

2 Random Matrix Theory

Random matrix theory, at its inception, primarily dealt with the eigenvalue distribution (also referred to as the spectral measure) of large-dimensional random matrices. One of the key technical tools to study these measures is the Stieltjes transform, often presented as the central object of the theory [Bai and Silverstein, 2010, Pastur and Shcherbina, 2011].

But signal processing and machine learning alike are often more interested in subspaces and eigenvectors (which often carry the structural information of the data) than in eigenvalues. Subspace or spectral methods, such as principal component analysis (PCA) [Wold et al., 1987], spectral clustering [Luxburg, 2007] and some semi-supervised learning techniques [Zhu, 2005] are built directly upon the eigenspace spanned by the several top eigenvectors.

Consequently, beyond the Stieltjes transform, a more general mathematical object, the *resolvent* of large random matrices will constitute the cornerstone of the book. The resolvent of a matrix gives access to its spectral measure, to the location of its isolated eigenvalues, to the statistical behavior of their associated eigenvectors when random, and consequently provides an entry-door to the performance analysis of numerous machine learning methods.

This chapter introduces the fundamental objects and tools necessary to characterize the behavior of large-dimensional random matrices (the resolvent, the Stieltjes transform method, etc.) in Section 2.1, with a particular focus on the modern and powerful technical approach of deterministic equivalents. Section 2.2 then presents some foundational random matrix results (under the form of deterministic equivalents), which will serve as cornerstones for the various machine learning applications discussed in the remainder of this book. Section 2.3 is next devoted to advanced considerations on the limiting spectrum of sample covariance matrix models, with applications to statistical inference in Section 2.4. Section 2.5 then introduces the family of spiked models which, as we will see, play a crucial role in statistics, signal processing, and machine learning applications. Section 2.6 lists and discusses other models and tools of interest in the random matrix literatures, with a short introduction to the alternative free probability approach and related techniques. Section 2.7 is finally devoted to the “modern” concentration of measure framework for random matrices, which, as we just elaborated in the previous chapter, provides a strong justification of the universality of random matrix results when applied to real data machine learning, and also provides a convenient mathematical framework to deal with neural networks. The chapter closes

with concluding remarks in Section 2.8 and exercises in Section 2.9, both intended to familiarize the reader with the tools introduced in the chapter as well as to provide supplementary results and proofs.

2.1 Fundamental Objects

2.1.1 The Resolvent

We first introduce the resolvent of a matrix.

Definition 1 (Resolvent). *For a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, the resolvent $\mathbf{Q}_{\mathbf{M}}(z)$ of \mathbf{M} is defined, for $z \in \mathbb{C}$ not an eigenvalue of \mathbf{M} , as*

$$\mathbf{Q}_{\mathbf{M}}(z) \equiv (\mathbf{M} - z\mathbf{I}_n)^{-1}. \quad (2.1)$$

The matrix $\mathbf{Q}_{\mathbf{M}}(z)$ will often simply be denoted $\mathbf{Q}(z)$ when there is no ambiguity.

The resolvent operator is in fact a very classical tool, the use of which goes far beyond random matrix theory. It is, for instance, exploited in the analysis of linear operators in general Hilbert space [Akhiezer and Glazman, 2013] as well as in monotone operator theory of importance to modern convex optimization theory [Bauschke and Combettes, 2017].

2.1.2 Spectral Measure and Stieltjes Transform

The first use of the resolvent $\mathbf{Q}_{\mathbf{M}}$ is in its relation to the *empirical spectral measure* $\mu_{\mathbf{M}}$ of the matrix \mathbf{M} under study, through the associated *Stieltjes transform* $m_{\mu_{\mathbf{M}}}$, which we all define next.

Definition 2 (Empirical spectral measure). *For a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, the spectral measure or empirical spectral measure or empirical spectral distribution (e.s.d.) $\mu_{\mathbf{M}}$ of \mathbf{M} is defined as the normalized counting measure of the eigenvalues $\lambda_1(\mathbf{M}), \dots, \lambda_n(\mathbf{M})$ of \mathbf{M} ,*

$$\mu_{\mathbf{M}} \equiv \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i(\mathbf{M})}. \quad (2.2)$$

Since $\int \mu_{\mathbf{M}}(dx) = 1$, the spectral measure $\mu_{\mathbf{M}}$ of a matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ (random or not) is a probability measure. For (probability) measures, we can define their associated Stieltjes transforms as follows.

Definition 3 (Stieltjes transform). *For a real probability measure μ with support $\text{supp}(\mu)$, the Stieltjes transform $m_{\mu}(z)$ is defined, for all $z \in \mathbb{C} \setminus \text{supp}(\mu)$, as*

$$m_{\mu}(z) \equiv \int \frac{1}{t-z} \mu(dt). \quad (2.3)$$

This definition and the Stieltjes transform framework in effect extend beyond probability measures to σ -finite real measures (i.e., measures μ such that $\mu(\mathbb{R}) < \infty$), which will occasionally be discussed in this book.

The Stieltjes transform m_μ has numerous interesting properties: it is complex analytic on its domain of definition $\mathbb{C} \setminus \text{supp}(\mu)$, it is bounded $|m_\mu(z)| \leq 1/\text{dist}(z, \text{supp}(\mu))$, it satisfies $\Im[z] > 0 \Rightarrow \Im[m_\mu(z)] > 0$, and it is an increasing function on all connected components of its restriction to $\mathbb{R} \setminus \text{supp}(\mu)$ (since $m'_\mu(x) = \int (t-x)^{-2} \mu(dt) > 0$) with $\lim_{x \rightarrow \pm\infty} m_\mu(x) = 0$ if $\text{supp}(\mu)$ is bounded.

As a transform, m_μ admits an inverse formula to recover μ , as per the following result.

Theorem 2.1 (Inverse Stieltjes transform). *For a, b continuity points of the probability measure μ , we have*

$$\mu([a, b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im [m_\mu(x + iy)] dx. \tag{2.4}$$

Besides, if μ admits a density f at x (i.e., $\mu(x)$ is differentiable in a neighborhood of x and $\lim_{\epsilon \rightarrow 0} (2\epsilon)^{-1} \mu([x - \epsilon, x + \epsilon]) = f(x)$),

$$f(x) = \frac{1}{\pi} \lim_{y \downarrow 0} \Im [m_\mu(x + iy)]. \tag{2.5}$$

Also, if μ has an isolated mass at x , then

$$\mu(\{x\}) = \lim_{y \downarrow 0} -iy m_\mu(x + iy). \tag{2.6}$$

Proof. Since $|\frac{y}{(t-x)^2 + y^2}| \leq \frac{1}{y}$ for $y > 0$, by Fubini's theorem,

$$\begin{aligned} \frac{1}{\pi} \int_a^b \Im [m_\mu(x + iy)] dx &= \frac{1}{\pi} \int_a^b \left[\int \frac{y}{(t-x)^2 + y^2} \mu(dt) \right] dx \\ &= \frac{1}{\pi} \int \left[\int_a^b \frac{y}{(t-x)^2 + y^2} dx \right] \mu(dt) \\ &= \frac{1}{\pi} \int \left[\arctan \left(\frac{b-t}{y} \right) - \arctan \left(\frac{a-t}{y} \right) \right] \mu(dt). \end{aligned}$$

As $y \downarrow 0$, the difference in brackets converges either to $\pm\pi$ or 0 depending on the relative position of a, b , and t . By the dominated convergence theorem, the limit, as $y \downarrow 0$, is $\int 1_{[a, b]} \mu(dt) = \mu([a, b])$. When μ has an isolated mass at x , say $\mu(dt) = a\delta_x(t)$, we similarly have, again by dominated convergence (using, in particular, $|y(t-x)| \leq \frac{1}{2}(y^2 + (t-x)^2)$) that

$$\lim_{y \downarrow 0} -iy m_\mu(x + iy) = -\lim_{y \downarrow 0} \int \frac{iy(t-x)\mu(dt)}{(t-x)^2 + y^2} + \lim_{y \downarrow 0} \int \frac{y^2 \mu(dt)}{(t-x)^2 + y^2} = a.$$

This concludes the proof of Theorem 2.1. □

The important relation between the empirical spectral measure $\mu_{\mathbf{M}}$ of $\mathbf{M} \in \mathbb{R}^{n \times n}$, the Stieltjes transform $m_{\mu_{\mathbf{M}}}(z)$, and the resolvent $\mathbf{Q}_{\mathbf{M}}(z)$ lies in the fact that

$$m_{\mu_{\mathbf{M}}}(z) = \frac{1}{n} \sum_{i=1}^n \int \frac{\delta_{\lambda_i(\mathbf{M})}(t)}{t-z} = \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i(\mathbf{M}) - z} = \frac{1}{n} \text{tr} \mathbf{Q}_{\mathbf{M}}(z). \tag{2.7}$$

Combining inverse Stieltjes transform in Theorem 2.1 and the relation above thus provides a link between \mathbf{Q}_M and the eigenvalue distribution of \mathbf{M} . While seemingly contorted at first sight, this link turns out to be a very efficient way to study the spectral measure of *large-dimensional random matrices* \mathbf{M} .

In particular, note that Theorem 2.1 raises an interesting fact: The Stieltjes transform $m_\mu(z) = \int (t - z)^{-1} \mu(dt)$ is defined on all $\mathbb{C} \setminus \text{supp}(\mu)$, and as z approaches the support $\text{supp}(\mu)$, the integrand $(t - z)^{-1}$ becomes singular. Yet, this is precise when $x = \Re[z] \in \text{supp}(\mu)$ while $\Im[z] \downarrow 0$ that one can retrieve the density of μ at x from the Stieltjes transform $m_\mu(z)$. This observation is key to the analysis of the spectrum (both eigenvalues and eigenvectors) of (random) matrices: The singular points of the resolvent of a (random) matrix provide the information about its spectrum.

Remark 2.1 (Resolvent as a matrix-valued Stieltjes transform), *As proposed in Hachem et al. [2007], it is convenient to extrapolate Definition 3 of Stieltjes transforms to $n \times n$ matrix-valued positive measures $\mathbf{M}(dt)$,¹ in which case Equation (2.7) can be generalized as*

$$\mathbf{Q}_M(z) = \int \frac{\mathbf{M}(dt)}{t - z} = \mathbf{U} \text{diag} \left\{ \frac{1}{\lambda_i(\mathbf{M}) - z} \right\}_{i=1}^n \mathbf{U}^T,$$

where we used the spectral decomposition $\mathbf{M} = \mathbf{U} \text{diag}\{\lambda_i(\mathbf{M})\}_{i=1}^n \mathbf{U}^T$. This definition coincides with the former definition of the resolvent of \mathbf{M} . As such, the resolvent $\mathbf{Q}_M(z)$ is an “improved” matrix-valued Stieltjes transform, which enjoys similar properties as Stieltjes transforms on real-valued measures: it is complex analytic on its domain of definition, it is bounded $\|\mathbf{Q}_M(z)\| \leq 1/\text{dist}(z, \text{supp}(\mu_M))$, and $x \mapsto \mathbf{Q}_M(x)$ for $x \in \mathbb{R} \setminus \text{supp}(\mu_M)$ is an increasing matrix-valued function with respect to symmetric matrix partial ordering (i.e., $\mathbf{A} \succeq \mathbf{B}$ whenever $\mathbf{z}^T(\mathbf{A} - \mathbf{B})\mathbf{z} \geq 0$ for all \mathbf{z}).

2.1.3 Cauchy’s Integral, Linear Eigenvalue Functionals, and Eigenspaces

Being complex analytic, the resolvent $\mathbf{Q}_M(z)$ can be assessed using advanced tools from complex analysis. Of particular interest to this book is the relation between the resolvent and Cauchy’s integral theorem.

Theorem 2.2 (Cauchy’s integral formula). *For $\Gamma \subset \mathbb{C}$, a positively (i.e., counterclockwise) oriented simple closed curve and a complex function $f(z)$ analytic in a region containing Γ and its inside, then*

- (i) if $z_0 \in \mathbb{C}$ is enclosed by Γ , $f(z_0) = -\frac{1}{2\pi i} \oint_{\Gamma} \frac{f(z)}{z_0 - z} dz$;
- (ii) if not, $\frac{1}{2\pi i} \oint_{\Gamma} \frac{f(z)}{z_0 - z} dz = 0$.

This result provides an immediate connection between the so-called *linear functionals of the eigenvalues* (also referred to as the *linear spectral statistics* [Bai and

¹ Defined by the fact that $\mu(dt; \mathbf{z}) = \mathbf{z}^T \mathbf{M}(dt) \mathbf{z} = \sum_{i,j} [\mathbf{z}]_i [\mathbf{z}]_j [\mathbf{M}]_{ij}(dt)$ is a positive real-valued measure for all \mathbf{z} . See Rozanov [1967] for an introduction.

Silverstein, 2004] or *linear eigenvalue statistics* [Lytova and Pastur, 2009]) of \mathbf{M} and the Stieltjes transform $m_{\mu_{\mathbf{M}}}(z)$ through

$$\frac{1}{n} \sum_{i=1}^n f(\lambda_i(\mathbf{M})) = -\frac{1}{2\pi i n} \oint_{\Gamma} f(z) \operatorname{tr}(\mathbf{Q}_{\mathbf{M}}(z)) dz = -\frac{1}{2\pi i} \oint_{\Gamma} f(z) m_{\mu_{\mathbf{M}}}(z) dz,$$

for all f complex analytic in a compact neighborhood of $\operatorname{supp}(\mu_{\mathbf{M}})$, by choosing the contour Γ to enclose $\operatorname{supp}(\mu_{\mathbf{M}})$ (i.e., all the eigenvalues $\lambda_i(\mathbf{M})$). More generally,

$$\frac{1}{n} \sum_{\lambda_i(\mathbf{M}) \in \Gamma^\circ} f(\lambda_i(\mathbf{M})) = -\frac{1}{2\pi i} \oint_{\Gamma} f(z) m_{\mu_{\mathbf{M}}}(z) dz,$$

for Γ° the inside of the contour Γ . Note that in this case it is sufficient for f to be analytic in a neighborhood of $\operatorname{supp}(\mu_{\mathbf{M}}) \cap \Gamma^\circ$; in particular, if one wishes to count the number of eigenvalues in an interval $[a, b]$, one may use the formula for $f(t) = 1_{t \in [a-\varepsilon, b+\varepsilon]}$ for some $\varepsilon > 0$ small, which is of course not analytic on \mathbb{C} but is analytic on an open neighborhood of $[a, b]$.

Another quantity of interest relates to eigenvectors and eigenspaces. Considering the spectral decomposition $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ with $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ and $\mathbf{\Lambda} = \operatorname{diag}\{\lambda_1(\mathbf{M}), \dots, \lambda_n(\mathbf{M})\}$, we have

$$\mathbf{Q}_{\mathbf{M}}(z) = \sum_{i=1}^n \frac{\mathbf{u}_i \mathbf{u}_i^T}{\lambda_i(\mathbf{M}) - z}$$

and thus the direct access to the i th eigenvector \mathbf{u}_i of \mathbf{M} through

$$\mathbf{u}_i \mathbf{u}_i^T = -\frac{1}{2\pi i} \oint_{\Gamma_{\lambda_i(\mathbf{M})}} \mathbf{Q}_{\mathbf{M}}(z) dz,$$

for $\Gamma_{\lambda_i(\mathbf{M})}$ a contour circling around $\lambda_i(\mathbf{M})$ only. More generally,

$$\mathbf{U}f(\mathbf{\Lambda}; \Gamma)\mathbf{U}^T = -\frac{1}{2\pi i} \oint_{\Gamma} f(z) \mathbf{Q}_{\mathbf{M}}(z) dz,$$

for f analytic in a neighborhood of Γ and its inside Γ° and $f(\mathbf{\Lambda}; \Gamma) = \operatorname{diag}\{f(\lambda_i(\mathbf{M})) \cdot 1_{\lambda_i(\mathbf{M}) \in \Gamma^\circ}\}_{i=1}^n$.

Of specific interest to this book will be the projection of an individual eigenvector \mathbf{u}_i of \mathbf{M} onto a *deterministic* vector \mathbf{v} . In particular, from the above,

$$|\mathbf{v}^T \mathbf{u}_i|^2 = -\frac{1}{2\pi i} \oint_{\Gamma_{\lambda_i(\mathbf{M})}} \mathbf{v}^T \mathbf{Q}_{\mathbf{M}}(z) \mathbf{v} dz.$$

In the real case $\mathbf{M} \in \mathbb{R}^{n \times n}$, this gives access to $\mathbf{v}^T \mathbf{u}_i$, up to a sign (which at any rate is not fixed since both \mathbf{u}_i and $-\mathbf{u}_i$ are valid eigenvectors). The formula extends in the complex case by replacing the transpose $(\cdot)^T$ with a Hermitian transpose $(\cdot)^*$, and thus providing access to the complex number $\mathbf{v}^* \mathbf{u}_i$ up to a “phase” $e^{i\theta}$ for $\theta \in [0, 2\pi)$.

To summarize, the resolvent $\mathbf{Q}_{\mathbf{M}}$ provides access to *scalar observations* of the eigenspectrum of \mathbf{M} through its *linear functionals*, that is, the scalar observations $\frac{1}{n} \sum_i f(\lambda_i(\mathbf{M}))$ and $|\mathbf{v}^T \mathbf{u}_i|$ accessible from $\frac{1}{n} \operatorname{tr} \mathbf{Q}_{\mathbf{M}}$ and $\mathbf{v}^T \mathbf{Q}_{\mathbf{M}} \mathbf{v}$, respectively.

Before proceeding to the application of these results to random matrices, it is worth noticing at this point that working with the resolvent automatically enables many

powerful tools from *complex analysis*, the Cauchy integral formula being only one instance. Analytic functions, such as the Stieltjes transform and the resolvent, are “extremely smooth” objects, and enjoy a host of convenient properties. One such important property is, as already mentioned in Theorem 2.2, that it suffices to know an analytic function locally to know it globally.

Theorem 2.3 (Vitali’s convergence theorem [Titchmarsh, 1939]). *Let f_1, f_2, \dots be a sequence of functions, analytic on a region $D \subset \mathbb{C}$, such that $|f_n(z)| \leq M$ uniformly on n and $z \in D$. Further, assume that $f_n(z_j)$ converges for a countable set of points $z_1, z_2, \dots \in D$ having a limit point inside D . Then, $f_n(z)$ converges uniformly to a limit in any region bounded by a contour interior to D . This limit is furthermore an analytic function of z .*

Vitali’s convergence theorem will be heavily exploited to study the behavior of resolvents $\mathbf{Q}_M(z)$ near the real axis (where it is almost singular but of utmost interest) by instead studying its properties away from the real axis (where it is mathematically more convenient). The theorem is in fact doubly interesting as it states that the knowledge of f_n at a countable number of points z_1, z_2, \dots is sufficient to fully characterize the limit f ; as we shall see later, this property will be used to prove the convergence of functionals $f_n(z) = g(\mathbf{Q}_M(z) - \bar{\mathbf{Q}}(z)) \rightarrow 0$ of random resolvents $\mathbf{Q}_M(z)$ to deterministic equivalents $\bar{\mathbf{Q}}(z)$ (here n is the growing size of the resolvents): if $f_n(z_j) \rightarrow 0$ almost surely for each z_1, z_2, \dots , then by the countable union of probability one events, $f_n(z_j) \rightarrow 0$ with probability one uniformly on the set $\{z_1, z_2, \dots\}$, and by Vitali we obtain that $f_n(z) \rightarrow 0$ with probability one uniformly on a (possibly very large) subset of \mathbb{C} .

2.1.4 Deterministic and Random Equivalents

This book is concerned with the situation, where \mathbf{M} is a *large-dimensional random matrix*, the eigenvalues and eigenvectors of which need be related to the statistical nature of the model design of \mathbf{M} .

In the early days of random matrix theory, the main focus was on the *limiting spectral measure* of $\mathbf{M} \in \mathbb{R}^{n \times n}$, that is, the characterization of a certain “limit” to the spectral measure μ_M of \mathbf{M} as the size of \mathbf{M} increases. For this purpose, a natural approach is to study the *random Stieltjes transform* $m_{\mu_M}(z)$ and to show that it admits a limit (in probability or almost surely) $m(z)$ as $n \rightarrow \infty$. However, this method has strong limitations: (i) it supposes that such a limit exists, therefore restricting the study to very regular models for \mathbf{M} and (ii) it only quantifies the Stieltjes transform $\frac{1}{n} \text{tr} \mathbf{Q}_M$, thereby discarding all subspace information about \mathbf{M} carried in the resolvent matrix \mathbf{Q}_M . As a consequence, a further study of the eigenvectors of \mathbf{M} often requires a complete rework.

To avoid these limitations, modern random matrix theory focuses instead on the notion of *deterministic equivalents*, which are deterministic matrices – thus finite dimensional objects rather than limits – having (in probability or almost surely)

asymptotically the same *scalar observations* as the random ones.² In particular, these scalar observations of deterministic equivalents (e.g., their normalized traces or their bilinear forms) need *not* themselves admit a limit as the matrix dimension n grows: What only matters is that they deterministically “track” the behavior of their random counterparts with increased accuracy as n grows large to infinity.

Definition 4 (Deterministic Equivalent). *We say that $\bar{\mathbf{Q}} \in \mathbb{R}^{n \times n}$ is a deterministic equivalent for the symmetric random matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ if, for (sequences of) deterministic matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ and vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ of unit norms (operator and Euclidean, respectively), we have, as $n \rightarrow \infty$,*

$$\frac{1}{n} \operatorname{tr} \mathbf{A}(\mathbf{Q} - \bar{\mathbf{Q}}) \rightarrow 0, \quad \mathbf{a}^\top (\mathbf{Q} - \bar{\mathbf{Q}}) \mathbf{b} \rightarrow 0,$$

where the convergence is either in probability or almost sure.

This definition³ has the advantage of bringing forth the two key elements that provide access to the spectral information of a random matrix \mathbf{M} : traces and bilinear forms (of its resolvent $\mathbf{Q}_{\mathbf{M}}(z)$ for some z). Deterministic equivalents for the resolvent $\mathbf{Q}_{\mathbf{M}}$ thus encode the necessary information to statistically quantify, at least spectrally, the random matrix \mathbf{M} .

A first and natural use of deterministic equivalents is to establish that, for a random matrix \mathbf{M} of interest, $\frac{1}{n} \operatorname{tr}(\mathbf{Q}_{\mathbf{M}}(z) - \bar{\mathbf{Q}}(z)) \rightarrow 0$, say almost surely, for all $z \in \mathcal{C}$ with $\mathcal{C} \subset \mathbb{C}$. Denoting $\bar{m}_n(z) = \frac{1}{n} \operatorname{tr} \bar{\mathbf{Q}}(z)$, this convergence implies that the Stieltjes transform of $\mu_{\mathbf{M}}$ “converges” in the sense that $m_{\mu_{\mathbf{M}}}(z) - \bar{m}_n(z) \rightarrow 0$. As we will see, this indicates that the spectral measure $\mu_{\mathbf{M}}$ gets increasingly well approximated, as n grows large, by a probability measure $\bar{\mu}_n$ having Stieltjes transform $\bar{m}_n(z)$. Identifying $\bar{m}_n(z)$, which uniquely defines $\bar{\mu}_n$ as per Theorem 2.1, will often be as far as *the Stieltjes transform method* will lead us. But in some rare cases (such as the Marčenko–Pastur and the semicircle laws), $\bar{\mu}_n$ will be explicitly identifiable.

In the remainder of the book, we will often characterize the large-dimensional (spectral) behavior of random matrix models \mathbf{M} through the “approximation” offered by the deterministic equivalents $\bar{\mathbf{Q}}(z)$ of their associated resolvents $\mathbf{Q}_{\mathbf{M}}(z)$, providing simultaneously access to their asymptotic spectral measures as well as to their eigenspaces. We will therefore extrapolate some of the core traditional results in random matrix theory, such as the Marčenko–Pastur law [Marčenko and Pastur, 1967],

² The wide spread of deterministic equivalents in the random matrix literature arose from application needs, primarily in signal processing and wireless communications, involving too structured matrix models for limiting eigenvalue distributions to be meaningful [Hachem et al., 2007, Couillet et al., 2011]. Yet, deterministic equivalents in fact originate from the (much earlier) works of Girko [2001]. They have recently been included as a new feature of free probability theory [Speicher and Vargas, 2012], an alternative approach to the resolvent method, which will be shortly discussed in Section 2.6.2.

³ The notion of “deterministic equivalent” has not been formally defined in the literature. The present definition is thus restricted to this book and is for the convenience of presentation. Section 2.7 will provide an alternative, possibly more satisfying, definition through the notion of *linear concentration* (Definition 8).

the sample covariance matrix model [Silverstein and Bai, 1995], etc., under this more general form of deterministic equivalents.

Remark 2.2 ($\bar{\mathbf{Q}}$ versus $\mathbb{E}[\mathbf{Q}]$). For $\bar{\mathbf{Q}}$ a deterministic equivalent for \mathbf{Q} , the (probabilistic) convergences $\frac{1}{n} \text{tr} \mathbf{A}(\mathbf{Q} - \bar{\mathbf{Q}}) \rightarrow 0$ and $\mathbf{a}^\top (\mathbf{Q} - \bar{\mathbf{Q}}) \mathbf{b} \rightarrow 0$ generally unfold from the deterministic relation that

$$\|\mathbb{E}[\mathbf{Q}] - \bar{\mathbf{Q}}\| \rightarrow 0,$$

and from a control of the variance of $\frac{1}{n} \text{tr}(\mathbf{A}\mathbf{Q})$ and $\mathbf{a}^\top \mathbf{Q} \mathbf{b}$; this will often be the strategy followed in our proofs. Note particularly that if the above relation is met, then $\mathbb{E}[\mathbf{Q}]$ itself is, by Definition 4, a deterministic equivalent for the random \mathbf{Q} . However, $\mathbb{E}[\mathbf{Q}]$ is often not convenient to work with and a “truly deterministic” matrix $\bar{\mathbf{Q}}$ involving no integration over probability spaces (and that can be numerically evaluated with ease) will be systematically preferred.

Deterministic equivalents will be used very regularly in the course of this book. To avoid heavy notations, particularly in the main theorems and their proofs, we will use the following shortcut notations, valid both for deterministic and random matrix equivalents.

Notation 1 (Matrix Equivalents). For $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times n}$ two random or deterministic matrices, we write

$$\mathbf{X} \leftrightarrow \mathbf{Y},$$

if, for all $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ of unit norms (respectively, operator and Euclidean), we have the simultaneous results

$$\frac{1}{n} \text{tr} \mathbf{A}(\mathbf{X} - \mathbf{Y}) \rightarrow 0, \quad \mathbf{a}^\top (\mathbf{X} - \mathbf{Y}) \mathbf{b} \rightarrow 0, \quad \|\mathbb{E}[\mathbf{X} - \mathbf{Y}]\| \rightarrow 0,$$

where, for random quantities, the convergence is either in probability or almost sure.

In many situations, deterministic equivalents \mathbf{Y} for a random matrix \mathbf{X} may not be directly accessible with classical random matrix techniques. In these cases, the introduction of an intermediary random matrix $\tilde{\mathbf{X}}$ satisfying $\|\tilde{\mathbf{X}} - \mathbf{X}\| \xrightarrow{\text{a.s.}} 0$ will help “propagate” the deterministic equivalent relations. Indeed, if $\tilde{\mathbf{X}} \leftrightarrow \mathbf{Y}$, then one necessarily has $\mathbf{X} \leftrightarrow \mathbf{Y}$. When the convergence $\|\tilde{\mathbf{X}} - \mathbf{X}\| \xrightarrow{\text{a.s.}} 0$ is too demanding, it may of course be sufficient in some cases to prove that $\mathbf{X} \leftrightarrow \tilde{\mathbf{X}}$ (in which case both matrices are random) to ensure that $\mathbf{X} \leftrightarrow \mathbf{Y}$. This justifies the need to apply the notation “ \leftrightarrow ” to arbitrary, random or deterministic, matrices.

2.2 Foundational Random Matrix Results

In this section, we introduce the main historical results of random matrix theory (appropriately updated under a deterministic equivalent form), which will serve as

supporting materials to most machine learning applications covered in this book.⁴ For readability and accessibility to the readers new to random matrix theory, we mostly stick to intuitive and short sketches of proofs. Yet, for the readers to have a glimpse on the technical details and modern tools of the field, some of the proof sketches will be appended with a complete and exhaustive proof.

Both sketches and detailed proofs rely on a set of elementary lemmas and identities, which need be introduced to understand their spirits and cornerstone arguments. This is done in Section 2.2.1. The detailed proofs differ from the sketches in having additional technical *probability* theory arguments to prove various convergence results. These arguments strongly depend on the underlying random matrix model hypotheses (Gaussian independent, i.i.d., concentrated random vectors, etc.); for readability, we will focus in our proofs on one specific line of arguments (that we claim to be the “historical” one) and will discuss alternative techniques in side remarks. In particular, the specific *concentration of measure* theoretic approach, which is both more “modern” (yet less mature) and more adapted to machine learning applications, will be given a separate treatment in Section 2.7.

2.2.1 Key Lemmas and Identities

Resolvent Identities

Most results discussed in this section consist in tools meant to help “approximate” random matrix resolvents $\mathbf{Q}(z)$ via deterministic resolvents $\bar{\mathbf{Q}}(z)$ in the sense of Definition 4. The following first identity provides a comparison of matrix inverses and is often used to compare the aforementioned resolvents.

Lemma 2.1 (Resolvent identity). *For invertible matrices \mathbf{A} and \mathbf{B} , we have*

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}.$$

Proof. This can be easily checked by multiplying both sides on the left by \mathbf{A} and on the right by \mathbf{B} . \square

Another useful lemma that helps directly connect the resolvent of \mathbf{BA} to that of \mathbf{AB} is given as follows:

Lemma 2.2. *For $\mathbf{A} \in \mathbb{R}^{p \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$, we have*

$$\mathbf{A}(\mathbf{BA} - z\mathbf{I}_n)^{-1} = (\mathbf{AB} - z\mathbf{I}_p)^{-1}\mathbf{A},$$

for $z \in \mathbb{C}$ distinct from 0 and from the eigenvalues of \mathbf{AB} .

Proof. Left-multiply both ends of the equality by $\mathbf{AB} - z\mathbf{I}_p$ to obtain $\mathbf{A} = \mathbf{A}$. \square

⁴ Although historically and technically, said “Wigner” models of symmetric random matrices with independent entries came first, are mathematically more accessible and have thus spurred more research efforts [Wigner, 1955, Mehta and Gaudin, 1960, Anderson et al., 2010], for the sake of machine learning applications, our focus is primarily on the slightly more involved sample covariance matrix models [Marčenko and Pastur, 1967, Bai and Silverstein, 2010].

For \mathbf{AB} and \mathbf{BA} symmetric, Lemma 2.2 is a special case of the more general relation $\mathbf{A} \cdot f(\mathbf{BA}) = f(\mathbf{AB}) \cdot \mathbf{A}$, with $f(\mathbf{M}) \equiv \mathbf{U}f(\Lambda)\mathbf{U}^T$ under the spectral decomposition $\mathbf{M} = \mathbf{U}\Lambda\mathbf{U}^T$ and f complex analytic. Since f is analytic, $f(\mathbf{BA}) = \sum_{i=0}^{\infty} c_i (\mathbf{BA})^i$ for some sequence $\{c_i\}_{i=0}^{\infty}$ and thus $\mathbf{A} \cdot f(\mathbf{BA}) = \sum_{i=0}^{\infty} c_i (\mathbf{AB})^i \cdot \mathbf{A} = f(\mathbf{AB}) \cdot \mathbf{A}$.

The next lemma, known as *Sylvester's identity* (also known as the Weinstein–Aronszajn identity), similarly relates the resolvents of \mathbf{AB} and \mathbf{BA} through their determinant.

Lemma 2.3 (Sylvester's identity). *For $\mathbf{A} \in \mathbb{R}^{p \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$ and $z \in \mathbb{C} \setminus \{0\}$,*

$$\det(\mathbf{AB} - z\mathbf{I}_p) = \det(\mathbf{BA} - z\mathbf{I}_n) (-z)^{p-n}.$$

Proof. It suffices to develop the block-matrix determinant (recall that $\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det \mathbf{D} \cdot \det(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}) = \det \mathbf{A} \cdot \det(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})$ when \mathbf{A}, \mathbf{D} are invertible)

$$\det \begin{pmatrix} z\mathbf{I}_p & \mathbf{zA} \\ \mathbf{B} & z\mathbf{I}_n \end{pmatrix} = \det(z\mathbf{I}_p) \cdot \det(z\mathbf{I}_n - \mathbf{BA}) = \det(z\mathbf{I}_n) \cdot \det(z\mathbf{I}_p - \mathbf{AB}).$$

□

An immediate consequence of Sylvester's identity is that \mathbf{AB} and \mathbf{BA} have the same *nonzero* eigenvalues (those nonzero z s for which both left- and right-hand sides vanish). Thus, say for $n \geq p$, $\mathbf{AB} \in \mathbb{R}^{p \times p}$ and $\mathbf{BA} \in \mathbb{R}^{n \times n}$ have the same spectrum, except for the additional $n - p$ zero eigenvalues of \mathbf{BA} . This remark implies the next identity.

Lemma 2.4 (Trace of resolvent and co-resolvent). *Let $\mathbf{A} \in \mathbb{R}^{p \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$, and $z \in \mathbb{C}$ not an eigenvalue of \mathbf{AB} nor zero. Then,*

$$\text{tr} \mathbf{Q}_{\mathbf{AB}}(z) = \text{tr} \mathbf{Q}_{\mathbf{BA}}(z) + \frac{n-p}{z}.$$

In particular, if \mathbf{AB} and \mathbf{BA} are symmetric,

$$m_{\mu_{\mathbf{AB}}}(z) = \frac{n}{p} m_{\mu_{\mathbf{BA}}}(z) + \frac{n-p}{pz},$$

for $\mu_{\mathbf{AB}}$ the empirical spectral measure of \mathbf{AB} defined in Definition 2.

It will be customary, if $\mathbf{Q}_{\mathbf{AB}}$ is the resolvent of the matrix \mathbf{AB} under study, to call $\mathbf{Q}_{\mathbf{BA}}$ the *co-resolvent* of \mathbf{AB} . We will see that the resolvent and co-resolvent of random matrix models (in particular, the resolvent and co-resolvent of \mathbf{XX}^T for \mathbf{X} some structured random matrix) often intervene together, and quite symmetrically, to define their associated deterministic equivalents.

Perturbation Identities

Quantifying the asymptotic global (e.g., spectral distribution) or local (e.g., isolated eigenvalues or projection on eigenvector) behavior of random matrices \mathbf{M} will systematically involve a *perturbation approach*. The idea often lies in comparing the behavior of the resolvent $\mathbf{Q} = \mathbf{Q}_{\mathbf{M}}$ to the resolvent \mathbf{Q}_{-i} of \mathbf{M}_{-i} , with \mathbf{M}_{-i} defined as \mathbf{M} with either i th row and/or column, or some i th contribution (e.g., $\mathbf{M}_{-i} = \sum_{j \neq i} \mathbf{x}_j \mathbf{x}_j^T$

if $\mathbf{M} = \sum_j \mathbf{x}_j \mathbf{x}_j^\top$), discarded. A number of so-called *perturbation* identities are then needed.

The first one involves the segmentation of \mathbf{M} under the form of subblocks, in general consisting of one large block and three small ones. The corresponding resolvent $\mathbf{Q}_\mathbf{M}$ can correspondingly be segmented in subblocks according to the following block inversion lemma.

Lemma 2.5 (Block matrix inversion). *For $\mathbf{A} \in \mathbb{R}^{p \times p}$, $\mathbf{B} \in \mathbb{R}^{p \times n}$, $\mathbf{C} \in \mathbb{R}^{n \times p}$ and $\mathbf{D} \in \mathbb{R}^{n \times n}$ with \mathbf{D} invertible, we have*

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{S}^{-1} & -\mathbf{S}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{S}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{S}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix},$$

where $\mathbf{S} \equiv \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is the Schur complement (for the block \mathbf{D}) of $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}$.⁵

As a consequence of Lemma 2.5, we get the following explicit form for all diagonal entries of an invertible matrix \mathbf{A} .

Lemma 2.6 (Diagonal entries of matrix inverse). *For invertible $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\mathbf{A}_{-i} \in \mathbb{R}^{(p-1) \times (p-1)}$ the matrix obtained by removing the i th row and column from \mathbf{A} ($i \in \{1, \dots, p\}$), we have*

$$[\mathbf{A}^{-1}]_{ii} = \frac{1}{[\mathbf{A}]_{ii} - \mathbf{A}_{i,-i}(\mathbf{A}_{-i})^{-1}\mathbf{A}_{-i,i}},$$

for $\mathbf{A}_{i,-i}, \mathbf{A}_{-i,i} \in \mathbb{R}^{p-1}$ the i th row and column of \mathbf{A} with i th entries removed, respectively.

The result follows from Lemma 2.5 for entry $(1, 1)$ and can then be generalized to an arbitrary diagonal entry (i, i) by pre- and post-multiplying by the permutation matrix \mathbf{P} which exchanges the first and the i th row and column. Alternatively, the result may be obtained from the fact that $\mathbf{A}^{-1} = \frac{\text{adj}(\mathbf{A})}{\det(\mathbf{A})}$, with $\text{adj}(\mathbf{A})$ the adjugate matrix of \mathbf{A} , together with the block determinant formula.

Perturbations by the addition or subtraction of low-rank matrices to \mathbf{M} induce modifications in its resolvent $\mathbf{Q}_\mathbf{M}$ that involve the following Woodbury identity.

Lemma 2.7 (Woodbury). *For $\mathbf{A} \in \mathbb{R}^{p \times p}$, $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{p \times n}$, such that both \mathbf{A} and $\mathbf{A} + \mathbf{U}\mathbf{V}^\top$ are invertible, we have*

$$(\mathbf{A} + \mathbf{U}\mathbf{V}^\top)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I}_n + \mathbf{V}^\top\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^\top\mathbf{A}^{-1}.$$

Note importantly that, while $(\mathbf{A} + \mathbf{U}\mathbf{V}^\top)^{-1}$ is of size $p \times p$, $\mathbf{I}_n + \mathbf{V}^\top\mathbf{A}^{-1}\mathbf{U}$ is of size $n \times n$. This will turn out useful, for $n \ll p$, to relate resolvents of large-dimensional matrices to resolvents of more elementary and small-size matrices. In particular, for $n = 1$, that is, $\mathbf{U}\mathbf{V}^\top = \mathbf{u}\mathbf{v}^\top$ for $\mathbf{U} = \mathbf{u} \in \mathbb{R}^p$ and $\mathbf{V} = \mathbf{v} \in \mathbb{R}^p$, the above identity specializes to the Sherman–Morrison formula.

⁵ The Schur complement $\mathbf{S} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is particularly known for its providing the block determinant formula $\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det(\mathbf{D})\det(\mathbf{S})$, already exploited in the proof of Sylvester’s identity, Lemma 2.3.

Lemma 2.8 (Sherman–Morrison). For $\mathbf{A} \in \mathbb{R}^{p \times p}$ invertible and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$, $\mathbf{A} + \mathbf{u}\mathbf{v}^\top$ is invertible if and only if $1 + \mathbf{v}^\top \mathbf{A}^{-1} \mathbf{u} \neq 0$ and

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^\top)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{u}\mathbf{v}^\top \mathbf{A}^{-1}}{1 + \mathbf{v}^\top \mathbf{A}^{-1} \mathbf{u}}.$$

Besides,

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^\top)^{-1} \mathbf{u} = \frac{\mathbf{A}^{-1} \mathbf{u}}{1 + \mathbf{v}^\top \mathbf{A}^{-1} \mathbf{u}}.$$

Letting $\mathbf{A} = \mathbf{M} - z\mathbf{I}_p$, $z \in \mathbb{C}$, and $\mathbf{v} = \tau \mathbf{u}$ for $\tau \in \mathbb{R}$ in the previous lemma leads to the following rank-one perturbation lemma for the resolvent of \mathbf{M} .

Lemma 2.9 (Silverstein and Bai [1995, Lemma 2.6]). For $\mathbf{A}, \mathbf{M} \in \mathbb{R}^{p \times p}$ symmetric, $\mathbf{u} \in \mathbb{R}^p$, $\tau \in \mathbb{R}$ and $z \in \mathbb{C} \setminus \mathbb{R}$,

$$\left| \text{tr} \mathbf{A}(\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top - z\mathbf{I}_p)^{-1} - \text{tr} \mathbf{A}(\mathbf{M} - z\mathbf{I}_p)^{-1} \right| \leq \frac{\|\mathbf{A}\|}{|\Im(z)|}.$$

Also, for $\mathbf{A}, \mathbf{M} \in \mathbb{R}^{p \times p}$ symmetric and nonnegative definite, $\mathbf{u} \in \mathbb{R}^p$, $\tau > 0$ and $z < 0$,⁶

$$\left| \text{tr} \mathbf{A}(\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top - z\mathbf{I}_p)^{-1} - \text{tr} \mathbf{A}(\mathbf{M} - z\mathbf{I}_p)^{-1} \right| \leq \frac{\|\mathbf{A}\|}{|z|}.$$

It is interesting (and possibly counterintuitive at first) to note that the norm $\|\mathbf{u}\|$ and the value τ do not intervene in the above inequality. In particular, irrespective of the amplitude of the rank-one perturbation $\tau \mathbf{u}\mathbf{u}^\top$, under the conditions of the lemma,

$$m_{\mu_{\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top}}(z) = m_{\mu_{\mathbf{M}}}(z) + O(p^{-1}),$$

and thus, by the link between spectrum and Stieltjes transform, the spectral measure of \mathbf{M} is asymptotically close to that of $\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top$ for any \mathbf{u} and τ , in the large p limit. This result can be understood through the following two arguments:

- (i) for large p , the spectrum of \mathbf{M} (say $\|\mathbf{M}\| = O(1)$ without loss of generality) is only nontrivial if the vast majority of the p eigenvalues of \mathbf{M} are of order $O(1)$: Thus, as p eigenvalues use a space of size $O(1)$, they tend to aggregate;
- (ii) by Weyl’s interlacing lemma presented next (Lemma 2.10) for symmetric matrices, the eigenvalues of \mathbf{M} and of $\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top$ are interlaced (i.e., $\dots \leq \lambda_i(\mathbf{M}) \leq \lambda_i(\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top) \leq \lambda_{i+1}(\mathbf{M}) \leq \dots$).

Together, Arguments (i) and (ii) thus indicate that the $\lambda_i(\mathbf{M})$ s and $\lambda_i(\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top)$ s are asymptotically the same, at the possible exception of *rightmost* eigenvalue $\lambda_p(\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top)$, which is free to be found away from $\lambda_p(\mathbf{M})$. The rank-one perturbation $\tau \mathbf{u}\mathbf{u}^\top$ of \mathbf{M} thus does not asymptotically affect the (limiting) spectral measure (the possible presence of an outlying eigenvalue having no effect on the normalized counting measure). In passing, this remark unveils the important fact that, by definition, the spectral measure, as well as its Stieltjes transform, is only able to capture the “bulk” behavior

⁶ Exercise 4 in Section 2.9 proposes a partial proof of Lemma 2.9 for the case $z < 0$.

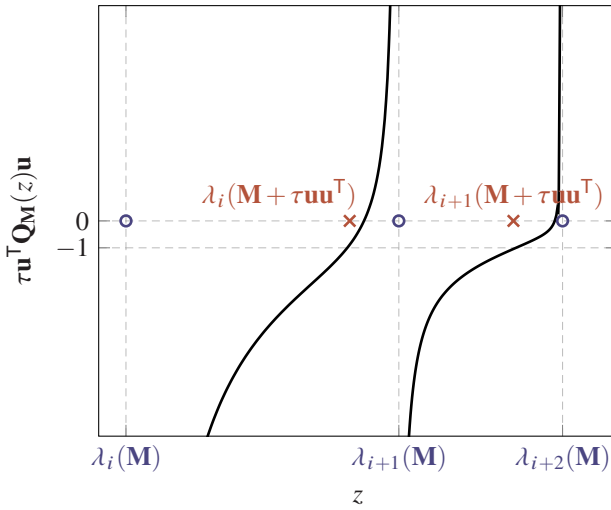


Figure 2.1 Illustration of the eigenvalues of \mathbf{M} and the rank-one perturbation $\mathbf{M} + \tau \mathbf{u} \mathbf{u}^T$ of \mathbf{M} , as well as the function $\tau \mathbf{u}^T \mathbf{Q}_M(z) \mathbf{u}$. Code on web: MATLAB and Python.

of the eigenvalues and *not* the behavior of individual eigenvalues. We will come back to this point in more detail in Section 2.5.

Unlike nonsymmetric matrices, symmetric matrices enjoy the nice property of having “stable” spectra with respect to rank-one perturbations. For $z \in \mathbb{R}$, an eigenvalue of $\mathbf{M} + \tau \mathbf{u} \mathbf{u}^T$ but not of \mathbf{M} with, say, $\tau > 0$, we have

$$\begin{aligned} 0 &= \det(\mathbf{M} + \tau \mathbf{u} \mathbf{u}^T - z \mathbf{I}_p) = \det(\mathbf{Q}_M^{-1}(z)) \cdot \det(\mathbf{I}_p + \tau \mathbf{Q}_M(z) \mathbf{u} \mathbf{u}^T) \\ &= \det(\mathbf{Q}_M^{-1}(z)) \cdot (1 + \tau \mathbf{u}^T \mathbf{Q}_M(z) \mathbf{u}), \end{aligned}$$

where the second equality unfolds from factoring out $\mathbf{Q}_M^{-1}(z) = \mathbf{M} - z \mathbf{I}_p$ (which is not singular as z is not an eigenvalue of \mathbf{M}) and the third from Sylvester’s identity, Lemma 2.3. As a consequence, z is one of the solutions to

$$-1 = \tau \mathbf{u}^T \mathbf{Q}_M(z) \mathbf{u} = \tau \sum_{i=1}^p \frac{|\mathbf{v}_i^T \mathbf{u}|^2}{\lambda_i(\mathbf{M}) - z}, \quad \text{with } \mathbf{M} = \sum_{i=1}^p \lambda_i(\mathbf{M}) \mathbf{v}_i \mathbf{v}_i^T,$$

which, seen as a function of z , has asymptotes at each $\lambda_i(\mathbf{M})$ and is increasing (from $-\infty$ to ∞) on the segments $(\lambda_i(\mathbf{M}), \lambda_{i+1}(\mathbf{M}))$ (eigenvalues being sorted in increasing order). The eigenvalues of $\mathbf{M} + \tau \mathbf{u} \mathbf{u}^T$ are therefore *interlaced* with those of \mathbf{M} , see Figure 2.1 for an illustration. This idea generalizes to generic low-rank perturbation in the following lemma.

Lemma 2.10 (Weyl’s inequality, [Horn and Johnson, 2012, Theorem 4.3.1]). *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{p \times p}$ be symmetric matrices and the respective eigenvalues of \mathbf{A} , \mathbf{B} and $\mathbf{A} + \mathbf{B}$*

be arranged in nondecreasing order, that is, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{p-1} \leq \lambda_p$. Then, for all $i \in \{1, \dots, p\}$,

$$\begin{aligned} \lambda_i(\mathbf{A} + \mathbf{B}) &\leq \lambda_{i+j}(\mathbf{A}) + \lambda_{p-j}(\mathbf{B}), \quad j = 0, 1, \dots, p - i, \\ \lambda_{i-j+1}(\mathbf{A}) + \lambda_j(\mathbf{B}) &\leq \lambda_i(\mathbf{A} + \mathbf{B}), \quad j = 1, \dots, i, \end{aligned}$$

In particular, taking $i = 1$ in the first equation and $i = p$ in the second equation, together with the fact $\lambda_j(\mathbf{B}) = -\lambda_{p+1-j}(-\mathbf{B})$ for $j = 1, \dots, p$, implies

$$\max_{1 \leq j \leq p} |\lambda_j(\mathbf{A}) - \lambda_j(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|.$$

This last implication is fundamental as it shows that the difference in operator norm $\|\mathbf{A} - \mathbf{B}\|$ controls (uniformly) the pairwise distance of the corresponding eigenvalues $|\lambda_j(\mathbf{A}) - \lambda_j(\mathbf{B})|$. Since $\|\mathbf{A} - \mathbf{B}\| \leq \|\mathbf{A} - \mathbf{B}\|_F$, the same holds for the (numerically simpler) Frobenius norm; however, it is in general too demanding to control the matrix differences in Frobenius norm which, as a result, is less used in practice (in particular, for most random matrix models $\mathbf{X} \in \mathbb{R}^{p \times p}$ considered in this book $\|\mathbf{X}\|_F$ is in general $O(\sqrt{p})$ larger than $\|\mathbf{X}\|$).⁷

Probability Identities

The results in the previous sections are algebraic identities useful to handle the resolvent \mathbf{Q}_M of the deterministic matrix \mathbf{M} . The second ingredient of random matrix analysis lies in (asymptotic) probability approximations as the dimensions of \mathbf{M} increase. Quite surprisingly, most results essentially revolve around the convergence of a certain quadratic form, which is often nothing more than a mere extension of the law of large numbers.

Those quadratic form convergence results come under multiple forms. The historical form, due to Bai and Silverstein, sometimes referred to as the “trace lemma,” is as follows.

Lemma 2.11 (Quadratic-form-close-to-the-trace, trace lemma, [Bai and Silverstein, 2010, Lemma B.26]). *Let $\mathbf{x} \in \mathbb{R}^p$ have independent entries x_i of zero mean, unit variance, and $\mathbb{E}[|x_i|^K] \leq \nu_K$ for some $K \geq 1$. Then for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $k \geq 1$,*

$$\mathbb{E} \left[\left| \mathbf{x}^\top \mathbf{A} \mathbf{x} - \text{tr} \mathbf{A} \right|^k \right] \leq C_k \left[\left(\nu_4 \text{tr}(\mathbf{A} \mathbf{A}^\top) \right)^{k/2} + \nu_{2k} \text{tr}(\mathbf{A} \mathbf{A}^\top)^{k/2} \right],$$

for some constant $C_k > 0$ independent of p . In particular, if $\|\mathbf{A}\| \leq 1$ and the entries of \mathbf{x} have bounded eighth-order moment,

⁷ This being said, the inequality $\|\mathbf{X}\| \leq (\text{tr}(\mathbf{X}\mathbf{X}^\top)^k)^{1/(2k)}$, which coincides with $\|\mathbf{X}\| \leq \|\mathbf{X}\|_F$ for $k = 1$ and becomes an equality in the $k \rightarrow \infty$ limit, is sometimes used (however with $k \geq 2$) to control the operator norm $\|\mathbf{X}\|$. Nonetheless, the approach is often quite cumbersome as it quickly becomes a heavy combinatorial calculus for not too small k .

$$\mathbb{E} \left[\left(\mathbf{x}^\top \mathbf{A} \mathbf{x} - \text{tr} \mathbf{A} \right)^4 \right] \leq Cp^2,$$

for some $C > 0$ independent of p , and consequently, as $p \rightarrow \infty$,

$$\frac{1}{p} \mathbf{x}^\top \mathbf{A} \mathbf{x} - \frac{1}{p} \text{tr} \mathbf{A} \xrightarrow{\text{a.s.}} 0.$$

This last result is rather intuitive. For $\mathbf{A} = \mathbf{I}_p$, this is simply an instance of the (strong) law of large numbers. For generic \mathbf{A} , first note that, by the independence of the entries of \mathbf{x} , $\mathbb{E}[\mathbf{x}^\top \mathbf{A} \mathbf{x}] = \text{tr} \mathbf{A}$. Exploiting the fact that $\text{Var}[\mathbf{x}^\top \mathbf{A} \mathbf{x}/p] = O(p^{-1})$ we have that $\mathbf{x}^\top \mathbf{A} \mathbf{x}/p - \text{tr} \mathbf{A}/p \rightarrow 0$, but only in probability; since the variance calculus involves exponentiating the entries x_i of \mathbf{x} up to power 4, they need to be of finite fourth moment. The almost sure convergence is achieved by showing the faster moment convergence $\mathbb{E}[(\mathbf{x}^\top \mathbf{A} \mathbf{x}/p - \text{tr} \mathbf{A}/p)^4] = O(p^{-2})$, which is the second statement of the lemma and requires eighth-order exponentiation of the x_i s. The request for \mathbf{A} to be of bounded norm with respect to p in this case “stabilizes” the quadratic form $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ by maintaining its concentration properties.

Recalling from Remark 2.1 that $\|\mathbf{Q}_\mathbf{M}(z)\| \leq 1/\text{dist}(z, \text{supp}(\mu_\mathbf{M}))$, Lemma 2.11 can be exploited for $\mathbf{A} = \mathbf{Q}_\mathbf{M}(z)$ for all z away from the support of $\mu_\mathbf{M}$ and all \mathbf{x} independent of $\mathbf{Q}_\mathbf{M}(z)$. The core of the proofs of the main random matrix results is essentially based on this last remark.

The quadratic-form-close-to-the-trace lemma is fundamental to already obtain *heuristics* on the main random matrix identities, using $\frac{1}{p} \mathbf{x}^\top \mathbf{A} \mathbf{x} \simeq \frac{1}{p} \text{tr} \mathbf{A}$ for \mathbf{x} independent of \mathbf{A} with independent zero-mean unit-variance entries. In the rigorous proof of many random matrix results presented in this book, the lemma allows for a careful control on the fluctuations of $\frac{1}{p} \mathbf{x}^\top \mathbf{A} \mathbf{x}$ for deterministic \mathbf{A} (or, conditioned on \mathbf{A}). However, \mathbf{A} may itself be random (as when $\mathbf{A} = \mathbf{Q}_\mathbf{M}(z)$ the resolvent of a random matrix \mathbf{M}). In this case, as a second step, the fluctuations of $\frac{1}{p} \text{tr} \mathbf{A}$ will also need be controlled. The difficulty here, especially when \mathbf{A} takes the form of an inverse matrix $\mathbf{A} = \mathbf{Q}_\mathbf{M}(z)$, is to exploit the independence in the, say, columns of \mathbf{M} nested *inside* the matrix inverse (or other more elaborate function of the random matrix \mathbf{M}). This can be elegantly and universally dealt with using Burkholder inequality: denoting $\mathbb{E}_i[\mathbf{M}]$ the expectation of the random matrix \mathbf{M} conditioned on its first (or last) i columns, the sequence $\{(\mathbb{E}_i - \mathbb{E}_{i-1})[\mathbf{M}]\}_{i=1}^p$ forms a so-called martingale difference sequence; the fluctuations of such objects (which in a way extend the notion of series of independent random variables) are well controlled by Burkholder inequality as follows.

Lemma 2.12 (Burkholder inequality, Bai and Silverstein [2010, Lemma 2.13]). *Let $\{X_i\}_{i=1}^\infty$ be a martingale difference for the increasing σ -field $\{\mathcal{F}_i\}$ and denote \mathbb{E}_k the expectation with respect to \mathcal{F}_k . Then, for $k \geq 2$, and some constant C_k only dependent on k ,*

$$\mathbb{E} \left[\left| \sum_{i=1}^n X_i \right|^k \right] \leq C_k \left(\mathbb{E} \left[\sum_{i=1}^n \mathbb{E}_{i-1}[|X_i|^2] \right]^{k/2} + \sum_{i=1}^n \mathbb{E}[|X_i|^k] \right).$$

Lemma 2.12 will mostly be used in the context of proof details on the fluctuations of technical random matrix functionals. It may however be substituted by other similar tools such as the Gaussian Nash–Poincaré inequality (Lemma 2.14 in the “Gaussian method” proof framework to be discussed in Section 2.2.2), which also involves moment bounds but restricted to Gaussian random variables, or more conveniently with concentration inequalities (see Section 2.7 for detail) which no longer involve moments (which can be cumbersome to compute) but (exponential) tail bounds.

These identities constitute the main technical ingredients needed to understand the proofs of both historical and more recent random matrix results. The next section introduces the most fundamental of those, which will be repeatedly recalled in the remainder of the book.

2.2.2 The Marčenko–Pastur and Semicircle Laws

We start by illustrating how the aforementioned tools can be used to prove the two most popular results in random matrix theory: the Marčenko–Pastur law and the Wigner semicircle law.

To simplify the exposition of the results, we will use the notation for deterministic equivalents introduced in Notation 1. That is, for $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times n}$, we will denote $\mathbf{X} \leftrightarrow \mathbf{Y}$ if, for all unit norm $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, $\frac{1}{n} \text{tr} \mathbf{A}(\mathbf{X} - \mathbf{Y}) \xrightarrow{\text{a.s.}} 0$, $\mathbf{a}^\top (\mathbf{X} - \mathbf{Y}) \mathbf{b} \xrightarrow{\text{a.s.}} 0$ and $\|\mathbb{E}[\mathbf{X} - \mathbf{Y}]\| \rightarrow 0$.

Most of the results involve Stieltjes transforms $m_\mu(z)$ of a real probability measure with support $\text{supp}(\mu) \subset \mathbb{R}$. Since Stieltjes transforms are such that $m_\mu(z) > 0$ for $z < \inf \text{supp}(\mu)$, $m_\mu(z) < 0$ for $z > \sup \text{supp}(\mu)$ and $\Im[z] \cdot \Im[m_\mu(z)] > 0$ if $z \in \mathbb{C} \setminus \mathbb{R}$ (see Definition 3 and the discussions thereafter), it will be convenient to introduce the following shortcut notation.

Notation 2 (“Valid” Stieltjes transform pair). For $\mathcal{A} \subset \mathbb{C}$, $z \in \mathcal{A}$ and $m \in \mathbb{C}$, we denote $\mathcal{Z}(\mathcal{A})$ the set of scalar pairs

$$\mathcal{Z}(\mathcal{A}) = \{(z, m) \in \mathcal{A} \times \mathbb{C}, \text{ such that } (\Im[z] \cdot \Im[m] > 0 \text{ if } \Im[z] \neq 0) \\ \text{or } (m > 0 \text{ if } z < \inf \mathcal{A}^c \cap \mathbb{R}) \text{ or } (m < 0 \text{ if } z > \sup \mathcal{A}^c \cap \mathbb{R})\}.$$

In particular, for convenient choices of \mathcal{A} (not always $\mathbb{C} \setminus \text{supp}(\mu)$), many results presented next will involve pairs $(z, m(z))$ defined as the unique solution of an implicit equation *within* $\mathcal{Z}(\mathcal{A})$ (while the implicit equation may, in general, have more than one solution in $\mathbb{C} \times \mathbb{C}$).

The Marčenko–Pastur Law

We present the Marčenko–Pastur law under the slightly modified form of a deterministic equivalent for the resolvent $\mathbf{Q}(z)$.

Theorem 2.4 (Marčenko and Pastur [1967]). Let $\mathbf{X} \in \mathbb{R}^{p \times n}$ with i.i.d. columns \mathbf{x}_i s such that \mathbf{x}_i has independent entries with zero mean, unit variance, and satisfying

some light tail condition⁸ and denote $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{X}\mathbf{X}^\top - z\mathbf{I}_p)^{-1}$ the resolvent of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$. Then, as $n, p \rightarrow \infty$ with $p/n \rightarrow c \in (0, \infty)$,

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z), \quad \bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_p, \tag{2.8}$$

with $(z, m(z))$ the unique solution in $\mathcal{Z}(\mathbb{C} \setminus [(1 - \sqrt{c})^2, (1 + \sqrt{c})^2])$ (see Notation 2) of

$$zcm^2(z) - (1 - c - z)m(z) + 1 = 0. \tag{2.9}$$

The function $m(z)$ is the Stieltjes transform of the probability measure μ given explicitly by

$$\mu(dx) = (1 - c^{-1})^+ \delta_0(x) + \frac{1}{2\pi cx} \sqrt{(x - E_-)^+(E_+ - x)^+} dx \tag{2.10}$$

where $E_\pm = (1 \pm \sqrt{c})^2$ and $(x)^+ = \max(0, x)$, and is known as the Marčenko–Pastur distribution. In particular, with probability one, the empirical spectral measure $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\top}$ converges weakly to μ .

Figure 2.2 depicts the density of the Marčenko–Pastur distribution for different values of $c = \lim p/n$. For a “fixed” dimension p , the ratio c decreases as the number of samples n grows large, so that the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$ become more “concentrated” (their spread is given by the length of the support $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$) around the (unique) population covariance matrix eigenvalue (when seeing \mathbf{X} as a collection $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ of p -dimensional data vectors with $\mathbb{E}[\mathbf{x}_i] = \mathbf{0}$ and $\text{Cov}[\mathbf{x}_i] = \mathbf{I}_p$), which is equal to 1.

Note that the “asymmetric bell” shape of the Marčenko–Pastur law gets increasingly skewed toward large values as c increases and that, for $c = 1$, the left-edge value has the very peculiar behavior to diverge. This $c = 1$ setting is referred to as the “hard edge” scenario explained by the fact that the limiting density becomes

$$\frac{1}{2\pi x} \sqrt{x^+(4 - x)^+} \sim \frac{1}{\pi\sqrt{x}},$$

as $x \downarrow 0$ and thus behaves as $1/\sqrt{x}$ near the left edge (the left edge being at $x = 0$) rather than as $\sqrt{x - (1 - \sqrt{c})^2}$ when $c \neq 1$ (the left edge being at $x = (1 - \sqrt{c})^2$, see

⁸ For this result, and those related, various tail conditions may be considered, for example, a uniform finite moment of order k for some $k > 2$ (usually $k = 4 + \varepsilon$ for any $\varepsilon > 0$ is sufficient). Depending on the proof approach though, stronger conditions may be requested, such as a sub-Gaussian tail behavior, a concentration of measure-type condition, etc. Determining the minimalistic conditions for the results to hold has been of long interest to mathematicians, as demonstrated by the huge impact of the complete proof by Tao and Vu [2008] of the full-circle law theorem under no other condition than the identical distribution of the zero-mean and unit-variance entries (see also [Bordenave and Chafaï, 2014]). Yet, for machine learning purposes, these are of minor interest: We shall systematically assume “sufficiently smooth” (and technically convenient) conditions to hold, without hampering the practical applicability of the results. This being said, it is already interesting to observe that, here and in the vast majority of the upcoming results, the matrix entries need *not* be identically distributed, and that only the statistical mean and cross-variance of the entries dictate the limiting spectral behavior. We presently assume that \mathbf{X} has i.i.d. columns for technical convenience in the proof – for instance, to exploit the (rough) union bound in (2.20); this condition can be generalized to “independent columns” by considering, for example, that the \mathbf{x}_i s are sub-Gaussian random vectors [Vershynin, 2018, Section 3.4].

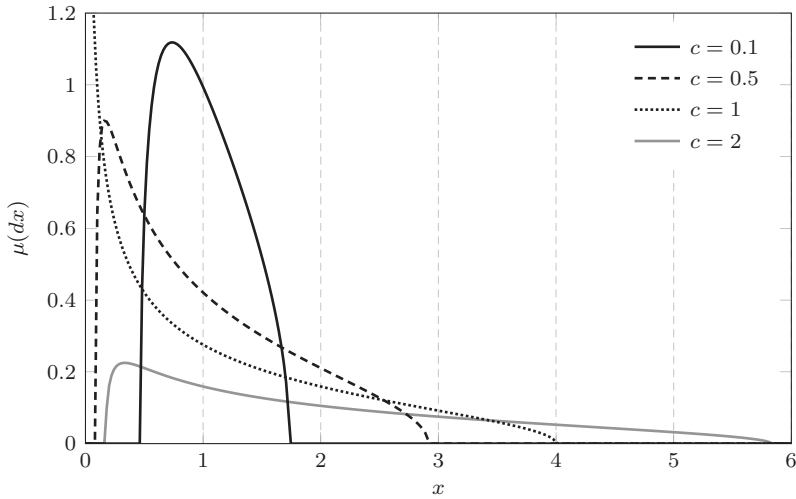


Figure 2.2 Marčenko–Pastur distribution for different values of c . Note the peculiar “hard-edge” behavior at $c = 1$, quite unlike other values of c .

also Exercise 6 for more detailed discussions on this point). When $c > 1$, a mass at zero is created (of weight $1 - c^{-1}$) while, possibly unexpectedly, the left edge of the main “bulk” of nonzero eigenvalues moves towards the right, leaving the open segment $(0, (1 - \sqrt{c})^2)$ empty.⁹

Proof of Theorem 2.4. Before going into the details of the proof, we first give a few intuitive arguments.

Intuitive idea

A first heuristic derivation, essentially due to Bai and Silverstein, consists in iteratively “guessing” the form of $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ for some matrix $\mathbf{F}(z)$. To this end, from Lemma 2.1, it first appears that, writing $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$,

$$\begin{aligned} \mathbf{Q}(z) - \bar{\mathbf{Q}}(z) &= \mathbf{Q}(z) \left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n}\mathbf{X}\mathbf{X}^\top \right) \bar{\mathbf{Q}}(z) \\ &= \mathbf{Q}(z) \left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top \right) \bar{\mathbf{Q}}(z). \end{aligned}$$

For $\bar{\mathbf{Q}}(z)$ to be a deterministic equivalent for $\mathbf{Q}(z)$, we wish, in particular, that $\frac{1}{p} \text{tr} \mathbf{A}(\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \xrightarrow{\text{a.s.}} 0$, for \mathbf{A} arbitrary, deterministic, and such that $\|\mathbf{A}\| = 1$. That is,

$$\frac{1}{p} \text{tr}(\mathbf{F}(z) + z\mathbf{I}_p)\bar{\mathbf{Q}}(z)\mathbf{A}\mathbf{Q}(z) - \frac{1}{n} \sum_{i=1}^n \frac{1}{p} \mathbf{x}_i^\top \bar{\mathbf{Q}}(z)\mathbf{A}\mathbf{Q}(z)\mathbf{x}_i \xrightarrow{\text{a.s.}} 0. \tag{2.11}$$

⁹ This hard-edge phenomenon is in fact not just an amusing artifact of the theory: It indeed has deep consequences in practice and notably explains the so-called *double-descent* phenomenon lately evidenced in large-dimensional statistical inference (see, e.g., [Nakkiran et al., 2020, Mei and Montanari, 2021, Deng et al., 2021, Liao et al., 2020]).

We recognize $\mathbf{x}_i^\top \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}(z) \mathbf{x}_i / p$ as a quadratic form on which we would like to use Lemma 2.11 to turn it into a trace term independent of \mathbf{x}_i . Yet, Lemma 2.11 *cannot* be applied directly as $\mathbf{Q}(z)$ depends on \mathbf{x}_i . To address this issue, we then use Lemma 2.8 to write

$$\mathbf{Q}(z) \mathbf{x}_i = \frac{\mathbf{Q}_{-i}(z) \mathbf{x}_i}{1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i}(z) \mathbf{x}_i},$$

where $\mathbf{Q}_{-i}(z) = (\frac{1}{n} \sum_{j \neq i} \mathbf{x}_j \mathbf{x}_j^\top - z \mathbf{I}_p)^{-1}$ is independent of \mathbf{x}_i . Now legitimately applying Lemma 2.11, we find that

$$\frac{1}{p} \mathbf{x}_i^\top \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}(z) \mathbf{x}_i = \frac{\frac{1}{p} \mathbf{x}_i^\top \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}_{-i}(z) \mathbf{x}_i}{1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i}(z) \mathbf{x}_i} \simeq \frac{\frac{1}{p} \text{tr} \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}_{-i}(z)}{1 + \frac{1}{n} \text{tr} \mathbf{Q}_{-i}(z)}. \tag{2.12}$$

From Lemma 2.9, normalized traces involving $\mathbf{Q}_{-i}(z)$ and $\mathbf{Q}(z)$ are asymptotically identical (since their inverse only differs by the rank-one matrix $\frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top$) and thus this further reads

$$\frac{1}{p} \mathbf{x}_i^\top \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}(z) \mathbf{x}_i \simeq \frac{\frac{1}{p} \text{tr} \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}(z)}{1 + \frac{1}{n} \text{tr} \mathbf{Q}(z)}.$$

Getting back to (2.11), we thus end up with the approximation

$$\frac{1}{p} \text{tr}(\mathbf{F}(z) + z \mathbf{I}_p) \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}(z) \simeq \frac{\frac{1}{p} \text{tr} \bar{\mathbf{Q}}(z) \mathbf{A} \mathbf{Q}(z)}{1 + \frac{1}{n} \text{tr} \mathbf{Q}(z)}, \tag{2.13}$$

(the argument of the right-hand side summation over i no longer depends on i , so the sum symbol vanishes). As a consequence, we can now “guess” the form of $\mathbf{F}(z)$: if it is to exist, $\mathbf{F}(z)$ must be of the type

$$\mathbf{F}(z) \simeq \left(-z + \frac{1}{1 + \frac{1}{n} \text{tr} \mathbf{Q}(z)} \right) \mathbf{I}_p,$$

for the approximation above to hold. To close the loop, taking $\mathbf{A} = \mathbf{I}_p$, $\frac{1}{p} \text{tr} \mathbf{Q}(z)$ appearing in this display must be well approximated by $m(z) \equiv \frac{1}{p} \text{tr} \bar{\mathbf{Q}}(z) = \frac{1}{p} \text{tr} \mathbf{F}^{-1}(z)$ so that

$$\frac{1}{p} \text{tr} \mathbf{Q}(z) \simeq m(z) = \frac{1}{-z + \frac{1}{1 + \frac{1}{n} \frac{1}{p} \text{tr} \mathbf{Q}(z)}} \simeq \frac{1}{-z + \frac{1}{1 + \frac{1}{n} m(z)}}, \tag{2.14}$$

and we thus finally have

$$\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z) = m(z) \mathbf{I}_p,$$

where, in the large n, p limit, $m(z)$ is solution to

$$m(z) = \left(-z + \frac{1}{1 + cm(z)} \right)^{-1},$$

or equivalently

$$zcm^2(z) - (1 - c - z)m(z) + 1 = 0.$$

This equation has two solutions defined via the two values of the complex square root function (letting $z = \rho e^{i\theta}$ for $\rho \geq 0$ and $\theta \in [0, 2\pi)$, $\sqrt{z} \in \{\pm\sqrt{\rho}e^{i\theta/2}\}$)

$$m(z) = \frac{1 - c - z}{2cz} + \frac{\sqrt{((1 + \sqrt{c})^2 - z)((1 - \sqrt{c})^2 - z)}}{2cz},$$

only one of which is such that $\Im[z]\Im[m(z)] > 0$ as imposed by the definition of Stieltjes transforms, see again Definition 3 and the discussion after that. Now, from the inverse Stieltjes transform theorem, Theorem 2.1, we find that $m(z)$ is the Stieltjes transform of the measure μ with

$$\mu([a, b]) = \frac{1}{\pi} \lim_{y \downarrow 0} \int_a^b \Im[m(x + iy)] dx,$$

for all continuity points $a, b \in \mathbb{R}$ of μ . The term under the square root in $m(z)$ being nonnegative only in the set $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$ (and thus of nonreal square root), the latter defines the support of the continuous part of the measure μ with density $\frac{\sqrt{((1 + \sqrt{c})^2 - x)(x - (1 - \sqrt{c})^2)}}{2c\pi x}$ at point x in the set. The case $x = 0$ brings a discontinuity in μ with weight equal to

$$\mu(\{0\}) = -\lim_{y \downarrow 0} iy m(iy) = \frac{c - 1}{2c} \pm \frac{c - 1}{2c},$$

where the sign is established by a second-order development of $zm(z)$ in the neighborhood of zero: that is, “+” for $c > 1$ inducing a mass $1 - 1/c$ for $p > n$, or “-” for $c < 1$ in which case $\mu(\{0\}) = 0$ and μ has no mass at zero.

Detailed proof of Theorem 2.4

Having heuristically identified $\bar{\mathbf{Q}}(z)$, we shall now use sound mathematical tools to prove that, indeed, $\bar{\mathbf{Q}}(z)$ is a deterministic equivalent for $\mathbf{Q}(z)$ in the sense of the theorem statement. Let us first show that

$$\mathbb{E}[\mathbf{Q}(z)] = \bar{\mathbf{Q}}(z) + o_{\|\cdot\|}(1), \tag{2.15}$$

where $o_{\|\cdot\|}(1)$ denotes a matrix term of vanishing operator norm as $n, p \rightarrow \infty$.

Convergence in mean. For mathematical convenience, we will take $z < 0$ in what follows. Since $\mathbf{Q}(z)$ and $\bar{\mathbf{Q}}(z)$ in the theorem statement are complex analytic functions for $z \notin \mathbb{R}^+$ (matrix-valued Stieltjes transforms are analytic), by Vitali’s convergence theorem, Theorem 2.3, obtaining the convergence results on \mathbb{R}^- (in fact even on a restricted local subset of \mathbb{R}^-) is equivalent to obtaining the result on all of $\mathbb{C} \setminus \mathbb{R}^+$.

We proceed in two steps by first introducing the intermediate deterministic quantities

$$\alpha(z) \equiv \frac{1}{n} \text{tr} \mathbb{E}[\mathbf{Q}_{-1}(z)], \quad \bar{\bar{\mathbf{Q}}}(z) \equiv \left(-z + \frac{1}{1 + \alpha(z)}\right)^{-1} \mathbf{I}_p, \tag{2.16}$$

where we denote $\mathbf{Q}_{-j}(z) \equiv (\frac{1}{n} \sum_{i \neq j} \mathbf{x}_i \mathbf{x}_i^T - z \mathbf{I}_p)^{-1}$ the “leave-one-out” version of $\mathbf{Q}(z)$ by removing the contribution from \mathbf{x}_j and use the fact that the distribution of \mathbf{Q}_{-j} is independent of the index j , as a consequence of the i.i.d. ness of the \mathbf{x}_i s.

From Lemma 2.1, we have (the argument z in $\alpha(z)$, $\mathbf{Q}(z)$ and $\bar{\mathbf{Q}}(z)$ is dropped when confusion is not possible)

$$\begin{aligned} \mathbb{E}[\mathbf{Q} - \bar{\mathbf{Q}}] &= \mathbb{E}\mathbf{Q} \left(\frac{\mathbf{I}_p}{1 + \alpha} - \frac{1}{n} \mathbf{X}\mathbf{X}^\top \right) \bar{\mathbf{Q}} = \frac{\mathbb{E}[\mathbf{Q}]}{1 + \alpha} \bar{\mathbf{Q}} - \frac{1}{n} \mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^\top] \bar{\mathbf{Q}} \\ &= \frac{\mathbb{E}[\mathbf{Q}]}{1 + \alpha} \bar{\mathbf{Q}} - \sum_{i=1}^n \frac{1}{n} \mathbb{E}[\mathbf{Q}\mathbf{x}_i\mathbf{x}_i^\top] \bar{\mathbf{Q}} = \frac{\mathbb{E}[\mathbf{Q}]}{1 + \alpha} \bar{\mathbf{Q}} - \sum_{i=1}^n \mathbb{E} \left[\frac{\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top}{1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i} \right] \bar{\mathbf{Q}}, \end{aligned}$$

where we applied Lemma 2.8 to obtain the last equality.

Since we expect $\frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i$ to be close to α (as a consequence of Lemmas 2.11 and 2.12), we rewrite

$$\frac{\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top}{1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i} = \frac{\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top}{1 + \alpha} - \frac{\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top (\frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i - \alpha)}{(1 + \alpha)(1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i)},$$

so that

$$\begin{aligned} \mathbb{E}[\mathbf{Q} - \bar{\mathbf{Q}}] &= \frac{\mathbb{E}[\mathbf{Q}]}{1 + \alpha} \bar{\mathbf{Q}} - \sum_{i=1}^n \frac{\mathbb{E}[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top] \bar{\mathbf{Q}}}{1 + \alpha} + \sum_{i=1}^n \frac{\mathbb{E}[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top d_i] \bar{\mathbf{Q}}}{1 + \alpha} \\ &= \frac{\mathbb{E}[\mathbf{Q}]}{1 + \alpha} \bar{\mathbf{Q}} - \sum_{i=1}^n \frac{\mathbb{E}[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top] \bar{\mathbf{Q}}}{1 + \alpha} + \frac{\mathbb{E}[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{X}\mathbf{D}\mathbf{X}^\top] \bar{\mathbf{Q}}}{1 + \alpha}, \end{aligned}$$

where we introduced $\mathbf{D} = \text{diag}\{d_i\}_{i=1}^n$ for $d_i = \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i - \alpha$, and used again Lemma 2.8 to write $\frac{\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top}{1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i} = \mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top$ in the first equality. Since $\mathbb{E}[\mathbf{Q}_{-i} \mathbf{x}_i \mathbf{x}_i^\top] = \mathbb{E}[\mathbf{Q}_{-i}]$, this further reads

$$\mathbb{E}[\mathbf{Q} - \bar{\mathbf{Q}}] = \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[\mathbf{Q}] - \mathbb{E}[\mathbf{Q}_{-i}]) \frac{\bar{\mathbf{Q}}}{1 + \alpha} + \frac{\mathbb{E}[\frac{1}{n} \mathbf{Q}\mathbf{X}\mathbf{D}\mathbf{X}^\top] \bar{\mathbf{Q}}}{1 + \alpha}. \tag{2.17}$$

For the first right-hand side term, again from Lemmas 2.1 and 2.8,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbf{Q} - \mathbf{Q}_{-i}] &= -\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top \mathbf{Q}_{-i} \right] \\ &= -\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^\top \mathbf{Q} \left(1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i \right) \right] \\ &= -\frac{1}{n} \mathbb{E} \left[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{X}\mathbf{D}_2 \mathbf{X}^\top \mathbf{Q} \right], \end{aligned} \tag{2.18}$$

where $\mathbf{D}_2 = \text{diag} \left\{ 1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i \right\}_{i=1}^n$ and thus

$$\mathbb{E}[\mathbf{Q} - \bar{\mathbf{Q}}] = -\frac{1}{n} \mathbb{E} \left[\mathbf{Q}_{-i} \frac{1}{n} \mathbf{X}\mathbf{D}_2 \mathbf{X}^\top \mathbf{Q} \right] \frac{\bar{\mathbf{Q}}}{1 + \alpha} + \frac{\mathbb{E}[\frac{1}{n} \mathbf{Q}\mathbf{X}\mathbf{D}\mathbf{X}^\top] \bar{\mathbf{Q}}}{1 + \alpha}. \tag{2.19}$$

It remains to show that the right-hand side terms vanish in the large p, n limit.

For the first term, note that

$$0 \leq \mathbf{Q}_{-i} \frac{1}{n} \mathbf{X}\mathbf{D}_2 \mathbf{X}^\top \mathbf{Q} \leq \mathbf{Q}_{-i} \frac{1}{n} \mathbf{X}\mathbf{X}^\top \mathbf{Q} \cdot \max_{1 \leq i \leq n} [\mathbf{D}_2]_{ii}$$

in the order of symmetric matrices. Since $\mathbf{Q} \frac{1}{n} \mathbf{X} \mathbf{X}^T = \mathbf{I}_p + z \mathbf{Q}$ which is of bounded operator norm (by 2) and $\|\mathbf{Q}\| \leq 1/|z|$, controlling $\|\mathbb{E}[\mathbf{Q} \frac{1}{n} \mathbf{X} \mathbf{D}_2 \mathbf{X}^T \mathbf{Q}]\|$ boils down to controlling $\mathbb{E}[\max_i [\mathbf{D}_2]_{ii}]$. This can be established in various ways. For instance, from the union bound and the i.i.d. nature of the \mathbf{x}_i s,

$$\mathbb{P}\left(\max_i [\mathbf{D}_2]_{ii} > t\right) \leq n \cdot \mathbb{P}([\mathbf{D}_2]_{11} > t). \tag{2.20}$$

Now, by Markov’s inequality $\mathbb{P}(X > t) \leq \mathbb{E}[X^k]/t^k$ for every k (with $X, t > 0$) and the moment inequality in Lemma 2.11 for, say $k = 4$, $\mathbb{P}(\max_i [\mathbf{D}_2]_{ii} > t)$ may be bounded by a function decreasing as t^{-4} , for all t large, and of order n^{-1} . Specifically, for k even,

$$\begin{aligned} \mathbb{P}\left([\mathbf{D}_2]_{11} > t + 1 + \frac{1}{n} \text{tr} \mathbf{Q}_{-1}\right) &\leq \frac{\mathbb{E}\left[\left(\frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \frac{1}{n} \text{tr} \mathbf{Q}_{-1}\right)^k\right]}{t^k} \\ &\leq \frac{\mathbb{E}_{\mathbf{Q}_{-1}} \mathbb{E}_{\mathbf{x}_1}\left[\left(\frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \frac{1}{n} \text{tr} \mathbf{Q}_{-1}\right)^k\right]}{t^k} \end{aligned}$$

where we isolated the expectation over \mathbf{Q}_{-1} from that over \mathbf{x}_1 to let appear the difference $\frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \frac{1}{n} \text{tr} \mathbf{Q}_{-1}$ which, conditionally on \mathbf{Q}_{-1} of bounded norm, we know is small and can be controlled using Lemma 2.11:

$$\mathbb{E}_{\mathbf{x}_1} \left[\left| \frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \frac{1}{n} \text{tr} \mathbf{Q}_{-1} \right|^4 \right] \leq \frac{C}{n^4} \text{tr}^2(\mathbf{Q}_{-1}^2),$$

for some constant $C > 0$, which depends on the fourth- and eighth-order moments of the entries of \mathbf{x} , but which is *independent* of n, p , according to Lemma 2.11 with $k = 4$. Since $\|\mathbf{Q}_{-1}\| \leq 1/|z|$ (note that this key boundedness property of the resolvent is used to simplify the analysis, here and in most random matrix proofs), we have $\text{tr}^2(\mathbf{Q}_{-1}^2) \leq p^2/|z|^4$, $1 + \frac{1}{n} \text{tr} \mathbf{Q}_{-1} \leq 1 + p/(n|z|)$, and therefore

$$\mathbb{P}([\mathbf{D}_2]_{11} > t) \leq \frac{Cp^2}{n^4|z|^4 \cdot t^4}$$

holds for all $t > C'$ for some $C' > 0$ that depends on n, p only via their ratio p/n . Finally, since

$$\begin{aligned} \mathbb{E}[\max_i [\mathbf{D}_2]_{ii}] &= \int_0^{C'} \mathbb{P}(\max_i [\mathbf{D}_2]_{ii} > t) dt + \int_{C'}^\infty \mathbb{P}(\max_i [\mathbf{D}_2]_{ii} > t) dt \\ &\leq C' + n \int_{C'}^\infty \mathbb{P}([\mathbf{D}_2]_{11} > t) dt \leq C' + \frac{Cp^2}{n^3|z|^4} \int_{C'}^\infty t^{-4} dt < \infty \end{aligned}$$

we find that $\mathbb{E}[\max_i [\mathbf{D}_2]_{ii}]$ is bounded. Note that this also proves, by (2.18), that $\|\mathbb{E}[\mathbf{Q} - \mathbf{Q}_{-1}]\| = O(n^{-1})$. Consequently, due to the leading $1/n$ factor in front of the first right-hand side term of (2.19), this term vanishes as $n, p \rightarrow \infty$.¹⁰

¹⁰ Another proof option could have been to derive a moment inequality for the random variable $|\mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \text{tr} \mathbb{E}[\mathbf{Q}_{-1}]|^k$ rather than for $\mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \text{tr} \mathbf{Q}_{-1}$, which would have involved Burkholder inequality used a bit later in the proof to control the fluctuations of $\text{tr} \mathbf{Q}_{-1} - \text{tr} \mathbb{E} \mathbf{Q}_{-1}$. But, as we saw, the fundamental boundedness of $\|\mathbf{Q}_{-1}\|$ discards *here* the need to control the fluctuations of \mathbf{Q}_{-1} .

To now handle the second right-hand side term in (2.19), one needs to control the norm of $\frac{1}{n}\mathbf{Q}\mathbf{X}\mathbf{D}\mathbf{X}^T\bar{\mathbf{Q}}$. This is not a symmetric matrix, but $\mathbb{E}[\mathbf{Q} - \bar{\mathbf{Q}}]$ is. We may rewrite (2.19) as the half-sum of itself and its transpose and we are thus left to controlling the operator norm of $\frac{1}{n}\mathbf{Q}\mathbf{X}\mathbf{D}\mathbf{X}^T\bar{\mathbf{Q}} + \frac{1}{n}\bar{\mathbf{Q}}\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{Q}$. Using the matrix inequalities $\mathbf{A}\mathbf{B}^T + \mathbf{B}\mathbf{A}^T \preceq \mathbf{A}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T$ (from $(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^T \succeq 0$) and $\mathbf{A}\mathbf{B}^T + \mathbf{B}\mathbf{A}^T \succeq -\mathbf{A}\mathbf{A}^T - \mathbf{B}\mathbf{B}^T$ (from $(\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B})^T \succeq 0$), it suffices to bound the norm of

$$\mathbb{E} \left[\frac{n^\varepsilon}{n} \mathbf{Q}\mathbf{X}\mathbf{D}^2\mathbf{X}^T\mathbf{Q} \right] + \mathbb{E} \left[\frac{n^{-\varepsilon}}{n} \bar{\mathbf{Q}}\mathbf{X}\mathbf{X}^T\bar{\mathbf{Q}} \right]$$

where the division of the n^{-2} constant into $n^{-1+\varepsilon}$ and $n^{-1-\varepsilon}$ for some $\varepsilon \in (0, 1/2]$ will appear as essential, since both terms may *not* have the same orders of magnitude (which depend on the so far unknown magnitude of the entries of \mathbf{D}). The second term above is easily seen to be of order $O(n^{-\varepsilon})$. As for the first term, we write, similar to the bound on \mathbf{D}_2 ,

$$\begin{aligned} n^\varepsilon \mathbb{E}[\|\mathbf{D}\|^2] &= n^\varepsilon \mathbb{E} \left[\max_i d_i^2 \right] \\ &\leq n^\varepsilon \int_0^{C'n^{-\theta-\varepsilon}} \mathbb{P} \left(\max_i d_i^2 > t \right) dt + n^{1+\varepsilon} \int_{C'n^{-\theta-\varepsilon}}^\infty \mathbb{P} (d_1^2 > t) dt \\ &\leq C'n^{-\theta} + n^{1+\varepsilon} \int_{C'n^{-\theta-\varepsilon}}^\infty \mathbb{P} \left(\left| \frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \alpha \right|^2 > t \right) dt, \end{aligned}$$

for some $C' > 0$ and $\theta \in (0, 1/2]$ to be determined, $d_i = \frac{1}{n} \mathbf{x}_i^T \mathbf{Q}_{-1} \mathbf{x}_i - \alpha$ and $\alpha = \frac{1}{n} \text{tr} \mathbb{E}[\mathbf{Q}_{-1}] > 0$. Here, since α involves an expectation over \mathbf{Q}_{-1} (and not \mathbf{Q}_{-1} itself as in the bound of $\|\mathbf{D}_2\|$), one needs to be more precise in the control of the fluctuations of *both* \mathbf{x}_1 and \mathbf{Q}_{-1} . Specifically, we write

$$\begin{aligned} &\mathbb{E} \left| \frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \frac{1}{n} \text{tr} \mathbb{E}[\mathbf{Q}_{-1}] \right|^4 \\ &= \mathbb{E} \left| \frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \frac{1}{n} \text{tr} \mathbf{Q}_{-1} + \frac{1}{n} \text{tr} (\mathbf{Q}_{-1} - \mathbb{E}[\mathbf{Q}_{-1}]) \right|^4 \\ &\leq \frac{8}{n^4} \mathbb{E} \left[\left| \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \text{tr} \mathbf{Q}_{-1} \right|^4 \right] + \frac{8}{n^4} \mathbb{E} \left[\left| \text{tr} \mathbf{Q}_{-1} - \text{tr} \mathbb{E}[\mathbf{Q}_{-1}] \right|^4 \right], \end{aligned}$$

which we will show to be of order $O(n^{-2})$. For the first right-hand side term, this follows from Lemma 2.11. For the second term, which does not involve a quadratic form but the fluctuations of the columns of \mathbf{X} inside the intricate functional $\text{tr} \mathbf{Q}_{-1}$, we will resort to Burkholder inequality, Lemma 2.12. For the sake of further reuse, we will prove a slightly more general result on $\mathbb{E}[\|\text{tr} \mathbf{Q}_{-1} - \text{tr} \mathbb{E}[\mathbf{Q}_{-1}]\|^4]$: First note that by Lemma 2.9 we may freely replace \mathbf{Q}_{-1} with \mathbf{Q} in the result without altering the desired control, and that we may generalize the control to $\mathbb{E}[\|\text{tr} \mathbf{A}\mathbf{Q}_{-1} - \text{tr} \mathbb{E}[\mathbf{A}\mathbf{Q}_{-1}]\|^4]$ for arbitrary \mathbf{A} deterministic of bounded norm (again, this will be useful later).

Specifically, under the notation of Lemma 2.12, observe that we may write

$$\begin{aligned} \frac{1}{p} \operatorname{tr} \mathbf{A}(\mathbb{E} \mathbf{Q} - \mathbf{Q}) &= \sum_{i=1}^n \mathbb{E}_i \left[\frac{1}{p} \operatorname{tr} \mathbf{A} \mathbf{Q} \right] - \mathbb{E}_{i-1} \left[\frac{1}{p} \operatorname{tr} \mathbf{A} \mathbf{Q} \right] \\ &= \frac{1}{p} \sum_{i=1}^n (\mathbb{E}_i - \mathbb{E}_{i-1}) [\operatorname{tr} \mathbf{A}(\mathbf{Q} - \mathbf{Q}_{-i})], \end{aligned}$$

(since $\mathbb{E}_i[\operatorname{tr} \mathbf{A} \mathbf{Q}_{-i}] = \mathbb{E}_{i-1}[\operatorname{tr} \mathbf{A} \mathbf{Q}_{-i}]$) for \mathcal{F}_i the σ -field generating the columns $\mathbf{x}_{i+1}, \dots, \mathbf{x}_n$ of \mathbf{X} and with the convention $\mathbb{E}_0[f(\mathbf{X})] = f(\mathbf{X})$. This forms a martingale difference sequence so that we fall under the scope of Burkholder inequality. Now, from the identity $\mathbf{Q} = \mathbf{Q}_{-i} - \frac{1}{n} \frac{\mathbf{Q}_{-i} \mathbf{x}_i \mathbf{x}_i^T \mathbf{Q}_{-i}}{1 + \frac{1}{n} \mathbf{x}_i^T \mathbf{Q}_{-i} \mathbf{x}_i}$ (Lemma 2.8),

$$(\mathbb{E}_i - \mathbb{E}_{i-1}) \left[\frac{1}{p} \operatorname{tr} \mathbf{A}(\mathbf{Q} - \mathbf{Q}_{-i}) \right] = -(\mathbb{E}_i - \mathbb{E}_{i-1}) \frac{\frac{1}{pn} \mathbf{x}_i^T \mathbf{Q}_{-i} \mathbf{A} \mathbf{Q}_{-i} \mathbf{x}_i}{1 + \frac{1}{n} \mathbf{x}_i^T \mathbf{Q}_{-i} \mathbf{x}_i},$$

which is order $O(p^{-1})$. As a consequence, from Lemma 2.12,

$$\mathbb{E} \left[\left| \frac{1}{p} \operatorname{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E} \mathbf{Q}) \right|^4 \right] = O(n^{-2}). \tag{2.21}$$

Of course, this, in particular, implies that $\mathbb{E}[|\frac{1}{p} \operatorname{tr}(\mathbf{Q}_{-1} - \mathbb{E} \mathbf{Q}_{-1})|^4] = O(n^{-2})$, as desired.

Having obtained this desired control on the moments, it finally follows from Markov’s inequality that

$$\mathbb{P} \left(\left| \frac{1}{n} \mathbf{x}_1^T \mathbf{Q}_{-1} \mathbf{x}_1 - \frac{1}{n} \operatorname{tr} \mathbb{E}[\mathbf{Q}_{-1}] \right|^2 > t \right) \leq C t^{-2} n^{-2},$$

for all $t > C'$ and for some constant $C', C > 0$. Therefore,

$$n^\varepsilon \mathbb{E}[\|\mathbf{D}\|^2] \leq C' n^{-\theta} + C C' n^{2\varepsilon + \theta - 1}.$$

By choosing, for instance, $\varepsilon = \theta = 1/4$, we thus conclude that¹¹

$$\|\mathbb{E}[\mathbf{Q}] - \bar{\bar{\mathbf{Q}}}\| \leq C n^{-1/4}, \quad \text{with } \bar{\bar{\mathbf{Q}}} = \left(-z + \frac{1}{1 + \alpha(z)} \right)^{-1} \mathbf{I}_p. \tag{2.22}$$

The introduction of the intermediate deterministic equivalent $\bar{\bar{\mathbf{Q}}}$ allowed us to compare \mathbf{Q} to $\bar{\bar{\mathbf{Q}}}$ by exploiting the more accessible statistical relation between \mathbf{Q} and $\mathbb{E}[\mathbf{Q}]$. We are now in position to compare the *deterministic* matrices $\bar{\bar{\mathbf{Q}}}$ and $\bar{\mathbf{Q}}$. To

¹¹ The obtained bound is of order $O(n^{-1/4})$, which is in fact suboptimal and could (at least) be improved to $O(n^{-1/2})$. It is interesting to note here that this loss in optimality follows from the very rough union bound $\mathbb{P}(\max_i d_i^2 > t) \leq n \mathbb{P}(d_i^2 > t)$, which the fourth-order moment bound in $O(n^{-2})$ applied in Markov’s inequality does not optimally compensates. Alternative approaches to avoid this suboptimality are (i) to either evaluate higher-order moments (in general, the moment of order $2k$ is bounded by $C n^{-k}$) but this may come at the cost of cumbersome calculus; or more conveniently (ii) to obtain *exponential* decay bounds of $\mathbb{P}(d_i^2 > t)$ of the order $O(e^{-n^\alpha})$, which automatically annihilate the polynomial loss induced by the extra factor n . Item (ii) partially justifies the relevance of a concentration of measure framework for random matrices, which we will detail in Section 2.7.

this end, recalling that $\bar{\mathbf{Q}}$ is defined implicitly through $\bar{\mathbf{Q}} = m(z)\mathbf{I}_p$ with $m(z) = (-z + \frac{1}{1+cm(z)})^{-1} = \frac{1}{p} \text{tr} \bar{\mathbf{Q}}(z)$, we write, again with Lemma 2.1,

$$\bar{\bar{\mathbf{Q}}} - \bar{\mathbf{Q}} = \frac{\alpha(z) - cm(z)}{(1 + cm(z))(1 + \alpha(z))} \bar{\bar{\mathbf{Q}}}\bar{\mathbf{Q}},$$

so that

$$\begin{aligned} |\alpha(z) - cm(z)| &= \left| \frac{1}{n} \text{tr} \left(\mathbb{E}[\mathbf{Q}_{-1}(z)] - \bar{\bar{\mathbf{Q}}}(z) \right) + \frac{1}{n} \text{tr} \left(\bar{\bar{\mathbf{Q}}}(z) - \bar{\mathbf{Q}}(z) \right) \right| \\ &= |\alpha(z) - cm(z)| \cdot \frac{\frac{1}{n} \text{tr}(\bar{\bar{\mathbf{Q}}}(z)\bar{\mathbf{Q}}(z))}{(1 + cm(z))(1 + \alpha(z))} + O(n^{-1/4}), \end{aligned}$$

where we used the fact that $\|\mathbb{E}[\mathbf{Q}_{-1}] - \bar{\bar{\mathbf{Q}}}\| \leq \|\mathbb{E}[\mathbf{Q}_{-1} - \mathbf{Q}]\| + \|\mathbb{E}[\mathbf{Q}] - \bar{\bar{\mathbf{Q}}}\| = O(n^{-1/4})$ from (2.22). Since $\alpha(z) > 0$ for $z < 0$, we have

$$\mathbf{0} \prec \frac{\bar{\bar{\mathbf{Q}}}(z)}{1 + \alpha(z)} \prec \frac{\mathbf{I}_p}{1 - z},$$

so that

$$0 < \frac{\frac{1}{n} \text{tr}(\bar{\bar{\mathbf{Q}}}(z)\bar{\mathbf{Q}}(z))}{(1 + cm(z))(1 + \alpha(z))} < \frac{1}{1 - z} \frac{cm(z)}{1 + cm(z)} < 1,$$

and therefore, since $m(z) > 0$ for $z < 0$,

$$|\alpha(z) - cm(z)| \rightarrow 0,$$

which concludes the proof of (2.15), and thus of the ‘‘convergence in mean’’ part of Theorem 2.4.

Concentration and almost sure convergence. To now prove the almost sure convergence $\frac{1}{p} \text{tr} \mathbf{A}(\mathbf{Q} - \bar{\mathbf{Q}}) \xrightarrow{\text{a.s.}} 0$ and $\mathbf{a}^\top(\mathbf{Q} - \bar{\mathbf{Q}})\mathbf{b} \xrightarrow{\text{a.s.}} 0$, it suffices to show

$$\frac{1}{p} \text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q}) \xrightarrow{\text{a.s.}} 0, \quad \mathbf{a}^\top(\mathbf{Q} - \mathbb{E}\mathbf{Q})\mathbf{b} \xrightarrow{\text{a.s.}} 0.$$

Both results can be proved similarly using Burkholder inequality, Lemma 2.12 (which is the historical approach proposed by Bai and Silverstein [2010]). We have indeed already proved in (2.21) that $\mathbb{E}[|\text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q})/p|^4] = O(n^{-2})$ so that, from Markov’s inequality (i.e., $\mathbb{P}(|X| > t) \leq \mathbb{E}[|X|^k]/t^k$) and the Borel–Cantelli lemma (i.e., $\mathbb{P}(|X_n| > t) = O(n^{-\ell})$ for some $\ell > 1$ for all $t > 0$ implies $X_n \xrightarrow{\text{a.s.}} 0$ as $n \rightarrow \infty$),

$$\frac{1}{p} \text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q}) \xrightarrow{\text{a.s.}} 0,$$

as requested. The convergence $\mathbf{a}^\top(\mathbf{Q} - \mathbb{E}\mathbf{Q})\mathbf{b} \xrightarrow{\text{a.s.}} 0$ can be obtained similarly. \square

A few remarks on Theorem 2.4 and its proof are in order.

Remark 2.3 (On the convergence rates). *In the course of the proofs above, we saw examples of a general concentration trend for linear statistics and bilinear/quadratic forms of random matrices. We shall indeed typically have for most of the models of random matrices $\mathbf{X} \in \mathbb{R}^{n \times n}$ under study in this book that*

- linear eigenvalue statistics $\frac{1}{n} \sum_{i=1}^n f(\lambda_i(\mathbf{X}))$ for sufficiently well-behaved f (so, for instance, $\frac{1}{n} \text{tr} \mathbf{Q}_X(z) = \frac{1}{n} \sum_i (\lambda_i(\mathbf{X}) - z)^{-1}$) converge at speed $O(1/n)$ (their variance scales like $O(1/n^2)$). From a central-limit theorem viewpoint, this is as fast as it can get. Indeed, \mathbf{X} is maximally composed of order $O(n^2)$ “degrees of freedom” and thus, by the central limit theorem, fluctuations are (at most) at speed $O(1/\sqrt{n^2}) = O(1/n)$.
- bilinear forms $\mathbf{a}^T f(\mathbf{X}) \mathbf{b}$, where $f(\mathbf{X}) = \mathbf{U} \text{diag}\{f(\lambda_i(\mathbf{X}))\}_{i=1}^n \mathbf{U}^T$ (in the spectral decomposition of \mathbf{X}) and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ of unit norm typically converge at a slower $O(1/\sqrt{n})$ speed. This weaker convergence speed can be understood by considering the case where $\mathbf{a} = \mathbf{b} = \mathbf{e}_1$ with \mathbf{e}_1 the canonical basis vector and $f(t) = (t - z)^{-1}$: In this case, by Lemma 2.6,

$$\mathbf{a}^T f(\mathbf{X}) \mathbf{b} = \mathbf{e}_1^T \mathbf{Q}(z) \mathbf{e}_1 = [\mathbf{Q}(z)]_{11} = (\mathbf{X}_{11} - z - \mathbf{X}_{-1,-1}(\mathbf{X}_{-1,-1} - z \mathbf{I}_{n-1})^{-1} \mathbf{X}_{-1,1})^{-1} \tag{2.23}$$

the fluctuation of which is dominated by that of \mathbf{X}_{11} and typically of order $O(1/\sqrt{n})$.

This remark is particularly interesting as it indicates, from a statistics viewpoint, that for data/feature matrix $\mathbf{X} \in \mathbb{R}^{p \times n}$, asymptotic approximations may gain accuracy by doubly exploiting the degrees of freedom in both the sample (n) and feature (p) sizes.

Remark 2.4 (On the assumptions on \mathbf{X}). Let us pursue here on footnote 8 to clarify the “light tail condition” phrase in Theorem 2.4. The Marčenko–Pastur law has been widely generalized and several times proven using different techniques. For instance, Adamczak [2011], O’Rourke [2012] assume that the \mathbf{X}_{ij} s are “weakly” dependent in the sense that their correlation or higher-order cross-moments vanish at a certain controlled speed as $n, p \rightarrow \infty$. Alternatively, the works of Bai and Silverstein [2010] tend to assume that the entries of \mathbf{X} are not necessarily identically distributed; in this case, an additional condition on the tails $\mathbb{P}(|\mathbf{X}_{ij}| > t)$ of the probability measures of the entries (for instance, a uniform bound on some moment higher than 2) is needed. El Karoui [2009] provides a first result, which assumes that the columns \mathbf{x}_i s of $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ are independent concentrated random vectors, an assumption that we will thoroughly discuss in Section 2.7; (very) roughly speaking, concentrated random vectors $\mathbf{x} \in \mathbb{R}^p$ can be written as $\mathbf{x} = \varphi(\mathbf{z})$, where $\mathbf{z} \in \mathbb{R}^p$ has i.i.d. entries either following a Gaussian law or of bounded support, and $\varphi: \mathbb{R}^p \rightarrow \mathbb{R}^p$ is any 1-Lipschitz function: This assumption essentially maintains the p degrees of freedom in \mathbf{x} (arising from \mathbf{z}), while allowing for strong correlation between the entries of \mathbf{x} . In this case, the Marčenko–Pastur law is indeed still valid if $\mathbf{x} = \varphi(\mathbf{z})$ has zero mean and identity covariance.

One may wonder how the (higher-order) moment conditions on the entries of \mathbf{X} could be relaxed as this seems to suggest that moment bounds can no longer be used. The approach historically proposed by Bai and Silverstein (well documented in Bai and Silverstein [2010]) relies on a truncation-and-centering approach which consists in replacing \mathbf{X} by a matrix $\tilde{\mathbf{X}}$ defined as $\tilde{\mathbf{X}}_{ij} = \mathbf{X}_{ij} \cdot 1_{|\mathbf{X}_{ij}| > t(n)}$ for a certain threshold t , typically (a well-chosen) function of n . Being “truncated,” the entries of $\tilde{\mathbf{X}}$ have moments of higher orders (of all orders if $t(n)$ is constant), so that moment bounds

can be used on $\tilde{\mathbf{X}}$. It then remains to show that the functional of \mathbf{X} of interest (e.g., the empirical spectral measure of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$) is asymptotically the same as that of $\tilde{\mathbf{X}}$ as $n, p \rightarrow \infty$. Other, possibly more convenient, techniques exist, which prove a result on \mathbf{X} having standard Gaussian entries (for instance, using Stein’s identity $\mathbb{E}[\xi f(\xi)] = \mathbb{E}[f'(\xi)]$ for $\xi \sim \mathcal{N}(0,1)$; see Lemma 2.13) before using specific controls on the deviations from the Gaussian case (such as generalized Stein’s lemma) to extrapolate between Gaussian and non-Gaussian cases. This is the subject of the next section.

The “Gaussian Method” Alternative

Pastur and Shcherbina [2011] propose an alternative proof scheme for Theorem 2.4, based on a two-step approach: (i) a proof for Gaussian \mathbf{X} and (ii) an interpolation method to non-Gaussian \mathbf{X} ; together known as the “Gaussian method.” Although less intuitive when compared to the Bai and Silverstein’s approach presented in the previous section, this method is much more flexible as it can handle more structured random matrix models, in particular, when the “guessing” part (of the ultimate deterministic equivalent $\tilde{\mathbf{Q}}$ for \mathbf{Q}) of Bai–Silverstein’s method is nontrivial.

The proof in the Gaussian case itself is handled in two steps (or more precisely is based on two ingredients): (i-a) convergence in mean of the resolvent with Stein’s lemma, Lemma 2.13, and (i-b) control of the variance with the Nash–Poincaré inequality, Lemma 2.14, to establish concentration and convergence (in probability or almost surely) of trace and bilinear forms.

Convergence in mean by Stein’s lemma.

Lemma 2.13 (Stein [1981]). *Let $x \sim \mathcal{N}(0,1)$ and $f: \mathbb{R} \rightarrow \mathbb{R}$ a continuously differentiable function having at most polynomial growth and such that $\mathbb{E}[f'(x)] < \infty$. Then,*

$$\mathbb{E}[xf(x)] = \mathbb{E}[f'(x)]. \tag{2.24}$$

In particular, for $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f: \mathbb{R}^p \rightarrow \mathbb{R}$ a continuously differentiable function with derivatives having at most polynomial growth with respect to p ,

$$\mathbb{E}[[\mathbf{x}]_i f(\mathbf{x})] = \sum_{j=1}^p [\mathbf{C}]_{ij} \mathbb{E} \left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j} \right], \tag{2.25}$$

where $\partial/\partial[\mathbf{x}]_i$ indicates differentiation with respect to the i th entry of \mathbf{x} ; or, in vector form

$$\mathbb{E}[\mathbf{x}f(\mathbf{x})] = \mathbf{C}\mathbb{E}[\nabla f(\mathbf{x})], \tag{2.26}$$

with $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to \mathbf{x} .

The lemma, sometimes referred to as the integration-by-parts formula for Gaussian variables, simply follows from

$$\begin{aligned} \mathbb{E}[xf(x)] &= \int xf(x)e^{-\frac{1}{2}x^2} dx \\ &= [-f(x)e^{-\frac{1}{2}x^2}]_{-\infty}^{\infty} + \int f'(x)e^{-\frac{1}{2}x^2} dx = \mathbb{E}[f'(x)] \end{aligned}$$

with integration by parts $\int u'v = [uv] - \int uv'$ for $u(x) = -e^{-\frac{1}{2}x^2}$ and $v(x) = f(x)$.

To prove (2.15) in the Gaussian case, let us thus assume \mathbf{X} Gaussian, that is, $\mathbf{X}_j \sim \mathcal{N}(0, 1)$ and exploit Lemma 2.13. First observe that $\mathbf{Q} = \frac{1}{z} \frac{1}{n} \mathbf{X}\mathbf{X}^T \mathbf{Q} - \frac{1}{z} \mathbf{I}_p$, so that

$$\mathbb{E}[\mathbf{Q}_{ij}] = \frac{1}{zn} \sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^T \mathbf{Q}]_{kj}] - \frac{1}{z} \delta_{ij},$$

in which $\mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^T \mathbf{Q}]_{kj}] = \mathbb{E}[xf(x)]$ for $x = \mathbf{X}_{ik}$ and $f(x) = [\mathbf{X}^T \mathbf{Q}]_{kj}$. Therefore, from Lemma 2.13 and the fact that $\partial \mathbf{Q} = -\frac{1}{n} \mathbf{Q} \partial(\mathbf{X}\mathbf{X}^T) \mathbf{Q}$,¹²

$$\begin{aligned} \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^T \mathbf{Q}]_{kj}] &= \mathbb{E} \left[\frac{\partial [\mathbf{X}^T \mathbf{Q}]_{kj}}{\partial \mathbf{X}_{ik}} \right] \\ &= \mathbb{E}[\mathbf{E}_{ik}^T \mathbf{Q}]_{kj} - \mathbb{E} \left[\frac{1}{n} \mathbf{X}^T \mathbf{Q} (\mathbf{E}_{ik} \mathbf{X}^T + \mathbf{X} \mathbf{E}_{ik}^T) \mathbf{Q} \right]_{kj} \\ &= \mathbb{E}[\mathbf{Q}_{ij}] - \mathbb{E} \left[\frac{1}{n} [\mathbf{X}^T \mathbf{Q}]_{ki} [\mathbf{X}^T \mathbf{Q}]_{kj} \right] - \mathbb{E} \left[\frac{1}{n} [\mathbf{X}^T \mathbf{Q} \mathbf{X}]_{kk} \mathbf{Q}_{ij} \right] \end{aligned}$$

for \mathbf{E}_{ij} the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il} \delta_{jm}$, so that, summing over k ,

$$\begin{aligned} \frac{1}{z} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}[\mathbf{X}^T \mathbf{Q}]_{kj}] &= \frac{1}{z} \mathbb{E}[\mathbf{Q}_{ij}] - \frac{1}{z} \frac{1}{n^2} \mathbb{E}[\mathbf{Q}_{ij} \text{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^T)] \\ &\quad - \frac{1}{z} \frac{1}{n^2} \mathbb{E}[\mathbf{Q}\mathbf{X}\mathbf{X}^T \mathbf{Q}]_{ij}. \end{aligned} \tag{2.27}$$

It is not too difficult to see that the term in the second line has vanishing operator norm (of order $O(n^{-1})$) as $n, p \rightarrow \infty$ (see later Remark 2.5, which shows that for complex-valued Gaussian \mathbf{X} this term does not even appear in the derivation). Also recall that $\text{tr}(\mathbf{Q}\mathbf{X}\mathbf{X}^T) = np + zn \text{tr} \mathbf{Q}$. As a result, matrix-wise, we obtain

$$\mathbb{E}[\mathbf{Q}] + \frac{1}{z} \mathbf{I}_p = \mathbb{E}[\mathbf{X}_{\cdot k}[\mathbf{X}^T \mathbf{Q}]_k] = \frac{1}{z} \mathbb{E}[\mathbf{Q}] - \frac{1}{z} \frac{1}{n} \mathbb{E}[\mathbf{Q}(p + z \text{tr} \mathbf{Q})] + o_{\|\cdot\|}(1),$$

where $\mathbf{X}_{\cdot k}$ and \mathbf{X}_k is the k th column and row of \mathbf{X} , respectively. As the random $\frac{1}{p} \text{tr} \mathbf{Q}$ is expected to converge to some *deterministic* $m(z)$ as $n, p \rightarrow \infty$, it can be taken out of the expectation in the limit so that, gathering all terms proportional to $\mathbb{E}[\mathbf{Q}]$ on the left-hand side, we finally have

$$\mathbb{E}[\mathbf{Q}](1 - p/n - z - p/n \cdot zm(z)) = \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which, taking the trace to identify $m(z)$, concludes the proof for the Gaussian case.

¹² This is the matrix version of $d(1/x) = -dx/x^2$.

Concentration and almost sure convergence by Nash–Poincaré inequality. To prove the concentration and the almost sure convergence of traces and bilinear forms of the resolvent in the case of Gaussian \mathbf{X} , one may then use the powerful Nash–Poincaré inequality as follows.

Lemma 2.14 (Nash–Poincaré inequality, [Pastur, 2005]). *For $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f: \mathbb{R}^p \rightarrow \mathbb{R}$ continuously differentiable with derivatives having at most polynomial growth with respect to p ,*

$$\text{Var}[f(\mathbf{x})] \leq \sum_{i,j=1}^p [\mathbf{C}]_{ij} \mathbb{E} \left[\frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_i} \frac{\partial f(\mathbf{x})}{\partial [\mathbf{x}]_j} \right] = \mathbb{E} \left[(\nabla f(\mathbf{x}))^\top \mathbf{C} \nabla f(\mathbf{x}) \right],$$

where we denote $\nabla f(\mathbf{x})$ the gradient of $f(\mathbf{x})$ with respect to \mathbf{x} .

The proof of Lemma 2.14 is quite elegant and is provided as an exercise, in Exercise 5 of Section 2.9.

In the present case, taking $f(\mathbf{X}) = \frac{1}{p} \text{tr} \mathbf{A} \mathbf{Q}$ for Gaussian \mathbf{X} with $\mathbf{X}_{ij} \sim \mathcal{N}(0, 1)$,

$$\text{Var} \left[\frac{1}{p} \text{tr} \mathbf{A} \mathbf{Q} \right] \leq \frac{1}{p^2} \sum_{i=1}^p \sum_{j=1}^n \mathbb{E} \left[\left| \frac{\partial \text{tr} \mathbf{A} \mathbf{Q}}{\partial [\mathbf{X}]_{ij}} \right|^2 \right].$$

Again using $\partial \mathbf{Q} = -\frac{1}{n} \mathbf{Q} \partial(\mathbf{X} \mathbf{X}^\top) \mathbf{Q}$, we find

$$\frac{\partial \text{tr} \mathbf{A} \mathbf{Q}}{\partial \mathbf{X}_{ij}} = -\frac{1}{n} [\mathbf{Q} \mathbf{A} \mathbf{Q} \mathbf{X} + \mathbf{Q} \mathbf{A}^\top \mathbf{Q} \mathbf{X}]_{ij},$$

so that, from $(a + b)^2 \leq 2(a^2 + b^2)$ and $\|\mathbf{A}\| = 1$,

$$\begin{aligned} \frac{1}{p^2} \sum_{i=1}^p \sum_{j=1}^n \mathbb{E} \left[\left| \frac{\partial \text{tr} \mathbf{A} \mathbf{Q}}{\partial \mathbf{X}_{ij}} \right|^2 \right] &\leq \frac{2}{p^2 n^2} \mathbb{E} [\text{tr}(\mathbf{Q} \mathbf{A} \mathbf{Q} \mathbf{X} \mathbf{X}^\top \mathbf{Q} \mathbf{A}^\top \mathbf{Q}) \\ &\quad + \text{tr}(\mathbf{Q} \mathbf{A}^\top \mathbf{Q} \mathbf{X} \mathbf{X}^\top \mathbf{Q} \mathbf{A} \mathbf{Q})] = O(n^{-2}). \end{aligned}$$

By Markov’s inequality and the Borel–Cantelli lemma, we thus have that $\frac{1}{p} \text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E} \mathbf{Q}) \xrightarrow{\text{a.s.}} 0$.

When it comes to evaluating the fluctuations of $\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b}$ with the same approach, it appears that $\text{Var}[\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b}] = O(n^{-1})$, which is enough to ensure convergence in probability (again by Markov’s inequality) but not in an almost sure sense (as the Borel–Cantelli lemma does not apply). Thus, one needs to resort to the evaluation of its higher-order moments, such as $\mathbb{E}[|\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b}|^4]$. To this end, we may use the fact that

$$\begin{aligned} &\mathbb{E}[|\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b}|^4] \\ &= \text{Var}[\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b}] + \left(\mathbb{E} \left[\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b} \right]^2 \right)^2 \\ &= \text{Var}[\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b}] + \left(\text{Var}[\mathbf{a}^\top (\mathbf{Q} - \mathbb{E} \mathbf{Q}) \mathbf{b}] \right)^2. \end{aligned}$$

Since we know that the rightmost term is of order $O(n^{-2})$, it remains to show, again through Nash–Poincaré inequality, that $\text{Var}[\mathbf{a}^\top(\mathbf{Q} - \mathbb{E}\mathbf{Q})\mathbf{b}]^2 = O(n^{-2})$, which is a cumbersome but easily obtained result as well.

Interpolation trick to non-Gaussian \mathbf{X} . To “interpolate” the obtained results from Gaussian \mathbf{X} to non-Gaussian \mathbf{X} , one may then use the following lemma, which can be viewed as a generalized version of Stein’s lemma to non-Gaussian distributions.

Lemma 2.15 (Interpolation trick, [Lytova and Pastur, 2009, Corollary 3.1]). *For $x \in \mathbb{R}$, a random variable with zero mean and unit variance, $y \sim \mathcal{N}(0, 1)$, and f a $(k + 2)$ times differentiable function with bounded derivatives,*

$$\mathbb{E}[f(x)] - \mathbb{E}[f(y)] = \sum_{l=2}^k \frac{\kappa_{l+1}}{2l!} \int_0^1 \mathbb{E}[f^{(l+1)}x(t)]t^{(l-1)/2} dt + \epsilon_k,$$

where κ_l is the l th cumulant of x , $x(t) = \sqrt{t}x + (1 - \sqrt{t})y$, and $|\epsilon_k| \leq C_k \mathbb{E}[|x|^{k+2}] \cdot \sup_t |f^{(k+2)}(t)|$ for some constant C_k only dependent on k .

All Gaussian expectations (means and variance) in the proof above can then be expressed as their non-Gaussian form up to a sum of moment control on the derivatives of f .

As mentioned above in (2.27), by considering complex Gaussian \mathbf{X} instead of real one, the derivation of Theorem 2.4 can be further simplified. This is detailed in the following remark.

Remark 2.5 (Simplification in the complex case). *The Marčenko–Pastur result presented in Theorem 2.4 has been proven universal with respect to the field (\mathbb{R} or \mathbb{C}) of the entries of \mathbf{X} , where the Gram matrix of interest in the complex case is $\mathbf{X}\mathbf{X}^*$ for \mathbf{X}^* the Hermitian conjugate (transpose conjugate) of \mathbf{X} . The resolvent now becomes $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{X}\mathbf{X}^* - z\mathbf{I}_p)^{-1}$. Interestingly, Stein’s lemma, Lemma 2.13, is simplified in the complex case into*

$$\mathbb{E}[\mathbf{X}_{ij} f(\mathbf{X}, \mathbf{X}^*)] = \mathbb{E}\left[\frac{d}{d\bar{\mathbf{X}}_{ij}} f(\mathbf{X}, \mathbf{X}^*)\right],$$

for $f(\mathbf{X}, \mathbf{X}^*)$ a (polynomially bounded) smooth function of both \mathbf{X} and \mathbf{X}^* , and $\bar{\mathbf{X}}_{ij}$ the complex conjugate of \mathbf{X}_{ij} , where the complex derivation rules become $(d/d\bar{x})(x) = 0$ and $(d/d\bar{x})\bar{x} = 1$ (see details in, for example Pastur and Shcherbina [2011]). As a consequence, we find that

$$\frac{d}{d\mathbf{X}_{ij}} \mathbf{X}\mathbf{X}^* = \mathbf{E}_{ij}\mathbf{X}^*,$$

for \mathbf{E}_{ij} , the indicator matrix with entry $[\mathbf{E}_{ij}]_{lm} = \delta_{il}\delta_{jm}$. This relation is more convenient to use than in the real case, where

$$\frac{d}{d\mathbf{X}_{ij}} \mathbf{X}\mathbf{X}^\top = \mathbf{E}_{ij}\mathbf{X}^\top + \mathbf{X}\mathbf{E}_{ij}^\top,$$

and two terms instead of one appear; in recollection of the derivation above of the Marčenko–Pastur theorem, Theorem 2.4, in the real case with Stein’s lemma, this extra term was anticipated to vanish as $n, p \rightarrow \infty$ (see Equation (2.27)).

This remark is particularly useful when universality is anticipated (essentially for all such “first order” deterministic equivalents) and when elaborate random matrix models are to be treated. That is, in these settings, it is convenient (at least as a preliminary exploration) to assume that \mathbf{X} has complex rather than real Gaussian entries.

Wigner Semicircle Law

While the Marčenko–Pastur law is at the heart of sample covariance matrix models and is thus a starting point in, for example, kernel methods for machine learning, Wigner semicircle law concerns symmetric matrices of independent entries (above and on the diagonal), which is more akin to random graphs and will be used in this book almost exclusively to this purpose.¹³

The main result, again presented under the form of a deterministic equivalent for the resolvent, is as follows.

Theorem 2.5 (Wigner [1955]). *Let $\mathbf{X} \in \mathbb{R}^{n \times n}$ be symmetric and such that the \mathbf{X}_{ij} s, $j \geq i$, are independent zero mean and unit variance random variables satisfying some light tail condition. Then, for $\mathbf{Q}(z) = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$, as $n \rightarrow \infty$,*

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z), \quad \bar{\mathbf{Q}}(z) = m(z)\mathbf{I}_n, \tag{2.28}$$

with $(z, m(z))$ the unique solution in $\mathcal{Z}(\mathbb{C} \setminus [-2, 2])$ of

$$m^2(z) + zm(z) + 1 = 0. \tag{2.29}$$

The function $m(z)$ is the Stieltjes transform of the probability measure

$$\mu(dx) = \frac{1}{2\pi} \sqrt{(4 - x^2)^+} dx, \tag{2.30}$$

which is known as the Wigner semicircle law.

Figure 2.3 compares the empirical spectral measure of \mathbf{X}/\sqrt{n} given in Theorem 2.5 with the (limiting) Wigner semicircle law (which, for a proper scaling of the axes, has a half circular shape as the name suggests), for $n = 1\,000$.

Sketch of proof of Theorem 2.5. Although not the historical method of Wigner,¹⁴ we propose here to follow exactly the two approaches detailed in the proof of the

¹³ Up to an important exception when dealing with “properly scaling kernels” in Section 4.3.

¹⁴ Wigner’s proof in Wigner [1955] relied on a method of moment approach: Having inferred that the limiting measure should be a semicircle, he proved via a combinatorial approach, that the successive “moments” $\frac{1}{n} \text{tr}(n^{-\frac{1}{2}} \mathbf{X})^k$ for $k = 1, 2, \dots$ must converge, as $n \rightarrow \infty$, to the moments of the semicircle measure $\int t^k \mu(dt)$. This method is simple but only useful if indeed the limiting measure μ can be inferred. In the Marčenko–Pastur case of Theorem 2.4 and even worse in more elaborate random matrix settings, the limiting measure μ is less obvious to anticipate.

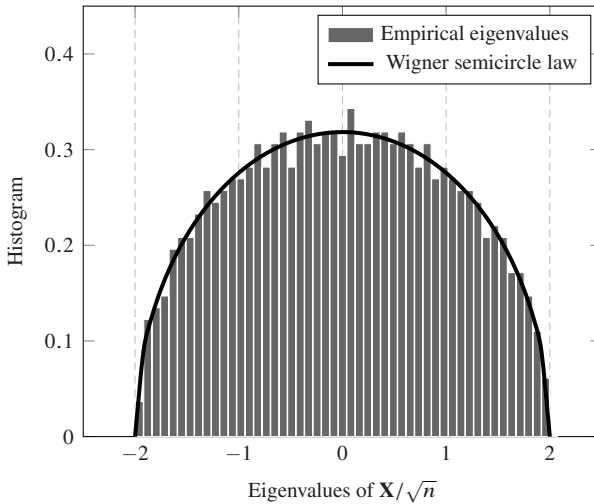


Figure 2.3 Histogram of the eigenvalues of \mathbf{X}/\sqrt{n} versus Wigner semicircle law, for \mathbf{X} having standard Gaussian entries and $n = 1000$. Code on web: MATLAB and Python.

Marčenko–Pastur theorem, Theorem 2.4. For pedagogical interest, we provide the main heuristic arguments both for the Bai–Silverstein and for the Gaussian method.

Bai–Silverstein heuristic. Let $\mathbf{Q} = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$ be the resolvent of interest, we write, by Lemma 2.6,

$$Q_{ii} = \frac{1}{\frac{1}{\sqrt{n}}\mathbf{X}_{ii} - z - \frac{1}{n}\mathbf{x}_i^T\mathbf{Q}_{-i}\mathbf{x}_i},$$

with $\mathbf{Q}_{-i} = (\mathbf{X}_{-i}/\sqrt{n} - z\mathbf{I}_{n-1})^{-1}$, $\mathbf{X}_{-i} \in \mathbb{R}^{(n-1) \times (n-1)}$ the matrix obtained by deleting the i th row and column from \mathbf{X} , and $\mathbf{x}_i \in \mathbb{R}^{n-1}$ the i th column (and thus the i th row by symmetry) of \mathbf{X} with its i th entry removed. Taking the sum over i we obtain

$$\frac{1}{n} \operatorname{tr} \mathbf{Q} = \frac{1}{n} \sum_{i=1}^n \frac{1}{\frac{1}{\sqrt{n}}\mathbf{X}_{ii} - z - \frac{1}{n}\mathbf{x}_i^T\mathbf{Q}_{-i}\mathbf{x}_i} = \frac{1}{n} \sum_{i=1}^n \frac{1}{-z - \frac{1}{n}\mathbf{x}_i^T\mathbf{Q}_{-i}\mathbf{x}_i} + o(1),$$

since $\frac{1}{\sqrt{n}}\mathbf{X}_{ii}$ asymptotically vanishes as $n \rightarrow \infty$. By Lemmas 2.9 and 2.11, we should have, for large n ,

$$\frac{1}{n}\mathbf{x}_i^T\mathbf{Q}_{-i}\mathbf{x}_i = \frac{1}{n} \operatorname{tr} \mathbf{Q}_{-i} + o(1) = \frac{1}{n} \operatorname{tr} \mathbf{Q} + o(1),$$

and thus the quadratic equation of $\frac{1}{n} \operatorname{tr} \mathbf{Q}$

$$\left(\frac{1}{n} \operatorname{tr} \mathbf{Q}\right)^2 + \frac{z}{n} \operatorname{tr} \mathbf{Q} + 1 = o(1).$$

With a concentration argument, for example, Lemma 2.12, we shall have, as $n \rightarrow \infty$, that $\frac{1}{n} \operatorname{tr} \mathbf{Q} - \frac{1}{n} \operatorname{tr} \mathbb{E}[\mathbf{Q}] \xrightarrow{\text{a.s.}} 0$ and therefore $\frac{1}{n} \operatorname{tr} \mathbf{Q}(z) - m(z) \xrightarrow{\text{a.s.}} 0$, with $m(z)$ the unique solution to

$$m^2(z) + zm(z) + 1 = 0,$$

the solution of which is explicitly given by

$$m(z) = \frac{1}{2}(-z + \sqrt{z^2 - 4}),$$

with $\sqrt{\cdot}$ again chosen as the branch of the square root for which $m(z)$ is a valid Stieltjes transform, see Notation 2. Taking the imaginary part and the limit when $z \rightarrow x \in \mathbb{R}$ (which is only nonzero if $x^2 - 4 < 0$) gives the form of the density $\mu(dx)$ in the theorem statement.

Note that in the above Bai–Silverstein heuristic, only the trace form $\frac{1}{n} \text{tr} \mathbf{Q}(z)$ was treated; when the more involved bilinear forms of the type $\mathbf{a}^\top \mathbf{Q}(z) \mathbf{b}$ are considered (in which case the *nondiagonal* entries of the inverse $\mathbf{Q}(z)$ need to be handled), it is often more convenient to resort to the Gaussian method proof approach as follows.

Gaussian method heuristic. Similar to the proof of the Marčenko–Pastur law with Gaussian methods in Section 2.2.2, observe that, for $\mathbf{Q} = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$, we have

$$\frac{1}{\sqrt{n}} \mathbb{E}[\mathbf{X}\mathbf{Q}] = \mathbf{I}_n + z\mathbb{E}[\mathbf{Q}], \tag{2.31}$$

so that by Lemma 2.13 and the fact that $\partial \mathbf{Q} = -\frac{1}{\sqrt{n}} \mathbf{Q}(\partial \mathbf{X})\mathbf{Q}$,

$$\begin{aligned} \mathbb{E}[\mathbf{Q}_{ij}] &= \frac{1}{z} \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik} \mathbf{Q}_{kj}] - \frac{1}{z} \delta_{ij} \\ &= \frac{1}{z} \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbb{E} \left[\frac{\partial \mathbf{Q}_{kj}}{\partial \mathbf{X}_{ik}} \right] - \frac{1}{z} \delta_{ij} \\ &= -\frac{1}{z} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\mathbf{Q}_{ki} \mathbf{Q}_{kj} + \mathbf{Q}_{kk} \mathbf{Q}_{ij}] - \frac{1}{z} \delta_{ij} \\ &= -\frac{1}{z} \frac{1}{n} \mathbb{E}[[\mathbf{Q}^2]_{ij} + \mathbf{Q}_{ij} \cdot \text{tr} \mathbf{Q}] - \frac{1}{z} \delta_{ij} \end{aligned}$$

which can be summarized in matrix form as

$$\mathbb{E}[\mathbf{Q}] = -\frac{1}{z} \frac{1}{n} \mathbb{E}[\mathbf{Q}^2] - \frac{1}{z} \mathbb{E}[\mathbf{Q}] \cdot \frac{1}{n} \text{tr} \mathbb{E}[\mathbf{Q}] - \frac{1}{z} \mathbf{I}_n + o_{\|\cdot\|}(1), \tag{2.32}$$

where we used the fact that $\frac{1}{n} \text{tr} \mathbf{Q} - \frac{1}{n} \text{tr} \mathbb{E} \mathbf{Q} \xrightarrow{\text{a.s.}} 0$ as $n \rightarrow \infty$ and can thus be asymptotically “taken out of the expectation.”

Since the first matrix on the right-hand side has asymptotically vanishing operator norm (of order $O(n^{-1})$) as $n, p \rightarrow \infty$,¹⁵ we reach

$$\mathbb{E}[\mathbf{Q}] = -\frac{1}{z} \left(1 + \frac{1}{z} \frac{1}{n} \text{tr} \mathbb{E}[\mathbf{Q}] \right)^{-1} \mathbf{I}_n + o_{\|\cdot\|}(1)$$

¹⁵ Again, we could even more simply have exploited Remark 2.5 to not even *produce* the term $\mathbb{E}[\mathbf{Q}_{ki} \mathbf{Q}_{kj}]$ in the early development of the calculus.

which, after taking the trace and using $\frac{1}{n} \operatorname{tr} \mathbb{E}[\mathbf{Q}(z)] - m(z) \rightarrow 0$, gives the limiting formula

$$m^2(z) + zm(z) + 1 = 0.$$

The rest of the development is then identical to the Bai–Silverstein approach. \square

2.2.3 Large-Dimensional Sample Covariance Matrices and Generalized Semicircles

The Marčenko–Pastur and semicircle theorems have long been the gold-standard in both theoretical and applied random matrix theory, in the sense that most mathematical studies and practical results concerned the Wishart and Wigner random matrix models.¹⁶ But the assumption of (the columns of) data \mathbf{X} having i.i.d., let alone standard Gaussian, entries has its limitation. In statistics where one is interested in the sample covariance matrix $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$, it is expected that the columns $\mathbf{x}_i \in \mathbb{R}^p$ of \mathbf{X} exhibit a correlation structure and even be nonnecessarily independent (in particular, when they are samples from a time series). In graph theory, where the affinity matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$ is the central object of study, one may wish to model graph patterns, degree heterogeneity, community structures, etc., which go against the i.i.d. (Bernoulli) assumption of so-called Erdős–Rényi graphs.

This section introduces generalizations of Marčenko–Pastur and semicircle theorems that go beyond the i.i.d. entries setting, to a level that is convenient to machine learning applications.¹⁷ As an example, in a machine learning classification context, \mathbf{X} will often be subdivided into subblocks that correspond to different classes, so as to model the existence of classes or communities within the data.

Large Sample Covariance Matrix Model and its Generalizations

Our first result generalizes the Marčenko–Pastur law, Theorem 2.4, to sample covariance matrices and is originally due to a long line of works by Silverstein and Bai [1995].

Theorem 2.6 (Sample covariance matrix, Silverstein and Bai [1995]). *Let $\mathbf{X} = \mathbf{C}^{\frac{1}{2}} \mathbf{Z} \in \mathbb{R}^{p \times n}$ with symmetric nonnegative definite $\mathbf{C} \in \mathbb{R}^{p \times p}$ of bounded operator norm (i.e., $\limsup_p \|\mathbf{C}\| < \infty$),¹⁸ $\mathbf{Z} \in \mathbb{R}^{p \times n}$ having independent zero mean and unit variance*

¹⁶ Among those studies are generalizations of the data model assumptions to matrices \mathbf{X} with dependent entries [Pajor and Pastur, 2009], refined studies and characterization of the limiting spectra [Silverstein and Choi, 1995] (to be discussed later in Section 2.3), deeper considerations on the local behavior of eigenvalues [Johnstone, 2001, 2008] (that will be briefly discussed in Section 2.5), just to name a few.

¹⁷ A host of other results for more elaborate random matrix models exists in the literature. Many are gathered in the books [Tulino and Verdú, 2004, Couillet and Debbah, 2011]: These books particularly focus on applications to wireless communication. Some of these results have effectively been reused to form the base ground of the current wave of machine learning-oriented random matrix models.

¹⁸ In the original article [Silverstein and Bai, 1995], the constraint on the bounded norm of $\|\mathbf{C}\|$ is relaxed and unnecessary. Yet, this complicates the proof and is never of actual use for the purpose of this book.

entries satisfying some light tail condition. Then, as $n, p \rightarrow \infty$ with $p/n \rightarrow c \in (0, \infty)$, letting $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{X}\mathbf{X}^\top - z\mathbf{I}_p)^{-1}$ and $\tilde{\mathbf{Q}}(z) = (\frac{1}{n}\mathbf{X}^\top\mathbf{X} - z\mathbf{I}_n)^{-1}$, we have

$$\begin{aligned} \mathbf{Q}(z) &\leftrightarrow \bar{\mathbf{Q}}(z) = -\frac{1}{z}(\mathbf{I}_p + \tilde{m}_p(z)\mathbf{C})^{-1}, \\ \tilde{\mathbf{Q}}(z) &\leftrightarrow \bar{\tilde{\mathbf{Q}}}(z) = \tilde{m}_p(z)\mathbf{I}_n, \end{aligned}$$

where $(z, \tilde{m}_p(z))$ is the unique solution in $\mathcal{Z}(\mathbb{C} \setminus \mathbb{R}^+)$ of¹⁹

$$\tilde{m}_p(z) = \left(-z + \frac{1}{n} \text{tr} \mathbf{C}(\mathbf{I}_p + \tilde{m}_p(z)\mathbf{C})^{-1}\right)^{-1}. \tag{2.33}$$

In particular, if the empirical spectral measure of \mathbf{C} converges, that is, $\mu_{\mathbf{C}} \rightarrow \nu$ as $p \rightarrow \infty$, then $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^\top} \xrightarrow{\text{a.s.}} \mu$, $\mu_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}} \xrightarrow{\text{a.s.}} \tilde{\mu}$ as $p, n \rightarrow \infty$ where $\mu, \tilde{\mu}$ are the unique measures having Stieltjes transforms $m(z)$ and $\tilde{m}(z)$, respectively, with

$$m(z) = \frac{1}{c}\tilde{m}(z) + \frac{1-c}{cz}, \quad \tilde{m}(z) = \left(-z + c \int \frac{t\nu(dt)}{1 + \tilde{m}(z)t}\right)^{-1}. \tag{2.34}$$

Before diving into the proof of Theorem 2.6, a few remarks are in order to better understand the statement of the theorem.

Remark 2.6 (On the implicit statement). As opposed to Theorem 2.4, the statement of the theorem is here implicit in the sense that μ is only defined through $m_\mu(z)$, itself implicitly defined as the solution of a fixed-point equation. The main reason for the explicit nature of Theorem 2.4 is that Equation (2.14), which provides the connection between $m(z)$ and a function of itself, boils down to a quadratic equation in $m(z)$, which can be solved explicitly and from which the inverse Stieltjes transform, Theorem 2.1, can be applied. Due to the presence of \mathbf{C} , in the present situation, the form equivalent to (2.14) here remains implicit. This will in fact be the case of almost all generalizations of the Marčenko–Pastur and semicircle theorems to be introduced in this book.

Note importantly that the uniqueness of the pair $(z, \tilde{m}_p(z))$ is stated within the set $\mathcal{Z}(\mathbb{C} \setminus \mathbb{R}^+)$, see Notation 2. In particular, for $z \in \mathbb{C}^+$ belonging to the upper half of the complex plane, there exists a unique $\tilde{m}_p(z) \in \mathbb{C}^+$ solution to the implicit equation; however, nothing prevents the existence of another solution (say in $\mathbb{C}^- = \{z \in \mathbb{C} \mid \Im[z] < 0\}$) to exist: This solution would not correspond to the sought-for $\tilde{m}_p(z)$. Possibly most importantly, we will see in Section 2.3 that, for $(z, \tilde{m}_p(z)) \in \{\mathbb{R}^+ \setminus \text{supp}(\mu)\} \times \mathbb{R}$ (a set excluded from $\mathcal{Z}(\mathbb{C} \setminus \mathbb{R}^+)$ but where $(z, \tilde{m}_p(z))$ can be formally defined by continuity), there may exist multiple solutions to the implicit equation! Fortunately, we will see that, here again, the correct solution can be identified.

¹⁹ Note that we denote the Stieltjes transform $\tilde{m}_p(z)$ with an additional subscript p , since, unlike Theorem 2.4, $\tilde{m}_p(z)$ is here defined as a function of the finite dimensional matrix \mathbf{C} , rather than as a function of the limiting spectral measure of \mathbf{C} . In particular, $\tilde{m}_p(z)$ needs *not* have a well-defined limit as $n, p \rightarrow \infty$. This again confirms the technical advantage of deterministic equivalents over limits (see again Definition 4): $\tilde{m}_p(z)$, instead of being a limit, is an increasingly accurate deterministic approximation of its *random* counterpart $\frac{1}{n} \text{tr} \tilde{\mathbf{Q}}(z)$, as n, p grow large.

Another fortunate realization is that the sought-for $\tilde{m}_p(z)$ solution also often happens to be the only “stable” one, in the sense that it will often be the only one discovered by numerical methods. See Remark 2.7 for detail.

In many applications, the value of $\tilde{m}_p(z)$ will rarely be a priority. Renaming $\tilde{\delta}_p(z) = \tilde{m}_p(z)$, we will instead be more often interested in the quantity $\delta_p(z) \equiv \frac{1}{n} \text{tr} \mathbf{C} \tilde{\mathbf{Q}}(z) = \frac{1}{n} \text{tr} \mathbf{C} (-z[\mathbf{I}_p + \tilde{\delta}_p(z)\mathbf{C}])^{-1}$ (in the vast majority of cases, for $z = -\gamma$, $\gamma \geq 0$ some deterministic parameter) which, from Lemma 2.11, corresponds to a deterministic equivalent for $\frac{1}{n} \mathbf{x}_0^T \mathbf{Q}(z) \mathbf{x}_0$ where $\mathbf{x}_0 = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_0$ for some $\mathbf{z}_0 \in \mathbb{R}^p$ independent of \mathbf{Z} having i.i.d. zero mean and unit variance entries: this quantity appears in the analysis of most regularized (not necessarily linear) regression problems. Interestingly, from the theorem statement, it can be checked that $\delta_p(z)$ satisfies the following very elegant symmetrically coupled equation

$$\begin{cases} \delta_p(z) = \frac{1}{n} \text{tr} \mathbf{C} (-z[\mathbf{I}_p + \tilde{\delta}_p(z)\mathbf{C}])^{-1} \\ \tilde{\delta}_p(z) = \frac{1}{n} \text{tr} \mathbf{I}_n (-z[\mathbf{I}_n + \delta_p(z)\mathbf{I}_n])^{-1} = -\frac{1}{z} \frac{1}{1+\delta_p(z)}. \end{cases} \tag{2.35}$$

Theorem 2.7 below will generalize this expression to the so-called *bi-correlated* model $\mathbf{C}^{\frac{1}{2}} \mathbf{Z} \tilde{\mathbf{C}}^{\frac{1}{2}}$ with \mathbf{I}_n replaced by an arbitrary nonnegative definite $\tilde{\mathbf{C}} \in \mathbb{R}^{n \times n}$ in the coupled equation above.

Remark 2.7 (Numerical evaluation of $m(z)$). *Due to its implicit nature, determining $m(z)$ for $z \in \mathbb{C} \setminus \mathbb{R}^+$ requires to solve an implicit equation. Using contraction and analyticity arguments, it can be shown that the standard fixed-point algorithm converges, that is,*²⁰

$$m(z) = \lim_{\ell \rightarrow \infty} m^{(\ell)}(z)$$

with say $\tilde{m}^{(0)}(z) = 0$ and for $\ell \geq 0$

$$m^{(\ell)}(z) = \frac{1}{c} \tilde{m}^{(\ell)}(z) + \frac{1-c}{cz}, \quad \tilde{m}^{(\ell+1)}(z) = \left(-z + c \int \frac{tv(dt)}{1 + \tilde{m}^{(\ell)}(z)t} \right)^{-1}, \tag{2.36}$$

or the equivalent finite-dimensional version with \mathbf{C} in (2.33).

One must be careful here that, since $m(z)$ is not formally defined for $z \in \text{supp}(\mu)$, the above argument does not hold in this set. Yet, the argument extends to $(\text{supp}(\mu), \infty)$, where the fixed-point iteration above is also numerically stable, but trying to solve (2.36) for $m(z)$ with $z \in \text{supp}(\mu)$ numerically leads to a nonconverging $m^{(\ell)}(z)$ sequence. This last remark can be effectively used in practice to numerically determine the right-edge $\text{supp}(\mu)$ of the support as being the smallest $z > 0$, starting from $+\infty$, for which the fixed-point iteration fails to converge (this can be done fast by dichotomy, starting from a left value $z_- > 0$ known to belong to the support and a large enough right value z_+).

²⁰ When carefully initialized, the convergence to the desired solution of standard fixed-point equations holds more generally (beyond the sample covariance model); see Couillet and Debbah [2011, Chapters 12–15] for examples of more involved models.

Numerically, when evaluating $m(z)$ for $z \in \mathbb{C}^+$ close to the real axis (say for $z = x + i\epsilon$, $|\epsilon| \ll 1$), the convergence can appear to be quite slow for $x \in \text{supp}(\mu)$. A convenient workaround is to sequentially evaluate $m(z)$ for all z s of the form $x + i\epsilon$, starting from some $z_0 = x_0 + i\epsilon$ away from the support, that is, for $x_0 \notin \text{supp}(\mu)$, then moving on to $z_1 = (x_0 \pm \epsilon') + i\epsilon$, then $z_2 = (x_0 \pm 2\epsilon') + i\epsilon$, etc., for some $\epsilon' \in \mathbb{R}$ small and, importantly, to systematically initialize the fixed-point iterations at position z_i with the value $m(z_{i-1})$ obtained at the previous position. Proceeding this way, the fixed-point iterations of $m(z_i)$ with $\Re[z_i] \in \text{supp}(\mu)$ are initialized close to the (non-real) solution and the convergence is in generally much faster than, for instance, the fixed initialization $m^{(0)}(z_i) = 0$. (However, note that the procedure may fail close to a mass of the spectrum of μ , typically at $z = 0$, and may keep accumulating errors if it happens to fail to converge at any given position of the spectrum.)

As a consequence of Remark 2.7, one can now numerically solve the implicit equation in Theorem 2.6 to draw, again numerically, the (limiting) spectrum μ .

Remark 2.8 (Drawing μ). *As shall be seen in Section 2.3, the limiting measure μ in Theorem 2.6 admits a density, which, from the inverse Stieltjes transform formula in Theorem 2.1 and Remark 2.7 above, can be approximated by solving for $m(z)$ with $z \in \mathbb{R} + i\epsilon$ for some $\epsilon > 0$ small (say $\epsilon = 10^{-5}$) and then retrieving the density at x as $\frac{1}{\pi} \Im[m(x + i\epsilon)]$.*

This procedure, however, only allows for a numerical approximation (rather than a theoretical evaluation) of μ and of its support (in particular, the support consists approximately in all values of x s such that $|\frac{1}{\pi} \Im[m(x + i\epsilon)]| \sim \epsilon \ll 1$). Section 2.3 will go beyond this imprecise numerical approach and provide an exact determination of (i) the limit $\lim_{z \in \mathbb{C}^+ \rightarrow x \in \mathbb{R} \setminus \{0\}} m(z)$ for all $x \in \mathbb{R} \setminus \{0\}$ and (ii) the support of μ .²¹

Figure 2.4 depicts the empirical versus limiting behavior of $\mu_{\frac{1}{n} \mathbf{X}\mathbf{X}^T}$ for \mathbf{C} having three distinct and evenly numerous eigenvalues. In this particular setting, the limiting spectrum is composed of several connected components, with shapes akin to the Marčenko–Pastur law. For sufficiently distinct eigenvalues of \mathbf{C} , these components are disjoint (Figure 2.4(a)) while for close eigenvalues they tend to merge (Figure 2.4(b)), and for $n < p$ a Dirac mass at zero is observed and the eigenvalues spread out even further into a single large component (Figure 2.4(c)).

Remark 2.9 (Deterministic equivalent for $\mu_{\frac{1}{n} \mathbf{X}\mathbf{X}^T}$). *The convergence result $\mu_{\frac{1}{n} \mathbf{X}\mathbf{X}^T} \xrightarrow{\text{a.s.}} \mu$ in Theorem 2.6 demands that there exists a limit v to which $\mu_{\mathbf{C}}$ converges as $p \rightarrow \infty$: this may not be practically meaningful. In generalized versions*

²¹ One may be surprised at the implicit statement that $\lim_{z \in \mathbb{C}^+ \rightarrow x \in \mathbb{R} \setminus \{0\}} m(z)$ exists for all $x \in \mathbb{R} \setminus \{0\}$, so in particular for $x \in \text{supp}(\mu)$ while we also stated, at the very beginning of this section in Definition 3, that $m(x) = \int (t - x)^{-1} \mu(dt)$ is *not* formally defined for $x \in \text{supp}(\mu)$. This is not a contradiction and is, we recall, at the core of the inverse Stieltjes transform formula in Theorem 2.1: The spectrum μ is precisely determined by looking at $\Im[m(z)]/\pi$ for z complex but arbitrarily close to the real axis. We will see in Section 2.3 that, at least for the sample covariance matrix model, $\lim_{z \in \mathbb{C}^+ \rightarrow x \in \mathbb{R} \setminus \{0\}} m(z)$ (as well as $\lim_{z \in \mathbb{C}^- \rightarrow x \in \mathbb{R} \setminus \{0\}} m(z)$ but whose value may be different!) indeed exists, while $m(x)$ itself need not be defined.

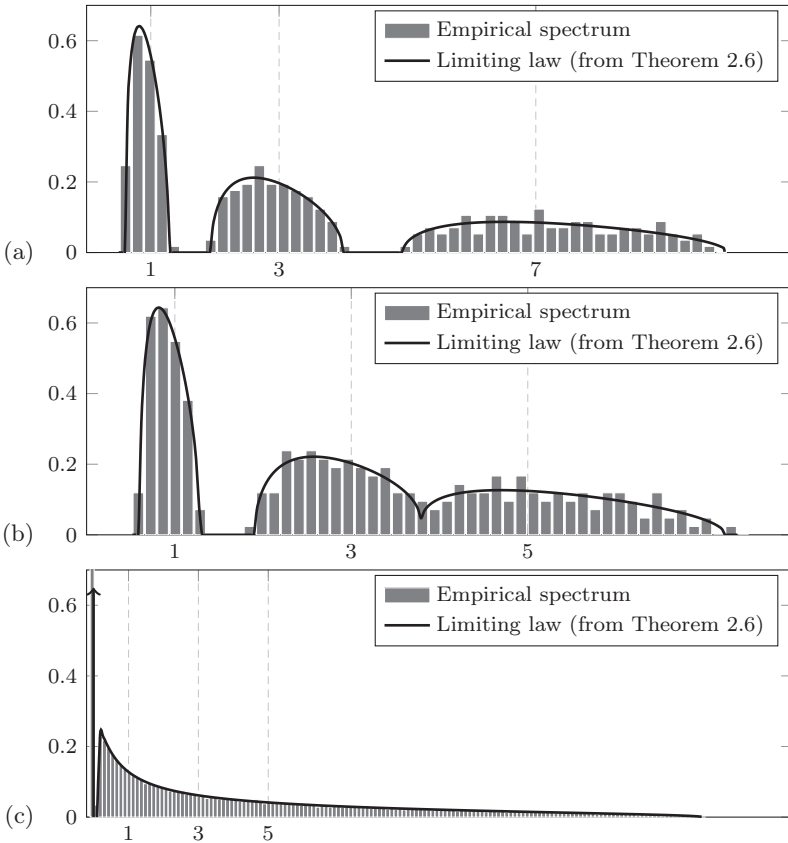


Figure 2.4 Histogram of the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$, $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z} \in \mathbb{R}^{p \times n}$, $[\mathbf{Z}]_{ij} \sim \mathcal{N}(0,1)$, $n = 3000$; for $p = 300$ and \mathbf{C} having spectral measure $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_7)$ (a), $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$ (b) and $p = 4500$ with $\mu_{\mathbf{C}} = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$ (c). Code on web: MATLAB and Python.

of Theorem 2.6 (see, for example, Theorem 2.8), even if the spectral measure of the population covariance matrix does converge, $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^T}$ may not have a limit.

One may instead consider the deterministic equivalent μ_p for $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^T}$, which is a sequence of probability measures for which $\text{dist}(\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^T}, \mu_p) \xrightarrow{\text{a.s.}} 0$ for some distance between probability measure (for instance, such that $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^T} - \mu_p \xrightarrow{\text{a.s.}} 0$ vaguely, so that for every bounded and continuous function f we have $\int f d\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^T} - \int f d\mu_p \xrightarrow{\text{a.s.}} 0$) as $n, p \rightarrow \infty$.

Practically speaking, since the data dimension p is in general a fixed quantity and \mathbf{C} a given covariance matrix (rather than specific values in a growing sequence of p s and \mathbf{C} s), one will always consider that the “effective” limiting measure ν actually coincides with (or is “frozen” to) $\mu_{\mathbf{C}} = \frac{1}{p} \sum_{i=1}^p \delta_{\lambda_i(\mathbf{C})}$.

Sketch of proof of Theorem 2.6. The proof of Theorem 2.6 generally follows the same line of arguments as that of Theorem 2.4. The main difference is that (2.12) here

becomes

$$\frac{1}{p} \mathbf{x}_i^\top \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q} \mathbf{x}_i = \frac{\frac{1}{p} \mathbf{x}_i^\top \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q}_{-i} \mathbf{x}_i}{1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i} = \frac{\frac{1}{p} \operatorname{tr} \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q}_{-i} \mathbf{C}}{1 + \frac{1}{n} \operatorname{tr} \mathbf{Q}_{-i} \mathbf{C}} + o(1),$$

where denoting $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$ for \mathbf{z}_i the i th column of $\mathbf{Z} \in \mathbb{R}^{p \times n}$ having independent zero mean and unit variance entries, we have by Lemma 2.11 that

$$\frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i = \frac{1}{n} \mathbf{z}_i^\top \mathbf{C}^{\frac{1}{2}} \mathbf{Q}_{-i} \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i = \frac{1}{n} \operatorname{tr} \mathbf{Q}_{-i} \mathbf{C} + o(1).$$

Again with Lemma 2.9 and the fact that $\frac{1}{n} \operatorname{tr} \mathbf{Q}_{-i} \mathbf{C}$ is bounded, we obtain the approximation

$$\frac{1}{p} \operatorname{tr} (\mathbf{F} + z \mathbf{I}_p) \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q} = \frac{\frac{1}{p} \operatorname{tr} \mathbf{C} \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q}}{1 + \frac{1}{n} \operatorname{tr} \mathbf{Q} \mathbf{C}} + o(1),$$

that should hold for any \mathbf{A} of unit norm, with $\mathbf{F}^{-1}(z) = \bar{\mathbf{Q}}(z)$ the sought-for deterministic equivalent, which then must admit the form

$$\mathbf{F}(z) = \frac{\mathbf{C}}{1 + \frac{1}{n} \operatorname{tr} \mathbf{Q} \mathbf{C}} - z \mathbf{I}_p + o_{\|\cdot\|}(1),$$

for the previous approximation to hold. Unlike in the proof of the Marčenko–Pastur theorem, Theorem 2.4, we see here the new term $\frac{1}{n} \operatorname{tr} \mathbf{Q} \mathbf{C}$ appears, which thus needs to be studied. Interestingly, note that taking $\mathbf{A} = \mathbf{C}$ in $\frac{1}{n} \operatorname{tr} \mathbf{A} (\mathbf{Q} - \bar{\mathbf{Q}}) \xrightarrow{\text{a.s.}} 0$ induces the following closed-form equation:

$$\frac{1}{n} \operatorname{tr} \mathbf{C} \mathbf{Q} = \frac{1}{n} \operatorname{tr} \mathbf{C} \bar{\mathbf{Q}} + o(1) = \frac{1}{n} \operatorname{tr} \mathbf{C} \left(-z \mathbf{I}_p + \frac{\mathbf{C}}{1 + \frac{1}{n} \operatorname{tr} \mathbf{C} \bar{\mathbf{Q}}} \right)^{-1} + o(1) \tag{2.37}$$

from which we obtain

$$\tilde{m}_p(z) = \left(-z + \frac{1}{n} \operatorname{tr} \mathbf{C} (\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C})^{-1} \right)^{-1},$$

if we denote $\tilde{m}_p(z) = -\frac{1}{z} \left(1 + \frac{1}{n} \operatorname{tr} \mathbf{C} \bar{\mathbf{Q}}(z) \right)^{-1}$, as requested.²²

With a deterministic equivalent $\bar{\mathbf{Q}} = \mathbf{F}^{-1}$ for \mathbf{Q} at hand, a corresponding deterministic equivalent for $\tilde{\mathbf{Q}} = \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} - z \mathbf{I}_n \right)^{-1}$ follows from the direct observation that $\tilde{\mathbf{Q}} = \frac{1}{z} \frac{1}{n} \mathbf{X}^\top \mathbf{Q} \mathbf{X} - \frac{1}{z} \mathbf{I}_n$, so that

$$\begin{aligned} [\tilde{\mathbf{Q}}]_{ij} &= \frac{1}{z} \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q} \mathbf{x}_j - \frac{1}{z} \delta_{ij} = \frac{1}{z} \frac{\frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_j}{1 + \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i} \mathbf{x}_i} - \frac{1}{z} \delta_{ij} \\ &= \frac{1}{z} \frac{\frac{1}{n} \operatorname{tr} \mathbf{C} \bar{\mathbf{Q}}}{1 + \frac{1}{n} \operatorname{tr} \mathbf{C} \bar{\mathbf{Q}}} \delta_{ij} - \frac{1}{z} \delta_{ij} + o(1) \\ &= -\frac{1}{z} \left(1 + \frac{1}{n} \operatorname{tr} \mathbf{C} \bar{\mathbf{Q}} \right)^{-1} \delta_{ij} + o(1) = \tilde{m}_p(z) \delta_{ij} + o(1), \end{aligned}$$

²² Note that we implicitly used here the fact that $\|\mathbf{C}\|$ is bounded.

and thus, for $\mathbf{A} \in \mathbb{R}^{n \times n}$ deterministic of bounded norm, applying the operator $\frac{1}{n} \sum_{i,j=1}^n [\mathbf{A}]_{ij} \cdot$ on both sides (one must be careful to ensure that the entry-wise “ $+o(1)$ ” approximation still holds under this operator), we confirm that $\tilde{\mathbf{Q}}(z) \equiv \tilde{m}_p(z) \mathbf{I}_n$ is indeed a deterministic equivalent for $\tilde{\mathbf{Q}}$. \square

Remark 2.10 (On singular population covariances). *It is interesting to note from Theorem 2.6 that, if the population covariance \mathbf{C} contains some zero eigenvalues, for example, if $\mu_{\mathbf{C}} \rightarrow \nu$ as $p \rightarrow \infty$ with*

$$\nu(dx) = (1 - c_\nu) \delta_0(x) + c_\nu \tilde{\nu}(dx)$$

for $c_\nu \in (0, 1)$ and $\tilde{\nu}$ some probability measure, then (as properly shown in Silverstein and Choi [1995]) $\tilde{\mu}(\{0\}) = \max(0, 1 - cc_\nu)$. This further implies

$$\mu(\{0\}) = \begin{cases} 1 - c_\nu & \text{for } cc_\nu \leq 1, \\ 1 - c^{-1} & \text{otherwise.} \end{cases}$$

This result diverges from the systematic $\mu(\{0\}) = \max(0, 1 - c^{-1})$ in the Marčenko–Pastur scenario, and takes into consideration the intrinsic dimension $c_\nu p$ of the random vector $\mathbf{C}^{\frac{1}{2}} \mathbf{z}_i \in \mathbb{R}^p$.

As we shall see later in this book, for machine learning applications, the data covariance structure \mathbf{C} may contain a wide range of very small eigenvalues, a behavior suggesting that the data representation is of much smaller effective dimension. It is interesting to observe that Theorem 2.6, in its expression in (2.33), in fact does not depend on the ratio p/n itself but on $\frac{1}{n} \text{tr} \mathbf{C}(\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C})^{-1}$: The effective data dimension is thus encapsulated within \mathbf{C} in (a nontrivial manner in) the fixed-point expression.

When the data $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ arise from a time series, or when each data sample is weighted by an independent coefficient (as shall be seen in Section 3.3 on robust statistical methods), the sample covariance matrix model is not sufficiently expressive but can be generalized to the so-called *bi-correlated* (or *separable covariance*) model as follows,

$$\frac{1}{n} \mathbf{C}^{\frac{1}{2}} \mathbf{Z} \tilde{\mathbf{C}} \mathbf{Z}^T \mathbf{C}^{\frac{1}{2}} \tag{2.38}$$

for $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $\tilde{\mathbf{C}} \in \mathbb{R}^{n \times n}$ two nonnegative definite matrices and $[\mathbf{Z}]_{ij}$ i.i.d. random variables with zero mean and unit variance. In particular, for \mathbf{Z} Gaussian and $\tilde{\mathbf{C}}^{\frac{1}{2}}$ Toeplitz (i.e., such that $[\tilde{\mathbf{C}}^{\frac{1}{2}}]_{ij} = \alpha_{|i-j|}$ for some sequence $\alpha_0, \dots, \alpha_{n-1}$), the columns of $\mathbf{Z} \tilde{\mathbf{C}}^{\frac{1}{2}}$ model a first-order auto-regressive process [Hamilton, 1994].²³

²³ In passing, Toeplitz matrices involved in time series analyses also exhibit interesting large-dimensional behavior. As an instance, Gray [2006] showed that, under some decay condition on the sequence $\{\alpha_i\}_{i=0}^{n-1}$, their spectral behavior is the same as that of equivalent *circulant* matrices, the latter having the nice property to be diagonalizable in the Fourier basis: The asymptotic eigenvalues of the Toeplitz matrix are, in particular, the coefficients of the discrete Fourier transform of the series $\{\alpha_i\}_{i=0}^{n-1}$.

For this model, we have the following theorem.

Theorem 2.7 (Bi-correlated model, separable covariance model, [Paul and Silverstein, 2009]). *Let $\mathbf{Z} \in \mathbb{R}^{p \times n}$ be a random matrix with i.i.d. zero mean, unit variance and light tail entries, and $\mathbf{C} \in \mathbb{R}^{p \times p}$, $\tilde{\mathbf{C}} \in \mathbb{R}^{n \times n}$ be symmetric nonnegative definite matrices with bounded operator norm. Then, as $n, p \rightarrow \infty$ with $p/n \rightarrow c \in (0, \infty)$, letting $\mathbf{Q}(z) = (\frac{1}{n} \mathbf{C}^{\frac{1}{2}} \mathbf{Z} \tilde{\mathbf{C}} \mathbf{Z}^T \mathbf{C}^{\frac{1}{2}} - z \mathbf{I}_p)^{-1}$ and $\tilde{\mathbf{Q}}(z) = (\frac{1}{n} \tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{Z}^T \mathbf{C} \mathbf{Z} \tilde{\mathbf{C}}^{\frac{1}{2}} - z \mathbf{I}_n)^{-1}$, we have*

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = -\frac{1}{z} (\mathbf{I}_p + \tilde{\delta}_p(z) \mathbf{C})^{-1}$$

$$\tilde{\mathbf{Q}}(z) \leftrightarrow \bar{\tilde{\mathbf{Q}}}(z) = -\frac{1}{z} (\mathbf{I}_n + \delta_p(z) \tilde{\mathbf{C}})^{-1}$$

with $(z, \delta_p(z)), (z, \tilde{\delta}_p(z)) \in \mathcal{Z}(\mathbb{C} \setminus \mathbb{R}^+)$ unique solutions to

$$\delta_p(z) = \frac{1}{n} \text{tr} \mathbf{C} \bar{\mathbf{Q}}(z), \quad \tilde{\delta}_p(z) = \frac{1}{n} \text{tr} \tilde{\mathbf{C}} \bar{\tilde{\mathbf{Q}}}(z).$$

In particular, if $\mu_{\mathbf{C}} \rightarrow \nu$ and $\mu_{\tilde{\mathbf{C}}} \rightarrow \tilde{\nu}$, then

$$\mu_{\frac{1}{n} \mathbf{C}^{\frac{1}{2}} \mathbf{Z} \tilde{\mathbf{C}} \mathbf{Z}^T \mathbf{C}^{\frac{1}{2}}} \xrightarrow{\text{a.s.}} \mu, \quad \mu_{\frac{1}{n} \tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{Z}^T \mathbf{C} \mathbf{Z} \tilde{\mathbf{C}}^{\frac{1}{2}}} \xrightarrow{\text{a.s.}} \tilde{\mu},$$

where $\mu, \tilde{\mu}$ are defined via their Stieltjes transforms $m(z)$ and $\tilde{m}(z)$ given by

$$m(z) = -\frac{1}{z} \int \frac{\nu(dt)}{1 + \tilde{\delta}(z)t}, \quad \tilde{m}(z) = -\frac{1}{z} \int \frac{\tilde{\nu}(dt)}{1 + \delta(z)t},$$

where $(z, \delta(z)), (z, \tilde{\delta}(z))$ are the unique solutions in $\mathcal{Z}(\mathbb{C} \setminus \mathbb{R}^+)$ to

$$\delta(z) = -\frac{c}{z} \int \frac{t\nu(dt)}{1 + \tilde{\delta}(z)t}, \quad \tilde{\delta}(z) = -\frac{1}{z} \int \frac{t\tilde{\nu}(dt)}{1 + \delta(z)t}.$$

Sketch of proof of Theorem 2.7. For simplicity and readability, only the case where both \mathbf{C} and $\tilde{\mathbf{C}}$ are diagonal is presented here.²⁴ In this case, similar to the decomposition performed in the proof of Theorem 2.6, one has the following symmetric re-expression of $\mathbf{Q}(z)$ and $\tilde{\mathbf{Q}}(z)$

$$\mathbf{Q}(z) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{C}^{\frac{1}{2}} \tilde{\mathbf{y}}_i (\mathbf{C}^{\frac{1}{2}} \tilde{\mathbf{y}}_i)^T - z \mathbf{I}_p \right)^{-1}$$

$$\tilde{\mathbf{Q}}(z) = \left(\frac{1}{n} \sum_{i=1}^p \tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{y}_i (\tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{y}_i)^T - z \mathbf{I}_n \right)^{-1}$$

where we denote $\tilde{\mathbf{y}}_i \in \mathbb{R}^p$ the i th column of $\mathbf{Z} \tilde{\mathbf{C}}^{\frac{1}{2}}$ and $\mathbf{y}_i \in \mathbb{R}^n$ the i th column of $\mathbf{Z}^T \mathbf{C}^{\frac{1}{2}}$ so that, for \mathbf{C} and $\tilde{\mathbf{C}}$ both diagonal, one has $\tilde{\mathbf{y}}_i = \tilde{\mathbf{C}}_{ii}^{\frac{1}{2}} \mathbf{z}_i$ and $\mathbf{y}_i = \mathbf{C}_{ii}^{\frac{1}{2}} \tilde{\mathbf{z}}_i$ with $\mathbf{z}_i \in \mathbb{R}^p$ the i th column and $\tilde{\mathbf{z}}_i \in \mathbb{R}^n$ the i th row of $\mathbf{Z} \in \mathbb{R}^{p \times n}$.

²⁴ Note that, if \mathbf{Z} is standard Gaussian, then $\mathbf{Z} \tilde{\mathbf{C}} \mathbf{Z}^T$ has the same distribution as $\mathbf{Z} \tilde{\mathbf{C}} \mathbf{U}^T \mathbf{Z}^T$ for any unitary matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ (since $\mathbf{Z} \sim \mathbf{Z} \mathbf{U}$ in law). We may then allow $\tilde{\mathbf{C}}$ to be diagonal by specifically choosing \mathbf{U} to be a matrix of eigenvectors of $\tilde{\mathbf{C}}$. By the universality of random matrix results with respect to the law of the independent entries of \mathbf{Z} (that we recall, can be rigorously established using, say, Lemma 2.15), this should be sufficient to retrieve the result for any \mathbf{Z} . The same remark symmetrically holds for \mathbf{C} .

As a consequence, with $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$ and $\tilde{\bar{\mathbf{Q}}}(z) = \tilde{\mathbf{F}}^{-1}(z)$, one obtains again with Lemmas 2.1 and 2.8 that

$$\begin{aligned} \mathbf{Q}(z) - \bar{\mathbf{Q}}(z) &= \mathbf{Q}(z) \left(\mathbf{F}(z) + z\mathbf{I}_p - \frac{1}{n} \sum_{i=1}^n \mathbf{C}^{\frac{1}{2}} \tilde{\mathbf{y}}_i (\mathbf{C}^{\frac{1}{2}} \tilde{\mathbf{y}}_i)^\top \right) \bar{\mathbf{Q}}(z) \\ &= \mathbf{Q}(\mathbf{F} + z\mathbf{I}_p) \bar{\mathbf{Q}} - \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{Q}_{-i} \mathbf{C}^{\frac{1}{2}} \tilde{\mathbf{C}}_{ii} \mathbf{z}_i \mathbf{z}_i^\top \mathbf{C}^{\frac{1}{2}} \bar{\mathbf{Q}}}{1 + \frac{1}{n} \tilde{\mathbf{C}}_{ii} \mathbf{z}_i^\top \mathbf{C}^{\frac{1}{2}} \mathbf{Q}_{-i} \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i}, \\ \tilde{\mathbf{Q}}(z) - \tilde{\bar{\mathbf{Q}}}(z) &= \tilde{\mathbf{Q}}(z) \left(\tilde{\mathbf{F}}(z) + z\mathbf{I}_n - \frac{1}{n} \sum_{i=1}^p \tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{y}_i (\tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{y}_i)^\top \right) \tilde{\bar{\mathbf{Q}}}(z) \\ &= \tilde{\mathbf{Q}}(\tilde{\mathbf{F}} + z\mathbf{I}_n) \tilde{\bar{\mathbf{Q}}} - \frac{1}{n} \sum_{i=1}^p \frac{\tilde{\mathbf{Q}}_{-i} \tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{C}_{ii} \tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^\top \tilde{\mathbf{C}}^{\frac{1}{2}} \tilde{\bar{\mathbf{Q}}}}{1 + \frac{1}{n} \mathbf{C}_{ii} \tilde{\mathbf{z}}_i^\top \tilde{\mathbf{C}}^{\frac{1}{2}} \tilde{\mathbf{Q}}_{-i} \tilde{\mathbf{C}}^{\frac{1}{2}} \tilde{\mathbf{z}}_i}, \end{aligned}$$

where we denote $\mathbf{Q}_{-i}(z) \equiv (\frac{1}{n} \sum_{j \neq i}^n \mathbf{C}^{\frac{1}{2}} \tilde{\mathbf{C}}_{jj} \mathbf{z}_j \mathbf{z}_j^\top \mathbf{C}^{\frac{1}{2}} - z\mathbf{I}_p)^{-1}$ and symmetrically $\tilde{\mathbf{Q}}_{-i}(z) \equiv (\frac{1}{n} \sum_{j \neq i}^p \tilde{\mathbf{C}}^{\frac{1}{2}} \mathbf{C}_{jj} \tilde{\mathbf{z}}_j \tilde{\mathbf{z}}_j^\top \tilde{\mathbf{C}}^{\frac{1}{2}} - z\mathbf{I}_n)^{-1}$, which are independent of \mathbf{z}_i and $\tilde{\mathbf{z}}_i$, respectively.

With this independence of \mathbf{Q}_{-i} on \mathbf{z}_i and $\tilde{\mathbf{Q}}_{-i}$ on $\tilde{\mathbf{z}}_i$, one deduces again with Lemma 2.11 that

$$\begin{aligned} \frac{1}{n} \tilde{\mathbf{C}}_{ii} \mathbf{z}_i^\top \mathbf{C}^{\frac{1}{2}} \mathbf{Q}_{-i} \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i &= \tilde{\mathbf{C}}_{ii} \cdot \frac{1}{n} \text{tr}(\mathbf{Q}_{-i} \mathbf{C}) + o(1), \\ \frac{1}{n} \mathbf{C}_{ii} \tilde{\mathbf{z}}_i^\top \tilde{\mathbf{C}}^{\frac{1}{2}} \tilde{\mathbf{Q}}_{-i} \tilde{\mathbf{C}}^{\frac{1}{2}} \tilde{\mathbf{z}}_i &= \mathbf{C}_{ii} \cdot \frac{1}{n} \text{tr}(\tilde{\mathbf{Q}}_{-i} \tilde{\mathbf{C}}) + o(1), \end{aligned}$$

so that $\mathbf{F}(z)$ and $\tilde{\mathbf{F}}(z)$ must take the followings forms

$$\begin{aligned} \mathbf{F}(z) &= \frac{1}{n} \sum_{i=1}^n \frac{\tilde{\mathbf{C}}_{ii} \cdot \mathbf{C}}{1 + \tilde{\mathbf{C}}_{ii} \cdot \frac{1}{n} \text{tr}(\mathbf{Q}_{-i} \mathbf{C})} - z\mathbf{I}_p = \frac{1}{n} \sum_{i=1}^n \frac{\tilde{\mathbf{C}}_{ii} \cdot \mathbf{C}}{1 + \tilde{\mathbf{C}}_{ii} \cdot \frac{1}{n} \text{tr} \mathbf{C} \bar{\mathbf{Q}}} - z\mathbf{I}_p + o_{\|\cdot\|}(1), \\ \tilde{\mathbf{F}}(z) &= \frac{1}{n} \sum_{i=1}^p \frac{\mathbf{C}_{ii} \cdot \tilde{\mathbf{C}}}{1 + \mathbf{C}_{ii} \cdot \frac{1}{n} \text{tr}(\tilde{\mathbf{Q}}_{-i} \tilde{\mathbf{C}})} - z\mathbf{I}_n = \frac{1}{n} \sum_{i=1}^p \frac{\mathbf{C}_{ii} \cdot \tilde{\mathbf{C}}}{1 + \mathbf{C}_{ii} \cdot \frac{1}{n} \text{tr} \tilde{\mathbf{C}} \tilde{\bar{\mathbf{Q}}}} - z\mathbf{I}_n + o_{\|\cdot\|}(1). \end{aligned}$$

Denoting $\delta_p(z) = \frac{1}{n} \text{tr} \mathbf{C} \bar{\mathbf{Q}}(z)$ and $\tilde{\delta}_p(z) = \frac{1}{n} \text{tr} \tilde{\mathbf{C}} \tilde{\bar{\mathbf{Q}}}(z)$, this can be further reduced to

$$\begin{aligned} \bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z) &= -\frac{1}{z} \left(\mathbf{I}_p - \frac{1}{z} \frac{1}{n} \sum_{i=1}^n \frac{\tilde{\mathbf{C}}_{ii}}{1 + \tilde{\mathbf{C}}_{ii} \delta_p(z)} \mathbf{C} \right)^{-1} + o_{\|\cdot\|}(1), \\ \tilde{\bar{\mathbf{Q}}}(z) = \tilde{\mathbf{F}}^{-1}(z) &= -\frac{1}{z} \left(\mathbf{I}_n - \frac{1}{z} \frac{1}{n} \sum_{i=1}^p \frac{\mathbf{C}_{ii}}{1 + \mathbf{C}_{ii} \tilde{\delta}_p(z)} \tilde{\mathbf{C}} \right)^{-1} + o_{\|\cdot\|}(1). \end{aligned}$$

To eventually close the loop and obtain the sought-for relation on $(\delta_p, \tilde{\delta}_p)$, one may plug the above approximation into the definition of δ_p and $\tilde{\delta}_p$ to obtain the following symmetric equation

$$\delta_p(z) = -\frac{1}{z} \frac{1}{n} \sum_{i=1}^p \frac{C_{ii}}{1 - \frac{1}{z} \frac{1}{n} \sum_{j=1}^n \frac{\tilde{C}_{jj} C_{ii}}{1 + \tilde{C}_{jj} \delta_p(z)}} + o(1),$$

$$\tilde{\delta}_p(z) = -\frac{1}{z} \frac{1}{n} \sum_{i=1}^n \frac{\tilde{C}_{ii}}{1 - \frac{1}{z} \frac{1}{n} \sum_{j=1}^p \frac{C_{jj} \tilde{C}_{ii}}{1 + C_{jj} \delta_p(z)}} + o(1),$$

which retrieves the expressions of Theorem 2.7. □

As already hinted at when commenting on Theorem 2.6 in (2.35), it is interesting to note the almost perfect symmetry in the equations for the resolvent and co-resolvent in the bi-correlated model. From a machine learning perspective, wherein $\mathbf{X} = \mathbf{C}^{\frac{1}{2}} \mathbf{Z} \tilde{\mathbf{C}}^{\frac{1}{2}}$ are the observed data, this symmetry between “space” and “time” correlations, or between the sample covariance matrix $\mathbf{X}\mathbf{X}^T$ and the Gram (kernel) matrix $\mathbf{X}^T\mathbf{X}$, will often allow for a natural connection between results in the spatial (e.g., PCA, subspace methods) and in the temporal (classification, regression) domains.

From a technical angle, by the trace lemma, Lemma 2.11, we immediately find that the functions $\delta_p(z)$ and $\tilde{\delta}_p(z)$ (which also happen to be Stieltjes transforms of finite measures on \mathbb{R}^+) are respectively deterministic equivalents for $\frac{1}{n} \mathbf{x}_0^T \mathbf{Q}(z) \mathbf{x}_0$ and $\frac{1}{n} \tilde{\mathbf{x}}_0^T \tilde{\mathbf{Q}}(z) \tilde{\mathbf{x}}_0$ for $\mathbf{x}_0 = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_0$, $\tilde{\mathbf{x}}_0 = \tilde{\mathbf{C}}^{\frac{1}{2}} \tilde{\mathbf{z}}_0$ and $\mathbf{z}_0 \in \mathbb{R}^p$, $\tilde{\mathbf{z}}_0 \in \mathbb{R}^n$ vectors of independent zero mean and unit variance entries, both independent of \mathbf{Z} . Similar to the remarks after Theorem 2.6, these quadratic forms will naturally arise in various applications of statistical inference and regression: particularly for $z = -\gamma$ with $\gamma \geq 0$ a regularization parameter, and \tilde{C}_{ii} (\tilde{C} will usually be diagonal) an effective weight parameter induced by the algorithm under study on data point $\mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$.

As pointed out above, the Gram matrix $\mathbf{X}^T\mathbf{X}$ is directly connected to kernel matrices of the type $\mathbf{K} = \{\mathbf{x}_i^T \mathbf{x}_j / p\}_{i,j=1}^n = \mathbf{X}^T\mathbf{X} / p$ (linear inner-product kernels) and $\mathbf{K} = \{\|\mathbf{x}_i - \mathbf{x}_j\|^2 / p\}_{i,j=1}^n$ (Euclidean distance kernels) since $\|\mathbf{x}_i - \mathbf{x}_j\|^2 / p = \|\mathbf{x}_i\|^2 / p + \|\mathbf{x}_j\|^2 / p - 2\mathbf{x}_i^T \mathbf{x}_j / p$, which also involves the matrix $\mathbf{X}^T\mathbf{X} / p$.²⁵ Assuming, as is the basic setting in a multi-class machine learning classification context, that the vectors \mathbf{x}_i arise from a mixture model, the following generalization of Theorem 2.6 is of more practical relevance to machine learning applications.

Theorem 2.8 (Sample covariance of k -class mixture models, [Benaych-Georges and Couillet, 2016]). *Let $\mathbf{X} = [\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(k)}] \in \mathbb{R}^{p \times n}$ with $\mathbf{X}^{(a)} = [\mathbf{x}_1^{(a)}, \dots, \mathbf{x}_{n_a}^{(a)}] \in \mathbb{R}^{p \times n_a}$ and $\mathbf{x}_i^{(a)} = \mathbf{C}_a^{\frac{1}{2}} \mathbf{z}_i^{(a)}$ for $\mathbf{z}_i^{(a)}$ a vector with i.i.d. zero mean, unit variance and light tail entries. Then, as $n_a, p \rightarrow \infty$ in such a way that k is fixed, $p/n \rightarrow c \in (0, \infty)$, and $n_a/n \rightarrow c_a \in (0, 1)$ for $a \in \{1, \dots, k\}$, letting $\mathbf{Q}(z) = (\frac{1}{n} \mathbf{X}\mathbf{X}^T - z\mathbf{I}_p)^{-1}$ and $\tilde{\mathbf{Q}}(z) = (\frac{1}{n} \mathbf{X}^T\mathbf{X} - z\mathbf{I}_n)^{-1}$, we have²⁶*

²⁵ The prefactor $1/p$ is necessary to ensure that the main eigenspectrum of \mathbf{K} remains of order $O(1)$ as p, n increase.

²⁶ Here, $\text{diag}\{\mathbf{v}_a\}_{a=1}^k$ is a diagonal matrix with the concatenated vector $\mathbf{v} = [\mathbf{v}_1^T, \dots, \mathbf{v}_k^T]$ on the diagonal; and $\mathbf{I}_{n_a} \in \mathbb{R}^{n_a}$ is the n_a -dimensional vector of all ones.

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = -\frac{1}{z} \left(\mathbf{I}_p + \sum_{a=1}^k c_a \tilde{g}_a(z) \mathbf{C}_a \right)^{-1}$$

$$\tilde{\mathbf{Q}}(z) \leftrightarrow \tilde{\bar{\mathbf{Q}}}(z) = \text{diag}\{\tilde{g}_a(z) \mathbf{1}_{n_a}\}_{a=1}^k$$

with $(z, \tilde{g}_a(z))$, $a \in \{1, \dots, k\}$, the unique solutions in $\mathcal{Z}(\mathbb{C} \setminus \mathbb{R}^+)$ to

$$\tilde{g}_a(z) = -\frac{1}{z} (1 + g_a(z))^{-1}, \quad g_a(z) = -\frac{1}{z} \frac{1}{n} \text{tr} \mathbf{C}_a \left(\mathbf{I}_p + \sum_{b=1}^k c_b \tilde{g}_b(z) \mathbf{C}_b \right)^{-1}.$$

Sketch of proof of Theorem 2.8. Similar to the proof of Theorem 2.6, we obtain, with the initial guess $\bar{\mathbf{Q}}(z) = \mathbf{F}^{-1}(z)$, that

$$\mathbf{Q} - \bar{\mathbf{Q}} = \mathbf{Q} \left(\mathbf{F} + z \mathbf{I}_p - \frac{1}{n} \sum_{a=1}^k \sum_{i=1}^{n_a} \mathbf{x}_i^{(a)} (\mathbf{x}_i^{(a)})^\top \right) \bar{\mathbf{Q}}$$

which, unlike in the proof of Theorem 2.6, contains a sum over a due to the different class covariances \mathbf{C}_a . To establish $\frac{1}{n} \text{tr} \mathbf{A}(\mathbf{Q} - \bar{\mathbf{Q}}) \xrightarrow{\text{a.s.}} 0$, one must have

$$\frac{1}{n} \text{tr}(\mathbf{F} + z \mathbf{I}_p) \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q} - \frac{1}{n} \sum_{a=1}^k \sum_{i=1}^{n_a} \frac{1}{n} (\mathbf{x}_i^{(a)})^\top \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q} \mathbf{x}_i^{(a)} \xrightarrow{\text{a.s.}} 0.$$

Applying Lemma 2.8 to remove the dependence in \mathbf{Q} of $\mathbf{x}_i^{(a)}$, together with Lemma 2.9, we deduce

$$\frac{1}{n} \sum_{a=1}^k \sum_{i=1}^{n_a} \frac{1}{n} (\mathbf{x}_i^{(a)})^\top \bar{\mathbf{Q}} \mathbf{A} \mathbf{Q} \mathbf{x}_i^{(a)} = \sum_{a=1}^k \frac{n_a}{n} \frac{\frac{1}{n} \text{tr} \mathbf{C}_a \bar{\mathbf{Q}} \mathbf{A} \bar{\mathbf{Q}}}{1 + \frac{1}{n} \text{tr} \bar{\mathbf{Q}} \mathbf{C}_a} + o(1),$$

so that \mathbf{F} must be written as the following sum over a :

$$\mathbf{F} = \sum_{a=1}^k c_a \frac{\mathbf{C}_a}{1 + \frac{1}{n} \text{tr} \bar{\mathbf{Q}} \mathbf{C}_a} - z \mathbf{I}_p + o_{\|\cdot\|}(1),$$

which produces the term $\frac{1}{n} \text{tr} \bar{\mathbf{Q}} \mathbf{C}_a$, $a = 1, \dots, k$. To identify these terms and close the loop, we take $\mathbf{A} = \mathbf{C}_b$ for each $b \in \{1, \dots, k\}$ to establish

$$\begin{aligned} \frac{1}{n} \text{tr} \mathbf{C}_b \mathbf{Q} &= \frac{1}{n} \text{tr} \mathbf{C}_b \bar{\mathbf{Q}} + o(1) \equiv g_b(z) + o(1) \\ &= \frac{1}{n} \text{tr} \mathbf{C}_b \left(-z \mathbf{I}_p + \sum_{a=1}^k c_a \frac{\mathbf{C}_a}{1 + \frac{1}{n} \text{tr} \bar{\mathbf{Q}} \mathbf{C}_a} \right)^{-1} + o(1) \\ &\equiv -\frac{1}{z} \frac{1}{n} \text{tr} \mathbf{C}_b \left(\mathbf{I}_p + \sum_{a=1}^k c_a \tilde{g}_a(z) \mathbf{C}_a \right) + o(1), \end{aligned}$$

where we denoted $\tilde{g}_a(z) \equiv -\frac{1}{z} (1 + \frac{1}{n} \text{tr} \bar{\mathbf{Q}} \mathbf{C}_a)^{-1} = -\frac{1}{z} (1 + g_a(z))^{-1}$, as desired. This thus produces a k -dimensional *vector* equation linking the $g_a(z)$ s rather than a scalar one as in the case of Theorem 2.6.

To finally derive a deterministic equivalent of $\bar{\mathbf{Q}}$ from that of \mathbf{Q} , we use again the fact that $\bar{\mathbf{Q}} = \frac{1}{z} \frac{1}{n} \mathbf{X}^\top \mathbf{Q} \mathbf{X} - \frac{1}{z} \mathbf{I}_n$ and therefore, indexing the set $\{1, \dots, n\}$ as $\{(1)1, \dots, (1)n_1, \dots, (k)1, \dots, (k)n_k\}$, we have

$$\begin{aligned} \mathbf{Q}_{(a)i,(b)j} &= \frac{1}{z} \frac{1}{n} (\mathbf{x}_i^{(a)})^\top \mathbf{Q} \mathbf{x}_j^{(b)} - \frac{1}{z} \delta_{(a)i,(b)j} \\ &= -\frac{1}{z} \left(1 + \frac{1}{n} \operatorname{tr} \bar{\mathbf{Q}} \mathbf{C}_a \right)^{-1} \delta_{(a)i,(b)j} + o(1) = \tilde{g}_a(z) \delta_{(a)i,(b)j} + o(1), \end{aligned}$$

which, after applying $\frac{1}{n} \operatorname{tr} \mathbf{A}(\cdot)$ on both sides for \mathbf{A} of unit norm, concludes the proof of Theorem 2.8. \square

With some further control, Theorem 2.8 may in fact be extended to $k = n$, that is, each data vector \mathbf{x}_i has its own, possibly distinct, covariance matrix, as shown in Wagner et al. [2012]. When the covariance matrices are diagonal, this is then equivalent to letting \mathbf{X} have a *variance profile*, that is, the entries $[\mathbf{X}]_{ij}$ s are all independent with zero mean and variance $\sigma_{ij}^2 \equiv [\mathbf{C}_i]_{jj}$ (with $\mathbf{C}_i = \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top]$), a setting studied in depth in Hachem et al. [2007] but originally found in Girko [2001].

The application of a variance profile to random matrices with independent entries finds an even more relevant application to Wigner matrices, as detailed next.

Generalized Semicircle Law with a Variance Profile

Similar to the large sample covariance matrix model, generalizations also exist for the Wigner semicircle law in Theorem 2.5. In the following theorem, a variance profile for the entries of the symmetric random matrix is considered.

Theorem 2.9 (Pastur and Shcherbina [2011]). *Let $\mathbf{X} \in \mathbb{R}^{n \times n}$ be symmetric and such that \mathbf{X}_{ij} , $j \geq i$, is of zero mean, bounded variance $\operatorname{Var}[\mathbf{X}_{ij}] = \sigma_{ij}^2$, and satisfies some light tail condition. Then, for $\mathbf{Q}(z) = (\mathbf{X}/\sqrt{n} - z\mathbf{I}_n)^{-1}$, we have*

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z), \quad \bar{\mathbf{Q}}(z) = \operatorname{diag} \left\{ \frac{1}{-z - g_i(z)} \right\}_{i=1}^n \tag{2.39}$$

with $(z, g_i(z)) \in \mathcal{Z}(\mathbb{C} \setminus \mathbb{R}^+)$, $i \in \{1, \dots, n\}$, uniquely determined by

$$g_i(z) = \frac{1}{n} \sum_{j=1}^n \frac{\sigma_{ij}^2}{-z - g_j(z)}.$$

Sketch of proof of Theorem 2.9. Basing ourselves on the Gaussian approach, the proof of Theorem 2.9 differs from that of Theorem 2.5 in the application of Lemma 2.13. Taking into consideration the variance $\mathbb{E}[\mathbf{X}_{ik}^2] = \sigma_{ik}^2$, Equation (2.31) gives

$$\begin{aligned} \mathbb{E}[\mathbf{Q}_{ij}] &= \frac{1}{z} \frac{1}{\sqrt{n}} \sum_{k=1}^n \mathbb{E}[\mathbf{X}_{ik}^2] \mathbb{E} \left[\frac{\partial \mathbf{Q}_{kj}}{\partial \mathbf{X}_{ik}} \right] - \frac{1}{z} \delta_{ij} \\ &= -\frac{1}{z} \frac{1}{n} \sum_{k=1}^n \sigma_{ik}^2 \mathbb{E}[\mathbf{Q}_{ki} \mathbf{Q}_{kj} + \mathbf{Q}_{kk} \mathbf{Q}_{ij}] - \frac{1}{z} \delta_{ij} \\ &= -\frac{1}{z} \frac{1}{n} \mathbb{E}[\mathbf{Q} \Sigma_i \mathbf{Q}]_{ij} - \frac{1}{z} \frac{1}{n} \mathbb{E}[\operatorname{tr}(\Sigma_i \mathbf{Q}) \mathbf{Q}_{ij}] - \frac{1}{z} \delta_{ij} \end{aligned}$$

with $\Sigma_i \equiv \operatorname{diag}\{\sigma_{ik}^2\}_{k=1}^n$, so that $\|\Sigma_i\| = O(1)$ uniformly over all i .

Note that the semicircle law in Theorem 2.5 is indeed a special case with $\sigma_{ij}^2 = \delta_{ij}$ and $\Sigma_i = \mathbf{I}_n$. As a consequence, similar to the term $\frac{1}{n}\mathbb{E}[\mathbf{Q}^2]$ in (2.32), the first term on the right-hand side vanishes as $n, p \rightarrow \infty$ (or, again, does not even appear if one considers complex Gaussian entries according to Remark 2.5). Following the same reasoning, the random variable $\frac{1}{n} \text{tr} \Sigma_i \mathbf{Q}(z)$ essentially plays the role of $\frac{1}{n} \text{tr} \mathbf{Q}(z)$ in (2.32) and is expected to converge to some deterministic $g_i(z) \equiv \frac{1}{n} \text{tr} \Sigma_i \bar{\mathbf{Q}}(z)$, which can be taken out of the expectation. This gives, in matrix form

$$\mathbb{E}[\mathbf{Q}(z)] = -\frac{1}{z} \text{diag}\{g_i(z)\}_{i=1}^n \mathbb{E}[\mathbf{Q}(z)] - \frac{1}{z} \mathbf{I}_n + o_{\|\cdot\|}(1).$$

Solving this equation for $\mathbb{E}[\mathbf{Q}(z)] \leftrightarrow \bar{\mathbf{Q}}(z)$ and applying $\frac{1}{n} \text{tr} \mathbf{A}(\cdot)$ on both sides for \mathbf{A} of unit norm, we conclude the proof of Theorem 2.9. □

Theorem 2.9 plays a significant role in the study of random graphs, with applications to community detection in large graphs or networks. We shall come back to this model in more detail later in Section 7.1.

Summarizing, this lengthy first technical section provided the necessary technical ingredients, along with several key results, to study the (large n, p) spectrum of “data sample matrices” from the data population statistics. In Section 2.4, we will seek to go backwards, trying to *infer* the population spectral statistics from the observed *empirical* spectrum of the available samples. To this end though, subtle supplementary results on the limiting spectra must be introduced. This is the objective of the next section.

The subsequent section, possibly the most technical of this part of the book, may be skipped at first read, the main ideas of Section 2.4 being understandable if some results are admitted. Yet, for a clear and rigorous treatment of the limitations of statistical inference in the large n, p regime, the readers will need to grasp the notions of Section 2.3.

2.3 Advanced Spectrum Considerations for Sample Covariances

As opposed to the Marčenko–Pastur law in Theorem 2.4, the generalized sample covariance matrix model of Theorem 2.6 (and beyond) only provides a characterization of the limiting spectral measure μ of $\mu_{\frac{1}{n} \mathbf{X} \mathbf{X}^T}$ (or a deterministic equivalent μ_p for it) through its Stieltjes transform $m(z)$ for $z \in \mathbb{C} \setminus \mathbb{R}^+$ (respectively, through a sequence $m_p(z)$ of Stieltjes transforms), which itself assumes an implicit form. Since the Stieltjes transform inversion formula (Theorem 2.1) involves the limit of $m(z)$ for $z \rightarrow x \in \mathbb{R}$, the sole information about $m(z)$ for all $z \in \mathbb{C} \setminus \mathbb{R}^+$ does not immediately quantify the measure μ .

From a theoretical standpoint, one may wonder whether the limiting μ admits a density as in the Marčenko–Pastur case and, if so, whether one can determine this density and its *exact* support. As recalled in Remarks 2.7 and 2.8, the density of μ (provided it exists) can be “numerically depicted” by solving for $m(z)$ with z close to,

but formally *away* from, the real axis. We aim here at a more theoretical and precise characterization of μ .

From a practical standpoint, a fundamental byproduct of this characterization is the introduction of the function $z \mapsto -\frac{1}{m(z)}$, which plays a key role in statistical inference. Indeed, we shall see in Section 2.4 and the many applications in Chapter 3 that the statistical information related to the population covariance \mathbf{C} (such as functionals of its eigenvalues, projections on its eigenvectors) can be accessed from the data matrix \mathbf{X} by means of a complex integral method involving the change of variable $z \mapsto -\frac{1}{m(z)}$.

2.3.1 Limiting Spectrum

In Silverstein and Choi [1995] (generalized later in Couillet and Hachem [2014] with a more systematic approach), the authors prove that, for any measure ν (the limiting spectral distribution of \mathbf{C}), the limiting measures μ and $\tilde{\mu}$ introduced in Theorem 2.6 indeed have a density with a well-defined support.²⁷

Density and Support of μ (and $\tilde{\mu}$)

Precisely, recall that $\mu = \frac{1}{c}\tilde{\mu} + (1 - \frac{1}{c})\delta_0$ (with δ_0 the Dirac mass at $x = 0$) with $\tilde{\mu}$ defined by its Stieltjes transform $\tilde{m}(z)$ solution to

$$\tilde{m}(z) = \left(-z + c \int \frac{t\nu(dt)}{1 + t\tilde{m}(z)} \right)^{-1}.$$

This functional expression has the interesting key property of being invertible, in the sense that it is formally equivalent to

$$z = -\frac{1}{\tilde{m}(z)} + c \int \frac{t\nu(dt)}{1 + t\tilde{m}(z)}.$$

As a consequence, the function $\tilde{m}(\cdot) : \mathbb{C} \setminus \text{supp}(\tilde{\mu}) \rightarrow \mathbb{C}$, $z \mapsto \tilde{m}(z)$ admits the functional inverse

$$z(\cdot) : \tilde{m}(\mathbb{C} \setminus \text{supp}(\tilde{\mu})) \rightarrow \mathbb{C}$$

$$\tilde{m} \mapsto -\frac{1}{\tilde{m}} + c \int \frac{t\nu(dt)}{1 + t\tilde{m}}.$$

The important point to notice here is that $z(\cdot)$, seen as the functional inverse of $\tilde{m}(\cdot)$, is only defined on the domain $\tilde{m}(\mathbb{C} \setminus \text{supp}(\tilde{\mu}))$. Yet, formally, this function could be

²⁷ It may come as very surprising but very few works in the random matrix literature have actually studied the *exact* behavior of the limiting measure μ of advanced random matrix models. The few exceptions are Silverstein and Choi [1995], Couillet and Hachem [2014], which study the defining equation of the Stieltjes transform m_μ of μ associated with the sample covariance matrix models $\mathbf{C}^{\frac{1}{2}}\mathbf{X}\mathbf{X}^T\mathbf{C}^{\frac{1}{2}}$ and $\mathbf{C}^{\frac{1}{2}}\mathbf{X}\mathbf{C}\mathbf{X}^T\mathbf{C}^{\frac{1}{2}}$, respectively, as well as the very extensive work [Ajanki et al., 2019] on the defining equation of m_μ attached to generalized Wigner models (for instance, the generalized semicircle law for Wigner models with a variance profile, Theorem 2.9). The small number of these studies testifies of the greater importance of the Stieltjes transform relation defining m_μ over the measure μ itself which, both in theory and in practice, is quite often of lesser interest.

extended to all values $\tilde{m} \in \mathbb{C}$ such that $0 \notin 1 + \tilde{m} \cdot \text{supp}(v)$ (i.e., all values that do not cancel the denominator $1 + t\tilde{m}$ for some $t \in \text{supp}(v)$).

The idea of Silverstein and Choi [1995], originally expressed in the seminal work of Marčenko and Pastur [1967], is twofold:

- **Outside the support.** (i) The Stieltjes transform $m_\mu(x) = \int (t-x)^{-1} \mu(dt)$ of a measure μ is well defined and an increasing function on its restriction to $x \in \mathbb{R} \setminus \text{supp}(\mu)$ (it has positive derivative there), hence (ii) so must be its functional inverse $x(\cdot)$ on its restriction to $m_\mu(\mathbb{R} \setminus \text{supp}(\mu))$, (iii) consequently, if $x(\cdot)$ admits an extension to some domain \mathcal{S} with $m_\mu(\mathbb{R} \setminus \text{supp}(\mu)) \subset \mathcal{S} \subset \mathbb{R}$, $x(\cdot)$ should only be increasing on $m_\mu(\mathbb{R} \setminus \text{supp}(\mu))$; ²⁸ (iv) therefore, the complementary $\mathbb{R} \setminus \text{supp}(\mu)$ to the support of μ can be determined as the union of the image of all increasing sections of $x(\cdot)$. See Figure 2.5, commented below, for a simplified visual understanding.

In our setting, this thus formally defines the support of the limiting measure μ of $\mu_{\frac{1}{n} \mathbf{X} \mathbf{X}^T}$.

- **In the support.** Inside this support, one then needs to determine the density of μ . To this end, one may first prove the existence of $\tilde{m}^\circ(x) = \lim_{\epsilon \rightarrow 0} \tilde{m}(x + i\epsilon)$. Upon existence, since $\Im[\tilde{m}^\circ(x)] > 0$ for $x \in \text{supp}(\mu)$, dominated convergence can be applied on the defining equation for $\tilde{m}(z)$ to find that $\tilde{m}^\circ(x)$ is a solution with positive imaginary part of

$$\tilde{m}^\circ(x) = \left(-x + c \int \frac{tv(dt)}{1 + \tilde{m}^\circ(x)t} \right)^{-1},$$

which is then shown to be unique.

These arguments are formally stated in the following theorem.

Theorem 2.10 (Silverstein and Choi [1995]). *Under the setting of Theorem 2.6 with $\mu_{\mathbf{C}} \rightarrow v$ as $p \rightarrow \infty$, define*

$$x(\cdot) : \mathbb{R} \setminus \{ \tilde{m} \mid (-1/\tilde{m}) \in \text{supp}(v) \} \rightarrow \mathbb{R}$$

$$\tilde{m} \mapsto -\frac{1}{\tilde{m}} + c \int \frac{tv(dt)}{1 + \tilde{m}t}.$$

Then, $\tilde{\mu}$ has a density \tilde{f} on $\mathbb{R} \setminus \{0\}$ and

- for $y \in \text{supp}(\tilde{\mu})$, $\tilde{f}(y) = \frac{1}{\pi} \Im[\tilde{m}^\circ(y)]$ with $\tilde{m}^\circ(y)$ the unique solution with positive imaginary part of $x(\tilde{m}^\circ(y)) = y$;
- the support $\text{supp}(\tilde{\mu}) \setminus \{0\}$, which coincides with $\text{supp}(\mu) \setminus \{0\}$, is defined by

$$\text{supp}(\mu) \setminus \{0\}$$

$$= \mathbb{R} \setminus \{ x(\tilde{m}) \mid (-1/\tilde{m}) \in \mathbb{R} \setminus \{ \text{supp}(v) \cup \{0\} \} \text{ and } x'(\tilde{m}) > 0 \}.$$

²⁸ Formally, it is clear that all decreasing sections of (the extended version of) $x(\cdot)$ cannot correspond to the functional inverse of a Stieltjes transform. It is less evident though that all increasing sections do correspond to the inverse of a Stieltjes transform; this was settled in Silverstein and Choi [1995].

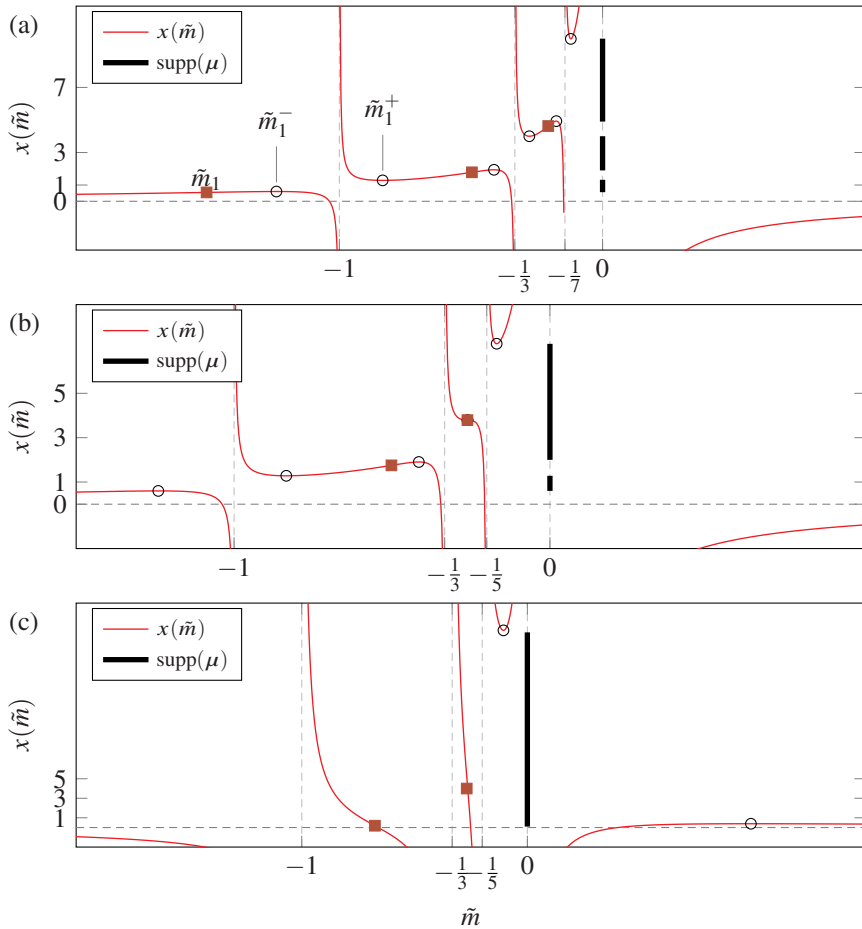


Figure 2.5 The functional inverse $x(\tilde{m})$ for $-1/\tilde{m} \in \mathbb{R} \setminus \text{supp}(\nu)$, with $\nu = \frac{1}{3}(\delta_1 + \delta_3 + \delta_7)$ (a) and $\nu = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$ (b), $c = 1/10$ in both cases, and $\nu = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$ with $c = 2$ (c). Local extrema are marked by circles, inflexion points by squares. The support of μ can be read on the vertical axes. Code on web: MATLAB and Python.

Figure 2.5 depicts the function $x(\tilde{m})$ under a similar setting as Figure 2.4 with ν composed of three Dirac masses. The top display, Figure 2.5(a), shows four increasing regions of $x(\cdot)$, thus corresponding (on the y-axis) to four connected components of $\mathbb{R} \setminus \text{supp}(\mu)$. The complementary, depicted in black on the y-axis, corresponds to the (three) connected components of $\text{supp}(\mu)$. The middle display, Figure 2.5(b), only shows three growing regions for $x(\cdot)$, thus restricting the support of μ to two connected components. Analogously, in the bottom display, Figure 2.5(c), there is only one growing region for $x(\cdot)$ (close to the y-axis from above), which now corresponds to a single connected component for $\text{supp}(\mu) \setminus \{0\}$. This is in accordance with the observations made in Figure 2.4, when altering either ν or c .

A careful analysis of the function $x(\cdot)$ actually reveals additional interesting properties:

- (i) the restriction of $x(\cdot)$ to its growing sections is a growing function. This follows from the fact that, there, $x(\cdot)$ is the functional inverse of $\tilde{m}(\cdot)$ restricted to $\mathbb{R} \setminus \text{supp}(\mu)$, which is a growing function.
- (ii) in the case of Figure 2.5, since ν is discrete, $x(\cdot)$ presents asymptotes at each $-1/t, t \in \text{supp}(\nu)$. Thus, from the previous item, $\text{supp}(\mu)$ is here determined by the union $\cup_k [\tilde{m}_k^-, \tilde{m}_k^+]$ for $\tilde{m}_1^- < \tilde{m}_1^+ < \tilde{m}_2^- < \dots$ the successive values of \tilde{m} such that $x'(\tilde{m}) = 0$. This remark may however *not* hold for ν with continuous support. Detailed conditions for this characterization to hold are discussed in Couillet and Hachem [2014], see also Exercise 8 in Section 2.9 for an example.
- (iii) the derivative of $x(\cdot)$ is given by

$$x'(\tilde{m}) = \frac{1}{\tilde{m}^2} - c \int \frac{t^2 \nu(dt)}{(1+t\tilde{m})^2}$$

and thus $\tilde{m}^2 x'(\tilde{m})$ converges to $1 - c$ as $|\tilde{m}| \rightarrow \infty$, while $x(\tilde{m}) \rightarrow 0$. Thus, $x(\cdot)$ is either decreasing or increasing at $\pm\infty$ depending on whether $c < 1$ or $c > 1$. In particular, the pre-image by $x(\cdot)$ of 0^+ is $-\infty$ if $c < 1$ (Figure 2.5(a) and Figure 2.5(b)) and some positive value if $c > 1$ (Figure 2.5(c)): This remark is fundamental for the next section.

Variable Change: relating $\text{supp}(\nu)$ and $\text{supp}(\mu)$

An important side consequence of the study above of $z(\cdot)$ (and its restriction $x(\cdot)$ to the real axis) is that the function

$$\begin{aligned} \gamma: \mathbb{C} \setminus \{\text{supp}(\mu) \cup \{0\}\} &\rightarrow \mathbb{C} \\ z = z(\tilde{m}) &\mapsto -\frac{1}{\tilde{m}} \end{aligned} \tag{2.40}$$

provides an *injective* mapping between points outside the support of μ and points outside the support of ν with the property that

$$\gamma(\mathbb{C} \setminus \mathbb{R}) \subset \mathbb{C} \setminus \mathbb{R} \quad \text{and} \quad \gamma(\mathbb{R} \setminus \text{supp}(\mu)) \subset \mathbb{R} \setminus \text{supp}(\nu)$$

but where the *inclusion is strict*, in general.

To understand this statement, first consider $z \in \mathbb{C} \setminus \mathbb{R}^+$. Then, by Theorem 2.6, there exists a unique pair $(z, \tilde{m}(z)) \in \mathcal{Z}$ and we may thus write $z = z(\tilde{m})$ for the value $\tilde{m} \in \mathbb{C} \setminus \mathbb{R}^-$ given by $\tilde{m} = \tilde{m}(z)$. For $z = x \in \mathbb{R}^+ \setminus \text{supp}(\mu)$, we have just seen in our discussion of Theorem 2.10 and Figure 2.5 that there also exists $\tilde{m} \in \mathbb{R}^-$ (it must be real because $\Im[\tilde{m}(x)] = 0$ outside the support) such that $x = x(\tilde{m})$. As a consequence, for $z \in \mathbb{C} \setminus \mathbb{R}$, $\tilde{m} = \tilde{m}(z) \in \mathbb{C} \setminus \mathbb{R}$ and thus $-1/\tilde{m} \in \mathbb{C} \setminus \mathbb{R}$. Similarly, for $x \in \mathbb{R} \setminus \text{supp}(\mu)$, from Figure 2.5, $-1/\tilde{m} \in \mathbb{R} \setminus \text{supp}(\nu)$. The map is however only *injective* (in general not *surjective*) as not all values of $\mathbb{C} \setminus \text{supp}(\nu)$ can be reached. For instance, in

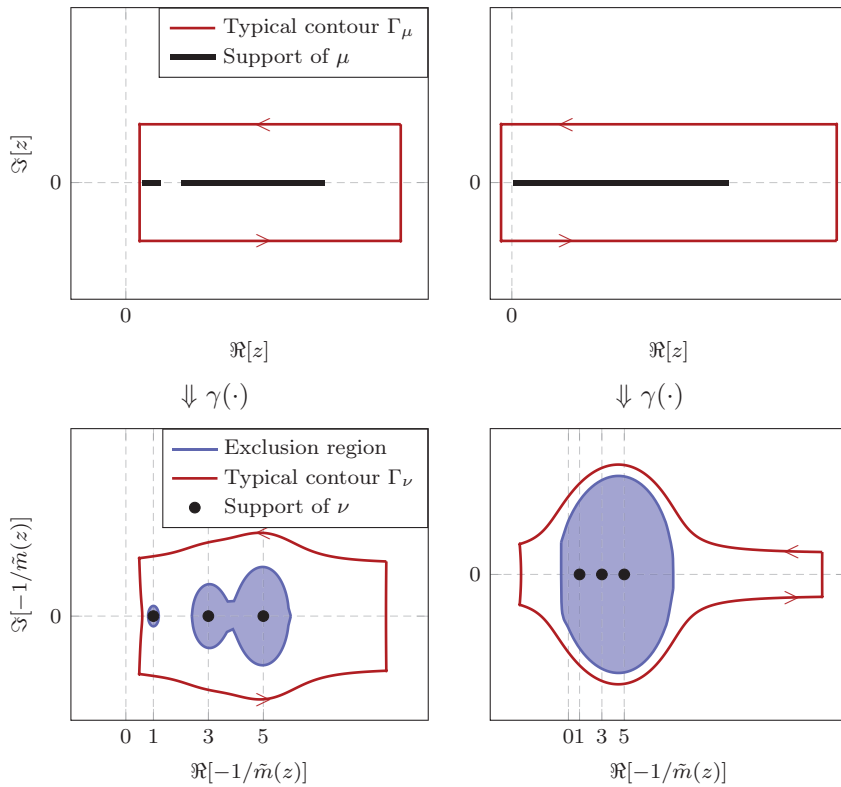


Figure 2.6 Domain of validity of variable changes, for $\nu = \frac{1}{3}(\delta_1 + \delta_3 + \delta_5)$, with $c = 1/10$ (left) and $c = 2$ (right). The filled blue regions in the bottom display are the (inaccessible) complementary to the image of $-1/\tilde{m}(\cdot)$. The red contour Γ_ν is the image by $-1/\tilde{m}(\cdot)$ of a rectangular contour Γ_μ surrounding $\text{supp}(\mu)$. Code on web: MATLAB and Python.

Figure 2.5, the sets $(-1/\tilde{m}_1^-, 1)$ and $(1, -1/\tilde{m}_1^+)$ cannot be reached by γ . This remark will constitute a fundamental limitation to statistical inference methods.

More visually, Figure 2.6 depicts in blue the complementary to the image $\gamma(\mathbb{C} \setminus \text{supp}(\mu))$. This blue region is *inaccessible* in the sense that no point in $\mathbb{C} \setminus \text{supp}(\mu)$ can have an image by $\gamma(\cdot)$ in it. In red are depicted typical images by $\gamma(\cdot)$ of rectangular contours surrounding $\text{supp}(\mu)$. Intuitively, we observe that, as c increases (compare left to right displays), the exclusion region increases in size and one thus cannot get “too close” to the support of ν (which is here the discrete union of three point masses): This “pushes” the image of the red contour further away from the real axis.

In particular, for $c > 1$, the exclusion region includes $\{0\}$. This is a consequence of Item (iii) in the remarks of the previous paragraph: While the right real crossing of a contour $\Gamma_\mu \subset \{z \in \mathbb{C}, \Re[z] > 0\}$ surrounding the support of μ will have an image by $\gamma(\cdot)$ somewhere on the right side of $\text{supp}(\nu)$, (i) for $c < 1$, the left real crossing will have 0^+ for image, and (ii) for $c > 1$, the left real crossing will have a negative value for image.

This, we shall see next in Section 2.4, is an important problem when it comes to estimating certain functionals $\int f d\nu$ of ν based on the sample measure $\mu_{\frac{1}{n}\mathbf{X}\mathbf{X}^T}$.

2.3.2 “No Eigenvalue Outside the Support”

Before exploiting the aforementioned change of variable (the mapping $z \mapsto -1/\tilde{m}(z)$) for statistical inference (in Section 2.4), an important extension of Theorem 2.6 is needed.

It must be stressed that the limiting results of Theorem 2.6 are *weak convergences* for the *normalized* counting measure $\frac{1}{p} \sum_{i=1}^p \delta_{\lambda_i(\frac{1}{n}\mathbf{X}\mathbf{X}^T)}$ (i.e., the spectral measure in Definition 2) of the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$. This, by definition, means that, for every continuous bounded f ,

$$\frac{1}{p} \sum_{i=1}^p f\left(\lambda_i\left(\frac{1}{n}\mathbf{X}\mathbf{X}^T\right)\right) - \int f(t)\mu(dt) \xrightarrow{\text{a.s.}} 0.$$

Letting, for instance, f be a smoothed version of the indicator $1_{[a,b]}$ for $a, b \in \text{supp}(\mu)$, this thus only says that the *averaged* number of eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ within $[a, b]$ converges to $\mu([a, b])$.

In the example of Figure 2.4(a) or Figure 2.4(b) if p_1 is the number of eigenvalues falling in the neighborhood of the leftmost connected component of μ (around one), it is thus *only* possible to know from Theorem 2.6 that $p_1/p = 1/3 + o(1)$ (almost surely), which is equivalent to $p_1 = p/3 + o(p)$. This, in particular, does *not* guarantee that $p_1 - p/3 \xrightarrow{\text{a.s.}} 0$ *exactly* as $n, p \rightarrow \infty$.

Worse, Theorem 2.6 only guarantees that, for $[a, b]$ a connected component of $\mathbb{R} \setminus \text{supp}(\mu)$, the number of eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ inside $[a, b]$ is asymptotically of order $o(p)$. As such, $[a, b]$ may *never* be empty, even for arbitrarily large n, p (it can contain a fixed finite number of eigenvalues or even a growing number of eigenvalues, so long that this number is much less than $O(p)$). In other words, Theorem 2.6 does *not* prevent a few eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ from “leaking” from the limiting support of μ , which, as we shall see in Figure 2.6 and Section 2.4, may cause problems in statistical inference.

The following result, again originally due to Bai and Silverstein, settles this nontrivial issue.

Theorem 2.11 (“No eigenvalue outside the support” and “exact separation”: [Bai and Silverstein, 1998, 1999, Bai et al., 1988]). *Under the setting of Theorem 2.6,²⁹ let $\|\mathbf{C}\|$ be bounded with $\mu_{\mathbf{C}} \rightarrow \nu$ and*

$$\max_{1 \leq i \leq p} \text{dist}(\lambda_i(\mathbf{C}), \text{supp}(\nu)) \rightarrow 0,$$

as $p \rightarrow \infty$. Consider also $-\infty \leq a < b \leq \infty$ such that $a, b \in \mathbb{R}^+ \setminus \text{supp}(\mu)$. Then the following results hold

²⁹ Here formally, the theorem statement must be understood with the “light tail condition” discarded, that is, the only condition on \mathbf{Z} is that it is composed of i.i.d. entries with zero mean and unit variance.

- if $\mathbb{E}[|\mathbf{Z}_{ij}|^4] < \infty$, then, for $|\mathcal{A}|$ the cardinality of set \mathcal{A} ,

$$\left| \left\{ \lambda_i \left(\frac{1}{n} \mathbf{X}\mathbf{X}^\top \right) \in [a, b] \right\} - |\{\lambda_i(\mathbf{C}) \in [\gamma(a), \gamma(b)]\}| \right| \xrightarrow{\text{a.s.}} 0$$

with $\gamma(\cdot)$ defined by (2.40). In particular, if $[a, b]$ is a connected component of $\mathbb{R}^+ \setminus \text{supp}(\mu)$, then

$$\left| \left\{ \lambda_i \left(\frac{1}{n} \mathbf{X}\mathbf{X}^\top \right) \in [a, b] \right\} \right| \xrightarrow{\text{a.s.}} 0.$$

That is, with probability one, no eigenvalues of $\frac{1}{n} \mathbf{X}\mathbf{X}^\top$ appears in $[a, b]$, for all n, p large.

- if $\mathbb{E}[\mathbf{Z}_{ij}^4] = \infty$, then

$$\max_{1 \leq i \leq p} \lambda_i \left(\frac{1}{n} \mathbf{X}\mathbf{X}^\top \right) \xrightarrow{\text{a.s.}} \infty.$$

In plain words, the theorem precisely states that

- under the condition that $\mathbb{E}[\mathbf{Z}_{ij}^4] < \infty$ and that no eigenvalue of \mathbf{C} isolates from its associated limiting spectrum ν , (i) there asymptotically exists no eigenvalue outside the support of μ and (ii) the eigenvalues assembled in asymptotically contiguous “bulks” are found in asymptotically expected numbers. For instance, in the setting of Figure 2.4, it can be verified that *not a single* eigenvalue is found away from the support of μ and, in addition, that the *exact* number of eigenvalues in the neighborhood of each connected component of μ is in *exact* proportion (for Figure 2.4(a), exactly $p/3$ eigenvalues in each component, and for Figure 2.4(b), exactly $2p/3$ eigenvalues in the rightmost and largest component). We emphasize that this is a much finer control (of order $O(1)$) of the eigenvalues than that offered by Theorem 2.6 (which is only of order $o(p)$).
- if $\mathbb{E}[\mathbf{Z}_{ij}^4] = \infty$ (for instance, for a Student t-distribution with low degree of freedom), this “exact separation” collapses: while in correct *asymptotic* proportion guaranteed by Theorem 2.6, up to $o(p)$ eigenvalues may be found away from the support of μ , with, in particular, the largest eigenvalue going to infinity.

For future reference, we insist on the condition

$$\max_{1 \leq i \leq p} \text{dist}(\lambda_i(\mathbf{C}), \text{supp}(\nu)) \rightarrow 0, \tag{2.41}$$

which is also fundamental for the above theorem to hold. Not surprisingly, if a single eigenvalue of \mathbf{C} were to diverge as $p \rightarrow \infty$, it is expected that an eigenvalue of $\frac{1}{n} \mathbf{X}\mathbf{X}^\top$ would also diverge. For instance, say $\lambda_1(\mathbf{C}) = p$ and $\lambda_2(\mathbf{C}) = \dots = \lambda_p(\mathbf{C}) = 1$; then, $\mu_{\mathbf{C}} \rightarrow \delta_1$ so that Theorems 2.4 and 2.6 ensure that $\mu_{\frac{1}{n} \mathbf{X}\mathbf{X}^\top}$ converges weakly to the Marčenko–Pastur law, while the largest eigenvalue of $\frac{1}{n} \mathbf{X}\mathbf{X}^\top$ is strongly expected to diverge to infinity (which it, indeed, does in this case). Section 2.5 on *spiked models* is strongly inspired by this remark.

2.4 Preliminaries on Statistical Inference

Section 2.3 provides the necessary (technical) ingredients for basic statistical inference considerations of large-dimensional sample covariance matrix models.

In this section, we will successively consider the estimation (i) of linear eigenvalue statistics³⁰ of the type $\frac{1}{p} \sum_{i=1}^p f(\lambda_i(\mathbf{C}))$ and (ii) of eigenvector projections $\mathbf{a}^\top \mathbf{u}_i$ (\mathbf{u}_i an eigenvector of \mathbf{C}) for deterministic vectors \mathbf{a} ; from the sample observation $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$, $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$ and \mathbf{z}_i with standard i.i.d. entries, as defined in Theorem 2.6.

Before entering the topic, it must be mentioned that *large-dimensional statistical inference*, from a random matrix approach, has stood for long as a complex problem. In particular, retrieving information about a population covariance \mathbf{C} from the samples $\mathbf{C}^{\frac{1}{2}} \mathbf{Z}$ may be seen as inverting Theorem 2.6, a problem tentatively tackled in El Karoui [2008] and later in Bun et al. [2017] using convex optimization (thus nonexact) schemes, but with limited success. Some specific objects, such as traces of powers of \mathbf{C} , traces of its resolvent, quadratic forms, etc., may be estimated by detoured means and formed the extensive database of the more-than-fifty *G-estimators* due to Girko [2001] (the phrase “G-estimator” should be understood as “generalized estimators,” according to Girko). In this section, we instead concentrate on a contour integral approach to *systematically* estimate a broad class of functionals of \mathbf{C} : The idea, found scattered in the literature, was revived by Mestre [2008]. The content of this section is not easily found in the existing literature but is strongly inspired by (a simplified treatment of) Mestre [2008].

2.4.1 Linear Eigenvalue Statistics

Relating Population and Sample Stieltjes Transforms

A first observation is that the defining equation for $\tilde{m}(z)$ in Theorem 2.6, that is,

$$\tilde{m}(z) = \left(-z + c \int \frac{tv(dt)}{1 + t\tilde{m}(z)} \right)^{-1}$$

can be equivalently rewritten under the form

$$m_v \left(-\frac{1}{\tilde{m}(z)} \right) = -zm(z)\tilde{m}(z) \tag{2.42}$$

where we recall that $m(z) = \frac{1}{c}\tilde{m}(z) + \frac{1-c}{c} \frac{1}{z}$. This simply follows from noticing that

³⁰ These are called *linear statistics* although f will in general not be linear. What is linear here is in fact the mapping $(f(\lambda_1), \dots, f(\lambda_p)) \mapsto \frac{1}{n} \sum_{i=1}^p f(\lambda_i)$.

$$\begin{aligned} \int \frac{t\nu(dt)}{1+t\tilde{m}(z)} &= \frac{1}{\tilde{m}(z)} \int \frac{t\tilde{m}(z)\nu(dt)}{1+t\tilde{m}(z)} \\ &= \frac{1}{\tilde{m}(z)} \left(1 - \int \frac{\nu(dt)}{1+t\tilde{m}(z)} \right) \\ &= \frac{1}{\tilde{m}(z)} \left(1 - \frac{1}{\tilde{m}(z)} \int \frac{\nu(dt)}{t - (-1/\tilde{m}(z))} \right) \end{aligned}$$

where, from Definition 3, we recognize $\int \frac{\nu(dt)}{t - (-1/\tilde{m}(z))}$ to be the Stieltjes transform m_ν of the measure ν evaluated at $-1/\tilde{m}(z)$.

Theorem 2.6 thus (indirectly) establishes a relation between the population statistics of \mathbf{C} and that of the (observed) sample covariance matrix $\frac{1}{n}\mathbf{X}\mathbf{X}^T$, through the Stieltjes transforms of their *limiting* measures, and we can already anticipate that $-1/\tilde{m}(z)$ will indeed play the role of a variable change to move from z in $m(z), \tilde{m}(z)$ to z' in $m_\nu(z')$ if $z' = -1/\tilde{m}(z)$.

Eigen-Inference

Now, observe that, for $f: \mathbb{C} \rightarrow \mathbb{C}$ a function analytic in a neighborhood of the eigenvalues of \mathbf{C} , by Cauchy’s integral theorem, the linear statistics $\frac{1}{p} \sum_{i=1}^p f(\lambda_i(\mathbf{C}))$ of the eigenvalues of \mathbf{C} can be expressed as³¹

$$\begin{aligned} \frac{1}{p} \sum_{i=1}^p f(\lambda_i(\mathbf{C})) &\simeq \int f(t)\nu(dt) \\ &= \int \left[\frac{1}{2\pi i} \oint_{\Gamma_\nu} \frac{f(z) dz}{z-t} \right] \nu(dt) \\ &= -\frac{1}{2\pi i} \oint_{\Gamma_\nu} f(z) \left[\int \frac{\nu(dt)}{t-z} \right] dz \\ &= -\frac{1}{2\pi i} \oint_{\Gamma_\nu} f(z)m_\nu(z) dz \end{aligned} \tag{2.43}$$

where $\Gamma_\nu \subset \mathbb{C}$ is a (positively oriented) contour encircling the support of ν but *no singularity of f* . Here, the integral exchange comes at no difficulty because Γ_ν is a closed compact contour carefully avoiding the support of ν (so that $t - z$ in the denominator is uniformly away from zero) and $\text{supp}(\nu)$ is bounded. Thus, one can express (smooth) linear statistics of the eigenvalues of \mathbf{C} by means of a complex integral involving the Stieltjes transform $m_\nu(z)$.

As a consequence of (2.42), it is now possible to relate the *nonobservable* $m_\nu(z)$ to $\tilde{m}(z)$, which is the large n, p limit of the *observable* Stieltjes transform $m_{\frac{1}{n}\mathbf{X}\mathbf{X}^T}(z)$. To be able to plug (2.42) into (2.43), one needs to perform the change of variable $z \mapsto -1/\tilde{m}(z)$. This is however *only possible* if there indeed exists a $\Gamma_\nu \subset \mathbb{C}$ (the contour in (2.43)) such that $\Gamma_\nu = -1/\tilde{m}(\Gamma_\mu)$ for some well-defined complex path Γ_μ . The discussions in Section 2.3.1 and in particular, around Figure 2.6, have clarified the conditions under which such a Γ_ν exists.

³¹ Here again the “ \simeq ” sign can be turned into an equality if one assumes $\nu = \frac{1}{p} \sum_{i=1}^p \delta_{\lambda_i(\mathbf{C})}$.

But let us assume that Γ_ν is indeed well defined as $\Gamma_\nu = -1/\tilde{m}(\Gamma_\mu)$ for some valid Γ_μ . Then, Equation (2.43) along with (2.42) imply

$$\begin{aligned} \int f(t)\nu(dt) &= -\frac{1}{2\pi i} \oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right) m_\nu\left(-\frac{1}{\tilde{m}(\omega)}\right) \frac{\tilde{m}'(\omega)}{\tilde{m}^2(\omega)} d\omega \\ &= \frac{1}{2\pi i} \oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right) \omega \frac{m(\omega)\tilde{m}'(\omega)}{\tilde{m}(\omega)} d\omega \end{aligned}$$

where we wrote $z = -1/\tilde{m}(\omega)$. Using that $m(\omega) = \frac{1}{c}\tilde{m}(\omega) + (1-c)/(c\omega)$, this further reads

$$\begin{aligned} \int f(t)\nu(dt) &= \frac{1}{2c\pi i} \oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right) \frac{(\omega\tilde{m}(\omega) + (1-c))\tilde{m}'(\omega)}{\tilde{m}(\omega)} d\omega \\ &= \frac{1}{2c\pi i} \oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right) \omega\tilde{m}'(\omega) d\omega - \frac{1-c}{c} f(0) \cdot 1_{\{0 \in \Gamma_\nu^\circ\}} \end{aligned}$$

where Γ_ν° is the inside of Γ_ν , and where for the last equality we used

$$\frac{1}{2\pi i} \oint_{\Gamma_\mu} f\left(-\frac{1}{\tilde{m}(\omega)}\right) \frac{\tilde{m}'(\omega)}{\tilde{m}(\omega)} d\omega = -\frac{1}{2\pi i} \oint_{\Gamma_\nu} z^{-1} f(z) dz = -f(0) \cdot 1_{\{0 \in \Gamma_\nu^\circ\}}$$

by residue calculus, assuming again that f is analytic on a sufficiently large region (in particular here around zero).

To complete the statistical inference framework, one finally needs to relate the above expression to the observation \mathbf{X} . The idea is to use the fact that $m_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}}(\omega) \xrightarrow{\text{a.s.}} \tilde{m}(\omega)$ from Theorem 2.6. To ensure that $\tilde{m}(\omega)$ can be replaced by $m_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}}(\omega)$ in the above expression, one however needs to ensure that dominated convergence on the compact set Γ_μ holds. For this, two ingredients are needed: (i) first guarantee that the convergence $m_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}}(\omega) \xrightarrow{\text{a.s.}} \tilde{m}(\omega)$ is uniform on Γ_μ , which easily follows from the analytic nature of Stieltjes transforms, and more importantly (ii) prove that the integrand $f(-1/m_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}}(\omega))\omega m'_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}}(\omega)$ is uniformly bounded on Γ_μ . This second item follows from Theorem 2.11 which guarantees that, for all n, p large, with probability one, all eigenvalues remain in the vicinity of $\text{supp}(\mu)$ under the additional conditions (i) $\mathbb{E}[|\mathbf{X}_{ij}|^4] < \infty$ and (ii) $\max_i \text{dist}(\lambda_i(\mathbf{C}), \text{supp}(\nu)) \rightarrow 0$.

As a consequence, accounting now for the conditions of validity of the variable change discussed in the previous section, we have the following statistical inference result, the original ideas of which are due to Mestre.

Theorem 2.12 (Inspired by Mestre [2008]). *Under the setting of Theorem 2.6 with $\mathbb{E}[|\mathbf{X}_{ij}|^4] < \infty$ and $\max_{1 \leq i \leq p} \text{dist}(\lambda_i(\mathbf{C}), \text{supp}(\nu)) \rightarrow 0$, let $f: \mathbb{C} \rightarrow \mathbb{C}$ be a complex function analytic on the complement of $\gamma(\mathbb{C} \setminus \text{supp}(\mu))$ in \mathbb{C} with γ defined in (2.40). Then,*

$$\frac{1}{p} \sum_{i=1}^p f(\lambda_i(\mathbf{C})) - \frac{1}{2c\pi i} \oint_{\Gamma_\mu} f\left(\frac{-1}{m_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}}(\omega)}\right) \omega m'_{\frac{1}{n}\mathbf{X}^\top\mathbf{X}}(\omega) d\omega \xrightarrow{\text{a.s.}} 0,$$

for some complex positively oriented contour $\Gamma_\mu \subset \mathbb{C}$ surrounding $\text{supp}(\mu) \setminus \{0\}$. In particular, if $c < 1$, the result holds for any f analytic on $\{z \in \mathbb{C}, \Re[z] > 0\}$ with Γ_μ chosen as any such contour within $\{z \in \mathbb{C}, \Re[z] > 0\}$.

From a numerical standpoint, for $c < 1$, Theorem 2.12 is rather straightforward: It indicates that any complex contour Γ_μ in $\{z \in \mathbb{C}, \Re[z] > 0\}$ guarantees the result. For $c > 1$, the choice of Γ_μ is less trivial. For safety, it is advised to take Γ_μ a contour closely fitting the support of $\mu \frac{1}{n} \mathbf{X}^T \mathbf{X}$, excluding zero (such as a small rectangle). Figures 2.5 and 2.6 visually explain the issue surrounding the case $c > 1$ and the technical request regarding the analytic nature of f : From Figure 2.6, since the “tightest-to-the-real-line” (red) contours Γ_ν in the bottom displays must avoid the blue areas (to be well-defined images of valid contours Γ_μ from the top displays), the minimal request is for f to be analytic on those blue areas *enclosed* in the red contour; if not analytic there, the complex integral would have additional residues, thereby altering the result of the theorem.

In practice, the most problematic case occurs when 0 falls within a blue area and one has to deal with functions $f(z)$ involving $\log(z)$, \sqrt{z} , $1/z$, all of which are singular at $z = 0$. A typical way out of this situation would be to add an extra term in the result to compensate for the extra residue; this compensating term would then have to be estimated. This however appears not always to be possible as discussed in the following remark.

Remark 2.11 (On the $c > 1$ case). *For $c > 1$ and for f not analytic at zero (for instance, $f(z) = \log(z)$, $f(z) = z^{-1}$, or $f(z) = \sqrt{z}$), Theorem 2.12 cannot be applied. That is, for these functions,*

$$\frac{1}{p} \sum_{i=1}^p f(\lambda_i(\mathbf{C}))$$

cannot be consistently estimated directly from the theorem statement. Using the above compensation by the residue at zero workaround, however, it appears that the compensating term is at least as hard to estimate as $\frac{1}{p} \sum_{i=1}^p f(\lambda_i(\mathbf{C}))$ itself. This somehow suggests that, when $p > n$ and thus the sample covariance matrix $\frac{1}{n} \mathbf{X} \mathbf{X}^T$ is of rank $n < p$, one lacks information to estimate some functionals of the p eigenvalues of \mathbf{C} . A similar problem will be discussed in Remark 3.3 on the application to between-covariance matrix distance estimation.

Application Example: Estimating Population Eigenvalues of Large Multiplicity

Figure 2.4 presents three scenarios where the population spectral measure $\mu_{\mathbf{C}}$ (or equivalently its limit ν) is a discrete sum of three distinct eigenvalues. A natural concern in the large n, p setting is whether it is possible to estimate these eigenvalues consistently from the sample data \mathbf{X} of size n .

In Figure 2.4(a), it a priori appears that averaging the sample eigenvalues of each component of μ_p may provide such a consistent estimator. This is however not the case: As can be checked below, this estimator is indeed biased. The framework devised in the previous section, on the contrary, will provide a consistent estimator: The idea is now to design a contour Γ_ν , which would encircle *only one* of the three masses

in the spectrum (rather than encircling the whole support of ν); one must then find a corresponding valid contour Γ_μ such that $\Gamma_\nu = -1/\tilde{m}(\Gamma_\mu)$; not surprisingly, this contour Γ_μ will encircle the “hump” in the empirical spectrum μ associated with the corresponding sought-for eigenvalue of \mathbf{C} (this being a consequence of the discussions in Section 2.3 and particularly of Figure 2.5). In Figure 2.4(b), a problem arises for the two population eigenvalues (3 and 5) of \mathbf{C} associated with the same connected component of μ : For these, no complex contour Γ_ν exists that would be a proper image $\Gamma_\nu = -1/\tilde{m}(\Gamma_\mu)$ and that would circle around either 3 or 5 alone, see also the left plots of Figure 2.6. We will see that a more involved procedure can nonetheless consistently estimate them both. In Figure 2.4(c), the difficulty further increases: Here again, it remains possible to estimate 1, 3, and 5 but at the cost of a more involved method.

Consider then the following generalized setting of Figure 2.4, where

$$\nu_{\mathbf{C}} = \frac{1}{p} \sum_{i=1}^k p_i \delta_{\ell_i} \rightarrow \sum_{i=1}^k c_i \delta_{\ell_i}$$

for $\ell_1 > \dots > \ell_k > 0$, k fixed with respect to n, p , and $p_i/p \rightarrow c_i > 0$ as $p \rightarrow \infty$ (i.e., each eigenvalue has a large multiplicity of order $O(p)$).

Fully Separable Case

We additionally assume for the moment that the sample size $n > p$ of $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ (where $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$, \mathbf{z}_i having standard i.i.d. entries and bounded fourth-order moment as demanded in Theorem 2.11) is sufficiently large for the number of connected components in μ to be exactly k , that is, each eigenvalue of \mathbf{C} is “mapped” to a single connected component of $\text{supp}(\mu)$ as in Figure 2.4(a) and Figure 2.5(a).

Then, to estimate the population ℓ_a , $a \in \{1, \dots, k\}$, Theorem 2.12 may be applied to the mere function $f(z) = z$, however for Γ_μ now changed into $\Gamma_\mu^{(a)}$, a contour containing only the a th connected component of $\text{supp}(\mu)$ (these connected components are sorted descendingly according to their values from ∞ to 0, for example, there are three connected components in Figure 2.4(a): the first component around 7, the second around 3, and the third around 1, respectively). Adapting Theorem 2.12 according to Theorem 2.11 and our previous line of reasoning, we then have

$$\ell_a - \hat{\ell}_a \xrightarrow{\text{a.s.}} 0, \quad \hat{\ell}_a = -\frac{n}{p_a} \frac{1}{2\pi i} \oint_{\Gamma_\mu^{(a)}} \omega \frac{m'_{\frac{1}{n} \mathbf{X}^T \mathbf{X}}(\omega)}{m_{\frac{1}{n} \mathbf{X}^T \mathbf{X}}(\omega)} d\omega \xrightarrow{\text{a.s.}} 0. \tag{2.44}$$

The estimator $\hat{\ell}_a$ can be numerically evaluated. However, recalling that $m_{\frac{1}{n} \mathbf{X}^T \mathbf{X}}(\omega)$ (and its derivative) are rational functions, this integral is prone to estimation by a simple residue calculus. Indeed, first observe that the integrand in the expression of $\hat{\ell}_a$ has two types of poles: (i) the $\lambda_i = \lambda_i(\frac{1}{n} \mathbf{X}^T \mathbf{X})$ falling inside the surface described by $\Gamma_\mu^{(a)}$, since in the neighborhood of λ_i ,

$$\begin{aligned}
 -\frac{n}{p_a} \omega \frac{m'_{\frac{1}{n}\mathbf{X}^T\mathbf{X}}(\omega)}{m_{\frac{1}{n}\mathbf{X}^T\mathbf{X}}(\omega)} &= -\frac{n}{p_a} \omega \frac{\frac{1}{n} \sum_{i=1}^n \frac{1}{(\lambda_i - \omega)^2}}{\frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i - \omega}} \\
 &\sim_{\omega \sim \lambda_i} -\frac{n}{p_a} \frac{\omega}{\lambda_i - \omega}
 \end{aligned}$$

and (ii) the zeros of $m_{\frac{1}{n}\mathbf{X}^T\mathbf{X}}$ falling within $\Gamma_\mu^{(a)}$.

For readability in what follows, we sort the eigenvalues of $\frac{1}{n}\mathbf{X}^T\mathbf{X}$ as $\lambda_1 \geq \dots \geq \lambda_n$ (these are almost surely distinct but for the possible zero eigenvalues). Dealing with the first type of poles is easy: The λ_i falling within $\Gamma_\mu^{(1)}$ are precisely the p_1 largest, within $\Gamma_\mu^{(2)}$ the next p_2 largest, etc., as per Theorem 2.11. The residue associated with λ_i is then

$$\lim_{\omega \rightarrow \lambda_i} (\omega - \lambda_i) \frac{n}{p_a} \frac{-\omega}{\lambda_i - \omega} = \frac{n}{p_a} \lambda_i.$$

The second set of poles is less immediate to retrieve. An important remark is that the zeros, call them η_j (sorted also as $\eta_1 \geq \eta_2 \geq \dots$), of $m_{\frac{1}{n}\mathbf{X}^T\mathbf{X}}(\omega)$ are necessarily real (since the Stieltjes transform has nonzero imaginary part for $\Im[\omega] \neq 0$, see Definition 3) and satisfy

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i - \eta_j} = 0.$$

Since the function

$$x \mapsto \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i - x}$$

is increasing and has ∞ and $-\infty$ asymptotes at $x = \lambda_i - 0$ and $x = \lambda_i + 0$, respectively, each η_j falls exactly in one of the intervals $[\lambda_i, \lambda_{i+1})$ and thus each λ_i pole is accompanied by its η_i pole (if sorted similarly; see Figure 2.7 for an illustration). The residue calculus then gives, by Taylor expanding the denominator,

$$\lim_{\omega \rightarrow \eta_j} (\omega - \eta_j) \frac{n}{p_a} \frac{-\omega m'_{\frac{1}{n}\mathbf{X}^T\mathbf{X}}(\omega)}{0 + m'_{\frac{1}{n}\mathbf{X}^T\mathbf{X}}(\eta_j)(\omega - \eta_j)} = -\frac{n}{p_a} \eta_j.$$

As a result, we finally have the estimator

$$\hat{\ell}_a = \frac{n}{p_a} \sum_{i=p_1+\dots+p_{a-1}+1}^{p_1+\dots+p_a} \lambda_i - \eta_i. \tag{2.45}$$

Surprisingly at first, it appears that the estimator is the sum of $p_a = O(p)$ terms, which may seem to conduct to an estimate of order $O(p)$. However, recall that $\lambda_1, \dots, \lambda_p$ are “compactified” in a support of size $O(1)$ and that $\lambda_{i-1} < \eta_i < \lambda_i$ so that $\lambda_i - \eta_i = O(p^{-1})$, which resolves the problem.

This formulation is nonetheless still not fully closed, in the sense that the η_i s are so far only provided in terms of the zeros of $m_{\frac{1}{n}\mathbf{X}^T\mathbf{X}}$. The following remark provides an explicit form.

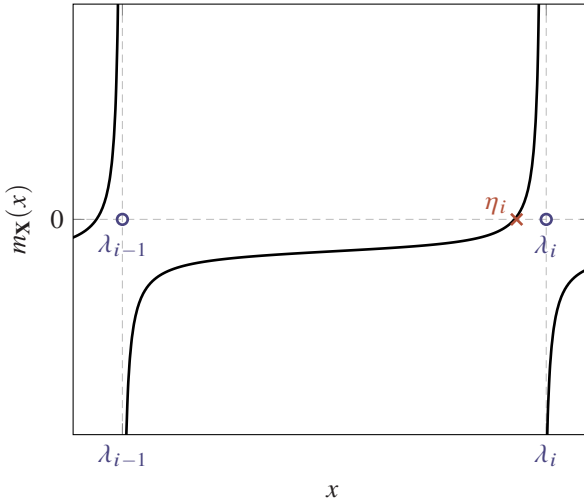


Figure 2.7 Illustration of the zeros (η_i) and poles (λ_i) of the (restriction to the real axis of the) Stieltjes transform $m_{\mathbf{X}}(x)$. Code on web: MATLAB and Python.

Remark 2.12 (Explicit expression for the zeros of $m_{\mathbf{X}}(z)$). For $\mathbf{X} \in \mathbb{R}^{n \times n}$ symmetric with eigenvalues $\lambda_1 > \dots > \lambda_n$, the zeros $\eta_1 > \eta_2 > \dots$ of $m_{\mathbf{X}}(z)$ satisfy the following equivalence relations

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i - \eta_j} = 0 &\Leftrightarrow \frac{1}{n} \sum_{i=1}^n \frac{-\eta_j}{\lambda_i - \eta_j} = 0 \\ &\Leftrightarrow \frac{1}{n} \sum_{i=1}^n \frac{\lambda_i}{\lambda_i - \eta_j} - 1 = 0 \\ &\Leftrightarrow \frac{1}{n} \sqrt{\lambda}^\top (\Lambda - \eta_j \mathbf{I}_n)^{-1} \sqrt{\lambda} - 1 = 0 \\ &\Leftrightarrow \det \left(\frac{1}{n} \sqrt{\lambda} \sqrt{\lambda}^\top (\Lambda - \eta_j \mathbf{I}_n)^{-1} - \mathbf{I}_n \right) = 0 \\ &\Leftrightarrow \det \left(\frac{1}{n} \sqrt{\lambda} \sqrt{\lambda}^\top - \Lambda + \eta_j \mathbf{I}_n \right) = 0 \end{aligned}$$

where we denoted $\sqrt{\lambda} \in \mathbb{R}^P$ the (column) vector of the $\sqrt{\lambda_i}$ s and $\Lambda \in \mathbb{R}^{P \times P}$ the diagonal matrix $\text{diag}\{\lambda_i\}_{i=1}^P$, sorted in the same way, and used Lemma 2.3 as well as the fact that $\det(\Lambda - \eta_j \mathbf{I}_n) \neq 0$ according to our discussion above.

Consequently, the zeros of $m_{\mathbf{X}}$ are exactly the eigenvalues of

$$\Lambda - \frac{1}{n} \sqrt{\lambda} \sqrt{\lambda}^\top.$$

Figure 2.8 depicts the estimation errors in the setting of two population eigenvalues ℓ_1 and ℓ_2 (with $\ell_1 = 1$ and $p/n = 1/4$), as a function of the difference $\Delta\lambda = \ell_2 - \ell_1$. Note first that the derived random matrix-based estimator significantly outperforms the naive approach of averaging the eigenvalues of each component of the sample covariance. Also, we observe that the estimator error of the proposed approach grows

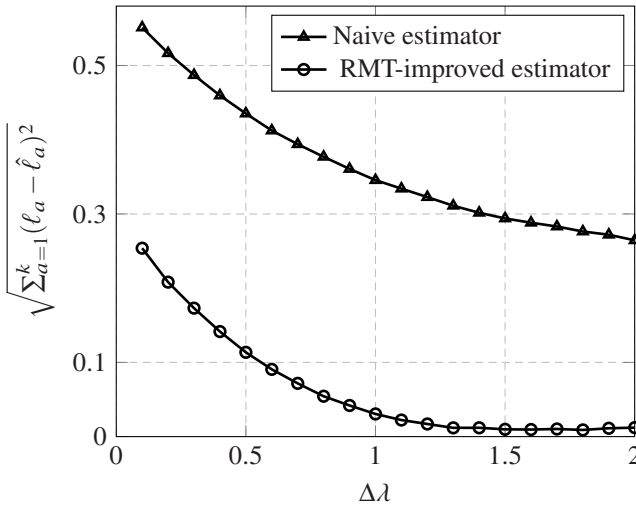


Figure 2.8 Eigenvalue estimation errors with naive and RMT-improved approach, as a function of $\Delta\lambda$, for $\ell_1 = 1$, $\ell_2 = 1 + \Delta\lambda$, $p = 256$ and $n = 1024$. Results averaged over 30 runs. Code on web: MATLAB and Python.

rapidly once $\Delta\lambda < 1$: This is a typical “avalanche effect,” which appears below the phase transition threshold when the two connected components of the support of the empirical measure μ are no longer separable and the estimator is thus, in theory, no longer consistent.

Nonseparable Case

The estimator introduced above is only valid if the contour $\Gamma_\mu^{(a)}$ is licit, in the sense that its image by the variable change $z \mapsto -1/\bar{m}(z)$ leads to a valid contour $\Gamma_\nu^{(a)}$ surrounding ℓ_a only. However, we have seen (in Figure 2.6 notably) that there may not exist any such licit $\Gamma_\mu^{(a)}$. In our present setting, Figure 2.5(b) and Figure 2.5(c), as well as Figure 2.6, reveal that, if say ℓ_1 and ℓ_2 are associated with a *single* connected component of $\text{supp}(\mu)$, then all contours $\Gamma_\mu^{(1)}$ surrounding *only* the p_1 largest empirical eigenvalues λ_i s are illicit.

In order to estimate both ℓ_1 and ℓ_2 individually, one must then resort to using at least *two* estimates of linear functionals of the couple (ℓ_1, ℓ_2) . One approach is to estimate simultaneously both $\frac{p_1}{p}\ell_1 + \frac{p_2}{p}\ell_2$ and $\frac{p_1}{p}\ell_1^2 + \frac{p_2}{p}\ell_2^2$, which are accessible from our present adaptation of Theorem 2.12 for $f(z) = z$ and $f(z) = z^2$, with a contour $\Gamma_\mu^{(1,2)}$ surrounding the connected component of μ encompassing the $p_1 + p_2$ largest λ_i s.

Assuming that p_1 and p_2 are known, this thus boils down to solving a second-order polynomial in $\hat{\ell}_1$ and $\hat{\ell}_2$. This procedure however has several limitations: (i) the polynomial equations may lead to nonreal solutions (recall that, while asymptotically this will not occur, the procedure is based on the *finite-dimensional* random realization \mathbf{X} , so that nonreal solutions may arise with nonzero probability), and (ii) assuming that p_1 and p_2 are known is, unlike the fully separable

case guaranteed by Theorem 2.11, in fact quite demanding as they cannot be easily estimated from the empirical eigenvalues λ_i themselves (an additional third equation is then needed), (iii) a further classical issue in statistics is that estimates of higher-order (second-order here) moments are increasingly prone to large variances as the moment order increases: As such, the need for additional equations to estimate the individual ℓ_a and their multiplicity must pass through generalized (nonpolynomial) moments, which are possibly cumbersome to estimate.

2.4.2 Eigenvector Projections and Subspace Methods

In the previous section on the inference methods for the linear statistics (of eigenvalues) of the population covariance \mathbf{C} , we exploited, as a immediate consequence of Theorem 2.6, the relation

$$m_\nu(-1/\tilde{m}(z)) = -zm(z)\tilde{m}(z)$$

between the Stieltjes transform m_ν of the population covariance measure ν and the Stieltjes transform m (and \tilde{m}) of the sample covariance measure μ (and $\tilde{\mu} = c\mu + (1-c)\delta_0$).

The deterministic equivalent statements $\mathbf{Q}(z) \leftrightarrow \tilde{\mathbf{Q}}(z)$ (as well as $\tilde{\mathbf{Q}}(z) \leftrightarrow \bar{\tilde{\mathbf{Q}}}(z)$) in Theorem 2.6 go beyond Stieltjes transform relations as they connect the whole resolvent matrix $\mathbf{Q}(z) = (\frac{1}{n}\mathbf{X}\mathbf{X}^\top - z\mathbf{I}_p)^{-1}$ of the sample covariance (almost directly) to the resolvent $(\mathbf{C} - z\mathbf{I}_p)^{-1}$ of the population covariance.

These relations can be used in the following ways: (i) when \mathbf{C} is known, they provide asymptotic characterizations of some functionals of \mathbf{X} involving its singular vectors (i.e., the eigenvectors $\hat{\mathbf{u}}_i(\mathbf{X}^\top\mathbf{X})$ of $\mathbf{X}^\top\mathbf{X}$ or $\hat{\mathbf{u}}_i(\mathbf{X}\mathbf{X}^\top)$ of $\mathbf{X}\mathbf{X}^\top$), in particular projections $\hat{\mathbf{u}}_i(\mathbf{X}\mathbf{X}^\top)^\top\mathbf{u}(\mathbf{C})$ onto the eigenvectors $\mathbf{u}(\mathbf{C})$ of \mathbf{C} ; (ii) when \mathbf{C} is unknown, they provide estimates for some functionals of the eigenvectors of \mathbf{C} , notably projections $\mathbf{a}^\top\mathbf{u}(\mathbf{C})$ onto deterministic vectors \mathbf{a} , using those of the empirical eigenvectors $\hat{\mathbf{u}}_i(\mathbf{X}\mathbf{X}^\top)$. The latter case is particularly suited to the so-called subspace methods, for instance, based on the fact that $\mathbf{u}(\mathbf{C})$ is known to be aligned (or be equal) to some vector \mathbf{a}_θ parametrized by θ and one aims to solve for θ maximizing this alignment. See Section 3.1.3 for an example of such methods in signal processing application. Another scenario of significance is spectral clustering, where the dominant eigenvectors of the kernel matrix $\mathbf{X}^\top\mathbf{X}$ are used to estimate the dominant population eigenvectors, themselves precisely providing the data classes: knowing their asymptotic alignment thus provides precise characterizations of the performance of spectral clustering.

Estimates of Functionals of \mathbf{X}

In some machine learning applications, the observed data \mathbf{X} will be processed in a nonlinear fashion that may nonetheless preserve its eigenvector structure. The spectral behavior of the resulting matrix may here be typically evaluated by means of its projection onto specific vector structures. This is, for instance, the case of some simple gradient descent mechanisms for supervised learning to be discussed in Section 5.2,

where the learning performance can be measured from the alignment between the gradient descent iterates and the classification vectors (such as the vector $[-\mathbf{1}_{n_1}, \mathbf{1}_{n_2}]$ in a binary classification setting).

For $\mathbf{M} \in \mathbb{R}^{p \times p}$, a symmetric matrix with spectral decomposition $\mathbf{M} = \mathbf{U}\Lambda\mathbf{U}^\top$ and $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_p\}$, and $f: \mathbb{R} \rightarrow \mathbb{R}$, we shall here denote

$$f(\mathbf{M}) = \mathbf{U} \text{diag}\{f(\lambda_i)\}_{i=1}^p \mathbf{U}^\top.$$

Assume f is extensible to a complex function $f: \mathbb{C} \rightarrow \mathbb{C}$, analytic on a neighborhood of $\lambda_1, \dots, \lambda_p$. Then, we have that

$$f(\mathbf{M}) = -\frac{1}{2\pi i} \oint_{\Gamma} f(z) \mathbf{Q}_{\mathbf{M}}(z) dz$$

for $\Gamma \subset \mathbb{C}$ a contour closely encompassing $\lambda_1, \dots, \lambda_p$ but *no singularity* of f . This result arises from a simple residue calculus. Indeed, writing

$$\mathbf{Q}_{\mathbf{M}} = \mathbf{U}(\Lambda - z\mathbf{I}_p)^{-1}\mathbf{U}^\top = \sum_{i=1}^p \frac{\mathbf{u}_i \mathbf{u}_i^\top}{\lambda_i - z}$$

with $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$, each eigenvalue λ_j is a pole of the integrand and the associated residue is

$$\lim_{z \rightarrow \lambda_j} (z - \lambda_j) f(z) \sum_{i=1}^p \frac{\mathbf{u}_i \mathbf{u}_i^\top}{\lambda_i - z} = -f(\lambda_j) \mathbf{u}_j \mathbf{u}_j^\top.$$

Summing the expression above over j gives the result.

Now, assuming that $\mathbf{Q}_{\mathbf{M}}(z)$ admits a deterministic equivalent $\bar{\mathbf{Q}}(z)$, we have, in particular, for $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$, deterministic and of bounded norms,

$$\begin{aligned} \frac{1}{p} \text{tr}(\mathbf{A}f(\mathbf{M})) &= -\frac{1}{2\pi i} \oint_{\Gamma} f(z) \frac{1}{p} \text{tr} \mathbf{A} \mathbf{Q}_{\mathbf{M}}(z) dz \\ &= -\frac{1}{2\pi i} \oint_{\Gamma} f(z) \frac{1}{p} \text{tr} \mathbf{A} \bar{\mathbf{Q}}(z) dz + o(1), \\ \mathbf{a}^\top f(\mathbf{M}) \mathbf{b} &= -\frac{1}{2\pi i} \oint_{\Gamma} f(z) \mathbf{a}^\top \mathbf{Q}_{\mathbf{M}}(z) \mathbf{b} dz \\ &= -\frac{1}{2\pi i} \oint_{\Gamma} f(z) \mathbf{a}^\top \bar{\mathbf{Q}}(z) \mathbf{b} dz + o(1), \end{aligned}$$

thereby giving access to the asymptotics of these eigenvector functionals.

Under the notations of Theorem 2.6, for $\mathbf{M} = \frac{1}{n} \mathbf{X} \mathbf{X}^\top$ the sample covariance matrix under study, we have, in particular,

$$\begin{aligned} \frac{1}{p} \text{tr} \mathbf{A} f\left(\frac{1}{n} \mathbf{X} \mathbf{X}^\top\right) &= \frac{1}{2\pi i} \oint_{\Gamma_\mu} \frac{f(z)}{z} \frac{1}{p} \text{tr} \mathbf{A} (\mathbf{I}_p + \tilde{m}(z)\mathbf{C})^{-1} dz + o(1) \quad (2.46) \\ \mathbf{a}^\top f\left(\frac{1}{n} \mathbf{X} \mathbf{X}^\top\right) \mathbf{b} &= \frac{1}{2\pi i} \oint_{\Gamma_\mu} \frac{f(z)}{z} \mathbf{a}^\top (\mathbf{I}_p + \tilde{m}(z)\mathbf{C})^{-1} \mathbf{b} dz + o(1) \end{aligned}$$

for Γ_μ a contour circling around the limiting spectral support $\text{supp}(\mu)$.

Example: Eigenspace Correlation

Returning to Figure 2.4, we have seen that, when the population covariance spectrum ν is a discrete measure $\nu = \sum_{a=1}^k \frac{p_a}{p} \delta_{\ell_a}$ and c is small enough, μ has a density that spreads in k connected components $\text{supp}(\mu) = \mathcal{S}_1 \cup \dots \cup \mathcal{S}_k$, with \mathcal{S}_a mapped to the atom ℓ_a of ν ; these connected components spread more when c increases. A natural subsequent question would be to know whether the eigenvectors $\hat{\mathbf{u}}_i$ s associated with the p_a eigenvalues of $\frac{1}{n} \mathbf{X} \mathbf{X}^T$ of a given connected component \mathcal{S}_a share the same eigenspace as that spanned by the eigenvectors \mathbf{u}_i s of \mathbf{C} corresponding to population eigenvalue ℓ_a (with multiplicity p_a) of ν .

This question can be answered by evaluating the following quantity

$$\frac{1}{p_a} \text{tr} \Pi_a \hat{\Pi}_a, \quad \Pi_a = \sum_{\lambda_i(\mathbf{C})=\ell_a} \mathbf{u}_i \mathbf{u}_i^T, \quad \hat{\Pi}_a = \sum_{j \sim \mathcal{S}_a} \hat{\mathbf{u}}_j \hat{\mathbf{u}}_j^T$$

and where the relation $j \sim \mathcal{S}_a$ stands for $\text{dist}(\lambda_j(\frac{1}{n} \mathbf{X} \mathbf{X}^T), \mathcal{S}_a) \rightarrow 0$, that is, those eigenvalues of $\frac{1}{n} \mathbf{X} \mathbf{X}^T$ converging to the limiting component \mathcal{S}_a of μ .

This quantity can be evaluated by letting $\mathbf{A} = \Pi_a$, $f(z) = 1$ and changing Γ_μ into $\Gamma_{\mathcal{S}_a}$, a contour surrounding *only* the component \mathcal{S}_a of $\text{supp}(\mu)$ in (2.46). We precisely get

$$\begin{aligned} \frac{1}{p_a} \text{tr} \Pi_a \hat{\Pi}_a &= \frac{1}{2\pi i} \oint_{\Gamma_{\mathcal{S}_a}} \frac{1}{z} \frac{1}{p_a} \text{tr} \Pi_a (\mathbf{I}_p + \tilde{m}(z) \mathbf{C})^{-1} dz + o(1) \\ &= \frac{1}{2\pi i} \oint_{\Gamma_{\mathcal{S}_a}} \frac{1}{z} \frac{1}{1 + \tilde{m}(z) \ell_a} dz + o(1), \end{aligned} \tag{2.47}$$

which, for a given population eigenvalue ℓ_a , can be evaluated numerically with the following two-step procedure:

- (i) with Theorem 2.10, determine the support of μ , which is assumed to have exactly k disjoint components, that is,

$$\text{supp}(\mu) = \bigcup_{a=1}^k \mathcal{S}_a, \quad \mathcal{S}_a = [s_a^-, s_a^+] \text{ with } s_a^+ < s_{a+1}^-, \tag{2.48}$$

- (ii) choose *any* licit contour $\Gamma_{\mathcal{S}_a}$ that carefully circles around *only* the component \mathcal{S}_a , for instance, the rectangular $\Gamma_{\mathcal{S}_a}$ as depicted in Figure 2.9 (which was adopted, for example, in Bai and Silverstein [2004]); then evaluate the integral numerically over this contour by solving the fixed-point defining equation of $\tilde{m}(z)$ in (2.33).

But we may go beyond this numerical evaluation and obtain an *explicit* expression of the integral. To this end, for the chosen rectangular contour $\Gamma_{\mathcal{S}_a}$ in Figure 2.9, this consists in evaluating the sum of four line integrals (two ‘‘horizontal’’ and two ‘‘vertical’’). We provide here the full derivation as it is instrumental of many such calculus arising in similar inference problems and, to the best of our knowledge, this

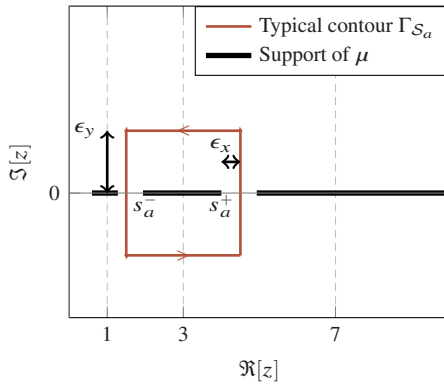


Figure 2.9 Typical contour Γ_{S_a} , for $\nu = \frac{1}{3}(\delta_1 + \delta_3 + \delta_7)$ with $c = 1/10$.

specific calculus was not derived elsewhere in the random matrix literature. Let us first focus on the sum of the two horizontal integrals

$$\int_{s_a^+ + \epsilon_x}^{s_a^- - \epsilon_x} g(x + i\epsilon_y) dx + \int_{s_a^- - \epsilon_x}^{s_a^+ + \epsilon_x} g(x - i\epsilon_y) dx$$

for $g(z) \equiv \frac{1}{z} \frac{1}{1 + \tilde{m}(z)\ell_a}$ our object of interest here. Note from the definition of Stieltjes transform, Definition 3, that³²

$$\Re[m(x + iy)] = \Re[m(x - iy)], \quad \Im[m(x + iy)] = -\Im[m(x - iy)],$$

for any Stieltjes transform $m(z)$ and, consequently,

$$\Re[g(x + iy)] = \Re[g(x - iy)], \quad \Im[g(x + iy)] = -\Im[g(x - iy)].$$

A direct consequence of this observation is that

$$\int_{s_a^+ + \epsilon_x}^{s_a^- - \epsilon_x} g(x + i\epsilon_y) dx + \int_{s_a^- - \epsilon_x}^{s_a^+ + \epsilon_x} g(x - i\epsilon_y) dx = -2i \int_{s_a^- - \epsilon_x}^{s_a^+ + \epsilon_x} \Im[g(x + i\epsilon_y)] dx$$

and thus only the imaginary part of $g(z) = \frac{1}{z} \frac{1}{1 + \tilde{m}(z)\ell_a}$ remains, which is explicitly given by

$$\Im[g(x + i\epsilon_y)] = -\frac{\epsilon_y + \ell_a (x\Im[\tilde{m}(x + i\epsilon_y)] + \epsilon_y\Re[\tilde{m}(x + i\epsilon_y)])}{(x^2 + \epsilon_y^2) (1 + 2\ell_a\Re[\tilde{m}(x + i\epsilon_y)] + \ell_a^2|\tilde{m}(x + i\epsilon_y)|^2)}.$$

As for the two vertical integrals (from $-\epsilon_y$ to ϵ_y), we expect that, in the limit $\epsilon_y \rightarrow 0$, they can be neglected. This is indeed the case as we know from Theorem 2.10 that the limit

$$\tilde{m}^\circ(x) = \lim_{\epsilon_y \downarrow 0} \tilde{m}(x + i\epsilon_y) = \lim_{\epsilon_y \downarrow 0} (\tilde{m}(x - i\epsilon_y))^*$$

(with $(\cdot)^*$ the complex conjugate) exists and is real for $x \notin \text{supp}(\mu)$, so that $g(z)$ is continuous on the vertical lines and the vertical integrals thus vanish as $\epsilon_y \rightarrow 0$. The

³² This second equality is reminiscent of the property $\Im[m(z)] \cdot \Im[z] > 0$ that immediately follows Definition 3.

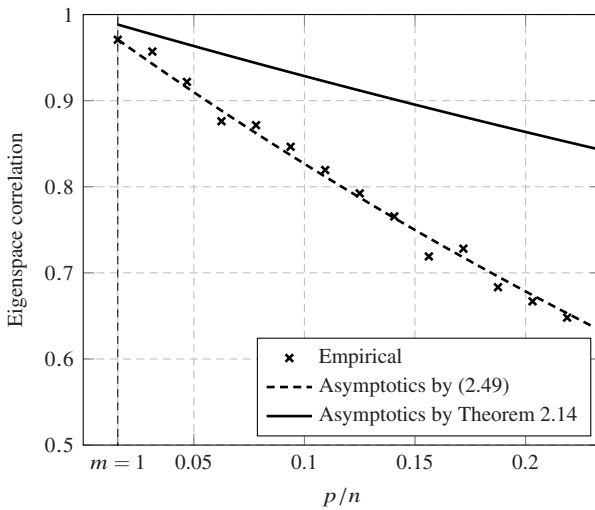


Figure 2.10 Empirical versus limiting eigenspace correlation as a function of $\frac{p}{n} = \frac{p}{m} \frac{m}{n}$ for $\nu = (1 - \frac{m}{p})\delta_1 + \frac{m}{p}\delta_2$, $\frac{m}{p} = \frac{1}{16}$ and $n = 1024$. Code on web: MATLAB and Python.

resulting complex integral thus corresponds to the limit of the horizontal integrals for $\epsilon_y \rightarrow 0$. As opposed to the vertical integrals though, for every $x \in \text{supp}(\mu)$, $\tilde{m}^\circ(x)$ is of positive imaginary part, so that the limits of $m(z)$ and thus of $g(z)$, for $z = x \pm i\epsilon_y$, come in conjugate pairs as $\epsilon_y \downarrow 0$. This finally leads to

$$\frac{1}{p_a} \text{tr} \Pi_a \hat{\Pi}_a = \frac{1}{\pi} \int_{s_a^-}^{s_a^+} \frac{\ell_a \Im[\tilde{m}^\circ(x)]}{1 + 2\ell_a \Re[\tilde{m}^\circ(x)] + \ell_a^2 |\tilde{m}^\circ(x)|^2} \frac{dx}{x} + o(1), \tag{2.49}$$

where we recall from Theorem 2.10 that, for x inside the support, $\tilde{m}^\circ(x)$ is the unique solution with positive imaginary part of

$$\tilde{m}^\circ(x) = \left(-x + c \int \frac{t\nu(dt)}{1 + \tilde{m}^\circ(x)t} \right)^{-1}.$$

We will show in Section 2.5 on “spiked models” that when the multiplicity p_a of atom ℓ_a is small – technically, if one assumes that $p_a = O(1)$ with respect to p – the alignment $\text{tr} \Pi_a \hat{\Pi}_a$ just derived takes a much simpler and fully explicit form, see Theorem 2.14. Yet, the present estimate, which we set under the scenario where $p_a = O(p)$, turns out (as numerical observations in Figure 2.10 suggest) to be as well precise even when p_a is small, at least in the setting of Figure 2.10.

To make this claim more visual, consider the setting where the population covariance $\mathbf{C} \in \mathbb{R}^{p \times p}$ has its $p - m$ eigenvalues equal to 1 and the remaining m eigenvalues equal to $\ell > 1$, so that the population spectral measure ν is a discrete measure having two components: $\nu = \frac{p-m}{p}\delta_1 + \frac{m}{p}\delta_\ell$. In the case where $m, n, p \rightarrow \infty$ with $\lim m/p, \lim p/n \in (0, \infty)$, the correlation of eigenspaces that corresponds to the leading eigenvalues of \mathbf{C} (equal to ℓ with multiplicity m) and those of $\frac{1}{n} \mathbf{X} \mathbf{X}^T$ can be fully characterized by (2.49). Figure 2.10 compares the empirical eigenspace correlation with

different limiting behaviors predicted by the “separate bulk” model in (2.49) versus the spiked model introduced later in Theorem 2.14. For small values of m , both limiting predictions are close, although (2.49) already shows a surprisingly marked advantage over the spiked model, even though $m \ll p$ (which goes against our assumptions). But as m increases, the spiked model-based Theorem 2.14 tends to overestimate the correlation, while the prediction (2.49) is a close match to the empirical output.

This observation, which appears to be quite systematic in random matrix theory, is interesting from an application perspective: In practice, \mathbf{C} is fixed (instead of growing size) and so are m, p , and n . Yet, the random matrix predictions based on simultaneously large m, p, n are always extremely accurate and, most importantly, *systematically more accurate* than when one assumes one of the dimensions (be it m, p , or n) is fixed.

As a side remark, if we only have access to the empirical covariance $\frac{1}{n}\mathbf{X}^T\mathbf{X}$ and its Stieltjes transform (i.e., if \mathbf{C} is unknown), then the contour integration in (2.47) asymptotically and practically reduces to residue calculus as

$$\begin{aligned} \frac{1}{p_a} \operatorname{tr} \Pi_a \hat{\Pi}_a &= \frac{1}{2\pi i} \oint_{\Gamma_{S_a}} \frac{1}{z} \frac{1}{1 + m \frac{1}{n} \mathbf{X}^T \mathbf{X}(z) \ell_a} dz + o(1) \\ &= \sum_{i \sim S_a} \frac{-m \frac{1}{n} \mathbf{X}^T \mathbf{X}(\zeta_i)}{\zeta_i m' \frac{1}{n} \mathbf{X}^T \mathbf{X}(\zeta_i)} + o(1), \end{aligned}$$

with ζ_i s the roots of $m \frac{1}{n} \mathbf{X}^T \mathbf{X}(\zeta_i) = -1/\ell_a$, which, by an argument similar to Remark 2.12, are the (sorted) eigenvalues of $\Lambda + \frac{\ell_a}{n} \mathbf{1}_n \mathbf{1}_n^T$, for Λ the diagonal matrix containing the eigenvalues of $\frac{1}{n} \mathbf{X}^T \mathbf{X}$. The residue calculus technique performed in the last equation was described in the previous section.

Eigenvector Inference and Subspace Methods

The second interest of the deterministic equivalent $\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z)$ of Theorem 2.6, already underlined in the previous example, now concerns the statistical inference of the eigenvectors and eigenspaces of \mathbf{C} . Unless a strong a priori structure is imposed, the eigenvectors themselves *cannot* be consistently estimated from \mathbf{X} (especially “large” eigenspaces involving $O(p^2)$ parameters, which cannot be estimated from the $O(pn)$ data observations). But their *scalar* projections onto some deterministic vectors are accessible. Precisely, for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$ of bounded Euclidean norm, denoting Π_i a projector on the eigenspace associated with the eigenvalue $\lambda_i(\mathbf{C})$,

$$\mathbf{a}^T \Pi_i \mathbf{b} = -\frac{1}{2\pi i} \oint_{\Gamma_v^i} \mathbf{a}^T (\mathbf{C} - z\mathbf{I}_p)^{-1} \mathbf{b} dz$$

for Γ_v^i a contour circling around $\lambda_i(\mathbf{C})$ only. From Theorem 2.6 and our subsequent discussions in Section 2.3, it is strongly desirable to use again the variable change $z = -1/\tilde{m}(\omega)$ in order to estimate $\mathbf{a}^T \Pi_i \mathbf{b}$ from an integral over $\mathbf{a}^T \mathbf{Q}(z) \mathbf{b}$ involving the resolvent $\mathbf{Q}(z)$. However, this is again *only possible* if there exists a pair of contours (Γ_v^i, Γ) such that $-1/\tilde{m}(\Gamma) = \Gamma_v^i$. This is, in general, not possible unless $\lambda_i(\mathbf{C})$ “induces” its own associated connected component in $\operatorname{supp}(\mu)$, see illustrations in

Figures 2.5 and 2.6. Assuming the validity of such variable change, we thus have

$$\begin{aligned} \mathbf{a}^\top \Pi_i \mathbf{b} &= -\frac{1}{2\pi i} \oint_{\Gamma} \mathbf{a}^\top \left(\mathbf{C} + \frac{1}{\tilde{m}(\omega)} \mathbf{I}_p \right)^{-1} \mathbf{b} \cdot \frac{\tilde{m}'(\omega)}{\tilde{m}^2(\omega)} d\omega \\ &= \frac{1}{2\pi i} \oint_{\Gamma} \mathbf{a}^\top \mathbf{Q}(\omega) \mathbf{b} \cdot \frac{\omega \tilde{m}'(\omega)}{\tilde{m}(\omega)} d\omega + o(1). \end{aligned}$$

This formula reveals handy when testing whether an expected “structure” vector $\mathbf{a} \in \mathbb{R}^p$ is present in the dominant subspace associated with, say, the largest eigenvalue (possibly with multiplicity) $\lambda_1(\mathbf{C})$ of the data covariance structure \mathbf{C} . The value $\mathbf{a}^\top \Pi \mathbf{a} / \|\mathbf{a}\|^2 \in [0, 1]$ precisely evaluates a score for the structure vector \mathbf{a} to be in the span of the dominant eigenvectors of \mathbf{C} .

This analysis finds several applications in detection and estimation, notably in the field of array processing. A concrete example, the G-MUSIC algorithm, is discussed in Section 3.1.3 but more results are available in the dedicated array processing literature [Mestre and Lagunas, 2008, Kammoun et al., 2017].

2.5 Spiked Models

The statistical methods discussed in the previous sections for the sample covariance matrix model offer a flexible estimation and inference framework, which can be extended to a large spectrum of random matrix models. However, they have a certain number of practical limitations: (i) they rely on the implicit nature of Theorem 2.6 and thus their behavior is not easily understood, (ii) the complex integration framework, while theoretically satisfying, may be difficult to handle in practice (conditions of the existence of valid contours need to be ensured, the complex integrals do not necessarily lend themselves to simple analytical evaluation, etc.).

In this section, we will consider a very special, yet practically far-reaching, case of sample covariance matrix models for which the limiting spectral measure coincides with the Marčenko–Pastur law, while the population covariance matrix has a nontrivial informative structure. Since the Marčenko–Pastur law assumes an explicit well-understood expression (recall Theorem 2.4), the various estimates of interest will be explicit, and thus intuitions on their behavior are easily derived. Besides, the various change of variable difficulties for contour integral methods met in the previous sections are greatly simplified in this setting.

These special models fundamentally rely on letting the covariance matrix \mathbf{C} be a *low-rank* perturbation of the identity matrix \mathbf{I}_p , that is, $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ for $\mathbf{P} \in \mathbb{R}^{p \times p}$ with $\text{rank}(\mathbf{P}) = k$ fixed with respect to n, p .

Such statistical models corresponding to a *low-rank update* of a classical random matrix model with well-known behavior are generically called *spiked models*.

2.5.1 Isolated Eigenvalues

Let us then consider again the model $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}} \mathbf{z}_i$, $\mathbf{z}_i \in \mathbb{R}^p$ with standard i.i.d. entries and where

$$\mathbf{C} = \mathbf{I}_p + \mathbf{P}, \quad \mathbf{P} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^\top$$

with k and $\ell_1 \geq \dots \geq \ell_k > 0$ fixed with respect to n, p .

According to Theorem 2.6, the spectral measure $\mu_{\frac{1}{n} \mathbf{X} \mathbf{X}^\top}$ admits a limit μ defined through the limiting spectral measure ν of \mathbf{C} . Note that here $\nu = \delta_1$ since

$$\mu_{\mathbf{C}} = \frac{p-k}{p} \delta_1 + \frac{1}{p} \sum_{i=1}^k \delta_{1+\ell_i} \rightarrow \delta_1$$

as $p \rightarrow \infty$. As a consequence, while \mathbf{C} is not the identity matrix, the limiting μ is the Marčenko–Pastur law introduced in Theorem 2.4. However, note importantly that the conditions for “no eigenvalue outside the support,” Theorem 2.11, do *not* hold here since $\text{dist}(1 + \ell_i, \text{supp}(\nu)) \not\rightarrow 0$ for $i \in \{1, \dots, k\}$. Therefore, one *cannot* claim that *all* the eigenvalues of $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ will lie within the support $\text{supp}(\mu)$.

We will precisely show here that, depending on the values of ℓ_i and the ratio $c = \lim p/n$, the i th largest eigenvalue $\hat{\lambda}_i$ of $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ may indeed *isolate* from $\text{supp}(\mu)$. As such, since most of the eigenvalues of $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ aggregate, except possibly for a few ones (up to k of them), the latter isolated eigenvalues are seen as isolated “spikes” in the histogram of eigenvalues. See Figure 2.11, commented next, for a visual representation of these “spikes.”

This specific result, due to Baik (not Bai) and Silverstein, is given in the following theorem.³³

Theorem 2.13 (Spiked eigenvalues, Baik and Silverstein [2006]). *Under the setting of Theorem 2.6 with $\mathbb{E}[Z_{ij}^4] < \infty$, let $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with $\mathbf{P} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^\top$ its spectral decomposition, where k and $\ell_1 \geq \dots \geq \ell_k > 0$ are fixed with respect to n, p . Then, denoting $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$ the eigenvalues of $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$, as $n, p \rightarrow \infty$ with $p/n \rightarrow c \in (0, \infty)$,*

$$\hat{\lambda}_i \xrightarrow{\text{a.s.}} \begin{cases} \lambda_i = 1 + \ell_i + c \frac{1+\ell_i}{\ell_i} > (1 + \sqrt{c})^2 & , \ell_i > \sqrt{c} \\ (1 + \sqrt{c})^2 & , \ell_i \leq \sqrt{c}. \end{cases}$$

The theorem thus identifies an abrupt change in the behavior of the i th dominant eigenvalue $\hat{\lambda}_i$ of $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$: if $\ell_i \leq \sqrt{c}$, $\hat{\lambda}_i$ converges to the right-edge $(1 + \sqrt{c})^2$ of the

³³ These same results were later retrieved using a free probability approach (so formally under slightly different assumptions on \mathbf{Z}) in the work of, for example, Benaych-Georges and Nadakuditi [2011] and were generalized to a larger class of random matrix models. This last article, a richer set of dedicated and better digested techniques (such as those exposed presently), as well as the growing evidence of fundamental applications of these results [Bianchi et al., 2011, Donoho et al., 2018, Candès et al., 2015, Couillet, 2015, Couillet and Hachem, 2013], triggered a renewed wave of interest for spiked models [Loubaton and Vallet, 2011, Capitaine, 2014].

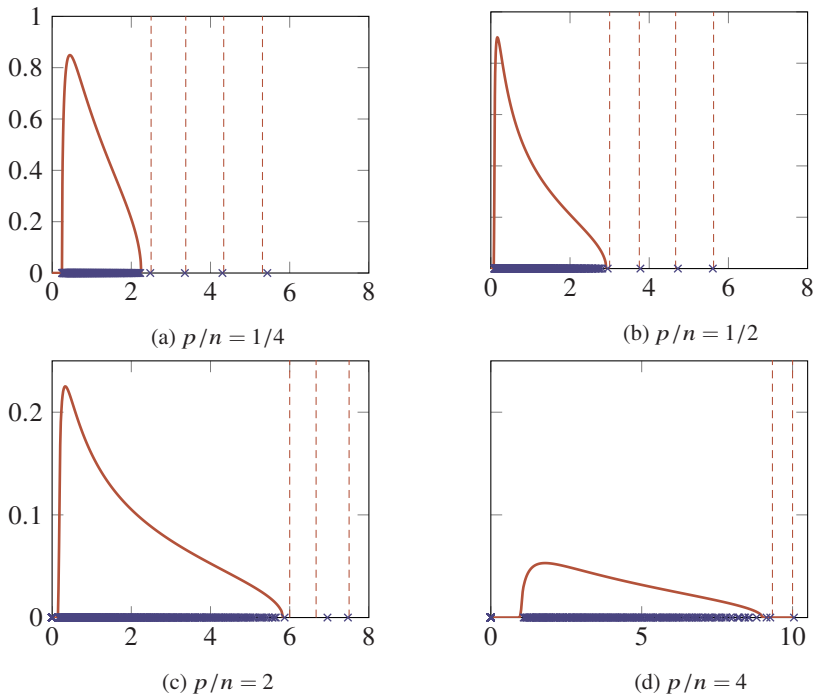


Figure 2.11 Eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$ (blue crosses), the Marčenko–Pastur law (red solid line), and asymptotic spike locations (red dashed line), for $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$, $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with $\mu_{\mathbf{P}} = \frac{p-4}{p}\delta_0 + \frac{1}{p}(\delta_1 + \delta_2 + \delta_3 + \delta_4)$, for $p = 1024$ and different values of n . (a) $p/n = 1/4$, (b) $p/n = 1/2$, (c) $p/n = 2$, and (d) $p/n = 4$. Code on web: MATLAB and Python.

support of the Marčenko–Pastur law μ and thus does *not* isolate. However, as soon as $\ell_i > \sqrt{c}$, $\hat{\lambda}_i$ converges to a limit *beyond* the right-edge of μ and thus *does* isolate from the Marčenko–Pastur support. Note in passing that the transition is smooth as the limit of λ_i as $\ell_i \rightarrow \sqrt{c}$ indeed coincides with $(1 + \sqrt{c})^2$, so there is no “sudden jump” of the limiting eigenvalue location at the $\ell_i = \sqrt{c}$ transition point.

With a physics inspiration, this phenomenon is often referred to as the *phase transition* of the spiked models.

From a statistical viewpoint, the fact that the i th eigenvalue $\hat{\lambda}_i$ of the sample covariance matrix $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$ “macroscopically” exceeds or not the other eigenvalues depending on whether $\ell_i > \sqrt{c}$ or $\ell_i \leq \sqrt{c}$ can be interpreted as a test of whether the “signal strength” ℓ_i of the low-rank structure exceeds the minimal *detectability* threshold \sqrt{c} : This can be achieved if the signal strength ℓ_i is itself strong enough, or alternatively if the number of observed independent data n is large enough (so that $c = \lim p/n$ is small), as common sense would suggest. Indeed, if $\ell_1 < \sqrt{c}$, the eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$ are all asymptotically compacted in the support $[(1 - \sqrt{c})^2, (1 + \sqrt{c})^2]$ and thus it is theoretically (asymptotically) impossible to tell whether $\mathbf{C} = \mathbf{I}_p$ or \mathbf{C} is more structured from the mere observation of the eigenspectrum $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$. This phase transition effect, for all successive spikes, is well illustrated in Figure 2.11.

Remark 2.13 (The case of negative ℓ_i s). *Baik and Silverstein [2006] in fact generalized the result in Theorem 2.13 to account for possibly negative ℓ_i s, that is, $\ell_i \in (-1, 0)$. In this situation, the following interesting phenomenon occurs: (i) if $c < 1$ and $\ell_i < -\sqrt{c}$, there exists an associated eigenvalue of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$, which converges to $1 + \ell_i + c(1 + \ell_i)/\ell_i \in (0, (1 - \sqrt{c})^2)$ so on the left-hand side of the limiting Marčenko–Pastur support; (ii) if $c < 1$ and $\ell_i \geq -\sqrt{c}$, the associated eigenvalue converges to the left edge $(1 - \sqrt{c})^2$; (iii) if $c > 1$ (so, in particular, since $\ell_i > -1$, one cannot have $\ell_i < -\sqrt{c}$), the corresponding eigenvalue tends to 0: So it is never possible to find isolated eigenvalues in the “empty space” $(0, (1 - \sqrt{c})^2)$ when $c > 1$.*

This negative- ℓ_i setting in effect finds no practical applications that we are aware of, and would additionally be cumbersome to integrate (due to heavier indexing) into a more general statement of Theorem 2.13.

Proof of Theorem 2.13. When it comes to assessing the eigenvalues of a given matrix \mathbf{M} , the first thing that comes to mind is to solve the determinant equation $\det(\mathbf{M} - \hat{\lambda}\mathbf{I}) = 0$. This approach is not convenient for $\mathbf{M} = \frac{1}{n}\mathbf{X}\mathbf{X}^\top$ of increasing dimensions and we have seen that the Stieltjes transform and resolvent method is an appropriate substitute in that case. Here, since the low-rank matrix \mathbf{P} only induces a low-rank perturbation of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top$, the use of Sylvester’s identity, Lemma 2.3, will turn the large-dimensional determinant equation into a small (fixed)-dimensional one, and the determinant equation method can then be applied. This is the approach we pursue here.

Specifically, let us seek for the presence of an eigenvalue $\hat{\lambda}$ of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top$ that is asymptotically *greater than* $(1 + \sqrt{c})^2$. Our approach is to “isolate” the low-rank contribution due to \mathbf{P} from the “whitened” sample covariance matrix model $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top$ with identity covariance. To this end, we write, with $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$,

$$\begin{aligned} 0 &= \det\left(\frac{1}{n}\mathbf{X}\mathbf{X}^\top - \hat{\lambda}\mathbf{I}_p\right) \\ &= \det\left(\frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^\top(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} - \hat{\lambda}\mathbf{I}_p\right) \\ &= \det(\mathbf{I}_p + \mathbf{P}) \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top - \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\right). \end{aligned}$$

Since $\det(\mathbf{I}_p + \mathbf{P}) \neq 0$, the first determinant can be discarded. For the second determinant, first recall from the resolvent identity, Lemma 2.1, that

$$(\mathbf{I}_p + \mathbf{P})^{-1} = \mathbf{I}_p - (\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P},$$

so that we can isolate the (now well-understood) resolvent of the “whitened” model. That is, letting $\mathbf{Q}(\hat{\lambda}) = (\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top - \hat{\lambda}\mathbf{I}_p)^{-1}$, we write

$$\begin{aligned} 0 &= \det\left(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top - \hat{\lambda}\mathbf{I}_p + \hat{\lambda}(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}\right) \\ &= \det\mathbf{Q}^{-1}(\hat{\lambda}) \det(\mathbf{I}_p + \hat{\lambda}\mathbf{Q}(\hat{\lambda})(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P}). \end{aligned} \tag{2.50}$$

Thanks to Theorem 2.11, inverting the matrix $\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top - \hat{\lambda}\mathbf{I}_p$ is (almost surely) licit for all large n, p as we demanded $\hat{\lambda} > (1 + \sqrt{c})^2$. Now, considering the spectral decomposition $\mathbf{P} = \mathbf{U}\mathbf{L}\mathbf{U}^\top$ with $\mathbf{L} = \text{diag}\{\ell_1, \dots, \ell_k\}$ and $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_k] \in \mathbb{R}^{p \times k}$, we further have

$$(\mathbf{I}_p + \mathbf{P})^{-1}\mathbf{P} = (\mathbf{I}_p + \mathbf{U}\mathbf{L}\mathbf{U}^\top)^{-1}\mathbf{U}\mathbf{L}\mathbf{U}^\top = \mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\mathbf{U}^\top.$$

Plugging into (2.50), this is

$$\begin{aligned} 0 &= \det \mathbf{Q}^{-1}(\hat{\lambda}) \det \left(\mathbf{I}_p + \hat{\lambda}\mathbf{Q}(\hat{\lambda})\mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L}\mathbf{U}^\top \right) \\ &= \det \mathbf{Q}^{-1}(\hat{\lambda}) \det \left(\mathbf{I}_k + \hat{\lambda}\mathbf{U}^\top\mathbf{Q}(\hat{\lambda})\mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L} \right), \end{aligned}$$

where in the last equality we applied Sylvester’s identity, Lemma 2.3. Since $\det \mathbf{Q}^{-1}(\hat{\lambda}) = \det(\frac{1}{n}\mathbf{Z}\mathbf{Z}^\top - \hat{\lambda}\mathbf{I}_p)$ does not vanish for all large n, p at $\hat{\lambda} > (1 + \sqrt{c})^2$, we finally have, for all large n, p , the following determinant equation for a much smaller matrix (of size $k \times k$)

$$0 = \det \left(\mathbf{I}_k + \hat{\lambda}\mathbf{U}^\top\mathbf{Q}(\hat{\lambda})\mathbf{U}(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L} \right).$$

Applying Theorem 2.4 entry-wise to each entry of the $k \times k$ matrix $\mathbf{U}^\top\mathbf{Q}(\hat{\lambda})\mathbf{U}$ (this is the step where it is fundamental that k remains finite as $n, p \rightarrow \infty$), we now know that

$$\mathbf{U}^\top\mathbf{Q}(\hat{\lambda})\mathbf{U} = m(\hat{\lambda})\mathbf{I}_k + o_{\|\cdot\|}(1)$$

almost surely, for $m(z)$ the Stieltjes transform of the Marčenko–Pastur law μ (the term \mathbf{I}_k arises from the fact that $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_k$). Consequently, by continuity of the determinant (this is a polynomial of its entries), we have

$$0 = \det \left(\mathbf{I}_k + \hat{\lambda}m(\hat{\lambda})(\mathbf{I}_k + \mathbf{L})^{-1}\mathbf{L} \right) + o(1)$$

and thus, if such a $\hat{\lambda}$ exists, it must satisfy

$$\hat{\lambda}m(\hat{\lambda}) = -\frac{1 + \ell_i}{\ell_i} + o(1),$$

for some $i \in \{1, \dots, k\}$.

We thus need to understand when the above equation has a solution. To this end, observe that the function $\mathbb{R} \setminus \text{supp}(\mu) \rightarrow \mathbb{R}, x \mapsto xm(x) = \int \frac{x}{t-x} \mu(dt)$ is increasing on its domain of definition and that $xm(x) \rightarrow -1$ as $x \rightarrow \infty$. Note from Theorem 2.4

$$zcm^2(z) - (1 - c - z)m(z) + 1 = 0 \Leftrightarrow zm(z) = -1 + \frac{1}{1 - z - czm(z)}, \tag{2.51}$$

so that we can express $zm(z)$ as a function of c and z (alternatively, we could use the explicit solution for $m(z)$ in the proof of the Marčenko–Pastur law, but this is slightly more cumbersome), so to obtain

$$\lim_{x \in \mathbb{R} \downarrow (1 + \sqrt{c})^2} xm(x) = -\frac{1 + \sqrt{c}}{\sqrt{c}}.$$

Thus, $xm(x)$ increases from $-\frac{1 + \sqrt{c}}{\sqrt{c}}$ to -1 on the set $((1 + \sqrt{c})^2, \infty)$. The equation $\hat{\lambda}m(\hat{\lambda}) = -\frac{1 + \ell_i}{\ell_i}$ thus has a solution if and only if $\ell_i > \sqrt{c}$ for some $i \in \{1, \dots, k\}$.

Assuming this holds, we may then use again (2.51) (replacing $zm(z)$ by $-(1 + \ell_i)/\ell_i$) to obtain

$$\hat{\lambda}_i \rightarrow \lambda_i = 1 + \ell_i + c \frac{1 + \ell_i}{\ell_i},$$

which concludes the proof of Theorem 2.13. □

Figure 2.11 depicts the eigenvalues of $\frac{1}{n}\mathbf{XX}^\top$ versus the Marčenko–Pastur law, in the scenario where $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with \mathbf{P} of rank four, for various ratios p/n . As predicted by Theorem 2.13, the number of visible “spikes” outside the limiting Marčenko–Pastur law support varies with p/n : As the ratio decreases, less spikes are visible. We also note that, for fixed p , the asymptotic characterization in Theorem 2.13 becomes less accurate as n decreases.

2.5.2 Isolated Eigenvectors

From a practical standpoint, we have seen that the presence of isolated eigenvalues in the spectrum of the sample covariance $\frac{1}{n}\mathbf{XX}^\top$ reveals the presence of some “structure” in the population covariance \mathbf{C} in the sense that $\mathbf{C} \neq \mathbf{I}_p$. We have however also seen that the converse is not true: assuming a spiked model for \mathbf{C} , the absence of isolated eigenvalue does not *always* imply $\mathbf{C} = \mathbf{I}_p$.

More interestingly, whether this “structure” is detected or not, one may wonder whether it can be estimated at all. More specifically, for $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with $\mathbf{P} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^\top$, are the eigenvectors $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_k$ of $\frac{1}{n}\mathbf{XX}^\top$ associated with its k largest eigenvalues $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_k$ good estimators of $\mathbf{u}_1, \dots, \mathbf{u}_k$?

Not surprisingly, as in Theorem 2.13 for the spiked eigenvalues, the answer is here again twofold: (i) if $\ell_i \leq \sqrt{c}$, then $\hat{\mathbf{u}}_i$ tends to be totally *uncorrelated* from and thus asymptotically orthogonal to \mathbf{u}_i ; ³⁴ while (ii) if $\ell_i > \sqrt{c}$, $\hat{\mathbf{u}}_i$ is, to some extent, *aligned* to \mathbf{u}_i . The following theorem, due to Paul [2007], quantifies this “to some extent.” ³⁵

Theorem 2.14 (Spiked eigenvector alignment, Paul [2007]). *Under the setting of Theorem 2.13, let $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_k$ be the eigenvectors associated with the largest k eigenvalues $\hat{\lambda}_1 > \dots > \hat{\lambda}_k$ of $\frac{1}{n}\mathbf{XX}^\top$. Further assume that $\ell_1 > \dots > \ell_k > 0$ are all distinct. Then, for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$ unit norm deterministic vectors*

$$\mathbf{a}^\top \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \mathbf{b} - \mathbf{a}^\top \mathbf{u}_i \mathbf{u}_i^\top \mathbf{b} \cdot \frac{1 - c\ell_i^{-2}}{1 + c\ell_i^{-1}} \cdot 1_{\ell_i > \sqrt{c}} \xrightarrow{\text{a.s.}} 0. \tag{2.52}$$

³⁴ In the “unstructured” case of $\mathbf{C} = \mathbf{I}_p$ and Gaussian \mathbf{Z} (i.e., the so-called Gaussian orthogonal ensemble, GOE), it is known that the eigenvectors of the resulting Wishart matrix are uniformly distributed on the unit sphere \mathbb{S}^{p-1} [Anderson et al., 2010, Section 2.5.1] (or equivalently, according to the Haar measure, see more details in Section 2.6.2) that is close to, for p large, a Gaussian distributed random vector with i.i.d. entries. The same holds for the eigenvectors of Wigner matrix in Theorem 2.5.

³⁵ Here again, a large body of literature and modernized tools were set in place to study asymptotic eigenvector behaviors. Some of them are only valid (or only convenient) for rank-one spike models [Paul, 2007, Benaych-Georges and Nadakuditi, 2011, 2012], but the techniques now widely used (such as the contour-integral method presented in this book) generally apply to an arbitrary (but fixed) number of spikes [Couillet and Hachem, 2013, Baik et al., 2005].

In particular, with $\mathbf{a} = \mathbf{b} = \mathbf{u}_i$, we obtain

$$|\mathbf{u}_i^\top \hat{\mathbf{u}}_i|^2 \xrightarrow{\text{a.s.}} \zeta_i \equiv \frac{1 - c\ell_i^{-2}}{1 + c\ell_i^{-1}} \cdot 1_{\ell_i > \sqrt{c}}. \tag{2.53}$$

Proof of Theorem 2.14. We first write that, for all large n, p almost surely and $\ell_i > \sqrt{c}$,

$$\mathbf{a}^\top \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \mathbf{b} = -\frac{1}{2\pi i} \oint_{\Gamma_{\lambda_i}} \mathbf{a}^\top \left(\frac{1}{n} \mathbf{X} \mathbf{X}^\top - z \mathbf{I}_p \right)^{-1} \mathbf{b} dz,$$

for Γ_{λ_i} a small contour enclosing *only* the almost sure limit $\lambda_i = 1 + \ell_i + c \frac{1+\ell_i}{\ell_i}$ of the eigenvalue $\hat{\lambda}_i$ of $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ given in Theorem 2.13. Isolating $\frac{1}{n} \mathbf{Z} \mathbf{Z}^\top$ from $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ as in the proof of Theorem 2.13, we have

$$\begin{aligned} & \mathbf{a}^\top \left(\frac{1}{n} \mathbf{X} \mathbf{X}^\top - z \mathbf{I}_p \right)^{-1} \mathbf{b} \\ &= \mathbf{a}^\top \left(\frac{1}{n} (\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} \mathbf{Z} \mathbf{Z}^\top (\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} - z \mathbf{I}_p \right)^{-1} \mathbf{b} \\ &= \mathbf{a}^\top (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \left(\frac{1}{n} \mathbf{Z} \mathbf{Z}^\top - z \mathbf{I}_p + z (\mathbf{I}_p + \mathbf{P})^{-1} \mathbf{P} \right)^{-1} (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{b} \end{aligned}$$

with $\mathbf{Q}(z) = (\frac{1}{n} \mathbf{Z} \mathbf{Z}^\top - z \mathbf{I}_p)^{-1}$, where we used $(\mathbf{I}_p + \mathbf{P})^{-1} = \mathbf{I}_p - (\mathbf{I}_p + \mathbf{P})^{-1} \mathbf{P}$ from Lemma 2.1. It then follows from the spectral decomposition that $(\mathbf{I}_p + \mathbf{P})^{-1} \mathbf{P} = \mathbf{U} (\mathbf{I}_k + \mathbf{L})^{-1} \mathbf{L} \mathbf{U}^\top$ for $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_k] \in \mathbb{R}^{p \times k}$ and $\mathbf{L} = \text{diag}\{\ell_i\}_{i=1}^k$ so that

$$\begin{aligned} & \mathbf{a}^\top \left(\frac{1}{n} \mathbf{X} \mathbf{X}^\top - z \mathbf{I}_p \right)^{-1} \mathbf{b} \\ &= \mathbf{a}^\top (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{Q}(z) (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{b} \\ &\quad - z \mathbf{a}^\top (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{Q}(z) \mathbf{U} \left(\mathbf{I}_k + \mathbf{L}^{-1} + z \mathbf{U}^\top \mathbf{Q}(z) \mathbf{U} \right)^{-1} \mathbf{U}^\top \mathbf{Q}(z) (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{b} \\ &= \mathbf{a}^\top (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{Q}(z) (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{b} \\ &\quad - z \mathbf{a}^\top (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{Q}(z) \mathbf{U} (\mathbf{L}^{-1} + (1 + zm(z)) \mathbf{I}_k)^{-1} \mathbf{U}^\top \mathbf{Q}(z) (\mathbf{I}_p + \mathbf{P})^{-\frac{1}{2}} \mathbf{b} + o(1), \end{aligned}$$

where we used Woodbury identity, Lemma 2.7, for the first equality, and $\mathbf{U}^\top \mathbf{Q}(z) \mathbf{U} = m(z) \mathbf{I}_k + o_{\|\cdot\|}(1)$, as per Theorem 2.4, for the second equality.

Note here that the complex integration of $\mathbf{Q}(z)$ on the contour Γ_{λ_i} only brings a nontrivial residue for the second right-hand side term owing to the inverse $(\mathbf{L}^{-1} + (1 + zm(z)) \mathbf{I}_k)^{-1}$, which is singular at $z = \lambda_i$ according to the proof of Theorem 2.13. We thus finally have

$$\begin{aligned} \mathbf{a}^\top \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \mathbf{b} &= \frac{1}{2\pi i} \oint_{\Gamma_{\lambda_i}} zm^2(z) \mathbf{a}^\top \mathbf{U} (\mathbf{I}_k + \mathbf{L})^{-\frac{1}{2}} (\mathbf{L}^{-1} + (1 + zm(z)) \mathbf{I}_k)^{-1} \\ &\quad \times (\mathbf{I}_k + \mathbf{L})^{-\frac{1}{2}} \mathbf{U}^\top \mathbf{b} dz + o(1). \end{aligned}$$

This expression can then be evaluated by residue calculus at $z = \lambda_i$, the only singularity of the integrand, as

$$\lim_{z \rightarrow \lambda_i} (z - \lambda_i)(\mathbf{L}^{-1} + (1 + zm(z))\mathbf{I}_k)^{-1} = \frac{\mathbf{e}_i \mathbf{e}_i^\top}{m(\lambda_i) + \lambda_i m'(\lambda_i)}$$

with $\mathbf{e}_i \in \mathbb{R}^k$ the canonical basis vector defined as $[\mathbf{e}_i]_j = \delta_{ij}$. Using the (here most convenient) form

$$m(z) = \frac{1}{-z + \frac{1}{1+cm(z)}}$$

of the Stieltjes transform of the Marčenko–Pastur law gives³⁶

$$m'(z) = \frac{m^2(z)}{1 - \frac{cm^2(z)}{(1+cm(z))^2}}$$

from which we obtain, in particular, that $m(\lambda_i) = -1/(\ell_i + c)$ and $m'(\lambda_i) = \ell_i^2(\ell_i + c)^{-2}(\ell_i^2 - c)^{-1}$. We finally get

$$\mathbf{a}^\top \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^\top \mathbf{b} = \mathbf{a}^\top \mathbf{u}_i \mathbf{u}_i^\top \mathbf{b} \cdot \frac{1 - c\ell_i^{-2}}{1 + c\ell_i^{-1}} + o(1),$$

which concludes the proof of Theorem 2.14. □

Figure 2.12 compares, in a single-spike scenario, the theoretical limit ζ_1 of $|\hat{\mathbf{u}}_1^\top \mathbf{u}_1|^2$ versus its empirical value for different ℓ_1 and different p, n with constant ratio p/n . It is important to note that the theoretical *asymptotic* phase transition phenomenon at $\ell_1 = \sqrt{c}$ corresponds to a sharp nondifferentiable change in the function $\ell_1 \mapsto \zeta_1 = (1 - c\ell_1^{-2})/(1 + c\ell_1^{-1}) \cdot \mathbf{1}_{\ell_1 \geq \sqrt{c}}$; a local analysis in the limit of $\ell_1 = \sqrt{c} + \epsilon$ reveals that ζ_1 (and thus $|\hat{\mathbf{u}}_1^\top \mathbf{u}_1|^2$ in the large n, p limit) is locally equal to $\zeta_1 \simeq \frac{2\epsilon}{\sqrt{c}(1+\sqrt{c})}$ and therefore, for sufficiently large n, p ,

$$|\hat{\mathbf{u}}_1^\top \mathbf{u}_1| =_{\ell_1 = \sqrt{c} + \epsilon} \sqrt{\frac{2}{\sqrt{c}(1 + \sqrt{c})}} \cdot \sqrt{\epsilon} + O(\epsilon),$$

which has an infinite derivative as $\ell_1 \downarrow \sqrt{c}$. On real data of finite size, this sharp transition is only observed for extremely large values of n, p . This, in particular, means that, in practice, residual information of \mathbf{u}_1 is still present in $\hat{\mathbf{u}}_1$ below the phase transition threshold.

2.5.3 Limiting Fluctuations

Theorem 2.13 on the *limiting* presence and position of isolated eigenvalues in the spectrum of $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ establishes that it suffices to evaluate whether the largest

³⁶ In passing, note that $m'(z)$ assumes an *explicit form* as a function of z and $m(z)$. While not surprising in the Marčenko–Pastur case, this turns out to be also true of more elaborate models, where $m(z)$ does *not* have an explicit expression.

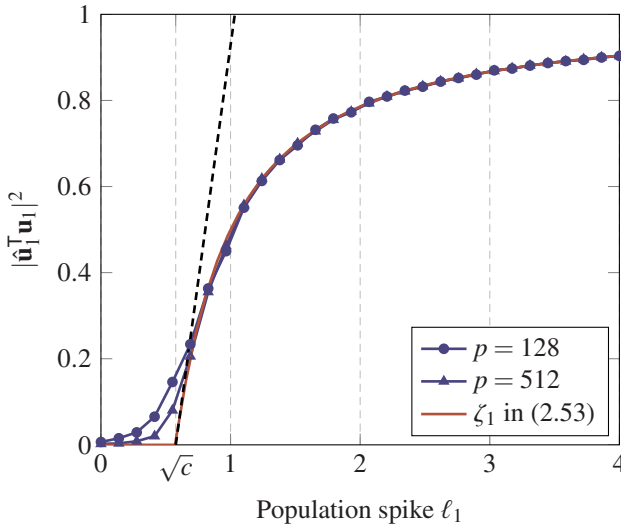


Figure 2.12 Empirical versus limiting $|\hat{\mathbf{u}}_1^T \mathbf{u}_1|^2$ for $\mathbf{X} = \mathbf{C}^{\frac{1}{2}} \mathbf{Z}$, $\mathbf{C} = \mathbf{I}_p + \ell_1 \mathbf{u}_1 \mathbf{u}_1^T$ and standard Gaussian \mathbf{Z} , $p/n = 1/3$, for different values of ℓ_1 . Results obtained by averaging over 200 runs. In **black** dashed line the local behavior around \sqrt{c} . Code on web: MATLAB and Python.

eigenvalue $\hat{\lambda}_1$ of $\frac{1}{n} \mathbf{X} \mathbf{X}^T$ “isolates from the other eigenvalues $\hat{\lambda}_2 > \dots > \hat{\lambda}_p$ ” to determine the presence of a structure in the population covariance \mathbf{C} (in the sense that $\mathbf{C} \neq \mathbf{I}_p$).

However, in practice, from the *finite-dimensional* observations $\hat{\lambda}_1, \dots, \hat{\lambda}_p$, how can one decide whether $\hat{\lambda}_1$ is isolated? On a random realization of \mathbf{X} , $\hat{\lambda}_1$ may haphazardly be found “rather far” from $\hat{\lambda}_2$ by a mere finite-dimensional probability effect. The natural question is then to determine whether the rate of occurrence of such “haphazard” events can be evaluated.

A whole line of works, based on rather different tools from the Stieltjes transform approach adopted in this book,³⁷ settles this question by evaluating, for $\mathbf{C} = \mathbf{I}_p$ or $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$ with the eigenvalues of \mathbf{P} below the phase transition threshold, the asymptotic probability for λ_1 to escape its limiting value $(1 + \sqrt{c})^2$. The main result of importance is the following.

³⁷ Unlike the Stieltjes transform method, these tools start from the *explicit* (finite-dimensional) formula of the joint eigenvalue distribution of Wishart or Wigner matrix, which is known in the Gaussian case (and only in this case) and given by (2.57). Exploiting the theory of orthogonal polynomials and determinantal processes, Equation (2.57) can be marginalized so to retrieve the exact (finite-dimensional) law of one or several specific eigenvalues (inside the bulk or on the edge). Taking the large-dimensional limit relates the law of the eigenvalues to the determinant of a specific kernel [Soshnikov, 2000] (Airy kernel for the edge eigenvalues [Johnstone, 2001, Soshnikov, 1999], sine kernel in the bulk [Arous and Pécché, 2005, Erdős et al., 2010]). Details on these techniques can be found in Anderson et al. [2010].

Theorem 2.15 (Fluctuation of the largest eigenvalue, Baik et al. [2005]). *Under the setting of Theorem 2.13, assume $0 \leq \ell_k < \dots < \ell_1 < \sqrt{c}$. Then,*

$$n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1 + \sqrt{c})^2}{(1 + \sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}} \rightarrow \text{TW}_1$$

in law, where TW_1 is the (real) Tracy–Widom distribution, historically defined in Tracy and Widom [1996].³⁸

Specifically, the theorem is placed here under the setting, where the “population spikes” ℓ_i s are below the phase transition and thus asymptotically *not isolated* (from the main bulk), as per Theorem 2.13. In this setting, the largest empirical eigenvalue $\hat{\lambda}_1$ thus converges to the right edge $(1 + \sqrt{c})^2$ of the Marčenko–Pastur law and more precisely behaves, according to the theorem, as $\hat{\lambda}_1 = (1 + \sqrt{c})^2 + n^{-\frac{2}{3}}T$, where T is a (scaled) Tracy–Widom random variable.

This result is of practical interest as it allows one to estimate, for sufficiently large n, p , the probability for $\hat{\lambda}_1$ to be found away from its theoretical limit $(1 + \sqrt{c})^2$ below the phase transition.

Precisely, the result shows that the limiting fluctuations of $\hat{\lambda}_1$ are not Gaussian but follow the Tracy–Widom distribution and that, possibly surprisingly, the rate of this fluctuation is of order $O(n^{-2/3})$ (instead of $O(n^{-1/2})$ or $O(n^{-1})$ as one would usually expect). This rate is strongly related to the following observation, initially made by Silverstein and Choi [1995]: Close to the right-edge of its support, the Marčenko–Pastur law behaves proportionally to $\sqrt{(1 + \sqrt{c})^2 - x}$. As such, the typical number of eigenvalues in a space of size ϵ in the neighborhood of the edge is

$$\int_{(1 + \sqrt{c})^2 - \epsilon}^{(1 + \sqrt{c})^2} \sqrt{(1 + \sqrt{c})^2 - x} dx \propto \epsilon^{\frac{3}{2}}.$$

This explains the typical $O(n^{-2/3})$ fluctuation of the eigenvalues in this neighborhood. See Exercises 6 and 7 for more discussions on this point.

The original result from Baik et al. [2005] also provides the limiting fluctuations of $\hat{\lambda}_1, \dots, \hat{\lambda}_k$ beyond the phase transition (i.e., when $\ell_i > \sqrt{c}$). Interestingly, above the transition, the fluctuation of $\hat{\lambda}_1$ is now a classical central limit-type of order $O(n^{-1/2})$. The surprising “transition” from $O(n^{-2/3})$ to $O(n^{-1/2})$ of the fluctuations of $\hat{\lambda}_1$ (which has little meaning or interpretability for finite n, p) is often referred to as the *BBP phase transition* after the names of Baik et al. [2005]. Couillet and Hachem [2013] go beyond these considerations by providing the joint fluctuations of the eigenvalues and eigenvector projections as follows.

³⁸ The Tracy–Widom distribution does not have an explicit form. Several works have provided approximated forms as well as tables of TW_1 [Chiani, 2014, Ma, 2012].

Theorem 2.16 (Joint fluctuations beyond the phase transition, Couillet and Hachem [2013]). *Under the setting and notations of Theorems 2.13 and 2.14, assume $\ell_1 > \dots > \ell_k > \sqrt{c}$ and define $L = \sqrt{p}[\hat{\lambda}_1 - \lambda_1, \dots, \hat{\lambda}_k - \lambda_k]^T$ and $V = \sqrt{p}[|\mathbf{u}_1^T \hat{\mathbf{u}}_1|^2 - \zeta_1, \dots, |\mathbf{u}_k^T \hat{\mathbf{u}}_k|^2 - \zeta_k]^T$. Then, as $p, n \rightarrow \infty$ with $p/n \rightarrow c \in (0, \infty)$,*

$$\begin{pmatrix} L \\ V \end{pmatrix} \rightarrow \mathcal{N} \left(\mathbf{0}_{2k}, \text{BlockDiag} \left\{ \left[\begin{array}{cc} \frac{c^2(1+\ell_i)^2 \left(1 + \frac{(1+\ell_i)^2}{(c+\ell_i)^2} \right)}{(c+\ell_i)^2(\ell_i^2-c)} & \frac{(1+\ell_i)^3 c^2}{(\ell_i+c)^2 \ell_i} \\ \frac{(1+\ell_i)^3 c^2}{(\ell_i+c)^2 \ell_i} & \frac{c(1+\ell_i)^2(\ell_i^2-c)}{\ell_i^2} \end{array} \right] \right\}_{i=1}^k \right)$$

in law, where $\text{BlockDiag}(\cdot)$ is the “block-diagonal” operator.

The theorem notably states that, in the large n, p limit, while each eigenvalue–eigenvector projector pair fluctuates together, the k pairs fluctuate independently (which would no longer be the case if some ℓ_i s had multiplicity larger than one; the request $\ell_1 > \dots > \ell_k > \sqrt{c}$ in the theorem statement avoids this technical difficulty, which is treated in Bai and Yao [2008], Couillet and Hachem [2013], but gives rise to more complex results).

Remark 2.14 (Tracy–Widom law: beyond the real field and universality). *The Tracy–Widom law was first introduced in the context of Wigner random matrices in Theorem 2.5. More precisely, Tracy and Widom [1996] showed that the fluctuation of the largest eigenvalue of a real Gaussian Wigner random matrix (i.e., $\frac{1}{\sqrt{n}}\mathbf{X}$ with $\mathbf{X} \in \mathbb{R}^{n \times n}$ of i.i.d. zero-mean and unit-variance Gaussian entries up to symmetry) asymptotically follows a Tracy–Widom distribution in the sense that*

$$n^{\frac{2}{3}}(\lambda_1 - 2) \rightarrow \text{TW}_1.$$

The Tracy–Widom law also extends beyond the largest eigenvalue: It holds true for the finitely many largest as well as smallest eigenvalues of the Wigner and the Wishart matrix (in the latter case only if $c = \lim p/n < 1$). It also goes beyond real-valued symmetric Gaussian matrices (often referred to as the GOE) and the real-valued Wishart random matrices, to complex (Gaussian unitary ensemble, GUE) and quaternionic (Gaussian symplectic ensemble, GSE) Gaussian matrices: In these scenarios, the limiting laws are respectively the TW_2 and TW_4 Tracy–Widom distributions [Tracy and Widom, 2000]. See Figure 2.13 for an illustration.

The Tracy–Widom law has also been proven, to some extent, to be universal with respect to the distribution (of the entries) of random matrices. Soshnikov [1999] and Erdős [2011] proved that, for fast decaying distributions, it is sufficient to match the first two moments of the entries to obtain asymptotic Tracy–Widom fluctuations.

Finally, while the fluctuations of the (finitely many) largest or smallest eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ are not independent (they give rise, both for the k largest or for the k smallest, to joint fluctuations), Bianchi et al. [2010] showed that the fluctuations of the one largest and one smallest eigenvalues of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ are independent. This last result has the interesting consequence that the fluctuations of the condition number

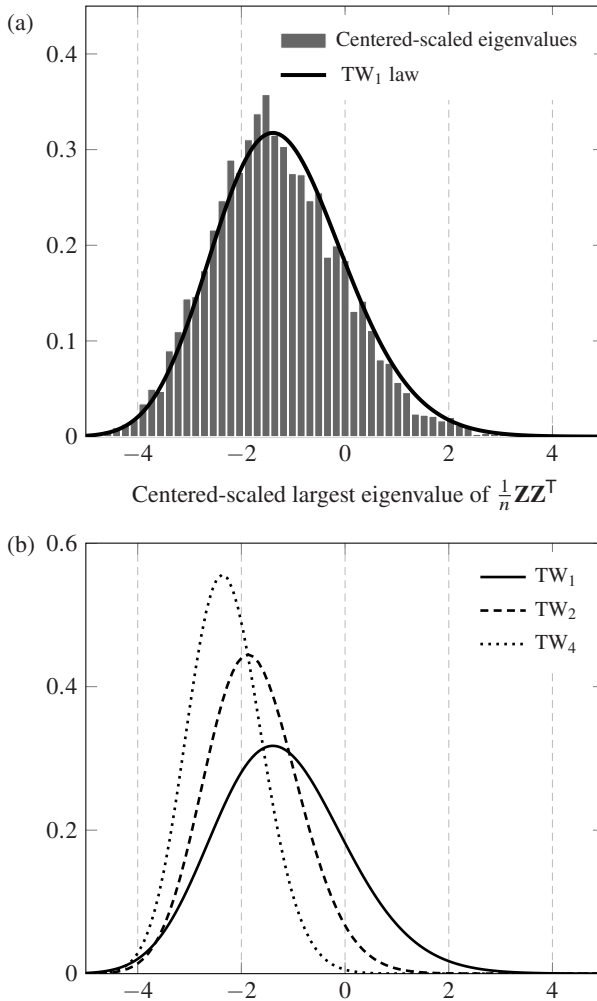


Figure 2.13 (a) Empirical histogram of $n^{\frac{2}{3}} \frac{\hat{\lambda}_1 - (1 + \sqrt{c})^2}{(1 + \sqrt{c})^{\frac{4}{3}} c^{-\frac{1}{6}}}$ for $p = 256$, $n = 512$ and standard Gaussian \mathbf{Z} , versus the real Tracy–Widom law TW_1 . Histogram obtained over 5000 independent runs. (b) Tracy–Widom distribution TW_β for $\beta = 1$ (real), 2 (complex), and 4 (symplectic). Code on web: MATLAB and Python.

of $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ (defined as the ratio between largest and smallest eigenvalues) around $(1 + \sqrt{c})^2 / (1 - \sqrt{c})^2$ are easily obtained, using, for instance, the so-called delta method [Vaart, 2000].

2.5.4 Further Discussions and Other Spiked Models

The “spiked model” terminology goes beyond sample covariance matrix models with $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$, for \mathbf{P} a low-rank matrix. In the literature, spiked models loosely refer to as “low rank perturbation” models in the following sense: There exists an underlying

random matrix model \mathbf{X} , the spectral measure of which converges to a well-defined measure with compact support (e.g., the Marčenko–Pastur or semicircle law) *and* having eigenvalues converging to the support (i.e., no single eigenvalue isolates as in Theorem 2.11), which is then modified in some way by a low-rank perturbation matrix \mathbf{P} ; the resulting matrix has the same limiting spectral measure as that of \mathbf{X} but with possibly a few spurious (isolated) eigenvalues.

Baik and Silverstein [2006] were the first to study spiked models, but their approach relied on the well-established results for sample covariance matrix models (i.e., Theorem 2.6) and was limited to the specific case of $\mathbf{C} = \mathbf{I}_p + \mathbf{P}$. This approach indeed requires a full understanding of a “more complex” statistical model before particularizing it to a low-rank perturbation. Pursuing on Footnote 33 that introduces Theorem 2.13, more modern tools launched a second wave of advances in spiked models, mostly triggered by the ideas found in Benaych-Georges and Nadakuditi [2012] (with a free probability approach), which is based on relating the spiked matrix model after perturbation to the underlying simple and nonperturbed matrix; this is mathematically simpler and opened the path to a broader scope of generalizations to more advanced random matrix models.

Among the popular spiked models, we have the following cases:

- the *information-plus-noise* model of the type

$$\frac{1}{n}(\mathbf{X} + \mathbf{P})(\mathbf{X} + \mathbf{P})^\top$$

with $\mathbf{X} \in \mathbb{R}^{p \times n}$ having i.i.d. standard entries (zero mean, unit variance, and finite fourth-order moment) and $\mathbf{P} \in \mathbb{R}^{p \times n}$ deterministic (or at least independent of \mathbf{X}) of fixed rank $k \ll \min(n, p)$;

- the *additive* model of the type

$$\mathbf{M} + \mathbf{P}$$

where $\mathbf{M} \in \mathbb{R}^{p \times p}$ is either of the type $\mathbf{M} = \frac{1}{n} \mathbf{X} \mathbf{X}^\top$, $\mathbf{X} \in \mathbb{R}^{p \times n}$ with standard i.i.d. entries, or of $\mathbf{M} = \mathbf{X} / \sqrt{n}$ with \mathbf{X} symmetric having standard i.i.d. entries above and on the diagonal and $\mathbf{P} \in \mathbb{R}^{n \times n}$ a deterministic matrix of low rank.

Each of these models has its own phase transition threshold (i.e., the value that eigenvalues of \mathbf{P} must exceed for a spike to be observed), dominant eigenvalue limits, and eigenvector projections. These can all be determined with the aforementioned proof approaches, see more examples in Exercises 11 and 12 of Section 2.9.

However, we will see in several applications in Chapter 4 that, in machine learning practice, we will be confronted with more general forms of low-rank perturbation models that do not fit this conventional “random matrix \mathbf{X} and deterministic perturbation \mathbf{P} ” assumption.

In particular, \mathbf{P} will often be a (possibly elaborate) function of \mathbf{X} . Also, the random matrix \mathbf{X} itself, which will often stand for the “noisy” part of the data model (while \mathbf{P} will in general comprises both the relevant information and possibly some extra noise), may induce its own isolated eigenvalues. For instance, we shall see later in Section 4.2.4 that, depending on the ratios p/n and $\text{tr} \mathbf{C}^4 / (\text{tr} \mathbf{C}^2)^2$, the random matrix

$\{[\mathbf{X}^T \mathbf{X}]_{ij}^2 \cdot \delta_{i \neq j}\}_{i,j=1}^n$, where $\mathbf{X} = \mathbf{C}^{\frac{1}{2}} \mathbf{Z}$ and \mathbf{Z} with i.i.d. standard entries, may have two isolated eigenvalues even when all the eigenvalues of \mathbf{C} remain in their limiting support. Also, in the context of robust estimation of covariance matrices to be discussed in Section 3.3, it will not be natural for the statistical model to impose that all its population eigenvalues converge to their limiting support (in particular, to mimic the action of a few outliers).

Yet, despite these technical differences, the proof approaches of Theorems 2.13 and 2.14 remain essentially valid. We thus propose here to generalize the notion of “spiked models” to models of the type $\mathbf{X} + \mathbf{P}$, where \mathbf{X} is some reference, well-understood, random matrix model (possibly inducing its own spikes) and \mathbf{P} is a low-rank matrix, possibly depending on \mathbf{X} .

With this definition, the aforementioned *sample covariance*, *information-plus-noise* and *additive* models are in fact all equivalent to an additive model. Precisely, we may write

$$\frac{1}{n}(\mathbf{X} + \mathbf{P})(\mathbf{X} + \mathbf{P})^T = \mathbf{M} + \mathbf{P}'$$

$$\text{with } \mathbf{M} = \frac{1}{n} \mathbf{X} \mathbf{X}^T, \quad \mathbf{P}' = \frac{1}{n} (\mathbf{X} \mathbf{P}^T + \mathbf{P} \mathbf{X}^T + \mathbf{P} \mathbf{P}^T)$$

and

$$\frac{1}{n}(\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} \mathbf{X} \mathbf{X}^T (\mathbf{I}_p + \mathbf{P})^{\frac{1}{2}} = \mathbf{M} + \mathbf{P}'$$

$$\text{with } \mathbf{M} = \frac{1}{n} \mathbf{X} \mathbf{X}^T, \quad \mathbf{P}' = \frac{1}{n} (\mathbf{X} \mathbf{P}'^T + \mathbf{P}'' \mathbf{X}^T + \mathbf{P}'' \mathbf{P}'^T)$$

where we introduced $\mathbf{P}'' = \mathbf{U}((\mathbf{I}_k + \mathbf{L})^{\frac{1}{2}} - \mathbf{I}_k) \mathbf{U}^T \mathbf{X}$ with $\mathbf{P} = \mathbf{U} \mathbf{L} \mathbf{U}^T$. In the remainder of the book, we shall systematically exploit this unified approach to treat all spiked models.

2.6 Information-plus-Noise, Deformed Wigner, and Other Models

2.6.1 Why Focus on the Sample Covariance Matrix Model?

The previous sections have mostly been concerned with the sample covariance matrix (as well as more marginally with Wigner matrices), as an instrumental statistical model for the introduction of the main technical tools of interest to the book: the Stieltjes transform and resolvent method, the spiked model approach, and statistical inference based on contour integrals, presented here in the form of their associated deterministic equivalents.

Several other classical random matrix models, of interest in statistics, will be listed in this section. The technical methods required to study these models are however not very different and thus not worth detailing in this book. Only pointers to relevant references will be provided here for the interested reader.

It is, in particular, important to stress that many statistical models arising in machine learning applications are so specific that they may not (strictly) fall in any of the conventional models discussed above. Yet, up to some additional fine-tuning and tricks, the analytical tools required to study these models are in general not much different from those presented in this chapter. Among examples met in the next chapters of this book, we may list:

- Graph Laplacian matrices (to be discussed in Chapter 7) of the form

$$\mathbf{D} - \mathbf{A}, \quad \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}, \quad \mathbf{D}^{-1}\mathbf{A}$$

for $\mathbf{A} \in \mathbb{R}^{n \times n}$ a symmetric matrix with independent entries (up to symmetry) and $\mathbf{D} = \text{diag}(\mathbf{A}\mathbf{1}_n)$. The dependence between \mathbf{A} and \mathbf{D} makes these random matrices slightly different from *deformed Wigner matrices* (see Section 2.6.2) of the type $\mathbf{A} + \mathbf{D}$, where \mathbf{A} has independent entries and \mathbf{D} is *deterministic*.

- Kernel random matrices of the inner-product or distance type

$$\mathbf{K} = \{f(\mathbf{x}_i^\top \mathbf{x}_j)\}_{i,j=1}^n, \quad \mathbf{K} = \{f(\|\mathbf{x}_i - \mathbf{x}_j\|^2)\}_{i,j=1}^n,$$

to be discussed in Chapter 4. There, the nontrivial dependence between the entries of \mathbf{K} differs significantly from sample covariance models (except of course for the linear kernel function $f(t) = t$ in the inner-product case).

- Robust estimators of scatter $\hat{\mathbf{C}}$ in Section 3.3 defined as the solutions to

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^n u \left(\frac{1}{p} \mathbf{x}_i^\top \hat{\mathbf{C}}^{-1} \mathbf{x}_i \right) \mathbf{x}_i \mathbf{x}_i^\top$$

for some nonincreasing function $u(t)$. There, due to the implicit nature of $\hat{\mathbf{C}}$, sample covariance matrix results cannot be applied directly.

- F-matrix models $\hat{\mathbf{C}}_1^{-1} \hat{\mathbf{C}}_2$ and product models $\hat{\mathbf{C}}_1 \hat{\mathbf{C}}_2$ for $\hat{\mathbf{C}}_a = \frac{1}{n} \mathbf{X}_a \mathbf{X}_a^\top$, $a \in \{1, 2\}$, with $\mathbf{X}_1, \mathbf{X}_2$ independent (notably Gaussian) random matrices, used in whitening methods [Yin et al., 1983], or in covariance matrix distance evaluation (e.g., Fisher distance, KL divergence, Wasserstein distance, etc., see Section 3.2). By successive conditioning, these models are more directly related to the sample covariance matrix models, although not strictly equivalent.
- Generalized sample covariance matrices of the type $\frac{1}{n} \mathbf{Z} \mathbf{D} \mathbf{Z}^\top$ for diagonal $\mathbf{D} \in \mathbb{R}^{n \times n}$ that *depends* on \mathbf{Z} (the independent case is handled in Theorem 2.6), but in an asymptotically “weak” manner, for instance, with $\mathbf{D} = \text{diag}\{f(\mathbf{w}^\top \mathbf{z}_i)\}_{i=1}^n$ for some deterministic $\mathbf{w} \in \mathbb{R}^p$, \mathbf{z}_i s columns of \mathbf{Z} , and $f: \mathbb{R} \rightarrow \mathbb{R}$. This family of random matrix models arises in many machine learning applications, for example, the Hessian matrix of the popular generalized linear model [Nelder and Wedderburn, 1972] can be shown to take this form [Liao and Mahoney, 2021], the spectral behavior of which is closely connected to the convergence rate of various optimization methods, see the concrete example of phase retrieval in Section 6.4.

The models and applications listed above appear to be strongly related, in one way or another, to sample covariance matrices. Among the examples above, kernel matrices, robust estimators, F-matrices and sample covariance products, as well as

Hessian-type matrices, all relate to sample covariance matrices. The graph Laplacian (as well, to some extent, as the kernel random matrix models) is more connected to Wigner matrices. This justifies the particularly focused vision of this chapter.³⁹

2.6.2 Other Models

Advanced Sample Covariance Matrices

From a historical standpoint, the model studied by Silverstein and Bai [1995] is slightly more general than that presented in Theorem 2.6. This model indeed assumes the presence of an additional deterministic matrix \mathbf{A} :

$$\mathbf{A} + \frac{1}{n} \mathbf{X}^\top \mathbf{C} \mathbf{X}$$

for random $\mathbf{X} \in \mathbb{R}^{p \times n}$ with independent entries and \mathbf{A}, \mathbf{C} deterministic matrices (in fact, \mathbf{C} was imposed to be diagonal in Silverstein and Bai [1995] but this assumption was later relaxed).

The bi-correlated (or separable covariance) model of the type $\frac{1}{n} \mathbf{C}^{\frac{1}{2}} \mathbf{X} \tilde{\mathbf{C}} \mathbf{X}^\top \mathbf{C}^{\frac{1}{2}}$ discussed in Theorem 2.7 was later studied in Paul and Silverstein [2009], where not only the limiting spectrum but also the condition for the exact separation of eigenvalues was derived. The extension of the spectral analysis of Silverstein and Choi [1995] for this model was then provided in Couillet and Hachem [2014]: A convenient *explicit* Stieltjes transform inverse $z(\tilde{m})$ no longer exists in this case (due to the presence of a coupled system of equations), but inverse mapping theorems guarantee its existence and lead to similar results.

For wireless communication purposes, the bi-correlated model was further extended in Couillet et al. [2011] to

$$\sum_{i=1}^k \frac{1}{n_i} \mathbf{R}_i^{\frac{1}{2}} \mathbf{X}_i \mathbf{T}_i \mathbf{X}_i^\top \mathbf{R}_i^{\frac{1}{2}}$$

where $\mathbf{T}_i \in \mathbb{R}^{n_i \times n_i}$ and $\mathbf{R}_i \in \mathbb{C}^{p \times p}$ are symmetric nonnegative definite matrices standing respectively for the transmit (T) and receive (R) correlation matrices at each end of a communication channel between k devices equipped with n_1, \dots, n_k antennas and a single receiver equipped with p antennas. Establishing the limiting spectral measure of this model allows one to estimate the maximally achievable communication rates between k simultaneously transmitting mobile terminals (phones, laptops, IoT devices) and a local base station. Further extensions of this model were then proposed to account for more involved wireless communication models, but they mostly consist in summing independent versions of Gram matrices $\mathbf{Z}_i \mathbf{Z}_i^\top$, where \mathbf{Z}_i is a Gaussian (or beyond Gaussian) random matrix with possibly nonzero mean, side correlations, a

³⁹ To the best of our knowledge, most, if not all, random matrix results directly related to machine learning applications boil down, in simple data model settings at least, to combinations (usually sums, sometimes products) of asymptotically independent matrices of the Wishart (Marčenko–Pastur related) and Wigner (semicircle related) types.

variance profile, etc.; see, for example Wen et al. [2013], Hachem et al. [2007], Wagner et al. [2012], Papazafeiropoulos and Ratnarajah [2015] out of a much longer list of articles on the topic.

Of interest to statistics is also the *information-plus-noise* model of the type

$$\frac{1}{n}(\mathbf{X} + \mathbf{A})(\mathbf{X} + \mathbf{A})^\top,$$

which is the sample “correlation” matrix between non-centered independent data $\mathbf{X} + \mathbf{A}$. This model also finds interest in wireless communications, where $[\mathbf{X} + \mathbf{A}]_{ij}$ models the statistical link between transmit antenna j and receive antenna i , which may not be at the same mean-distance (controlled by \mathbf{A}_{ij}) than another antenna pair. This model was first studied by Dozier and Silverstein [2007] who established the (unique) canonical equation ruling the limiting spectral measure of the model, as a function of the limiting Stieltjes transform of $\mu_{\mathbf{A}}$. Surprisingly enough, this model induces specific technical difficulties that left open for long the question of the exact location of the eigenvalues. Only much later in Loubaton and Vallet [2011] for the Gaussian case and then in Capitaine [2014] for the generic i.i.d. setting was the result fully obtained: that is, as for the sample covariance matrix in Theorem 2.11, under compactness assumption on the eigenvalues of \mathbf{A} , none of the eigenvalues of $\frac{1}{n}(\mathbf{X} + \mathbf{A})(\mathbf{X} + \mathbf{A})^\top$ asymptotically escapes the limiting support with high probability.

Yet, for practical applications, if the vectors of means $\mathbf{A}_1, \dots, \mathbf{A}_n$ in the model $\mathbf{X} + \mathbf{A}$ are equal (to say vector $\boldsymbol{\mu} \in \mathbb{R}^p$), then $\mathbf{A} = \boldsymbol{\mu} \mathbf{1}_n^\top$ reduces to a rank-one matrix, and $\frac{1}{n}(\mathbf{X} + \mathbf{A})(\mathbf{X} + \mathbf{A})^\top$ is merely a spiked model, which does not necessitate the technical intricacies in the aforementioned articles. If instead the entries \mathbf{A}_{ij} s are distinct with no specific (e.g., low-rank) structure, then it is in general not natural to assume that the \mathbf{X}_{ij} s have equal variance (as the variance should scale with the mean). To handle this setting, Hachem et al. [2007], Dumont et al. [2010] studied the generic *noncentered variance profile* model

$$\frac{1}{n}(\mathbf{B} \odot \mathbf{X} + \mathbf{A})(\mathbf{B} \odot \mathbf{X} + \mathbf{A})^\top$$

where \mathbf{B} is a symmetric matrix and \odot is the entry-wise Hadamard product. There is no natural limiting spectral measure for this model (even when the spectra of \mathbf{A}, \mathbf{B} are assumed to converge) but deterministic equivalents (e.g., of its resolvent matrix and of the associated Stieltjes transform) can be established, which generally rely on a set of pn fixed-point equations. To our knowledge, no result on the conditions for the exact spectrum separation has been obtained in this setting. In the “separable case” where $\mathbf{B} = \mathbf{b}_1 \mathbf{b}_2^\top$ for some vectors $\mathbf{b}_1, \mathbf{b}_2$ (in which case $\mathbf{B} \odot \mathbf{X} = \text{diag}(\mathbf{b}_1) \mathbf{X} \text{diag}(\mathbf{b}_2)$), the solution reduces to two fixed-point equations and the exact asymptotic location of the eigenvalues is almost a direct application of the “no-eigenvalue outside the support” theorem for the bi-correlated and the information-plus-noise models (i.e., the extension of Theorem 2.11 to these models).

Advanced Wigner Matrices

The generalizations of the Wigner random matrix model ($\frac{1}{\sqrt{n}}\mathbf{X}$ with \mathbf{X} having i.i.d. zero mean and unit variance entries) have been studied quite in parallel to the generalizations of the sample covariance matrix model $\frac{1}{n}\mathbf{X}\mathbf{X}^T$, as the technical tools and proofs are quite alike (if not simpler).

The first extended model of historical interest was that of the deformed (i.e., nonzero mean) Wigner model of the type

$$\frac{1}{\sqrt{n}}(\mathbf{X} + \mathbf{A})$$

for \mathbf{A} symmetric and deterministic [Khorunzhy and Pastur, 1994]. Yet again, of utmost interest in practice is the case where the independent entries of \mathbf{X} have differing variances, which brings forth the model

$$\frac{1}{\sqrt{n}}(\mathbf{B} \odot \mathbf{X} + \mathbf{A}).$$

The set of the n^2 canonical implicit (Stieltjes transform-related) equations induced for this model, or for its separable version ($\mathbf{B} = \mathbf{b}_1\mathbf{b}_2^T$), have been thoroughly investigated in Ajanki et al. [2019].

In practice, these models are directly applicable to the adjacency matrices of random graphs ($[\mathbf{X} + \mathbf{A}]_{ij}$ is the connectivity between node i and node j) with independent linking probabilities. The elementary case of such random graph models is the so-called Erdős–Rényi graph for which $\mathbf{X} + \mathbf{A}$ has i.i.d. Bernoulli $\{0, 1\}$ entries with parameter p . In this case, $\mathbf{A} = p\mathbf{1}_n\mathbf{1}_n^T$ is a rank-one matrix and \mathbf{X} has i.i.d. $\{-p, 1 - p\}$ entries such that $\mathbb{P}(\mathbf{X}_{ij} = 1 - p) = p, \mathbb{P}(\mathbf{X}_{ij} = -p) = 1 - p$ and therefore $\mathbb{E}[\mathbf{X}_{ij}] = 0$ with $\text{Var}[\mathbf{X}_{ij}] = p(1 - p)$. $\mathbf{X} + \mathbf{A}$ thus boils down to a spiked model. Assuming that the graph has *heterogeneous degrees*, in the sense that every particular node has its own probability q_i to connect to any other arbitrary node in the graph, we end up with the model $\text{diag}(\mathbf{q})\mathbf{X}\text{diag}(\mathbf{q}) + \mathbf{A}$ with $\mathbf{q} = [q_1, \dots, q_n]$, $\mathbf{X}_{ij} \in \{-q_iq_j, 1 - q_iq_j\}$ and $\mathbf{A}_{ij} = q_iq_j$. Here again $\mathbf{A} = \mathbf{q}\mathbf{q}^T$ is a rank-one matrix. See Chapter 7 for more detailed discussions on these random graph models.

(Real) Haar Random Matrices

Many algorithms and techniques in machine learning and data processing involve random projections, in general onto a lower dimensional subspace. This naturally calls for the study of random isometric matrices $\mathbf{U} \in \mathbb{R}^{p \times n}$, $n \leq p$, such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$ (because then $\mathbf{U}\mathbf{U}^T \in \mathbb{R}^{p \times p}$ is a projector on the n -dimensional subspace spanned by the columns of \mathbf{U}). These can be alternatively seen as concatenating the n columns of an underlying orthogonal matrix $\tilde{\mathbf{U}} \in \mathbb{R}^{p \times p}$.

Assuming $\tilde{\mathbf{U}}$ to be drawn uniformly in the space of unitary $p \times p$ matrices (this is called the *Haar* measure), $\mathbf{U} \in \mathbb{R}^{p \times n}$ is an *orthogonally invariant* random matrix, in the sense that $\mathbf{V}_1\mathbf{U}\mathbf{V}_2$ has the same law as \mathbf{U} for any pair of deterministic *orthogonal* matrices $\mathbf{V}_1 \in \mathbb{R}^{p \times p}, \mathbf{V}_2 \in \mathbb{R}^{n \times n}$. However, unlike Gaussian random matrices $\mathbf{Z} \in \mathbb{R}^{p \times n}$, which are also orthogonally invariant, the entries of \mathbf{U} are *not* independent

as they must satisfy $\mathbf{U}^T \mathbf{U} = \mathbf{I}_n$. This makes the study of the family of *Haar random matrices* more involved than the standard Gaussian (or i.i.d.) case.

Yet, strong analogies exist between the Gaussian and the Haar random matrices. To start with, note that \mathbf{U} can be constructed from standard Gaussian random matrices by letting $\mathbf{U} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-\frac{1}{2}}$ where $\mathbf{Z} \in \mathbb{R}^{p \times n}, n \leq p$, is a random matrix with i.i.d. standard Gaussian entries (it suffices to verify that $\mathbf{U}^T \mathbf{U} = \mathbf{I}_n$). Using this property, the fundamental trace lemma, Lemma 2.11, can be extended to a Haar-matrix equivalent [Debbah et al., 2003, Couillet et al., 2012].

Lemma 2.16 (Trace lemma for isometric matrices [Couillet et al., 2012, Lemma 5]). *Let $\mathbf{U} \in \mathbb{R}^{p \times n}$ be $n < p$ columns of a $p \times p$ Haar random matrix and $\mathbf{u} \in \mathbb{R}^p$ be a column of \mathbf{U} . Then, for $\mathbf{X} \in \mathbb{R}^{p \times p}$ a matrix function of the columns of \mathbf{U} , except \mathbf{u} , and of bounded operator norm,*

$$\mathbb{E} \left[\left| \mathbf{u}^T \mathbf{X} \mathbf{u} - \frac{1}{p-n} \text{tr} \Pi \mathbf{X} \right|^4 \right] \leq \frac{C}{p^2},$$

where $\Pi = \mathbf{I}_p - \mathbf{U} \mathbf{U}^T + \mathbf{u} \mathbf{u}^T$ (i.e., a projector on the complementary to the subspace spanned by the columns of \mathbf{U} , except \mathbf{u}) and C a constant depending only on the operator norm $\|\mathbf{X}\|$ and the ratio n/p .

Of course, since $\mathbf{U} \mathbf{U}^T$ is a projection matrix, all its eigenvalues are 1 and 0 and there is thus no interest in studying the spectrum of $\mathbf{U} \mathbf{U}^T$ itself. The above trace lemma however becomes handy when dealing with more structured models, such as $\mathbf{C}^{\frac{1}{2}} \mathbf{U} \mathbf{U}^T \mathbf{C}^{\frac{1}{2}}$ for some deterministic \mathbf{C} matrix; the latter may be seen as a generalization of the sample covariance matrix model of Theorem 2.6. Specifically, we have the following result, which provides a deterministic equivalent for this model.

Theorem 2.17 (Haar sample covariance [Couillet et al., 2012, Theorem 1]). *Let $\mathbf{X} = \mathbf{C}^{\frac{1}{2}} \mathbf{U} \in \mathbb{R}^{p \times n}$, where $\mathbf{U} \in \mathbb{R}^{p \times n}$ are the $n < p$ columns of a $p \times p$ Haar random matrix, and let $\mathbf{C} \in \mathbb{R}^{p \times p}$ be symmetric nonnegative definite with bounded operator norm. Then, for $z < 0$, as $p/n \rightarrow c \in (1, \infty)$, letting $\mathbf{Q}(z) = (\frac{p}{n} \mathbf{X} \mathbf{X}^T - z \mathbf{I}_p)^{-1}$, we have*

$$\mathbf{Q}(z) \leftrightarrow \bar{\mathbf{Q}}(z) = -\frac{1}{z} (\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C})^{-1},$$

where $\tilde{m}_p(z)$ is the unique positive solution to

$$\tilde{m}_p(z) = \left(-z + (1 + zc^{-1} \tilde{m}_p(z)) \cdot \frac{1}{n} \text{tr} \mathbf{C} (\mathbf{I}_p + \tilde{m}_p(z) \mathbf{C})^{-1} \right)^{-1}. \tag{2.54}$$

In the statement of the theorem, we used a ‘‘correction’’ factor $\frac{p}{n}$ in front of $\mathbf{X} \mathbf{X}^T$ to ensure the correspondence between $\mathbb{E}[\mathbf{U} \mathbf{U}^T] = \frac{n}{p} \mathbf{I}_p$ and the setting of Theorem 2.6, where $\mathbb{E}[\frac{1}{n} \mathbf{Z} \mathbf{Z}^T] = \mathbf{I}_p$. Indeed, it is quite interesting to observe the close relation between Theorems 2.6 and 2.17 which, despite the major difference imposed by the strongly *dependent* structure of \mathbf{U} versus the *independent* structure of \mathbf{Z} , leads almost to the same deterministic equivalent. The only difference lies in the extra term $zc^{-1} \tilde{m}_p(z)$ in the defining equation (2.54) for $\tilde{m}_p(z)$.

Similar to the case of random matrices with i.i.d. entries versus Gaussian entries, it is also, in the case of Haar matrix models, sometimes more convenient to work with Gaussian-specific identities rather than the “independence”-related trace lemma above. Specifically, an equivalent for Stein’s lemma, Lemma 2.13, also exists for Haar matrices.

Lemma 2.17 (Stein’s lemma for Haar matrices [Pastur and Shcherbina, 2011, Chapter 8]). *Let $\tilde{\mathbf{U}} \in \mathbb{R}^{p \times p}$ be a Haar matrix and $f: \mathbb{R}^{p \times p} \rightarrow \mathbb{R}$ a function admitting an analytic extension in the neighborhood of the set of unitary matrices in $\mathbb{R}^{p \times p}$. Then we have, for all $j, j' \in \{1, \dots, p\}$,*

$$\mathbb{E} \left[\sum_{i=1}^p f'_{ij}(\tilde{\mathbf{U}}) \tilde{\mathbf{U}}_{ij'} - f'_{ij'}(\tilde{\mathbf{U}}) \tilde{\mathbf{U}}_{ij} \right] = 0,$$

where f'_{ij} is the classical derivative with respect to $\tilde{\mathbf{U}}_{ij}$ (not accounting for the dependence of the other entries in $\tilde{\mathbf{U}}$). In the complex case ($\tilde{\mathbf{U}} \in \mathbb{C}^{p \times p}$ and $f(\tilde{\mathbf{U}}) \in \mathbb{C}$), this reduces to⁴⁰

$$\mathbb{E} \left[\sum_{i=1}^p f'_{ij}(\tilde{\mathbf{U}}) \tilde{\mathbf{U}}_{ij'} \right] = 0.$$

Similarly a Nash–Poincaré inequality, Lemma 2.14, for Haar matrix models is defined.

Lemma 2.18 (Nash–Poincaré for Haar matrices). *Under the setting of Lemma 2.17, we have*

$$\text{Var}(f(\tilde{\mathbf{U}})) \leq \frac{1}{p} \sum_{i,j=1}^p \mathbb{E} [|f'_{ij}(\tilde{\mathbf{U}})|^2].$$

Although seemingly less exploitable, the above Stein’s lemma for Haar matrices is in fact quite convenient and easily leads to results such as the aforementioned Theorem 2.17 (for instance, by considering matrix functions of the form $f(\tilde{\mathbf{U}}\mathbf{D})$ for $\mathbf{D} \in \mathbb{R}^{p \times p}$ diagonal with $\mathbf{D}_{ii} = \delta_{i \leq n}$ – so that $f(\tilde{\mathbf{U}}\mathbf{D})$ only selects $n < p$ columns of $\tilde{\mathbf{U}} \in \mathbb{R}^{p \times p}$). Exercise 13 proposes to retrieve the result of Theorem 2.17 using both Gaussian (so applying Lemmas 2.17 and 2.18) and i.i.d. (so applying Lemma 2.16) approaches.

As a major difference between the i.i.d. (as in Theorem 2.6) and the Haar settings, note that Theorem 2.17 is stated under the constraint that z be real negative. In effect, Couillet et al. [2012] showed that it is far from trivial to extend the result to $z \in \mathbb{C}$ away from the negative real axis: in particular, unlike in the classical sample covariance setting of Theorem 2.6, in the “Haar sample covariance” of Theorem 2.17, the fixed-point iteration in (2.54) fails to converge for $z = x + i\epsilon$ with $x > 0$ and $\epsilon \ll 1$. This

⁴⁰ Similar to what we saw in Remark 2.5, it is in practice more convenient to work under a complex (unitary) \mathbf{U} setting, even in the real (orthogonal) case, as deterministic equivalents are universal with respect to the underlying field (real or complex) of the entries of \mathbf{U} .

particularly makes it difficult to exploit the result to retrieve (both theoretically⁴¹ and numerically) the limiting spectral measure of $\mathbf{C}^{\frac{1}{2}}\mathbf{U}\mathbf{U}^{\top}\mathbf{C}^{\frac{1}{2}}$, at least in its present form of (2.54).

There in fact exists a whole other branch of tools in the random matrix literature, called *free probability theory* [Voiculescu et al., 1992], which much more easily recovers Theorem 2.17 (under a different formulation though) and as well obtains the limiting spectral measure of $\mathbf{C}^{\frac{1}{2}}\mathbf{U}\mathbf{U}^{\top}\mathbf{C}^{\frac{1}{2}}$: Specifically, using an interesting extension to random matrices of classical probability theory on scalar random variables, free probability theory demonstrates that this limiting measure is the so-called *free multiplicative convolution* (see Section 2.6.2 for a proper definition) of the limiting spectral measure ν of \mathbf{C} and of the limiting spectral measure of $\mathbf{U}\mathbf{U}^{\top}$ (i.e., the discrete measure $\delta_1 + (c - 1)\delta_0$). The next section provides a short introduction to free probability theory.

Turning to machine learning applications of results on (derivatives of) Haar random matrix models, to the best of our knowledge, very few works have so far fully exploited the strength of these identities. For this reason, we will not elaborate much more on these aspects and will only, in the following section, briefly introduce free probability theory, which has many advantages (especially when dealing with Haar or permutation-invariant random matrices) but also strong limitations when compared to the Stieltjes transform and resolvent approach. We thus point the interested reader to the (in fact rich) literature for more details on this topic. Those tools may nonetheless reveal fundamental insight in the future into specific random projection or random permutation-based methods with isometric constraints in machine learning and AI.

The Free Probability Approach

Free probability theory is a drastically different approach to study random matrices. It is particularly efficient in some scenarios, such as when the sum or product of random matrices are involved. The theory was developed in parallel to the Stieltjes transform method discussed in this book and originates from the works of Voiculescu et al. [1992], who originally aimed to describe a theory of probabilities on noncommutative algebras. A detailed introduction of the theory is beyond the scope of this book and we refer the interested readers to Hiai and Petz [2006], Biane [1998] and Couillet and Debbah [2011, Chapters 4 and 5]. Although free probability theory is rooted in a combinatorial approach (see, e.g., Nica and Speicher [2006]), it also contains some elegant analytic results, which can be related to the Stieltjes transform: In the sequel, we emphasize those useful results.

For μ and ν two probability measures compactly supported on $[0, \infty)$, Hiai and Petz [2006] proved that there always exist two *free random variables* a and b in some noncommutative probability space having distributions μ and ν , respectively. The distribution of $a + b$ and ab depend solely on μ, ν and can be associated with probability

⁴¹ One may claim that, since convergence holds for all $z < 0$, as per Vitali's convergence theorem (Theorem 2.3), it can then be extended to all of $\mathbb{C} \setminus \mathbb{R}^+$. This is however not so simple as it is difficult to ensure that $\tilde{m}_p(z)$ in Theorem 2.17, as defined through its fixed-point equation, is indeed analytic in a certain cone $\{z = e^{i\theta} \mid \theta \in (-\theta^\circ, \theta^\circ) \setminus \{0\}\}$ ($\theta^\circ \in (0, \pi)$).

measures called *free additive convolution* and *free multiplicative convolution* of the distributions μ and ν , denoted $\mu \boxplus \nu$ and $\mu \boxtimes \nu$, respectively. These measures are both compactly supported on $[0, \infty)$ [Voiculescu et al., 1992].

These free additive and multiplicative convolutions satisfy convenient analytic expressions, through the so-called *R*- and *S*-transforms introduced below.

Definition 5 (*R*- and *S*-transform). *Let μ be a probability measure with support $\text{supp}(\mu)$ and Stieltjes transform $m_\mu(z)$, for $z \in \mathbb{C}^+$. The R-transform of μ , denoted R_μ , is defined as the solution to*

$$m_\mu(R_\mu(z) + z^{-1}) = -z$$

or equivalently

$$m_\mu(z) = \frac{1}{R_\mu(-m_\mu(z)) - z}.$$

Next, let $\psi_\mu(z)$ be defined as

$$\psi_\mu(z) = \int \frac{zt}{1-zt} \mu(dt) = -1 - z^{-1}m_\mu(z^{-1})$$

and let χ_μ be its unique functional inverse, analytic in the neighborhood of zero, that is, $\chi_\mu(\psi_\mu(z)) = z$ for $|z|$ small enough. Then, the *S*-transform of μ , denoted S_μ , is given by

$$S_\mu(z) = \chi_\mu(z) \frac{1+z}{z}.$$

In particular, $S_\mu(z)$ satisfies

$$m_\mu\left(\frac{z+1}{zS_\mu(z)}\right) = -zS_\mu(z).$$

The main property of *R*- and *S*-transforms is summarized below, and requires the notion of *freeness* between noncommutative random variables. Freeness is not an easy notion, and is defined through a series of moment conditions and combinatorial calculus, which we will not go into detail on here (see again Hiai and Petz [2006], Biane [1998]). One needs to just remember at this point that freeness extends the notion of independence to noncommutative random variables.

Lemma 2.19 (*R*- and *S*-transforms of sums and products). *For a and b two free random variables with compactly supported distributions μ and ν , respectively, the law $\mu \boxplus \nu$ of $a + b$ satisfies*

$$R_{\mu \boxplus \nu}(z) = R_\mu(z) + R_\nu(z).$$

Similarly, the law $\mu \boxtimes \nu$ of ab satisfies

$$S_{\mu \boxtimes \nu}(z) = S_\mu(z)S_\nu(z).$$

Of interest to the present book is that “asymptotically large random matrices” are typical examples of noncommutative random variables for which freeness can be ensured. To avoid dealing with infinite-size linear operators, it is more appropriate to

define a notion of *asymptotic freeness* for finite-dimensional random matrices, which translates the freeness of their respective limiting operators.⁴²

As such, the main result of interest to us is the following: for $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times n}$ two asymptotically free random matrices with respective *limiting spectral measures* μ_A and μ_B , the limiting spectral measure μ_{A+B} ($A+B$ is here merely a notation with no formal meaning) of $\mathbf{A} + \mathbf{B}$ exists and satisfies

$$\mu_{A+B} = \mu_A \boxplus \mu_B, \quad R_{A+B}(z) = R_A(z) + R_B(z),$$

for $R_A(z), R_B(z)$, and $R_{A+B}(z)$ the R -transforms of μ_A, μ_B , and μ_{A+B} , respectively. Similarly, μ_{AB} , the limiting spectral measure of the matrix product \mathbf{AB} , exists and satisfies

$$\mu_{AB} = \mu_A \boxtimes \mu_B, \quad S_{AB}(z) = S_A(z)S_B(z),$$

for $S_A(z), S_B(z)$, and $S_{AB}(z)$ the S -transforms of μ_A, μ_B , and μ_{AB} . The above equalities should be understood to hold in the almost sure sense.

Clearly, the asymptotically freeness assumption plays a key role in relating the limiting spectrum of $\mathbf{A} + \mathbf{B}$ or \mathbf{AB} to that of \mathbf{A} and \mathbf{B} , which unfortunately in practice only applies easily to a limited range of random matrices. In essence, \mathbf{A} and \mathbf{B} are asymptotically free if they are both independent *and* if the distribution of their respective eigenvectors are sufficiently “isotropic” with respect to one another: So essentially, when one of the two matrices is invariant by left and right multiplying by arbitrary unitary matrices. As a consequence, the two major cases of matrix pairs known to be asymptotically free are: (i) a standard Gaussian random matrix and any other independent random matrix (for instance, a deterministic matrix or another standard Gaussian random matrix, independent of the first), and (ii) a Haar random matrix and any other independent random matrix. One may, for instance, easily determine the limiting spectral measure of models of the type $\mathbf{X} + \mathbf{A}$ for \mathbf{X} a Wigner matrix or a Wishart matrix and \mathbf{A} deterministic, or of \mathbf{XAX}^T with \mathbf{X} Gaussian or Haar distributed. These objects are however limited and it is technically difficult to establish asymptotic freeness, the formal definition of which is a matter of heavy combinatorial calculus (see, e.g., Biane [1998, Section 3]). As a result, free probability theory can be more complex to use when summing or multiplying two random matrices with structured eigenvectors, such as simple models like $\mathbf{X} \odot \mathbf{B} + \mathbf{A}$ for \mathbf{X} a Wigner matrix and \mathbf{B} a deterministic variance profile: These matrices are *not* free with respect to deterministic matrices, so that the R - and S -transform formulas cannot be exploited, at least directly; this

⁴² One must be careful that, in the whole book, we never define large-dimensional random matrices as being of “infinite” dimensions (which would turn them, when correctly defined, into operators in an infinite-dimensional Hilbert space): All the objects treated throughout the book are finite-dimensional objects, some functionals of which are studied when the size of the matrices increases. For actual works on operators, seen as limit of random matrices in infinite-dimensional spaces, see Pastur and Figotin [1992] on almost-periodic random operators. Aside from a few exceptions though [Hachem et al., 2015], these elegant works find little practical applications in systems and software engineering. This being said, it is theoretically interesting to observe that the Stieltjes transform approach thoroughly developed in the present book shares many common grounds with the more general theory of linear operators in Hilbert spaces [Akhiezer and Glazman, 2013].

very fact has strongly limited the (rigorous) reach of the free probability approach in the past decade.

A fundamental result to efficiently use the addition and product rules in Lemma 2.19 are the basic forms of the R - and S -transforms of elementary random matrix models. Specifically, the R - and S -transforms of the Marčenko–Pastur and semicircle distributions are known in closed forms.

Lemma 2.20 (*R - and S -transforms of Marčenko–Pastur and semicircle law*). *The R -transform $R_{MP,c}(z)$ and S -transform $S_{MP,c}(z)$ of the Marčenko–Pastur law $\mu_{MP,c}$ of parameter c , that is, of the limiting spectral measure of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$, $\mathbf{Z} \in \mathbb{R}^{p \times n}$ with i.i.d. zero-mean, unit-variance entries, as $p/n \rightarrow c \in (0, \infty)$, given explicitly by (2.10), read*

$$R_{MP,c}(z) = \frac{1}{1 - cz}, \quad S_{MP,c}(z) = \frac{1}{1 + cz}. \tag{2.55}$$

As for the R -transform $R_{SC}(z)$ and S -transform $S_{SC}(z)$ of the semicircle law μ_{SC} , given by (2.30), we have

$$R_{SC}(z) = z, \quad S_{SC}(z) = \frac{1}{\sqrt{z}}. \tag{2.56}$$

With Lemma 2.20, one is able to derive, with a free probability approach, the limiting spectral measure of the information-plus-noise-type random matrix model $\mathbf{M} = \mathbf{A} + \frac{1}{n}\mathbf{X}\mathbf{X}^T$ for $\mathbf{X} \in \mathbb{R}^{p \times n}$ having i.i.d. standard Gaussian entries and $\mathbf{A} \in \mathbb{R}^{p \times p}$ a deterministic matrix. Specifically, calling μ_A and μ_M the limiting spectral measure of \mathbf{A} and \mathbf{M} as $n, p \rightarrow \infty$ with $p/n \rightarrow c$, we have

$$\mu_M = \mu_A \boxplus \mu_{MP,c}, \quad R_M(z) = R_A(z) + R_{MP,c}(z)$$

so that, by Definition 5 and Lemma 2.20,

$$m_M(z) = \frac{1}{R_M(-m_M(z)) - z} = \frac{1}{R_A(-m_M(z)) + \frac{1}{1 + cm_M(z)} - z}$$

or equivalently

$$R_A(-m_M(z)) + \frac{1}{-m_M(z)} = z - \frac{1}{1 + cm_M(z)}$$

which, by taking the Stieltjes transform $m_A(\cdot)$ of the limiting law of \mathbf{A} on both sides, together with Definition 5, gives

$$m_M(z) = m_A\left(z - \frac{1}{1 + cm_M(z)}\right).$$

The same result would have been more painstaking to derive using a purely Stieltjes transform approach (see, e.g., Silverstein and Bai [1995]). However, since very few matrix models can be easily shown to be asymptotically free, the free probability framework quickly fails to operate for more structured random matrix models.

Recent works try to cope with these limitations as well as open the range of applicability of free probability theory to handle sums and products of matrices under weaker forms of asymptotic freeness conditions (to characterize, for instance, the limiting

spectrum of the sum of random matrices with row and columns permutation invariance [Au et al., 2018], or to extend the notion of deterministic equivalents to a free probability setting [Speicher and Vargas, 2012]). Despite these efforts, when dealing with random matrix models with involved structures arising from machine learning applications, the resolvent and Stieltjes transform approaches turn out more flexible; they are thus the focus of this book.

Full Circle Law, β -Ensembles, Sparse Random Matrices, etc.

Mathematicians have long been intrigued by the “simplest” random matrix model in appearance, that is, $\mathbf{X}/\sqrt{n} \in \mathbb{R}^{n \times n}$ (nonsymmetric) with i.i.d. zero-mean and unit-variance entries. Being a nonsymmetric matrix (at least with high probability), the eigenvalues of \mathbf{X}/\sqrt{n} are complex and they have long been known to spread uniformly on the unit complex disc $\{z \in \mathbb{C}, |z| < 1\}$. Surprisingly though, despite its simple statement, this result, known as the *full circle law* or the *circular law*, has only been proven in full generality very recently by Tao and Vu [2008]. To explain the difficulties: (i) the Stieltjes transform method cannot be applied directly as the spectrum is complex (and thus taking a limit $z \rightarrow z_0$ for z_0 in the support does not allow to “enter” the complex support as in the real eigenvalue case); there the solution was provided earlier by Girko [1985] who introduced the alternative V-transform; (ii) the V-transform involves the limit of an integral form on the logarithm of the singular values of \mathbf{X} which, being square, tends to have a lot of singular values tending to zero (the singular values of \mathbf{X} are the square roots of the eigenvalues of $\mathbf{X}\mathbf{X}^T$ with \mathbf{X} of size $p \times n$, where $p = n$: that is, this is the technically most difficult *hard-edge* scenario of the Marčenko–Pastur law depicted in Figure 2.2 with $c = 1$); this technical difficulty, previously worked around by invoking the existence of high-order moments for \mathbf{X}_{ij} was solved by Tao and Vu by means of the *ϵ -net technique*, popular today in compressive sensing and high-dimensional statistics [Vershynin, 2018].

From the perspective of the present book, the eigenvalues of nonsymmetric models are of marginal interest. These could be used for the analysis of directed random graphs although, to our knowledge, not much work exists in this direction.

Another more mathematical interest relates to the fact that Gaussian random matrices are much better known than random matrices with i.i.d. entries and, consequently, come along with a host of other technical tools. In particular, not only the limiting spectral measure, but actually the exact *finite-dimensional joint distribution* $\mathbb{P}(\lambda_1, \dots, \lambda_n)$ of (real, complex, or quaternionic) Gaussian symmetric random matrix \mathbf{X} and $\frac{1}{n}\mathbf{X}\mathbf{X}^T$ (with \mathbf{X} having i.i.d. standard Gaussian entries) is known. The expressions of $\mathbb{P}(\lambda_1, \dots, \lambda_n)$ for these different cases are quite related.

In particular, the joint eigenvalue distribution for the Gaussian Wigner matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$ is explicitly given by

$$\mathbb{P}(\lambda_1, \dots, \lambda_n) \propto \prod_{i=1}^n e^{-\frac{1}{4}\beta n \lambda_i} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta, \quad (2.57)$$

for real Gaussian \mathbf{X} when $\beta = 1$ (recall from Remark 2.14 that this is the GOE), complex Gaussian \mathbf{X} when $\beta = 2$ (GUE), and quaternionic Gaussian \mathbf{X} when $\beta = 4$

(GSE). Much work has been devoted to the study of the asymptotics of the joint law of this now called β -ensemble of random matrices. In particular, the Tracy–Widom law for the largest eigenvalue introduced in Theorem 2.15 is obtained by marginalizing the joint measure to obtain the probability $\mathbb{P}(\lambda_1 > x)$. See Anderson et al. [2010] for an introduction to these quite different methods.

Most of the aforementioned random matrix models however share as a common denominator their relying on $O(n^2)$ “degrees of freedom,” in the sense that they are designed out of order $O(n^2)$ independent random variables. For the sample covariance matrix model $\frac{1}{n}\mathbf{X}\mathbf{X}^T$, we have $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z} \in \mathbb{R}^{p \times n}$ with \mathbf{Z} made of np independent entries. For the Wigner model, $\mathbf{X} \in \mathbb{R}^{n \times n}$ has $n(n+1)/2$ independent entries on and above the diagonal. This large number of degrees of freedom is the *major asset of random matrix theory*, as presented in this book: they trigger (i) concentration properties that do not appear if p is fixed and only $n \rightarrow \infty$ (such as quadratic form concentration $\frac{1}{p}\|\mathbf{z}_i\|^2 \xrightarrow{\text{a.s.}} 1$ for $\mathbf{z}_i \in \mathbb{R}^p$ having i.i.d. entries with zero mean and unit variance as in Lemma 2.11), (ii) fast convergence rates with typical central limit theorem of order up to $O(1/n)$ and, possibly most importantly, (iii) *universality* with respect to the underlying distribution of the independent entries (i.e., asymptotic statistics loosely depend on the actual law of the entries), which simplifies the analysis and provides robustness of the studied objects to deviations from the statistical model.

Yet, a host of practical random matrix models demand less degrees of freedom. Realistic networks, for instance, (social nets, brain connectivity, molecular networks, etc.) are naturally modeled by *sparse* random (say symmetric) adjacency matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ with typical number of nonzero elements scaling as $O(n)$ rather than $O(n^2)$. Every row/column \mathbf{a}_i of \mathbf{A} typically has $O(1)$ nonzero elements (corresponding to the neighbors or contacts of node i in the underlying graph), and thus $\|\mathbf{a}_i\|$ does *not* concentrate as $n \rightarrow \infty$. Kernel random matrices $\mathbf{K} = \{f(\|\mathbf{x}_i - \mathbf{x}_j\|^2)\}_{i,j=1}^n$ of finite-dimensional vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ with p small (e.g., in the context of classification or clustering of 2D or 3D data points) are also more challenging to study than their large- p counterpart, as every entry of \mathbf{K} remains a random variable, which does not concentrate in the large- n alone limit. The consequences are numerous: (i) the analysis of these objects is more difficult, if doable at all, (ii) universality and robustness to model assumptions are lost: large- n asymptotics remain a function of the law of, not only the “statistical structure,” but also the *precise* distribution of the entries of \mathbf{a}_i and \mathbf{x}_i . These hard-to-obtain results thus hardly lead to simple and rich insight offered by the proposed random matrix analysis.

Nevertheless, a branch of random matrix theory focuses on these important models. Stieltjes transform methods are here mostly ineffective and one has to rely on moment approaches and combinatorics. A particularly interesting approach when it comes to sparse random graphs of size n is that, as $n \rightarrow \infty$, the graph has a “tree-like” structure; indeed, with a probability $O(1/n)$ for each node to reach out to any another node, the probability of the presence of cycles in the graph is vanishingly small. This has motivated the independent development of a graph-based random matrix framework, strongly pushed by Bordenave and Lelarge [2010], Bordenave et al. [2011]. The results are however generally “weak” from a practical standpoint. For instance, while

it has long been known that the spectral measure of a *dense* Erdős–Rényi random graph \mathbf{A} with Bernoulli i.i.d. entries (i.e., with $O(n^2)$ degrees of freedom) converges to the semicircle law, it is still unknown to which measure a *sparse* random graph \mathbf{A} converges: the limiting law is known to exist, to be decomposed as the sum of a (known) discrete measure and a (unknown) continuous measure, and to have an unbounded support (as opposed to the semicircular distribution as in Theorem 2.5) [Salez, 2011].

The specific kernel random matrix $\mathbf{K} = \{\|\mathbf{x}_i - \mathbf{x}_j\|^2\}_{i,j=1}^n$ with \mathbf{x}_i of fixed dimension, known as a Euclidean random matrix, has also been studied in Bordenave [2008], but again with results of limited practical reach.

Aside from side comments, the book will not dig into these fundamentally different problems, tools, and results. We exclusively concentrate on dense random matrix models.

2.6.3 Other Statistics

Most of the statistics of practical interest in the application chapters are directly related to deterministic equivalents of the resolvent of random matrices and to their linear statistics. For instance, we shall see that the performance of classification methods (measured by classification accuracy) of n data vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ in a k -class $(\mathcal{C}_1, \dots, \mathcal{C}_k)$ problem can in general be estimated from the k -dimensional matrix of quadratic forms

$$\frac{1}{n} \mathbf{J}^T \mathbf{Q}(z) \mathbf{J}$$

(or some closely related statistics), where $\mathbf{Q}(z)$ is the resolvent of the underlying affinity matrix of the data (kernel, graph Laplacian, etc.) and $\mathbf{J} = [\mathbf{j}_1, \dots, \mathbf{j}_k] \in \mathbb{R}^{n \times k}$ with $[\mathbf{j}_a]_i = \delta_{\mathbf{x}_i \in \mathcal{C}_a}$ the canonical vector of class \mathcal{C}_a .

Yet, some specific results (such as the classification rate of some random neural networks, the exact proof of the asymptotic Gaussian behavior of the entries of the dominant eigenvector in graph adjacency and kernel matrices, etc.) demand more than just *first-order* limiting statistics. A further common statistics of interest lies in the *second-order fluctuations*, that is, in central limit theorems, of the objects under study.

These statistics have long been studied in the random matrix literature, starting from the works of Bai and Silverstein [2004] who, under the sample covariance setting of Theorem 2.6, established a central limit of the type

$$n \int f(t) (\mu_{\frac{1}{n} \mathbf{X} \mathbf{X}^T} - \mu)(dt) \rightarrow \mathcal{N}(M(f), \sigma^2(f))$$

for all analytic functions f . This result (and all similar results for related models) has the following noteworthy properties:

- the convergence rate is of order $O(n^{-1})$. This however only holds for linear statistics of the eigenvalues; bilinear forms $\mathbf{a}^T (\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)) \mathbf{b}$ fluctuate at a slower $O(n^{-\frac{1}{2}})$ rate;

- the mean (or bias term) $M(f)$ and variance $\sigma^2(f)$ depend on $\mathbb{E}[|Z_{ij}|^4]$ (which, thus, must be assumed finite). Both write as the sum $A + \kappa B$ with κ the kurtosis of the entries Z_{ij} . The mean $M(f)$, in particular, vanishes in the complex Gaussian case and the variance in the complex Gaussian case is twice as large as that in the real Gaussian case.

Many results on central limit theorems for a vast spectrum of linear statistics of random models have been established, for instance, in Hachem et al. [2008] for sample covariance matrices $\frac{1}{n} \mathbf{X} \mathbf{X}^T$ with \mathbf{X} having a variance profile, in Lytova and Pastur [2009] to Wigner matrix model for less smooth functions f (five times differentiable), or in Zheng et al. [2017] for F-matrix models of the type $(\frac{1}{n_1} \mathbf{X}_1 \mathbf{X}_1^T)^{-1} \frac{1}{n_2} \mathbf{X}_2 \mathbf{X}_2^T$. A generalization to three times differentiable f is proposed in Najim and Yao [2016]. Central limit theorems for bilinear forms are found for instance in Kammoun et al. [2009]. Fluctuations of the isolated eigenvalues and eigenvector projections in a spiked random matrix model can also be found in Baik et al. [2005], Bai and Yao [2008], Couillet and Hachem [2013]. These fluctuations are at a slower $O(n^{-1/2})$ rate.

A central limit result for the linear statistical inference method of Theorem 2.12 has also been established in Yao et al. [2013]. There again it is shown that the convergence speed is of order $O(n^{-1})$ with a bias and a variance of the form $A + \kappa B$ with κ the kurtosis of the underlying distribution (and, again, the bias vanishes in the complex Gaussian case). An estimation method is also proposed for the means and variances, which is of practical interest to empirically assess the confidence interval of the estimator.

Due to a strong motivation from the field of wireless communications, some specific linear statistics have been particularly widely studied in the random matrix literature. This is notably the case of the logarithm function. Statistics of the type

$$\int \log(1 + st) \mu_{\frac{1}{n} \mathbf{X} \mathbf{X}^T}(dt)$$

for $s > 0$ are particularly important in wireless communications as they give access to the achievable communication rate over a linear wireless communication channel \mathbf{X} . This $\log(1 + st)$ term arises from the entropy of Gaussian random variables and is also found in many other applications, such as with the estimation of the Kullback–Leibler divergence between two multivariate Gaussian vectors to be discussed in Section 3.2. A particularly convenient feature of the integral form $\int \log(1 + st) \mu(dt)$ is that its derivative with respect to s (i.e., $\int t/(1 + st) \mu(dt)$) is immediately related to the Stieltjes transform of μ .⁴³ It is thus not required to use a complex contour integral method to assess these quantities (a real integration is sufficient). See Tulino and Verdú [2004], Couillet and Debbah [2011] for a detailed account of these findings.

In technical terms, there are essentially two major methods to obtain central limit theorems of random matrix quantities. Recalling that linear functionals $u(\mathbf{Q})$ of the resolvent $\mathbf{Q} = (\mathbf{X} - z \mathbf{I}_n)^{-1}$ of the random matrix \mathbf{X} under study (e.g., bilinear forms $\mathbf{a}^T \mathbf{Q} \mathbf{b}$ or traces $\text{tr} \mathbf{A} \mathbf{Q}$), as our central object of interest, cannot in general be expressed

⁴³ Specifically, $\int t/(1 + st) \mu(dt) = s^{-1}(1 - s^{-1} m_\mu(-s^{-1}))$.

as a sum of independent random variables. Instead, Bai and Silverstein [2010] propose to use the martingale difference approach, which we previously exploited in the detailed proof of the Marčenko–Pastur theorem, Theorem 2.4. More precisely, for \mathbf{X} having independent columns, it is convenient to write

$$u(\mathbf{Q}) - \mathbb{E}[u(\mathbf{Q})] = \sum_{i=1}^n \mathbb{E}_{i-1}[u(\mathbf{Q})] - \mathbb{E}_i[u(\mathbf{Q})],$$

where \mathbb{E}_i is the expectation conditioned on the columns $\mathbf{x}_1, \dots, \mathbf{x}_i$ of \mathbf{X} , with the convention $\mathbb{E}_0[u(\mathbf{Q})] = u(\mathbf{Q})$. This is a sum of martingale differences, for which Billingsley [2012, Theorem 35.12] provides a central limit theorem (see also Bai and Silverstein [2010, Chapter 9]).

Alternatively, Pastur proposes to use Gaussian techniques, which we also explored in the alternative proof of Theorem 2.4, along with a *characteristic function approach* (see examples in Pastur and Shcherbina [2011]) to show that

$$\mathbb{E} \left[e^{-tu(\mathbf{Q})} \right] \rightarrow e^{-tM - \frac{1}{2}t^2\sigma^2},$$

which is the Gaussian characteristic function. To reach this convergence, the approach consists in exploiting Stein's lemma, Lemma 2.13, on the differentiated (along t) left-hand expectation, that is,

$$\mathbb{E} \left[-tu(\mathbf{Q})e^{-tu(\mathbf{Q})} \right].$$

Exploiting the fact that u is linear and that $\mathbf{Q} = -\frac{1}{z}\mathbf{I}_n + \frac{1}{z}\mathbf{Q}\mathbf{X}$, this expectation can be reduced as a function of the type $\mathbb{E}[\mathbf{X}f(\mathbf{X})]$ on which Lemma 2.13 can be applied. The objective is then to show that this differentiated characteristic function converges to the derivative of the limiting Gaussian characteristic function, that is, $(-tM - t\sigma^2)e^{-tM - \frac{1}{2}t^2\sigma^2}$. This can be achieved, for instance, by controlling the difference using the Nash–Poincaré inequality, Lemma 2.14.

2.7 Beyond Vectors of Independent Entries: Concentration of Measure in RMT

2.7.1 Limitations of the i.i.d. Assumption

In the previous sections, we have shown that the Stieltjes transform and resolvent approaches are quite versatile tools which, in a way, form a surrounding “complex analysis and linear algebra core” for random matrix theory analysis. This core, however, must be *independently supplemented* by appropriate probabilistic tools (which ensure the necessary convergences for linear algebra and complex analysis methods to be applied).

When it comes to these probabilistic methods, we have seen that a major driver for most of the results lies in exploiting the *independence both in samples (n) and features (p)* of the underlying random matrix \mathbf{X} . It is thus no wonder that a natural and long-standing assumption in the early works in random matrix theory was to

request for \mathbf{X} to have all independent (or “linearly dependent” as in models of the type $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}\tilde{\mathbf{C}}^{\frac{1}{2}} + \mathbf{A}$) entries. Most generalizations of these results usually assume mere deviations from this setting (by allowing weak, or asymptotically vanishing, correlation between the entries, for instance).

However, while for random graphs it is largely conceivable to request independent “noise” associated with each link, and for random vector observations it is natural to ask for these observations to be independent, requesting that *every single observation made of independent entries* is very constraining. Note, in particular, that what we referred to as the sample covariance matrix model in Theorem 2.6 is in fact a very restricted model, where each observation \mathbf{x}_i needs to be of the form $\mathbf{x}_i = \mathbf{C}^{\frac{1}{2}}\mathbf{z}_i$ for some random vector \mathbf{z}_i having independent entries. This model is mostly convenient only in the Gaussian case where $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ and as a result $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. Most multivariate random vectors \mathbf{x}_i with zero mean and covariance \mathbf{C} (elliptical distributions, correlated vectors of Bernoulli entries, etc.) cannot be factorized under this form.

Most importantly, the “real data” \mathbf{x}_i (images, sounds, videos, DNA sequences, population features, etc.) met in machine learning applications tend to live in (possibly very contorted) manifolds that cannot be linearly “whitened” into a vector of independent entries by merely operating $\mathbf{C}^{-\frac{1}{2}}\mathbf{x}_i$.

2.7.2 Concentrated Random Vectors as the Answer

El Karoui [2009] and Pajor and Pastur [2009] were the first to realize (or at least to fully exploit the fact) that, from a probability standpoint, the proof of the sample covariance matrix result in Theorem 2.6 from Silverstein and Bai [1995] only relies on (i) the independence between the (*column*) vectors \mathbf{x}_i composing $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$ (and thus *not necessarily* of all the entries), and (ii) the convergence

$$\frac{1}{n}\mathbf{x}_i^{\top}\mathbf{Q}_{-i}(z)\mathbf{x}_i - \frac{1}{n}\text{tr}\mathbf{Q}_{-i}\mathbf{C} \rightarrow 0 \tag{2.58}$$

in some probabilistic sense, where $\mathbf{Q}_{-i}(z) = (\frac{1}{n}\mathbf{X}\mathbf{X}^{\top} - \frac{1}{n}\mathbf{x}_i\mathbf{x}_i^{\top} - z\mathbf{I}_p)^{-1}$. For the latter, it is sufficient but *not necessary* for $\mathbf{z}_i = \mathbf{C}^{-\frac{1}{2}}\mathbf{x}_i$ to have standard i.i.d. entries. In particular, El Karoui showed that this convergence also holds if \mathbf{x}_i is a *concentrated random vector*: A fundamental property at the core of our present concern and which, we will show, has far-reaching consequences to the application in real-world machine learning and AI.

In a nutshell, the concentration of measure theory, extensively developed by Ledoux [2005], considers random vectors $\mathbf{x} \in \mathbb{R}^p$ having the property that every 1-Lipschitz functional $\phi: \mathbb{R}^p \rightarrow \mathbb{R}$ of \mathbf{x} is “predictable,” in the sense that there exists a deterministic value $M_{\phi} \in \mathbb{R}$ such that the random variable $\phi(\mathbf{x})$ remains in the neighborhood of M_{ϕ} , and that the diameter of this neighborhood vanishes as $p \rightarrow \infty$. This notion must *not* be confused with the fact that the random vector \mathbf{x} itself converges, which is in general largely wrong: only the *scalar observations* $\phi(\mathbf{x})$ of \mathbf{x} converge, and we will

say in this case that \mathbf{x} “concentrates.”⁴⁴ More formally, assuming $M_\phi = O(1)$ with respect to p (otherwise, it needs to be appropriately scaled), there exists a function $\alpha(t, p)$ decreasing to zero in both t and p such that

$$\mathbb{P}(|\phi(\mathbf{x}) - M_\phi| > t) \leq \alpha(t, p). \quad (2.59)$$

Of particular interest is the case $\alpha(t, p) = e^{-t^\beta p^\gamma}$ for some $\beta, \gamma > 0$ which, since the exponential grows faster than any polynomial, provides a more powerful and much more flexible inequality than the moment bounds introduced in the proof of the Marčenko–Pastur law.⁴⁵ The mapping $\mathbf{x}_i \rightarrow \frac{1}{n} \mathbf{x}_i^\top \mathbf{Q}_{-i}(z) \mathbf{x}_i$ in (2.58) is however not Lipschitz, and thus more profound technical considerations are requested to show that Theorem 2.6 indeed extends to the case where the \mathbf{x}_i s are independent concentrated random vectors. This is performed in an intricate manner in El Karoui [2009]. A more systematic approach has been recently developed in Louart and Couillet [2018], the basics of which will be discussed in the next section.

Paradoxically, very few “classical” multivariate distributions are known to produce concentrated random vectors, and yet, this is enough to bring an outstanding practical competitive advantage against vectors with independent entries, when it comes to modeling real data in machine learning practice.

Among popular distributions, only the Gaussian random vector $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, the uniformly distributed vector on the unit sphere $\mathbf{x} \sim \mathbb{S}^{p-1}$, and the vector \mathbf{x} with i.i.d. entries with bounded support (i.e., $|x_i| < K$ for some $K > 0$) are known to be concentrated random vectors. Worse, for the latter, the definition (2.59) only holds for all 1-Lipschitz *and convex* maps, which is practically inconvenient (since, as opposed to Lipschitz maps, Lipschitz-convex functions are not stable through composition).

Let us thus stick for the moment to the example of $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. The major advantage of being a concentrated random vector is that this concentration property is stable under any 1-Lipschitz map $f: \mathbb{R}^p \rightarrow \mathbb{R}^q$. So, if \mathbf{x} is concentrated, so is $\mathbf{x}' = f(\mathbf{x})$, which, as opposed to $\mathbf{C}^{\frac{1}{2}} \mathbf{x}$, can be a vector with *intricate nonlinear dependence* between its entries (as we shall see right after, this intricate dependence may be such that photo-realistic images, able to deceive the human eyes, can be generated from Lipschitz maps of standard Gaussian random vectors).

Now, the key reasons why the class of random vectors $\{f(\mathbf{x})\}$ spanned by 1-Lipschitz maps f is so fundamental to machine learning are that

- (i) there exist machine learning techniques that *learn* to produce artificial but highly realistic data, exclusively based on Lipschitz maps. The most popular of these methods are the generative adversarial networks proposed by Goodfellow et al.

⁴⁴ As a matter of fact, as we will see, the concept of concentration is even more general in that it allows one to control the fluctuations of $\phi(\mathbf{x})$, for arbitrary $\phi: \mathbb{R}^p \rightarrow \mathbb{R}^q$ for generic $q \geq 1$, even when $\phi(\mathbf{x})$ does not converge. Lipschitz operators being stable through composition, iterated controls of Lipschitz functions with various Lipschitz constants enable a thin tracking of the behavior of sometimes intricate nonlinear functionals of \mathbf{x} (such as through the layers of a neural network).

⁴⁵ Of course, as a compensation for this simplification, this imposes more technical constraints on the entries of the random vector \mathbf{x} , such as the existence of moments of all orders. But, as far as practical statistical machine learning considerations are concerned, this is far from a heavy request.

[2014]. Those are feedforward neural networks which, after training, generate highly realistic data $f(\mathbf{x})$ from a standard Gaussian input $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ (so realistic that even human beings cannot tell synthetic data from real ones, see again samples in Figure 1.8). Since a feedforward neural network is a sequence of linear operators (inter-layer connections and convolution operators) and Lipschitz nonlinear activation functions (sigmoid, rectified linear, etc.), $f(\cdot)$ is indeed Lipschitz (as the composition of Lipschitz operators remains Lipschitz); (ii) feature extraction procedures in machine learning are also mostly Lipschitz maps. The most popular of these today are convolutional neural networks, which are again feedforward neural nets and thus, by definition, Lipschitz maps of the input data. But this is also mostly true for many “classical” machine learning methods, such as support vector machines, semi-supervised graph learning, spectral clustering, etc.

As a consequence of (i) and (ii), since real data can be trust-worthily approximated by outputs \mathbf{x}' of some Lipschitz function $\mathbf{x}' = f(\mathbf{x})$ of random Gaussian vectors \mathbf{x} , the class of concentrated random vectors encompasses a broad “set of (almost) realistic data.” Furthermore, in practice, the features exploited by most machine learning algorithms can be seen as yet another Lipschitz mapping $g(\mathbf{x}')$ of the data \mathbf{x}' . Since \mathbf{x}' takes the form of $\mathbf{x}' = f(\mathbf{x})$ for standard Gaussian $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{x}'' = (g \circ f)(\mathbf{x})$ is again a Lipschitz map of a standard Gaussian vector and thus a concentrated random vector.

It then becomes natural to model a wide range of realistic data, and their corresponding features extracted by, say, modern neural networks, as concentrated random vectors, for example, as Lipschitz functions of standard Gaussian vectors.

Remark 2.15 (Concentration inequalities versus concentration of measure theory). *The concentration of measure theory developed by Ledoux [2005] provides as corollaries a list of popular concentration inequalities such as Gaussian concentration inequalities, Bernstein’s and Talagrand’s inequalities for random variables with bounded entries,⁴⁶ McDiarmid’s inequalities for functionals of bounded deviations of independent random variables, etc. These results, quite popular in statistics, can however only marginally be used as a full-fledged concentration of measure-oriented random matrix framework. As an instance, quadratic forms of the type $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ are not naturally handled by these concentration inequalities (for which the Hanson–Weight inequality provides an answer, see Rudelson and Vershynin [2013] and Exercise 15). More importantly, while quadratic form concentration is essentially sufficient to prove the convergence of Stieltjes transforms, proving the resolvent convergence $\mathbb{E}[\mathbf{Q}] - \bar{\mathbf{Q}} \rightarrow 0$ under a concentration inequality setting actually demands to further expand the works of Ledoux, as will be shown next.*

It must also be stressed that Tao [2012], Vershynin [2012] provided an introduction to what Vershynin refers to as nonasymptotic random matrix theory based on concentration inequalities. The approach followed by the authors however significantly

⁴⁶ To be more exact, Talagrand’s work was developed in parallel to Ledoux’s theory and is rather complementary than a consequence of one another.

differs from the present $n, p \rightarrow \infty$ with $p/n \rightarrow c \in (0, \infty)$ large-dimensional random matrix considerations. In nonasymptotic random matrix theory, the variables n, p are left “free” (to grow at any relative speed to infinity) and the use of concentration inequalities aims at retrieving bounds on, for instance, the largest or smallest eigenvalue or singular value of the underlying random matrices, without resorting to the Stieltjes transform approach. For Tao, this control step is the crux of the proof of the circular law (based on the ϵ -net theory developed by the author) for nonsymmetric matrices \mathbf{X} with i.i.d. entries. For Vershynin, these nonasymptotic spectrum controls are exploited in applications to compressive sensing, where random matrix theory also plays a key role – for instance, in providing “typical” matrices fulfilling the popular restricted isometry property [Candès, 2008].

The approach proposed in this book also provides a set of inequalities, where n, p have an untied growth to infinity, but the application of these convergence results is mostly of interest in a joint growth rate for n, p . Besides, additional tools to Ledoux’s original framework, such as the notion of linear concentration, will be needed.

Remark 2.16 (Limitations of the concentration of measure framework). *It is important to raise here (somewhat ironically) that the concentration of measure framework, which finds important corollaries to the field of compressive sensing [Donoho, 2006, Baraniuk, 2007], is, as presented here, at odds with the compressive sensing framework. Indeed, compressive sensing is a major field of research in large-dimensional statistics and machine learning, which assumes that large-dimensional data are intrinsically of low dimension. That is, in the simplest linear setting, data vectors $\mathbf{x} \in \mathbb{R}^p$ can be written as $\mathbf{x} = \mathbf{A}\mathbf{y}$ for some matrix $\mathbf{A} \in \mathbb{R}^{p \times q}$ (generally unknown) and $\mathbf{y} \in \mathbb{R}^q$ for $q \ll p$. From there, the idea of compressive sensing is that meaningful statistical inference on \mathbf{y} can be performed based on few independent realizations $n \ll p$ (which is convenient if p is extremely large). There, concentration inequalities are mostly used to deal with the (usually random) observation matrix \mathbf{A} , rather than with the underlying (low-dimensional) \mathbf{y} .*

In the present random matrix framework, concentration of measure is used to model the data, not the data operating matrices. These data however must not be of intrinsic low dimension $q \ll n$. Or, at least, if they were, we would impose in our framework that $n \sim q$ and $n, q, p \rightarrow \infty$ with a small but $O(1)$ ratio q/p . If instead $q = O(1) \ll n$, then we would fall back under the (technically more difficult) sparse regime discussed at the end of Section 2.6.2, where the present framework is mostly ineffective.

As shall be seen in concrete applications presented in this book, high-resolution images are very appropriately modeled by concentrated random vectors of intrinsically large dimensions. However, feature vectors such as bag-of-words (also known as tf*idf features) for text classification [Manning et al., 2008], which are very large but extremely sparse vectors, cannot be handled by the random matrix framework presented here.

This however does not mean that compressive sensing is complementary to random matrix theory. Compressive sensing indeed tackles the “difficult” problem analyzing sparse recovery algorithms by somehow “loose” inequalities and bounds: that is, it

cannot accurately predict the exact performance of a given algorithm (however, it can ensure its convergence and its efficiency as $n \rightarrow \infty$ at a certain rate with respect to p , while q is in general fixed). Random matrix theory instead requests that the intrinsic dimension $q \rightarrow \infty$, even slowly so, but manages in exchange (by exploiting the q degrees of freedom in the feature space) to provide accurate performance estimates of machine learning algorithms for all finite (but at least moderately large) n, q .

2.7.3 Elements of Concentration of Measure for Random Matrices

We recall here basic elements of the concentration of measure theory of immediate interest to random matrix applications. More advanced considerations can be found in Ledoux [2005] from a mathematical standpoint, and in Louart and Couillet [2018] with a more random matrix-oriented flavor.

Concentration of Random Variables

Before getting into generic multivariate concentration of measure theory, we need to start with the concept of concentration of a (uni-variate) random *variable*. Concentration of measure can be defined in two parallel ways.

Definition 6 (Concentration of a random variable). *Let $\alpha: \mathbb{R}^+ \rightarrow [0, 1]$ be a nonincreasing function with $\alpha(\infty) = 0$. A random variable x is α -concentrated and we write $x \propto \alpha$ if, for an independent copy x' of x , and all $t > 0$,*

$$\mathbb{P}(|x - x'| > t) \leq \alpha(t).$$

The definition suggests that any two *independent* realizations of x cannot live far from one another. Alternatively, we may define x as concentrated if there exists a deterministic *pivot* a close to which x remains.

Definition 7 (Concentration around a pivot). *Let $\alpha: \mathbb{R}^+ \rightarrow [0, 1]$ be a nonincreasing function and $a \in \mathbb{R}$. Then, x is α -concentrated around the pivot a , denoted $x \in a \pm \alpha$, if for all $t > 0$,*

$$\mathbb{P}(|x - a| > t) \leq \alpha(t).$$

These two definitions are not formally equivalent. However, we have the implication

$$x \propto \alpha \Rightarrow x \in M_x \pm 2\alpha \Rightarrow x \propto 4\alpha(\cdot/2),$$

where $M_x \in \mathbb{R}$ is a *median* of x , that is such that $\mathbb{P}(x \geq M_x) \geq 1/2$ and $\mathbb{P}(x \leq M_x) \geq 1/2$. The loss of a factor $1/2$ arises here from the bound $\mathbb{P}(|x - x'| > t) \leq \mathbb{P}(|x - a| > t/2) + \mathbb{P}(|x' - a| > t/2)$. As a result, up to constants, it is then possible to use either definition interchangeably (the proofs of subsequent results are usually more accessible to one or the other definition).

A particularly appealing result is that 1-Lipschitz maps $f: \mathbb{R} \rightarrow \mathbb{R}$ of a concentrated random variable x maintain the concentration, that is,

$$x \in a \pm \alpha \Rightarrow f(x) \in f(a) \pm \alpha. \tag{2.60}$$

This is a particularly fundamental result which suggests that every “smooth” function of sub-linear growth of x satisfies the same concentration property. This result naturally arises from the fact that $|f(x) - f(x')| \leq |x - x'|$ and thus $\mathbb{P}(|f(x) - f(x')| > t) \leq \mathbb{P}(|x - x'| > t)$.

Evidently, sums of concentrated random variables are also concentrated:

$$\begin{aligned} x_1 \in a \pm \alpha, x_2 \in b \pm \beta &\Rightarrow (x_1 + x_2) \in (a + b) \pm [\alpha(\cdot/2) + \beta(\cdot/2)] \\ x_1 \in a \pm \alpha, x_2 \in b \pm \beta &\Rightarrow (x_1 + x_2) \in (a + b) \pm [\alpha(\cdot/2) + \beta(\cdot/2)], \end{aligned}$$

where in the first line the factor 1/2 again unfolds from the bound $\mathbb{P}(|x_1 + x_2 - x'_1 - x'_2| > t) \leq \mathbb{P}(|x_1 - x'_1| > t/2) + \mathbb{P}(|x_2 - x'_2| > t/2)$, and similarly for the second line.

However, products, particularly of *dependent* random variables, are less obvious to tackle, as one needs to avoid conditioning. The problem can be worked around using the following two relations

$$\begin{aligned} x_1 x_2 - ab &= (x_1 - a)(x_2 - b) + a(x_2 - b) + b(x_1 - a) \\ |x_1 - a||x_2 - b| > t &\Rightarrow (|x_1 - a| > \sqrt{t}) \text{ or } (|x_2 - b| > \sqrt{t}) \end{aligned}$$

so to obtain

$$\begin{aligned} x_1 \in a \pm \alpha, x_2 \in b \pm \beta &\Rightarrow x_1 x_2 \in ab \pm \\ \begin{cases} \alpha(\sqrt{\cdot/3}) + \alpha(\cdot/3|b|) + \beta(\sqrt{\cdot/3}) + \beta(\cdot/3|a|), & a, b \neq 0 \\ \alpha(\sqrt{\cdot/2}) + \alpha(\cdot/2|b|) + \beta(\sqrt{\cdot/2}), & a = 0, b \neq 0 \\ \alpha(\sqrt{\cdot}) + \beta(\sqrt{\cdot}), & a = b = 0. \end{cases} \end{aligned}$$

For large t , the probability $\mathbb{P}(|x_1 x_2| > t)$ is here dominated by the terms $\alpha(\sqrt{\cdot})$ and $\beta(\sqrt{\cdot})$, which is not surprising. In the particular case where $x_1 = x_2 = x$, or more generally for powers x^k of concentrated random variables x , we have

$$x \in a \pm \alpha \Rightarrow x^k \in a^k \pm \left[\alpha(\cdot/2^k |a|^{k-1}) + \alpha((\cdot/2)^{\frac{1}{k}}) \right] \tag{2.61}$$

with $\alpha(\cdot/0) = \alpha(\infty)$ by convention, which is based on noticing that

$$|x^k - a^k| \leq (2|a|)^k \left(\frac{|x - a|}{|a|} + \frac{|x - a|^k}{|a|^k} \right).$$

This result will be particularly useful for random matrix applications to quadratic forms.

Remark 2.17 (Exponential concentration). *Of utmost interest is the case where $\alpha(t) = Ce^{-(t/\sigma)^q}$ for some $C, \sigma, q > 0$. In particular, it is known that standard random Gaussian variables x satisfy*

$$x \sim \mathcal{N}(0, 1) \Rightarrow x \in 0 \pm 2e^{-(\cdot)^2/2}.$$

Exponential concentrations are fast and induce a lot of convenient properties. In particular, using the formula $\mathbb{E}[|x|^k] = \int_0^\infty \mathbb{P}(|x|^k > t) dt$, it appears that all (absolute) moments of exponentially concentrated random variables exist. In particular,

$$x \in C e^{-(\cdot/\sigma)^q} \Rightarrow x \in \mathbb{E}[x] \pm e^{\frac{Cq}{q}} e^{-(\cdot/2\sigma)^q} \tag{2.62}$$

so that an exponentially concentrated random variable concentrates around its mean. But most importantly, we have the implications

$$\begin{aligned} x \in a \pm C e^{-(\cdot/\sigma)^q} &\Rightarrow \forall r \geq q, \mathbb{E}[|x - a|^r] \leq C\Gamma(r/q + 1)\sigma^r \\ &\Rightarrow x \in a \pm C e^{-(\cdot/\sigma)^q/e} \end{aligned}$$

with Γ the gamma-distribution. Thus, exponential concentration is “equivalent” to controlled growth by σ^r of all moments $r \geq q$. This is particularly appealing when moments occasionally turn out more convenient to deal with than bounds on tail probabilities.

Concentration of Random Vectors

The concept of concentration of random variable x , stating that x does not deviate much from a given pivot a , cannot be straightforwardly extended to that of random vectors. Indeed, random vectors (in particular, large dimensional ones) rather tend to “avoid” their statistical means or medians: for example, Gaussian random vectors $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ are of zero mean but they “concentrate” on a $O(1)$ -thick layer around the sphere in \mathbb{R}^p of diameter \sqrt{p} (see, e.g., Figure 1.6 for an illustration).

Instead, for a normed vector space $(E, \|\cdot\|)$, we will consider that a random vector $\mathbf{x} \in E$ is concentrated for some class of functions $\mathcal{F}: \mathbb{R}^p \rightarrow \mathbb{R}$ if, for all $f \in \mathcal{F}$, $f(\mathbf{x})$ is a concentrated random variable. Depending on the “broadness” of the class, being a concentrated random vector can be more demanding. Ledoux [2005] originally defined two such classes \mathcal{F} : the class of 1-Lipschitz maps (appropriate for Gaussian or random unitary vectors) and the class of convex (or weakly convex) 1-Lipschitz maps (adapted to vectors of independent bounded entries). There, the Lipschitz property (i.e., the fact that $|f(\mathbf{x}) - f(\mathbf{y})| \leq \|\mathbf{x} - \mathbf{y}\|$) is with respect to the norm $\|\cdot\|$ in E , and thus the concentration rates may depend on $\|\cdot\|$. In order to better encompass random matrices in the concentration of measure framework, a looser additional class \mathcal{F} will be introduced here: that of *unit-norm linear functionals*.

Linear Concentration

Linear concentration is an important concept in random matrix theory as it provides a quite general and flexible definition for the key notion of *deterministic equivalents* (recall Definition 4) of great significance in this book.

Definition 8 (Linear concentration). *A random vector $\mathbf{x} \in E$ is linearly α -concentrated around the deterministic equivalent $\bar{\mathbf{x}}$, with respect to the norm $\|\cdot\|$ in E , if, for all unit norm linear functional $u: E \rightarrow \mathbb{R}$ (i.e., $|u(\mathbf{x})| \leq \|\mathbf{x}\|$),*

$$u(\mathbf{x}) \in u(\bar{\mathbf{x}}) \pm \alpha.$$

The expectation being a linear operator (from E to E), an advantage of linear concentration is that, upon existence, $\mathbb{E}[\mathbf{x}]$ is a deterministic equivalent for the concentrated random vector \mathbf{x} . In particular, if \mathbf{Q} is a random matrix (e.g., the resolvent of some other underlying random matrix) in the “vector space” $(\mathbb{R}^{P \times P}, \|\cdot\|)$, with $\|\cdot\|$ the operator norm, and that \mathbf{Q} is linearly concentrated with respect to $\|\cdot\|$, then, as already mentioned in Remark 2.2, $\mathbb{E}\mathbf{Q}$ is a deterministic equivalent for \mathbf{Q} and we have, in particular, for all $\mathbf{A} \in \mathbb{R}^{P \times P}$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^P$ of bounded (operator and Euclidean) norms,

$$\frac{1}{p} \text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q}) \rightarrow 0, \quad \mathbf{a}^\top (\mathbf{Q} - \mathbb{E}\mathbf{Q}) \mathbf{b} \rightarrow 0,$$

where the convergence is in probability and, if $\alpha(t) = Ce^{-t^q}$ for some $q > 0$, the convergence is also almost sure.⁴⁷ This result implies that the newly defined notion of *deterministic equivalents* from a linear concentration standpoint automatically induces the former Definition 4.

Lipschitz Concentration

Lipschitz concentration is the most popular type of concentrations (due to its compatibility with (2.60)). This notion is even in general merely called “concentration” (rather than Lipschitz concentration) and is defined as follows.

Definition 9 (Lipschitz concentration). *A random vector $\mathbf{x} \in E$ is Lipschitz α -concentrated with respect to the norm $\|\cdot\|$ if, for every 1-Lipschitz function $f: E \rightarrow \mathbb{R}$, we have either of the conditions*

$$\begin{aligned} f(\mathbf{x}) &\propto \alpha, \text{ denoted } \mathbf{x} \propto \alpha \\ f(\mathbf{x}) &\in M_f \pm \alpha, \text{ denoted } \mathbf{x} \stackrel{M}{\propto} \alpha \\ f(\mathbf{x}) &\in \mathbb{E}[f(\mathbf{x})] \pm \alpha, \text{ denoted } \mathbf{x} \stackrel{\mathbb{E}}{\propto} \alpha \end{aligned}$$

holds, where M_f is a median of $f(\mathbf{x})$.

Similar to the concentration of random variables, the three notions are not fully equivalent. For generic α -concentration, we have

$$\mathbf{x} \propto \alpha \Rightarrow \mathbf{x} \stackrel{M}{\propto} 2\alpha \Rightarrow \mathbf{x} \propto 4\alpha(\cdot/2)$$

and, in the case of exponential concentrations, the expectation is well defined and we further have

$$\mathbf{x} \stackrel{M}{\propto} Ce^{-(\cdot/\sigma)^q} \Rightarrow \mathbf{x} \stackrel{\mathbb{E}}{\propto} e^{C^q/q} e^{-(\cdot/2\sigma)^q} \Rightarrow \mathbf{x} \stackrel{M}{\propto} 2e^{C^q/q} e^{-(\cdot/4\sigma)^q}.$$

The most fundamental result at the very heart of the concentration of measure theory is that Gaussian random vectors $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ are Lipschitz concentrated in $(\mathbb{R}^p, \|\cdot\|)$ for $\|\cdot\|$ the Euclidean norm, that is

⁴⁷ Here we exploit the fact that, for $u(\mathbf{Z}) = \frac{1}{p} \text{tr} \mathbf{AZ}$, $|u(\mathbf{Z})| \leq \|\mathbf{Z}\|$ when $\|\mathbf{A}\| \leq 1$ and that, for $u(\mathbf{Z}) = \mathbf{a}^\top \mathbf{Z} \mathbf{b} = \text{tr}(\mathbf{b}\mathbf{a}^\top \mathbf{Z})$, $|u(\mathbf{Z})| \leq \|\mathbf{Z}\|$ for $\|\mathbf{a}\|, \|\mathbf{b}\| \leq 1$.

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p) \Rightarrow \mathbf{x} \stackrel{M}{\infty} 2e^{-(\cdot)^2/2} \text{ and } \mathbf{x} \stackrel{\mathbb{E}}{\infty} 2e^{-(\cdot)^2/2}.$$

A fundamental fact about the above concentration is that it does *not* depend on the size p of the ambient space (neither in the tail nor in the head parameters). As such, arbitrarily large standard Gaussian vectors (and thus concatenation of *independent* n such vectors, as well as matrices $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ built from *independent* standard Gaussian vectors \mathbf{x}_i endowed with the Frobenius norm) also concentrate with *no dependence* on p, n .

This is in fact far from natural as, even for independent vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$, all of which being concentrated, the joint concentration of $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ with respect to the Euclidean norm in the product space generally comes along with a *loss of concentration rate proportional to n* . Besides, if the vectors \mathbf{x}_1 and \mathbf{x}_2 are both concentrated but *not* independent, the concatenation vector $(\mathbf{x}_1, \mathbf{x}_2)$ may not even be concentrated.

Remark 2.18 (On the location of Gaussian vectors). *To clearly understand the relation between a standard Gaussian random vector $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ and its dimension, note that, $\|\mathbf{x}\|$ having a chi-distribution with median $\sqrt{p} + O(1/\sqrt{p})$, its exponential concentration precisely implies*

$$\mathbb{P}(\|\mathbf{x}\| - \sqrt{p} > t) \leq 2e^{-(t+O(1/p))^2/4}.$$

Thus, $\mathbf{x} \in \mathbb{R}^p$ is a random vector that essentially lives close to a sphere of radius $O(\sqrt{p})$ and thickness $O(1)$ or, equivalently, \mathbf{x}/\sqrt{p} is a random vector distributed close to \mathbb{S}^{p-1} , the unit sphere in \mathbb{R}^p , with actual distance to the sphere vanishing as $O(1/\sqrt{p})$. The vector \mathbf{x} is thus nowhere near its expected value $\mathbf{0}$ (see again Figure 1.6 for an illustration).

This remark is fundamental as it disrupts with the small-dimensional mental image, where \mathbf{x} lives close to its mean. In 1D to 3D, one indeed visualizes that (independent) Gaussian random vectors are densely “concentrated” around their mean (close to the center of the bell-shaped distribution). The intuitive extension of this visualization to larger dimensions would, however, be erroneous.

As for concentrated random variables, Lipschitz concentrated random vectors are stable through Lipschitz mapping in the sense that, for all 1-Lipschitz $\phi: E \rightarrow E'$ with respect to norms $\|\cdot\|_E$ and $\|\cdot\|_{E'}$,

$$\mathbf{x} \stackrel{M}{\infty} \alpha \Rightarrow \phi(\mathbf{x}) \stackrel{M}{\infty} \alpha. \tag{2.63}$$

Convex (Lipschitz) Concentration

To define convex concentration, we need to recall the notion of quasi-convex functions: $f: E \rightarrow \mathbb{R}$ is quasi-convex if, for all $t \in \mathbb{R}$, the sets $\{\mathbf{x} \in E \mid f(\mathbf{x}) \leq t\}$ are convex sets, that is, for all $t \in [0, 1]$ and $\mathbf{x}, \mathbf{y} \in E$, $f(t\mathbf{x} + (1-t)\mathbf{y}) \leq \max\{f(\mathbf{x}), f(\mathbf{y})\}$. In particular, convex functions are quasi-convex (thus the notion generalizes convexity) and, for $E = \mathbb{R}$, all monotonous functions (even concave ones) are quasi-convex.

Then, we have the following definition of convex concentration.

Definition 10 (Convex concentration). *A vector $\mathbf{x} \in E$ is (Lipschitz) convexly concentrated for the norm $\|\cdot\|$ if, for any 1-Lipschitz and quasi-convex function $f: E \rightarrow \mathbb{R}$, we have either of the conditions*

$$\begin{aligned} f(\mathbf{x}) &\propto \alpha, \text{ denoted } \mathbf{x} \propto_c \alpha \\ f(\mathbf{x}) &\in M_f \pm \alpha, \text{ denoted } \mathbf{x} \stackrel{M}{\propto}_c \alpha \\ f(\mathbf{x}) &\in \mathbb{E}[f(\mathbf{x})] \pm \alpha, \text{ denoted } \mathbf{x} \stackrel{\mathbb{E}}{\propto}_c \alpha \end{aligned}$$

holds, where M_f is a median of $f(\mathbf{x})$.

Obviously, all Lipschitz convex functions being Lipschitz, Lipschitz concentration implies convex concentration (which itself implies the even less demanding linear concentration); for instance, in the case of exponential concentration,

$$\mathbf{x} \stackrel{\mathbb{E}}{\propto}_c C e^{-(\cdot/\sigma)^q} \Rightarrow \mathbf{x} \stackrel{\mathbb{E}}{\propto}_c C e^{-(\cdot/\sigma)^q} \Rightarrow \mathbf{x} \in \mathbb{E}[\mathbf{x}] \pm e^{-(\cdot/\sigma)^q}.$$

The interest for convex concentration is related to the following result due to Talagrand [1995, Theorem 4.1.1]: Let $\mathbf{x} \in \{0,1\}^p$ be a random vector of independent entries, then

$$\mathbf{x} \stackrel{M}{\propto}_c 4e^{-(\cdot)^2/4}.$$

However, convex concentration has the major limitation that quasi-convex functions are *not* stable by composition. This prevents the simple adaptation of numerous results obtained for Lipschitz (or linear) concentration. Yet, for f quasi-convex and g affine, $f \circ g$ is still quasi-convex.

Nonetheless, the results necessary to our present random matrix analysis of sample covariance matrix models can fortunately be extended.

Convex Concentration Transversally to a Group Action

A last convenient notion of concentration, dedicated to random matrix theory, consists in transferring concentration from \mathbf{X} to the vector of its singular values. This will help transfer concentration from the data to linear statistics of the eigenvalues of the sample covariance matrix. To this end though, convex concentration is too demanding and we need to further restrict the space of functions as follows.

Definition 11 (Convex concentration transversally to group action). *Let $\mathbf{x} \in E$ and G a group acting on E . Then, \mathbf{x} is convexly α -concentrated transversally to the action of G if, for all quasi-convex 1-Lipschitz and G -invariant function f (i.e., $f(g \cdot \mathbf{x}) = f(\mathbf{x})$ for $g \in G$), $f(\mathbf{x}) \propto \alpha$. This is denoted $\mathbf{x} \propto_G^T \alpha$.*

In particular, denote $\sigma(\mathbf{X}) = (\sigma_1(\mathbf{X}), \dots, \sigma_{\min\{p,n\}}(\mathbf{X}))$ the vector of the singular values of $\mathbf{X} \in \mathbb{R}^{p \times n}$ (i.e., $\sigma_i(\mathbf{X}) = \sqrt{\lambda_i(\mathbf{X}\mathbf{X}^T)}$ for $i \leq \min\{p,n\}$), and define the group $\mathcal{O}_{p,n} = \{(\mathbf{U}, \mathbf{V}) \in \mathbb{R}^{p \times p} \times \mathbb{R}^{n \times n} \text{ orthonormal}\}$ acting on $\mathbb{R}^{p \times n}$ by $(\mathbf{U}, \mathbf{V}) \cdot \mathbf{M} = \mathbf{U}\mathbf{M}\mathbf{V}^T$ and the group \mathcal{S}_p of permutations of size p acting on \mathbb{R}^p by $\tau \cdot \mathbf{y} = (\mathbf{y}_{\tau(1)}, \dots, \mathbf{y}_{\tau(p)})$.

Then, we have the following result, inspired by Davis [1957],

$$\mathbf{X} \underset{\mathcal{O}_{p,n}}{\infty} \alpha \Leftrightarrow \sigma(\mathbf{X}) \underset{\mathcal{S}_{\min\{p,n\}}}{\infty} \alpha. \tag{2.64}$$

2.7.4 A Concentration Inequality Version of Theorem 2.6

Equipped with these elementary results, we can now provide an extension of the fundamental Theorem 2.6 to the case of concentrated (random) data vectors.

Before getting to the main result, we introduce some preliminary lemmas, which generalize classical random matrix results to the concentration of measure framework. Most of these results and their corresponding proofs can be found in Louart and Couillet [2018].

Trace Lemma

A first result of importance concerns the extension of the “quadratic-form-close-to-the-trace” lemma, Lemma 2.11, from a moment-based version to a concentration of measure setting. The result consists in a generalization of a popular result in concentration of measure theory known as Hanson–Wright’s theorem (see, e.g., Vershynin [2018, Theorem 6.2.1] for a version of random vectors having independent subGaussian entries).

Lemma 2.21 (Trace lemma for concentrated vectors). *Let $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\mathbf{x} \in \mathbb{R}^p$ such that $\mathbf{x} \underset{c}{\infty} C e^{-(\cdot/\sigma)^q}$. Then,*

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \in \text{tr}(\mathbb{E}[\mathbf{x}\mathbf{x}^T] \mathbf{A}) \pm C' \left(e^{-(\cdot/4\sigma \|\mathbf{A}\| \cdot \mathbb{E}[\|\mathbf{x}\|])^q} + e^{-(\cdot/2 \|\mathbf{A}\| \sigma^2)^{\frac{q}{2}}} \right)$$

for some constant $C' > 0$ depends only on C and q .

This lemma follows almost automatically from two elementary ingredients of the concentration of measure theory: (i) assuming first that \mathbf{A} is nonnegative definite, $\mathbf{x}^T \mathbf{A} \mathbf{x} = \|\mathbf{A}^{\frac{1}{2}} \mathbf{x}\|^2$ with $\|\mathbf{A}^{\frac{1}{2}} \mathbf{x}\|$ a concentrated random variable (it is a Lipschitz and convex function of \mathbf{x}) which, (ii) from the concentration of powers of concentrated random variables (2.61) for $k = 2$, gives the concentration result, however around $(\mathbb{E}[\|\mathbf{A}^{\frac{1}{2}} \mathbf{x}\|])^2$. It then suffices to apply, for example Ledoux [2005, Proposition 1.9] which states that, if a random variable exponentially concentrates around some constant $(\mathbb{E}[\|\mathbf{A}^{\frac{1}{2}} \mathbf{x}\|])^2$ here), then up to a change of constant, it also exponentially concentrates around its expectation. For generic \mathbf{A} , it suffices to write \mathbf{A} as the sum of its symmetric nonnegative and symmetric negative parts.

This lemma particularly stresses the technical convenience of the concentration of measure framework. The key random matrix results, such as Lemma 2.11 for vectors of i.i.d. entries, often rely on dedicated tools and possibly heavy (combinatorial) proof techniques. Here, the concentration of measure alternative to Lemma 2.11 follows from a mere few-line argument (once the elementary tools of the theory are in place). Besides, the exponential rate of convergence is very versatile and particularly ensures the uniform convergence of $\{\mathbf{x}_i^T \mathbf{A} \mathbf{x}_i, i = 1, \dots, n\}$, for n any polynomial in p ; using

the moment method would demand to systematically compute high-order moments of $\mathbf{x}_i^T \mathbf{A} \mathbf{x}_i$ to obtain uniform convergence over large n (e.g., with Markov’s inequality).

Concentration of the Stieltjes Transform

Next, we generalize the convergence of Stieltjes transforms in a generic concentration of measure form.

Lemma 2.22 (Trace of Resolvent). *For $\mathbf{X} \in \mathbb{R}^{p \times n}$ equipped with the Frobenius norm, and $\mathbf{Q}(z) = (\frac{1}{n} \mathbf{X} \mathbf{X}^T - z \mathbf{I}_p)^{-1}$ for $z < 0$,*

$$\mathbf{X} \infty_c \alpha \text{ in } (\mathbb{R}^{p \times n}, \|\cdot\|_F) \Rightarrow \text{tr} \mathbf{Q}(z) \infty 2\alpha \left(\frac{\sqrt{n|z|^3(\cdot)}}{8 \min\{p, n\}} \right).$$

To prove this lemma, first recall that $\mathbf{X} \infty_c \alpha \Rightarrow \sigma(\mathbf{X}) \infty_{\mathcal{S}_d}^T \alpha$ with $d = \min\{p, n\}$. Also, $\text{tr} \mathbf{Q}(z) = \sum_{i=1}^d f(\sigma_i(\mathbf{X})/\sqrt{n})$ for $f: \mathbb{R}_+ \rightarrow \mathbb{R}, s \mapsto 1/(s^2 - z)$. This function f is $(2|z|^{-3/2})$ -Lipschitz (checked by bounding its derivative) and the mapping $(s_1, \dots, s_d) \mapsto \sum_{i=1}^d s_i$ is evidently \mathcal{S}_d -invariant. However, f is not quasi-convex but can be written as the sum $f = g - h$ of two quasi-convex $4|z|^{-3/2}$ -Lipschitz functions ($h(s) = (s/|z| - 1/\sqrt{|z|})^2 \cdot 1_{\{s \in [0, \sqrt{|z|}]\}}$ and $g = f + h$). Consequently, since $\mathbf{X} \infty_c \alpha \Rightarrow \mathbf{X} \infty_{\mathcal{O}_{p,n}}^T \alpha$, we have from (2.64) both the concentration of $\sum_i g(\sigma_i(\mathbf{X}))$ and of $\sum_i h(\sigma_i(\mathbf{X}))$, and it then remains to apply the result on the concentration of the sum of two concentrated random variables to obtain the result.

Again here, the proof is elegant and immediate, although the mapping $\mathbf{X} \mapsto \text{tr} \mathbf{Q}(z)$ is highly nontrivial from a statistical standpoint. Note, in particular, that the technical difficulty raised by the nonconvexity of f would not have been a problem if we had rather assumed Lipschitz concentration $\mathbf{X} \infty \alpha$ for \mathbf{X} (which we recall is more demanding for \mathbf{X} and would in particular *exclude* the case of \mathbf{X} with bounded i.i.d. entries).

Concentration of the Resolvent Q and its Deterministic Equivalents

The approach followed in the previous lemma uses the convenient decomposition of $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ as $f = g - h$ for two convex and Lipschitz functions g and h . It does not seem that the mapping $f(\mathbf{X}) = \mathbf{Q}_{\frac{1}{n} \mathbf{X} \mathbf{X}^T}(z)$ from $\mathbb{R}^{p \times n}$ to $\mathbb{R}^{p \times p}$ can be treated similarly, as no such Lipschitz function division can be exploited. One must there resort to the additional strength of exponential concentration to divide the space $\mathbb{R}^{p \times n}$ into a compact space for the operator norm $\{\mathbf{X} \mid \|\mathbf{X}\| \leq K\sqrt{n}\}$, where f will be shown to be automatically Lipschitz (as its image is bounded) and the complement space $\{\mathbf{X} \mid \|\mathbf{X}\| > K\sqrt{n}\}$ which is of vanishing probability for all large $K > 0$.

Regrouping these two results, we have the following concentration for the resolvent.

Lemma 2.23 (Concentration of $\mathbf{Q}_{\frac{1}{n} \mathbf{X} \mathbf{X}^T}$). *For $\mathbf{X} \in \mathbb{R}^{p \times n}$ and $z < 0$, let $\mathbf{Q}(z) = (\frac{1}{n} \mathbf{X} \mathbf{X}^T - z \mathbf{I}_p)^{-1}$. Then, we have the following two results*

$$\begin{aligned} \mathbf{X} \infty \alpha &\Rightarrow \mathbf{Q}(z) \infty \alpha \left(\sqrt{n|z|^3(\cdot)}/2 \right) \\ \mathbf{X} \infty_c \mathbb{E} C e^{-(\cdot/\sigma)^q} &\Rightarrow \mathbf{Q}(z) \in \mathbb{E} \mathbf{Q}(z) \pm 2C e^{-\left(\sqrt{n|z|^3(\cdot)}/4\sigma\right)^q} \end{aligned}$$

where the left-hand side concentrations are understood in $(\mathbb{R}^{p \times n}, \|\cdot\|_F)$ and the right-hand side in $(\mathbb{R}^{p \times p}, \|\cdot\|_F)$.

This result is in fact quite powerful and automatically induces (and vastly generalizes) the notion of deterministic equivalent of Definition 4, that is, it implies that $\frac{1}{n} \text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q}) \xrightarrow{\text{a.s.}} 0$ and $\mathbf{a}^T(\mathbf{Q} - \mathbb{E}\mathbf{Q})\mathbf{b} \xrightarrow{\text{a.s.}} 0$ for all $\mathbf{A}, \mathbf{a}, \mathbf{b}$ of unit norm, as $n, p \rightarrow \infty$. Indeed, first recall that the first statement (of Lipschitz concentration) implies that

$$\mathbf{X} \stackrel{\mathbb{E}}{\infty} \alpha \Rightarrow \mathbf{Q}(z) \in \mathbb{E}\mathbf{Q}(z) \pm \alpha \left(\sqrt{n|z|^3}(\cdot)/2 \right)$$

(since Lipschitz concentration around the mean implies linear concentration around the mean). Next, note that the linear concentrations of \mathbf{Q} (under either Lipschitz or convex-Lipschitz-exponential concentration for \mathbf{X}) hold here with respect to the Frobenius norm of $\mathbf{X} \in \mathbb{R}^{p \times n}$. That is, for $\mathbf{A} \in \mathbb{R}^{p \times p}$ of unit Frobenius (rather than only spectral) norm,⁴⁸

$$\text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q}) = O(n^{-\frac{1}{2}}).$$

In particular, letting $p/n \rightarrow c > 0$, from $\|\mathbf{A}\|_F \leq \sqrt{\text{rank}(\mathbf{A})} \cdot \|\mathbf{A}\|$ (with $\|\cdot\|$ the operator norm) and $\|\mathbf{A}\| \leq \|\mathbf{A}\|_F$, we have (i) if $\mathbf{A} = \mathbf{a}\mathbf{b}^T$ is of unit rank with \mathbf{a}, \mathbf{b} of unit norm, then $\text{tr} \mathbf{a}\mathbf{b}^T(\mathbf{Q} - \mathbb{E}\mathbf{Q}) = \mathbf{a}^T(\mathbf{Q} - \mathbb{E}\mathbf{Q})\mathbf{b} = O(n^{-1/2})$, while (ii) if \mathbf{A} is of arbitrary rank (say $\text{rank}(\mathbf{A}) = p$) and of unit spectral norm, then we have $p^{-\frac{1}{2}} \text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q}) = O(n^{-1/2})$ so that $\frac{1}{p} \text{tr} \mathbf{A}(\mathbf{Q} - \mathbb{E}\mathbf{Q}) = O(n^{-1})$.

Of course, since $\|\cdot\| \leq \|\cdot\|_F$ in $\mathbb{R}^{p \times p}$, Lemma 2.23 applies to \mathbf{Q} in $(\mathbb{R}^{p \times p}, \|\cdot\|)$ in a spectral norm sense as well.

The proof of the first part of the lemma is again rather straightforward, once the basic concentration of measure arguments are in place. Here, we simply use the fact that the mapping $f: \mathbb{R}^{p \times n} \rightarrow \mathbb{R}^{p \times p}$, $\mathbf{X} \mapsto \mathbf{Q}(z)$ is $(2/\sqrt{|z|^3 n})$ -Lipschitz. Indeed, by the resolvent identity, Lemma 2.1,

$$f(\mathbf{X} + \mathbf{H}) - f(\mathbf{X}) = -\frac{1}{n} f(\mathbf{X} + \mathbf{H})((\mathbf{X} + \mathbf{H})\mathbf{H}^T + \mathbf{H}\mathbf{X}^T) f(\mathbf{X})$$

so that, from $\|f(\mathbf{X})\mathbf{X}\| \leq \sqrt{n/|z|}$, $\|f(\mathbf{X})\| \leq 1/|z|$, and $\|\mathbf{A}\mathbf{B}\|_F \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|_F$ (where $\|\cdot\|$ is the operator norm), we have $\|f(\mathbf{X} + \mathbf{H}) - f(\mathbf{X})\|_F \leq 2\|\mathbf{H}\|_F/\sqrt{|z|^3 n}$ and thus the result.

The proof of the second part is less immediate. Since the result is a linear concentration of the resolvent, one needs to control the concentration of the random variable $\text{tr} \mathbf{A}\mathbf{Q}$ obtained for arbitrary $\mathbf{A} \in \mathbb{R}^{p \times p}$ with $\|\mathbf{A}\|_F \leq 1$. This is obtained by considering the mapping $f: \mathbf{X} \mapsto \text{tr} \mathbf{A}\mathbf{Q}$, with the major difference from Lemma 2.22 that $f(\mathbf{Q})$ is now *not* a mere combination of the singular values of \mathbf{Q} . The function f is not convex (as already discussed in Lemma 2.22) but can again be divided as $f = h - g$ with $g: \mathbf{X} \mapsto \frac{1}{n|z|^2} \text{tr} \mathbf{X}\mathbf{X}^T$ and $h = f + g$ both convex, with h Lipschitz and g Lipschitz

⁴⁸ One must be careful not to confuse the steps of the proof which use a smart division of $\mathbb{R}^{p \times n}$ into bounded and unbounded operator norm $\|\mathbf{X}\|/\sqrt{n}$, and the fact that the ultimate concentration results hold with respect to the Frobenius (instead of spectral) norm.

on the bounded region $\{\mathbf{X} \mid \|\mathbf{X}\| \leq K\sqrt{n}\}$. Using a truncation method by considering $(\mathbf{X}^K)_{ij} = \min\{1, \frac{K\sqrt{n}|z|^3}{\|\mathbf{X}\|_F}\}\mathbf{X}_{ij}$ for growing K , one obtains that the sequence of concentrated random variables $\text{tr}\mathbf{A}\mathbf{Q}^K = \text{tr}\mathbf{A}(\frac{1}{n}\mathbf{X}^K(\mathbf{X}^K)^\top - z\mathbf{I}_p)^{-1}$ converges in law to $\text{tr}\mathbf{A}\mathbf{Q}$, which can then be shown to imply that $\text{tr}\mathbf{A}\mathbf{Q}$ is also a concentrated random variable.

Main Result

Let us rephrase the setting of Theorem 2.6 by letting $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ be n i.i.d. random vectors with law \mathcal{L} such that

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \infty C e^{-(\cdot)^q/c}$$

for some $C, c, q > 0$ with respect to the Frobenius norm (which implies in particular, by the action of the 1-Lipschitz mapping $f: (\mathbf{x}_1, \dots, \mathbf{x}_n) \mapsto \mathbf{x}_i$, that each \mathbf{x}_i is itself concentrated). This request of *joint* rather than *individual* (vector) concentration may be considered demanding, but is at least satisfied by (i) $\mathbf{x}_i = \phi(\mathbf{y}_i)$ with 1-Lipschitz maps $\phi: \mathbb{R}^{p'} \rightarrow \mathbb{R}^p$ for (i-a) $\mathbf{y}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p'})$ or (i-b) \mathbf{y}_i uniformly distributed on the $\sqrt{p'}$ -radius sphere of $\mathbb{R}^{p'}$, or (ii) for \mathbf{x}_i composed of (an affine mapping of) i.i.d. entries supported on $[-1, 1]$, see Louart and Couillet [2018, Remark 3.2].

With the above results, and some specific technical arguments, we have the following *concentration of measure version* of Theorem 2.6.

Theorem 2.18 (Sample covariance of concentrated random vectors). *Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \infty C e^{-(\cdot)^q/c}$ with i.i.d. $\mathbf{x}_i \in \mathbb{R}^p$, and $z < 0$. Further assume that $\mathbb{E}\|\mathbf{x}_i\|/\sqrt{p}$ (or, if $q \geq 2$, simply $\|\mathbb{E}[\mathbf{x}_i]\|/\sqrt{p}$), $\text{tr}\Phi/p$ with $\Phi = \frac{1}{n}\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$, as well as p/n are all bounded. Then, for all large n ,*

$$\mathbf{Q}(z) \in \bar{\mathbf{Q}}(z) \pm C' e^{-(\sqrt{n}\cdot)^q/c'} \text{ in } (\mathbb{R}^{p \times n}, \|\cdot\|)$$

for some $C', c' > 0$, where

$$\bar{\mathbf{Q}}(z) = \left(\frac{\Phi}{1 + \delta(z)} - z\mathbf{I}_p \right)^{-1}$$

and $\delta(z)$ is the unique positive solution to $\delta(z) = \frac{1}{n} \text{tr}\Phi\bar{\mathbf{Q}}(z)$.

Remark 2.19 (On real $z < 0$). *It must be noted here that the concentration framework devised in this section is only valid for real-valued matrices and thus Theorem 2.18 holds here for $z < 0$ only. Using additional arguments (of complex analytic extension of $\mathbf{Q}(z)$ and $\bar{\mathbf{Q}}(z)$), Theorem 2.18 can be naturally extended to all $z \in \mathbb{C} \setminus \mathbb{R}^+$.*

Denoting $\delta(z) = -1 - \frac{1}{z\tilde{m}_p(z)}$ and $\Phi = \mathbf{C}$, it comes immediately that the deterministic equivalent $\bar{\mathbf{Q}}$ in Theorem 2.18 above has the same ‘‘formal statement’’ as in Theorem 2.6; we shall see that using $\delta(z)$ rather than $\tilde{m}_p(z)$ is more convenient under the concentration of measure framework. Yet, there are a few key differences to raise between both theorems. First, $\Phi = \frac{1}{n}\mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ is *not* a covariance matrix as the present concentration of measure on \mathbf{X} does not impose that $\mathbb{E}[\mathbf{X}] = \mathbf{0}$. Also, the deterministic

equivalent $\bar{\mathbf{Q}}(z)$ comes along with a convergence speed and an exponential tail, which are both more practical than a mere almost sure convergence of specific statistics.

Theorem 2.18 unfolds from the same idea introduced in the proof of the Marčenko–Pastur law (Theorem 2.4), by successively introducing two deterministic equivalents. We provide here the basic arguments of the proof. We already know from Lemma 2.23 that $\mathbf{Q}(z) \in \mathbb{E}\mathbf{Q}(z) \pm Ce^{-c(\sqrt{n}\cdot)^q}$ for some $C, c > 0$ and it only remains to show that $\|\mathbb{E}\mathbf{Q}(z) - \bar{\mathbf{Q}}(z)\|$ is small.

To this end, we introduce the first deterministic equivalent

$$\bar{\bar{\mathbf{Q}}}(z) = \left(\frac{\Phi}{1 + \delta'(z)} - z\mathbf{I}_p \right)^{-1},$$

where $\delta'(z) = \frac{1}{n}\mathbb{E}[\mathbf{x}^\top \mathbf{Q}_-(z)\mathbf{x}] = \frac{1}{n}\text{tr}(\Phi\mathbb{E}\mathbf{Q}_-)$ for $\mathbf{Q}_- \in \mathbb{R}^{p \times p}$ the resolvent of $\frac{1}{n}\mathbf{X}\mathbf{X}^\top - \frac{1}{n}\mathbf{x}\mathbf{x}^\top$ and \mathbf{x} any column of \mathbf{X} . Applying the same ideas as in the proof of Theorem 2.4, we obtain (we discard the argument z for readability)

$$\begin{aligned} \mathbb{E}\mathbf{Q} - \bar{\bar{\mathbf{Q}}} &= \mathbb{E} \left[\mathbf{Q} \left(\frac{\Phi}{1 + \delta'} - \frac{1}{n}\mathbf{X}\mathbf{X}^\top \right) \right] \bar{\bar{\mathbf{Q}}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\mathbf{Q} \left(\frac{\Phi}{1 + \delta'} - \mathbf{x}_i\mathbf{x}_i^\top \right) \right] \bar{\bar{\mathbf{Q}}} = \mathbb{E} \left[\mathbf{Q} \left(\frac{\Phi}{1 + \delta'} - \mathbf{x}\mathbf{x}^\top \right) \right] \bar{\bar{\mathbf{Q}}} \end{aligned}$$

which, along with $\mathbf{Q} = \mathbf{Q}_- - \frac{1}{n} \frac{\mathbf{Q}_-\mathbf{x}\mathbf{x}^\top\mathbf{Q}_-}{1 + \frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x}}$ and $\mathbf{Q}\mathbf{x} = \frac{\mathbf{Q}_-\mathbf{x}}{1 + \frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x}}$ from Lemma 2.8, gives

$$\begin{aligned} \mathbb{E}\mathbf{Q} - \bar{\bar{\mathbf{Q}}} &= \mathbb{E}[\mathbf{E}_1] - \mathbb{E}[\mathbf{E}_2], \\ \mathbf{E}_1 &= \mathbf{Q}_- \left(\frac{\Phi}{1 + \delta'} - \frac{\mathbf{x}\mathbf{x}^\top}{1 + \frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x}} \right) \bar{\bar{\mathbf{Q}}}, \quad \mathbf{E}_2 = \frac{1}{n(1 + \delta')} \mathbf{Q}_-\mathbf{x}\mathbf{x}^\top\mathbf{Q}_-\Phi\bar{\bar{\mathbf{Q}}}. \end{aligned}$$

To bound $\|\mathbb{E}\mathbf{Q} - \bar{\bar{\mathbf{Q}}}\|$ it suffices to bound $|\mathbf{a}^\top(\mathbb{E}\mathbf{Q} - \bar{\bar{\mathbf{Q}}})\mathbf{a}|$ for any unit norm \mathbf{a} . Applying Cauchy–Schwarz inequality twice we have

$$\begin{aligned} |\mathbf{a}^\top \mathbb{E}[\mathbf{E}_1]\mathbf{a}| &= \left| \mathbb{E} \left[\mathbf{a}^\top \mathbf{Q}_- \mathbf{x} \mathbf{x}^\top \bar{\bar{\mathbf{Q}}}\mathbf{a} \cdot \frac{\frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x} - \delta'}{(1 + \delta')(1 + \frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x})} \right] \right| \\ &\leq \mathbb{E} \left[|\mathbf{a}^\top \mathbf{Q}_-\mathbf{x}| \cdot |\mathbf{x}^\top \bar{\bar{\mathbf{Q}}}\mathbf{a}| \cdot \left| \frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x} - \delta' \right| \right] \\ &\leq \sqrt{\mathbb{E} \left[|\mathbf{a}^\top \mathbf{Q}_-\mathbf{x}|^2 \cdot \left| \frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x} - \delta' \right| \right] \cdot \mathbb{E} \left[|\mathbf{x}^\top \bar{\bar{\mathbf{Q}}}\mathbf{a}|^2 \cdot \left| \frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x} - \delta' \right| \right]} \\ &= O(n^{-\frac{1}{2}}) \end{aligned}$$

where we used here: (i) $\mathbf{a}^\top \bar{\bar{\mathbf{Q}}}\mathbf{x} \propto Ce^{-(\cdot)^q}$ and $\mathbf{a}^\top \mathbf{Q}_-\mathbf{x} \propto Ce^{-c(\cdot)^q}$ (from which $\mathbb{E}[|\mathbf{a}^\top \bar{\bar{\mathbf{Q}}}\mathbf{x}|^k] = O(1)$ and $\mathbb{E}[|\mathbf{a}^\top \mathbf{Q}_-\mathbf{x}|^k] = O(1)$) and (ii) $\frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x} \in \delta' \pm Ce^{-c(n\cdot)^{q/2}} + Ce^{-c(\sqrt{n}\cdot)^q}$ (from which $\mathbb{E}[|\frac{1}{n}\mathbf{x}^\top\mathbf{Q}_-\mathbf{x} - \delta'|^k] = O(n^{-\frac{k}{2}})$). The concentration results (i) and (ii) themselves unfold from the previous generic results on concentration of vectors and bilinear forms. Similarly,

$$|\mathbf{a}^\top \mathbb{E}[\mathbf{E}_2]\mathbf{a}| \leq \frac{1}{n} \sqrt{\mathbb{E}[|\mathbf{a}^\top \mathbf{Q}_-\mathbf{x}|^2] \cdot \mathbb{E}[|\mathbf{x}^\top \mathbf{Q}_-\Phi\bar{\bar{\mathbf{Q}}}\mathbf{a}|^2]} = O(n^{-1}).$$

We thus find that $\|\mathbb{E}\mathbf{Q} - \bar{\mathbf{Q}}\| = O(n^{-\frac{1}{2}})$. Integrated into $\mathbf{Q}(z) \in \mathbb{E}\mathbf{Q}(z) \pm Ce^{-c(\sqrt{n})^q}$, this gives $\mathbf{Q}(z) \in \bar{\mathbf{Q}} \pm Ce^{-c(\sqrt{n})^q}$.

It thus remains to show similarly that $\|\bar{\mathbf{Q}} - \bar{\bar{\mathbf{Q}}}\|$ is small. Note that

$$\|\bar{\mathbf{Q}} - \bar{\bar{\mathbf{Q}}}\| = \frac{|\delta' - \delta|}{(1 + \delta)(1 + \delta')} \|\bar{\mathbf{Q}}\Phi\bar{\bar{\mathbf{Q}}}\| \leq \frac{|\delta - \delta'|}{|z|}$$

and it thus suffices to control $\delta - \delta'$, which, by the implicit form of δ , satisfies

$$\begin{aligned} |\delta - \delta'| &= \frac{1}{n} |\text{tr} \Phi(\bar{\mathbf{Q}} - \bar{\bar{\mathbf{Q}}} + \bar{\bar{\mathbf{Q}}} - \mathbb{E}\mathbf{Q} + \mathbb{E}[\mathbf{Q} - \mathbf{Q}_-])| \\ &\leq \frac{1}{n} |\text{tr} \Phi(\bar{\mathbf{Q}} - \bar{\bar{\mathbf{Q}}})| + \frac{1}{n} \text{tr} \Phi \|\bar{\bar{\mathbf{Q}}} - \mathbb{E}\mathbf{Q}\| + \frac{1}{n} \text{tr} \Phi \|\mathbb{E}[\mathbf{Q} - \mathbf{Q}_-]\| \\ &\leq \sqrt{\frac{1}{n(1 + \delta)^2} \text{tr} \Phi^2 \bar{\mathbf{Q}}^2} \cdot \sqrt{\frac{1}{n(1 + \delta')^2} \text{tr} \Phi^2 \bar{\bar{\mathbf{Q}}}^2} \cdot |\delta - \delta'| + O(n^{-\frac{1}{2}}) \end{aligned}$$

where we used $\text{tr} \mathbf{A}\mathbf{B} \leq \|\mathbf{B}\| \cdot \text{tr} \mathbf{A}$ for symmetric and nonnegative definite $\mathbf{A} \in \mathbb{R}^{p \times p}$, and $\|\mathbb{E}[\mathbf{Q} - \mathbf{Q}_-]\| = O(n^{-1/2})$, which unfolds from

$$\|\mathbb{E}[\mathbf{Q} - \mathbf{Q}_-]\| = \frac{1}{n} \left\| \mathbb{E} \frac{\mathbf{Q}_- \mathbf{x} \mathbf{x}^T \mathbf{Q}_-}{1 + \frac{1}{n} \mathbf{x}^T \mathbf{Q}_- \mathbf{x}} \right\| = \frac{1}{n} \left\| \frac{\mathbb{E}[\mathbf{Q}_- \Phi \mathbf{Q}_-]}{1 + \delta'} \right\| + O(n^{-\frac{1}{2}}).$$

The prefactor of $|\delta - \delta'|$ is strictly less than 1 for all large n , and thus $|\delta - \delta'| = O(n^{-\frac{1}{2}})$, which concludes the proof.

2.8 Concluding Remarks

This section explored basic to advanced spectral properties of a family of random matrix models, with a strong emphasis on the sample covariance matrix model (Theorem 2.6), in the regime of large and commensurable data number n and dimension p . Despite the simplicity of its definition, we saw that the limiting spectral measure of the sample covariance matrix is far from trivial, that advanced techniques from complex analysis can be used to perform statistical inference, and that, unlike in the classical $n \rightarrow \infty$ and p fixed regime, phase transition phenomena arise, below which some inference problems are asymptotically insoluble.

Fortunately, even if the statistical models used in concrete machine learning applications are often more involved, we will see, in the remainder of the book, that the main techniques and tools used to understand and improve various machine learning methods are essentially the same as those presented so far. In particular, we will see in the following sections that:

- in (not necessarily linear) regression problems, the resolvent (of sample covariance matrices, of kernel matrices, of the Gram matrix of nonlinear random feature maps, etc.) will systematically appear as the central object of interest (which is reminiscent of the fact that regression is an inverse problem);

- in classification problems, the spectrum of kernel random matrices and Laplacian random matrices (for spectral clustering or spectral community detection), or different types of functionals of these kernel and Laplacian random matrices (for supervised or semi-supervised graph-based learning) will play an important role; the performance achieved by these methods, given in terms of misclassification rates, probability of false alarms, etc., will, in particular, demand the evaluation of the limiting means and variances of these functionals;
- in the specific case of spectral or subspace methods, such as PCA, manifold-based clustering, spectral clustering, or community detection, the aforementioned phase transition phenomena will arise and show that there exist “strict” limitations for these methods: In particular, a minimal samples-over-dimension ratio exists below which no detection or classification is possible;
- even for optimization-based machine learning problems, such as generalized linear models [Nelder and Wedderburn, 1972], that rarely offer a solution explicitly defined from (the resolvent of) a particular random matrix, their large-dimensional (limiting) performance will be shown ultimately to depend, in an almost explicit way, on some slightly more involved random matrices; there, the twist will be to realize that some random quantities (not always easy to identify) converge and can asymptotically be replaced by deterministic equivalents obtained from a perturbation analysis (e.g., some sort of a local “linearization”).

Before delving into these applications, it is important to recall that we shall purposely place ourselves under the “realistic” situation, where the number of samples n cannot be chosen arbitrarily large (samples never really come for free in practice) and particularly not overwhelmingly larger than the typical dimension p of the data. More importantly, we also impose that the problem being addressed is not “asymptotically trivial,” that is, for p, n realistically large, the misclassification probability or the cost to be minimized will not vanish. This way, the asymptotic analysis ($n, p \rightarrow \infty$) will be a realistic representative of the finite (but not too small) dimensional and moderately difficult machine learning problem. This is quite different from many parallel theoretical machine learning works, which often aim at concluding (usually through the evaluation of error bounds, rather than exact results) that the algorithm under analysis provides an asymptotic perfect performance (vanishing misclassification rate or cost) in a certain growth regime of n with respect to p . Our vision instead is that, in the (more) realist large dimensional regime, n and p must be considered as both fixed (only not to too small values).

As such, to best appreciate the many results to come in the next chapters, these must be seen through this “finite-dimensional and realistic” lens.

2.9 Exercises

In this section, we provide short exercises to familiarize the reader with various useful notions and properties of random matrix calculus discussed this far in Chapter 2, with detailed solutions provided at <https://zhenyu-liao.github.io/book>.

2.9.1 Properties of the Stieltjes Transform

Exercise 1 (Stieltjes transform and moments). Show that the Stieltjes transform $m_\mu(z)$ defined in Definition 3, of a probability measure μ with bounded support (and thus finite moments), is a moment generating function in the sense that, for all $z \in \mathbb{C}$ such that $|z| > \max\{|\inf(\text{supp}(\mu))|, |\sup(\text{supp}(\mu))|\}$,

$$m_\mu(z) = -\frac{1}{z} \sum_{k=0}^{\infty} M_k z^{-k},$$

where $M_k = \int t^k \mu(dt)$.

From this formulation, propose a method to evaluate the successive moments of μ using m_μ .

Exercise 2 (Nonimmediate Stieltjes transforms). Let $\mathbf{X} \in \mathbb{R}^{n \times n}$ be a symmetric matrix and $\mathbf{Q}(z) = (\mathbf{X} - z\mathbf{I}_n)^{-1}$ its resolvent. Show that, for any $\mathbf{u} \in \mathbb{R}^n$ of unit norm $\|\mathbf{u}\| = 1$ and any \mathbf{A} nonnegative definite and such that $\text{tr} \mathbf{A} = 1$, the quantities $\mathbf{u}^\top \mathbf{Q}(z) \mathbf{u}$ and $\text{tr} \mathbf{A} \mathbf{Q}(z)$ are also Stieltjes transform of probability measures.

What are these measures and what are their supports?

Exercise 3 (Stieltjes transform and singular values). Let μ be a probability measure on \mathbb{R}^+ and ν, ν' be the measures defined by

$$\begin{aligned} \int f(t) \nu(dt) &= \int f(\sqrt{t}) \mu(dt) \\ \int f(t) \nu'(dt) &= \frac{1}{2} \left(\int f(t) \nu(dt) + \int f(-t) \nu(dt) \right) \end{aligned}$$

for all bounded continuous f .

What are ν and ν' when $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}$ for some $\lambda_1, \dots, \lambda_n \geq 0$?

Show that the Stieltjes transform $m_{\nu'}$ of ν' satisfies

$$m_{\nu'}(z) = z m_\mu(z^2).$$

Letting $\mathbf{X} \in \mathbb{R}^{n \times p}$ and μ be the empirical spectral measure of $\mathbf{X}\mathbf{X}^\top$ as in Definition 2, relate the Stieltjes transform of the empirical spectral measure of the matrix

$$\Gamma = \begin{bmatrix} \mathbf{0}_{n \times n} & \mathbf{X} \\ \mathbf{X}^\top & \mathbf{0}_{p \times p} \end{bmatrix} \in \mathbb{R}^{(n+p) \times (n+p)}$$

to that of the measure μ , and conclude on the nature of this Stieltjes transform for $n = p$.

Exercise 4 (Proof of Lemma 2.9: a special case). For $\mathbf{A}, \mathbf{M} \in \mathbb{R}^{p \times p}$ symmetric nonnegative definite matrices, $\mathbf{u} \in \mathbb{R}^p$, $\tau > 0$ and $z < 0$, show that

$$\left| \text{tr} \mathbf{A} \left(\mathbf{M} + \tau \mathbf{u}\mathbf{u}^\top - z \mathbf{I}_p \right)^{-1} - \text{tr} \mathbf{A} \left(\mathbf{M} - z \mathbf{I}_p \right)^{-1} \right| \leq \frac{\|\mathbf{A}\|}{|z|}.$$

Exercise 5 (Proof of Nash–Poincaré inequality, Lemma 2.14). The objective of the exercise is to show that, for $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} \in \mathbb{R}^{p \times p}$ and $f: \mathbb{R}^p \rightarrow \mathbb{R}$ of bounded

first- and second-order derivatives,

$$\text{Var}[f(\mathbf{x})] \leq \mathbb{E} \left[(\nabla f(\mathbf{x}))^\top \mathbf{C} \nabla f(\mathbf{x}) \right].$$

To this end, it is convenient to first define an “interpolating” Gaussian vector $\mathbf{x}(t) = \sqrt{t}\mathbf{x}_1 + \sqrt{1-t}\mathbf{x}_2$ for $t \in [0,1]$ with $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_1)$, $\mathbf{x}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_2)$ independent, and show, by applying successively the chain rule and Stein’s lemma, Lemma 2.13, that for twice differentiable g ,

$$\begin{aligned} \mathbb{E}[g(\mathbf{x}_1)] - \mathbb{E}[g(\mathbf{x}_2)] &= \int_0^1 \frac{d}{dt} \mathbb{E}[g(\mathbf{x}(t))] dt \\ &= \frac{1}{2} \int_0^1 \mathbb{E} \left[\nabla g(\mathbf{x}(t))^\top \mathbf{C}_1 \nabla g(\mathbf{x}(t)) - \nabla g(\mathbf{x}(t))^\top \mathbf{C}_2 \nabla g(\mathbf{x}(t)) \right] dt. \end{aligned}$$

From there, apply the result to the vectors $\mathbf{x}_1 = [\mathbf{y}^\top, \mathbf{y}^\top]^\top \in \mathbb{R}^{2p}$ and $\mathbf{x}_2 = [\mathbf{y}_1^\top, \mathbf{y}_2^\top]^\top \in \mathbb{R}^{2p}$ for $\mathbf{y}, \mathbf{y}_1, \mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ independent, and $g([\mathbf{a}^\top, \mathbf{b}^\top]^\top) = f(\mathbf{a})f(\mathbf{b})$. Conclude by an application of Cauchy–Schwarz inequality on the expectation under the resulting integrand and the observation that the bound on the integrand is constant with respect to t .

2.9.2 On Limiting Laws

Exercise 6 (The $\sqrt{|x - E|}$ behavior of the edges). Show that both the semicircle law (Theorem 2.5) and the Marčenko–Pastur law (Theorem 2.4, for $c \neq 1$) have a local $\sqrt{|x - E|}$ behavior at each of the edges E of their support.

Conclude on the typical number of eigenvalues of the Wishart matrix $\frac{1}{n} \mathbf{X} \mathbf{X}^\top \in \mathbb{R}^{p \times p}$ (with $\mathbf{X}_{ij} \sim \mathcal{N}(0,1)$ independent) and the Wigner $\frac{1}{\sqrt{n}} \mathbf{X} \in \mathbb{R}^{n \times n}$ (with say $\mathbf{X}_{ij} = \mathbf{X}_{ji} \sim \mathcal{N}(0,1)$ independent up to symmetry) found near the edges of their respective supports.

Relate this finding to the Tracy–Widom distribution of the fluctuations of the largest and smallest eigenvalues in Theorem 2.15.

What happens for the left edge of the support of the Marčenko–Pastur law and to the associated smallest eigenvalues of Wishart matrices when $\lim p/n = c = 1$? How many eigenvalues are then found close to the left edge in this so-called “hard-edge” setting? Conclude on the typical fluctuations of these eigenvalues and confirm numerically.

Exercise 7 (The $\sqrt{|x - E|}$ behavior in elaborate models). We here seek to extend the results in Exercise 6 to the sample covariance matrix model $\frac{1}{n} \mathbf{X} \mathbf{X}^\top$ where $\mathbf{X} = \mathbf{C}^{\frac{1}{2}} \mathbf{Z}$ with \mathbf{Z} having independent entries of zero mean, unit variance and \mathbf{C} having a bounded limiting spectral measure ν with fast decaying tails. We denote $\tilde{m}(z)$ the Stieltjes transform of the limiting spectral measure $\tilde{\mu}$ of $\frac{1}{n} \mathbf{X}^\top \mathbf{X}$.

Using Figure 2.5 as a reference and recalling the formulation for functional inverse

$$x(\tilde{m}) = -\frac{1}{\tilde{m}} + c \int \frac{t\nu(dt)}{1 + t\tilde{m}}$$

extensively discussed in Section 2.3.1, visually justify that $x''(\tilde{m})$ can be (complex) analytically extended in the neighborhood of the local extrema of $x(\tilde{m})$ (i.e., each

point \tilde{m} where $x'(\tilde{m}) = 0$) into a function $z(\tilde{m})$, which must locally coincide with the inverse Stieltjes transform of $\tilde{m}(z)$.

Deduce that $\tilde{m}(z)$ must be of the form $\sqrt{z - E}$ near an edge E and conclude.

Exercise 8 (Further results on $x(\tilde{m})$). We aim in this exercise to justify some of the visual observations in Figure 2.5 with the help of

$$x(\tilde{m}) = -\frac{1}{\tilde{m}} + c \int \frac{tv(dt)}{1+t\tilde{m}}.$$

Show that, for $\tilde{m}_1 \neq \tilde{m}_2$ such that $x'(\tilde{m}_1), x'(\tilde{m}_2) > 0$, we cannot have $x(\tilde{m}_1) = x(\tilde{m}_2)$: that is, the increasing segments of $x(\tilde{m})$ never “overlap.”

Besides, show that, if $\tilde{m}_1 < \tilde{m}_2$ are both of the same sign, and $x'(\tilde{m}_1), x'(\tilde{m}_2) > 0$, then $x(\tilde{m}_1) < x(\tilde{m}_2)$: that is, the increasing segments of $x(\tilde{m})$ never “swap.” To this end, we may prove the intermediary result

$$(\tilde{m}_1 - \tilde{m}_2) \left(1 - \int \frac{c\tilde{m}_1\tilde{m}_2t^2v(dt)}{(1+t\tilde{m}_1)(1+t\tilde{m}_2)} \right) = \tilde{m}_1\tilde{m}_2(x(\tilde{m}_1) - x(\tilde{m}_2))$$

and use Cauchy–Schwarz inequality to control the left-hand side term.

Finally show that, if v has bounded support, then $x(\tilde{m}) \rightarrow 0$ as $\tilde{m} \rightarrow \pm\infty$.

As a final remark, note that the only important observation about Figure 2.5, which we have not shown here is the fact that the points \tilde{m} where $x'(\tilde{m}) = 0$ must exist. In fact, this is not always the case and heavily depends on the nature of the tails of the measure v . Justify in particular that, for some v , there may be no asymptote on the edges of the domain of definition of $x(\cdot)$ (as opposed to what is seen in Figure 2.5).

2.9.3 On Eigen-Inference

Exercise 9 (Alternative estimates of $\frac{1}{p} \text{tr}(\frac{1}{n} \mathbf{X}\mathbf{X}^\top)^2$). Let $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Z}$ for $\mathbf{Z} \in \mathbb{R}^{p \times n}$ with independent standard Gaussian entries, and \mathbf{C} deterministic symmetric nonnegative definite, of bounded operator norm, and limiting spectral measure v .

Determine the limit, as $n, p \rightarrow \infty$ and $p/n \rightarrow c \in (0, \infty)$ of the (empirical) second-order moment

$$M_2 = \frac{1}{p} \text{tr} \left(\frac{1}{n} \mathbf{X}\mathbf{X}^\top \right)^2$$

as a function of the moments of v .

Retrieve the same result from the results of Exercise 1 along with the expression of the Stieltjes transform $m(z)$ of the limiting spectrum μ of $\frac{1}{n} \mathbf{X}\mathbf{X}^\top$. It may be useful to first show that $m(z)$ is also solution to

$$m(z) = \int \frac{v(dt)}{-z(1+ctm(z)) + (1-c)t}.$$

Exercise 10 (Location of the zeros of $\tilde{m}(z)$). Figure 2.7 and Remark 2.12 both show that the zeros η_1, \dots, η_n of $m_{\mathbf{X}}(z)$, the Stieltjes transform of a symmetric matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$, are interlaced with the eigenvalues $\lambda_1, \dots, \lambda_n$ of \mathbf{X} .

In the sample covariance matrix case $\mathbf{X} = \frac{1}{n} \mathbf{Z}^T \mathbf{C} \mathbf{Z}$ with $\mathbf{Z} \in \mathbb{R}^{p \times n}$ having independent standard Gaussian entries and \mathbf{C} with limited spectral measure ν of bounded and connected support, this means that (up to zero eigenvalues) the roots η_i of $m_{\frac{1}{n} \mathbf{Z}^T \mathbf{C} \mathbf{Z}}(z)$ are all found in the limiting support of the empirical spectral measure $\tilde{\mu}$ of $\frac{1}{n} \mathbf{Z}^T \mathbf{C} \mathbf{Z}$, at the possible exception of the leftmost η_1 .

Using a change of variable involving $\tilde{m}(z)$ of the formula

$$0 = \frac{1}{2\pi i} \oint_{\Gamma} \frac{dw}{w}$$

for all Γ not enclosing zero, then the approximation $\tilde{m}(z) = m_{\frac{1}{n} \mathbf{Z}^T \mathbf{C} \mathbf{Z}}(z) + o(1)$ and finally the residue theorem show that no zero of $m_{\frac{1}{n} \mathbf{Z}^T \mathbf{C} \mathbf{Z}}(z)$ can be found at macroscopic distance from the limiting support of $\tilde{\mu}$.

This conclusion is of practical interest to statistical inference applications discussed in Section 2.4.1 and in particular, to the explicit expression in (2.45) from (2.44), for which case this result ensures the existence of a valid contour that circles around all the λ_i poles and η_i poles, at least almost surely for sufficiently large n, p . (And the leftmost η_1 is not a problem.)

2.9.4 Spiked Models

Exercise 11 (Additive spiked model). *Similar to Theorem 2.13, the phase transition threshold for the additive model $\frac{1}{n} \mathbf{X} \mathbf{X}^T + \mathbf{P}$ for \mathbf{X} having i.i.d. entries of zero mean, unit variance and low rank $\mathbf{P} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^T$ with $\ell_1 > \dots > \ell_k > 0$ is determined by the condition*

$$\ell_i > \sqrt{c}(1 + \sqrt{c})$$

with $c = \lim p/n$ as $p, n \rightarrow \infty$. Under this condition, show that the (almost sure) limiting value of the corresponding isolated eigenvalue $\hat{\lambda}_i$ of $\frac{1}{n} \mathbf{X} \mathbf{X}^T + \mathbf{P}$ is given by

$$\hat{\lambda}_i \xrightarrow{\text{a.s.}} \lambda_i = 1 + \ell_i + \frac{c}{\ell_i - c}.$$

Further show, similar to Theorem 2.14 that, letting $\hat{\mathbf{u}}_i$ be the eigenvector associated with $\hat{\lambda}_i$, we have

$$|\hat{\mathbf{u}}_i^T \mathbf{u}_i|^2 \xrightarrow{\text{a.s.}} 1 - \frac{c}{(\ell_i - c)^2}.$$

Exercise 12 (Additive spiked model: the Wigner case). *Let \mathbf{X} be symmetric with $[\mathbf{X}]_{ij}$, $i \geq j$, i.i.d. with zero mean and unit variance. As in Exercise 11, show that the “spiked” phase transition threshold for the model $\mathbf{X}/\sqrt{n} + \mathbf{P}$ with $\mathbf{P} = \sum_{i=1}^k \ell_i \mathbf{u}_i \mathbf{u}_i^T$, with $\ell_1 > \dots > \ell_k > 0$ is determined by the condition*

$$\ell_i > 1$$

and that, under this condition, the isolated eigenvalue $\hat{\lambda}_i$ of $\frac{1}{n} \mathbf{X} + \mathbf{P}$ associated with ℓ_i satisfies

$$\hat{\lambda}_i \xrightarrow{\text{a.s.}} \lambda_i = \ell_i + \frac{1}{\ell_i}.$$

Show finally that, for $\hat{\mathbf{u}}_i$ the eigenvector associated with $\hat{\lambda}_i$, we have

$$|\hat{\mathbf{u}}_i^\top \mathbf{u}_i|^2 \xrightarrow{\text{a.s.}} 1 - \frac{1}{\ell_i^2}.$$

2.9.5 Deterministic Equivalent

Exercise 13 (Sketch of proof of Theorem 2.17). *Inspired by the (sketch of) proof of Theorem 2.6, prove Theorem 2.17 using*

- (i) *the trace lemma adapted to Haar random matrices, Lemma 2.16; and*
- (ii) *Stein’s lemma adapted to Haar random matrices, Lemma 2.17.*

Exercise 14 (Higher-order deterministic equivalent). *Theorem 2.4 provides a deterministic equivalent for the resolvent $\mathbf{Q} = (\frac{1}{n}\mathbf{X}\mathbf{X}^\top - z\mathbf{I}_p)^{-1}$ for $\mathbf{X} \in \mathbb{R}^{p \times n}$ having i.i.d. zero-mean and unit-variance entries, which, according to Notation 1, provides access to the asymptotic behavior of $\mathbf{a}^\top \mathbf{Q} \mathbf{b}$. In many machine learning applications, however, the object of natural interest (e.g., the mean squared error in a regression context and the variance in a classification context) often involves the asymptotic behavior of $\mathbf{a}^\top \mathbf{Q} \mathbf{A} \mathbf{Q} \mathbf{b}$, which requires a deterministic equivalent for random matrices of the type $\mathbf{Q} \mathbf{A} \mathbf{Q}$, for some \mathbf{A} independent of \mathbf{Q} . In particular, for $\mathbf{Q} \leftrightarrow \bar{\mathbf{Q}}$ (such that $\|\mathbb{E}[\mathbf{Q}] - \bar{\mathbf{Q}}\| \rightarrow 0$), $\bar{\mathbf{Q}} \mathbf{A} \bar{\mathbf{Q}}$ is in general not a deterministic equivalent for $\mathbf{Q} \mathbf{A} \mathbf{Q}$. This is due to the fact that*

$$\mathbb{E}[\mathbf{Q} \mathbf{A} \mathbf{Q}] \not\approx \mathbb{E}[\mathbf{Q}] \mathbf{A} \mathbb{E}[\mathbf{Q}].$$

Instead, prove that, in the setting of Theorem 2.4, one has

$$\mathbf{Q}(z) \mathbf{A} \mathbf{Q}(z) \leftrightarrow m^2(z) \mathbf{A} + \frac{1}{n} \text{tr} \mathbf{A} \cdot \frac{m'(z) m^2(z)}{(1 + cm(z))^2} \mathbf{I}_p.$$

As a sanity check, using the fact that $\partial \mathbf{Q}(z) / \partial z = \mathbf{Q}^2(z)$ and taking $\mathbf{A} = \mathbf{I}_p$ in the equation above, confirm that

$$\mathbf{Q}^2(z) \leftrightarrow m'(z) \mathbf{I}_p$$

for $m'(z) = \frac{m^2(z)}{1 - \frac{cm^2(z)}{(1+cm(z))^2}}$ obtained from differentiating the Marčenko–Pastur equation (2.9).

2.9.6 Concentration of Measure

Exercise 15 (Concentration of matrix quadratic forms). *Recalling the definitions and notations of Section 2.7, let $\mathbf{X} \in \mathbb{R}^{p \times n}$ be a random matrix satisfying*

$$\mathbf{X} \infty C e^{c \cdot 2}, \text{ and } \|\mathbb{E}[\mathbf{X}]\| \leq K$$

for some $K, C, c > 0$. Given $\mathbf{A} \in \mathbb{R}^{p \times p}$ deterministic, we aim to prove the linear concentration of $\mathbf{X}^\top \mathbf{A} \mathbf{X}$ in $(\mathbb{R}^{n \times n}, \|\cdot\|_F)$. To this end, we consider a deterministic matrix

$\mathbf{B} \in \mathbb{R}^{n \times n}$ such that $\|\mathbf{B}\|_F \leq 1$ and study the behavior of $\text{tr}(\mathbf{B}\mathbf{X}^\top\mathbf{A}\mathbf{X})$. Consider first the singular value decomposition

$$\mathbf{A} = \mathbf{U}_A \Lambda_A \mathbf{V}_A^\top, \quad \mathbf{B} = \mathbf{U}_B \Lambda_B \mathbf{V}_B^\top,$$

with $\mathbf{U}_A, \mathbf{V}_A \in \mathbb{R}^{p \times p}$ and $\mathbf{U}_B, \mathbf{V}_B \in \mathbb{R}^{n \times n}$ orthogonal matrices, $\Lambda_A \in \mathbb{R}^{p \times p}$, $\Lambda_B \in \mathbb{R}^{n \times n}$ diagonal matrices, and define $\tilde{\mathbf{X}}_1 = \mathbf{U}_A^\top \mathbf{X} \mathbf{V}_B$, $\tilde{\mathbf{X}}_2 = \mathbf{V}_A^\top \mathbf{X} \mathbf{U}_B \in \mathbb{R}^{p \times n}$. In the sequel, the constants $K', C', c' > 0$ are understood only depending on K, C, c and might change from line to line.

First show that there exist $K', C', c' > 0$ such that, for $t > K' \sqrt{\log(np)}$ and $\tilde{\mathbf{X}} \in \{\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2\}$

$$\mathbb{P}(\|\tilde{\mathbf{X}} - \mathbb{E}[\tilde{\mathbf{X}}]\|_\infty \geq t) \leq C' e^{-c't^2/\log(np)}.$$

Deduce from the bound $\|\mathbb{E}[\mathbf{X}]\| \leq K$ that there exists a constant $K' > 0$ depending only on K, C, c such that

$$\mathbb{E}[\|\tilde{\mathbf{X}}\|_\infty] \leq K' \sqrt{\log(np)}.$$

This established, introduce the set $\mathcal{A}_\theta = \{\mathbf{X} \in \mathbb{R}^{p \times n}, \max\{\|\tilde{\mathbf{X}}_1\|_\infty, \|\tilde{\mathbf{X}}_2\|_\infty\} \leq \theta\} \subset \mathbb{R}^{p \times n}$ and show that for all $\theta \geq K' \sqrt{\log(np)}$ with $K' > 1$, we have

$$\mathbb{P}(\mathbf{X} \in \mathcal{A}_\theta^c) \leq C' e^{-c'\theta^2}$$

and that the mapping $\mathbf{X} \mapsto \text{tr}(\mathbf{B}\mathbf{X}^\top\mathbf{A}\mathbf{X})$ is $\theta\|\mathbf{A}\|_F$ -Lipschitz on \mathcal{A}_θ .

Introduce M , a median of $\text{tr}(\mathbf{B}\mathbf{X}^\top\mathbf{A}\mathbf{X})$, and note that

$$\mathbb{P}\left(\left|\text{tr}(\mathbf{B}\mathbf{X}^\top\mathbf{A}\mathbf{X}) - M\right| \geq t, \mathbf{X} \in \mathcal{A}_\theta\right) \leq C' e^{-c't^2/(\theta\|\mathbf{A}\|_F)^2}.$$

Conclude by carefully choosing the parameter $\theta \geq K' \sqrt{\log(np)}$ and showing that

$$\mathbf{X}^\top\mathbf{A}\mathbf{X} \in \mathbb{E}[\mathbf{X}^\top\mathbf{A}\mathbf{X}] \pm C' e^{-c'.2/(\log(np)\|\mathbf{A}\|_F^2)} + C' e^{-c'./\|\mathbf{A}\|_F}.$$

2.9.7 Beyond Matrices

Exercise 16 (Towards spiked models in random tensors). Let $\mathcal{Y} \in \mathbb{R}^{n \times n \times n}$ be a three-way symmetric tensor, i.e., such that $[\mathcal{Y}]_{ijk}$ is constant to exchanges of its indexes, defined by

$$\mathcal{Y} = \ell \mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x} + \frac{1}{\sqrt{n}} \mathcal{W}$$

where $\mathcal{W} \in \mathbb{R}^{n \times n \times n}$ has independent $\mathcal{N}(0,1)$ entries up to symmetry, deterministic $\mathbf{x} \in \mathbb{R}^n$ of unit norm, and $[\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}]_{ijk} = a_i b_j c_k$.

A possible definition of the ‘‘eigenvalue-eigenvector’’ pair $(\hat{\lambda}, \hat{\mathbf{u}})$ (without loss of generality such that $\hat{\lambda} \geq 0$ and $\|\hat{\mathbf{u}}\| = 1$) of a symmetric tensor \mathcal{Y} is the solution to Lim [2005]

$$\mathcal{Y} \cdot \hat{\mathbf{u}} \cdot \hat{\mathbf{u}} = \hat{\lambda} \hat{\mathbf{u}},$$

where $\mathcal{A} \cdot \mathbf{a} \cdot \mathbf{b} = \sum_{ij} [\mathcal{A}]_{ij} a_i b_j \in \mathbb{R}^n$ is the contraction of the tensor \mathcal{A} on the vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. The objective here is to characterize the (possible) spike $\hat{\lambda}$ as well as the associated eigenvector alignment $|\hat{\mathbf{u}}^T \mathbf{x}|$ between the dominant eigenvector and \mathbf{x} .

Show first that the matrix $\mathbf{Y}_x = \mathcal{Y} \cdot \mathbf{x} = \sum_{i=1}^n \mathcal{Y}_{i..} x_i \in \mathbb{R}^{n \times n}$ takes the form

$$\mathbf{Y}_x = \ell \mathbf{x} \mathbf{x}^T + \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \mathbf{W}_i,$$

where $\mathbf{W}_i \in \mathbb{R}^{n \times n}$ is the i th “layer” matrix of the tensor \mathcal{W} such that $[\mathbf{W}_i]_{ab} = \mathcal{W}_{iab}$.

Using the Gaussian method discussed in Section 2.2.2, show that the limiting spectral measure of \mathbf{Y}_x is the semicircle law supported on $[-2, 2]$ (we may discard the rank-one matrix $\ell \mathbf{x} \mathbf{x}^T$ to retrieve this result). Then, using a spiked model analysis as in Section 2.5, show that

- for all $\ell > 0$, there must exist an isolated eigenvalue $\hat{\lambda}_x$ of \mathbf{Y}_x (thus no phase transition) asymptotically equal to (with high probability)

$$\hat{\lambda}_x \rightarrow \lambda_x = \sqrt{\ell^2 + 4};$$

- the eigenvector $\hat{\mathbf{u}}_x$ associated with $\hat{\lambda}_x$ satisfies (again with high probability)

$$|\hat{\mathbf{u}}_x^T \mathbf{x}|^2 \rightarrow \frac{\ell}{\sqrt{\ell^2 + 4}}.$$

Conclude on an asymptotic upper bound for the quantity $\hat{\lambda} |\hat{\mathbf{u}}^T \mathbf{x}|$.