7 Spectrum analysis

In this chapter, we further study the spectra of the important random matrix models for wireless communications that are the sample covariance matrix and the information plus noise models. It has already been shown in Chapter 3 that, as the e.s.d. of the population covariance matrix (or of the information matrix) converges, the e.s.d. of the sample covariance matrix (or the information plus noise matrix) converges almost surely. The limiting d.f. can then be fully characterized as a function of the l.s.d. of the population covariance matrix (or of the information matrix). It is however not convenient to invert the problem and to describe the l.s.d. of the population covariance matrix (or of the information matrix) as a function of the l.s.d. of the observed matrices. The answer to this inverse problem is provided in Chapter 8, which however requires some effort to be fully accessible. The development of the tools necessary for the statistical eigen-inference methods of Chapter 8 is one of the motivations of the current chapter.

The starting motivation, initiated by the work of Silverstein and Choi [Silverstein and Choi, 1995], which resulted in the important Theorem 7.4 (accompanied later by an important corollary, due to Mestre [Mestre, 2008a], Theorem 7.5), was to characterize the l.s.d. of the sample covariance matrix in closed-form. Remember that, up to this point, we can only characterize the l.s.d. F of a sample covariance matrix through the expression of its Stieltjes transform, as the unique solution $m_F(z)$ of some fixed-point equation for all $z \in \mathbb{C} \setminus \mathbb{R}^+$. To obtain an explicit expression of F, it therefore suffices to use the inverse Stieltjes transform formula (3.2). However, this suggests having a closer look at the limiting behavior of $m_F(z)$ as z approaches the positive real half-line, about which we do not know much yet. Therefore, up to this point in our analysis, it is impossible to describe the support of the l.s.d., apart from rough estimations based on the expression of $\Im[m_F(z)]$, for $z = x + iy$, y being small. It is also not convenient to depict $F'(x)$: the solution is to take $z = x + iy$, with y small and x spanning from zero to infinity, and to draw the curve $z \mapsto \frac{1}{\pi}\Im[m_F(z)]$ for such z. In the following, we will show that, as z tends to $x > 0$, $m_F(z)$ has a limit which can be characterized in two different ways, depending on whether x belongs to the support of F or not. In any case, this limit is also characterized as the solution to an implicit equation, although particular care must be taken as to which of the multiple solutions of this implicit equation needs to be considered.

Before we detail this advanced spectrum characterization, we provide a different set of results, fundamental to the validation of the eigen-inference methods proposed in Chapter 8. These results, namely the asymptotic absence of eigenvalues outside the support of F, Theorem 7.1, and the exact separation of the support into disjoints clusters, Theorem 7.2, are once more due to Bai and Silverstein [Bai and Silverstein, 1998, 1999]. Their object is the analysis, on top of the characterization of F, of the behavior of the particular eigenvalues of the e.s.d. of the sample covariance matrix as the dimensions grow large. It is fundamental to understand here, and this will be reminded again in the next section, that the convergence of the e.s.d. toward F, as the matrix size N grows large, does *not* imply the convergence of the largest eigenvalue of the sample covariance matrix towards the right edge of the support. Indeed, the largest eigenvalues, having weight $1/N$ in the spectrum, do not contribute asymptotically to the support of F. As such, it may well be found outside the support of F for all finite N, without invalidating Theorem 3.13. This particular case in the Marčenko–Pastur model where eigenvalues are found outside the support almost surely when the entries of the random i.i.d. matrix $\mathbf{X}_N$ in Theorem 3.13, $\mathbf{T}_N = \mathbf{I}_N$, have infinite fourth order moment. In this scenario, it is even proved in [Silverstein *et al.*, 1988] that the largest eigenvalue grows without bound as the system dimensions grow to infinity, while all the mass of the l.s.d. is asymptotically kept in the support; if the fourth order moment is finite. Under finite fourth moment assumption though [Bai and Silverstein, 1998; Yin *et al.*, 1988], the important result to be detailed below is that no eigenvalue is to be found outside the limiting support and that the eigenvalues are found where they ought to be. This last statement is in fact slightly erroneous and will be adequately corrected when discussing the *spiked* models that lead some eigenvalues to leave the limiting support. To be more precise, when the moment of order four of the entries of $\mathbf{X}_N$ exists, we can characterize exactly the subsets of $\mathbb{R}^+$ where no eigenvalue is asymptotically found, almost surely. Further discussions on the extreme eigenvalues of sample covariance matrices are provided in Chapter 9, where (non-central) limiting theorems for the distribution of these eigenvalues are provided.

7.1 Sample covariance matrix

7.1.1 No eigenvalues outside the support

As observed in the previous sections, most early results of random matrix theory dealt with the limiting behavior of e.s.d. For instance, the Marčenko–Pastur law ensures that the e.s.d. of the sample covariance matrix $\mathbf{R}_N$ of vectors with i.i.d. entries of zero mean and unit variance converges almost surely towards a limit distribution function F. However, the Marčenko–Pastur law does not say anything about the behavior of any specific eigenvalue, say for instance the extreme lowest and largest eigenvalues $\lambda_{\min}$ and $\lambda_{\max}$ of $\mathbf{R}_N$. It is relevant in particular to wonder whether $\lambda_{\min}$ and $\lambda_{\max}$ can be asymptotically found *outside* the support of F. Indeed, if all eigenvalues but the extreme two are in the support of F, then the l.s.d. of $\mathbf{R}_N$ is still F, which is still consistent with the Marčenko–Pastur law. It turns out that this is not the case in general. Under some mild assumption on the entries of the sample covariance matrix, no eigenvalue is found outside the support. We specifically have the following theorem.

Theorem 7.1 ([Bai and Silverstein, 1998; Yin *et al.*, 1988])**.** *Let the matrix* $\mathbf{X}_N = \left(\frac{1}{\sqrt{n}}X_{N,ij}\right) \in \mathbb{C}^{N\times n}$ *have i.i.d. entries, such that* $X_{N,11}$ *has zero mean, unit variance, and finite fourth order moment. Let* $\mathbf{T}_N \in \mathbb{C}^{N\times N}$ *be non-random, with uniformly bounded spectral norm* $\|\mathbf{T}_N\|$*, whose e.s.d.* $F^{\mathbf{T}_N}$ *converge weakly to* H*. From Theorem 3.13, the e.s.d. of* $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}} \in \mathbb{C}^{N\times N}$ *converges weakly and almost surely towards some distribution function* F*, as* N*,* n *go to infinity with ratio* $c_N = N/n \to c$*,* $0 < c < \infty$*. Similarly, the e.s.d. of* $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N\mathbf{X}_N \in \mathbb{C}^{n\times n}$ *converges towards* $\underline{F}$ *given by:*

$$\underline{F}(x) = cF(x) + (1-c)1_{[0,\infty)}(x).$$

Denote $\underline{F}_N$ *the distribution with Stieltjes transform* $m_{\underline{F}_N}(z)$*, which is solution, for* $z \in \mathbb{C}^+$*, of the following equation in* m

$$m = -\left(z - \frac{N}{n}\int \frac{\tau}{1+\tau m} dF^{\mathbf{T}_N}(\tau)\right)^{-1} \tag{7.1}$$

and define F_N *the d.f. such that*

$$\underline{F}_N(x) = \frac{N}{n}F_N(x) + \left(1 - \frac{N}{n}\right)1_{[0,\infty)}(x).$$

Let $N_0 \in \mathbb{N}$*, and choose an interval* $[a,b]$*,* $a,b \in (0,\infty]$*, lying in an open interval outside the union of the supports of* F *and* F_N *for all* $N \geq N_0$*. For* $\omega \in \Omega$*, the random space generating the series* $\mathbf{X}_1, \mathbf{X}_2, \ldots$*, denote* $\mathcal{L}_N(\omega)$ *the set of eigenvalues of* $\mathbf{B}_N(\omega)$*. Then*

$$P(\{\omega, \mathcal{L}_N(\omega) \cap [a,b] \neq \emptyset \text{ i.o.}\}) = 0.$$

This means concretely that, given a segment $[a,b]$ outside the union of the supports of F and $F_{N_0}, F_{N_0+1}, \ldots$, for all series $\mathbf{B}_1(\omega), \mathbf{B}_2(\omega), \ldots$, with ω in some set of probability one, there exists $M(\omega)$ such that, for all $N \geq M(\omega)$, there will be no eigenvalue of $\mathbf{B}_N(\omega)$ in $[a,b]$. By definition, F_K is the l.s.d. of an hypothetical $\mathbf{B}_N$ with $H = F^{\mathbf{T}_K}$. The necessity to consider the supports of $F_{N_0}, F_{N_0+1}, \ldots$ is essential when a few eigenvalues of $\mathbf{T}_N$ are isolated and eventually contribute with probability zero to the l.s.d. H. Indeed, it is rather intuitive that, if the largest eigenvalue of $\mathbf{T}_N$ is large compared to the rest, at least one eigenvalue of $\mathbf{B}_N$ will also be large compared to the rest (take $n \gg N$ to be convinced). Theorem 7.1 states exactly here that there will be neither any eigenvalue outside the support of the main mass of $F^{\mathbf{B}_N}$, nor any eigenvalue around the largest one. Those models in which some eigenvalues of $\mathbf{T}_N$ are isolated are referred to as *spiked models.* These are thoroughly discussed in Chapter 9. In wireless communications and modern signal processing, Theorem 7.1 is of key importance for signal sensing and hypothesis testing methods since it allows us to verify whether the eigenvalues empirically found in sample covariance matrix spectra originate either from noise contributions or from signal sources. In the simple case where signals sensed at an antenna array originate either from white noise or from a coherent signal source impaired by white noise, this can be performed by simply verifying if the extreme eigenvalue of the sample covariance matrix is inside or outside the support of the Marčenko–Pastur law (Figure 1.1); see further Chapter 16.

We give hereafter a sketch of the proof, which again only involves the Stieltjes transform.

Proof. Surprisingly, the proof unfolds from a mere (though non-trivial) refinement of the Stieltjes transform relation proved in Theorem 3.13. Let F_N be defined as above and let m_N be its Stieltjes transform. It is possible to show that, for $z = x + iv_N$, with $v_N = N^{-1/68}$

$$\sup_{x\in[a,b]} |m_{\mathbf{B}_N}(z) - m_N(z)| = o\left(\frac{1}{N}v_N\right)$$

almost surely. This result is in fact also true when $\Im[z]$ equals $\sqrt{2}v_N, \sqrt{3}v_N, \ldots$ or $\sqrt{34}v_N$. Note that this refines the known statement that the difference is of order $o(1)$. We take this property, which requires more than ten pages of calculus, for granted. We now have that

$$\max_{1\leq k\leq 34} \sup_{x\in[a,b]} \left| m_{\mathbf{B}_N}(x + ik^{\frac{1}{2}}v_N) - m_N(x + ik^{\frac{1}{2}}v_N)\right| = o(v_N^{67})$$

almost surely. Expanding the Stieltjes transforms and considering only the imaginary parts, we obtain

$$\max_{1\leq k\leq 34} \sup_{x\in[a,b]} \left| \int \frac{d(F^{\mathbf{B}_N}(\lambda) - F_N(\lambda))}{(x-\lambda)^2 + kv_N^2}\right| = o(v_N^{66})$$

almost surely. Taking successive differences over the 34 values of k, we end up with

$$\sup_{x\in[a,b]} \left| \int \frac{(v_N^2)^{33} d(F^{\mathbf{B}_N}(\lambda) - F_N(\lambda))}{\prod_{k=1}^{34}((x-\lambda)^2 + kv_N^2)}\right| = o(v_N^{66}) \tag{7.2}$$

almost surely, from which the term v_N^{66} simplifies on both sides. Consider now $a' < a$ and $b' > b$ such that $[a', b']$ is outside the support of F. We then divide

(7.2) into two terms, as (remember that $1/N = v_N^{68}$)

$$\sup_{x\in[a,b]} \left| \int \frac{1_{\mathbb{R}^+\setminus[a',b']}(\lambda) d(F^{\mathbf{B}_N}(\lambda) - F_N(\lambda))}{\prod_{k=1}^{34}((x-\lambda)^2 + kv_N^2)} + \sum_{\lambda_j\in[a',b']} \frac{v_N^{68}}{\prod_{k=1}^{34}((x-\lambda_j)^2 + kv_N^2)} \right|$$
$$= o(1)$$

almost surely. Assume now that, for a subsequence $\phi(1), \phi(2), \ldots$ of $1, 2, \ldots$, there always exists at least one eigenvalue of $\mathbf{B}_{\phi(N)}$ in $[a, b]$. Then, for x taken equal to this eigenvalue, one term of the discrete sum above (whose summands are all non-negative) is exactly $1/34!$, which is uniformly bounded away from zero. This implies that the integral must also be bounded away from zero. However the integrand of the integral is clearly uniformly bounded on $[a', b']$ and, from Theorem 3.13, $F^{\mathbf{B}_N} - F \Rightarrow 0$. Therefore the integral tends to zero as $N \to \infty$. This is a contradiction. Therefore, the probability that there is an eigenvalue of $\mathbf{B}_N$ in $[a, b]$ infinitely often is null. Now, from [Yin *et al.*, 1988], the largest eigenvalue of $\frac{1}{n}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ is almost surely asymptotically bounded. Therefore, since $\|\mathbf{T}_N\|$ is also bounded by hypothesis, the theorem applies also to $b = \infty$. □

Note that the finiteness of the fourth order moment of the entries $X_{N,ij}$ is fundamental for the validity of Theorem 7.1. It is indeed proved in [Yin *et al.*, 1988] and [Silverstein *et al.*, 1988] that:

- if the entries $X_{N,ij}$ have finite fourth order moment, with probability one, the largest eigenvalue of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ tends to the edge $(1 + \sqrt{c})^2$, $c = \lim_N N/n$ of the support of the Marčenko–Pastur law, which is an immediate corollary of Theorem 7.1 with $\mathbf{T}_N = \mathbf{I}_N$;
- if the entries $X_{N,ij}$ do not have a finite fourth order moment then, with probability one, the limit superior of the largest eigenvalue of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ is infinite, i.e. with probability one, for all $A > 0$, there exists N such that the largest eigenvalue of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ is larger than A. It is therefore important never to forget the underlying assumption made on the tails of the distribution of the entries in $\mathbf{X}_N$.

We now move to an extension of Theorem 7.1.

7.1.2 Exact spectrum separation

Now assume that the e.s.d. of $\mathbf{T}_N$ converges to the distribution function of, say, three evenly weighted masses in λ_1, λ_2, and λ_3. For not-too-large ratios $c_N = N/n$, it is observed that the support of F is divided into up to three clusters of eigenvalues. In particular, when n becomes large while N is kept fixed, the clusters consist of three punctual masses in λ_1, λ_2, and λ_3, as required by classical probability theory. This is illustrated in Figure 7.1 in the case of a three-fold clustered and a two-fold clustered support of F. The reason why we observe sometimes three and sometimes less clusters is linked to the spreading of each cluster due to the limiting ratio c; the smaller c, the thinner the clusters, as already observed in the simple case of the Marčenko–Pastur law, Figure 2.2. Considering Theorem 7.1, it is tempting to assume that, in addition to each cluster of F being composed of one third of the total spectrum mass, each cluster of $\mathbf{B}_N$ contains exactly one third of the eigenvalues of $\mathbf{B}_N$. However, Theorem 7.1 only ensures that no eigenvalue is found outside the support of F for all N larger than a given M, and does not say how the eigenvalues of $\mathbf{B}_N$ are distributed in the various clusters. The answer to this question is provided in [Bai and Silverstein, 1999] in which the exact separation properties of the l.s.d. of such matrices $\mathbf{B}_N$ is discussed.

Theorem 7.2 ([Bai and Silverstein, 1999]). *Assume the hypothesis of Theorem 7.1 with $\mathbf{T}_N$ non-negative definite. Consider similarly $0 < a < b < \infty$ such that $[a, b]$ lies in an open interval outside the support of F and F_N for all large N. Denote additionally λ_k and τ_k, $k \in \{1, \ldots, iN\}$ the eigenvalues of $\mathbf{B}_N$ and $\mathbf{T}_N$, respectively. Then we have:*

1. *If $c(1 - H(0)) > 1$, then the smallest value x_0 in the support of F is positive and $\lambda_N \to x_0$ almost surely, as $N \to \infty$.*
2. *If $c(1 - H(0)) \leq 1$, or $c(1 - H(0)) > 1$ but $[a, b]$ is not contained in $[0, x_0]$, then, with probability one, and for all large N*[1]

$$\# \{k,\ \lambda_k < a\} = \# \{k,\ \tau_k < -1/m_F(a)\}$$
$$\# \{k,\ \lambda_k > b\} = \# \{k,\ \tau_k > -1/m_F(b)\}.$$

The original statement of [Bai and Silverstein, 1999], equivalent to the above but less explicit, is that, with the ordering $\tau_1 \geq \ldots \geq \tau_N$ and $\lambda_1 \geq \ldots \geq \lambda_N$,

$$P(\lambda_{i_N} > b, \lambda_{i_N+1} < a \text{ for all large } N) = 1$$

where i_N is the unique integer such that

$$\tau_{i_N} > -1/m_F(b),$$
$$\tau_{i_N+1} < -1/m_F(a).$$

To understand this statement, consider for instance the first plot in Figure 7.1 and an interval $[a, b]$ comprised between the second and third clusters. What the above statement claims is that, if i_N and $i_N + 1$ are the indexes of the right and left eigenvalues when $F^{\mathbf{B}_N}$ jumps from one cluster to the next, and N is

[1] The expression "the set (or event) $A_N \subset \Omega$ holds with probability one, for all large N" is used in place of "there exists $B \subset \Omega$, with $P(B) = 1$, such that, for $\omega \in B$, there exists $N_0(\omega)$ for which $N > N_0(\omega)$ implies $\omega \in A_N$." It is particularly important to note that "for all large N" is somewhat misleading as it does *not* indicate the existence of a universal N_0 such that $N > N_0$ implies $\omega \in A_N$ for all $\omega \in B$, but rather the existence of an $N_0(\omega)$ for each such ω. Here, for instance, $A_N = \{\omega,\ \#\{\lambda_k(\omega) < a\} = \#\{\tau_k(\omega) < -1/m_F(a)\}\}$ and the space Ω is the generator of the series $\mathbf{B}_1(\omega), \mathbf{B}_2(\omega), \ldots$.

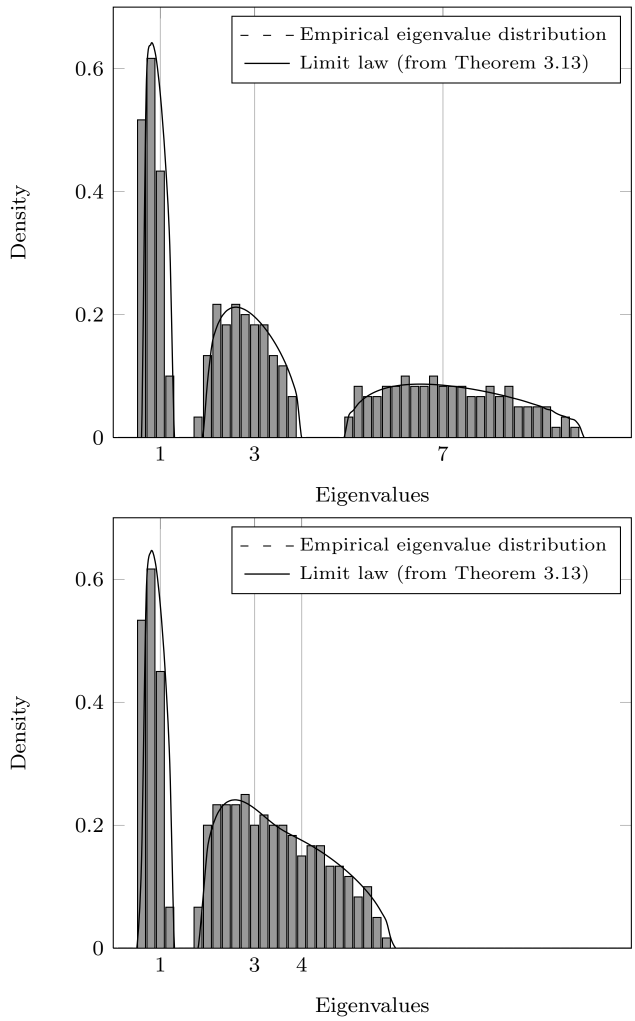

Figure 7.1 Histogram of the eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, $N = 300$, $n = 3000$, with $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in (i) 1, 3, and 7 on top, (ii) 1, 3, and 4 on the bottom.

large enough, then there is an associated jump from the corresponding i_Nth and $(i_N + 1)$th eigenvalues of $\mathbf{T}_N$ (for instance, at the position of the discontinuity from eigenvalue 7 to eigenvalue 3).

This bears some importance for signal detection. Indeed, consider the problem of the transmission of information plus noise. Given the dimension p of the signal space and $n - p$ of the noise space, for large c, Theorem 7.2 allows us to isolate the eigenvalues corresponding to the signal space from those corresponding to the noise space. If both eigenvalue spaces are isolated in two distinct clusters, then we can exactly determine the dimension of each space and infer, e.g. the number of transmitting entities. The next question that then naturally arises is to determine for which values of $c = \lim_N n/N$ the support of F separates into 1, 2, or more clusters.

7.1.3 Asymptotic spectrum analysis

For better understanding in the following, we will take the convention that the (hypothetical) single mass at zero in the spectrum of F is *not* considered as a 'cluster'. We will number the successive clusters from left to right, from one to K_F with K_F the number of clusters in F, and we will denote k_F the cluster generated by the population eigenvalue t_k, to be introduced shortly. For instance, if two sample eigenvalues t_i and $t_{i+1} \neq t_i$ generate a unique cluster in F (as in the bottom graph in Figure 7.1, where $t_2 = 3$ and $t_3 = 4$ generate the same cluster), then $i_F = (i+1)_F$). The results to come will provide a unique way to define k_F mathematically and not only visually. To this end, we need to study in more depth the properties of the limiting spectrum F of the sample covariance matrix.

Remember first that, for the model $\underline{\mathbf{B}}_N = \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N \mathbf{X}_N \in \mathbb{C}^{n \times n}$ of l.s.d. $\underline{F}$, where $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ has i.i.d. entries of zero mean and variance $1/n$, $\mathbf{T}_N$ has l.s.d. H and $N/n \to c$, $m_{\underline{F}}(z)$, $z \in \mathbb{C}^+$, Equation (3.22) has an inverse formula, given by:

$$z_{\underline{F}}(\underline{m}) = -\frac{1}{\underline{m}} + c \int \frac{t}{1 + t\underline{m}} dH(t) \tag{7.3}$$

for $\underline{m} \in \mathbb{C}^+$. The equation $z_{\underline{F}}(\underline{m}) = z \in \mathbb{C}^+$ has a unique solution $\underline{m}$ with positive imaginary part and this solution equals $m_{\underline{F}}(z)$ by Theorem 3.13. Of course, $\mathbf{B}_N$ and $\underline{\mathbf{B}}_N$ only differ from $|N - n|$ zero eigenvalues, so it is equivalent to study the l.s.d. of $\mathbf{B}_N$ or that of $\underline{\mathbf{B}}_N$. The link between their respective Stieltjes transforms is given by:

$$m_{\underline{F}}(z) = c m_F(z) + (c-1)\frac{1}{z}$$

from (3.16). Since $\underline{F}$ turns out to be simpler to study, we will focus on $\underline{\mathbf{B}}_N$ instead of the sample covariance matrix $\mathbf{B}_N$ itself.

Now, according to the Stieltjes inversion formula (3.2), for every continuity points a, b of $\underline{F}$

$$\underline{F}(b) - \underline{F}(a) = \lim_{y \to 0^+} \frac{1}{\pi} \int_a^b \Im[m_{\underline{F}}(x + iy)] dx.$$

To determine the distribution $\underline{F}$, and therefore the distribution F, we must determine the limit of $m_{\underline{F}}(z)$ as $z \in \mathbb{C}^+$ tends to $x \in \mathbb{R}^*$. It can in fact be shown that this limit exists.

Theorem 7.3 ([Silverstein and Choi, 1995]). *Let $\underline{\mathbf{B}}_N \in \mathbb{C}^{n \times n}$ be defined as previously, with almost sure l.s.d. $\underline{F}$. Then, for $x \in \mathbb{R}^*$*

$$\lim_{\substack{z \to x \\ z \in \mathbb{C}^+}} m_{\underline{F}}(z) \triangleq m^\circ(x) \tag{7.4}$$

exists and the function m° *is continuous on* $\mathbb{R}^*$*. For* x *in the support of* $\underline{F}$*, the density* $\underline{f}(x) \triangleq \underline{F}'(x)$ *equals* $\frac{1}{\pi}\Im[m^\circ(x)]$*. Moreover,* $\underline{f}$ *is analytic for all* $x \in \mathbb{R}^*$ *such that* $\underline{f}(x) > 0$*.*

The study of m° makes it therefore possible to describe the complete support $S_{\underline{F}}$ of $\underline{F}$ as well as the limiting density $\underline{f}$. Since $S_{\underline{F}}$ equals S_F but for an additional mass in zero, this is equivalent to determining the support of S_F. Choi and Silverstein provided an accurate description of the function m°, as follows.

Theorem 7.4 ([Silverstein and Choi, 1995])**.** *Let* $B = \{\underline{m} \mid \underline{m} \neq 0, -1/\underline{m} \in S_H^c\}$*, with* S_H^c *the complementary of* S_H*, and* $x_{\underline{F}}$ *be the function defined on* B *by*

$$x_{\underline{F}}(\underline{m}) = -\frac{1}{\underline{m}} + c\int \frac{t}{1+t\underline{m}}dH(t). \tag{7.5}$$

For $x \in \mathbb{R}^*$*, we can determine the limit* $m^\circ(x)$ *of* $m_{\underline{F}}(z)$ *as* $z \to x$*,* $z \in \mathbb{C}^+$*, along the following rules:*

1. *If* $x \in S_{\underline{F}}$*, then* $m^\circ(x)$ *is the unique solution in* B *with positive imaginary part of the equation* $x = x_{\underline{F}}(\underline{m})$ *in the dummy variable* $\underline{m}$*.*
2. *If* $x \in S_{\underline{F}}^c$*, then* $m^\circ(x)$ *is the unique real solution in* B *of the equation* $x = x_{\underline{F}}(\underline{m})$ *in the dummy variable* $\underline{m}$ *such that* $x'_{\underline{F}}(m_0) > 0$*. Conversely, for* $\underline{m} \in B$*, if* $x'_{\underline{F}}(\underline{m}) > 0$*, then* $x_{\underline{F}}(\underline{m}) \in S_{\underline{F}}^c$*.*

From rule 1, along with Theorem 7.3, we can evaluate for every $x > 0$ the limiting density $\underline{f}(x)$, hence $F(x)$, by finding the complex solution with positive imaginary part of $x = x_{\underline{F}}(\underline{m})$.

Rule 2 makes it simple to determine analytically the exact support of F. It indeed suffices to draw $x_{\underline{F}}(\underline{m})$ for $-1/\underline{m} \in S_H^c$. Whenever $x_{\underline{F}}$ is increasing on an interval I, $x_{\underline{F}}(I)$ is outside $S_{\underline{F}}$. The support $S_{\underline{F}}$ of $\underline{F}$, and therefore of F (modulo the mass in zero), is then defined exactly by the complementary set

$$S_{\underline{F}} = \mathbb{R} \setminus \bigcup_{\substack{a,b\in\mathbb{R}\\ a<b}} \left\{ x_{\underline{F}}((a,b)) \mid \forall \underline{m} \in (a,b), x'_{\underline{F}}(\underline{m}) > 0 \right\}.$$

This is depicted in Figure 7.2 in the case when H is composed of three evenly weighted masses t_1, t_2, t_3 in $\{1, 3, 5\}$ or $\{1, 3, 10\}$ and $c = 1/10$. Notice that, in the case where $t_3 = 10$, F is divided into three clusters, while, when $t_3 = 5$, F is divided into only two clusters, which is due to the fact that $x_{\underline{F}}$ is non-increasing in the interval $(-1/3, -1/5)$. For applicative purposes, we will see in Chapter 17 that it might be essential that the consecutive clusters be disjoint. This is one reason why Theorem 7.6 is so important.

We do not provide a rigorous proof of Theorem 7.4. In fact, while thoroughly proved in 1995, this result was already intuited by Marčenko and Pastur in 1967 [Marčenko and Pastur, 1967]. The fact that $x_{\underline{F}}(\underline{m})$ increases outside the spectrum of $\underline{F}$ and is not increasing elsewhere is indeed very intuitive, and is not

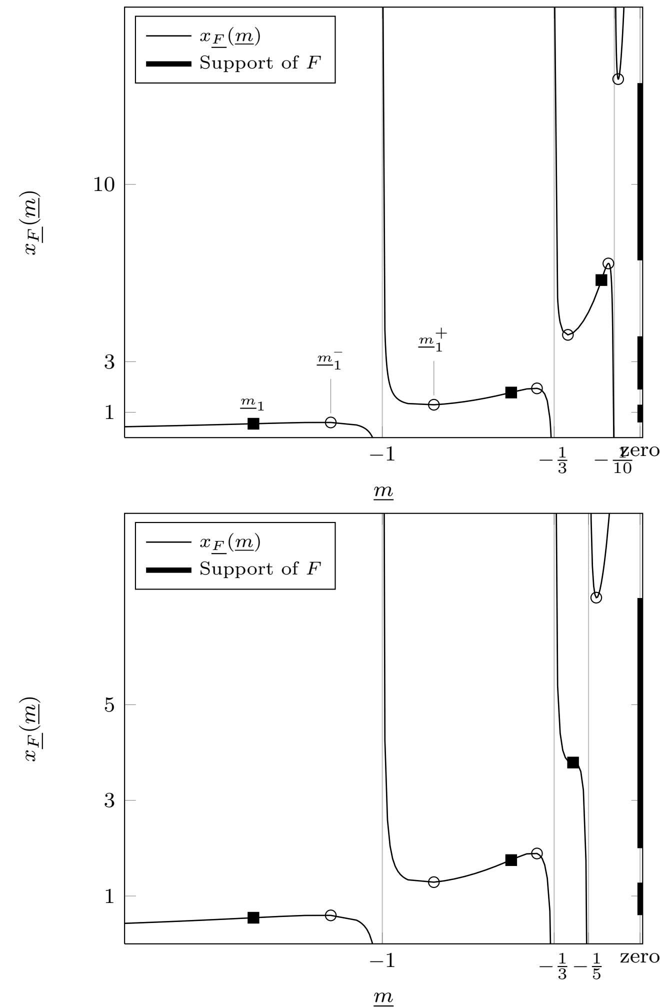

Figure 7.2 $x_{\underline{F}}(\underline{m})$ for $\underline{m}$ real, $\mathbf{T}_N$ diagonal composed of three evenly weighted masses in 1, 3, and 10 (top) and 1, 3, and 5 (bottom), $c = 1/10$ in both cases. Local extrema are marked in circles, inflexion points are marked in squares. The support of F can be read on the right vertical axises.

actually limited to the sample covariance matrix case. Observe indeed that, for any F, and any $x_0 \in \mathbb{R}^*$ outside the support of F, $m_F(x_0)$ is clearly well defined and

$$m'_F(x_0) = \int \frac{1}{(\lambda - x_0)^2} dF(\lambda) > 0.$$

Therefore $m_F(x)$ is continuous and increasing on an open neighborhood of x_0. This implies that it is locally a one-to-one mapping on this neighborhood and therefore admits an inverse $x_F(m)$, which is also continuous and increasing. This explains why $x_F(m)$ increases when its image is outside the spectrum of F.

Now, if for some real m_0, $x_F(m_0)$ is continuous and increasing, then it is locally invertible and its inverse ought to be $m_F(x)$, continuous and increasing, in which case x is outside the spectrum of F. Obviously, this reasoning is far from being a proof (at least the converse requires much more work).

From Figure 7.2 and Theorem 7.4, we now observe that, when the e.s.d. of population matrix is composed of a few masses, $x'_{\underline{F}}(\underline{m}) = 0$ has exactly $2K_F$ solutions with K_F the number of clusters in F. Denote these roots in increasing order $\underline{m}_1^- < \underline{m}_1^+ \leq \underline{m}_2^- < \underline{m}_2^+ < \ldots \leq \underline{m}_{K_F}^- < \underline{m}_{K_F}^+$. Each pair $(\underline{m}_j^-, \underline{m}_j^+)$ is such that $x_{\underline{F}}([\underline{m}_j^-, \underline{m}_j^+])$ is the jth cluster in F. We therefore have a way to determine the support of the asymptotic spectrum through the function $x'_{\underline{F}}$. This is presented in the following result.

Theorem 7.5 ([Couillet *et al.*, 2011c; Mestre, 2008a]). *Let $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ be defined as in Theorem 7.1. Then the support S_F of the l.s.d. F of $\mathbf{B}_N$ is*

$$S_F = \bigcup_{j=1}^{K_F} [x_j^-, x_j^+]$$

where $x_1^-, x_1^+, \ldots, x_{K_F}^-, x_{K_F}^+$ are defined by

$$\begin{aligned} x_j^- &= -\frac{1}{\underline{m}_j^-} + \sum_{r=1}^{K} c_r \frac{t_r}{1 + t_r \underline{m}_j^-} \\ x_j^+ &= -\frac{1}{\underline{m}_j^-} + \sum_{r=1}^{K} c_r \frac{t_r}{1 + t_r \underline{m}_j^+} \end{aligned}$$

with $\underline{m}_1^- < \underline{m}_1^+ \leq \underline{m}_2^- < \underline{m}_2^+ \leq \ldots \leq \underline{m}_{K_F}^- < \underline{m}_{K_F}^+$ the $2K_F$ (possibly counted with multiplicity) real roots of the equation in $\underline{m}$

$$\sum_{r=1}^{K} c_r \frac{t_r^2 \underline{m}^2}{(1 + t_r \underline{m}^2)^2} = 1.$$

Note further from Figure 7.2 that, while $x'_{\underline{F}}(\underline{m})$ might not have roots on some intervals $(-1/t_{k-1}, -1/t_k)$, it always has a unique inflexion point there. This is proved in [Couillet *et al.*, 2011c] by observing that $x''_{\underline{F}}(\underline{m}) = 0$ is equivalent to

$$\sum_{r=1}^{K} c_r \frac{t_r^3 \underline{m}^3}{(1 + t_r \underline{m})^3} - 1 = 0$$

the left-hand side of which has always positive derivative and shows asymptotes in the neighborhood of t_r; hence the existence of a unique inflexion point on every interval $(-1/t_{k-1}, -1/t_k)$, for $1 \leq k \leq K$, with convention $t_0 = 0^+$. When $x_{\underline{F}}$ increases on an interval $(-1/t_{k-1}, -1/t_k)$, it must have its inflexion point in a point of positive derivative (from the concavity change induced by the asymptotes). Therefore, to verify that cluster k_F is disjoint from clusters $(k-1)_F$ and $(k+1)_F$ (when they exist), it suffices to verify that the $(k-1)$th and kth roots $\underline{m}_{k-1}$ and $\underline{m}_k$ of $x''_{\underline{F}}(\underline{m})$ are such that $x'_{\underline{F}}(\underline{m}_{k-1}) > 0$ and $x'_{\underline{F}}(\underline{m}_k) > 0$. From this observation, we therefore have the following result.

Theorem 7.6 ([Couillet *et al.*, 2011c; Mestre, 2008b]). *Let $\mathbf{B}_N$ be defined as in Theorem 7.1, with $\mathbf{T}_N = \operatorname{diag}(\tau_1, \ldots, \tau_N) \in \mathbb{R}^{N\times N}$, diagonal containing K distinct eigenvalues $0 < t_1 < \ldots < t_K$, for some fixed K. Denote N_k the multiplicity of the* kth *largest distinct eigenvalue (assuming ordering of the τ_i, we may then have $\tau_1 = \ldots = \tau_{N_1} = t_1, \ldots, \tau_{N-N_K+1} = \ldots = \tau_N = t_K$). Assume also that, for all $1 \leq r \leq K$, $N_r/n \to c_r > 0$, and $N/n \to c$, with $0 < c < \infty$. Then the cluster k_F associated with the eigenvalue t_k in the l.s.d. F of $\mathbf{B}_N$ is distinct from the clusters $(k-1)_F$ and $(k+1)_F$ (when they exist), associated with t_{k-1} and t_{k+1} in F, respectively, if and only if*

$$\sum_{r=1}^{K} c_r \frac{t_r^2 \underline{m}_k^2}{(1+t_r\underline{m}_k^2)^2} < 1$$
$$\sum_{r=1}^{K} c_r \frac{t_r^2 \underline{m}_{k+1}^2}{(1+t_r\underline{m}_{k+1}^2)^2} < 1 \tag{7.6}$$

where $\underline{m}_1, \ldots, \underline{m}_K$ are such that $\underline{m}_{K+1} = 0$ and $\underline{m}_1 < \underline{m}_2 < \ldots < \underline{m}_K$ are the K solutions of the equation in $\underline{m}$

$$\sum_{r=1}^{K} c_r \frac{t_r^3 \underline{m}^3}{(1+t_r\underline{m})^3} = 1.$$

For $k=1$, this condition ensures $1_F = 2_F - 1$. For $k = K$, this ensures $K_F = (K-1)_F + 1$. For $1 < k < K$, this ensures $(k-1)_F + 1 = k_F = (k+1)_F - 1$.

Remark now that the conditions of Equation (7.6) are left unchanged if all $t_1, \ldots, t_K$ are scaled by a common constant. Indeed, if t_j becomes αt_j for all j, then $\underline{m}_1, \ldots, \underline{m}_K$ become $\underline{m}_1/\alpha, \ldots, \underline{m}_K/\alpha$ and the scaling effects cancel out in Equation (7.6). Therefore, in the case $K = 2$, the separability condition only depends on the ratios c_1, c_2 and on t_1/t_2. If $c_1 = c_2 = c/2$, then we can depict the plot of the critical ratio $1/c$ as a function of t_1/t_2 for which cluster separability happens. This is depicted in Figure 7.3. Since $1/c$ is the limit of the ratio n/N, Figure 7.3 determines, for a fixed observation size N, the limiting number of samples per observation size required to achieve cluster separability. Observe how steeply the plot of $1/c$ increases when t_1 gets close to t_2; this suggests that the tools to be presented later that require this cluster separability will be very inefficient when it comes to separate close sources (the definition of 'closeness' depending on each specific study, e.g. close directions of signal arrivals in radar applications, close transmit powers in signal sensing, etc.). Figure 7.4 depicts the regions of separability of all clusters in the case $K = 3$, for fixed $c = 0.1$, $c_1 = c_2 = c_3$, as a function of the ratios t_3/t_1 and t_2/t_1. Observe that the triplets $(1, 3, 7)$ and $(1, 3, 10)$ are well inside the separability region as suggested, respectively, by Figure 7.1 (top) and Figure 7.2 (top); on the contrary, notice that

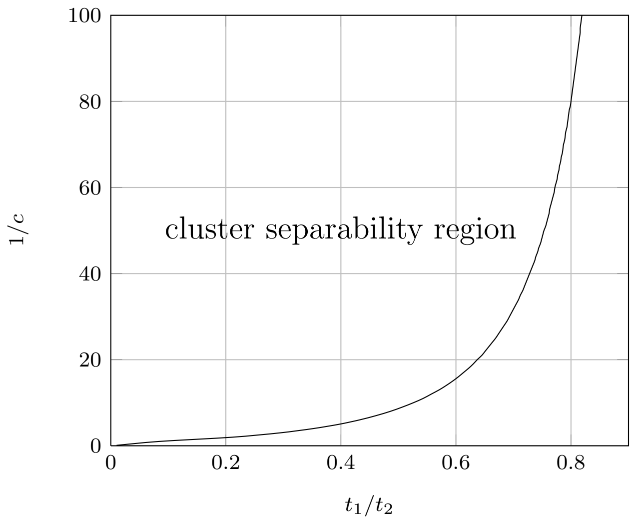

Figure 7.3 Limiting ratio c to ensure separability of (t_1, t_2), $t_1 \leq t_2$, $K = 2$, $c_1 = c_2$.

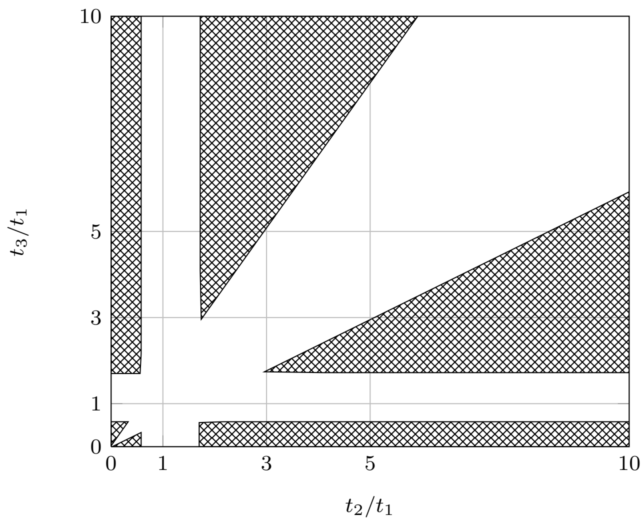

Figure 7.4 Subset of (t_1, t_2, t_3) that satisfy cluster separability condition, $c_1 = c_2 = c_3$, $c = 0.1$, in crosshatched pattern.

the triplets $(1, 3, 4)$ and $(1, 3, 5)$ are outside the separability region, confirming then the observations of Figure 7.1 (bottom) and Figure 7.2 (bottom).

After establishing these primary results for the sample covariance matrix models, we now move to the information plus noise model. According to the previous remark borrowed from Marčenko and Pastur in [Marčenko and Pastur, 1967], we infer that it will still be the case that the Stieltjes transform $m_F(x)$, extended to the real axis, has a local inverse $x_F(m)$, which is continuous and increasing, and that the range where $x_F(m)$ increases is exactly the complementary to the support of F. This statement will be shown to be somewhat correct. The main difference with the sample covariance matrix model is that there does not exist an explicit inverse $x_F(m)$, as in (7.5) and therefore $m_F(x)$ may have various inverses $x_F(m)$ for different subsets in the complementary of the support of F.

7.2 Information plus noise model

The asymptotic absence of eigenvalues outside the support of unconstrained information plus noise matrices (when the e.s.d. of the information matrix converges), i.e. with i.i.d. noise matrix components, is still at the stage of conjecture. While promising developments are being currently carried out, there exists to this day no proof of this fact, let alone a proof of the exact separation of information plus noise clusters. Nonetheless, in the particular case where the noise matrix is Gaussian, the two results have been recently proved [Vallet *et al.*, 2010]. Those results are given hereafter.

7.2.1 Exact separation

We recall that an information plus noise matrix $\mathbf{B}_N$ is defined by

$$\mathbf{B}_N = \frac{1}{n}(\mathbf{A}_N + \sigma\mathbf{X}_N)(\mathbf{A}_N + \sigma\mathbf{X}_N)^{\mathsf{H}} \tag{7.7}$$

where $\mathbf{A}_N$ is deterministic, representing the deterministic signal, $\mathbf{X}_N$ is random and represents the noise matrix, and $\sigma > 0$.

We start by introducing the theorem which states that, for all large N, no eigenvalue is found outside the asymptotic spectrum of the information plus noise model.

Theorem 7.7. *Let $\mathbf{B}_N$ be defined as in (7.7), with $\mathbf{A}_N \in \mathbb{C}^{N\times n}$ such that $H_N \triangleq F^{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}} \Rightarrow H$ and $\sup_N \|\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}\| < \infty$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ with entries $X_{N,ij}$ independent for all i, j, N, Gaussian with zero mean and unit variance. Further denote $c_N = N/n$ and assume $c_N \to c$, positive and finite. From Theorem 3.15, we know that $F^{\mathbf{B}_N}$ converges almost surely to a limit distribution F with Stieltjes transform $m_F(z)$ solution of the equation in m*

$$\frac{m}{1+\sigma^2 c_N m} = m_H\left(z(1+\sigma^2 cm)^2 - \sigma^2(1-c)(1+\sigma^2 cm)\right) \tag{7.8}$$

this solution being unique for $z \in \mathbb{C}^+$, $m \in \mathbb{C}^+$ and $\Im[zm] \geq 0$. Denote now $m_N(z)$ this solution when m_H is replaced by m_{H_N} and c by c_N, and denote F_N the distribution function with Stieltjes transform $m_N(z)$.

Let $N_0 \in \mathbb{N}$, and choose an interval $[a, b]$ outside the union of the supports of F and F_N for all $N \geq N_0$. For $\omega \in \Omega$, the probability space generating the *sequences* $\mathbf{X}_1, \mathbf{X}_2, \ldots$*, denote* $\mathcal{L}_N(\omega)$ *the set of eigenvalues of* $\mathbf{B}_N(\omega)$*. Then*

$$P(\omega, \mathcal{L}_N(\omega) \cap [a,b] \neq \emptyset \text{ i.o.}) = 0.$$

In fact, an even stronger result is proved in [Vallet, 2011]. It is precisely shown there that the probability that there exists an eigenvalue of $\mathbf{B}_N$ in $[a,b]$ if of order $O(N^{-l})$ for every integer l. The next theorem ensures that the repartition of the eigenvalues in the consecutive clusters is exactly as expected.

Theorem 7.8 ([Vallet *et al.*, 2010])**.** *Let* $\mathbf{B}_N$ *be as in Theorem 7.7. Let* $a < b$ *be such that* $[a,b]$ *lies outside the support of* F*. Denote* λ_k *and* a_k *the* kth *eigenvalues smallest of* $\mathbf{B}_N$ *and* $\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}$*, respectively. Also denote*

$$w_N(z) = z(1 + \sigma^2 c_N m_N(z))^2 - \sigma^2(1 - c_N)(1 + \sigma^2 c_N m_N(z))$$

with m_N *and* c_N *as in Theorem 7.7. Then, with probability one, for all large* N*,*

$$\begin{aligned}\#\left\{k,\ \lambda_k < a\right\} &= \#\left\{k,\ a_k < w_N(a)\right\} \\ \#\left\{k,\ \lambda_k > b\right\} &= \#\left\{k,\ a_k > w_N(b)\right\}.\end{aligned}$$

Similarly to Theorem 7.2, the above result merely says that each cluster of eigenvalues of $\mathbf{B}_N$ possesses as many eigenvalues as in the clusters of eigenvalues of $\frac{1}{n}\mathbf{A}_N A_N^{\mathsf{H}}$ which generated them (up to cluster overlaps as shown in Figure 7.5).

We provide hereafter a sketch of the proofs of both Theorem 7.7 and Theorem 7.8 where considerations of complex integration play a fundamental role. In the following chapter, Chapter 8, we introduce in detail the methods of complex integration for random matrix theory and particularly for statistical inference. The stronger proof of [Vallet, 2011] is slightly different and uses more fundamentally the Gaussian tools that are the integration by part formula and the Nash–Poincaré inequality.

Proof of Theorem 7.7 and Theorem 7.8. As already mentioned, these results are only known to hold for the Gaussian case for the time being. The way these results are achieved is similar to the way Theorem 7.1 and Theorem 7.2 were obtained, although the techniques are radically different. Indeed, somewhat similarly to Theorem 7.1, the first objective is to show that the difference $m_N(z) - \mathrm{E}[m_{\mathbf{B}_N}(z)]$ between the deterministic equivalent $m_N(z)$ of the empirical Stieltjes transform $m_{\mathbf{B}_N}(z)$ and $\mathrm{E}[m_{\mathbf{B}_N}(z)]$ goes to zero at a sufficiently fast rate. In the Gaussian case, this rate is of order $O(1/N^2)$. Remember from Theorem 6.5 that such a convergence rate was already observed for doubly correlated Gaussian models and allowed us to ensure that $N(m_N(z) - \mathrm{E}[m_{\mathbf{B}_N}(z)]) \to 0$. Using the fact, established precisely in Chapter 8, that, for holomorphic functions f and a distribution function G

$$\int f(x)dG(x) = -\frac{1}{2\pi i}\oint f(z)m_G(z)dz$$

on a positively oriented contour encircling the support of F, we can infer the recent result from [Haagerup *et al.*, 2006]

$$\mathrm{E}\left[\int f(x)[F^{\mathbf{B}_N} - F_N](dx)\right] = O\left(\frac{1}{N^2}\right).$$

Take f any infinitely differentiable function that is identically one on $[a,b] \subset \mathbb{R}$ and identically zero outside $(a-\varepsilon, b+\varepsilon)$ for some small positive ε, such that $(a-\varepsilon, b+\varepsilon)$ is outside the support of F. From the convergence rate above, we first have.

$$\mathrm{E}\left[\sum_{k=1}^{N} \lambda_k 1_{(a-\varepsilon,b+\varepsilon)}(\lambda_k)\right] = N(F_N(b) - F_N(a)) + O\left(\frac{1}{N}\right)$$

and therefore, for large N, we have in expectation the correct mass of eigenvalues in $(a-\varepsilon, b+\varepsilon)$. But we obviously want more than that: i.e., we want to determine the asymptotic exact number of these eigenvalues. Using the Nash–Poincaré inequality, Theorem 6.7, we can in fact show that, for this choice of f

$$\mathrm{E}\left[\left(\int f(x)[F^{\mathbf{B}_N} - F_N](dx)\right)^2\right] = O\left(\frac{1}{N^4}\right).$$

This is enough to prove, thanks to the Markov inequality, Theorem 3.5, that

$$P\left(\left|\int f(x)[F^{\mathbf{B}_N} - F_N](dx)\right| > \frac{1}{N^{\frac{4}{3}}}\right) < \frac{K}{N^{\frac{4}{3}}}$$

for some constant K. From there, the Borel–Cantelli lemma, Theorem 3.6, ensures that the above event is infinitely often true with probability zero; i.e. the event

$$\left|\sum_{k=1}^{N} \lambda_k 1_{(a-\varepsilon,b+\varepsilon)}(\lambda_k) - N(F_N(b) - F_N(a))\right| > \frac{K}{N^{\frac{1}{3}}}$$

is infinitely often true with probability zero. Therefore, with probability one, there exists N_0 such that, for $N > N_0$ there is no eigenvalue in $(a-\varepsilon, b+\varepsilon)$. This proves the first result.

Take now $[a,b]$ not necessarily outside the support of F and ε such that $(a-\varepsilon, a) \cup (b, b+\varepsilon)$ is outside the support of F. Then, repeating the same procedure as above but to characterize now

$$\left|\sum_{k=1}^{N} \lambda_k 1_{[a,b]}(\lambda_k) - N(F_N(b) - F_N(a))\right|$$

we find that this term equals

$$\left|\sum_{k=1}^{N} \lambda_k 1_{(a-\varepsilon,b+\varepsilon)}(\lambda_k) - N(F_N(b) - F_N(a))\right|$$

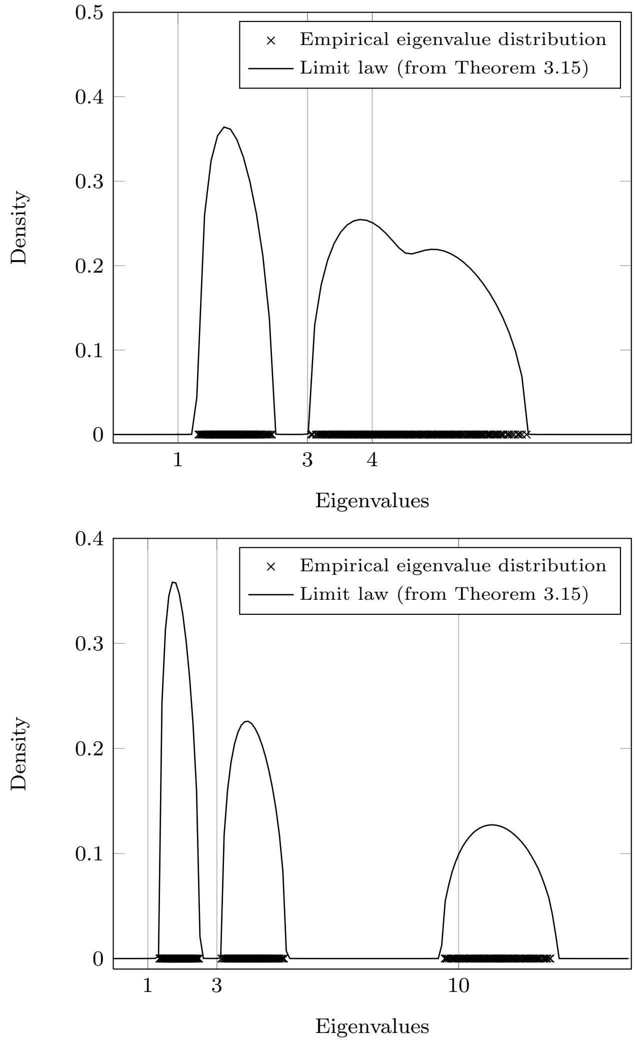

Figure 7.5 Empirical and limit eigenvalue distribution of the information plus noise model $\mathbf{B}_N = \frac{1}{n}(\mathbf{A}_N + \sigma\mathbf{X}_N)(\mathbf{A}_N + \sigma\mathbf{X}_N)^{\mathsf{H}}$, $N = 300$, $n = 3000$ ($c = 1/10$), $F^{\frac{1}{N}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}}$ has three evenly weighted masses at $1, 3, 4$ (top) and $1, 3, 10$ (bottom).

almost surely in the large N limit since there is asymptotically no eigenvalue in $(a - \varepsilon, a) \cup (b, b + \varepsilon)$. This now says that the asymptotic number of eigenvalues in $[a, b]$ is $N(F_N(b) - F_N(a))$ almost surely. The fact that the indexes of these eigenvalues are those expected is obvious. If it were not the case, then we can always find an interval on the left or on the right of $[a, b]$ which does not contain the right amount of eigenvalues, which is contradictory from this proof. This completes the proof of both results. $\square$

7.2.2 Asymptotic spectrum analysis

A similar spectrum analysis as in the case of sample covariance matrices when the population covariance matrix has a finite number of distinct eigenvalues can be performed for the information plus noise model. As discussed previously, the extension of $m_F(z)$ to the real positive half-line is locally invertible and increasing when outside the support of F. The semi-converse is again true: if $x_F(m)$ is an inverse function for $m_F(x)$ continuous with positive derivative, then its image is outside the support of F. However here, $x_F(m)$ is not necessarily unique, as will be confirmed by simulations. Let us first state the main result.

Theorem 7.9 ([Dozier and Silverstein, 2007b])**.** *Let $\mathbf{B}_N = \frac{1}{n}(\mathbf{A}_N + \sigma\mathbf{X}_N)(\mathbf{A}_N + \sigma\mathbf{X}_N)^{\mathsf{H}}$, with $\mathbf{A}_N \in \mathbb{C}^{N\times n}$ such that $H_N \triangleq F^{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}} \Rightarrow H$ and $\sup_N \|\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}\| < \infty$, $\mathbf{X}_N = (X_{N,ij}) \in \mathbb{C}^{N\times n}$ with $X_{N,ij}$ independent for all i, j, N with zero mean and unit variance (we release here the non-necessary Gaussian hypothesis). Denote S_F and S_H the supports of F and H, respectively. Take $(h_1, h_2) \subset S_H^{\mathsf{c}}$. Then there is a unique interval $(m_{F,1}, m_{F,2}) \subset (-\frac{1}{\sigma^2 c}, \infty)$ such that the function*

$$m \mapsto \frac{m}{1+\sigma^2 cm}$$

maps $(m_{F,1}, m_{F,2})$ to $(m_{H,1}, m_{H,2}) \subset (-\infty, \frac{1}{\sigma^2 c})$, where we introduced $(m_{H,1}, m_{H,2}) = m_H((h_1, h_2))$. On (h_1, h_2), m_H is invertible, and then we can define

$$x_F(m) = \frac{1}{b^2} m_H^{-1}\left(\frac{1}{\sigma^2 c}\left(1 - \frac{1}{b}\right)\right) + \frac{1}{b}\sigma^2(1-c)$$

with $b = 1 + \sigma^2 cm$.

Then:

1. *if for $m \in (m_{F,1}, m_{F,2})$, $x(m) \in S_F^{\mathsf{c}}$, then $x'(m) > 0$;*
2. *if $x_F'(m) > 0$ for $b \in (m_{F,1}, m_{F,2})$, then $x_F(m) \in S_F^{\mathsf{c}}$ and $m = m_F(x_F(m))$.*

Similar to the sample covariance matrix case, Theorem 7.9 gives readily a way to determine the support of F: for m varying in $(m_{F,1}, m_{F,2})$, whenever $x_F(m)$ increases, its image is outside the support of F. The support of F is therefore the complementary set to the union of all such intervals. We must nonetheless be aware that the definition of $x_F(m)$ is actually linked to the choice of the interval $(h_1, h_2) \subset S_H^{\mathsf{c}}$. In Theorem 7.4, we had a unique explicit inverse for $x_F(m)$ as a function of m, whatever the choice of the pre-image of m_H (the Stieltjes transform of the l.s.d. of the population covariance matrix); this statement no longer holds here.

In fact, if S_H is subdivided into K_H clusters, we can expect at most $K_H + 1$ different local inverses for $x_F(m)$ as m varies along $\mathbb{R}$. This is in fact exactly what is observed. Figure 7.6 depicts the situation when H is composed of three

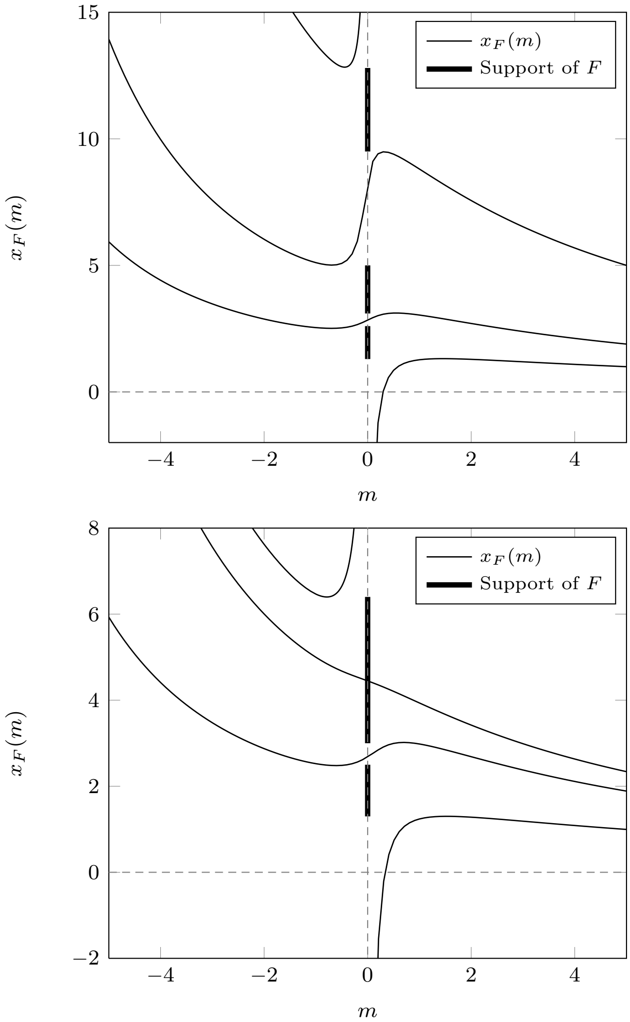

Figure 7.6 Information plus noise model, $x_F(m)$ for m real, $F^{\frac{1}{N}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}} \Rightarrow H$, where H has three evenly weighted masses in 1, 3, and 10 (top) and 1, 3, and 4 (bottom), $c = 1/10$, $\sigma = 0.1$ in both cases. The support of F can be read on the central vertical axises.

evenly weighted masses in $(1, 3, 4)$, then $(1, 3, 10)$. Observe that $K_H + 1$ different inverses exist that have the aforementioned behavior.

Now, also similar to the sample covariance matrix model, a lot more can be said in the case where H is composed of a finite number of masses. The exact determination of the boundary of F can be determined. The result is summarized as follows.

Theorem 7.10 ([Vallet *et al.*, 2010]). *Let* $\mathbf{B}_N$ *be defined as in Theorem 7.9, where* $F^{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}} = H$ *is composed of* K *eigenvalues* $h_1, \ldots, h_K$ *(we implicitly* *assume N takes only values consistent with $F^{\frac{1}{n}\mathbf{A}_N\mathbf{A}_N^{\mathsf{H}}} = H$). Let ϕ be the function on $\mathbb{R} \setminus \{h_1, \ldots, h_K\}$ defined by*

$$\phi(w) = w(1 - \sigma^2 c m_H(w))^2 + (1-c)\sigma^2(1 - \sigma^2 c m_H(w)).$$

Then $\phi(w)$ has $2K_F$, $K_F \leq K$, local maxima, such that $1 - \sigma^2 c m_H(w) > 0$ and $\phi(w) > 0$. We denote these maxima $w_1^-, w_1^+, w_2^-, w_2^+, \ldots, w_{K_F}^-, w_{K_F}^+$ in the order

$$w_1^- < 0 < w_1^+ \leq w_2^- < w_2^+ \leq \ldots \leq w_{K_F}^- < w_{K_F}^+.$$

Furthermore, denoting $x_k^- = \phi(w_k^-)$ and $x_k^+ = \phi(w_k^+)$, we have:

$$0 < x_1^- < x_1^+ \leq x_2^- < x_2^+ \leq \ldots \leq x_{K_F}^- < x_{K_F}^+.$$

The support S_F of F is the union of the compact sets $[x_k^-, x_k^+]$, $k \in \{1, \ldots, K_F\}$

$$S_F = \bigcup_{k=1}^{K_F} [x_k^-, x_k^+].$$

Note that this alternative approach, via the function $\phi(w)$, allows us to give a deterministic expression of the subsets $[x_k^-, x_k^+]$ without the need to explicitly invert m_H in $K+1$ different inverses, which is more convenient.

A cluster separability condition can also be established, based on the results of Theorem 7.10. Namely, we say that the cluster in F corresponding to the eigenvalue h_k is disjoint from the neighboring clusters if there exists $k_F \in \{1, \ldots, K_F\}$ such that

$$h_{k-1} < w_{k_F}^- < h_k < w_{k_F}^+ < h_{k+1}$$

with convention $h_0 = 0$, $h_{K+1} = \infty$, and we say that k_F is the cluster associated with h_k in F.

This concludes this chapter on spectral analysis of the sample covariance matrix and the information plus noise models. As mentioned in the Introduction of this chapter, these results will be applied to solve eigen-inference problems, i.e. inverse problems concerning the eigenvalue or eigenvector structure of the underlying matrix models. We will then move to the last chapter, Chapter 9, of the theoretical part, which is concerned with limiting results on the extreme eigenvalues for both the sample covariance matrix and information plus noise models. These results will push further the theorems of exact separation by establishing the limiting distributions of the extreme eigenvalues (although solely in the Gaussian case) and also some properties on the corresponding eigenvectors.