)$ from Assumption 4 similar to (Hachem et al., 2013), while Point (ii) is obtained thanks to a classical union bound on N^2 events, the validity of which follows from considering sufficiently high order moment bounds on the vanishing random quantities involved in $\hat{\eta}_{\mathrm{RG}}(\theta) - \eta(\theta)$. In our setting, the latter moment bounds are obtained by selecting p large enough in Lemma 1 of Section 5 (in a similar fashion as is performed for the technical proof that $\min_j \lambda_N(\hat{S}_{(j)}) > \varepsilon$ for all large n a.s. in Section 5). It is easily seen that, this being ensured, the proof of Corollary 1 unfolds similar to that of (Hachem et al., 2013, Theorem 3), which as a consequence we do not further detail.

References

Bai, Z. D., Silverstein, J. W., 1998. No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. The Annals of Probability 26 (1), 316–345.

Bai, Z. D., Silverstein, J. W., 2009. Spectral analysis of large dimensional random matrices, 2nd Edition. Springer Series in Statistics, New York, NY, USA.

Baik, J., Silverstein, J. W., 2006. Eigenvalues of large sample covariance matrices of spiked population models. Journal of Multivariate Analysis 97 (6), 1382–1408.

Benaych-Georges, F., Guionnet, A., Maida, M., 2010. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices.
URL http://arxiv.org/abs/1009.0145

Benaych-Georges, F., Rao, R., 2011. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Advances in Mathematics 227 (1), 494–521.

Chapon, F., Couillet, R., Hachem, W., Mestre, X., 2012. On the isolated eigenvalues of large Gram random matrices with a fixed rank deformation. Electronic Journal of ProbabilitySubmitted for publication.

Chitour, Y., Couillet, R., Pascal, F., 2014. Uniqueness of maronna's m-estimators of scatter. arXiv preprint arXiv:1403.5977.

Couillet, R., Hachem, W., 2012. Fluctuations of spiked random matrix models and failure diagnosis in sensor networks. IEEE Transactions on Information TheoryTo appear.

Couillet, R., Hachem, W., 2013. Analysis of the limit spectral measure of large random matrices of the separable covariance type. Journal of Multivariate Analysis.

Couillet, R., McKay, M., 2013. Large dimensional analysis and optimization of robust shrinkage covariance matrix estimators. to appear in Journal of Multivariate Analysis.

Couillet, R., Pascal, F., Silverstein, J. W., 2013a. Robust Estimates of Covariance Matrices in the Large Dimensional Regime. IEEE Transactions on Information Theory.
URL http://arxiv.org/abs/1204.5320

Couillet, R., Pascal, F., Silverstein, J. W., 2013b. The random matrix regime of Maronna's M-estimator with elliptically distributed samples. Journal of Multivariate Analysis.
URL http://arxiv.org/abs/1311.7034

Girko, V. L., 1987. Introduction to general statistical analysis. Theory of Probability & Its Applications 32 (2), 229–242.

Hachem, W., Loubaton, P., Mestre, X., Najim, J., Vallet, P., 2013. A subspace estimator for fixed rank perturbations of large random matrices. Journal of Multivariate Analysis 114, 427–447.

Horn, R. A., Johnson, C. R., 1985. Matrix Analysis. Cambridge University Press.

Huber, P. J., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics 35 (1), 73–101.

Loubaton, P., Vallet, P., 2010. Almost sure localization of the eigenvalues in a Gaussian information plus noise model. Application to the spiked models. Electronic Journal of Probability 16, 1934–1959.

Maronna, R. A., 1976. Robust M-estimators of multivariate location and scatter. The annals of statistics, 51–67.

Maronna, R. A., Martin, R. D., Yohai, V. J., 2006. Robust statistics. J. Wiley.

Marčenko, V. A., Pastur, L. A., 1967. Distribution of eigenvalues for some sets of random matrices. Math USSR-Sbornik 1 (4), 457–483.

Mestre, X., Nov. 2008a. Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates. IEEE Transactions on Information Theory 54 (11), 5113–5129.

Mestre, X., Nov. 2008b. On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. IEEE Transactions on Signal Processing 56 (11), 5353–5368.

Nadler, B., 2010. Nonparametric detection of signals by information theoretic criteria: performance analysis and an improved estimator. IEEE Transactions on Signal Processing 58 (5), 2746–2756.

Pastur, L., Ŝerbina, M., 2011. Eigenvalue distribution of large random matrices. American Mathematical Society.

Paul, D., 2007. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Statistica Sinica 17 (4), 1617.

Schmidt, R., 1986. Multiple emitter location and signal parameter estimation. IEEE Transactions on Antennas and Propagation 34 (3), 276–280.

Silverstein, J. W., Bai, Z. D., 1995. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. Journal of Multivariate Analysis 54 (2), 175–192.

Silverstein, J. W., Choi, S., 1995. Analysis of the limiting spectral distribution of large dimensional random matrices. Journal of Multivariate Analysis 54 (2), 295–309.

Titchmarsh, E. C., 1939. The Theory of Functions. Oxford University Press, New York, NY, USA.

Tyler, D. E., 1987. A distribution-free M-estimator of multivariate scatter. The Annals of Statistics 15 (1), 234–251.