

# Signal Processing in Large Systems: A New Paradigm

Romain Couillet<sup>1</sup> and Mérouane Debbah<sup>2</sup>

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

<sup>2</sup> *Alcatel Lucent Chair on Flexible Radio, Gif sur Yvette, Supélec, France.*

**Abstract**—For a long time, detection and parameter estimation methods for signal processing have relied on asymptotic statistics as the number  $n$  of observations of a population grows large comparatively to the population size  $N$ , i.e.  $n/N \rightarrow \infty$ . Modern technological and societal advances now demand the study of sometimes extremely large populations and simultaneously require fast signal processing due to accelerated system dynamics. This results in not-so-large practical ratios  $n/N$ , sometimes even smaller than one. A disruptive change in classical signal processing methods has therefore been initiated in the past ten years, mostly spurred by the field of large dimensional random matrix theory. The early works in random matrix theory for signal processing applications are however scarce and highly technical. This tutorial provides an accessible methodological introduction to the modern tools of random matrix theory and to the signal processing methods derived from them, with an emphasis on simple illustrative examples.

## I. INTRODUCTION

Most signal processing methods, e.g. statistical tests or parameter estimators, are based on asymptotic statistics of  $n$  observed random signal vectors [1]. This is because a deterministic behavior often arises when  $n \rightarrow \infty$ , which simplifies the problem as exemplified by the celebrated law of large numbers and the central limit theorem. With the increase of the systems dimension, denoted by  $N$ , and the need for even faster dynamics, the scenario  $n \gg N$  becomes less and less likely to occur in practice. As a consequence, the large  $n$  properties are no longer valid, as will be discussed in Section II. This is the case for instance in array processing where a large number of antennas is used to detect incoming signals with short time stationarity properties (e.g. radar detection of fast moving objects) [2]. Other examples are found for instance in mathematical finance where numerous stocks show price index correlation over short observable time periods [3] or in evolutionary biology where the joint presence of multiple genes in the genotype of a given species is analyzed from a few DNA samples [4].

This tutorial introduces recent tools to cope with the  $n \simeq N$  limitation of classical approaches. These tools are based on the study of large dimensional random matrices, which originates from the works of Wigner [5] in 1955 and, more importantly here, of Marčenko and Pastur [6] in 1967. Section III presents a brief introduction to random matrix theory. More precisely, we will explain the basic tools used in this field which differ greatly from classical approaches, based on which we will introduce statistical inference tools in the  $n \simeq N$  regime. These tools use random matrix theory jointly with complex analysis

methods [7] and are often known as G-estimation, named after the G-estimator formulas from Girko [8]. Since the techniques presented in Section III are expected to be rather new to the non-expert reader, we will elaborate on toy examples to introduce the methods rather than on a list of theoretical results. More realistic application examples, taken from the still limited literature, are then presented in Section IV. These will further familiarize the reader with the introduced concepts. These examples span from signal detection in array processing to failure localisation in large dimensional networks. Finally, in Section V, we will draw some conclusions and provide future prospects of this emerging field.

*Notations:* Boldface lowercase (uppercase) characters stand for vectors (matrices), with  $\mathbf{I}_N$  the  $N \times N$  identity matrix. The notations  $(\cdot)^T$  and  $(\cdot)^H$  denote transpose and Hermitian transpose, respectively. The value  $\sqrt{-1}$  is denoted by  $i$ . The notation  $\mathbb{E}$  is the expectation operator. The notations  $\xrightarrow{\text{a.s.}}$  and  $\Rightarrow$  denote almost sure and weak convergence of random variables, respectively. The symbol  $\mathcal{N}(\mu, \sigma^2)$  indicates a real Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . The norm  $\|\cdot\|$  will be understood as the Euclidean norm for vectors and the spectral norm for Hermitian matrices ( $\|\mathbf{X}\| = \max_i |\lambda_i|$  for  $\mathbf{X}$  with eigenvalues  $\lambda_1, \dots, \lambda_N$ ). Finally,  $(x)^+ = \max(x, 0)$  for real  $x$ ,  $1_{x \leq y}$  is the function of  $x$  equal to zero for  $x > y$  and equal to one for  $x \leq y$ , and  $\delta_x^y = 1_{x \leq y} 1_{y \leq x}$ .

## II. LIMITATIONS OF CLASSICAL SIGNAL PROCESSING

We start our exposition with a simple example explaining the compelling need for the study of large dimensional random matrices.

### A. The Marčenko-Pastur law

Consider  $\mathbf{y} = \mathbf{T}^{\frac{1}{2}} \mathbf{x} \in \mathbb{C}^N$  a complex random variable with  $\mathbf{x}$  having independent and identically distributed (i.i.d.) entries with zero mean and unit variance. If  $\mathbf{y}_1, \dots, \mathbf{y}_n$  are  $n$  independent realizations of  $\mathbf{y}$ , then, from the strong law of large numbers,

$$\left\| \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^H - \mathbf{T} \right\| \xrightarrow{\text{a.s.}} 0 \quad (1)$$

as  $n \rightarrow \infty$ . For further use, we will denote  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$  with  $(i, j)$  entry  $Y_{ij}$ , and realize that with this notation,

$$\hat{\mathbf{T}} \triangleq \frac{1}{n} \mathbf{Y} \mathbf{Y}^H = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^H.$$

Here  $\hat{\mathbf{T}}$  is often called the *sample covariance matrix*, obtained from the observations  $\mathbf{y}_1, \dots, \mathbf{y}_n$  of the random variable  $\mathbf{y}$  with *population covariance matrix*  $E[\mathbf{y}\mathbf{y}^H] = \mathbf{T}$ . The convergence (1) suggests that, for  $n$  sufficiently large,  $\hat{\mathbf{T}}$  is an accurate estimate of  $\mathbf{T}$ . The natural reason why  $\mathbf{T}$  can be asymptotically well approximated is because  $\hat{\mathbf{T}}$  originates from a total of  $Nn \gg N^2$  observations, while the parameter  $\mathbf{T}$  to be estimated is of size  $N^2$ .

Suppose now for simplicity that  $\mathbf{T} = \mathbf{I}_N$ . If the relation  $Nn \gg N^2$  is not met in practice, e.g. if  $n$  cannot be taken very large compared to the system size  $N$ , then a peculiar behaviour of  $\hat{\mathbf{T}}$  arises, as both  $N, n \rightarrow \infty$  while  $N/n$  does not tend to zero. Observe that the entry  $(i, j)$  of  $\hat{\mathbf{T}}$  is given by

$$[\hat{\mathbf{T}}]_{ij} = \frac{1}{n} \sum_{k=1}^n Y_{ik} Y_{jk}^*.$$

Since  $\mathbf{T} = \mathbf{I}_N$ ,  $E[Y_{ik} Y_{jk}^*] = \delta_{ij}$ . Therefore, the law of large numbers ensures that  $[\hat{\mathbf{T}}]_{ij} \rightarrow 0$  if  $i \neq j$ , while  $[\hat{\mathbf{T}}]_{ii} \rightarrow 1$ , for all  $i, j$ . Since  $N$  grows along with  $n$ , we might then say that  $\hat{\mathbf{T}}$  converges point-wise to an “infinitely large identity matrix”. This stands, irrelevant of  $N$ , as long as  $n \rightarrow \infty$ , so in particular if  $N = 2n \rightarrow \infty$ . However, under this hypothesis,  $\hat{\mathbf{T}}$  is of maximum rank  $N/2$  and is therefore rank-deficient.<sup>1</sup> Altogether, we may then conclude that  $\hat{\mathbf{T}}$  converges point-wise to an “infinite-size identity matrix” which has the property that half of its eigenvalues equal zero. The convergence (1) therefore clearly does not hold here as the spectral norm (or the absolute largest eigenvalue) of  $\hat{\mathbf{T}} - \mathbf{I}_N$  is greater or equal to 1 for all  $N$ .

This convergence paradox (convergence of  $\hat{\mathbf{T}}$  to an identity matrix with zero eigenvalues) arises obviously from an incorrect reasoning when taking the limits  $N, n \rightarrow \infty$ . Indeed, the convergence of matrices with increasing sizes only makes sense if a proper measure, here the spectral norm, in the space of such matrices is considered. The outcome of our previous observation is then that the *random* (or empirical) eigenvalue distribution  $F^{\hat{\mathbf{T}}}$  of  $\hat{\mathbf{T}}$ , defined as

$$F^{\hat{\mathbf{T}}}(x) = \frac{1}{N} \sum_{i=1}^N 1_{x \leq \lambda_i}$$

with  $\lambda_1, \dots, \lambda_N$  the eigenvalues of  $\hat{\mathbf{T}}$ , does not converge (weakly) to a Dirac mass in 1, and this is all what can be said at this point. This remark has fundamental consequences for basic signal processing detection and estimation procedures. In fact, although the eigenvalue distribution of  $\hat{\mathbf{T}}$  does not converge to a Dirac in 1, it turns out that in many situations of practical interest, it does converge towards a limiting distribution, as  $N, n \rightarrow \infty$ ,  $N/n \rightarrow c \geq 0$ . In the particular case where  $\mathbf{T} = \mathbf{I}_N$ , the limiting distribution is known as the Marčenko-Pastur law, initially studied by Marčenko and Pastur in [6]. This distribution is depicted in Figure 1 and compared to the empirical eigenvalue distribution of a matrix  $\hat{\mathbf{T}}$  of large dimensions. From this figure, we observe that the empirical distribution of the eigenvalues is a good match to

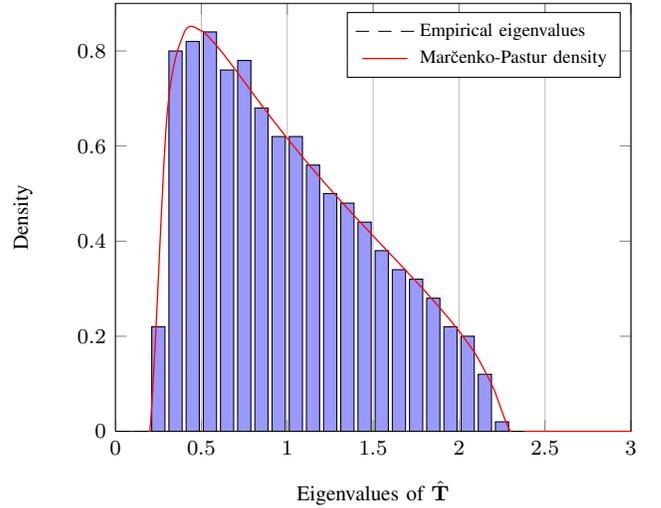


Fig. 1. Histogram of the eigenvalues of a single realization of  $\hat{\mathbf{T}} = \frac{1}{n} \sum_{k=1}^n \mathbf{y}_k \mathbf{y}_k^H$ ,  $\mathbf{y}_k$  has standard Gaussian entries, for  $n = 2000$ ,  $N = 500$ .

the limiting distribution for  $N, n$  large. It is also observed that no eigenvalue seems to lie outside the support of the limiting distribution even for finite  $N, n$ . This fundamental property can be proved to hold for any matrix  $\mathbf{Y}$  with independent zero mean and unit variance entries, along with some mild assumptions on the higher order moments. The shape of the Marčenko-Pastur law for different limiting ratios  $c$  is depicted in Figure 2. We see from this figure that, as  $c \rightarrow 0$ , the support of the Marčenko-Pastur law tends to concentrate into a single mass in 1. This is in line with our expectations from the discussion above for finite  $N$ , while the support tends to spread and eventually reaches 0 when  $c$  is large, which is compliant with the existence of a mass of eigenvalues at 0 when  $N < n$ , i.e.  $c > 1$ .

The general tools employed to derive the Marčenko-Pastur law and its generalizations to some more advanced random matrix models are briefly introduced in Section III.

The main implication of the above observations on signal processing methods is that, in general, the tools developed in view of applications for small  $N$  and large  $n$  are no longer adequate if either  $N$  is taken much larger, or if  $n$  cannot be afforded to be large. Observe for instance in Figure 2 that, even for  $n$  being ten times larger than  $N$ , the spread of the Marčenko-Pastur law around 1 is significant enough for the classical large  $n$  assumption to be quite disputable. In the remainder of this section, we introduce a practical inference problem which assumes the setup  $n \gg N$  and which can be proved asymptotically biased if  $N/n$  does not converge to zero.

### B. An eigenvalue inference problem

The following example deals with the statistical inference of eigenvalues of a large population covariance matrix. This problem is very generic and finds applications in array processing, such as radar detection, where the objective is to enumerate the sources and to infer their distances to the radar, or in cognitive radios where the objective is to detect and estimate

<sup>1</sup>Indeed,  $\hat{\mathbf{T}} = \frac{1}{N} \sum_{i=1}^{N/2} \mathbf{x}_i \mathbf{x}_i^H$ , which is the sum of  $N/2$  rank-1 matrices.

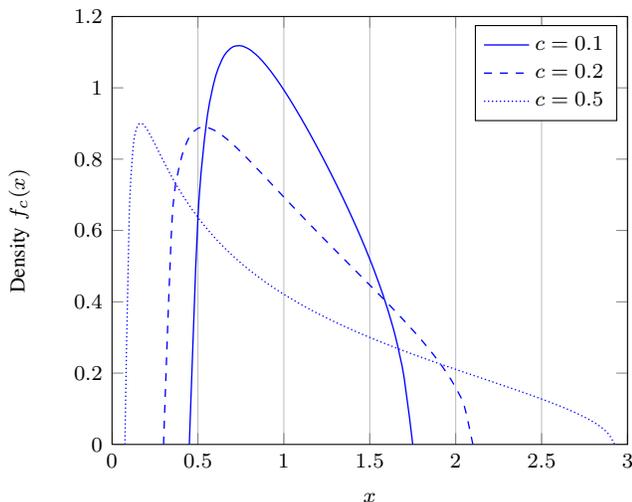


Fig. 2. Marčenko-Pastur density  $f_c$  for different limit ratios  $c = \lim N/n$ .

the power of concurrent signal transmissions. This example will be used throughout the article to demonstrate the transition from the classical signal processing considerations towards the advanced random matrix approaches.

We assume that one has access to successive independent realizations  $\mathbf{y}_1, \dots, \mathbf{y}_n$  of a stationary process  $\mathbf{y}_t = \mathbf{U}\mathbf{x}_t \in \mathbb{C}^N$ , with  $\mathbf{U} \in \mathbb{C}^{N \times N}$  an unknown unitary matrix and  $\mathbf{x}_t \in \mathbb{C}^N$  a complex circularly symmetric Gaussian vector with zero mean and covariance matrix  $\mathbf{P}$ . We further assume that  $\mathbf{P}$  is diagonal and composed of  $N_i$  eigenvalues equal to  $P_i$ , for  $1 \leq i \leq K$ , where the  $P_i$  are distinct, and where  $N = \sum_{i=1}^K N_i$ . The presence of the unknown matrix  $\mathbf{U}$  translates the fact that, in many applications, the eigenvector structure of the population covariance matrix  $\mathbf{T} = \mathbf{U}\mathbf{P}\mathbf{U}^H$  of  $\mathbf{y}_t$  is not necessarily known. Hence, the diagonal entries of  $\mathbf{P}$  are not directly accessible. Note that this is in sharp contrast to covariance matrix estimation of structured signals, as in the recent works [9], [10] where population covariance matrices have a Toeplitz-like structure. Here, no such assumption on  $\mathbf{U}$  is made, so that alternative techniques need be used in order to estimate the entries of  $\mathbf{P}$ .

Our objective is to infer the values of  $P_1, \dots, P_K$  in the setting where both  $n$  and  $N$  are of similar order of magnitude or, stated in signal processing terms, when the number of observations of the process  $\mathbf{y}_t$  is *not* large compared to the system dimension  $N$ . For readability, we denote  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$  and recall that  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  is the sample covariance matrix of these data.

Let us first consider the classical approach which assumes that  $n \gg N$ . We then have from a direct application of the law of large numbers that

$$\left\| \frac{1}{n}\mathbf{Y}\mathbf{Y}^H - \mathbf{U}\mathbf{P}\mathbf{U}^H \right\| \xrightarrow{\text{a.s.}} 0 \quad (2)$$

as  $n \rightarrow \infty$ . Since the convergence is in spectral norm and since the eigenvalues of  $\mathbf{U}\mathbf{P}\mathbf{U}^H$  are the same as those of  $\mathbf{P}$ ,

an (asymptotically)  $n$ -consistent estimate  $\hat{P}_k^\infty$  for  $P_k$  reads

$$\hat{P}_k^\infty = \frac{1}{N_k} \sum_{i \in \mathcal{N}_k} \lambda_i \quad (3)$$

where  $\lambda_1 \leq \dots \leq \lambda_N$  are the ordered eigenvalues of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  gathered in successive clusters  $\mathcal{N}_k = \{N - \sum_{i=k}^K N_i + 1, N - \sum_{i=k+1}^K N_i\}$ . Indeed, the eigenvalues of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  indexed by  $\mathcal{N}_k$  can be directly mapped to those with same indexes in  $\mathbf{P}$ , as can already be observed in Figure 3.

However, for  $N$  and  $n$  of comparable sizes, the convergence (2) no longer holds. In this scenario, the spectrum of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  converges weakly to the union of a maximum of  $K$  compact clusters of eigenvalues concentrated somewhat around  $P_1, \dots, P_K$ , but whose centers of mass are not located at  $P_1, \dots, P_K$ . This is depicted in Figure 3, where it is clearly seen that eigenvalues gather somewhat around the empirical values of  $P_1, P_2, P_3$  in clusters. Therefore, as  $N, n \rightarrow \infty$ , the estimator  $\hat{P}_k^\infty$  is no longer (asymptotically) consistent. Note also that, depending on the values of  $P_1, \dots, P_K$  (and in fact also on the ratio  $N/n$ ), the number of clusters varies. From the second plot of Figure 3, we already anticipate the well-known problem of order selection, i.e. determining the number  $K$  of distinct input sources from an observation  $\mathbf{Y}$ . For these reasons, more advanced analyses of the spectrum of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  must be carried out, from which  $(N, n)$ -consistent alternatives to (3) can be derived, i.e. estimators which are consistent as both  $N, n$  grow large. This introduction of a methodological approach to  $(N, n)$ -consistent estimators, called G-estimators, is one of the objectives of this article.

### III. LARGE DIMENSIONAL RANDOM MATRIX THEORY

In this section, we introduce the basic notions of large dimensional random matrix theory and describe two elementary tools: the resolvent and the Stieltjes transform. These tools are necessary to understand the stark difference between the large random matrix techniques and the conventional statistical methods.

Random matrix theory studies the properties of matrices whose entries follow a joint probability distribution. The two main axes deal with the exact statistical characterization of small-size matrices and the asymptotic behaviour of large dimensional matrices. For the analysis of small-size matrices, one usually seeks closed form expressions of the exact statistics. However, these expressions become quickly intractable for practical analysis as one departs from white Gaussian matrices [11]–[13]. The mathematical methods used are usually very standard, mostly based on standard tools from probability theory. In the context of large dimensional random matrices, the tools are very different: asymptotic statistics and probability are obviously fundamental, but also linear algebra, as well as real and complex analysis, and combinatorics. The focus of these techniques are either based on invariance properties of the underlying matrices (free probability theory [14], [15], combinatorics [16], [17], Gaussian integration by part methods [18], [19]) or on independence properties of their entries [6], [20], [21]. In the following, we will mostly concentrate on tools for large dimensional random matrices

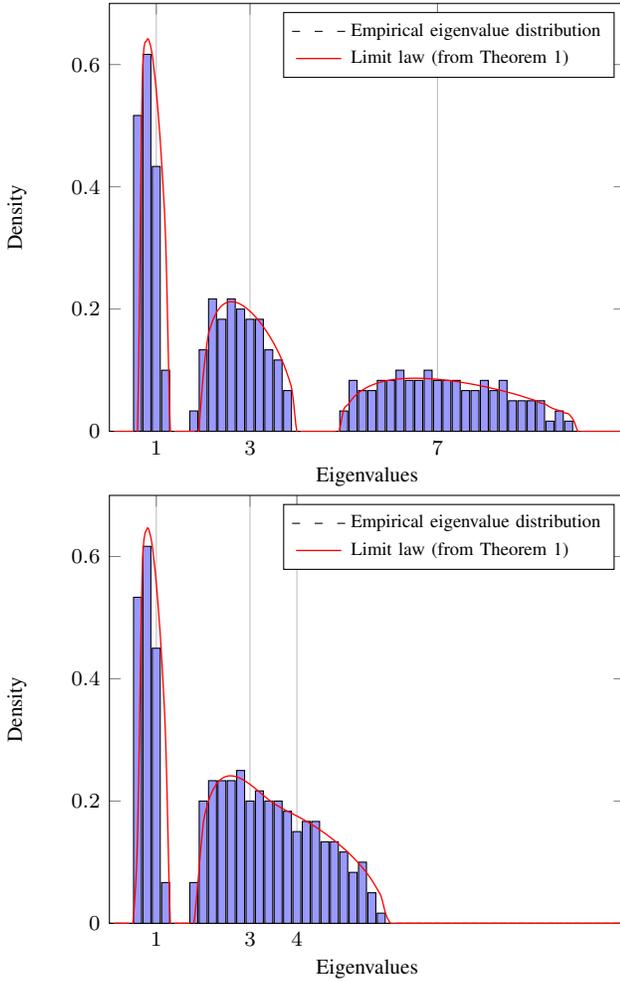


Fig. 3. Histogram of the eigenvalues of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  for  $N = 300$ ,  $n = 3000$ , with  $\mathbf{P}$  diagonal composed of three evenly weighted masses in (i) 1, 3 and 7 at the top, (ii) 1, 3, and 4 on the bottom.

with independent entries, although some results we present originate from other techniques. The main tool of interest in the spectral study of these matrices is the *Stieltjes transform*, which we introduce hereafter.

#### A. The Stieltjes transform

Similar to the celebrated Fourier transform in classical probability theory and signal processing which allows one to perform simpler analysis in the Fourier (or frequency) domain than in the initial (or time) domain, the *spectral analysis* of large dimensional random matrices<sup>2</sup> is often carried out with the help of the Stieltjes transform.

*Definition 1 (Stieltjes transform):* Let  $F$  be a real probability distribution function and  $z \in \mathbb{C}$  taken outside the support  $\mathcal{S}$  of  $F$ . Then the (Cauchy-)Stieltjes transform  $m_F(z)$  of  $F$  at point  $z$  is defined as

$$m_F(z) \triangleq \int_{\mathcal{S}} \frac{1}{t-z} dF(t).$$

<sup>2</sup>What one refers here to as spectrum is the eigenvalue distribution of the underlying matrix.

Note importantly that, for  $\mathbf{X} \in \mathbb{C}^{N \times N}$  Hermitian with eigenvalues  $\lambda_1, \dots, \lambda_N$ , and eigenvalue distribution  $F^{\mathbf{X}}$ , the Stieltjes transform  $m_{\mathbf{X}}$  of  $\mathbf{X}$  is

$$m_{\mathbf{X}}(z) = \int \frac{dF^{\mathbf{X}}(t)}{t-z} = \frac{1}{N} \sum_{k=1}^N \frac{1}{\lambda_k - z} = \frac{1}{N} \text{tr}(\mathbf{X} - z\mathbf{I}_N)^{-1}.$$

The Stieltjes transform, in the same way as the Fourier transform, has an inverse formula, given by

$$F(b) - F(a) = \lim_{y \downarrow 0} \int_a^b m_F(x + iy) dx. \quad (4)$$

This formula is however rarely used in practice. What this says for random matrices is that, if the Stieltjes transform of a matrix  $\mathbf{X}$  is known, then one can retrieve its eigenvalue distribution, which is often more difficult to obtain in the spectral domain than in the Stieltjes transform domain.

Since we will constantly deal with matrices of the form  $(\mathbf{X} - z\mathbf{I}_N)^{-1}$ , called the *resolvent* matrix of  $\mathbf{X}$ , the analysis of random matrices is fundamentally based on the exploitation of classical matrix inversion formulas. In order to better capture the essence of the Stieltjes transform approach, we hereafter outline the core arguments of the proof of the Marčenko-Pastur law, as performed in the original article [6] back in 1967.

#### B. The Marčenko-Pastur law

We recall that we wish to determine the expression of the limit of the eigenvalue distribution of  $\mathbf{Y}\mathbf{Y}^H$ , given in Figure 2, as  $N, n \rightarrow \infty$ ,  $N/n \rightarrow c$ , where  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{C}^{N \times n}$  has i.i.d. entries of zero mean and variance  $1/n$ . In the Stieltjes transform domain, we therefore need to evaluate  $m_{\mathbf{Y}\mathbf{Y}^H}(z) = \frac{1}{N} \text{tr}(\mathbf{Y}\mathbf{Y}^H - z\mathbf{I}_N)^{-1}$ .

The idea is to concentrate on the value of each individual diagonal entry of the resolvent  $(\mathbf{Y}\mathbf{Y}^H - z\mathbf{I}_N)^{-1}$ . Because of its obvious symmetrical structure, it suffices to study the asymptotic behaviour of the first diagonal entry. Writing  $(\mathbf{Y}\mathbf{Y}^H - z\mathbf{I}_N)^{-1}$  as a block matrix with upper-left corner composed of the unique (1, 1) entry and applying the classical Schur complement [22], it is easy to see that

$$[(\mathbf{Y}\mathbf{Y}^H - z\mathbf{I}_N)^{-1}]_{11} = \frac{1}{-z - z\mathbf{y}_1^H(\tilde{\mathbf{Y}}^H\tilde{\mathbf{Y}} - z\mathbf{I}_n)^{-1}\mathbf{y}_1} \quad (5)$$

where we have defined  $\tilde{\mathbf{Y}}$  and  $\mathbf{y}_1$  such that  $\mathbf{Y}^H = [\mathbf{y}_1 \tilde{\mathbf{Y}}^H]$ . Due to the independence between  $\mathbf{y}_1$  and  $\tilde{\mathbf{Y}}$  and the fact that  $\mathbf{y}_1$  has i.i.d. entries, the quantity  $\mathbf{y}_1^H(\tilde{\mathbf{Y}}^H\tilde{\mathbf{Y}} - z\mathbf{I}_n)^{-1}\mathbf{y}_1$  approaches  $\frac{1}{n} \text{tr}(\tilde{\mathbf{Y}}^H\tilde{\mathbf{Y}} - z\mathbf{I}_n)^{-1}$  as the system sizes  $N, n$  grow large. Indeed, this is clearly true in the expectation over  $\mathbf{y}_1$  (with equality) and, conditioned on  $\tilde{\mathbf{Y}}$ , the convergence somewhat recalls a law of large numbers. The interesting part is to notice that the latter trace expression is tightly connected to  $\frac{1}{N} \text{tr}(\mathbf{Y}\mathbf{Y}^H - z\mathbf{I}_N)^{-1}$ , as the only major difference between the two lies in a single column change (a rank-one perturbation) which asymptotically will not alter the limiting normalized trace. A complete exposition of this derivation can be found in many random matrix books, e.g. [23]–[25].

Since this remark is valid for all diagonal entries of the resolvent  $(\mathbf{Y}\mathbf{Y}^H - z\mathbf{I}_N)^{-1}$ , it holds also true for their average  $\frac{1}{N} \text{tr}(\mathbf{Y}\mathbf{Y}^H - z\mathbf{I}_N)^{-1}$ , i.e. for the Stieltjes transform  $m_{\mathbf{Y}\mathbf{Y}^H}(z)$

of  $\mathbf{Y}\mathbf{Y}^H$ . Recalling (5), the Stieltjes transform can therefore be asymptotically well approximated as a function of itself. This naturally leads  $m_{\mathbf{Y}\mathbf{Y}^H}(z)$  to be approximately the solution of a *fixed-point equation* (or implicit equation). In many random matrix models, this is as far as the Stieltjes transform would go, and we would conclude that the limiting spectrum of the matrix under study is defined through its Stieltjes transform, whose expression is only known through a fixed-point equation (with usually a unique solution<sup>3</sup>).

In the case of the Marčenko-Pastur law, this fixed-point equation reads precisely

$$m_{\mathbf{Y}\mathbf{Y}^H}(z) \simeq \frac{1}{1 - c - z - zcm_{\mathbf{Y}\mathbf{Y}^H}(z)} \quad (6)$$

which turns out to be equivalent to a second order polynomial in  $m_{\mathbf{Y}\mathbf{Y}^H}(z)$  which can be solved explicitly. Using the inverse Stieltjes transform formula (4), we then obtain the limiting spectrum of  $\mathbf{Y}\mathbf{Y}^H$  with density  $f(x)$  given by

$$f(x) = (1 - c^{-1})^+ \delta(x) + \frac{1}{2\pi cx} \sqrt{(x-a)(b-x)}$$

with  $a = (1 - \sqrt{c})^2$ ,  $b = (1 + \sqrt{c})^2$ , and we recall that  $c$  is the limiting ratio  $\lim N/n$ , where the square-root appearing in the expression originates from the roots of the second order polynomial defining  $m_{\mathbf{Y}\mathbf{Y}^H}(z)$ .

It is interesting to see that this expression is only defined for  $a \leq x \leq b$  and possibly  $x = 0$ , which defines the support of the distribution. In particular, we confirm, as already discussed that, as  $c \rightarrow 0$ ,  $[a, b]$  tends to a singleton in 1. The convergence to  $\{1\}$  is however quite slow as it is of order  $\sqrt{c}$  for  $c$  small (since  $(1 + \sqrt{c})^2 \simeq 1 + 2\sqrt{c}$  in this regime). This suggests that  $n$  must be much larger than  $N$  for the classical large- $n$  approximation to be acceptable. Typically, in array processing with 100 antennas, no less than 10 000 signal observations must be available for subspace methods relying on the large- $n$  approximation to be valid with an error of order 1/10 on the assumed eigenvalues of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$ . The resulting error may already lead to quite large fluctuations of the classical estimators.

This observation strongly suggests that the classical techniques which assume  $n \gg N$  must be revisited, as will be detailed in Section III-D. In the next section, we go a step further and introduce a generalization of the Marčenko-Pastur law for more structured random matrices.

### C. Spectrum of large matrices

As should have become clear from the discussion above, a central quantity of interest in many signal processing applications is the *sample covariance matrix* of  $n$  realizations  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$  of a process  $\mathbf{y}_t = \mathbf{T}\mathbf{x}_t$  with  $\mathbf{x}_t$  having i.i.d. zero mean and unit variance entries. The realizations may be independent, in which case  $\mathbf{Y} = \mathbf{T}\mathbf{X}$ ,  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  with i.i.d. entries, or have a (linear) time dependence structure, in which case  $\mathbf{y}_t$  is often modelled as an autoregressive moving-average (ARMA) process; if so, we can write  $\mathbf{Y} = \mathbf{T}\mathbf{Z}\mathbf{R}$

<sup>3</sup>Those fixed-point equations, for  $z < 0$ , are often *standard interference functions*, as described in [26], allowing for simple algorithmic methods to obtain the solutions.

with  $\mathbf{Z}$  a random matrix with i.i.d. zero mean entries and  $\mathbf{R}$  a deterministic time correlation matrix (generally Toeplitz). The limiting eigenvalue distribution of many of such random matrix structures have been studied in the past ten years and have led to many fundamental results which only recently have found their way to signal processing applications.

In the following, we treat the simplest case of a sample covariance matrix based on  $n$  independent observations of the process  $\mathbf{y}_t$  described above, when  $\mathbf{x}_t$  has i.i.d. entries with zero mean. The outline of the techniques below is in general the same when it comes to more complicated models.

We start with the important generalization of the Marčenko-Pastur law to matrices with left-sided correlation, whose general proof is due to Bai and Silverstein [20].

*Theorem 1:* Consider the matrix

$$\hat{\mathbf{T}} = \mathbf{T}^{\frac{1}{2}} \mathbf{X} \mathbf{X}^H \mathbf{T}^{\frac{1}{2}} = \sum_{i=1}^n \mathbf{T}^{\frac{1}{2}} \mathbf{x}_i \mathbf{x}_i^H \mathbf{T}^{\frac{1}{2}}$$

where  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{C}^{N \times n}$  has i.i.d. entries with zero mean and variance  $1/n$ , and  $\mathbf{T} \in \mathbb{C}^{N \times N}$  is nonnegative Hermitian whose eigenvalue distribution  $F^{\mathbf{T}}$  converges weakly to  $F^T$  as  $N \rightarrow \infty$ . Then, the eigenvalue distribution  $F^{\hat{\mathbf{T}}}$  of  $\hat{\mathbf{T}}$  converges weakly and almost surely to the distribution  $F$  with Stieltjes transform  $m_F(z) = cm_{\underline{F}}(z) + (c-1)\frac{1}{z}$ , where  $m_{\underline{F}}(z)$  is solution to

$$m_{\underline{F}}(z) = - \left( z - c \int \frac{t}{1 + tm_{\underline{F}}(z)} dF^T(t) \right)^{-1} \quad (7)$$

for all  $z \in \mathbb{C}^+ \triangleq \{z \in \mathbb{C}, \Im[z] > 0\}$ .

Note here that, contrary to the case treated previously where  $\mathbf{T} = \mathbf{I}_N$ , the limiting eigenvalue distribution of  $\hat{\mathbf{T}}$  does not take an explicit form, but is only defined *implicitly* through its Stieltjes transform which satisfies a fixed-point equation. This result is nonetheless usually sufficient to derive *explicit* detection tests and estimators for our application needs. Although not mentioned for readability in the statement of the theorem, it is usually true that the fixed-point equation has a unique solution on some restricted set, and that classical fixed-point algorithms do converge surely to the proper solution; this is particularly convenient for numerical evaluations.

An important remark for the subsequent discussion is that the integral in (7) can be easily related to the Stieltjes transform of  $m_T$  of  $F^T$  to give the equivalent representation

$$m_{\underline{F}}(z) = - \left( z - \frac{c}{m_{\underline{F}}(z)} \left[ 1 - \frac{1}{m_{\underline{F}}(z)} m_T \left( -\frac{1}{m_{\underline{F}}(z)} \right) \right] \right)^{-1}. \quad (8)$$

Although the original formula of [20] is given by (7), Equation (8) is more interesting in several aspects. The important observation in (8) is that the Stieltjes transform  $m_T$  can be related to the limiting Stieltjes transform  $m_{\underline{F}}$  of  $\hat{\mathbf{T}}$  through this simple equation. This is the fundamental anchor for *inference techniques*, where the *observable data* ( $\hat{\mathbf{T}}$ ) get connected to the *hidden parameters* (information on  $\mathbf{T}$ ) to be estimated.

Indeed, in the sample covariance matrix model, eigen-inference methods used in subspace estimators consist in

retrieving information about the population covariance matrix  $\mathbf{T}$  from the sample covariance matrix  $\hat{\mathbf{T}}$ . Therefore, if one can relate the information to be estimated in  $\mathbf{T}$  to  $m_{\mathbf{T}}$ , then (8) provides the fundamental link between  $\mathbf{T}$  and  $\hat{\mathbf{T}}$  through their respective limiting Stieltjes transform, from which estimators can be obtained. This procedure is detailed in the next section.

Note that Theorem 1 provides the value of  $m_F(z)$  for all  $z \in \mathbb{C}^+$ . Therefore, if one desires to numerically evaluate the limiting eigenvalue distribution of  $\hat{\mathbf{T}}$ , it suffices to evaluate  $m_F(z)$  for  $z = x + i\varepsilon$ , with  $\varepsilon > 0$  small and for all  $x > 0$  and then use the inverse Stieltjes transform formula (4) to describe the limiting spectrum density  $f$  by

$$f(x) \simeq \frac{1}{\pi} m_F(x + i\varepsilon).$$

We used this technique to produce Figure 3 in which  $\mathbf{T} = \text{UPU}^H$  was assumed composed of three evenly weighted masses in  $\{1, 3, 7\}$  (top) or  $\{1, 3, 4\}$  (bottom).

As mentioned above, it is possible to derive the limiting eigenvalue distribution of a wide range of random matrix models or, if a limit does not exist for the model (this may arise for instance if the eigenvalues of  $\mathbf{T}$  are not assumed to converge in law), to derive *deterministic equivalents* for the large dimensional matrices. For the latter, the eigenvalue distribution  $F^{\hat{\mathbf{T}}}$  of  $\hat{\mathbf{T}}$  can be approximated for each  $N$  by a *deterministic* distribution  $F_N$  such that  $F^{\hat{\mathbf{T}}} - F_N \Rightarrow 0$  almost surely. These distribution functions  $F_N$  usually have a very practical form for analysis; see e.g. [27]–[29] for examples of random matrix models that do not admit limits but only deterministic equivalents.

The knowledge of the limiting eigenvalue distribution of such matrices is not of particular interest to signal processing, but it is a required prior step before performing statistical inference and detection methods. In wireless communication problems however, such as that of evaluating the capacity

$$\begin{aligned} C &= \frac{1}{N} \log \det \left( \mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{H}\mathbf{H}^H \right) \\ &= \int_{\sigma^2}^{\infty} \left( \frac{1}{t} - \frac{1}{N} \text{tr}(\mathbf{H}\mathbf{H}^H + t\mathbf{I}_N)^{-1} \right) dt \end{aligned}$$

of a multi-antenna Gaussian channel  $\mathbf{H}$  with noise variance  $\sigma^2$ , these tools alone are fundamental, see e.g. [28]–[32]. In particular, we recognize in the above formula that the Stieltjes transform  $\frac{1}{N} \text{tr}(\mathbf{H}\mathbf{H}^H - z\mathbf{I}_N)^{-1}$  of  $\mathbf{H}\mathbf{H}^H$ , and not its eigenvalue distribution, is the quantity of central interest. This is probably the main explanation why wireless communication applications of random matrix results have appeared as early as in 1999 [33], [34], while applications for signal processing methods emerged only very recently.

In the following, we pursue our study of the spectral properties of large dimensional random matrices to applications in statistical inference and introduce the eigen-inference (or G-estimation) method.

#### D. G-estimation

As introduced in the previous section, one of the major objectives of random matrix theory for signal processing is to improve the statistical tests and inference techniques derived

under the assumption of infinitely many observations for scenarios where the system population size  $N$  and the number of observations  $n$  are of the same order of magnitude. For this, we will rely on the connections between the Stieltjes transforms of the population and sample covariance matrices.

The G-estimation method, originally due to Girko [8], intends to provide such  $(N, n)$ -consistent estimators. The general technique consists in a three-step approach along the following lines. First, the parameter to be estimated, call it  $\theta$ , is written under the form of a functional of the Stieltjes transform of the deterministic matrices of the model, say here  $\theta = f(m_{\mathbf{T}})$  for a population matrix  $\mathbf{T}$  with Stieltjes transform  $m_{\mathbf{T}}$ . Then,  $m_{\mathbf{T}}$  is connected to the Stieltjes transform  $m_{\hat{\mathbf{T}}}$  of an observed matrix  $\hat{\mathbf{T}}$  through a formula similar to that described earlier in (8), e.g.  $m_{\mathbf{T}} = g(m_{\underline{F}})$  for a certain function  $g$ . Finally, connecting the pieces together, we have a link between  $\theta$  and the observations  $\hat{\mathbf{T}}$ , this link being only exact asymptotically. For finite  $N, n$  dimensions, this naturally produces an (asymptotically)  $(N, n)$ -consistent estimator  $\hat{\theta} = g(m_{\hat{\mathbf{T}}})$ .

While the fundamental tool for connecting population and observation spaces is the Stieltjes transform, the second important tool, that will help connecting the parameter  $\theta$  to the Stieltjes transform  $m_{\mathbf{T}}$ , is the Cauchy integral formula and complex analysis in general.

For better understanding, we hereafter elaborate on a concrete example to describe the eigen-inference framework within the context of the estimation of the source powers discussed in Section II-B.

We recall that one observes a matrix of the type  $\mathbf{Y} = \text{UP}^{\frac{1}{2}}\mathbf{X}$ , where  $\mathbf{U} \in \mathbb{C}^{N \times N}$  is unitary,  $\mathbf{X} \in \mathbb{C}^{N \times n}$  is filled with i.i.d. entries with zero mean and unit variance and  $\mathbf{P}$  is diagonal with  $N_1$  entries equal to  $P_1$ , ..., and  $N_K$  entries equal to  $P_K$ , all values of  $P_i$  being distinct.

While the objective in the previous section was to characterize the asymptotic eigenvalue distribution of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  when  $\mathbf{P}$  is known, the target of this section is instead to infer  $P_1, \dots, P_K$  from the observation  $\mathbf{Y}$ . For this, we make the following fundamental observation. From Cauchy's complex integration formula [7],

$$P_k = -\frac{1}{2\pi i} \oint_{\mathcal{C}_k} \frac{z}{P_k - z} dz$$

for a complex positively oriented contour  $\mathcal{C}_k$  circling once around  $P_k$ . If  $\mathcal{C}_k$  does not enclose any of the other  $P_i$ ,  $i \neq k$ , then it is also true, again from Cauchy's integration formula, that

$$\begin{aligned} P_k &= -\frac{1}{2\pi i} \frac{N}{N_k} \oint_{\mathcal{C}_k} z \frac{1}{N} \sum_{i=0}^K N_i \frac{1}{P_i - z} dz \\ &= -\frac{1}{2\pi i} \frac{N}{N_k} \oint_{\mathcal{C}_k} z m_{\mathbf{P}}(z) dz. \end{aligned} \quad (9)$$

Hence, we are able to write  $P_k$  as a function of the Stieltjes transform of the non-observable matrix  $\mathbf{P}$ . The next step is to link  $m_{\mathbf{P}}$  to the Stieltjes transform of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$ . For this, we use (8) with  $\mathbf{T} = \text{UPU}^H$  (which has the same eigenvalue spectrum as  $\mathbf{P}$ ) and  $\underline{F}$  the limiting spectrum of  $\frac{1}{N}\mathbf{Y}^H\mathbf{Y}$ , which

after the variable change  $z = m_{\underline{F}}(u)$  leads to the elementary expression

$$P_k = \frac{1}{c} \frac{1}{2\pi i} \oint_{\mathcal{C}'_k} u \frac{m'_{\underline{F}}(u)}{m_{\underline{F}}(u)} du$$

for some contour  $\mathcal{C}'_k$  which happens to circle exactly around the cluster of eigenvalues of  $\hat{\mathbf{T}}$  associated to  $P_k$  only, and with  $m'_{\underline{F}}$  the complex derivative of  $m_{\underline{F}}(z)$  with respect to  $z$ . It is therefore fundamental that there exists a contour that circles around the cluster associated to  $P_k$  only. It may happen, depending on the ratio  $c$  and the values of the  $P_k$ , that the limiting spectrum of  $\hat{\mathbf{T}}$  generates less than  $K$  clusters, some clusters being associated with multiple power values, as depicted in the second graph of Figure 3. In this scenario, the above result *does not* hold and the technique collapses. In signal processing applications, this effect is linked to the problem of *source separation*. From now on, we therefore assume that the cluster of eigenvalues associated to  $P_k$  is perfectly isolated. Necessary and sufficient conditions for this to hold are clearly established [57].

Under the cluster separability assumption, an estimator  $\hat{P}_k$  for  $P_k$  is now straightforward to obtain. Indeed, replacing  $m_{\underline{F}}$  by its empirical estimate  $m_{\frac{1}{N}\mathbf{Y}\mathbf{Y}^H}$ ,  $P_k$  is approximately equal to a complex integral whose integrand is constituted of rational functions. A final step of residue calculus [7] completes the calculus and we obtain explicitly the estimator

$$\hat{P}_k = \frac{1}{c_k} \sum_{m \in \mathcal{N}_k} (\lambda_m - \mu_m) \quad (10)$$

where  $\mathcal{N}_k$  is defined as in (3),  $\lambda_1 < \dots < \lambda_N$  are the eigenvalues of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$ , and  $\mu_1 < \dots < \mu_N$  are the ordered eigenvalues of  $\text{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^T$ , with  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)^T$ . This estimator therefore improves the  $n$ -consistent estimator (3) to an  $(N, n)$ -consistent estimator which surprisingly turns out to have much better performance statistics for all couples  $(N, n)$ . Performance figures will be provided in Section IV for signal processing scenarios of more practical interest.

Multiple estimators can be derived similarly for different types of models and problems, a large number of which have been investigated by Girko and named the G-estimators. A list of more than fifty estimators for various applications can be found in [35]. In Section IV, some examples in the context of path loss estimation in wireless communications and angle of arrival estimation for array processing will be discussed.

We now turn to a finer analysis of random matrices which is no longer concerned with the weak limit of eigenvalue distributions but rather with the limiting fluctuations of individual eigenvalues (in particular the extreme eigenvalues). This problem is at the core of some standard statistical tests for signal detection, such as the generalized likelihood ratio test (GLRT).

### E. Extreme eigenvalues

It is important to note at this point that Theorem 1 and all theorems deriving from the Stieltjes transform method only provide the weak convergence of the empirical eigenvalue distribution to a limit  $F$  but do not say anything about the

behaviour of the individual eigenvalues. In particular, the fact that the support of  $F$  is compact does not mean that the empirical eigenvalues will asymptotically *all* lie in this compact support. Indeed, consider for instance the distribution  $F_N(x) = \frac{N-1}{N}1_{x \leq 1} + \frac{1}{N}1_{x \leq 2}$ . Then the largest value taken by a random variable distributed as  $F_N$  equals 2 for all  $N$  although it is clear that  $F_N$  converges weakly to  $F = 1_{x \leq 1}$  (and then the support of  $F$  is the singleton  $\{1\}$ ). In the context of Theorem 1, this reasoning indicates that, although the weak limit of  $F^{\hat{\mathbf{T}}}$  is compactly supported, some or even many eigenvalues of  $F^{\hat{\mathbf{T}}}$  may still be found outside the support.

As a matter of fact, it is proved in [36] that, if the entries of  $\mathbf{X}$  in Theorem 1 have infinite fourth order moment, then the largest eigenvalue of  $\mathbf{X}\mathbf{X}^H$  tends to infinity. For non-degenerate scenarios though, i.e. when the fourth moment of the entries of  $\mathbf{X}$  is finite, we have the following much expected result [37], [38].

*Theorem 2 (No eigenvalue outside the support):* Let  $\hat{\mathbf{T}}$  be defined as in Theorem 1 with  $\mathbf{X}$  having entries with finite fourth order moment and with  $F^{\mathbf{T}} \Rightarrow F^T$ , such that  $\mathbf{T}$  has no eigenvalue outside the support of  $F^T$ . We recall that  $F^{\hat{\mathbf{T}}} \Rightarrow F$  for  $F$  defined in Theorem 1. Take now  $[a, b] \subset \mathbb{R} \cup \{\pm\infty\}$  strictly away from the support of  $F$ . Then, almost surely, there is no eigenvalue of  $\hat{\mathbf{T}}$  found in  $[a, b]$  for all large  $N$ .

This result says in particular that, for  $\mathbf{T} = \mathbf{I}_N$ , and for all large  $N$ , there is no eigenvalue of  $\hat{\mathbf{T}}$  found away from the support of the Marčenko-Pastur law. This further generalizes to the scenario where  $\mathbf{T}$  has a few distinct eigenvalues, each with large multiplicities, in which case  $F$  is formed of multiple compact clusters, as depicted in Figure 3. In this context, it is proved in [39] that the number of eigenvalues asymptotically found in each cluster matches exactly the multiplicity of the corresponding mass in  $\mathbf{T}$  (assuming cluster separability). This is often referred to as the property of *exact spectrum separation*. Note that this point fully justifies the last step in the derivation of the estimator (10) where we somewhat hid the fact that the residue calculus naturally imposes to know where the eigenvalues of  $\hat{\mathbf{T}}$  are precisely located. As we will see in Section IV, these results on the asymptotic location of the eigenvalues allow us to derive new hypothesis tests for the detection of signals embedded in white noise. In particular, we will present tests on the received sample covariance matrix  $\hat{\mathbf{T}}$  that confront a pure white noise against a signal-plus-noise hypotheses, i.e.  $\mathbf{T} = \mathbf{I}_N$  against  $\mathbf{T} \neq \mathbf{I}_N$ .

Based on the earlier reasoning on the weak limit interpretations, we insist that, while Theorem 1 does state that  $F^{\mathbf{T}} \Rightarrow 1_{x \leq 1}$  implies that  $F$  is the Marčenko-Pastur law, it does *not* state that  $F^{\hat{\mathbf{T}}} \Rightarrow 1_{x \leq 1}$  implies that no eigenvalue is asymptotically found away from the support of the Marčenko-Pastur law. Indeed, if  $\mathbf{T} = \text{diag}(1, \dots, 1, a)$ , with  $a \neq 1$ , the theorem cannot be applied. Conversely, for  $\mathbf{T}$  described above, it is not clear whether an eigenvalue of  $\hat{\mathbf{T}}$  will be seen outside the limiting support. This interrogation is of fundamental importance for the performance evaluation of asymptotic detection tests and has triggered a recent interest for this particular model for  $\mathbf{T}$ , called the *spike model*, which will be discussed next.

### F. Spike models

It is of particular interest for signal processing applications to model low rank signal spaces embedded in white noise. This is the case in array processing where a single signal source with unique propagation path is received during several consecutive time instants, hence producing a rank-one signal matrix to be added to the ambient full rank white noise. The objective of this particular model is to derive simple detection and identification procedures that do not fall in the rather involved scheme of Section III-D.

Under the generic denomination of “spike models” we understand all random matrix models showing small rank perturbations of some classical random matrix models. For instance, for  $\mathbf{X} \in \mathbb{C}^{N \times n}$  with i.i.d. entries with zero mean and variance  $1/n$ , the matrices  $\hat{\mathbf{T}} = (\mathbf{X} + \mathbf{E})(\mathbf{X} + \mathbf{E})^H$ ,  $\mathbf{E} \in \mathbb{C}^{N \times n}$  such that  $\mathbf{E}\mathbf{E}^H$  is of small rank  $K$ , or  $\hat{\mathbf{T}} = (\mathbf{I}_N + \mathbf{E})^{\frac{1}{2}} \mathbf{X}\mathbf{X}^H (\mathbf{I}_N + \mathbf{E})^{\frac{1}{2}}$ ,  $\mathbf{E} \in \mathbb{C}^{N \times N}$  of small rank  $K$ , fall into the generic scheme of spike models. For all of these models, Theorem 1 and similar results, see e.g. [40], claim that  $F^{\hat{\mathbf{T}}}$  converges weakly to the Marčenko-Pastur law, which is the major motivation of these models as they are rather simple to analyze. The interest here though is to study the behaviour of the extreme eigenvalues of  $\hat{\mathbf{T}}$ .

To carry on with the same models as before, we concentrate here on  $\hat{\mathbf{T}}$  modeled as in Theorem 1, with  $\mathbf{T}$  taken to be a small rank perturbation of the identity matrix. Similar properties can be found for other models e.g. in [41], [42]. The first result deals with first order limits of eigenvalues and eigenvector projections of the spike model [41], [43], [44].

*Theorem 3:* Let  $\hat{\mathbf{T}}$  be defined as in Theorem 1,  $\mathbf{X}$  have i.i.d. entries with zero mean and variance  $1/n$ , and  $\mathbf{T} = \mathbf{I}_N + \mathbf{E}$  with  $\mathbf{E} = \sum_{i=1}^K \omega_i \mathbf{u}_i \mathbf{u}_i^H$  its spectral decomposition (i.e.  $\mathbf{u}_i^H \mathbf{u}_j = \delta_i^j$ ), where  $K$  is fixed and  $\omega_1 > \dots > \omega_K > 1$ . Denote  $\lambda_1 \geq \dots \geq \lambda_N$  the ordered eigenvalues of  $\hat{\mathbf{T}}$  and  $\hat{\mathbf{u}}_i \in \mathbb{C}^N$  the eigenvector associated with the eigenvalue  $\lambda_i$ . Then, for  $1 \leq i \leq K$ , as  $N, n \rightarrow \infty$  with  $N/n \rightarrow c$ ,

$$\lambda_i \xrightarrow{\text{a.s.}} \begin{cases} (1 + \sqrt{c})^2 & , \text{ if } \omega_i \leq \sqrt{c} \\ \rho_i \triangleq 1 + \omega_i + c \frac{1 + \omega_i}{\omega_i} & , \text{ if } \omega_i > \sqrt{c} \end{cases}$$

and

$$|\mathbf{u}_i^H \hat{\mathbf{u}}_i| \xrightarrow{\text{a.s.}} \begin{cases} 0 & , \text{ if } \omega_i \leq \sqrt{c} \\ \xi_i \triangleq \frac{1 - c\omega_i^{-2}}{1 + c\omega_i^{-1}} & , \text{ if } \omega_i > \sqrt{c}. \end{cases}$$

The main observation of this result is that, if  $\omega_1 \leq \sqrt{c}$ , then the largest eigenvalue of  $\hat{\mathbf{T}}$  asymptotically converges to the right edge of the Marčenko-Pastur law and is therefore *hidden* in the main bulk of the eigenvalues, while if  $\omega_1 > \sqrt{c}$ , the largest eigenvalue of  $\hat{\mathbf{T}}$  is found outside the main cluster. This has fundamental consequences. When  $\hat{\mathbf{T}}$  is the sample covariance matrix of signal-plus-noise data with signal strength  $\omega_1$ , then, if the signal is strong enough or conversely if  $N/n$  is small enough, it is possible to detect this signal based on the presence of an eigenvalue exterior to the main eigenvalue cluster of  $\hat{\mathbf{T}}$ . Otherwise, neither the eigenvalues nor the eigenvectors of  $\hat{\mathbf{T}}$  provide any information on the presence of a signal, at least asymptotically. As a conclusion,

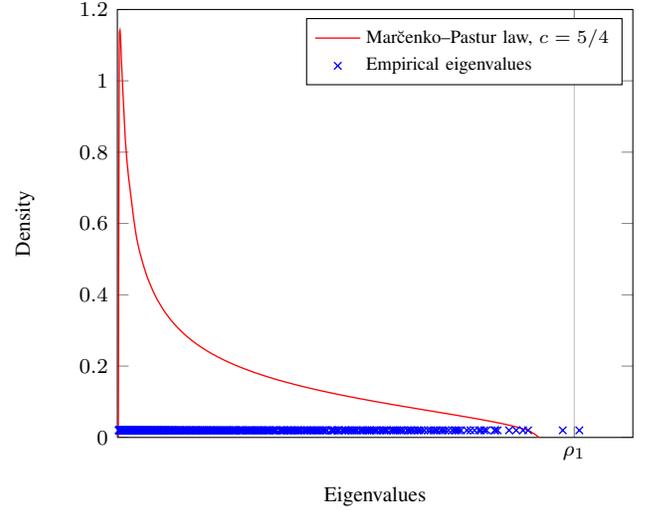


Fig. 4. Eigenvalues of  $\hat{\mathbf{T}} = \mathbf{T}^{\frac{1}{2}} \mathbf{X}\mathbf{X}^H \mathbf{T}^{\frac{1}{2}}$ , where  $\mathbf{T}$  is a diagonal of ones but for the first four entries set to  $\{3, 3, 2, 2\}$ ,  $N = 500$ ,  $n = 400$ . The theoretical limiting eigenvalue of  $\hat{\mathbf{T}}$  is emphasized.

in the latter scenario, there is no way to decide if a signal is indeed present or not. This provides *fundamental limits* of signal detection tests. This subject is further discussed in the application Section IV.

In Figure 4, we depict the eigenvalues of a spike model as in Theorem 3, with four population eigenvalues greater than one. Two of them exceed the detectability threshold  $\omega_i > \sqrt{c}$ , while the other two do not. As expected, only two eigenvalues are visible outside the bulk, with values close to their theoretical limit.

Although this was not exactly the approach followed initially in [43], the same Stieltjes transform and complex integration framework can be used to prove this result. We hereafter give a short description of this proof in the simpler case where  $\mathbf{E} = \omega \mathbf{u}\mathbf{u}^H$  and  $\omega > \sqrt{c}$ .

1) *Extreme eigenvalues:* By definition of an eigenvalue,  $\det(\hat{\mathbf{T}} - z\mathbf{I}_N) = 0$  for  $z$  an eigenvalue of  $\hat{\mathbf{T}}$ . Some algebraic manipulation based on determinant product formulas and Woodbury’s identity [22] lead to

$$\det(\hat{\mathbf{T}} - z\mathbf{I}_N) = f_N(z) \det(\mathbf{I}_N + \mathbf{E}) \det(\mathbf{X}\mathbf{X}^H - z\mathbf{I}_N) \\ \text{with } f_N(z) = 1 + z \frac{\omega}{1 + \omega} \mathbf{u}^H (\mathbf{X}\mathbf{X}^H - z\mathbf{I}_N)^{-1} \mathbf{u}.$$

An eigenvalue of  $\hat{\mathbf{T}}$  not inside the main cluster cannot cancel the right-hand determinant in the first line and must therefore cancel  $f_N(z)$ . Standard random matrix lemmas then show that, since  $\mathbf{u}^H (\mathbf{X}\mathbf{X}^H - z\mathbf{I}_N)^{-1} \mathbf{u}$  is asymptotically close to  $m_F(z)$ , the Stieltjes transform of the Marčenko-Pastur law,

$$f_N(z) \xrightarrow{\text{a.s.}} f(z) \triangleq 1 + z \frac{\omega}{1 + \omega} m_F(z).$$

Substituting  $m_F(z)$  by its exact formulation (that is, the Stieltjes transform of the Marčenko-Pastur law), we then obtain that  $f(z) = 0$  is equivalent to  $z = 1 + \omega + c \frac{1 + \omega}{\omega}$  if  $\omega > \sqrt{c}$  and has no solution otherwise, which is the expected result.

2) *Eigenvector projections*: For the result on eigenvector projections, we use again the Cauchy integration scheme which states that, for any deterministic vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{C}^N$  and for all large  $N$ ,

$$\mathbf{a}^H \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^H \mathbf{b} = -\frac{1}{2\pi i} \oint_{\mathcal{C}} \mathbf{a}^H (\hat{\mathbf{T}} - z \mathbf{I}_N)^{-1} \mathbf{b} \quad (11)$$

where  $\mathcal{C}$  is a positively oriented contour circling around  $\omega$  and excluding 1, and where we again recognize a quadratic form involving the resolvent of  $\hat{\mathbf{T}}$ . This is an immediate consequence of the fact that  $(\hat{\mathbf{T}} - z \mathbf{I}_N)^{-1}$  has a pole in  $\lambda_1$  with residue  $-\hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^H$ . We then use the same matrix manipulations as in the previous section (by isolating the term  $\mathbf{X}\mathbf{X}^H$ ) to finally obtain the result of Theorem 3 after residue calculus.

These first order results are important since they suggest detection and identification tests in signal processing scenarios with extremely large arrays. However, for realistic small system sizes, these results are not sufficient to provide efficient statistical tests. To overcome this limitation, we need to go beyond the first order limits and characterize the fluctuations of the extreme eigenvalues and eigenvector projections. Recent works have provided such statistics for the results of Theorem 3 as well as for the fluctuations of the estimator (10), which we introduce presently.

### G. Fluctuations

We first mention that the fluctuations of functionals of  $F^{\hat{\mathbf{T}}} - F$  in Theorem 1 have been derived in [45]. More precisely, for any well-behaved function  $f$  (at least holomorphic on the support of  $F$ ) and under some mild technical assumptions,

$$N \int f(t) \left( dF^{\hat{\mathbf{T}}} - dF(t) \right) \Rightarrow \mathcal{N}(0, \sigma^2)$$

with  $\sigma^2$  known. This in particular applies to  $f(t) = (t - z)^{-1}$  for  $z \in \mathbb{C} \setminus \mathbb{R}^+$ , which gives the asymptotic fluctuations of the Stieltjes transform of  $\hat{\mathbf{T}}$ . From there, classical asymptotic statistics tools such as the delta-method [1] allow one to transfer the Gaussian fluctuations of the Stieltjes transform  $m_{\hat{\mathbf{T}}}$  to the fluctuations of any function of it, e.g. estimators based on  $m_{\hat{\mathbf{T}}}$ . The following result on the fluctuations of the estimator (10) therefore comes with no surprise [46].

*Theorem 4*: Consider the estimator  $\hat{P}_k$  in (10). Assuming the entries of  $\mathbf{X}$  have finite fourth order moment (to ensure Theorem 2) and that  $P_k$  generates an isolated cluster,

$$N(\hat{P}_k - P_k) \Rightarrow \mathcal{N}(0, \sigma^2)$$

where  $\sigma^2$  is evaluated precisely in [46] as a complex integral form involving derivatives of the limiting Stieltjes transform of  $\hat{\mathbf{T}}$ .

In the case of the spike model, it is of interest to derive the fluctuations of the extreme eigenvalues and eigenvector projections, whose limits were given in Theorem 3. Surprisingly, the fluctuations of these variables are not always Gaussian, [44], [47], [48].

*Theorem 5*: Let  $\hat{\mathbf{T}}$  be defined as in Theorem 3. Then, for  $1 \leq k \leq K$ , if  $\omega_k < \sqrt{c}$ ,

$$N^{\frac{2}{3}} \frac{\lambda_k - (1 + \sqrt{c})^2}{(1 + \sqrt{c})^{\frac{4}{3}} \sqrt{c}} \Rightarrow T_2$$

where  $T_2$  is known as the *complex Tracy-Widom* distribution, described in [49] as the solution of a Painlevé differential equation. If, on the other hand,  $\omega_k > \sqrt{c}$ , then

$$\sqrt{N} \begin{pmatrix} |\mathbf{u}_k^H \hat{\mathbf{u}}_k|^2 - \xi_k \\ \lambda_k - \rho_k \end{pmatrix} \Rightarrow \mathcal{N}(0, \boldsymbol{\Sigma}_k)$$

where the entries of the  $2 \times 2$  matrix  $\boldsymbol{\Sigma}_k$  can be expressed in a simple closed form as a function of  $\omega_k$  only. Moreover, for  $k \neq k'$  such that  $\omega_k, \omega_{k'} > \sqrt{c}$  and distinct, the centered-scaled eigenvalues and eigenvector projections are asymptotically independent.

The second result of Theorem 5 can be obtained using the same Stieltjes transform and complex integration framework as described in the sketch of proof of Theorem 3, see [44] for details. Again in that case, these are mainly the fluctuations of the Stieltjes transform of  $\mathbf{X}\mathbf{X}^H$  and an application of the delta-method which lead to the result. The first result of Theorem 5 is proved with very different methods which we do not introduce here (since the extreme eigenvalues are inside the base support, complex contour integration approaches cannot be performed). These methods involve techniques such as orthogonal polynomials and Fredholm determinants which rely on the study of the *finite* dimensional distribution of the eigenvalues of  $\mathbf{X}\mathbf{X}^H$  before ultimately providing asymptotic results, see [24], [50] for details.

The Tracy-Widom distribution is depicted in Figure 5. It is interesting to see that the Tracy-Widom law is centered on a negative value and that the probability for positive values is low. This suggests that, if the largest eigenvalue of  $\hat{\mathbf{T}}$  is greater than  $(1 + \sqrt{c})^2$ , then it is very likely that  $\mathbf{T}$  is not the identity matrix. A similar remark can be made for the smallest eigenvalue of  $\hat{\mathbf{T}}$  which has a mirrored Tracy-Widom fluctuation [51], when  $c < 1$ .

Theorem 5 allows for asymptotically accurate test statistics for the decision on signal-plus-noise against pure noise models, thanks to the result on the extreme eigenvalues. Moreover, the results on the eigenvector projections provide even more information to the observer. In the specific context of failure diagnosis [44], introduced in Section IV, this can be used for the identification of the (asymptotically) most likely assumption from a set of  $M$  failure models of the type  $(\mathbf{I}_N + \mathbf{E}_k)^{\frac{1}{2}} \mathbf{X}\mathbf{X}^H (\mathbf{I}_N + \mathbf{E}_k)^{\frac{1}{2}}$ , for  $k \in \{1, \dots, M\}$ .

### H. Practical limitations

For all techniques derived above, it is important to keep in mind that the results are only valid under the assumption of large matrix dimensions. As it turns out from convergence speed considerations, estimators based on weak convergence properties of the eigenvalue distribution are accurate even for very small system sizes ( $N = 8$  is often sufficient, if not less), while second order statistics of functionals of the eigenvalue distribution roughly require  $N$  to be at least of order 64. When

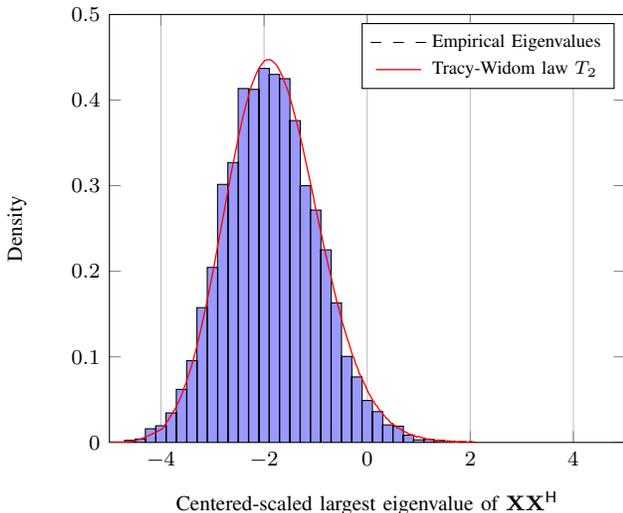


Fig. 5. Distribution of  $N^{\frac{2}{3}}c^{-\frac{1}{2}}(1+\sqrt{c})^{-\frac{4}{3}}[\lambda_1 - (1+\sqrt{c})^2]$  against the Tracy-Widom law for  $N = 500$ ,  $n = 1500$ ,  $c = 1/3$ , for the covariance matrix model  $\mathbf{X}\mathbf{X}^H$ ,  $\mathbf{X} \in \mathbb{C}^{N \times n}$  with Gaussian i.i.d. entries. Empirical distribution from 10,000 Monte-Carlo simulations.

it comes to asymptotic properties of the extreme eigenvalues, due to the loss of the collective averaging effect of all the eigenvalues, the convergence speed is much slower so that  $N = 16$  is often a minimum for first order convergence, while  $N$  of order 128 is required for asymptotic second order statistics to become accurate. Note also that these empirical values assume “non-degenerated” conditions; for instance, for  $N = 128$ , the second order statistics of the largest eigenvalue in a spike model with population eigenvalue  $\omega = \sqrt{c} + \varepsilon$  for a small  $\varepsilon > 0$  are often far from Gaussian with zero mean (since they should be simultaneously close to follow the very negatively-centered Tracy-Widom distribution) so that, depending on how small  $\varepsilon$  is chosen, much larger  $N$  are needed for an accurate Gaussian approximation to arise. This suggests that all the above methods have to be manipulated with extreme care depending on the application at hand.

The results introduced so far are only a small subset of the important contributions resulting from more than ten years of applied random matrix theory. A large exposition of sketches of proofs, main strategies to address applied random matrix problems, as well as a large amount of applications are analyzed in the book [25] in the joint context of wireless communications and signal processing. An extension of some notions introduced in the present tutorial can also be found in [52]. Deeper random matrix consideration about the Stieltjes transform approach can be found in [23] and references therein, while discussions on extreme eigenvalues and the tools needed to derive the Tracy-Widom fluctuations, based on the theory of orthogonal polynomials and Fredholm determinants, can be found in e.g. [24], [50], [53] and references therein. In the next section, we discuss several application examples, already partly introduced above, which apply Theorems 1-5 in various contexts of hypothesis testing and subspace statistical inference.

## IV. CASE STUDIES

In this section, several classical detection and estimation schemes are revisited for the regime  $N \simeq n$  studied in Section III.

### A. Multi-source power inference in i.i.d. channels

Our first case study originates from the field of wireless communication, closely related to the previous example of eigenvalue inference. We consider a cognitive sensing scenario [54] in which an  $N$ -antenna sensing device (a mobile terminal for instance) collects data from  $K$  multi-antenna transmitters. Transmitter  $k$  is equipped with  $M_k$  antennas and sends the signals  $\mathbf{x}_k(t) \in \mathbb{C}^{M_k}$  to the receiver at time  $t$  through the channel  $\mathbf{H}_k \in \mathbb{C}^{N \times M_k}$ , containing i.i.d. Gaussian entries with zero mean and variance  $1/N$ .<sup>4</sup> We assume  $M_1 + \dots + M_K \triangleq M \leq N$  and that  $\mathbf{H} \triangleq [\mathbf{H}_1, \dots, \mathbf{H}_K]$  remains static during at least  $n$  symbol periods. The receiver also captures white Gaussian noise  $\sigma\mathbf{w}(t)$  with zero mean and variance  $\sigma^2$ . The receive signal  $\mathbf{y}(t) \in \mathbb{C}^N$  at time  $t$  is then modelled as

$$\mathbf{y}(t) = \sum_{i=1}^K \sqrt{P_i} \mathbf{H}_i \mathbf{x}_i(t) + \sigma\mathbf{w}(t).$$

Gathering  $n$  realizations of  $\mathbf{y}(t) \in \mathbb{C}^N$  in the matrix  $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(n)]$ ,  $\mathbf{Y}$  takes the form

$$\mathbf{Y} = (\mathbf{H}\mathbf{P}^{\frac{1}{2}} \quad \sigma\mathbf{I}_N) \begin{pmatrix} \mathbf{X} \\ \mathbf{W} \end{pmatrix}$$

where  $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(n)]$ ,  $\mathbf{x}(t) = [\mathbf{x}_1(t)^T, \dots, \mathbf{x}_K(t)^T]^T$ ,  $\mathbf{P} = \text{diag}(P_1\mathbf{I}_{M_1}, \dots, P_K\mathbf{I}_{M_K})$ , and  $\mathbf{W} = [\mathbf{w}(1), \dots, \mathbf{w}(n)]$ . We wish to estimate the power values  $P_1, \dots, P_K$  from the observation  $\mathbf{Y}$ . We see that the sample covariance matrix approach developed in Section III-D is no longer valid as the population matrix  $\mathbf{T} \triangleq \mathbf{H}\mathbf{P}\mathbf{H}^H + \sigma^2\mathbf{I}_N$  is now random.

Following the general methodology developed in Section III-D, assuming the asymptotic cluster separation property for a certain  $P_k$ , we first use the connection between  $P_k$  and  $m_P(z)$ , the limiting Stieltjes transform of  $\mathbf{P}$ , given by (9). The innovation from the study in Section III-D is that the link between  $m_P$  and  $m_F$ , the limiting Stieltjes transform of  $\frac{1}{N}\mathbf{Y}^H\mathbf{Y}$ , is more involved and requires some more random matrix arguments than Theorem 1 alone. This link, along with a generalization of Theorem 2 to account for the randomness in  $\mathbf{T}$ , induces the following  $(N, n)$ -consistent estimator  $\hat{P}_k$  of  $P_k$  [44]

$$\hat{P}_k = \frac{Nn}{M_k(n-N)} \sum_{i \in \mathcal{N}_k} (\eta_i - \mu_i)$$

where  $\mathcal{N}_k$  is the set of indexes of the eigenvalues of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  associated to  $P_k$ ,  $\eta_1 < \dots < \eta_N$  are the eigenvalues of  $\text{diag}(\boldsymbol{\lambda}) - \frac{1}{N}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^T$  and  $\mu_1 < \dots < \mu_N$  are the eigenvalues of  $\text{diag}(\boldsymbol{\lambda}) - \frac{1}{n}\sqrt{\boldsymbol{\lambda}}\sqrt{\boldsymbol{\lambda}}^T$ , where  $\lambda_1 < \dots < \lambda_N$  are the ordered eigenvalues of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$ .

In Figure 6, the performance comparison between the classical large  $n$  approach and the Stieltjes transform approach for

<sup>4</sup>The normalization of the channel entries by  $1/N$  along with  $\|\mathbf{x}_k(t)\|$  bounded allows one to keep the transmit signal power bounded as  $N$  grows.

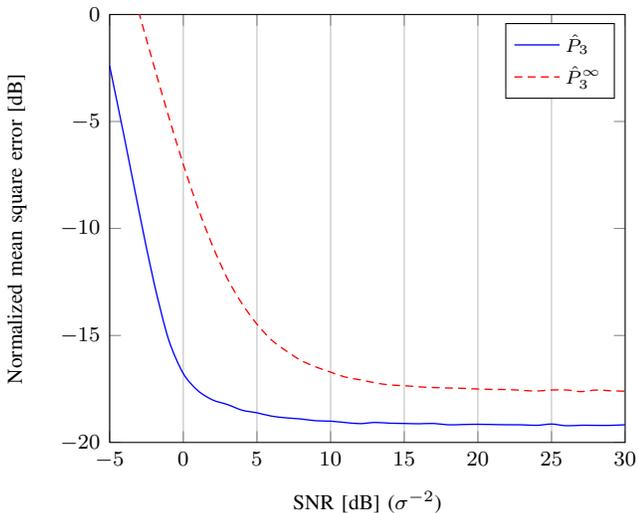


Fig. 6. Normalized mean square error of largest estimated power  $P_3$  for  $P_1 = 1/16$ ,  $P_2 = 1/4$ ,  $P_3 = 1$ ,  $M_1 = M_2 = M_3 = 4$ ,  $M = 12$ ,  $N = 24$ ,  $n = 128$ . Comparison between conventional and Stieltjes transform approaches.

power estimation is depicted, based on 10,000 Monte Carlo simulations. Similar to the estimator (3), the classical approach of  $P_k$ , denoted  $\hat{P}_k^\infty$ , consists in assuming  $n \gg N$  and  $N \gg M_i$  for each  $i$  (recall that  $M_i$  is the multiplicity of  $P_i$ ). A clear performance gain is observed for the whole range of signal-to-noise ratios (SNR), defined as  $\sigma^{-2}$ , although a saturation level is observed which translates the approximation error due to the asymptotic analysis. For low SNR, an *avalanche* effect is observed, which is caused by the inaccuracy of the Stieltjes transform approach when clusters tend to merge (here the cluster associated to the value  $\sigma^2$  grows large as  $\sigma^2$  increases and covers the clusters associated to  $P_1$ , then  $P_2$ , and finally  $P_3$ ). Nonetheless, this apparently strong cluster separability limitation generates an avalanche below the SNR level of the well-known avalanche effect produced by the classical large  $n$  estimator. The random matrix method therefore yields a twofold performance gain compared to the classical method, as it is both more accurate (which was expected) and is also more robust to noise which is an interesting, not necessarily anticipated, outcome.

Our next example is based on a contribution from Mestre [2] which initiated a series of statistical inference methods exploiting complex integration, among which the source detection methods described above.

### B. G-MUSIC

The MUSIC method, originally due to Schmidt and later extended by McCloud and Scharf [55], [56], is a direction of arrival detection technique based on a subspace approach. The scenario consists of an array of  $N$  sensors (e.g. a radar antenna) receiving the signals originating from  $K$  sources at angles  $\theta_1, \dots, \theta_K$  with respect to the array. The objective is to estimate those angles. The data vector  $\mathbf{y}(t)$  received at time

$t$  by the array can be written as

$$\mathbf{y}(t) = \sum_{k=1}^K \mathbf{s}(\theta_k) x_k(t) + \sigma \mathbf{w}(t)$$

where  $\mathbf{s}(\theta) \in \mathbb{C}^N$  is a deterministic vector-valued function of the angle  $\theta \in [0, 2\pi)$ ,  $x_k(t) \in \mathbb{C}$  is the standard Gaussian i.i.d. data sent by source  $k$  at time  $t$  and  $\mathbf{w}(t) \in \mathbb{C}^N$  is the additive standard Gaussian noise. The vector  $\mathbf{y}(t)$  is Gaussian with covariance

$$\mathbf{T} = \mathbf{S}(\Theta) \mathbf{S}(\Theta)^H + \sigma^2 \mathbf{I}_N$$

where  $\mathbf{S}(\Theta) = [\mathbf{s}(\theta_1), \dots, \mathbf{s}(\theta_K)] \in \mathbb{C}^{N \times K}$ . We denote as usual  $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(n)] \in \mathbb{C}^{N \times n}$  the concatenated matrix of  $n$  independent observations.

We call  $\omega_1 \leq \dots \leq \omega_N$  the eigenvalues of  $\mathbf{T}$  and  $\mathbf{u}_1, \dots, \mathbf{u}_N$  their associated eigenvectors. Similarly, we will denote  $\lambda_1 \leq \dots \leq \lambda_N$  the eigenvalues of  $\hat{\mathbf{T}} \triangleq \frac{1}{n} \mathbf{Y} \mathbf{Y}^H$ , with respective eigenvectors  $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_N$ . Assuming  $N \geq K$ , the smallest  $N - K$  eigenvalues of  $\mathbf{T}$  equal  $\sigma^2$  and we can represent  $\mathbf{T}$  under the form

$$\mathbf{T} = (\mathbf{U}_W \quad \mathbf{U}_S) \begin{pmatrix} \sigma^2 \mathbf{I}_{N-K} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_S \end{pmatrix} \begin{pmatrix} \mathbf{U}_W^H \\ \mathbf{U}_S^H \end{pmatrix}$$

$\mathbf{\Omega}_S = \text{diag}(\omega_{N-K+1}, \dots, \omega_N)$ ,  $\mathbf{U}_S = [\mathbf{u}_{N-K+1}, \dots, \mathbf{u}_N]$  the so-called *signal space* and  $\mathbf{U}_W = [\mathbf{u}_1, \dots, \mathbf{u}_{N-K}]$  the so-called *noise space*.

The basic idea of the MUSIC method is to observe that any vector lying in the signal space is orthogonal to the noise space. This leads in particular to

$$\eta(\theta_k) \triangleq \mathbf{s}(\theta_k)^H \mathbf{U}_W \mathbf{U}_W^H \mathbf{s}(\theta_k) = 0$$

for  $k \in \{1, \dots, K\}$ .

A natural estimator  $\hat{\theta}_k$  of  $\theta_k$  in the neighborhood of  $\theta_k$  consists in the minimal argument of

$$\hat{\eta}(\theta) \triangleq \mathbf{s}(\theta)^H \hat{\mathbf{U}}_W \hat{\mathbf{U}}_W^H \mathbf{s}(\theta)$$

where  $\hat{\mathbf{U}}_W = [\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_{N-K}]$  is the eigenvector space corresponding to the smallest  $N - K$  eigenvalues of  $\hat{\mathbf{T}}$ .

This estimator is however proven inconsistent for  $N, n$  growing large simultaneously in [57], which is now not surprising. To produce an  $(N, n)$ -consistent estimator, recall (11). Similar to the derivation of the eigenvector projection in Section III-F, we can write

$$\mathbf{s}(\theta)^H \mathbf{U}_W \mathbf{U}_W^H \mathbf{s}(\theta) = \frac{1}{2\pi i} \oint_{\mathcal{C}_k} \mathbf{s}(\theta)^H (\mathbf{T} - z \mathbf{I}_N)^{-1} \mathbf{s}(\theta) dz$$

for a contour  $\mathcal{C}_k$  circling around  $\sigma^2$  only. Connecting  $\mathbf{T}$  to  $\hat{\mathbf{T}}$  from a theorem similar to Theorem 1 and performing residue calculus, we finally obtain a good approximation of  $\eta(\theta)$ , and then an  $(N, n)$ -consistent estimator for the direction of arrival. This estimator consists precisely in determining the  $K$  deepest minima (zeros may not exist) of the function [2]

$$\mathbf{s}(\theta)^H \left( \sum_{i=1}^N \phi(i) \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \right) \mathbf{s}(\theta)$$

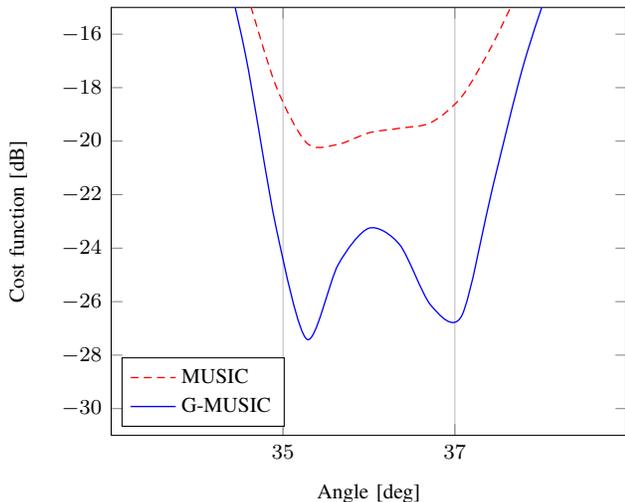


Fig. 7. MUSIC against G-MUSIC for DoA detection of  $K = 2$  signal sources,  $N = 20$  sensors,  $n = 150$  samples, SNR of 10 dB. Angles of arrival of  $35^\circ$  and  $37^\circ$ .

with  $\phi(i)$  defined as

$$\phi(i) = \begin{cases} 1 + \sum_{k=N-K+1}^N \left[ \frac{\lambda_k}{\lambda_i - \lambda_k} - \frac{\mu_k}{\lambda_i - \mu_k} \right], & i \leq N - K \\ - \sum_{k=1}^{N-K} \left[ \frac{\lambda_k}{\lambda_i - \lambda_k} - \frac{\mu_k}{\lambda_i - \mu_k} \right], & i > N - K \end{cases}$$

and where, similar to above,  $\mu_1 \leq \dots \leq \mu_N$  are the eigenvalues of  $\text{diag}(\boldsymbol{\lambda}) - \frac{1}{n} \sqrt{\boldsymbol{\lambda}} \sqrt{\boldsymbol{\lambda}}^\top$ , with  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)^\top$ ,  $\lambda_1 < \dots < \lambda_N$  the ordered eigenvalues of  $\hat{\mathbf{T}}$ .

Observe that, contrary to the classical MUSIC method, not only the noise subspace but all the eigenvectors of  $\hat{\mathbf{T}}$  are used, although they are divided in two subsets with different weight policies applied to the eigenvectors in each set. This result is generalized to the scenario where  $x_k(t)$  is non random. In this case, the technique involves different reference results than the sample covariance matrix theorems required here but the final estimator is obtained similarly [58].

In Figure 7, the performance comparison between the traditional MUSIC and the G-MUSIC algorithms is depicted, when two sources have close angles which the traditional MUSIC algorithm is not able to distinguish. It is clear on this single-shot realization that the G-MUSIC algorithm shows deeper minima and allows for a better source resolution.

Note that both inference methods described above, be it the power or angle of arrival estimation schemes, assume a priori knowledge of the cluster indexes to be able to implement the estimators. The proposed random matrix method therefore does not solve the *order selection* problem. New techniques are therefore being investigated that deal with the generalization of the Akaike principle [59] and the minimum description length technique [60] for improved order selection using random matrix theory, see e.g. [61].

Remark also that, so far, only regular sample covariance matrix models have been analyzed to address questions of array processing. In reality, due (in particular) to the non-Gaussian noise structure in radar detection, more advanced estimators of the population covariance matrix are used based on the celebrated Huber and Maronna robust M-estimators

[62], [63]. We are currently investigating these M-estimators within the random matrix framework. Further intricate models involving multi-path propagation and the technique consisting in stacking successive vector observations are also under study.

This completes the set of examples using results based on weak limits of large dimensional random matrices and G-estimation. We now turn to examples involving the results for the spike models, starting with simple signal detection procedures.

### C. Detection

Consider the hypothesis test which consists in deciding whether a received signal  $\mathbf{y}(t) \in \mathbb{C}^N$  consists of pure noise,  $\mathbf{y}(t) = \sigma \mathbf{w}(t)$ , with  $\mathbf{w}(t) \in \mathbb{C}^N$  with Gaussian entries with zero mean and unit variance, for some unknown  $\sigma$  (hypothesis  $\mathcal{H}_0$ ), or containing a signal plus noise,  $\mathbf{y}(t) = \mathbf{h}x(t) + \sigma \mathbf{w}(t)$ , for a time-independent vector channel  $\mathbf{h} \in \mathbb{C}^N$ , a scalar standard Gaussian signal  $x(t)$  (hypothesis  $\mathcal{H}_1$ ) and a noise variance  $\sigma^2$ , which are all unknown. As usual, we gather  $n$  i.i.d. received data in  $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(n)]$ .

Since  $\mathbf{h}$  is unknown, we cannot compare directly  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . Instead, we will accept or reject  $\mathcal{H}_0$  based on how  $\mathbf{Y}$  fits the pure noise hypothesis. It is a known result that the generalized likelihood ratio test (GLRT) for the decision between  $\mathcal{H}_0$  and  $\mathcal{H}_1$  boils down to the test [64]

$$\lambda'_1 \triangleq \frac{\lambda_1}{\left(\frac{1}{Nn} \text{tr} \mathbf{Y} \mathbf{Y}^H\right)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} f(\varepsilon)$$

with  $\lambda_1$  the largest eigenvalue of  $\frac{1}{n} \mathbf{Y} \mathbf{Y}^H$ , for  $f$  a given monotonic function and  $\varepsilon$  the maximally acceptable false alarm rate. Evaluating the statistical properties of  $\lambda'_1$  for finite  $N$  is however rather involved and leads to impractical tests, see e.g. [13]. Instead, we consider here a very elementary statistical test based on Theorem 5.

It is clear that  $\frac{1}{Nn} \text{tr} \mathbf{Y} \mathbf{Y}^H \xrightarrow{\text{a.s.}} \sigma^2$  under both  $\mathcal{H}_0$  or  $\mathcal{H}_1$ . Therefore, an application of Slutsky's lemma [1] ensures that the asymptotic fluctuations of  $\lambda'_1$  follow a Tracy-Widom distribution around  $(1 + \sqrt{N/n})^2$ . The GLRT method therefore leads asymptotically to test the Tracy-Widom statistics for the appropriately centered and scaled version of  $\lambda'_1$ . More precisely, for a false alarm rate  $\varepsilon$ , this is

$$N^{\frac{2}{3}} \frac{\lambda'_1 - (1 + \sqrt{c})^2}{(1 + \sqrt{c})^{\frac{4}{3}} \sqrt{c}} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} T_2^{-1}(1 - \varepsilon)$$

with  $c = N/n$  and  $T_2$  the Tracy-Widom distribution. Further properties of the above statistical test, and in particular theoretical expression of false alarm rates and test powers are derived in [65] using the theory of large deviations. Due to the relatively slow convergence speed of the largest eigenvalue distribution to the Tracy-Widom law, some recent works have proposed refinements, see e.g. [66].

In Figure 8, the performance of the GLRT detector given by the receiver operating curve for different false alarm rate levels is depicted, for small system sizes. We present a comparison between the GLRT approach, the empirical technique [67] which consists in using the conditioning number of  $\frac{1}{n} \mathbf{Y} \mathbf{Y}^H$  (ratio largest and the smallest eigenvalues) as a signal detector

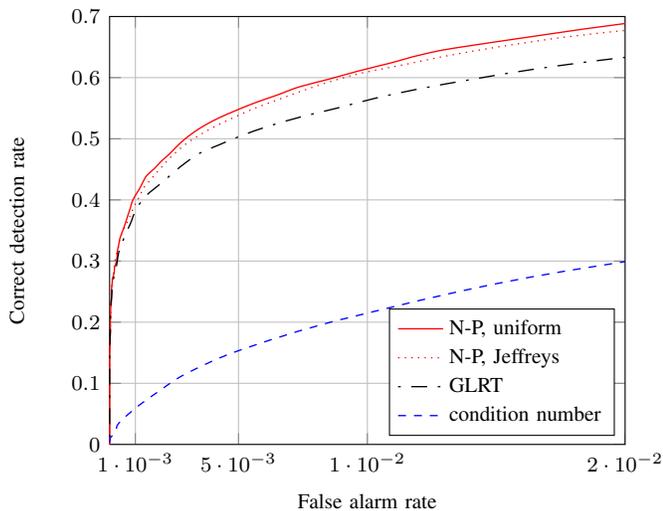


Fig. 8. Receiver operating curve for a priori unknown  $\sigma^2$  of the Neyman-Pearson test (N-P), condition number method and GLRT,  $N = 4$ ,  $n = 8$ , SNR= 0 dB,  $\mathbf{h}$  has Gaussian entries of zero mean and unit variance. For the Neyman-Pearson test, both uniform and Jeffreys prior, with exponent  $\beta = 1$ , are provided.

and the optimal Neyman-Pearson detector using Bayesian marginalization (with maximum entropy priors on  $\mathbf{h}$  and either uniform or  $\beta$ -Jeffreys prior on  $\sigma^2$ ) derived in [68]. It turns out that the suboptimal asymptotic GLRT approach is quite close in performance to the finite dimensional exact optimum, the latter being however quite complex to implement.

Our last application uses the second order statistics of eigenvalues and eigenvectors for a spike model in the context of failure localization in large dimensional networks.

#### D. Failure detection in large networks

The method presented here allows for fast detection and localization of local failures (few links or few nodes) in a large sensor network of  $N$  sensors gathering data about  $M$  system parameters  $\theta_1, \dots, \theta_M$ . Assume the linear scenario

$$\mathbf{x}(t) = \mathbf{H}\boldsymbol{\theta}(t) + \sigma\mathbf{w}(t)$$

for  $\boldsymbol{\theta}(t) = [\theta_1(t), \dots, \theta_M(t)]^T \in \mathbb{C}^M$  and  $\mathbf{w}(t) \in \mathbb{C}^N$  two standard Gaussian signals with independent entries of zero mean and unit variance. The objective here is to detect a small rank perturbation of  $\mathbb{E}[\mathbf{x}(t)\mathbf{x}(t)^H] \triangleq \mathbf{T} = \mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I}_N$ , originating from a local failure or change in the parameters, assuming  $\mathbf{T}$  known. In particular, when the entry  $k$  of  $\boldsymbol{\theta}(t)$  shows a sudden change in variance, the model changes as

$$\mathbf{x}'(t) = \mathbf{H}(\mathbf{I}_M + \alpha_k \mathbf{e}_k \mathbf{e}_k^H) \boldsymbol{\theta}(t) + \sigma\mathbf{w}(t)$$

for a given  $\alpha_k \geq -1$  and with  $\mathbf{e}_k \in \mathbb{C}^M$  the vector of all zeros but for  $\mathbf{e}_k(k) = 1$ . Taking for instance  $\alpha_k = -1$  turns the entry  $k$  of  $\boldsymbol{\theta}(t)$  into zero, corresponding to a complete collapse of the parameter under control. Denoting  $\mathbf{y}(t) = \mathbf{T}^{-\frac{1}{2}}\mathbf{x}(t)$  and  $\mathbf{y}'(t) = \mathbf{T}^{-\frac{1}{2}}\mathbf{x}'(t)$ , we have  $\mathbb{E}[\mathbf{y}(t)\mathbf{y}(t)^H] = \mathbf{I}_N$ , while

$$\mathbb{E}[\mathbf{y}'(t)\mathbf{y}'(t)^H] = \mathbf{I}_N + [(1 + \alpha_k)^2 - 1]\mathbf{T}^{-\frac{1}{2}}\mathbf{H}\mathbf{e}_k\mathbf{e}_k^H\mathbf{H}^H\mathbf{T}^{-\frac{1}{2}}$$

which is a rank-1 perturbation of  $\mathbf{I}_N$ .

Now assume that we do not know whether the model follows the expression of  $\mathbf{y}(t)$  or  $\mathbf{y}'(t)$ , and let us generically denote both by  $\mathbf{y}(t)$ . A simple off-line failure (or change) detection test consists, as in Section IV-C, to decide whether the largest eigenvalue of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$ , with  $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(n)]$ , has a Tracy-Widom distribution. However, we wish now to go further and to be able to decide, upon failure detection, which entry  $\theta_k$  of  $\boldsymbol{\theta}$  was altered. For this, we need to provide a statistical test for the hypothesis  $\mathcal{H}_k = \text{"parameter } \theta_k \text{ failed"}$ . A mere maximum likelihood procedure consisting in testing the Gaussian distribution of  $\mathbf{Y}$ , assuming  $\alpha_k$  known for each  $k$ , is however costly for  $N$  large and becomes impractical if  $\alpha_k$  is unknown. Instead, one can perform a statistical test on the extreme eigenvalues and eigenvector projections of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$ , which mainly costs the computational time of eigenvalue decomposition (already performed in the detection test). For this, we need to assume that the number  $n$  of observations is sufficiently large for a failure of any parameter of  $\boldsymbol{\theta}$  to be detectable. As an immediate application of Theorem 5, denoting  $\lambda$  the largest eigenvalue of  $\frac{1}{n}\mathbf{Y}\mathbf{Y}^H$  and  $\hat{\mathbf{u}}$  its corresponding eigenvector, the estimator  $\hat{k}$  for  $k$  is then given by [44]

$$\hat{k} = \arg \max_{1 \leq i \leq M} -N \left( \frac{|\mathbf{u}_i^H \hat{\mathbf{u}}|^2 - \xi_i}{\lambda - \rho_i} \right)^T \boldsymbol{\Sigma}_i^{-1} \left( \frac{|\mathbf{u}_i^H \hat{\mathbf{u}}|^2 - \xi_i}{\lambda - \rho_i} \right) - \log \det \boldsymbol{\Sigma}_i$$

with  $\mathbf{u}_i$  such that  $\|\mathbf{u}_i\| = 1$ ,  $\omega_i \mathbf{u}_i \mathbf{u}_i^H = [(1 + \alpha_i)^2 - 1]\mathbf{T}^{-\frac{1}{2}}\mathbf{H}\mathbf{e}_i\mathbf{e}_i^H\mathbf{H}^H\mathbf{T}^{-\frac{1}{2}}$ , and  $\xi_i, \rho_i, \boldsymbol{\Sigma}_i$  defined as a function of  $\omega_i$  in Theorem 5 (the index  $i$  refers here to a change in parameter  $\theta_i$  and not to the  $i$ -th largest eigenvalue of the small rank perturbation matrix).

This procedure still assumes that  $\alpha_i$  is known for each  $i$ . If not, an estimator of  $\alpha_i$  can be derived and Theorem 5 adapted accordingly to account for the fluctuations of the estimator.

In Figure 9, we depict the detection and localization performance for a sudden drop to zero of the parameter  $\theta_1(t)$  in a scenario with  $M = N = 10$ . We see in Figure 9 that, compared to the detection performance, the localization performance is rather poor for small  $n$  but reaches the same level as the detection performance for larger  $n$ , which is explained by the inaccuracy of the Gaussian approximation for  $\omega_i$  close to  $\sqrt{c}$ .

## V. CONCLUSIONS

In this short tutorial about statistical inference using large dimensional random matrix theory, we have argued that many traditional signal processing methods are inconsistent when both the population and system dimensions are large. We then introduced notions of random matrix theory which provide results on the spectrum of large random matrices. These results were then used to adjust some of these inconsistent signal processing methods to new consistent estimates. To this end, we presented a recent method based (i) on the Stieltjes transform to derive weak convergence properties of the spectrum of large matrices and (ii) on complex integration to derive estimators. This somewhat parallels the Fourier transform and M-estimator framework usually met in classical asymptotic signal processing [1].

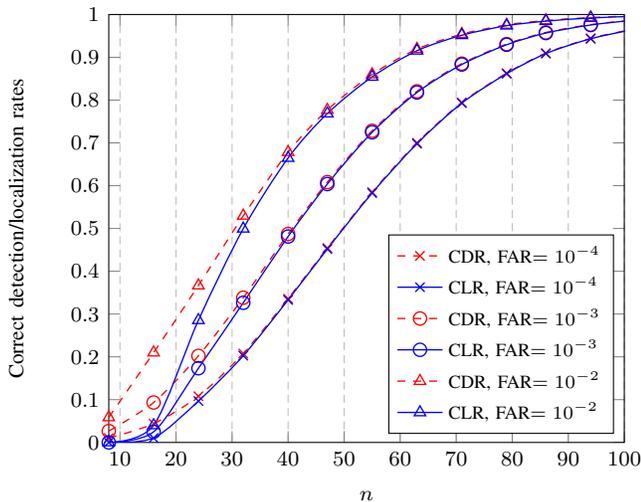


Fig. 9. Correct detection (CDR) and localization (CLR) rates for different levels of false alarm rates (FAR) and different values of  $n$ , for drop of variance of  $\theta_1$ . The minimal theoretical  $n$  for observability is  $n = 8$ .

Nonetheless, while classical signal processing tools are very mature, statistical inference based on random matrix analysis still lacks many fundamental mathematical results that only experts can provide to this day for advanced system models. The main consequence is the slow appearance of methods to derive various estimators for more involved signal processing problems. The emergence of the aforementioned Stieltjes transform and complex integration framework is a rare exception to this rule which, we believe, is the foundation for many future breakthroughs in this area.

## REFERENCES

- [1] A. W. Van der Vaart, *Asymptotic Statistics*. New York: Cambridge University Press, 2000.
- [2] X. Mestre and M. Lagunas, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 598–614, Feb. 2008.
- [3] L. Laloux, P. Cizeau, M. Potters, and J. P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance*, vol. 3, no. 3, pp. 391–397, Jul. 2000.
- [4] N. Hansen and A. Ostermeier, "Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation," in *Evolutionary Computation, Proceedings of IEEE International Conference on*. IEEE, 1996, pp. 312–317.
- [5] E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *The Annals of Mathematics*, vol. 62, no. 3, pp. 548–564, Nov. 1955.
- [6] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.
- [7] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill Series in Higher Mathematics, May 1986.
- [8] V. L. Girko, *Theory of Random Determinants*. Dordrecht, The Netherlands: Kluwer, Kluwer Academic Publishers, 1990.
- [9] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.
- [10] W. B. Wu and M. Pourahmadi, "Banding sample autocovariance matrices of stationary processes," *Statistica Sinica*, vol. 19, no. 4, pp. 1755–1768, 2009.
- [11] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20, no. 1-2, pp. 32–52, Dec. 1928.
- [12] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 475–501, 1964.
- [13] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Eigenvalues and condition numbers of complex random matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 2, pp. 441–456, 2005.
- [14] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy - Mathematical Surveys and Monographs No. 77*. Providence, RI, USA: American Mathematical Society, 2006.
- [15] P. Biane, "Free probability for probabilists," *Quantum Probability Communications*, vol. 11, pp. 55–71, 2003.
- [16] A. Masucci, Ø. Ryan, S. Yang, and M. Debbah, "Finite dimensional statistical inference," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2457–2473, 2011.
- [17] N. R. Rao and A. Edelman, "The polynomial method for random matrices," *Foundations of Computational Mathematics*, vol. 8, no. 6, pp. 649–702, Dec. 2008.
- [18] L. A. Pastur and V. Vasilchuk, "On the law of addition of random matrices," *Communications in Mathematical Physics*, vol. 214, no. 2, pp. 249–286, 2000.
- [19] W. Hachem, O. Khorunzhy, P. Loubaton, J. Najim, and L. A. Pastur, "A new approach for capacity analysis of large dimensional multi-antenna channels," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3987–4004, 2008.
- [20] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.
- [21] X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008.
- [22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [23] Z. D. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, 2nd ed. New York, NY, USA: Springer Series in Statistics, 2009.
- [24] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*. Cambridge University Press, 2010.
- [25] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.
- [26] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, 1995.
- [27] F. Dupuy and P. Loubaton, "Mutual information of frequency selective MIMO systems: an asymptotic approach," 2009. [Online]. Available: <http://www-syscom.univ-mlv.fr/fdupuy/publications.php>
- [28] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011.
- [29] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.
- [30] R. Couillet, J. Hoydis, and M. Debbah, "Deterministic equivalents for the analysis of unitary precoded systems," *IEEE Transactions on Information Theory*, 2011, submitted for publication. [Online]. Available: <http://arxiv.org/abs/1011.3717>
- [31] J. Hoydis, R. Couillet, and M. Debbah, "Random beamforming over correlated fading channels," *IEEE Transactions on Information Theory*, 2011, submitted for publication.
- [32] F. Dupuy and P. Loubaton, "On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach," *IEEE Transactions on Information Theory*, 2010, to appear. [Online]. Available: <http://arxiv.org/abs/1001.3102>
- [33] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: effective interference, effective bandwidth and user capacity," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 641–657, Feb. 1999.
- [34] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 622–640, Feb. 1999.
- [35] V. L. Girko, "Ten years of general statistical analysis." [Online]. Available: [www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf](http://www.general-statistical-analysis.girko.freewebspace.com/chapter14.pdf)
- [36] J. W. Silverstein, Z. D. Bai, and Y. Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," *Journal of Multivariate Analysis*, vol. 26, no. 2, pp. 166–168, 1988.
- [37] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah, "On the limit of the largest eigenvalue of the large dimensional sample covariance matrix,"

- Probability Theory and Related Fields*, vol. 78, no. 4, pp. 509–521, 1988.
- [38] Z. D. Bai and J. W. Silverstein, “No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices,” *The Annals of Probability*, vol. 26, no. 1, pp. 316–345, Jan. 1998.
- [39] —, “Exact separation of eigenvalues of large dimensional sample covariance matrices,” *The Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.
- [40] B. Dozier and J. W. Silverstein, “On the empirical distribution of eigenvalues of large dimensional information plus noise-type matrices,” *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 678–694, 2007.
- [41] F. Benaych-Georges and R. Rao, “The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices,” *Advances in Mathematics*, vol. 227, no. 1, pp. 494–521, 2011.
- [42] Z. D. Bai and J. F. Yao, “Limit theorems for sample eigenvalues in a generalized spiked population model,” 2008. [Online]. Available: <http://arxiv.org/abs/0806.1141>
- [43] J. Baik and J. W. Silverstein, “Eigenvalues of large sample covariance matrices of spiked population models,” *Journal of Multivariate Analysis*, vol. 97, no. 6, pp. 1382–1408, 2006.
- [44] R. Couillet and W. Hachem, “Local failure detection and diagnosis in large sensor networks,” *IEEE Transactions on Information Theory*, 2011, submitted for publication.
- [45] Z. D. Bai and J. W. Silverstein, “CLT of linear spectral statistics of large dimensional sample covariance matrices,” *The Annals of Probability*, vol. 32, no. 1A, pp. 553–605, 2004.
- [46] J. Yao, R. Couillet, J. Najim, and M. Debbah, “Fluctuations of an Improved Population Eigenvalue Estimator in Sample Covariance Matrix Models,” *IEEE Transactions on Information Theory*, 2011, submitted for publication.
- [47] I. M. Johnstone, “On the distribution of the largest eigenvalue in principal components analysis,” *Annals of Statistics*, vol. 99, no. 2, pp. 295–327, 2001.
- [48] J. Baik, G. Ben Arous, and S. Péché, “Phase transition of the largest eigenvalue for non-null complex sample covariance matrices,” *The Annals of Probability*, vol. 33, no. 5, pp. 1643–1697, 2005.
- [49] C. A. Tracy and H. Widom, “On orthogonal and symplectic matrix ensembles,” *Communications in Mathematical Physics*, vol. 177, no. 3, pp. 727–754, 1996.
- [50] J. Faraut, “Random matrices and orthogonal polynomials,” Lecture Notes, CIMPA School of Merida, 2006. [Online]. Available: [www.math.jussieu.fr/faraut/Merida.Notes.pdf](http://www.math.jussieu.fr/faraut/Merida.Notes.pdf)
- [51] O. N. Feldheim and S. Sodin, “A universality result for the smallest eigenvalues of certain sample covariance matrices,” *Geometric And Functional Analysis*, vol. 20, no. 1, pp. 88–123, 2010.
- [52] T. Chen, D. Rajan, and E. Serpedin, *Mathematical foundations for signal processing, communications and networking*. Cambridge University Press, 2011.
- [53] Y. V. Fyodorov, “Introduction to the random matrix theory: Gaussian unitary ensemble and beyond,” *Recent Perspectives in Random Matrix Theory and Number Theory*, vol. 322, pp. 31–78, 2005.
- [54] J. Mitola III and G. Q. Maguire Jr, “Cognitive radio: making software radios more personal,” *IEEE Personal Communication Magazine*, vol. 6, no. 4, pp. 13–18, 1999.
- [55] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [56] M. L. McCloud and L. L. Scharf, “A new subspace identification algorithm for high-resolution DOA estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 10, pp. 1382–1390, 2002.
- [57] X. Mestre, “On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices,” *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5353–5368, Nov. 2008.
- [58] P. Vallet, P. Loubaton, and X. Mestre, “Improved subspace estimation for multivariate observations of high dimension: the deterministic signals case,” *IEEE Transactions on Information Theory*, 2010, submitted for publication. [Online]. Available: <http://arxiv.org/abs/1002.3234>
- [59] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [60] J. Rissanen, “A universal prior for integers and estimation by minimum description length,” *The Annals of Statistics*, vol. 11, no. 2, pp. 416–431, 1983.
- [61] B. Nadler, “Nonparametric detection of signals by information theoretic criteria: performance analysis and an improved estimator,” *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2746–2756, 2010.
- [62] R. A. Maronna, “Robust M-estimators of multivariate location and scatter,” *The Annals of Statistics*, pp. 51–67, 1976.
- [63] J. T. Kent and D. E. Tyler, “Redescending M-estimates of multivariate location and scatter,” *The Annals of Statistics*, pp. 2102–2119, 1991.
- [64] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Transactions on Signal, Speech and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.
- [65] P. Bianchi, J. Najim, M. Maida, and M. Debbah, “Performance of some eigen-based hypothesis tests for collaborative sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2400–2419, 2011.
- [66] B. Nadler, “On the distribution of the ratio of the largest eigenvalue to the trace of a wishart matrix,” *Journal of Multivariate Analysis*, vol. 102, no. 2, pp. 363–371, 2011.
- [67] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, “Cooperative spectrum sensing using random matrix theory,” in *IEEE Pervasive Computing (ISWPC’08)*, Santorini, Greece, May 2008, pp. 334–338.
- [68] R. Couillet and M. Debbah, “A Bayesian framework for collaborative multi-source signal detection,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5186–5195, Oct. 2010.



**Romain Couillet** received his MSc in Mobile Communications at the Eurecom Institute and his MSC in Communication Systems in Telecom ParisTech, France in 2007. From 2007 to 2010, he worked with ST-Ericsson as an Algorithm Development Engineer on the Long Term Evolution Advanced project, where he prepared his PhD with Supelec, France, which he graduated in November 2010. He is currently an assistant professor at Supelec. His research topics are in information theory, signal processing, and random matrix theory. He is the recipient of the Valuetools 2008 best student paper award and of the 2011 EEA/GdR ISIS/GRETSI best PhD thesis award.



**Mérouane Debbah** was born in Madrid, Spain. He entered the Ecole Normale Supérieure de Cachan (France) in 1996 where he received his M.Sc and Ph.D. degrees respectively in 1999 and 2002. From 1999 to 2002, he worked for Motorola Labs on Wireless Local Area Networks and prospective fourth generation systems. From 2002 until 2003, he was appointed Senior Researcher at the Vienna Research Center for Telecommunications (FTW) (Vienna, Austria). From 2003 until 2007, he joined the Mobile Communications de-partment of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor. He is presently a Professor at Supelec (Gif-sur-Yvette, France), holder of the Alcatel-Lucent Chair on Flexible Radio. His research interests are in information theory, signal processing and wireless communications. Mérouane Debbah is the recipient of the “Mario Boella” prize award in 2005, the 2007 General Symposium IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award as well as the Valuetools 2007, Valuetools 2008 and CrownCom2009 best student paper awards. He is a WWRF fellow.