Spectral Theory of Large Dimensional Random Matrices Applied to Signal Detection*

Jack W. Silverstein[1] and Patrick L. Combettes[2]

[1]North Carolina State University
Department of Mathematics
Raleigh, NC 27695, USA

[2]The City College of The City University of New York
Department of Electrical Engineering
New York, NY 10031

Fall 1990

Abstract

Results on the spectral behavior of random matrices as the dimension increases are applied to the problem of detecting the number of sources impinging on an array of sensors. A common strategy to solve this problem is to estimate the multiplicity of the smallest eigenvalue of the spatial covariance matrix R of the sensed data from the sample covariance matrix $\widehat{R}$. Existing approaches, such as that based on information theoretic criteria, rely on the closeness of the noise eigenvalues of $\widehat{R}$ to each other and, therefore, the sample size has to be quite large when the number of sources is large in order to obtain a good estimate. The analysis presented in this report focuses on the splitting of the spectrum of $\widehat{R}$ into noise and signal eigenvalues. It is shown that, when the number of sensors is large, the number of signals can be estimated with a sample size considerably less than that required by previous approaches. The practical significance of the main result is that detection can be achieved with a number of samples comparable to the number of sensors in large dimensional array processing.

*This technical report from 1990 is the long version of the paper:
J. W. Silverstein and P. L. Combettes, Signal detection via spectral theory of large dimensional random matrices, *IEEE Transactions on Signal Processing,* vol. 40, no. 8, pp. 2100–2105, August 1992,
which was published without the proofs. Since several colleagues have requested those proofs over the years, we make them available now. Current emails of the authors: jack@math.ncsu.edu and plc@math.ncsu.edu.

1 Introduction

In many signal processing applications, a fundamental problem is the determination of the number of signals impinging on an array of sensors. Under the assumption that the vector of sensed data consists of superimposed random signals corrupted by additive white noise, the number of signals present in the scene is related to the multiplicity of the smallest eigenvalue of the spatial covariance matrix R of the data process, this eigenvalue being equal to the power of the noise. Since R is unknown, its spectrum must be approximated by observing that of the sample covariance matrix $\widehat{R}$ of the data process sampled across time. The eigenvalues of $\widehat{R}$ being typically distinct[1], the detection problem is that of deciding which of the smallest eigenvalues are associated with the noise. An approach is to use hypothesis tests on the multiplicity of the smallest eigenvalue of a random matrix, such as that discussed in [10]. An alternative strategy based on information theoretic criteria for model selection was proposed in [18] and was further studied in [9, 20, 22, 23, 24].

All of these detection methods rely on the ergodic theorem and their performance strongly depends on R being closely approximated by $\widehat{R}$, requiring the sample size to be quite large. In applications where the number of signals and, consequently, the number of sensors, is sizable, the required number of samples may be prohibitive. The purpose of this report is to bring into play elements of the spectral theory of random matrices, more specifically, results on the limiting distribution of the eigenvalues of random matrices as the dimension increases. This analysis will show that, when the number of sensors is large, the number of signals can be estimated with a sample size considerably less than that required by invoking the ergodic theorem.

The report is organized as follows. Results from the spectral theory of random matrices are introduced in Section 2. The application to signal detection is presented in Section 3 and numerical results are provided in Section 4. Our concluding remarks appear in Section 5. All of our results are proved in the Appendix (Section 6).

2 Spectral Theory of Random Matrices

Throughout this report, $\mathbb{N}^*$ will denote the set of strictly positive integers and $\mathbb{R}_+^*$ the set of strictly positive real numbers. All the random variables (r.v.'s) are defined on a probability space $(\Omega, \Sigma, \mathsf{P})$. A r.v. X is said to be in $L^r(\mathsf{P})$ $(0 < r < +\infty)$ if $\mathsf{E}|X|^r < +\infty$. For r.v.'s, almost sure convergence is denoted by $\xrightarrow{\text{a.s.}}$ and, for distribution functions[2] (d.f.'s), weak convergence is denoted by $\Rightarrow$. The transpose of a matrix A is denoted by $A^\top$, its conjugate transpose by A^*, and its trace by $\mathrm{tr}A$.

Let M be an $m \times m$ random matrix with real-valued eigenvalues $\{\Lambda_1, \ldots, \Lambda_m\}$. The empirical d.f. of the r.v.'s $\{\Lambda_1, \ldots, \Lambda_m\}$ is the stochastic process defined by[3]

$$(\forall \omega \in \Omega)(\forall x \in \mathbb{R}) \quad F^M(x, \omega) = \frac{1}{m} \sum_{i=1}^{m} 1_{]-\infty, x]}(\Lambda_i(\omega)). \tag{2.1}$$

We now review the main result, a limit theorem found in [19].

[1] For a sufficient condition under which the eigenvalues of $\widehat{R}$ would be almost surely distinct, see [13].

[2] By a d.f. we mean a right-continuous nondecreasing function F on $\mathbb{R}$ with $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$. The support of F is the closed set $\mathcal{S} = \{x \in \mathbb{R} \mid (\forall \varepsilon \in \mathbb{R}_+^*) \;\; F(x+\varepsilon) > F(x-\varepsilon)\}$.

[3] The characteristic function of a set S is denoted by 1_S.

Theorem 1 [19]. Let $(Y_{ij})_{i,j\geqslant 1}$ be i.i.d. real-valued r.v.'s with $\mathsf{E}|Y_{11} - \mathsf{E}Y_{11}|^2 = 1$. For each m in $\mathbb{N}^*$, let $Y_m = [Y_{ij}]_{m\times n}$, where $n = n(m)$ and $m/n \to y > 0$ as $m \to +\infty$, and let T_m be an $m \times m$ symmetric nonnegative definite random matrix independent of the Y_{ij}'s for which there exists a sequence of strictly positive numbers $(\mu_k)_{k\in\mathbb{N}^*}$ such that for each k in $\mathbb{N}^*$

$$\int_0^{+\infty} x^k dF^{T_m}(x) = \frac{1}{m}\mathrm{tr}T_m^k \xrightarrow{\text{a.s.}} \mu_k \quad \text{as} \quad m \to +\infty \tag{2.2}$$

and where the μ_k's satisfy Carleman's sufficiency condition, $\sum_{k\in\mathbb{N}^*} \mu_{2k}^{-1/2k} = +\infty$, for the existence and the uniqueness of the d.f. H having moments $(\mu_k)_{k\in\mathbb{N}^*}$. Let $M_m = (1/n)Y_m Y_m^\top T_m$. Then, almost surely, $(F^{M_m})_{m\geqslant 1}$ converges weakly to a nonrandom d.f. F having moments

$$(\forall k \in \mathbb{N}^*) \quad \nu_k = \sum_{w=1}^{k} y^{k-w} \sum \frac{k!}{m_1! \cdots m_w! w!} \mu_1^{m_1} \cdots \mu_w^{m_w} \tag{2.3}$$

where the inner sum extends over all w-tuples of positive integers $(m_1, \ldots, m_w)$ such that $\sum_{i=1}^w m_i = k - w + 1$ and $\sum_{i=1}^w i m_i = k$. Moreover, these moments uniquely determine F.

Similar results are given in [12] and [17] with varying degrees of assumptions, although in both papers the matrices studied can have complex-valued entries. However, the proof in [19] can easily be modified to allow complex-valued entries in Y_m and T_m, giving the same result, provided T_m is Hermitian and we take $M_m = (1/n)Y_m Y_m^* T_m$.

Although it does not appear likely a general explicit expression for F in terms of y and arbitrary H can be derived, useful qualitative information can be found from the different methods used in [12], [14], and [17] to express transforms of F (transforms of Stieltjes type in [12] and [17], the characteristic function in [14]). For example, in [12], it is shown that the endpoints of the connected components of the support of F are given by the extrema of the function

$$f(\alpha) = -\frac{1}{\alpha} + y \int_0^{+\infty} \frac{dH(x)}{\alpha + 1/x}. \tag{2.4}$$

The analysis in [14] shows how one can prove that F is absolutely continuous on $\mathbb{R}_+^*$ and express its derivative, provided the inverse of a certain function defined by y and H can be analytically extended in the real part of the complex plane.

We now provide additional results apropos of the limiting behavior of $(F^{M_m})_{m\geqslant 1}$.

Theorem 2. The limiting d.f. F in Theorem 1 is continuous on $\mathbb{R}_+^*$. Moreover, if H places no mass at 0 then, almost surely, $(F^{M_m})_{m\geqslant 1}$ converges to F uniformly in $\mathbb{R}$.

Proposition 1. With the same notation and hypotheses as in Theorem 1, the following hold:

(i) F and y uniquely determine H.

(ii) Almost surely, $(F^{T_m})_{m\geqslant 1}$ converges to H weakly.

(iii) $F \Rightarrow H$ as $y \to 0$.

Statement (iii) has a direct bearing on the problem of estimating the spectrum of a covariance matrix from observing that of a sample covariance matrix. Indeed, the matrix $(1/n)T_m^{1/2} Y_m Y_m^* T_m^{1/2}$ (whose eigenvalues are identical to those of M_m[4]) encompasses a broad class of sample covariance matrices stemming from n i.i.d. samples distributed as an m-dimensional random vector X with

[4]The reader is reminded that given two matrices $A_{p\times q}$ and $B_{q\times p}$, where $p \geqslant q$, the spectrum of AB is that of BA augmented by $p - q$ zeros.

$\mathsf{E}X = 0$ and $\mathsf{E}XX^* = T_m$ (including the Wishart case when X is multivariate complex Gaussian). In estimating the spectrum of T_m from the sample covariance matrix, there seems to be no mention in the literature as to the dependence of n on m, that is, how large the sample size should be vis-à-vis the vector dimension in order to estimate the eigenvalues to within a certain degree of accuracy. Indeed, asymptotic results are expressed only in terms of the sample size (see e.g. [1]). The fact that F differs from H for $y > 0$ while $F \Rightarrow H$ as $y \to 0$, which complements the fact that, for fixed m, $M_m \xrightarrow{\text{a.s.}} T_m$ as $n \to +\infty$, confirms the intuitively apparent statement that, for m large, n should be much larger, in the sense that $m = o(n)$.

3 Application to Signal Detection

3.1 Description of the Problem and Assumptions

Let p be the number of sensors in the array, q the unknown number of signals ($q < p$), and $[0, \tau]$ be the observation interval. At each time t in $[0, \tau]$, the j-th signal present in the scene, the additive noise at the i-th sensor, and the received data at the i-th sensor are respectively represented by the $L^2(\mathsf{P})$ complex-valued r.v.'s $S_j(t)$, $N_i(t)$, and $X_i(t)$. The random vectors $(S(t) = [S_1(t) \dots S_q(t)]^\top)_{t\in[0,\tau]}$ are identically distributed (i.d.) with nonsingular spatial covariance matrix $R_S = \mathsf{E}S(0)S(0)^*$. Moreover, it is assumed that the r.v.'s $(N_i(t) \mid 1 \leqslant i \leqslant p,\ t \in [0, \tau]\}$ are independent and identically distributed (i.i.d.) with $\mathsf{E}N_1(0) = 0$ and $\mathsf{E}|N_1(0)|^2 = \sigma^2$, where σ^2 is unknown, and independent from the r.v.'s $(S_j(t) \mid 1 \leqslant j \leqslant q,\ t \in [0, \tau])$. Let

$$N(t) = \sigma W(t) = \sigma[W_1(t) \dots W_p(t)]^\top \tag{3.1}$$

(so that the $W_i(t)$'s are standardized) and $X(t) = [X_1(t) \dots X_p(t)]^\top$. The data collected by the array of sensors are modeled as observations of the random vector

$$X(t) = AS(t) + N(t), \qquad t \in [0, \tau], \tag{3.2}$$

where A is a $p \times q$ complex matrix depending on the geometry of the array and the parameters of the signals, and is assumed to have rank q. The detection problem is to estimate q from the observation of n snapshots $\{X(t_1), \dots, X(t_n)\}$ of the data process. Under the above assumptions, the random vectors $(X(t))_{t\in[0,\tau]}$ are i.d. with spatial covariance matrix

$$R = \mathsf{E}X(0)X(0)^* = AR_SA^* + \sigma^2 I_p, \tag{3.3}$$

where I_p denotes the $p \times p$ identity matrix. Moreover, the $p - q$ smallest eigenvalues of R are equal to σ^2. These eigenvalues will be referred to as the noise eigenvalues and the remainder of the spectrum will be referred to as the signal eigenvalues. In practice, R is not known, and its spectrum must be inferred from observing that of the sample covariance matrix

$$\widehat{R} = \frac{1}{n}\sum_{i=1}^{n} X(t_i)X(t_i)^*. \tag{3.4}$$

Loosely speaking, one must then decide where the observed spectrum splits into noise and signal eigenvalues.

3.2 General Analysis

For every t in $[0, \tau]$, let us assume that the signal vector is given by

$$S(t) = CV(t) \quad \text{with} \quad V(t) = [V_1(t), \ldots, V_q(t)]^\top, \tag{3.5}$$

where C is $q \times q$, nonsingular, and the r.v.'s $\{V_1(t), \ldots, V_q(t)\}$ are i.i.d. with the same d.f. as $W_1(0)$. It is worth noting that this general formulation comprises the special case when $S(0)$ is multivariate complex Gaussian, which is a common assumption in array signal processing. Let $B = AC$. Then (3.2) yields

$$X(t) = [B \;\; \sigma I_p] \begin{bmatrix} V(t) \\ W(t) \end{bmatrix}. \tag{3.6}$$

Notice that $R_S = CC^*$ and $R = BB^* + \sigma^2 I_p$. If we further assume that the n vectors $\{S(t_1), \ldots, S(t_n)\}$ are independent, then the n data samples $\{X(t_1), \ldots, X(t_n)\}$ will also be independent and the corresponding sample covariance matrix $\widehat{R}$ takes on the form

$$\widehat{R} = \frac{1}{n}[B \;\; \sigma I_p] VV^* [B \;\; \sigma I_p]^*, \tag{3.7}$$

where $V = [V_{ij}]_{(p+q)\times n}$ consists of i.i.d. standardized entries.

Theorem 3. If $W_1(0)$ is standard complex Gaussian[5], the joint distribution of the eigenvalues of $\widehat{R}$ in (3.7) is the same as the joint distribution of the eigenvalues of $\widehat{R}' = (1/n) Y_p Y_p^* (BB^* + \sigma^2 I_p)$, where Y_p is any $p \times n$ random matrix with i.i.d. standardized complex Gaussian entries. In general, for p and n sufficiently large, with high probability, the empirical d.f.'s $F^{\widehat{R}}$ and $F^{\widehat{R}'}$ are close to the d.f. F of Theorem 1 for $m = p$, $y = p/n$, and $H = F^{BB^* + \sigma^2 I_p}$.

The importance of Theorem 3 becomes immediately apparent. The observations of the empirical d.f. $F^{\widehat{R}}$, for suitably large p and n, will not vary very much from one realization to another, even if n is not large relative to p. In fact, by Theorem 2, with high probability, $F^{\widehat{R}}$ will be uniformly close to a d.f. F that depends only on y and the eigenvalues of $BB^* + \sigma^2 I_p$. Hence, a realization of $F^{\widehat{R}}$ and the ratio p/n can be used to describe, to within a certain degree of accuracy, $F^{BB^* + \sigma^2 I_p}$, which will yield σ^2 and the ratio $y_1 = q/p$ which corresponds to the q strictly positive eigenvalues of BB^*.

Much of the information on the spectrum of $BB^* + \sigma^2 I_p$ can be directly observed from plotting histograms of the eigenvalues of $\widehat{R}$, in particular, the ratio y_1 of signal eigenvalues. Let G denote the empirical d.f. of the eigenvalues of $BB^* + \sigma^2 I_p$ which are greater than σ^2, and let b_1 and b_2 denote, respectively, the smallest and largest of these values. Then, for every x in $\mathbb{R}$, we can write

$$H(x) = F^{BB^* + \sigma^2 I_p}(x) = (1 - y_1) 1_{[\sigma^2, +\infty[}(x) + y_1 G(x) \tag{3.8}$$

Proposition 2. When $y < 1$, the smallest interval $[x_1, x_4]$ containing the support of F satisfies $0 < x_1 < x_4 < +\infty$ with $x_1 \uparrow \sigma^2$ and $x_4 \downarrow b_2$ as $y \downarrow 0$. In addition, there exists an α in $]-1/\sigma^2, -1/b_1[$ such that

$$g(\alpha) = y \left((1 - y_1) \left(\frac{\alpha}{\alpha + 1/\sigma^2} \right)^2 + y_1 \int_{b_1}^{b_2} \left(\frac{\alpha}{\alpha + 1/x} \right)^2 dG(x) \right) < 1 \tag{3.9}$$

[5] A r.v. is said to be standardized complex Gaussian if its real and imaginary parts are i.i.d. with mean zero and variance 1/2.

(which can always be found for y sufficiently small) if and only if the support of F splits into at least two separate components, with the leftmost interval $[x_1, x_2]$ being a connected component of the support containing mass $1 - y_1$ from F. Furthermore, for y sufficiently small, $x_2 \downarrow \sigma^2$ as $y \downarrow 0$ and, if $[x_3, x_4]$ denotes the smallest interval containing the remaining support of F, then $x_3 \uparrow b_1$ as $y \downarrow 0$. Regardless of the respective location of x_2 and x_3 vis-à-vis σ^2 and b_1, the separation between the noise and signal portions of the spectrum, i.e. $x_3 - x_2$, increases as y decreases. When $y > 1$, F places mass $1 - 1/y$ at the origin, but the remaining support will lie to the right of a strictly positive value x_1. It is still possible for the support of F to split further provided (3.9) holds. In this case the leftmost interval $[x_1, x_2]$ will carry mass $(1/y) - y_1$, leaving mass y_1 to the remaining support of F to the right of x_2. When $y = 1$ the latter situation applies, except now $x_1 = 0$, and there will be no mass at 0.

Thus, if p and n are large enough so that $F^{\widehat{R}}$ is close to F with high probability, then for $y = p/n$ suitably small, an appropriately constructed histogram of the eigenvalues of $\widehat{R}$ will display clustering on the left separated from the rest of the figure. The proportion of the number of eigenvalues associated with the histogram to the right of the clustering will then be close to q/p, with high probability.

Although the theory merely guarantees that the proportion of signal eigenvalues of $\widehat{R}$ is close to that of R, extensive simulation strongly suggests that the spectrum of $\widehat{R}$ splits into two portions containing the exact number of noise and signal eigenvalues, and that the endpoints of these portions agree very closely with the ones predicted by the theory. This point, which will be illustrated in Section 4, leads to the possibility of the existence of a much stronger underlying spectral theory deepening the results of Theorem 1. Results along these lines are known for the extreme eigenvalues when $T_m = \sigma^2 I_m$. Such specific cases will be discussed in Section 3.3.

Intuitively, the above procedure has advantages over other methods used to estimate q, in particular, those adapted from information theoretic criteria discussed in [9], [18], [20], [22], [23], and [24]. The latter methods try to exploit the closeness of the noise eigenvalues of $\widehat{R}$ to each other as well as their separation from the remaining signal eigenvalues. Usually the sample size has to be quite large for the smaller eigenvalues to cluster. On the other hand, only the separation of the two classes of eigenvalues is needed when viewing the spectrum, so a suitable n can conceivably be much smaller, sometimes even smaller than p. In other words, previous methods require $\widehat{R}$ to be near $BB^* + \sigma^2 I_p$, while, for situations where p is sizable, the present analysis requires n to be large enough so that the support of F separates.

3.3 Specific Cases

An important case to consider is the one for which no signal is present, that is, when $B = 0$, or equivalently, when $T_m = \sigma^2 I_m$. Then it is known [7, 8, 12] that, for $y \leqslant 1$, F is continuously differentiable, where

$$F'(x) = \begin{cases} \dfrac{\left((x - \sigma^2(1-\sqrt{y})^2)(\sigma^2(1+\sqrt{y})^2 - x)\right)^{1/2}}{2\pi\sigma^2 yx} & \text{if } \sigma^2(1-\sqrt{y})^2 < x < \sigma^2(1+\sqrt{y})^2; \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

and for $y > 1$, F has derivative (3.10) on $\mathbb{R}_+^*$ and mass $1 - 1/y$ at 0. Furthermore, the largest eigenvalue of M_m converges almost surely [respect. in probability] to $\sigma^2(1 + \sqrt{y})^2$ as $m \to +\infty$ if and only if $\mathsf{E}Y_{11} = 0$ and $Y_{11} \in L^4(\mathsf{P})$ [respect. $x^4\mathsf{P}\{\omega \in \Omega \mid |Y_{11}(\omega)| \geq x\} \to 0$ as $x \to +\infty$] [2, 6, 16, 21]. The almost sure convergence of the smallest eigenvalue of M_m to $\sigma^2(1 - \sqrt{y})^2$ when

$y < 1$ has thus far been shown only for Y_{11} standardized Gaussian [15] (it is remarked here that the results on the extreme eigenvalues have been verified for Y_{11} real-valued, but, again, the proofs can be extended to the complex case).

The above results can be used to investigate the possibility of no signals arriving at the sensors. Certainly, the existence of at least one signal would be in doubt if the number of samples were quite large but histograms indicate only one connected component away from 0. But, for any y, comparisons can be made between histograms of the eigenvalues, (3.10), and F' when $B \neq 0$, to infer whether or not signals are present, provided the latter densities exist and appear different enough from (3.10) to make a distinction.

For this reason it is mentioned briefly here the case when G places mass at one value, $b > \sigma^2$. Except for the situation of only one signal, this case is not typically found in practice. However, the d.f. F can be completely determined and its properties strongly suggest the smoothness and appearance of F for general G. Only the case $y < 1$ will be outlined, the remaining cases for y following as above. From the analysis in [14] it can be shown that F is continuously differentiable with derivative of the form

$$F'(x) = \begin{cases} k\dfrac{\left(p_3(x) + x\sqrt{p_4(x)}\right)^{1/3} - \left(p_3(x) - x\sqrt{p_4(x)}\right)^{1/3}}{x} & \text{if } p_4(x) \geqslant 0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

Here, p_3 and p_4 are, respectively, third and fourth degree polynomials depending continuously on y, y_1, σ^2, b, and the leading coefficient of p_4 is negative. The latter polynomial has either two real roots, $0 < x_1 < x_4$, so that F has support on $[x_1, x_4]$, or four real roots, $0 < x_1 < x_2 < x_3 < x_4$, which is the above mentioned case where the support of F splits into two intervals, $[x_1, x_2]$ and $[x_3, x_4]$, with $F(x_2) - F(x_1) = 1 - y_1$. Using (3.9), it is straightforward to show F splits if and only if

$$y\frac{\left((b^2 y_1)^{1/3} + (\sigma^4(1-y_1))^{1/3}\right)^3}{(b-\sigma^2)^2} < 1. \tag{3.12}$$

When the left-hand side of (3.12) is equal to 1, then p_4 still has four real roots, but $x_2 = x_3$. When (3.12) holds, F' is unimodal on each of the intervals, with infinite slopes at each endpoint. If there is a $y < 1$, say y_o, for which the left-hand side of (3.12) is equal to 1, then, since the graph of F' varies continuously with y, as y increases from 0, the separate curves eventually join (at $y = y_o$) and the single curve will display two relative maxima, at least for y near y_o. Thus, although y may not be small enough to split F', it may still be possible to infer the number of signal eigenvalues from the shape of a histogram.

4 Simulation Results

The objective of this section is to illustrate some aspects of our analysis through their application to the case of a linear array with p sensors receiving noisy signals from q narrow-band far-field sources. The sensors are assumed to be omnidirectional with unity gain and uniform spacing $\lambda/2$, where λ is the signal wavelength[6].

[6] In this context, the matrix A in (3.2) has a Vandermonde structure with $A_{ki} = \exp(-\imath\pi(k-1)\sin\theta_i)$ $(1 \leqslant k \leqslant p,$ $1 \leqslant i \leqslant q)$, where θ_i is the angle of arrival of the i-th signal with respect to the normal to the array.

Our analysis applies to cases where p is large. Simulations have supported its applicability for values of p as low as 30. In the simulation presented here, the number of sensors is set to $p = 50$ and the noise is zero mean, white, complex Gaussian, with power $\sigma^2 = 1$. The signal scenario consists of $q = 35$ partially correlated sources with angles of arrivals uniformly spaced between -70^o and 70^o and power selected at random from a uniform distribution so as to yield signal-to-noise ratios ranging from 0dB to 10dB. The signal vector is multivariate complex Gaussian and obtained according to (3.5), where C is a randomly generated banded matrix.

The spectrum $\mathcal{L}$ of $R = BB^* + I_{50}$, where $B = AC$, was computed in order to obtain an explicit expression for the functions $f(\cdot)$ of (2.4) and $g(\cdot)$ of (3.9). Newton's method was used to find the minimum of $g(\cdot)$ over $]-1/\sigma^2, -1/b_1[$ and, whence, it was found that the largest value of y for which the splitting of the spectrum occurs (i.e. (3.9) holds) is $\tilde{y} = 1.058$. Then, with the above configuration, four experiments were performed with the following number of samples n: 50, 100, 250, and 1500 (which corresponds to values of y of 1, 1/2, 1/5, and 1/30, respectively). In each experiment, 10 realizations $\mathcal{L}_1, \ldots, \mathcal{L}_{10}$ of the spectrum of the sample covariance matrix $\widehat{R}$ were observed, the eigenvalues being arranged in nondecreasing order. The results of these experiments are shown in Tables 1 through 4. Even for $y = 1$, the $p - q = 15$ smallest eigenvalues are seen to cluster to the left of most of the observed spectra. This confinement delimitates exactly the noise portion of the spectrum and, thereby, detects the exact number of signals. As discussed earlier, for a given value of y, the theoretical endpoints of the supports of the noise and signal portions of the spectrum can be determined from the location of the relative extrema of $f(\cdot)$. Newton's method was used to this end and gave the results shown in Table 5. In agreement with Proposition 2, it is seen that, as y decreases, the separation $x_3 - x_2$ increases while the endpoints converge towards the theoretical values.

Table 1. Observed Spectra - $y = 1$**.**

	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$	$\mathcal{L}_4$	$\mathcal{L}_5$	$\mathcal{L}_6$	$\mathcal{L}_7$	$\mathcal{L}_8$	$\mathcal{L}_9$	$\mathcal{L}_{10}$	$\mathcal{L}$
λ_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1
λ_2	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	1
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{10}	0.31	0.33	0.32	0.32	0.28	0.28	0.33	0.35	0.34	0.34	1
λ_{11}	0.43	0.40	0.37	0.41	0.38	0.36	0.38	0.40	0.45	0.50	1
λ_{12}	0.45	0.47	0.48	0.49	0.44	0.49	0.41	0.52	0.48	0.55	1
λ_{13}	0.57	0.57	0.50	0.60	0.58	0.61	0.58	0.63	0.64	0.64	1
λ_{14}	0.67	0.64	0.64	0.74	0.80	0.73	0.78	0.75	0.74	0.73	1
λ_{15}	0.86	0.87	0.83	1.05	0.96	0.95	0.95	0.86	0.87	0.90	1
λ_{16}	1.38	1.64	1.40	1.90	1.18	1.45	2.35	1.71	1.61	1.70	5.34
λ_{17}	2.59	2.72	2.41	2.81	1.82	3.41	3.20	2.50	2.85	2.40	6.20
λ_{18}	5.61	5.21	4.74	4.97	3.52	4.11	5.85	4.88	4.67	5.32	21.4
λ_{19}	7.98	7.64	8.22	7.37	9.66	6.06	7.03	8.47	6.70	6.07	23.1
λ_{20}	11.4	9.87	10.8	9.67	11.6	8.16	11.1	11.1	12.2	11.8	25.7
λ_{21}	14.8	13.3	11.9	11.7	16.6	13.7	14.3	12.9	14.5	13.3	49.2
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{49}	1159	1074	1137	1065	1067	1154	1128	1123	1135	1229	756
λ_{50}	1470	1309	1233	1390	1458	1556	1547	1306	1347	1522	932

Table 2. Observed Spectra - $y = 1/2$**.**

	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$	$\mathcal{L}_4$	$\mathcal{L}_5$	$\mathcal{L}_6$	$\mathcal{L}_7$	$\mathcal{L}_8$	$\mathcal{L}_9$	$\mathcal{L}_{10}$	$\mathcal{L}$
λ_1	0.22	0.23	0.23	0.23	0.22	0.21	0.21	0.22	0.21	0.22	1
λ_2	0.28	0.29	0.25	0.25	0.28	0.25	0.29	0.26	0.26	0.26	1
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{10}	0.77	0.79	0.75	0.71	0.79	0.71	0.75	0.73	0.76	0.70	1
λ_{11}	0.97	0.91	0.83	0.77	0.84	0.79	0.78	0.84	0.82	0.79	1
λ_{12}	1.05	0.96	0.90	0.91	0.89	0.88	0.89	0.93	0.90	0.89	1
λ_{13}	1.11	1.06	0.96	0.99	0.99	0.95	0.96	0.94	0.93	0.97	1
λ_{14}	1.20	1.16	1.07	1.10	1.09	1.11	1.14	1.13	1.01	1.03	1
λ_{15}	1.41	1.32	1.20	1.30	1.16	1.24	1.31	1.31	1.21	1.33	1
λ_{16}	3.21	3.21	3.35	3.81	3.01	3.14	3.35	3.29	2.97	3.54	5.34
λ_{17}	4.04	4.27	4.49	4.78	4.07	4.25	4.82	5.36	3.65	4.61	6.20
λ_{18}	13.2	10.2	11.8	11.7	13.4	11.8	12.1	11.5	11.2	11.6	21.4
λ_{19}	13.7	14.7	15.9	16.5	14.4	13.7	15.3	16.8	13.4	17.7	23.1
λ_{20}	16.4	18.1	17.5	21.1	18.4	19.2	19.1	19.5	15.7	19.8	25.7
λ_{21}	29.4	30.1	28.1	32.7	29.8	25.9	32.0	31.8	28.1	28.7	49.2
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{49}	1019	959	879	1137	1012	941	1083	894	910	946	756
λ_{50}	1457	1174	1030	1360	1117	1053	1137	1186	1149	1084	932

Table 3. Observed Spectra - $y = 1/5$**.**

	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$	$\mathcal{L}_4$	$\mathcal{L}_5$	$\mathcal{L}_6$	$\mathcal{L}_7$	$\mathcal{L}_8$	$\mathcal{L}_9$	$\mathcal{L}_{10}$	$\mathcal{L}$
λ_1	0.54	0.52	0.48	0.53	0.51	0.52	0.52	0.47	0.50	0.49	1
λ_2	0.60	0.56	0.55	0.57	0.57	0.58	0.54	0.59	0.57	0.54	1
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{10}	0.94	0.93	0.94	0.90	0.92	0.95	0.94	0.96	0.91	0.92	1
λ_{11}	1.01	0.98	1.00	0.97	0.99	0.97	0.95	1.06	1.00	0.97	1
λ_{12}	1.09	1.03	1.06	1.02	1.02	1.04	1.01	1.10	1.09	1.01	1
λ_{13}	1.12	1.06	1.13	1.11	1.07	1.10	1.11	1.14	1.12	1.10	1
λ_{14}	1.18	1.12	1.24	1.16	1.17	1.17	1.19	1.18	1.25	1.17	1
λ_{15}	1.36	1.27	1.31	1.32	1.21	1.38	1.26	1.27	1.34	1.19	1
λ_{16}	4.35	4.78	4.37	4.84	4.96	4.03	4.26	4.19	4.71	4.36	5.34
λ_{17}	5.26	6.05	6.31	6.13	5.64	5.54	5.61	5.71	5.29	4.94	6.20
λ_{18}	17.4	17.5	18.1	18.5	17.0	17.4	16.8	17.2	17.7	16.5	21.4
λ_{19}	19.4	20.1	18.7	20.4	18.9	20.0	19.0	20.3	19.2	18.8	23.1
λ_{20}	22.9	25.1	21.3	22.3	22.7	21.9	23.4	22.4	22.9	23.3	25.7
λ_{21}	36.7	40.2	39.6	39.8	40.8	39.8	42.4	40.9	41.9	37.0	49.2
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{49}	902	856	887	875	800	818	893	889	878	831	756
λ_{50}	1063	985	1004	1064	1142	980	1043	1068	1008	993	932

Table 4. Observed Spectra - $y = 1/30$**.**

	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$	$\mathcal{L}_4$	$\mathcal{L}_5$	$\mathcal{L}_6$	$\mathcal{L}_7$	$\mathcal{L}_8$	$\mathcal{L}_9$	$\mathcal{L}_{10}$	$\mathcal{L}$
λ_1	0.82	0.82	0.81	0.82	0.83	0.81	0.83	0.82	0.81	0.83	1
λ_2	0.84	0.84	0.84	0.86	0.85	0.83	0.85	0.85	0.85	0.85	1
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{10}	1.02	1.02	1.01	1.01	1.01	1.00	1.02	1.02	1.02	1.02	1
λ_{11}	1.04	1.04	1.03	1.03	1.04	1.03	1.04	1.04	1.04	1.06	1
λ_{12}	1.06	1.04	1.06	1.06	1.07	1.05	1.06	1.06	1.05	1.07	1
λ_{13}	1.08	1.07	1.09	1.08	1.09	1.08	1.09	1.07	1.08	1.09	1
λ_{14}	1.10	1.11	1.11	1.10	1.12	1.11	1.10	1.08	1.10	1.11	1
λ_{15}	1.16	1.13	1.13	1.14	1.17	1.14	1.14	1.11	1.15	1.13	1
λ_{16}	5.32	5.09	5.23	5.27	5.18	5.12	5.03	5.27	5.12	5.52	5.34
λ_{17}	5.97	6.34	6.08	6.10	5.94	6.16	6.26	5.96	6.05	5.92	6.20
λ_{18}	20.5	20.6	20.1	19.4	20.2	19.8	21.7	20.5	20.1	20.9	21.4
λ_{19}	22.0	22.1	21.7	21.3	22.4	22.4	23.1	23.0	22.6	22.2	23.1
λ_{20}	24.8	25.3	26.0	25.5	25.2	24.2	25.5	25.3	25.1	24.5	25.7
λ_{21}	48.6	48.8	49.6	47.7	48.3	48.9	46.8	47.9	47.7	48.4	49.2
$\vdots$			$\vdots$					$\vdots$			$\vdots$
λ_{49}	764	807	818	766	772	765	779	797	766	788	756
λ_{50}	944	978	929	947	896	948	966	956	991	962	932

Table 5. Theoretical Bounds for Noise and Signal Spectrum Supports.

	$y = 1$	$y = 1/5$	$y = 1/30$	$y = 0$
x_1	0.000	0.4642	0.789	1.000
x_2	1.124	1.369	1.184	1.000
x_3	1.167	3.970	5.785	5.342
x_4	1586	1137	995.4	931.6

5 Conclusion

We have applied results from the spectral theory of large dimensional random matrices to the signal detection problem in situations where the number of sources is sizable. A theoretical foundation was established for the analysis of the splitting of the spectrum of the sample covariance matrix between a connected noise component and a remaining signal component. While conventional methods require that the sample size be impracticably large in order to closely approximate the spatial covariance matrix, the present analysis shows that the observed spectrum will split with high probability with a number of samples comparable to the number of sensors. As far as the detection problem is concerned, the eigenvalues of the spatial covariance matrix R need not be estimated with a high degree of precision; only the accurate splitting of the spectrum is required.

This work should suggest to the engineering community that by simply observing the spectrum of a large dimensional sample covariance matrix, highly relevant information can be extracted when the sample size is not exceedingly large. In the context of large dimensional array processing, the practical significance of our main result is that detection can be achieved when the sample size is only on the same order of magnitude as the number of sensors.

6 Appendix: Proofs

The imaginary part of a complex number z is denoted by $\Im z$.

Proof of Proposition 1. (i) is established by noting that the sequence $(\nu_k)_{k\in\mathbb{N}^*}$ can be derived from F and, from (2.3), the sequence $(\mu_k)_{k\in\mathbb{N}^*}$, which uniquely determines H, can be computed unambiguously; (ii) follows from (2.2) and the Fréchet-Shohat theorem[7]; (iii) is a direct consequence of the Fréchet-Shohat theorem since (2.3) implies that $(\forall k \in \mathbb{N}^*)\ \nu_k \to \mu_k$ as $y \to 0$.

Proof of Proposition 2. In [12] the matrix corresponding to $M_m = (1/n)X_m X_m^* T_m$ is $B_n = (1/n)X_m^* T_m X_m$. The spectra of these two matrices differ only by an additional number of zeros, $n-m$ extra zeros to the spectrum of B_n when $m < n$, $m-n$ to the spectrum of M_m when $m > n$. Let K denote the limiting empirical d.f. of the eigenvalues of B_n. It follows that

$$(\forall x \in \mathbb{R}) \quad K(x) = (1-y)1_{[0,+\infty[}(x) + yF(x). \tag{6.1}$$

Let $\mathcal{A}(\cdot)$ be the Stieltjes transform of K, i.e.

$$(\forall z \in \mathbb{C} \setminus \mathbb{R}) \quad \mathcal{A}(z) = \int_{-\infty}^{+\infty} \frac{dK(x)}{x-z}, \tag{6.2}$$

and let $\mathcal{B}(\cdot)$ denote the Stieltjes transform of F. From (6.1) and (6.2), it follows that

$$\mathcal{A}(z) = \frac{1-y}{-z} + y\mathcal{B}(z). \tag{6.3}$$

In [12] it is shown that $\mathcal{A}(z)$, for $\Im\, z > 0$, is the unique solution to the equation

$$z = -\frac{1}{\mathcal{A}(z)} + y\int_0^{+\infty} \frac{xdH(x)}{1+x\mathcal{A}(z)}, \tag{6.4}$$

from which K can be calculated from

$$K(x_2) - K(x_1) = \lim_{\eta\downarrow 0} \frac{1}{\pi} \int_{x_1}^{x_2} \Im\mathcal{A}(\xi + \imath\eta)d\xi, \tag{6.5}$$

where x_1 and x_2 are continuity points of K. Associated with the above inversion formula is the following. If $w_1 < w_2$ are values lying outside the support of K, then

$$K(w_2) - K(w_1) = -\frac{1}{2\pi\imath} \oint_C \mathcal{A}(z)dz, \tag{6.6}$$

where C can be taken as the circle in the complex plane having a diameter with endpoints w_1 and w_2 on the real axis. It is remarked in [12] that, on the union of intervals on the real axis outside the support of K, $\mathcal{A}$ is real and strictly increasing, and is continuous on each interval. Therefore, its inverse exists on the range of these intervals and is given by (6.4) for $\mathcal{A}$ and z real. This inverse, denoted by $f(\cdot)$, is given by (2.4). Writing H as in (3.8), $f(\cdot)$ takes on the form

$$f(\alpha) = -\frac{1}{\alpha} + \frac{y(1-y_1)}{\alpha + 1/\sigma^2} + yy_1 \int_{b_1}^{b_2} \frac{dG(x)}{\alpha + 1/x}. \tag{6.7}$$

[7] The Fréchet-Shohat theorem [11] states that if $(F_n)_{n\geqslant 1}$ is a sequence of d.f.'s having moments of all orders with $(\forall k \in \mathbb{N}^*)\ \lim_{n\to+\infty} \int_{-\infty}^{+\infty} x^k dF_n(x) = \mu_k$ finite, and if F is the only d.f. with moments $(\mu_k)_{k\in\mathbb{N}^*}$, then $F_n \Rightarrow F$ as $n \to +\infty$.

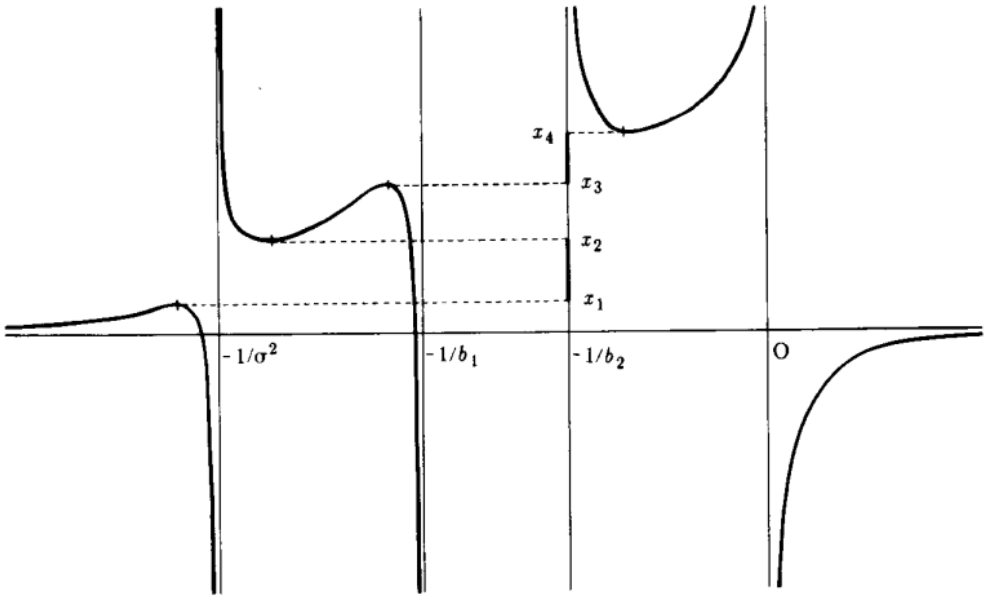

Figure 1: Graph of $f(\cdot)$.

The asymptotic properties of x_1 through x_4, the existence of x_2 and x_3, when $g(\alpha)$ in (3.9) is less than 1, and the existence of mass at 0 when $y > 1$, all follow from elementary calculus[8], together with the fact that $f'(\alpha) = (1 - g(\alpha))/\alpha^2$. Figure 1 shows a typical graph of f when separation occurs.

As for the mass F assigns to this interval and to the remaining portions of the support, (6.6) can be used together with a change of variables from (6.4). We will only derive the mass for $[x_1, x_2]$ when $y < 1$, the other portions of the support and cases ($y > 1$, $y = 1$) being similar. With $0 < w_1 < x_1$, and w_2 lying slightly to the right of x_2 we have

$$K(w_2) - K(w_1) = -\frac{1}{2\pi i} \oint_C \left(\frac{1}{\alpha} - \frac{\alpha y(1-y_1)}{(\alpha + 1/\sigma^2)^2} - y y_1 \int_{b_1}^{b_2} \frac{\alpha dG(x)}{(\alpha + 1/x)^2} \right) d\alpha, \tag{6.8}$$

where C is a simple closed positively oriented contour enclosing $-1/\sigma^2$ but not the origin nor any part of $[-1/b_1, -1/b_2]$. From Cauchy's integral formula, we have

$$K(w_2) - K(w_1) = y(1 - y_1), \tag{6.9}$$

so that from (6.1) we conclude $[x_1, x_2]$ contains mass $1 - y_1$ from F. Let us mention here that the values of $F(0)$ are obtained in the proof of Theorem 2.

The following lemma will be needed in the proof of Theorem 2.

Lemma 1. Let A and B be Hermitian, nonnegative definite matrices. Then for every α and β in $\mathbb{R}_+^*$, $F^{AB}(\alpha\beta) \leqslant F^A(\alpha) + F^B(\beta)$.

[8] In the case when the integral $I(\alpha)$ in (6.7) satisfies $I(b_1-) = -\infty$ and $I(b_2+) = +\infty$, the results are straightforward. If $I(b_1-)$ [respect. $I(b_2+)$] is finite, the additional fact that $f(\cdot)$ does not exist in any interval $]b_1, b_1 + \varepsilon[$ [respect. $]b_2 - \varepsilon, b_2[$] is needed. The latter is proven by using the fact that whenever $I(\alpha)$ exists, $G'(\alpha) = 0$, which can be verified in a straightforward manner. Thus, if $f(\cdot)$ were to exist on, say $]b_1, b_1 + \varepsilon[$, then necessarily G would place mass at b_1, resulting in $I(b_1-) = -\infty$, a contradiction.

Proof. For any $m \times m$ matrix C with real eigenvalues, let $\ell^C(i)$ be the i-th largest eigenvalue of C if $1 \leqslant i \leqslant m$, and 0 otherwise. From a routine extension of a result in [4], for every positive integers r_A and r_B

$$\ell^{AB}(m-(r_A+r_B)) \geqslant \ell^A(m-r_A)\ell^B(m-r_B). \tag{6.10}$$

In particular, if we let r_A [respect. r_B] be the number of eigenvalues of A [respect. B] less than or equal to α [respect. β], the result follows.

Proof of Theorem 2. We shall use here an argument similar to that used in [17]. The fact that K, and therefore F, is continuous on $\mathbb{R}_+^*$ can be proven by contradiction. First we notice from (6.4) that $\mathcal{A}(z)$ satisfies

$$\mathcal{A}(z) = \left(-z + y\int_0^{+\infty} \frac{xdH(x)}{1+x\mathcal{A}(z)}\right)^{-1}. \tag{6.11}$$

Suppose $x_0 > 0$ is a discontinuity point of K with jump μ. Then (6.2) gives

$$\Im\mathcal{A}(x_0+\imath\eta) = \int_{-\infty}^{+\infty} \frac{\eta dK(x)}{(x-x_0)^2+\eta^2} \geqslant \int_{\{x_0\}} \frac{\eta dK(x)}{(x-x_0)^2+\eta^2} = \frac{\mu}{\eta} \to +\infty \quad \text{as} \quad \eta \downarrow 0. \tag{6.12}$$

On the other hand,

$$\left|\int_0^{+\infty} \frac{xdH(x)}{1+x\mathcal{A}(x_0+\imath\eta)}\right| \leqslant \frac{1}{\Im\mathcal{A}(x_0+\imath\eta)} \to 0 \quad \text{as} \quad \eta \downarrow 0. \tag{6.13}$$

The contradiction then arises from the fact that as $\eta \downarrow 0$ the left-hand side of (6.11) becomes unbounded, while the right-hand side approaches $-1/x_0$. The last assertion is proved by noting that if H places no mass at 0, $(F^{M_m}(0\pm))_{m\geqslant 1}$ converges almost surely to $F(0\pm)$. Indeed, trivially, $(\forall m \in \mathbb{N}^*)\ F^{M_m}(0-) = F(0-) = 0$. Moreover, since the eigenvalues of M_n, $(1/n)Y_mY_m^*$, and T_m are all positive, it follows from Sylvester's inequality [5] that

$$\max\{F^{(1/n)Y_mY_m^*}(0), F^{T_m}(0)\} \leqslant F^{M_n}(0) \leqslant F^{(1/n)Y_mY_m^*}(0) + F^{T_m}(0). \tag{6.14}$$

Since 0 is not a mass point of H, it is a point a continuity of H and, therefore, from (ii) in Proposition 1, almost surely, $(F^{T_m}(0))_{m\geqslant 1}$ converges to 0. Thus, from (6.14),

$$|F^{(1/n)Y_mY_m^*}(0) - F^{M_m}(0)| \xrightarrow{\text{a.s.}} 0 \tag{6.15}$$

When $y \leqslant 1$, from (3.10), $F^{(1/n)Y_mY_m^*}(0) \xrightarrow{\text{a.s.}} 0$. Fix an arbitrary δ in $\mathbb{R}_+^*$. Let ε be in $]0,(1-\sqrt{y})^2[$ such that $F^H(\varepsilon) < \delta$. By (3.10) again, $F^{(1/n)Y_mY_m^*}(\varepsilon) \xrightarrow{\text{a.s.}} 0$. Moreover, from (ii) in Proposition 1, almost surely, $\limsup_{m\to+\infty} F^{T_m}(\varepsilon) < \delta$. Then, by Lemma 1, the almost sure limit of $F^{M_n}(\varepsilon^2)$ is less than δ and it follows that $F(0) < \delta$. Since δ can be made arbitrarily small, we conclude that

$$F^{M_m}(0) \xrightarrow{\text{a.s.}} 0 = F(0). \tag{6.16}$$

When $y > 1$, from Section 4.2, $F^{(1/n)Y_mY_m^*}(0) \xrightarrow{\text{a.s.}} 1-(1/y)$. Therefore, it will follow from (6.15) and the same argument that

$$F^{M_m}(0) \xrightarrow{\text{a.s.}} 1 - \frac{1}{y} = F(0). \tag{6.17}$$

The proof is complete since, if a sequence of d.f.'s $(F_n)_{n\geqslant 1}$ converges weakly to a d.f. F and if $(F_n(x\pm))_{n\geqslant 1}$ converges to $F(x\pm)$ at every point x of discontinuity of F, then $(F_n)_{n\geqslant 1}$ converges to F uniformly in $\mathbb{R}$ [3].

Proof of Theorem 3. First, notice that the eigenvalues of $\widehat{R}$ are the same as the p largest eigenvalues of

$$\frac{1}{n}VV^*[B \;\; \sigma I_p]^*[B \;\; \sigma I_p]. \tag{6.18}$$

Thus, for p and n sufficiently large, we see that the empirical d.f. $F^{\widehat{R}}$ of the eigenvalues of $\widehat{R}$, together with q zeros, is close to the nonrandom limiting d.f. guaranteed by Theorem 1 where $m = p+q$, $y = (p+q)/n$, and $H = F^{[B \;\; \sigma I_p]^*[B \;\; \sigma I_p]}$. For the purpose of removing the q singularities, note that the limiting d.f. in Theorem 1 does not depend on the d.f. of Y_{11}. Therefore, without loss of generality, we may assume that the entries of V are standardized complex Gaussian. Now, if we let $O^*\Lambda O$ denote the spectral decomposition of $[B \;\; \sigma I_p]^*[B \;\; \sigma I_p]$, where the eigenvalues are arranged in nonincreasing order along the diagonal of Λ, then the eigenvalues of the matrix in (6.18) are the same as those of

$$\frac{1}{n}\Lambda^{1/2}OVV^*O^*\Lambda^{1/2}. \tag{6.19}$$

Since the entries of V are i.i.d. standardized complex Gaussian, so are the entries of OV. Notice the entries of the matrix in (6.19) outside the upper left $p \times p$ submatrix are zero. Therefore, the spectrum of $\widehat{R}$ is the same as that of the $p \times p$ upper block. Note also that the p largest eigenvalues of $[B \;\; \sigma I_p]^*[B \;\; \sigma I_p]$ are the same as those of

$$[B \;\; \sigma I_p][B \;\; \sigma I_p]^* = BB^* + \sigma^2 I_p. \tag{6.20}$$

Let $P^*\Lambda' P$ denote the spectral decomposition of $BB^* + \sigma^2 I_p$, and let Z denote the first p rows of OV. Then, the spectrum of $\widehat{R}$ is the same as that of

$$\frac{1}{n}ZZ^*\Lambda' = \frac{1}{n}ZZ^*P\big(BB^* + \sigma^2 I_p\big)P^*, \tag{6.21}$$

which is the same as that of

$$\frac{1}{n}Y_pY_p^*(BB^* + \sigma^2 I_p) = \widehat{R}', \tag{6.22}$$

where $Y_p = P^*Z$ is $p \times n$ and contains i.i.d. standardized complex Gaussian entries.

References

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis,* second edition. New York: John Wiley & Sons, 1984.

[2] Z. D. Bai, J. W. Silverstein, 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, August 1988.

[3] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales,* second edition. New York: Springer-Verlag, 1988.

[4] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," *Proceedings of the National Academy of Sciences (USA),* vol. 37, no. 11, pp. 760–766, November 1951.

[5] F. R. Gantmacher, *The Theory of Matrices,* vol. 1. New York: Chelsea, 1977.

[6] S. Geman, "A limit theorem for the norm of random matrices," *The Annals of Probability,* vol. 8, no. 2, pp. 252–261, April 1980.

[7] U. Grenander and J. W. Silverstein, "Spectral analysis of networks with random topologies," *SIAM Journal on Applied Mathematics,* vol. 32, no. 2, pp. 499–519, March 1977.

[8] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *Journal of Multivariate Analysis,* vol. 12, no. 1, pp. 1–38, March 1982.

[9] M. Kaveh, H. Wang, and H. Hung, "On the theoretical performance of a class of estimators of the number of narrow-band sources," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. ASSP-35, no. 9, pp. 1350–1352, September 1987.

[10] P. R. Krishnaiah, "Some recent developments on complex multivariate distributions," *Journal of Multivariate Analysis,* vol. 6, no. 1, pp. 1–30, March 1976.

[11] Michel Loève, *Probability Theory,* vol. 1, fourth edition. New York: Springer-Verlag, 1977.

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

[13] M. Okamoto, "Distinctness of the eigenvalues of a quadratic form in a multivariate sample," *The Annals of Statistics,* vol. 1, no. 4, pp. 763–765, July 1973.

[14] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate F matrix," *SIAM Journal on Mathematical Analysis,* vol. 16, no. 3, pp. 641–646, May 1985.

[15] J. W. Silverstein, "The smallest eigenvalue of a large dimensional Wishart matrix," *The Annals of Probability,* vol. 13, no. 4, pp. 1364–1368, November 1985.

[16] J. W. Silverstein, "On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix," *Journal of Multivariate Analysis,* vol. 30, no. 2, pp. 307–311, August 1989.

[17] K. W. Wachter, "The strong limits of random matrix spectra for sample matrices of independent elements," *The Annals of Probability,* vol. 6, no. 1, pp. 1–18, February 1978.

[18] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. ASSP-33, no. 2, pp. 387–392, April 1985.

[19] Y. Q. Yin, "Limiting spectral distribution for a class of random matrices," *Journal of Multivariate Analysis,* vol. 20, no. 1, pp. 50–68, October 1986.

[20] Y. Q. Yin and P. R. Krishnaiah, "On some nonparametric methods for detection of the number of signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. ASSP-35, no. 11, pp. 1533–1538, November 1987.

[21] 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, August 1988.

[22] Q.-T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. ASSP-37, no. 10, pp. 1557–1567, October 1989.

[23] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis,* vol. 20, no. 1, pp. 1–25, October 1986.

[24] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. ASSP-35, no. 2, pp. 129–132, February 1987.