9 Extreme eigenvalues and eigenspace projections

This last chapter of Part I introduces very recent mathematical advances of deep interest to the field of wireless communications, related to the limiting behavior of the extreme eigenvalues and of their corresponding eigenvectors. Again, the main objects which have been extensively studied in this respect are derivatives of the sample covariance matrix and of the information plus noise matrix.

This chapter will be divided into two sections, whose results emerge from two very different random matrix approaches. The first results, about the limiting extreme eigenvalues of the spiked models, unfold from the previous exact separation results described in Chapter 7. It will in particular be proved that in a sample covariance matrix model, when all population eigenvalues are equal but for the few largest ones, the l.s.d. of the sample covariance matrix is still the Marčenko–Pastur law, but a few eigenvalues may now be found outside the support of the l.s.d. The second set of results concerns mostly random matrix models with Gaussian entries, for which limiting results on the second-order behavior of extreme eigenvalues are available. These results use very different approaches than those proposed so far, in particular the theory of orthogonal polynomials and determinantal representations. This subject, which requires many additional tools, is briefly introduced in this chapter. For more information about these tools, see, e.g. the tutorial [Johnstone, 2006] or the book [Mehta, 2004].

We start this section with the spiked models.

9.1 Spiked models

9.1.1 First order statistics of spiked models

As an introduction to the spiked models, let us consider the now standard covariance matrix model $\mathbf{B}_N = \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$ with $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ random with i.i.d. entries of zero mean, variance $1/n$, and finite fourth order moment. From Theorem 2.13 and Theorem 7.1, we know that, as $N, n \to \infty$ with $N/n \to c > 0$, the e.s.d. $F^{\mathbf{B}_N}$ of $\mathbf{B}_N$ converges weakly and almost surely to the Marčenko–Pastur law with d.f. F, i.e. $F^{\mathbf{B}_N} \Rightarrow F$, *and* that, for all large N, no eigenvalue of $\mathbf{B}_N$ can be found outside the support of F.

We also know, from Theorem 3.13, that, if $\mathbf{B}_N$ is now $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, and $F^{\mathbf{T}_N}(x) \Rightarrow 1_{\{x \leq 1\}}$, then $F^{\mathbf{B}_N} \Rightarrow F$ almost surely, as well. However, as mentioned in the discussion following Theorem 7.1, it is not necessarily true that no eigenvalue of $\mathbf{B}_N$ is found outside the support of F for all large N. Intuitively, if $\mathbf{T}_N = \mathrm{diag}(1, \ldots, 1, 100)$, with $N-1$ ones and a single value at hundred, it is expected that the largest eigenvalue of $\mathbf{B}_N$ does not remain inside F for all large N, especially so for N/n small.

In the preceding example, $\mathbf{B}_N$ follows a *spiked model*, in the sense that it appears as a small perturbation of a standard matrix model, small meaning "low rank" here. Obviously, this spiked model is nothing but an instance of the sample covariance matrix model of Theorem 3.13. However, the fact that $\mathbf{B}_N$ is a low rank perturbation of $\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$ leads to extremely interesting simplifications of the study of such model, starting with the fact that the limiting positions of the extreme eigenvalues, as well as their fluctuations around these positions, can be characterized in *closed form*.

The importance of spiked models in wireless communications and signal processing arises when performing signal sensing or statistical inference, when the (hypothetical) signal space has small dimension compared to the noise space. In this case, we might want to be able to decide on the presence of a signal based on the spectrum of the sample covariance matrix $\mathbf{B}_N$. Typically, if an eigenvalue is found outside the limiting predicted noise spectrum F of $\mathbf{B}_N$ (i.e. outside the support of the Marčenko–Pastur law above), then this must indicate the presence of a signal bearing informative data. If, on the opposite, all the eigenvalues are inside the limiting support F, then this should indicate the absence of such a signal. The story is however not so simple and, while the former statement is correct, the latter is not. Indeed, as we will see in the following, it might not always be true that a *spike* (i.e. a large eigenvalue) in $\mathbf{T}_N$ results in a spike in $\mathbf{B}_N$ found outside the support of F, in the sense that the support of F may "hide" the spike in some sense. This is especially true when the size of the main clusters of eigenvalues (linked to the ratio N/n) is large enough to "absorb" the spike of $\mathbf{B}_N$ that would have resulted from the population spike of $\mathbf{T}_N$. In this case, for signal detection purposes, whether a signal bearing informative data is present or not, there is no way to decide on the presence of this signal by simply looking at the asymptotic spectrum. The condition for decidability is given in the following fundamental initial result.

Theorem 9.1 ([Baik and Silverstein, 2006])**.** *Let* $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$*, where* $\mathbf{X}_N \in \mathbb{C}^{N \times n}$ *has i.i.d. entries of zero mean, variance* $1/n$ *and order moment of order* $O(1/n^2)$*, and* $\mathbf{T}_N \in \mathbb{R}^{N \times N}$ *is diagonal given by:*

$$\mathbf{T}_N = \mathrm{diag}(\underbrace{1, \ldots, 1}_{N-r}, \underbrace{1+\omega_1, \ldots, 1+\omega_1}_{j_1}, \ldots, \underbrace{1+\omega_r, \ldots, 1+\omega_t}_{j_t})$$

with $\omega_1 > \ldots > \omega_t > -1$ *for some* t *fixed and* $j_1 + \ldots + j_t = r$*.*

Take $c > 0$ and denote $p = \#\{i, \omega_i > \sqrt{c}\}$ and $q = \#\{i, \omega_i < -\sqrt{c}\}$. Denote additionally $\lambda_1 \geq \ldots \geq \lambda_N$ the ordered eigenvalues of $\mathbf{B}_N$. We then have, as $N, n \to \infty$, with $N/n \to c$:

- *if $\omega_k > 0$, for $1 \leq i \leq j_k$*

$$\lambda_{j_1+\ldots+j_{k-1}+i} \xrightarrow{\text{a.s.}} \begin{cases} \rho_k & , \ 1 \leq k \leq p \\ (1+\sqrt{c})^2 & , \ k > p \end{cases}$$

- *if $\omega_k < 0$ and $c < 1$, for $1 \leq i \leq j_k$*

$$\lambda_{N-j_t-\ldots-j_k+i} \xrightarrow{\text{a.s.}} \begin{cases} \rho_k & , \ t-q+1 \leq k \leq t \\ (1-\sqrt{c})^2 & , \ k < t-q+1. \end{cases}$$

where we defined $\rho_k = 1 + \omega_k + c\omega_k^{-1}(1+\omega_k)$.

Note that the function $\rho : \omega \mapsto 1 + \omega + c\omega^{-1}(1+\omega)$ is increasing on $(-\infty, \sqrt{c})$ and $(\sqrt{c}, \infty)$, and equals $(1+\sqrt{c})^2$ for $\omega = \sqrt{c}$. Therefore, the theorem can be interpreted as follows: if $\mathbf{T}_N$ has a diagonal entry $\omega > \sqrt{c}$, then $\mathbf{B}_N$ will asymptotically have a corresponding eigenvalue ρ away from the support of the Marčenko–Pastur law, with ρ being larger for larger ω and ρ getting close to the right-edge of the support for ω close to $\sqrt{c}$ and eventually collapsing within the support for $0 < \omega \leq \sqrt{c}$. When $\omega < -\sqrt{c}$, it is fundamental that $c \leq 1$, otherwise the eigenvalue of $\mathbf{B}_N$ associated to ω will fall within the set of null eigenvalues of $\mathbf{B}_N$.

The value $\sqrt{c}$ therefore corresponds to a universal boundary above which it is possible to observe eigenvalues outside the support of the Marčenko–Pastur law. From a signal processing viewpoint, this boundary corresponds to a threshold above which the presence of a signal of power ω can be decided from the observation of $\mathbf{B}_N$ or not. This phenomenon is depicted in Figure 9.1 where we consider $t = 2$ population spikes $\omega_1 = 2$ and $\omega_2 = 1$, both of multiplicity $j_1 = j_2 = 2$, hence $r = 4$. We illustrate the decidability condition depending on c by considering first $c = 1/3$, in which case $\sqrt{c} \simeq 0.57 < \omega_2 < \omega_1$ and then we expect two spikes of $\mathbf{B}_N$ at position $\rho_2 \simeq 2.67$ and two spikes of $\mathbf{B}_N$ at position $\rho_1 = 3.5$. We then increase c to $c = 5/4$ for which $\omega_2 < \sqrt{c} \simeq 1.12 < \omega_1$; we therefore expect only the two eigenvalues associated with α_1 at position $\rho_1 \simeq 4.88$ to lie outside the spectrum of F. This is approximately what is observed.

The fact that spikes are non-discernible for large c leads to a seemingly paradoxical situation. Consider indeed that the sample space is fixed to n samples while the population space of dimension N increases, so that we increase the collection of input data to improve the quality of the experiment. In the context of signal sensing, if we rely only on a global analysis of the empirical eigenvalues of the input covariance matrix to declare that "if eigenvalues are found outside the support, a signal is detected," then it seems that we are better off limiting N to a minimal value and therefore we are better off with a mediocre quality of the experiment; otherwise the decidability threshold is severely impacted. The

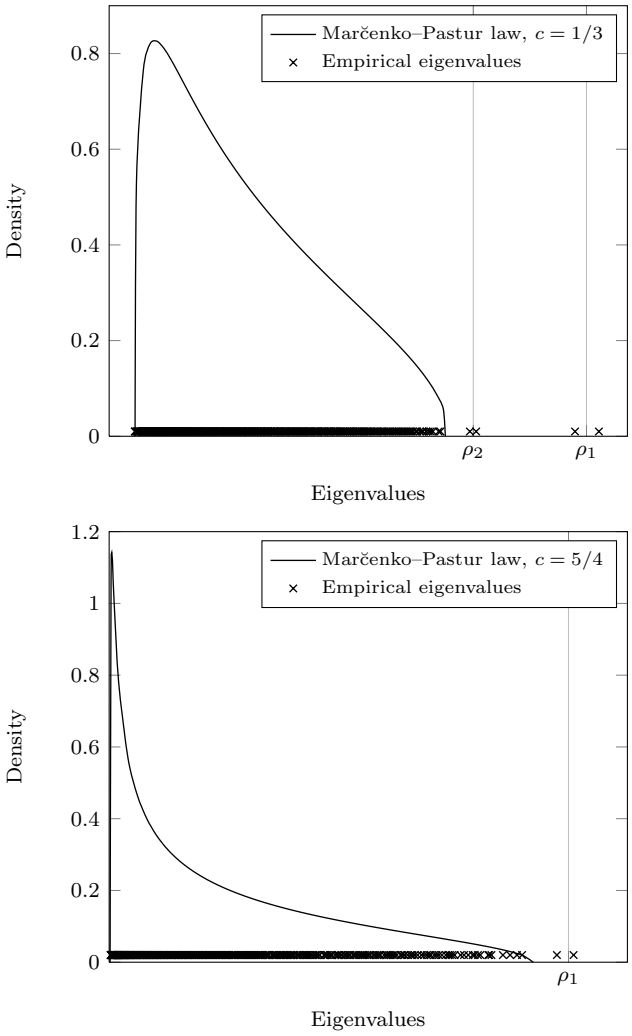

Figure 9.1 Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $\mathbf{T}_N$ is a diagonal of ones but for the first four entries set to $\{3, 3, 2, 2\}$ ($\omega_1 = 3$, $\omega_2 = 2$). On top, $N = 500$, $n = 1500$. On the bottom, $N = 500$, $n = 400$. Theoretical limit eigenvalues $\rho_k = 1 + \omega_k + c\omega_k^{-1}(1 + \omega_k)$ of $\mathbf{B}_N$ are stressed.

reasoning error however is that, in standard signal processing models, the larger N, the larger the signal power collected on the antenna array, and therefore the larger ω. These practical aspects are discussed in Chapter 16.

Another way of observing practically when the e.s.d. at hand is close to the Marčenko–Pastur law F is to plot the empirical eigenvalues against the quantiles $F^{-1}(\frac{k-1/2}{N})$ for $k = 1, \ldots, N$. This is depicted in Figure 9.2, for the case $c = 1/3$ with the same set of population spikes $\{2, 2, 3, 3\}$ in $\mathbf{T}_N$ as before. We observe again the presence of four outlying eigenvalues in the e.s.d. of $\mathbf{B}_N$.

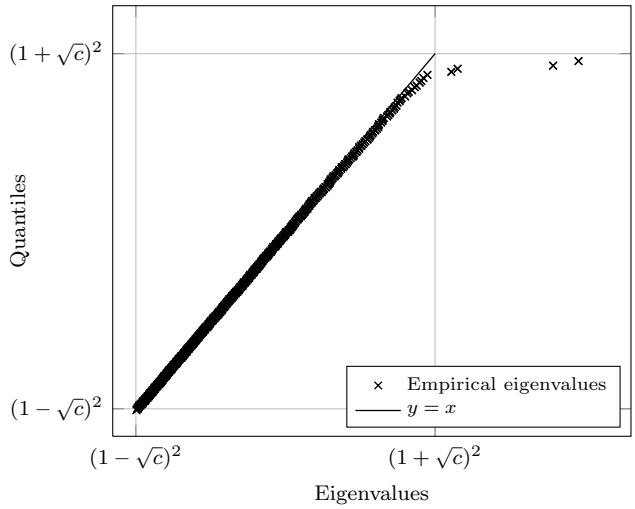

Figure 9.2 Eigenvalues of $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}} \mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} \mathbf{T}_N^{\frac{1}{2}}$, where $\mathbf{T}_N$ is a diagonal of ones but for the first four entries set to $\{3, 3, 2, 2\}$, against the quantiles of the Marčenko–Pastur law, $N = 500$, $n = 15\,000$, $c = 1/3$.

The original proof of Theorem 9.1 relies heavily on the tools used in [Bai and Silverstein, 1998, 1999] for the exact separation of the sample covariance matrix. Indeed, in [Baik and Silverstein, 2006], population spikes are seen as clusters of eigenvalues of supports sufficiently small to asymptotically admit only $j_1, \ldots, j_t$ eigenvalues, respectively. When $\mathbf{X}_N$ is chosen to be Gaussian, or at least left-unitarily invariant, different techniques can be used which greatly simplify the proof. In the remainder of this section, we focus on this scenario. We start with a short outlook of the proof of Theorem 9.1 in this context.

Proof of Theorem 9.1 with $\mathbf{X}_N$ Gaussian. Denote $\mathbf{\Omega} = \operatorname{diag}(\omega_1 \mathbf{I}_{j_1}, \ldots, \omega_t \mathbf{I}_{j_t})$. Take N sufficiently large so that $\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$ has no eigenvalue above $(1+\sqrt{c})^2 + \varepsilon$ for some $\varepsilon > 0$. Take $x > (1+\sqrt{c})^2 + \varepsilon$. If x is an eigenvalue of $\mathbf{B}_N$, then, by definition $\det(\mathbf{B}_N - z\mathbf{I}_N) = 0$. Developing this expression and using the fact that x is not an eigenvalue of $\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}}$, we have

$$\begin{aligned}
&\det(\mathbf{B}_N - z\mathbf{I}_N) \\
&= \det \mathbf{T}_N \det\left(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N + x\left[\mathbf{I}_N - \mathbf{T}_N^{-1}\right]\right) \\
&= \det \mathbf{T}_N \det\left(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N\right) \det\left(\mathbf{I}_N + x(\mathbf{T}_N - \mathbf{I}_N)\mathbf{T}_N^{-1}(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\right)
\end{aligned}$$

where we used $\mathbf{I}_N - \mathbf{T}_N^{-1} = (\mathbf{T}_N - \mathbf{I}_N)\mathbf{T}_N^{-1}$. Equating the above to zero implies that the third determinant equals zero. Using the fact that $\mathbf{T}_N - \mathbf{I}_N$ has rank r, there exists $\mathbf{U} \in \mathbb{C}^{N \times r}$ such that $\mathbf{T}_N - \mathbf{I}_N = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}$. From $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$, we therefore have

$$\begin{aligned}
&\det\left(\mathbf{I}_N + x(\mathbf{T}_N - \mathbf{I}_N)\mathbf{T}_N^{-1}(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\right) \\
&= \det\left(\mathbf{I}_r + x\mathbf{\Omega}\mathbf{U}^{\mathsf{H}}(\mathbf{I}_N + \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\mathsf{H}})^{-1}(\mathbf{X}_N \mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}\right)
\end{aligned}$$

$$= \det\left(\mathbf{I}_r + x\boldsymbol{\Omega}\left(\mathbf{I}_r + \boldsymbol{\Omega}\right)^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}\right)$$

where in the last equality, we used $\mathbf{U}^{\mathsf{H}}(\mathbf{I}_N + \mathbf{U}\boldsymbol{\Omega}\mathbf{U}^{\mathsf{H}})^{-1} = (\mathbf{I}_r + \boldsymbol{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}$. The interest of the above derivation lies in the isolation of the resolvent $(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}$ for which much is known. The remaining terms matrices are deterministic and therefore easily dealt with.

Since r is fixed and $\mathbf{X}_N$ is left-unitarily invariant, an immediate extension of the trace lemma, Lemma 3.4, ensures that $\mathbf{U}^{\mathsf{H}}(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U} \xrightarrow{\text{a.s.}} m(x)\mathbf{I}_r$ (each diagonal entry of the left-hand side is a quadratic form and each non-diagonal entry a bilinear form based on orthogonal vectors), where $m(z)$ is the Stieltjes transform of the l.s.d. of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$. As a consequence, an eigenvalue x of $\mathbf{B}_N$ outside the support, if any, satisfies asymptotically

$$xm(x)\omega_k(1+\omega_k)^{-1} \xrightarrow{\text{a.s.}} -1$$

for some k. Replacing $m(x)$ by the expression of the Stieltjes transform of the Marčenko–Pastur law (3.20) gives immediately the expected result $x \xrightarrow{\text{a.s.}} 1 + \omega_k + c\omega_k^{-1}(1+\omega_k)$, only if $\omega_k > \sqrt{c}$. Otherwise, the equation has no solution outside $[(1-\sqrt{c})^2, (1+\sqrt{c})^2]$. □

This proof approach is originally due to Benaych-Georges in [Benaych-Georges and Rao, 2011]. Since only in the last step of the proof is $m(z)$ particularized to the Stieltjes transform of the Marčenko-Pastur law, his result is more general than Theorem 9.1. A proper study of the existence outside the limiting support of solutions to the equation $xm(x)\omega_k(1+\omega_k)^{-1} + 1 = 0$ then gives precisely the following result.

Theorem 9.2 ([Benaych-Georges and Rao, 2011])**.** *Let $\mathbf{R}_N \in \mathbb{C}^{N\times N}$ be a Hermitian random matrix with ordered eigenvalues $\nu_1 \geq \ldots \geq \nu_N$ for which we assume that the e.s.d. $F^{\mathbf{R}_N}$ converges almost surely toward F with compact support with infimum a and supremum b, such that $\nu_1 \xrightarrow{\text{a.s.}} b$ and $\nu_N \xrightarrow{\text{a.s.}} a$. Consider also a perturbation matrix $\mathbf{P}_N$ of rank $r = j_1 + \ldots + j_t$, with ordered non-zero eigenvalues $\omega_1 > \ldots > \omega_t < -1$, ω_i having multiplicity j_i, these values being fixed for all N. Denote s the integer such that $\omega_s > 0 > \omega_{s+1}$. We further assume that either $\mathbf{R}_N$ is bi-unitarily invariant. Denote $\mathbf{B}_N$ the matrix*

$$\mathbf{B}_N = \left(\mathbf{I}_N + \mathbf{P}_N\right)^{\frac{1}{2}}\mathbf{R}_N\left(\mathbf{I}_N + \mathbf{P}_N\right)^{\frac{1}{2}}$$

with ordered eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N$. Then, as N grows large, for $1 \leq i \leq s$, $1 \leq l \leq j_i$

$$\lambda_{j_1+\ldots+j_{i-1}+l} \xrightarrow{\text{a.s.}} \begin{cases} \rho_i \ , & \text{if } b^+ m_F(b^+)\omega_i(1+\omega_i)^{-1} < -1 \\ b \ , & \text{otherwise} \end{cases}$$

with ρ_i the unique solution $x \in (b, \infty)$ to $x m_F(x)\omega_i(1+\omega_i)^{-1} = 1$, and, for $s+1 \leq i \leq t$,

$$\lambda_{n-j_t-\ldots-j_i+l} \xrightarrow{\text{a.s.}} \begin{cases} \rho_i \ , \text{ if } a^- m_F(a^-)\omega_i(1+\omega_i)^{-1} > -1 \\ a \ , \text{ otherwise} \end{cases}$$

with ρ_i the unique solution $x \in (-\infty, a)$ to $x m_F(x)\omega_i(1+\omega_i)^{-1} = 1$.

Another result following the same line of proof for a different perturbation model is also provided in [Benaych-Georges and Rao, 2011], as follows.

Theorem 9.3 ([Benaych-Georges and Rao, 2011])**.** *Let $\mathbf{R}_N, \mathbf{P}_N \in \mathbb{C}^{N\times N}$ be defined as in Theorem 9.2 and $\mathbf{B}_N$ the matrix defined as*

$$\mathbf{B}_N = \mathbf{R}_N + \mathbf{P}_N$$

with ordered eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N$. Then, as N grows large, for $1 \leq i \leq s$, $1 \leq l \leq j_i$,

$$\lambda_{j_1+\ldots+j_{i-1}+l} \xrightarrow{\text{a.s.}} \begin{cases} \rho_i \ , \text{ if } \omega_i m_F(-b^+) > 1 \\ b \ , \text{ otherwise} \end{cases}$$

with ρ_i the unique solution $x \in (b, \infty)$ to $\omega_i m_F(x) + 1 = 0$, and, for $s+1 \leq i \leq t$,

$$\lambda_{n-j_t-\ldots-j_i+l} \xrightarrow{\text{a.s.}} \begin{cases} \rho_i \ , \text{ if } \omega_i m_F(-a^-) < 1 \\ a \ , \text{ otherwise} \end{cases}$$

with ρ_i the unique solution $a \in (-\infty, a)$ to $\omega_i m_F(x) + 1 = 0$.

As mentioned above, Theorem 9.2 encompasses the case where $\mathbf{R}_N = \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$, with $\mathbf{X}_N$ filled with i.i.d. Gaussian entries, perturbed in the sense of Theorem 9.1. In this case, this result generalizes Theorem 9.1 for unitarily invariant matrices, although it does not encompass the general i.i.d. case. It also contains the case $\mathbf{R}_N = \mathbf{X}_N\mathbf{T}_N\mathbf{X}_N^{\mathsf{H}}$ under some conditions on the Hermitian matrix $\mathbf{T}_N \in \mathbb{C}^{n\times n}$. Since, in this scenario, m_F is known exactly from Theorem 3.13, the limiting position of the sample spikes unfold immediately.

When the unitary invariance of $\mathbf{X}_N$ is no longer ensured in the model, the proof technique from Benaych-Georges can still be used, but some more work on the characterization of the matrices $\mathbf{U}^{\mathsf{H}}(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}$ is required. This is performed in [Chapon *et al.*, 2012] for the information plus noise model with right-side correlation of the noise matrix. The results are as follows.

Theorem 9.4 ([Chapon *et al.*, 2012])**.** *Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ be a random matrix with Gaussian i.i.d. entries of zero mean and variance $1/n$, $\mathbf{T}_N = \operatorname{diag}(t_1^N, \ldots, t_n^N)$ such that $F^{\mathbf{T}_N} \Rightarrow H$ with compact support $\mathcal{S}_H$ and that the maximum distance from the t_i^N to $\mathcal{S}_H$ tends to zero. Let also $\mathbf{P}_N = \mathbf{U}_N\mathbf{\Omega}_N^{\frac{1}{2}}\mathbf{V}_N \in \mathbb{C}^{N\times n}$ of rank r fixed, with $\mathbf{U}_N \in \mathbb{C}^{N\times r}$, $\mathbf{V}_N \in \mathbb{C}^{r\times n}$ isometric and $\mathbf{\Omega}_N$ diagonal such that*

$\mathbf{\Omega}_N \to \text{diag}(\omega_1 \mathbf{I}_{j_1}, \ldots, \omega_t \mathbf{I}_{j_t})$, $\omega_i > 0$. *Further assume that* $\mathbf{V}_N^{\mathsf{H}} f(\mathbf{T}_N)\mathbf{V}_N \to \int f(t)dH(t)\mathbf{I}_r$ *for every continuous bounded* f.[1] *Define* $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ *as*

$$\mathbf{B}_N = (\mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}} + \mathbf{P}_N)(\mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}} + \mathbf{P}_N)^{\mathsf{H}}.$$

Then, as $N, n \to \infty$, $N/n \to c$, $F^{\mathbf{B}_N} \Rightarrow F$ *where* F *has Stieltjes transform* $m_F(z)$ *given by the analytic solution such that* $m_F(z) \in \mathbb{C}^+$ *when* $z \in \mathbb{C}^+$ *of the equation in* m

$$m = \left(-z + \int \frac{t}{1 + cmt} dH(t) \right)^{-1}.$$

Moreover, for (a, b) *outside the support* $\mathcal{S}_F$ *of* F *with* $a \geq 0$, *there exists* $0 \leq k \leq r$ *eigenvalues of* $\mathbf{B}_N$ *in* (a, b) *for all large* N, *almost surely, with limits* $\rho_1^{[a,b]}, \ldots, \rho_k^{[a,b]}$ *given by the* k *solutions in* $[a, b]$ *of the* t *equations in* $\rho > 0$

$$\omega_i m_F(\rho)(c m_F(\rho) - 1 + c) = 1$$

for $i \in \{1, \ldots, t\}$. *Moreover, there is no eigenvalue of* $\mathbf{B}_N$ *in* $(a, b) \subset [0, \inf(\mathcal{S}_F)]$, *for all large* N, *almost surely.*

As an immediate corollary, we have the following result, initially proved in [Loubaton and Vallet, 2010].

Corollary 9.1 ([Loubaton and Vallet, 2010])**.** *Let* $\mathbf{B}_N = (\mathbf{X}_N + \mathbf{P}_N)(\mathbf{X}_N + \mathbf{P}_N)^{\mathsf{H}}$ *with eigenvalues* $\lambda_1 \geq \ldots \lambda_N$ *and with* $\mathbf{X}_N$ *and* $\mathbf{P}_N$ *as in Theorem 9.4. Then, as* $N, n \to \infty$, $N/n \to c$, *if* $\omega_i > \sqrt{c}$, *for* $1 \leq l \leq j_i$

$$\lambda_{j_1 + \ldots + j_{i-1} + l} \xrightarrow{\text{a.s.}} \omega_i^{-1}(c + \omega_i)(1 + \omega_i).$$

Note that, although in [Chapon *et al.*, 2012], a more general form of the theorem above is provided, the result relies on seemingly stringent underlying assumptions. The fact that $\mathbf{T}_N$ is diagonal is not a problem as any unitary product on the right of $\mathbf{X}_N \mathbf{T}_N^{\frac{1}{2}} + \mathbf{P}_N$, compensated by a product on the right of $\mathbf{X}_N$ can always turn $\mathbf{T}_N$ into a diagonal matrix while not changing the statistics of the model. However, the eigenvalues of $\mathbf{T}_N$ are imposed, similar to Theorem 9.2 and Theorem 9.3, not to escape the l.s.d. of $\mathbf{T}_N$. This is merely to avoid the generation of spikes by $\mathbf{T}_N$ itself. In practical applications, if $\mathbf{T}_N$ models the time correlation between successive stationary noise samples gathered in the columns of $\mathbf{X}_N \mathbf{T}_N$, then it is usual to model $\mathbf{T}_N$ as a Toeplitz matrix (e.g. in an autoregressive moving average assumption) with limiting spectrum given by a compact connected power spectrum density. Now, the condition $\mathbf{V}_N^{\mathsf{H}} f(\mathbf{T}_N)\mathbf{V}_N \to \int f(t)dH(t)\mathbf{I}_r$ may seem arbitrary as $\mathbf{T}_N$ and $\mathbf{V}_N$ are (sequences of) deterministic matrices which have, as such, no reason to have

[1] We denote $f(\mathbf{T})$ for a diagonal $\mathbf{T}$ matrix the diagonal matrix containing the $f(t_i^N)$ in the same order.

any specific limiting behavior. For application purposes though, if $\mathbf{P}_N$ is taken to be random, but independent of $\mathbf{X}_N$ with $\mathbf{V}_N$ isometric (e.g. $\mathbf{P}_N = \mathbf{h}\mathbf{x}^{\mathsf{H}}$ is the rank-1 product of a deterministic channel vector $\mathbf{h} \in \mathbb{C}^N$ and random i.i.d. white Gaussian signals $(x_1, \ldots, x_n) = \mathbf{x}^{\mathsf{H}}$), then the condition is met from the trace lemma, Theorem 3.4. Note importantly here that the trace lemma, usually applied to quadratic forms, is here applied to the matrix product $\mathbf{V}_N^{\mathsf{H}} f(\mathbf{T}_N)\mathbf{V}_N$ which is obviously valid as each entry of this *finite* rank matrix is a bilinear form.

Although not mentioned in the statement of Theorem 9.4, the total number of spiked eigenvalues may exceed r in this situation (see a precise discussion in [Chapon *et al.*, 2012]). In fact, if $\mathcal{S}_F$ is composed of several disjoint clusters, as many as r eigenvalues can be found in-between *each* of these clusters, with total number more than r. In [Chapon *et al.*, 2012], it is shown that one can always find a matrix $\mathbf{P}_N$ such that exactly r eigenvalues are found between two clusters. However, it is not shown if it is possible to find a matrix $\mathbf{P}_N$ having exactly r eigenvalues between each of these clusters. This triggers an open question raised in [Chapon *et al.*, 2012] regarding the numbering of these spiked eigenvalues. Indeed, while in the previous theorems, the spikes happened to be the extreme eigenvalues, so either the largest or the smallest, in the setting of Theorem 9.4, it is difficult to know "from which cluster the spikes jumped out", hence a problem in ordering them.

As noticed from the various discussions above, the study of extreme eigenvalues carries some importance in problems of detection of signals embedded in white or colored noise, but not only. Fields such as speech recognition, statistical learning, or finance also have interests in extreme eigenvalues of covariance matrices. For the particular case of finance, see, e.g., [Laloux *et al.*, 2000; Plerous *et al.*, 2002], consider $\mathbf{X}_N$ is the $N \times n$ matrix in which each row stands for a market product, while every column stands for a time period, say a month, as already presented in Chapter 1. The (i, j)th entry of $\mathbf{X}_N$ contains the evolution of the market index for product i in time period j. If all time-product evolutions are independent in the sense that the evolution of the value of product A for a given month does not impact the evolution of the value of product B, then it is expected that the rows of $\mathbf{X}_N$ are statistically independent. Also, if the time scale is chosen such that the evolution of the price of product A over a given time period is roughly uncorrelated with its evolution on the subsequent time period, then the columns will also be statistically independent. Therefore, after proper centralization and normalization of the entries (to ensure constant variance), it is expected that, if N, n are large, the empirical eigenvalue distribution of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ follows the Marčenko–Pastur law. If not, i.e. if some eigenvalues are found outside the support, then there exist some non-trivial correlation patterns in $\mathbf{X}_N$. The largest eigenvalue here allows the trader to anticipate the largest possible gain to be made if he aligns his portfolio on the corresponding eigenvector. But the sample eigenvector associated to the largest eigenvalue has no reason to be connected to the population eigenvector (which the trader wants to retrieve) associated with the largest population spike.

Nonetheless, as one may expect from the classical $n \gg N$ regime, the eigenvector $\mathbf{u}_1$ associated to the largest population eigenvalue ω_1 (in case of multiplicity one) should be much aligned to the corresponding sample eigenvector $\hat{\mathbf{u}}_1$ associated to the largest sample eigenvalue λ_1, in the sense that $|\hat{\mathbf{u}}_1^{\mathsf{H}}\mathbf{u}_1| \simeq 1$. This being said, if ω_1 is so small that λ_1 does not appear as a spike, there is intuitively little hope that $\hat{\mathbf{u}}_1$ be at all connected to $\mathbf{u}_1$, and we might even expect $|\hat{\mathbf{u}}_1^{\mathsf{H}}\mathbf{u}_1| \simeq 0$. This is what we study next.

Prior to this, let us provide some engineering motivation for this study, beside portfolio optimization in finance. In signal processing, the connection between the population (or signal) eigenspace and the observation has a key importance in subspace methods such as the MUSIC algorithm, which is known to have problems in the $N \simeq n$ regime, exactly due to the difficulty to evaluate the noise subspace in this scenario. Knowing more about the connection between the signal (or noise) and the observation subspaces allows one to propose improved subspace methods, as will be discussed in Section 17.1. In wireless communications, and particularly in cognitive radio or interference alignment schemes, it is classical to use precoders based on the channel state information between two signal sources so to avoid or minimize self-interference. These precoders intend to align their eigenvectors orthogonally to the subspace spanned by the interference channel. However, in fast fading channels, obtaining channel state information is rate-expensive so that only partial information is known, leading to residual interference, which it is fundamental to properly measure.

The first result, initially due to Paul in the real Gaussian case [Paul, 2007], then generalized to unitarily-invariant matrices in [Couillet and Hachem, 2012] is given as follows.

Theorem 9.5 ([Couillet and Hachem, 2012])**.** *Let $\mathbf{R}_N \in \mathbb{C}^{N\times N}$ be a random unitarily invariant Hermitian matrix with eigenvalues $\nu_1 \geq \ldots \geq \nu_N$ such that $F^{\mathbf{R}_N} \Rightarrow F$ with compact support $[a,b]$ with $\nu_1 \xrightarrow{\text{a.s.}} b$ and $\nu_N \xrightarrow{\text{a.s.}} b$. For t and $j_1,\ldots,j_t$ fixed with $\sum_i j_i = r$, let $\mathbf{P}_N = \mathbf{U}_N\mathbf{\Omega}\mathbf{U}_N^{\mathsf{H}} \in \mathbb{C}^{N\times N}$ with $\mathbf{U}_N = [\mathbf{U}_1^N,\ldots,\mathbf{U}_t^N] \in \mathbb{C}^{N\times r}$ isometric, $\mathbf{U}_i^N \in \mathbb{C}^{N\times j_i}$, and $\mathbf{\Omega} = \operatorname{diag}(\omega_1\mathbf{I}_{j_1},\ldots,\omega_t\mathbf{I}_{j_t})$, $\omega_1 > \ldots > \omega_t > -1$. Define*

$$\mathbf{B}_N = (\mathbf{I}_N + \mathbf{P}_N)^{\frac{1}{2}}\mathbf{R}_N(\mathbf{I}_N + \mathbf{P}_N)^{\frac{1}{2}}$$

with largest eigenvalues $\lambda_1 \geq \ldots \geq \lambda_r$ and associated eigenvectors the ordered columns of $[\hat{\mathbf{U}}_1^N,\ldots,\hat{\mathbf{U}}_t^N]$, respectively. Then, as $N \to \infty$,

$$(\mathbf{U}_l^N)^{\mathsf{H}}\hat{\mathbf{U}}_k^N(\hat{\mathbf{U}}_k^N)^{\mathsf{H}}\mathbf{U}_l^N \xrightarrow{\text{a.s.}} \begin{cases} \xi_k\delta(k-l)\mathbf{I}_{j_l} & , \ \omega_k \text{ isolated} \\ 0 & , \text{ otherwise} \end{cases}$$

where, by "ω_k isolated", we mean $b^+ m_F(b^+)\omega_k(1+\omega_k)^{-1} < -1$ or $a^- m_F(a^-)\omega_k(1+\omega_k)^{-1} > -1$, and with

$$\xi_k = \frac{m_F(\rho_k)(1+\rho_k m_F(\rho_k))}{m_F(\rho_k) + \rho_k m_F'(\rho_k)}$$

and ρ_k as in Theorem 9.2.

An interesting corollary, often sufficient for application purposes, is the scenario where $\mathbf{R}_N = \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$, normalized Gaussian, as follows.

Corollary 9.2 ([Couillet and Hachem, 2012; Paul, 2007])**.** *In the conditions of Theorem 9.5, let $\mathbf{R}_N = \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$, $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ with independent real or complex Gaussian entries with zero mean and variance $1/n$. Then, as $N, n \to \infty$, $N/n \to c$, the following holds,*

$$(\mathbf{U}_l^N)^{\mathsf{H}}\hat{\mathbf{U}}_k^N(\hat{\mathbf{U}}_k^N)^{\mathsf{H}}\mathbf{U}_l^N \xrightarrow{\text{a.s.}} \begin{cases} \xi_k\delta(k-l)\mathbf{I}_{j_l} \ , & \omega_k > \sqrt{c} \text{ or } \omega_k < -\sqrt{c} \\ 0 \ , & \text{otherwise} \end{cases}$$

where $\xi_k = (1 - c\omega_k^{-2})(1 + \omega_k^{-1})^{-1}$.

What the above result says is that, when a population spike ω_k is above the detectability threshold $\sqrt{c}$ (or below its opposite), the eigenspace associated to this eigenvalue will align to the corresponding eigenspace of the j_k corresponding sample eigenvalues and to these only. This alignment tends to one as $\omega_k \to \infty$ or as $c \to 0$, as predicted by the $n \gg N$ regime, and tends to zero as $\omega_k \downarrow \sqrt{c}$ as expected from the previous discussion. This means in particular that, in the asymptotic regime, nothing can be said about the population eigenvectors if $\omega_k < \sqrt{c}$.

In Figure 9.3, the situation of a single population spike ω_1 with multiplicity one is considered. The matrix dimensions N and n are taken to be such that $N/n = 1/3$, and $N \in \{100, 200, 400\}$. We compare the averaged empirical projections $\left|(\hat{\mathbf{u}}_1^N)^{\mathsf{H}}\mathbf{u}_1^N\right|$ (with $\hat{\mathbf{u}}_1^N = \hat{\mathbf{U}}_1^N$ and $\mathbf{u}_1^N = \mathbf{U}_1^N$ in this vector case) against Theorem 9.5. We observe that the convergence rate of the limiting projection is very slow. Therefore, although nothing can be said asymptotically on the eigenvectors of a spiked model, when $\omega_1 < \sqrt{c}$, there exists a large range of values of N and n for which this is not so.

In the following, we provide a proof of Theorem 9.5, based on the complex integration method discussed in Section 8.1.2.

Proof of Theorem 9.5. We only prove here the case of main interest where $\omega_k > \sqrt{c}$ and discard the indexes N for readability. We also assume for simplicity of notation that $\omega_t > \sqrt{c}$. We shall prove that, for $\mathbf{a}_N, \mathbf{b}_N \in \mathbb{C}^N$ two vectors of bounded spectral norm,

$$\mathbf{a}_N^{\mathsf{H}}\hat{\mathbf{U}}_k\hat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{b}_N - \xi_k\mathbf{a}_N^{\mathsf{H}}\mathbf{U}_k\mathbf{U}_k^{\mathsf{H}}\mathbf{b}_N \xrightarrow{\text{a.s.}} 0$$

with $\xi_k = (1 - c\omega_k^{-2})(1 + c\omega_k^{-1})^{-1}$.

The starting point is the Cauchy integration formula (8.14), which states that

$$\mathbf{a}_N^{\mathsf{H}}\hat{\mathbf{U}}_k\hat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{b}_N = -\frac{1}{2\pi i}\oint_{\mathcal{C}_k} \mathbf{a}_N^{\mathsf{H}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{b}_N dz$$

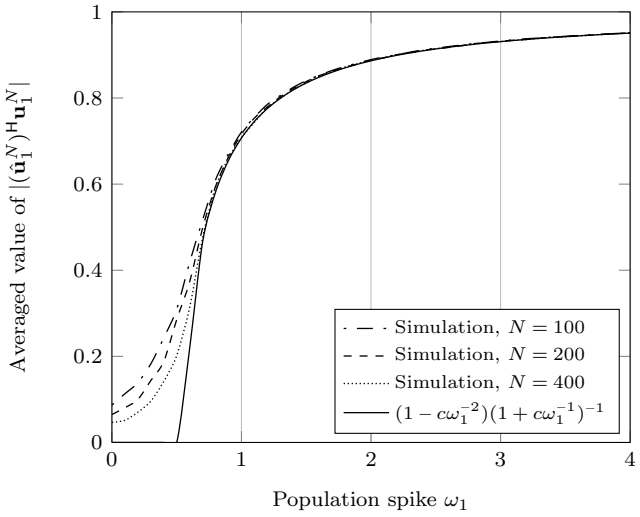

Figure 9.3 Averaged absolute value of the projector $|(\hat{\mathbf{u}}_1^N)^{\mathsf{H}}\mathbf{u}_1^N|$ of the eigenvectors corresponding to the single population and sample spikes in the model $\mathbf{B}_N = (\mathbf{I}_N + \mathbf{P}_N)^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}(\mathbf{I}_N + \mathbf{P}_N)^{\frac{1}{2}}$, $\mathbf{P}_N = \omega_1\mathbf{u}_1^N(\mathbf{u}_1^N)^{\mathsf{H}}$, with $N/n = 1/3$ and varying ω_1.

for a positively-oriented contour $\mathcal{C}_k$ circling around the j_k sample spike eigenvalues of $\mathbf{B}_N$ associated with ω_k (so $\mathcal{C}_k$ lies away from the limiting support of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$). The idea of the following is similar to the proof of Theorem 9.1. We wish to rewrite the expression above so to isolate the resolvent $\mathbf{Q}(z) = (\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - z\mathbf{I}_N)^{-1}$. Factorizing $(\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}}$ on each side, then $\mathbf{Q}(z)$ on the left, and using Woodbury's matrix inversion formula, we can write

$$
\begin{aligned}
&(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\
&= (\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}} \left(\mathbf{X}_N\mathbf{X}_N - z\mathbf{I}_N + z\mathbf{P}_N(\mathbf{I}_N + \mathbf{P}_N)^{-1}\right)^{-1} (\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}} \\
&= (\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}}\mathbf{Q}(z)\left(\mathbf{I}_N - z\mathbf{U}_N\hat{\mathbf{H}}(z)^{-1}\mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1}\mathbf{U}_N^{\mathsf{H}}\mathbf{Q}(z)\right)(\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}}
\end{aligned}
$$

where $\hat{\mathbf{H}}(z) = \mathbf{I}_r + z\mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1}\mathbf{U}_N^{\mathsf{H}}\mathbf{Q}(z)\mathbf{U}_N$. This matrix of small size and featuring $\mathbf{Q}(z)$ and deterministic matrices is now our center of interest. With the notations

$$
\begin{aligned}
\tilde{\mathbf{a}}_N^{\mathsf{H}} &= z\mathbf{a}_N^{\mathsf{H}}(\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}}\mathbf{Q}(z)\mathbf{U}_N \\
\tilde{\mathbf{b}}_N &= \mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1}\mathbf{U}_N^{\mathsf{H}}\mathbf{Q}(z)(\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}}
\end{aligned}
$$

we can then write our initial equation

$$
\begin{aligned}
&\mathbf{a}_N^{\mathsf{H}}\hat{\mathbf{U}}_k\hat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{b}_N \\
&= -\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\mathbf{a}_N^{\mathsf{H}}(\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}}\mathbf{Q}(z)(\mathbf{I}_N + \mathbf{P}_N)^{-\frac{1}{2}}\mathbf{b}_N dz + \frac{1}{2\pi i}\oint_{\mathcal{C}_k}\tilde{\mathbf{a}}_N^{\mathsf{H}}\hat{\mathbf{H}}(z)\tilde{\mathbf{b}}_N.
\end{aligned}
$$

Of course, the interesting term is the second term, as, for N large, with probability one, the first term has no pole within $\mathcal{C}_k$ (because the eigenvalues of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ are asymptotically found within their limiting support). Since $\tilde{\mathbf{a}}_N$, $\tilde{\mathbf{b}}_N$, and $\hat{\mathbf{H}}(z)$ are small vectors or matrices with simple expressions in terms of $\mathbf{Q}(z)$, their large dimensional behavior is easy to study and we obtain, through dominated convergence arguments,

$$\mathbf{a}_N^{\mathsf{H}}\hat{\mathbf{U}}_k\hat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{b}_N - \frac{1}{2\pi i}\oint_{\mathcal{C}_k} zm(z)^2\mathbf{a}_N^{\mathsf{H}}\mathbf{U}_N(\mathbf{I}_r+\mathbf{\Omega})^{-\frac{1}{2}}\mathbf{H}(z)^{-1}(\mathbf{I}_r+\mathbf{\Omega})^{-\frac{1}{2}}\mathbf{U}_N^{\mathsf{H}}\mathbf{b}_N \xrightarrow{\text{a.s.}} 0$$

where

$$\mathbf{H}(z) = \operatorname{diag}((1+zm(z)\omega_1(1+\omega_1)^{-1})\mathbf{I}_{j_1},\ldots,(1+zm(z)\omega_t(1+\omega_t)^{-1})\mathbf{I}_{j_t})$$

and $m(z)$ is the Stieltjes transform of the Marčenko-Pastur law (3.20). Decomposing $\mathbf{U}_N$ as $\mathbf{U}_N = [\mathbf{U}_1^N,\ldots,\mathbf{U}_t^N]$, we then have

$$\mathbf{a}_N^{\mathsf{H}}\hat{\mathbf{U}}_k\hat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{b}_N - \sum_{i=1}^{t}\frac{\mathbf{a}_N^{\mathsf{H}}\mathbf{U}_i^N(\mathbf{U}_i^N)^{\mathsf{H}}\mathbf{b}_N}{1+\omega_i}\frac{1}{2\pi i}\oint_{\mathcal{C}_k}\frac{zm(z)^2}{(1+\omega_i)\omega_i^{-1}+zm(z)}dz \xrightarrow{\text{a.s.}} 0.$$

From the proof of Theorem 9.1 (or equivalently from Theorem 9.2), $(1+\omega_i)\omega_i^{-1}+zm(z)=0$ for z in the interior of $\mathcal{C}_k$ only for $i=k$. Residue calculus then gives

$$\mathbf{a}_N^{\mathsf{H}}\hat{\mathbf{U}}_k\hat{\mathbf{U}}_k^{\mathsf{H}}\mathbf{b}_N - \frac{m(\rho_k)(1+\rho_k m(\rho_k))}{m(\rho_k)+\rho_k m'(\rho_k)}\mathbf{a}_N^{\mathsf{H}}\mathbf{U}_k^N(\mathbf{U}_k^N)^{\mathsf{H}}\mathbf{b}_N \xrightarrow{\text{a.s.}} 0$$

with ρ_k the real solution to $(1+\omega_k)\omega_k^{-1}+zm(z)=0$. Replacing m by its explicit formula and using $\rho_k = 1+\omega_k+c(1+\omega_k)\omega_k^{-1}$ (from Theorem 9.1), we finally find the desired result. □

Similar results were derived in [Benaych-Georges and Rao, 2011] for the models $\mathbf{B}_N = \mathbf{R}_N(\mathbf{I}_N+\mathbf{P}_N)$ and $\mathbf{B}_N = \mathbf{R}_N+\mathbf{P}_N$ with $\mathbf{R}_N$ bi-unitarily invariant, similar to those of Theorem 9.3 and Theorem 9.2. Being of less interest in practice (in most practical scenarios, one is interested either in $\mathbf{B}_N = (\mathbf{I}_N+\mathbf{P}_N)^{\frac{1}{2}}\mathbf{R}_N(\mathbf{I}_N+\mathbf{P}_N)^{\frac{1}{2}}$ or $\mathbf{B}_N = (\mathbf{X}_N+\mathbf{P}_N)(\mathbf{X}_N+\mathbf{P}_N)^{\mathsf{H}}$), we only refer the reader to the original article.

As the spiked models are of particular interest for signal detection and statistical inference purposes, it is desirable to derive second order statistics of the above results, in order to be able to derive statistical tests. Indeed, taking for instance the result on the largest eigenvalues, the theorems above only give conditions for their asymptotic presence or absence outside a given limiting support. For practical applications however, with finite system dimensions, and therefore with a non-zero probability of finding eigenvalues outside the support even without perturbation, those results do not give any hint on an appropriate test to decide on the presence or the absence of a population spike based on the observation of the sample eigenvalues. For this, one needs to go further and come up with central limit theorems for the quantities studied above. This is the target of the following section.

9.1.2 Second order statistics of spiked models

Although it might seem more natural to first of all study the fluctuations of the extreme eigenvalues in matrix models without perturbations before looking into spiked models, it turns out that the former is much more complicated than the later. As a matter of fact, so far, no characterization of the asymptotic fluctuations of the extreme eigenvalues for the non-perturbed case, even for the simplest Wishart model, was obtained using the tools developed so far. These results will be presented instead in Section 9.2, with the help of new tools which will be shortly introduce.

In the present section, we characterize the fluctuations of the largest eigenvalues and associated eigenspace projections in the case of detectable spikes (i.e. when sample spikes are found outside the limiting support of the underlying model distribution). Similar to the previous section, the techniques follow the original ideas of Benaych-Georges [Benaych-Georges *et al.*, 2010] but also in parallel of Capitaine *et al.* [Capitaine *et al.*, 2009] who characterized the limiting first and second order statistics of deformed Wigner matrices rather than deformed sample covariance matrices. The techniques employed for the proof of Theorem 9.1 and Theorem 9.5, based in particular on the complex integration approach, can in fact be used to characterize the limiting second order statistics of individual eigenvalues and eigenvector projections, but fail to characterize the *joint* fluctuations when the population spikes have multiplicity greater than one. This is due to the fact that the contour integral technique cannot isolate individual eigenvalues which converge to the same limit. The solution to this is brought in [Benaych-Georges *et al.*, 2010] via an elegant technique which we shall discuss below.

The most general result, characterizing the joint fluctuations of eigenvalues and eigenvectors for the model $\mathbf{B}_N = (\mathbf{I}_N + \mathbf{P}_N)^{\frac{1}{2}} \mathbf{R}_N (\mathbf{I}_N + \mathbf{P}_N)^{\frac{1}{2}}$ with $\mathbf{R}_N$ following the conditions of Theorem 9.2, is given in [Couillet and Hachem, 2012], as follows.

Theorem 9.6 ([Couillet and Hachem, 2012])**.** *Let $\mathbf{B}_N \in \mathbb{C}^{N \times N}$ be defined as in Theorem 9.5, and assume $\omega_1, \ldots, \omega_t$ isolated in the sense of Theorem 9.5. Define the following quantities*

$$\mathbf{H}_i^N = \sqrt{N} (\mathbf{U}_i^N)^{\mathsf{H}} \left(\hat{\mathbf{U}}_i^N (\hat{\mathbf{U}}_i^N)^{\mathsf{H}} - \xi_i \mathbf{I}_N \right) \mathbf{U}_i^N$$

$$\mathbf{L}_i^N = \sqrt{N} \begin{pmatrix} \lambda_{j_1+\ldots+j_{i-1}+1} - \rho_i \\ \vdots \\ \lambda_{j_1+\ldots+j_i} - \rho_i \end{pmatrix}$$

with ξ_i and ρ_i defined as in Theorem 9.2 and Theorem 9.5. Then, as $N \to \infty$,

$$\left((\mathbf{H}_i^N, \mathbf{L}_i^N) \right)_{i=1}^t \Rightarrow \left((\mathbf{H}_i, \mathbf{L}_i) \right)_{i=1}^t$$

with $\mathbf{L}_i \in \mathbb{C}^N$ *the vector of decreasingly-ordered eigenvalues of* $\mathbf{K}_i \in \mathbb{C}^{N\times N}$*, where*

$$\begin{pmatrix}\mathbf{H}_i\\ \mathbf{K}_i\end{pmatrix} = \left((\mathbf{D}(\rho_i)\mathbf{F}(\rho_i)\mathbf{D}(\rho_i)^{\mathsf{H}})^{\frac{1}{2}} \otimes \mathbf{I}_{j_i}\right)\begin{pmatrix}\mathbf{G}_{1,i}\\ \mathbf{G}_{2,i}\end{pmatrix}$$

where '$\otimes$' is the matrix Kronecker product, $\mathbf{G}_{1,1}, \ldots, \mathbf{G}_{2,t}$ *are independent GUE matrices*[2] *with* $\mathbf{G}_{1,k}, \mathbf{G}_{2,k} \in \mathbb{C}^{j_k\times j_k}$*, and* $\mathbf{D}(\rho), \mathbf{F}(\rho)$ *are defined, for* $\rho \in \mathbb{R}\setminus[a,b]$ *(recall that* $[a,b]$ *is the support of* F*), as*

$$\mathbf{D}(\rho) = \begin{pmatrix}\frac{h(\rho)(1+h(\rho))h''(\rho)}{h'(\rho)^3} & -\frac{h(\rho)(1+h(\rho))}{h'(\rho)^2}\\ -\frac{\rho}{h'(\rho)} & 0\end{pmatrix}$$

$$\mathbf{F}(\rho) = \begin{pmatrix}m'_F(\rho) - m_F(\rho)^2 & \frac{1}{2}m''_F(\rho) - m_F(\rho)m'_F(\rho)\\ \frac{1}{2}m''_F(\rho) - m_F(\rho)m'_F(\rho) & \frac{1}{6}m'''_F(\rho) - m'_F(\rho)^2\end{pmatrix}$$

where $h(x) = xm_F(x)$*.*

Although the result is not easy to appreciate, a few important observations can already be made. First, the scaled eigenvalues and eigenspace projections of the t groups of sample spikes are asymptotically independent. Then, if ω_k is a population spike with multiplicity greater than one, the joint distribution of the eigenvalues (or of the eigenspace projections) is, up to a scaling factor, the joint distribution of the eigenvalues of a GUE matrix. This being said, when the multiplicity of each population spikes is one and $\mathbf{R}_N = \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ with $\mathbf{X}_N$ normalized Gaussian, we have the following, more practical, corollary.

Corollary 9.3 ([Couillet and Hachem, 2012])**.** *Assume the conditions and notations of Theorem 9.6, and take* $j_i = 1$ *for* $i \in \{1,\ldots,t\}$ *and* $\mathbf{R}_N = \mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ *with* $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ *with independent Gaussian entries with zero mean and variance* $1/n$*. Then, as* $N, n \to \infty$*,* $N/n \to c$*, assuming* $\omega_i \in (-1, -\sqrt{c}) \cup (\sqrt{c}, \infty)$ *for all* $i \in \{1,\ldots,t\}$*,*

$$\left((\mathbf{H}_i^N, \mathbf{L}_i^N)\right)_{i=1}^t \Rightarrow X \sim \mathcal{N}\left(0, \operatorname{diag}(\mathbf{M}(\omega_1), \ldots, \mathbf{M}(\omega_t))\right)$$

where

$$\mathbf{M}(\omega) = \begin{pmatrix}\frac{c^2(1+\omega)^2}{(c+\omega)^2(\omega^2-c)}\left(c\frac{(1+\omega)^2}{(c+\omega)^2}+1\right) & \frac{(1+\omega)^3c^2}{(c+\omega)^2\omega}\\ \frac{(1+\omega)^3c^2}{(c+\omega)^2\omega} & \frac{c(1+\omega)^2(\omega^2-c)}{\omega^2}\end{pmatrix}.$$

In the following, we provide a short idea of the technique at play for the proof of the results above (alternative approaches can be found in [Bai and Yao, 2008a,b]).

[2] We recall that a GUE matrix $\mathbf{G} \in \mathbb{C}^{j\times j}$ (from the Gaussian unitary ensemble) is a random Hermitian matrix with independent entries, whose diagonal entries are $\mathcal{N}(0,1)$ and upper-diagonal entries are $\mathcal{CN}(0,1)$.

Proof. As already mentioned, the main difficulty is to deal with the sample spikes all converging to the same limiting value and for which complex integration methods are inadequate as contours cannot be simply designed that isolate these eigenvalues for all N. We therefore use a technique from Benaych-Geroges here. The key idea is to relate the eigenvalues $\lambda_1, \ldots, \lambda_r$ to the determinantal equation

$$\det\left(\mathbf{I}_r + x\mathbf{\Omega}(\mathbf{I}_r + \mathbf{\Omega})^{-1}\mathbf{U}^{\mathsf{H}}(\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}\mathbf{U}\right) = \det\hat{\mathbf{H}}(x) = 0$$

to which they are the solutions (recall the proof of Theorem 9.1). Note in particular that, for N sufficiently large, and with probability one, letting $x_1(i) > y_1(i) > x_2(i) > \ldots > y_{j_i}(i)$, for $i = 1, \ldots, t$,

$$\begin{aligned} &P\left(x_l(i) < \sqrt{N}(\lambda_{j_1+\ldots+j_{i-1}+l} - \rho_i) < y_l(i), \text{ for all } i, l\right) \\ &= P\left(\det\hat{\mathbf{H}}(\rho_i + N^{-\frac{1}{2}}x_l(i))\det\hat{\mathbf{H}}(\rho_i + N^{-\frac{1}{2}}y_l(i)) < 0, \text{ for all } i, l\right). \end{aligned}$$

Indeed, since $\det\hat{\mathbf{H}}(x)$ has zeros only at $\lambda_1, \ldots, \lambda_r$ for all large N, it changes sign between each of them. The product of two determinants taken at the intermediate positions

$$\begin{aligned} &\lambda_{j_1+\ldots+j_{i-1}+(l-1)} < \rho_i + N^{-\frac{1}{2}}x_l(i) < \lambda_{j_1+\ldots+j_{i-1}+l} \\ &< \rho_i + N^{-\frac{1}{2}}y_l(i) < \lambda_{j_1+\ldots+j_{i-1}+(l+1)} \end{aligned}$$

is therefore negative. The converse is also clear. Therefore, it suffices to study the behavior of the determinant of $\hat{\mathbf{H}}$. From now on, let us focus on the j_i eigenvalues $\lambda_{j_1+\ldots+j_{i-1}+1}$ to $\lambda_{j_1+\ldots+j_i}$ for simplicity of notations. Developing the determinant and using the convergence relation already established in Theorem 9.1 and its proof, one can show that

$$\begin{aligned} &N^{\frac{j_i}{2}}\det\hat{\mathbf{H}}(\rho_i + N^{-\frac{1}{2}}x) - \prod_{j\neq i}\left(\frac{\omega_i - \omega_k}{\omega_i(1+\omega_k)}\right)^{j_k} \\ &\times \det\left(\sqrt{N}\frac{\omega_i\rho_i}{1+\omega_i}\mathbf{U}_i^{\mathsf{H}}(\mathbf{Q}(\rho_i) - m_F(\rho_i)\mathbf{I}_N)\mathbf{U}_i + \frac{\omega_i h'(\rho_i)}{1+\omega_i}x\mathbf{I}_{j_i}\right) \to 0 \end{aligned} \tag{9.1}$$

in probability, with $\mathbf{U}_i = \mathbf{U}_i^N$ and $\mathbf{Q}(x) = (\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} - x\mathbf{I}_N)^{-1}$. The source of fluctuations of $\det\hat{\mathbf{H}}(\rho_i + N^{-\frac{1}{2}}x)$ therefore lies entirely in the fluctuations of $\sqrt{N}\mathbf{U}_i^{\mathsf{H}}(\mathbf{Q}(\rho_i) - m_F(\rho_i)\mathbf{I}_N)\mathbf{U}_i$, which we need to study.

Using the left-unitary invariance of $\mathbf{X}_N$, denoting $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}} = \mathbf{V}\mathbf{L}\mathbf{V}^{\mathsf{H}}$ its spectral decomposition, it is clear that $\mathbf{U}\mathbf{V}$ is a Haar matrix, which we can write under the form $\mathbf{U}\mathbf{V} = \mathbf{Z}(\mathbf{Z}^{\mathsf{H}}\mathbf{Z})^{-\frac{1}{2}}$ (from Definition 4.6), where $\mathbf{Z} \in \mathbb{C}^{N\times r}$ is a standard Gaussian matrix. Clearly $N^{-1}\mathbf{Z}^{\mathsf{H}}\mathbf{Z} \in \mathbb{C}^{r\times r} \xrightarrow{\text{a.s.}} \mathbf{I}_r$ and therefore the term can be ignored. We conclude that $\sqrt{N}\mathbf{U}_i^{\mathsf{H}}(\mathbf{Q}(\rho_i) - m_F(\rho_i)\mathbf{I}_N)\mathbf{U}_i$ is asymptotically

equal in distribution to

$$N^{-\frac{1}{2}}\mathbf{Z}_i^{\mathsf{H}}\left((\mathbf{L}-z\mathbf{I}_N)^{-1}-\frac{1}{N}\operatorname{tr}(\mathbf{L}-z\mathbf{I}_N)^{-1}\mathbf{I}_N\right)\mathbf{Z}_i$$
$$=N^{-\frac{1}{2}}\sum_{k=1}^{N}\left(\frac{1}{l_k-z}-\frac{1}{N}\operatorname{tr}(\mathbf{L}-z\mathbf{I}_N)^{-1}\right)(\mathbf{z}_{i,k}\mathbf{z}_{i,k}^{\mathsf{H}}-\mathbf{I}_{j_i})$$

where $\mathbf{Z}_i\in\mathbb{C}^{N\times j_i}$ is such that $\mathbf{Z}=[\mathbf{Z}_1,\ldots,\mathbf{Z}_t]$, $\mathbf{Z}_i^*=[\mathbf{z}_{i,1},\ldots,\mathbf{z}_{i,N}]$, and $l_1,\ldots,l_N$ are the diagonal entries of $\mathbf{L}$. Note the addition of the term $-\mathbf{I}_{j_i}$ in the right-hand side term which has no effect in the result but will be fundamental to exhibit the GUE limit. Conditioned on $\mathbf{L}$, this $j_i\times j_i$ matrix clearly has Gaussian entries with zero mean (real on the diagonal and complex outside) with variance a function of $\mathbf{L}$. Using the convergence of the variance (see e.g. [Bai and Silverstein, 2004])

$$\frac{1}{N}\sum_{k=1}^{N}\left(\frac{1}{l_k-z}-\frac{1}{N}\sum_{k'=1}^{N}\frac{1}{l_{k'}-z}\right)^2\xrightarrow{\text{a.s.}}m_F'(z)-m_F(z)^2$$

then gives the limiting GUE fluctuations of the matrix $(m_F'(\rho_i)-m_F(\rho_i)^2)^{-\frac{1}{2}}\sqrt{N}\mathbf{U}_i^{\mathsf{H}}(\mathbf{Q}(\rho_i)-m_F(\rho_i)\mathbf{I}_N)\mathbf{U}_i$.

Coming back to (9.1), and along with the discussion above, we then conclude that $\det\hat{\mathbf{H}}(\rho_i+N^{-\frac{1}{2}}x)$ changes sign asymptotically when

$$\det\left(\frac{\omega_i\rho_i}{1+\omega_i}(m_F'(\rho_i)-m_F(\rho_i)^2)^{\frac{1}{2}}\mathbf{G}_i+\frac{\omega_i h'(\rho_i)}{1+\omega_i}x\mathbf{I}_{j_i}\right)$$

changes sign, where $\mathbf{G}_i\in\mathbb{C}^{j_i\times j_i}$ is a GUE matrix. With the same reasoning as previously, this corresponds to positions where x crosses one of the r eigenvalues of the matrix

$$-\frac{\rho_i}{h'(\rho_i)}(m_F'(\rho_i)-m_F(\rho_i)^2)^{\frac{1}{2}}\mathbf{G}_i.$$

This is exactly the claimed result for this block of eigenvalues. The independence across the t blocks follows from the fact that the matrices $\mathbf{Z}_i$ are independent.

For the fluctuations of the eigenspace projections, using the same complex integration tools as in the proof of Theorem 9.5 (here we do not need to isolate the eigenvectors attached to a given eigenspace), one can show that the fluctuations of the eigenspace projections depend both on the fluctuations of $\sqrt{N}\mathbf{U}_i^{\mathsf{H}}(\mathbf{Q}(\rho_i)-m_F(\rho_i)\mathbf{I}_N)\mathbf{U}_i$ as well as on the fluctuations of $\sqrt{N}\mathbf{U}_i^{\mathsf{H}}(\mathbf{Q}'(\rho_i)-m_F'(\rho_i)\mathbf{I}_N)\mathbf{U}_i$, which does not present major additional difficulties. Details are provided in [Couillet and Hachem, 2012] which we do not further discuss here. □

The above results are particularly interesting in signal processing in the applicative context of local failure localization in large dimensional systems, e.g. node failure or sudden parameter change in large sensor networks. The underlying idea is that a local failure may change the network topology, modeled though the covariance matrix of the nodal observations, by a small rank perturbation

of a base scenario. The perturbation matrix is a *signature of the failure* which is often easier to identify from its eigenvector properties than from its eigenvalues, particularly so in homogeneous networks where each failure leads to similar amplitudes of the extreme eigenvalues. See [Couillet and Hachem, 2012] for more details.

This completes this section on spiked models. Some of the above results will be recalled in the following section, when mentioning the phenomenon of phase transition, corresponding to a dramatic change of the fluctuations of the extreme eigenvalues when the population spikes move from non-isolated to isolated. In the following, we move back to the standard random matrix models with no perturbation and discuss the asymptotic statistics of the extreme eigenvalues which turn out not to behave as Gaussian random variables in the large dimensional regime.

9.2 Distribution of extreme eigenvalues

For motivation, we first discuss a scenario of application using the results on the distribution of the extreme eigenvalues in a standard random matrix model. Let us consider the case of the observation of a sample covariance matrix, upon which the experimenter would like to decide whether the population eigenvalue matrix is either an identity matrix or a perturbed identity matrix. A first idea is then to decide whether the observed largest eigenvalue is inside or outside the support of the Marčenko–Pastur law. However, due to the finite dimensionality of the matrix in practice, there is a non-zero probability for the largest observed eigenvalue to lie outside the support in the non-perturbed scenario. Further information on the statistical distribution of the largest eigenvalue of sample covariance matrices is then required, to be able to design adequate hypothesis tests. We will come back to these practical considerations in Chapter 16.

The study of the second order statistics for the extreme eigenvalues of the classically known random matrices took off in the mid-nineties, spurred by the work of Tracy and Widom [Tracy and Widom, 1996] on $N \times N$ (Wigner) Hermitian matrices with i.i.d. Gaussian entries above the diagonal. As their work does not use the same tools as developed so far, before introducing the results from Tracy and Widom, we first provide some notions of orthogonal polynomials in order to understand how these results are derived. This introduction is based on the tutorial [Fyodorov, 2005] and on [Anderson *et al.*, 2010, Chapter 3].

9.2.1 Introduction to the method of orthogonal polynomials

To study the behavior of some particular eigenvalue of a random matrix, it is required to study its marginal distribution. Calling $P^{\leq}_{(\lambda_1^N,\ldots,\lambda_N^N)}(\lambda_1^N,\ldots,\lambda_N^N)$ the joint density of the ordered eigenvalues $\lambda_1^N \leq \ldots \leq \lambda_N^N$ of some Hermitian

random matrix $\mathbf{X}_N \in \mathbb{C}^{N\times N}$, the largest eigenvalue λ_N^N has density

$$P_{\lambda_N^N}(\lambda_N^N) = \int_{\lambda_1^N} \dots \int_{\lambda_{N-1}^N} P^{\leq}_{(\lambda_1^N,\dots,\lambda_N^N)}(\lambda_1^N,\dots,\lambda_N^N) d\lambda_1^N \dots d\lambda_N^N. \tag{9.2}$$

In the case where the order of the eigenvalues is irrelevant, we have that

$$P_{(\lambda_1^N,\dots,\lambda_N^N)}(\lambda_1^N,\dots,\lambda_N^N) = \frac{1}{N!}\, P^{\leq}_{(\lambda_1^N,\dots,\lambda_N^N)}(\lambda_1^N,\dots,\lambda_N^N)$$

with $P_{(\lambda_1^N,\dots,\lambda_N^N)}$ the density of the unordered eigenvalues.

From now on, the eigenvalue indexes $1,\dots,N$ are considered to be just labels instead of ordering indexes. From the above equality, it is equivalent, and as will turn out actually simpler, to study the unordered eigenvalue distribution rather than the ordered eigenvalue distribution. In the particular case of a zero Wishart matrix with $n \geq N$ degrees of freedom, this property holds and we have from Theorem 2.3 that

$$P_{(\lambda_1^N,\dots,\lambda_N^N)}(\lambda_1^N,\dots,\lambda_N^N) = e^{-\sum_{i=1}^N \lambda_i^N} \prod_{i=1}^N \frac{(\lambda_i^N)^{n-N}}{(n-i)!i!} \prod_{i<j} (\lambda_i^N - \lambda_j^N)^2.$$

Similarly, we have for Gaussian Wigner matrices [Tulino and Verdú, 2004], i.e. Wigner matrices with upper-diagonal entries complex standard Gaussian and diagonal entries real standard Gaussian

$$P_{(\lambda_1^N,\dots,\lambda_N^N)}(\lambda_1^N,\dots,\lambda_N^N) = \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\sum_{i=1}^N (\lambda_i^N)^2} \prod_{i=1}^N \frac{1}{i!} \prod_{i<j} (\lambda_i^N - \lambda_j^N)^2. \tag{9.3}$$

The problem now is to be able to compute the multi-dimensional marginalization for either of the above distributions, or for more involved distributions. We concentrate on the simpler Gaussian Wigner case in what follows.

To be able to handle the marginalization procedure, we will use the *reproducing kernel* property, given below which can be found in [Deift, 2000].

Theorem 9.7. *Let $\mathbf{K}_n \in \mathbb{C}^{n\times n}$ with (i,j) entry $K_{ij} = f(x_i, x_j)$ for some complex-valued function f of two real variables and a real vector $\mathbf{x} = (x_1,\dots,x_n)$. The function f is said to satisfy the reproducing kernel property with respect to a real measure μ if*

$$\int f(x,y) f(y,z) d\mu(y) = f(x,z).$$

Under this condition, we have that

$$\int \det \mathbf{K}_n d\mu(x_n) = (q-(n-1)) \det \mathbf{K}_{n-1} \tag{9.4}$$

with

$$q = \int f(x,x) d\mu(x).$$

The above property is interesting in the sense that, if such a reproducing kernel property can be exhibited, then we can successively iterate (9.4) in order to perform marginalization calculus such as in (9.2).

By working on the expression of the eigenvalue distribution of Gaussian Wigner matrices (9.3), it is possible to write $P_{(\lambda_1^N,\dots,\lambda_N^N)}$ under the form

$$P_{(\lambda_1^N,\dots,\lambda_N^N)}(\lambda_1^N,\dots,\lambda_N^N) = C \det\left(\left\{e^{-\frac{1}{2}(\lambda_j^N)^2}\pi_{i-1}(\lambda_j^N)\right\}_{1\leq i,j\leq N}\right)^2$$

for *any* set of polynomials $(\pi_0,\dots,\pi_{N-1})$ with π_k of degree k and leading coefficient 1, and for some normalizing constant C. A proof of this fact stems from similar arguments as for the proof of Lemma 16.1, namely that the matrix above can be written under the form of the product of a diagonal matrix with entries $e^{\frac{1}{2}(\lambda_j^N)^2}$ and a matrix with polynomial entries $\pi_i(x_j)$, the determinant of which is proportional to the product of $e^{\sum_j(\lambda_j^N)^2}$ times the Vandermonde determinant $\prod_{i<j}(\lambda_i^N-\lambda_j^N)$.

Now, since we have the freedom to take any set of polynomials $(\pi_0,\dots,\pi_{N-1})$ with leading coefficient 1, we choose a set of *orthogonal polynomials* with respect to the weighting coefficient e^{-x^2}, i.e. we define $(\pi_0,\dots,\pi_{N-1})$ to be such that

$$\int e^{-x^2}\pi_i(x)\pi_j(x)dx = \delta_i^j.$$

Denoting now $\mathbf{K}_N \in \mathbb{C}^{N\times N}$ the matrix with (i,j) entry

$$K_{ij} = k_N(\lambda_i^N,\lambda_j^N) \triangleq \sum_{k=0}^{N-1}\left[e^{-\frac{1}{2}(\lambda_i^N)^2}\pi_k(\lambda_i^N)\right]\left[e^{-\frac{1}{2}(\lambda_j^N)^2}\pi_k(\lambda_j^N)\right]$$

we observe easily, from the fact that $\det(\mathbf{A}^2) = \det(\mathbf{A}^\mathsf{T}\mathbf{A})$, that

$$P_{(\lambda_1^N,\dots,\lambda_N^N)}(\lambda_1^N,\dots,\lambda_N^N) = C\det\mathbf{K}_N.$$

From the construction of $\mathbf{K}_N$, through the orthogonality of the polynomials $\pi_0(x),\dots,\pi_{N-1}(x)$, we have that

$$\int k_N(x,y)k_N(y,z)dy = k_N(x,y)$$

and the function k_N has the reproducing kernel property.

This ensures that

$$\int\dots\int \det\mathbf{K}_N dx_{k+1}\dots dx_N = (N-k)!\det\mathbf{K}_k$$

where the term $(N-k)!$ follows from the computation of $\int k_n(x,x)dx$ for $n\in\{k+1,\dots,N\}$.

To finally compute the probability distribution of the largest eigenvalue, note that the probability that it is greater than ξ is complementary to the probability that there is no eigenvalue in $B=(\xi,\infty)$. The latter, called the *hole probability*,

expresses as

$$\begin{aligned} Q_N(B) &= P\left(\lambda_i^N < \xi,\ i = 1, \dots, N\right) \\ &= \int \dots \int P_{(\lambda_1^N, \dots, \lambda_N^N)}(\lambda_1^N, \dots, \lambda_N^N) \prod_{k=1}^{N} (1 - 1_B(\lambda_k^N)) d\lambda_1^N \dots d\lambda_N^N. \end{aligned}$$

From the above discussion, expanding the product term, this can be shown to express as

$$Q_N(B) = \sum_{i=0}^{N} (-1)^i \frac{1}{i!} \int_B \dots \int_B \det \mathbf{K}_i \ d\lambda_1^N \dots d\lambda_i^N.$$

This last expression is in fact a *Fredholm determinant*, denoted $\det(\mathcal{I} - \mathcal{K}_N|_B)$, where $\mathcal{K}_N$ is called an integral operator with kernel k_N acting on square integrable functions on B. These Fredholm determinants are well-studied objects which generalize the classical matrix determinants. In particular, the kernel k_N is a largely known object. The main interest of the following section is that, up to an appropriate centering and scaling, $Q_N(B)$ exhibits limiting properties as $N \to \infty$. In the present scenario, since we investigate the fluctuations of the properly centered and scaled largest eigenvalue of $\mathbf{X}_N$, we are precisely interested into $P(N^{\frac{2}{3}}(N^{-\frac{1}{2}}\lambda_i^N - 2) < \xi,\ i = 1, \dots, N)$, and therefore, after an immediate variable change, into the kernel

$$a_N(x, y) = N^{-\frac{1}{6}} k_N(2\sqrt{N} + N^{-\frac{1}{6}}x, 2\sqrt{N} + N^{-\frac{1}{6}}y).$$

The kernel properties of k_N entail that a_N can be written under the form

$$a_N(x, y) = \frac{\psi_N(x)\psi_N'(y) - \psi_N(y)\psi_N'(x)}{x - y} - \frac{1}{2} N^{-\frac{1}{3}} \psi_N(x)\psi_N(y)$$

where $\psi_N(z)$ is a holomorphic function on $\mathbb{C}$ satisfying in particular $\psi_N(z) \to \mathrm{Ai}(z)$, with $\mathrm{Ai}(z)$ the Airy function. This implies that the probability $P(N^{\frac{2}{3}}(N^{-\frac{1}{2}}\lambda_i^N - 2) < \xi,\ \ i = 1, \dots, N)$ can be expressed as the Fredholm determinant $\det(\mathcal{I} - \mathcal{A}|_B)$ with $\mathcal{A}$ the integral operator with kernel

$$a(x, y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x - y}.$$

That is,

$$P\left(N^{\frac{2}{3}}(N^{-\frac{1}{2}}\lambda_i^N - 2) < \xi,\ i = 1, \dots, N\right) \to \det\left(\mathcal{I} - \mathcal{A}|_{(\xi, \infty)}\right).$$

This result, implicitly known for a long time, is however seemingly unsatisfying as the computation of Fredholm determinants is not an easy task.

The major contribution of Tracy and Widom, which we will introduce next, is to show that the above Fredholm determinant intimately relates to differential Painlevé equations, which conveys a closed form expression for the above limiting probability distribution, known now as the Tracy-Widom law. However, and quite interestingly, it was shown by Bornemann [Bornemann, 2009a,b] that the

numerical evaluation of Fredholm determinants is in fact much lighter than the actual resolution of Painlevé equations.

Further information on the tools above can be found in the early book from Mehta [Mehta, 2004], the very clear tutorial from Fyodorov [Fyodorov, 2005], and the specialized book from Anderson, Guionnet, and Zeitouni [Anderson *et al.*, 2010], among others. In the following, we introduce the main results concerning limit laws of extreme eigenvalues known to this day.

9.2.2 Limiting laws of the extreme eigenvalues

The major result on the limiting density of extreme eigenvalues is due to Tracy and Widom. It comes as follows.

Theorem 9.8 ([Tracy and Widom, 1996])**.** *Let $\mathbf{X}_N \in \mathbb{C}^{N\times N}$ be Hermitian with independent Gaussian off-diagonal entries of zero mean and variance $1/N$. Denote λ_N^- and λ_N^+ the smallest and largest eigenvalues of $\mathbf{X}_N$, respectively. Then, as $N \to \infty$*

$$N^{\frac{2}{3}}\left(\lambda_N^+ - 2\right) \Rightarrow X^+ \sim F^+$$
$$N^{\frac{2}{3}}\left(\lambda_N^- + 2\right) \Rightarrow X^- \sim F^-$$

where F^+ is the Tracy–Widom law given by:

$$F^+(t) = \exp\left(-\int_t^\infty (x-t)^2 q^2(x)dx\right) \tag{9.5}$$

with q the Painlevé II function that solves the differential equation

$$q''(x) = xq(x) + 2q^3(x)$$
$$q(x) \sim_{x\to\infty} \mathrm{Ai}(x)$$

in which $\mathrm{Ai}(x)$ is the Airy function, and F^- is defined as

$$F^-(x) \triangleq 1 - F^+(-x).$$

This theorem is in fact extended in [Tracy and Widom, 1996] to a more general class of matrix spaces, including the space of real symmetric matrices and that of quaternion-valued symmetric matrices, with Gaussian i.i.d. entries. Those are therefore all special cases of Wigner matrices. The space of Gaussian real symmetric matrices is referred to as the *Gaussian orthogonal ensemble* (denoted GOE), that of complex Gaussian Hermitian matrices is referred to as the *Gaussian unitary ensemble* (GUE) (already met previously in Theorem 9.6), and that of quaternion-valued symmetric Gaussian matrices is referred to as the *Gaussian symplectic ensemble* (GSE). The seemingly strange "orthogonal" and "unitary" denominations arise from deeper considerations on these ensembles, involving orthogonal polynomials, see, e.g., [Faraut, 2006] for details.

It was later shown [Bianchi *et al.*, 2010] that the random variables λ_N^+ and λ_N^- are asymptotically independent, giving therefore a simple description of their ratio, the *condition number* of $\mathbf{X}_N$.

Theorem 9.9 ([Bianchi *et al.*, 2010])**.** *Under the assumptions of Theorem 9.8*

$$\left(N^{\frac{2}{3}}\left(\lambda_N^+ - 2\right), N^{\frac{3}{2}}\left(\lambda_N^- + 2\right)\right) \Rightarrow (X^+, X^-)$$

where X^+ and X^- are independent random variables with respective distributions F^+ and F^-. The random variable λ_N^+/λ_N^- satisfies

$$N^{\frac{2}{3}}\left(\frac{\lambda_N^+}{\lambda_N^-} + 1\right) \Rightarrow -\frac{1}{2}\left(X^+ + X^-\right).$$

The result of interest to our study of extreme eigenvalues of Wishart and perturbed Wishart matrices was proposed later on by Johansson for the largest eigenvalue in the complex case [Johansson, 2000], followed by Johnstone [Johnstone, 2001] for the largest eigenvalue in the real case, while it took ten years before Feldheim and Sodin provided a proof of the result on the smallest eigenvalue in both real and complex cases [Feldheim and Sodin, 2010]. We only mention here the complex case.

Theorem 9.10 ([Feldheim and Sodin, 2010; Johansson, 2000])**.** *Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ be a random matrix with i.i.d. Gaussian entries of zero mean and variance $1/n$. Denoting λ_N^+ and λ_N^- the largest and smallest eigenvalues of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$, respectively, we have:*

$$N^{\frac{2}{3}}\frac{\lambda_N^+ - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}}\sqrt{c}} \Rightarrow X \sim F^+$$
$$N^{\frac{2}{3}}\frac{\lambda_N^- - (1-\sqrt{c})^2}{-(1-\sqrt{c})^{\frac{4}{3}}\sqrt{c}} \Rightarrow X \sim F^+$$

as $N, n \to \infty$ with $c = \lim_N N/n < 1$ and F^+ the Tracy–Widom distribution defined in Theorem 9.8. Moreover, the convergence result for λ_N^+ holds also for $c \geq 1$.

As we shall see in Section 16.4, a quantity of particular interest for signal detection under unknown noise variance is the ratio $T = \lambda_N^+(\frac{1}{N}\operatorname{tr}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}})^{-1}$. Indeed, while we assume above the entries of $\mathbf{X}_N$ of variance $1/n$, in practice, one is often led to considering noise observations $\mathbf{X}_N$ with entries of variance σ^2/n for some unknown σ^2. Taking the above ratio cancels this unknown parameter. Now, a mere application of the Slutsky lemma, Theorem 8.12, along with the above result, shows that $T \Rightarrow F^+$ in the large dimensional regime, and we have a first practical result of deep interest in signal processing.

However interesting since simple, this result, along with many limiting second order statistics, has the major drawback that the rate of convergence towards

the limit is rather slow (due to the slow rate of convergence of λ_N^+ to the Tracy–Widom). That is, N, n have to be taken rather large for the Tracy–Widom approximation to be relevant, especially in the tails. To address this question, Nadler [Nadler, 2011] provides a refined result, by introducing further statistical orders, as follows

Theorem 9.11 ([Nadler, 2011])**.** *Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ be a random matrix with i.i.d. Gaussian entries of zero mean and variance $1/n$ with largest eigenvalue λ_N^+. Denote*

$$T = \frac{\lambda_N^+}{\frac{1}{N}\operatorname{tr}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}}$$

Then, as $N, n \to \infty$ with $N/n \to c$,

$$P\left(\frac{T-\mu_{N,n}}{\sigma_{N,n}} > y\right) = 1 - F^+(y) + \frac{1}{2Nn}\frac{\mu_{Nn}^2}{\sigma_{Nn}^2}(F^+)''(y) + \frac{C(y)}{N^{\frac{2}{3}}}$$

where we defined

$$\mu_{N,n} = \frac{1}{n}\left(\sqrt{n-\frac{1}{2}} + \sqrt{N-\frac{1}{2}}\right)^2$$

$$\sigma_{N,n} = \sqrt{\frac{\mu_{N,n}}{n}}\left(\frac{1}{\sqrt{n-\frac{1}{2}}} + \frac{1}{\sqrt{p-\frac{1}{2}}}\right)^{\frac{1}{3}}$$

and where $C(y)$ is a constant term.

It is easy to see in the above formulation that *all* terms in the right-hand side are $O(N^{-\frac{2}{3}})$ and that, therefore, the expression may seem meaningless. It however turns out that $C(y)$ is relatively small for rather large values of y which makes the approximation without the $C(y)N^{-\frac{2}{3}}$ term very useful in practice.

The empirical against theoretical distributions of the largest eigenvalues of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ are depicted in Figure 9.4, for $N = 500$, $c = 1/3$.

Observe that the Tracy–Widom law is largely weighted on the negative half line. This means that the largest eigenvalue of $\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}$ has a strong tendency to lie much inside the support of the l.s.d. rather than outside. For the same scenario $N = 500$, $c = 1/3$, we now depict in Figure 9.5 the Tracy–Widom law against the empirical distribution of the largest eigenvalue of $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ in the case where $\mathbf{T}_N \in \mathbb{R}^{N\times N}$ is diagonal composed of all ones but for $T_{11} = 1.5$. From Theorem 7.2, no eigenvalue is found outside the asymptotic spectrum of the Marčenko–Pastur law. Figure 9.5 suggests that the largest eigenvalue of $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ does not converge to the Tracy–Widom law since it shows a much heavier tail in the positive side; this is however not true asymptotically. The asymptotic limiting distribution of the largest eigenvalue of $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ is still the Tracy–Widom law, but the convergence towards the second order limit

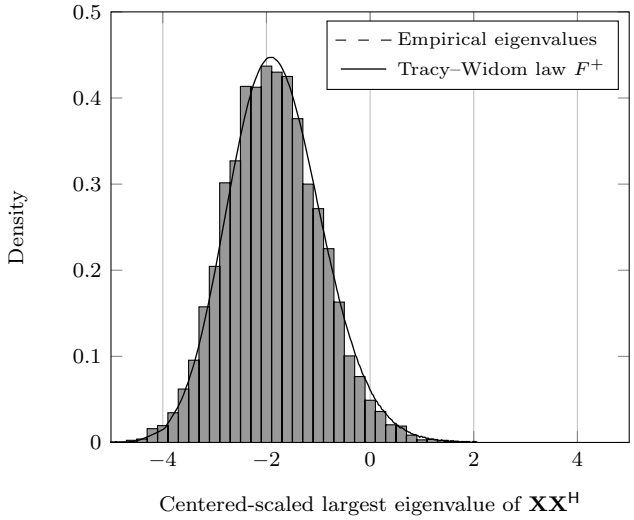

Figure 9.4 Density of $N^{\frac{2}{3}}c^{-\frac{1}{2}}(1+\sqrt{c})^{-\frac{4}{3}}\left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the Tracy–Widom law for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{X}\mathbf{X}^\mathsf{H}$ of Theorem 9.9. Empirical distribution taken over 10 000 Monte-Carlo simulations.

arises at a seemingly much slower rate. This is proved in the following theorem. To appreciate the convergence towards the Tracy–Widom law, N must then be taken much larger.

Theorem 9.12 ([Baik *et al.*, 2005])**.** *Let $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ have i.i.d. Gaussian entries of zero mean and variance $1/n$ and $\mathbf{T}_N = \mathrm{diag}(\tau_1, \ldots, \tau_N) \in \mathbb{R}^{N\times N}$. Assume, for some fixed r and k, $\tau_{r+1} = \ldots = \tau_N = 1$ and $\tau_1 = \ldots = \tau_k = 1+\omega$ while $\tau_{k+1}, \ldots, \tau_r$ lie in a compact subset of $(0, \tau_1)$. Assume further that the ratio N/n is constant, equal to $c < 1$ as N, n grow. Denoting λ_N^+ the largest eigenvalue of $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^\mathsf{H}\mathbf{T}_N^{\frac{1}{2}}$, we then have*

- *If $\omega < \sqrt{c}$*

$$N^{\frac{2}{3}}\frac{\lambda_N^+ - (1+\sqrt{c})^2}{(1+\sqrt{c})^{\frac{4}{3}}\sqrt{c}} \Rightarrow X^+ \sim F^+$$

  with F^+ the Tracy–Widom distribution.
- *If $\omega > \sqrt{c}$*

$$\left(c\frac{(1+\omega)^2(\omega^2-c)}{\omega^2}\right)^{-\frac{1}{2}} N^{\frac{1}{2}}\left[\lambda_N^+ - \left(1+\omega+c\frac{1+\omega}{\omega}\right)\right] \Rightarrow X_k \sim G_k$$

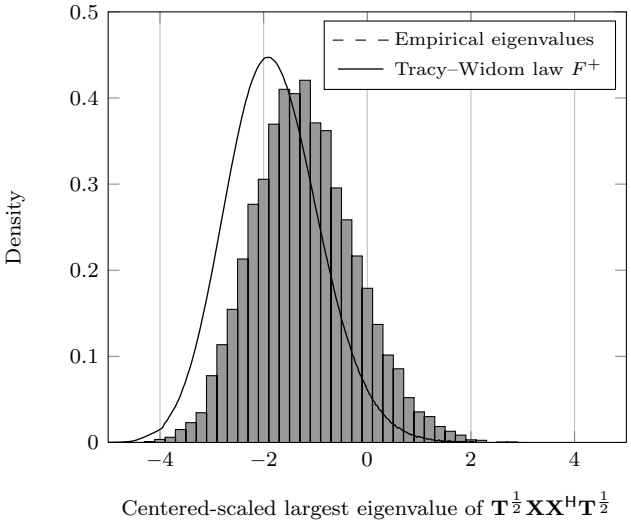

Figure 9.5 Distribution of $N^{\frac{2}{3}}c^{-\frac{1}{2}}(1+\sqrt{c})^{-\frac{4}{3}}\left[\lambda_N^+ - (1+\sqrt{c})^2\right]$ against the Tracy–Widom law for $N = 500$, $n = 1500$, $c = 1/3$, for the covariance matrix model $\mathbf{T}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\frac{1}{2}}$ with $\mathbf{T}$ diagonal with all entries 1 but for $T_{11} = 1.5$. Empirical distribution taken over 10 000 Monte-Carlo simulations.

with G_k the distribution function of the largest eigenvalue of the $k \times k$ GUE, given by:

$$G_k(x) = \frac{1}{Z_k}\int_{-\infty}^{x}\cdots\int_{-\infty}^{x}\prod_{1\le i<j\le k}|\xi_i-\xi_j|^2\prod_{i=1}^{k}e^{-\frac{1}{2}\xi_i^2}d\xi_1\ldots d\xi_k.$$

with Z_k a normalization constant. In particular, $G_1(x)$ is the Gaussian zero mean and unit variance distribution function.

Obviously, the second part of this result is not new, since it recalls Theorem 9.6. The corollary of Theorem 9.12 is that, if the largest population eigenvalue (i.e. the largest population spike) is not large enough for any eigenvalue of the sample covariance matrix to escape the support of the Marčenko–Pastur law, whatever its multiplicity k, then the asymptotic distribution of the largest eigenvalue is the Tracy–Widom law. This confirms that the behavior observed in Figure 9.5 has not reached its asymptotic limit. Theorem 9.12 goes further by stating that, if on the contrary τ_1 is larger than the transition limit $1+\sqrt{c}$, where an eigenvalue will be found outside the support of the Marčenko–Pastur law, then, if $k = 1$, the largest eigenvalue in $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ has a central limit with convergence rate $O(n^{-\frac{1}{2}})$ instead of the Tracy–Widom rate $O(n^{-\frac{2}{3}})$ when $\omega < \sqrt{c}$. This sudden convergence rate change is referred to by the author in [Baik *et al.*, 2005] as a *phase transition*. The case $\omega = \sqrt{c}$ is also treated in [Baik *et al.*, 2005] that shows that λ_N^+ converges in distribution to yet another law F_k, depending on the

multiplicity k of ω, with rate $O(n^{\frac{2}{3}})$; the law F_k does not have a simple expression though. Since the case $\omega = \sqrt{c}$ is highly improbable for practical applications, this case was deliberately discarded from Theorem 9.12.

On the other hand, this nice result is yet another disappointment for signal detection applications. Remember that one of our initial motivations to investigate further the asymptotic distribution of the largest eigenvalue of $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ was the inability to visually determine the presence of a spike $\omega < \sqrt{c}$ from the asymptotic spectrum of $\mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$. Now, it turns out that even the distribution of the largest eigenvalue in that case is asymptotically the same as that when $\mathbf{T}_N = \mathbf{I}_N$. There is therefore not much left to be done in the asymptotic regime to perform signal detection under the detection threshold $\sqrt{c}$. In that case, we may resort to further limit orders, or derive exact expressions of the largest eigenvalue distribution. Similar considerations are addressed in Chapter 16.

We also mention that, in the real case $\mathbf{X}_N \in \mathbb{R}^{N\times n}$, if $\omega > \sqrt{c}$ has multiplicity one, Paul proves that the limiting distribution of $\lambda_N^+ - \left(1+\omega+c(1+\omega)\omega^{-1}\right)$ is still Gaussian but with variance double that of the complex case, i.e. $2n^{-1}\left((1+\omega)^2 - c(1+\omega)^2\omega^{-2}\right)$ [Paul, 2007]. Theorem 9.12 is also extended in [Karoui, 2007] to more general Gaussian sample covariance matrix models, where it is proved that under some conditions on the population covariance matrix, for any integer k fixed, the largest k eigenvalues of the sample covariance matrix have a Tracy–Widom distribution with the same scaling factor but different centering and scaling coefficients.

Theorem 9.13 ([Karoui, 2007]). *Let $\mathbf{B}_N = \mathbf{T}_N^{\frac{1}{2}}\mathbf{X}_N\mathbf{X}_N^{\mathsf{H}}\mathbf{T}_N^{\frac{1}{2}}$ with $\mathbf{X}_N \in \mathbb{C}^{N\times n}$ having independent Gaussian entries with zero mean and variance $1/n$, and $\mathbf{T}_N$ Hermitian nonnegative with eigenvalues $\tau_1 \geq \ldots \geq \tau_n$. Assume $F^{\mathbf{T}_N} \Rightarrow H$ with compact support within $[a,b]$, and $\tau_1 \to b$, $\tau_N \to a$. Further assume that H has a density h in a left neighborhood I of b such that $h(x) \geq C(b-x)$ for some $C > 0$ and for all $x \in I$. Then, denoting λ_N^+ the largest eigenvalue of $\mathbf{B}_N$, as $N, n \to \infty$ with $N/n = c_N$ such that $0 < \liminf c_N \leq \limsup c_N < \infty$,*

$$\frac{n^{\frac{2}{3}}}{\sigma_N}\left(\lambda_N^+ - \mu_N\right) \Rightarrow X^+ \sim F^+$$

with F^+ the complex Tracy-Widom distribution, and

$$\mu_N = \frac{1}{\zeta_N}\left(1 + c_N\zeta_N \int \frac{t}{1-t\zeta_N} dF^{\mathbf{T}_N}(t)\right)$$
$$\sigma_N^3 = \frac{1}{\zeta_N^3}\left(1 + c_N\zeta_N^3 \int \left(\frac{t}{1-t\zeta_N}\right)^3 dF^{\mathbf{T}_N}(t)\right)$$

where ζ_N is the unique solution in $[0, 1/\tau_1)$ of the equation in x

$$cx^2 \int \left(\frac{t}{1-tx} \right)^2 dF^{\mathbf{T}_N}(t) = 1.$$

Of course, assuming that c_N converges (fast enough) to a constant c, the statement of the result above can be changed by taking the limits in μ_N and σ_N. Note the additional, up to now uncommon, assumption of the local existence of a density for h with the local bound $h(x) \geq C(b-x)$. This condition is in fact essential for the existence of the solution ζ as the defining integral tends to ∞ as $x \uparrow 1/\tau_1$. This assumption, which translates the fact that H must have a "sharp edge" in b, is completely in line with the Marčenko-Pastur case and the observations from Silverstein and Choi [Silverstein and Choi, 1995] on the square-root behaviour at the edge of the spectrum when H is discrete.

This concludes this chapter on the analysis of the extreme eigenvalues of standard random matrix models. In the next section, we summarize the different chapters of the theoretical part before moving to the applications of (most of) these results.