On spectral properties of large dimensional correlation matrices and covariance matrices computed from elliptically distributed data

Noureddine El Karoui *
Department of Statistics,
University of California, Berkeley

Abstract

We place ourselves in the setting of high-dimensional statistical inference, where the number of variables p in a dataset of interest is of the same order of magnitude as the number of observations n. More formally we study the asymptotic properties of correlation and covariance matrices under the setting that $p/n \to \rho \in (0, \infty)$, for general population covariance.

We show that spectral properties for large dimensional correlation matrices are similar to those of large dimensional covariance matrices, for a large class of models studied in random matrix theory.

We also derive a Marčenko-Pastur type system of equations for the limiting spectral distribution of covariance matrices computed from elliptically distributed data. The motivation for this study comes from the possible relevance of such distributional assumptions to problems in econometrics and portfolio optimization. From a theoretical standpoint, we show that our approach can be extended beyond elliptically distributed data to more general geometric frameworks.

A mathematical theme of the paper is the important use we make of concentration inequalities.

1 Introduction

It is increasingly common in multivariate Statistics to have to work with datasets where the number of variables, p, is of the same order of magnitude as the number of observations, n. When studying asymptotic properties of estimators in this setting, usually under the assumption that p/n has a finite limit, we often obtain convergence results that differ from those obtained under the classical assumptions that p is fixed and n goes to infinity.

This realization is not recent: the first paper in the area is probably Marčenko and Pastur (1967), where the authors studied the behavior of the eigenvalues of large dimensional sample covariance matrices, for diagonal population covariance matrices, and with some assumptions on the structure of the data. The surprising result they found was, in the case of i.i.d data, that the eigenvalues of the sample covariance matrix X^*X/n do not concentrate around 1, but rather were spread out on the interval $[(1-\sqrt{p/n})^2, (1+\sqrt{p/n})^2]$, when $p \leq n$. Moreover their distribution is asymptotically non-random. We note that this seminal paper is much richer than just described, and refer the reader there for more details.

Since this result there has been a flurry of activity, especially in recent years, concerning the behavior of the largest eigenvalue of sample covariance matrices (Geman (1980), Yin et al. (1988)), their fluctuation behavior in the null case (Forrester (1993), Johansson (2000), Johnstone (2001), El Karoui (2003)) and under alternatives (Baik et al. (2005), El Karoui (2007), Paul (2007)), as well as fluctuation results for linear spectral statistics of those matrices (Jonsson (1982), Bai and Silverstein (2004), Anderson and Zeitouni (2006)). Even more recently, some of these results have started to be used to develop better estimators of these large dimensional covariance matrices (Burda et al. (2004), El Karoui (2006) and Rao et al. (2007)).

*Support from NSF grant DMS-0605169 and hospitality and support from SAMSI in the Fall of 2006 are gratefully acknowledged. I would like to thank James Mingo and Jack Silverstein for an interesting conversation while there. **AMS 2000 SC:** 62H10. **Key words and Phrases :** covariance matrices, correlation matrices, eigenvalues of covariance matrices, multivariate statistical analysis, high-dimensional inference, random matrix theory, elliptical distributions. **Contact :** `nkaroui@stat.berkeley.edu`

We note also that from a statistical point of view, other approaches to estimation using regularization have been taken, with sometime striking results (Bickel and Levina (2007), Ledoit and Wolf (2004)).

As noted above, the random matrix results in question concern somewhat exclusively sample covariance matrices. However, in practice, sample correlation matrices are often used, for instance for Principal Component Analysis (PCA). A question we were asked several times by practitioners is how much of the random matrix results would hold if one were concerned with correlation matrices and not covariance matrices. Part of the answer is already known from a paper due to Jiang (2004), where he considered the case of i.i.d data. The answer was that spectral distribution results as well as a.s convergence of extreme eigenvalue results held in this situation. However, in practice, the assumption of i.i.d data is not very reasonable, and in most cases, practitioners would actually hope to be in the presence of an interesting covariance structure, away from the no-information case represented by the identity covariance matrix. In this paper we tackle the case where the population covariance is not Id_p, and show that classic random matrix results hold then too, with the population covariance matrix replaced by the population correlation matrix. This means that recently developed methods that make use of random matrix theory to better estimate the eigenvalues of population covariance matrices can also be used to estimate the spectrum of population correlation matrices.

As explained below, such results can be shown for Gaussian and some non-Gaussian data. A natural question is therefore to wonder how robust to these distributional assumptions the results are. In particular, a recent paper (Frahm and Jaekel (2005)) and a recent monograph (McNeil et al. (2005)) make an interesting case for modeling financial data through elliptically distributed data. As explained in Frahm and Jaekel (2005) and McNeil et al. (2005), this has to do with certain tail-dependence properties that are absent from Gaussian data and present in certain class of elliptically distributed data. We show in the second part of the paper that for elliptically distributed data, the spectrum of the sample covariance matrix is asymptotically non-random, and we characterize it through the use of Stieltjes transforms. In particular, it shows that the Marčenko-Pastur equation is not robust to deviation from the "Gaussian+" model usually considered in random matrix theory. The result explains some of the numerical results obtained by Frahm and Jaekel (2005). From a more theoretical standpoint, our approach allows us to break away from models for data where some independence between entries of a (observed or unobserved) data vector is required. Rather, what we need are concentration properties for convex 1-Lipschitz (with respect to Euclidian norm) functionals of these data vectors. Hence, our approach will show that some results in random matrix theory hold in wider generality than was previously known.

As it turns out, a central element of the proofs to be presented are the concentration properties of certain quadratic forms. We make use below of a number of concentration inequalities, recent and less recent. The usefulness of these inequalities in random matrix theory has already been illustrated, in another context than what we develop below, in Guionnet and Zeitouni (2000). A very good reference on the topic of concentration is Ledoux (2001).

2 On correlation matrices

We now turn to our study of correlation matrices. The main result is Theorem 1, which says that under the model considered there - related to the classical one in random matrix theory - results concerning the spectral distribution and the largest eigenvalue pass without much modifications from the sample covariance matrix to the sample correlation matrix.

Before we proceed, we need to set some notations. In the rest of the paper, we call $\mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Im}\,[z] > 0\}$.

2.1 A simple lemma

In what follows, we use $|||\cdot|||_2$ to denote the spectral norm of a matrix.

Lemma 1. *Suppose that M_p is a $p \times p$ Hermitian random matrix, whose spectral characteristics (spectral distribution F_p or largest eigenvalue $\lambda_1(M_p)$) converge a.s to a limit, and whose spectral norm is (a.s) bounded as $p \to \infty$. Suppose that D_p is a $p \times p$ diagonal matrix and that $|||D_p - \mathrm{Id}_p|||_2 \to 0$ a.s. Then the spectral characteristics of $D_p M_p D_p$ or $D_p^{-1} M_p D_p^{-1}$ have the same limit as that of M_p.*

Proof. The assumption $|||D_p - \mathrm{Id}_p|||_2 \to 0$ implies that for p large enough, D_p is invertible. Now

$$|||M_p - D_p M_p D_p|||_2 = |||M_p - M_p D_p + M_p D_p - D_p M_p D_p|||_2$$
$$\leq |||M_p|||_2 |||D_p - \mathrm{Id}_p|||_2 + |||D_p - \mathrm{Id}_p|||_2 |||D_p|||_2 |||M_p|||_2 \to 0 \text{ a.s}$$

Using Weyl's inequality (see Bhatia (1997), Corollary III.2.6), i.e the fact that for Hermitian matrices A and B and any i, if λ_i denotes the i-th eigenvalue of A, ordered in decreasing order, $|\lambda_i(A)-\lambda_i(B)| \leq |||A-B|||_2$, we conclude that for any fixed k

$$\max_k |\lambda_k(M_p) - \lambda_k(D_p M_p D_p)| \to 0 \text{ a.s} \quad .$$

Now because $|||M_p|||_2$ is bounded a.s, the two sequences are a.s asymptotically distributed (see Grenander and Szegö (1958), p.62, or Gray (2002)). Therefore, if $F_p(M_p)$ weakly converges to F, then $F_p(D_p M_p D_p)$ also converges to F.

Now if $|||D_p - \mathrm{Id}_p|||_2 \to 0$, then $|||D_p^{-1} - \mathrm{Id}_p|||_2 \to 0$, too. So the same results hold when we replace D_p by D_p^{-1}. □

The previous lemma is helpful in our context thanks to the following elementary fact, which is standard in multivariate Statistics.

Fact 1 (Correlation matrix as function of covariance matrix)**.** *Call C_p the correlation matrix of our data and S_p the covariance matrix of the data. Call $D_p(S_p)$ the diagonal matrix consisting of the diagonal of S_p. Then we have:*

$$C_p = [D_p(S_p)]^{-1/2}\, S_p\, [D_p(S_p)]^{-1/2} \quad .$$

Proof. This is just a simple consequence of the fact if D is a diagonal matrix

$$(DHD)_{i,j} = d_{i,i} H_{i,j} d_{j,j} \ .$$

Note that $C_p(i,j) = S_p(i,j)/\sqrt{S_p(i,i)S_p(j,j)}$ and the assertion follows. □

The consequence of the previous remark is that we will deduce the asymptotic spectral properties of correlation matrices from that of covariance matrices by simply showing convergence of the diagonal of S_p (or a scaled version of it) to Id_p in operator norm.

We are now ready to state the main theorem of this section.

Theorem 1. *Suppose X is $n \times p$ matrix of i.i.d random variables with mean 0, variance 1. Assume further that $\mathbf{E}\left(|X_{i,j}|^4 (\log(|X_{i,j}|))^{2+2\epsilon}\right) < \infty$. Suppose Σ_p is a $p \times p$ covariance matrix and call Γ_p the corresponding correlation matrix. Assume that $|||\Gamma_p|||_2 < K$, for all p. Call $Y_1 = X\Sigma_p^{1/2}$ and $Y = X\Gamma_p^{1/2}$. Then the spectral properties of corr(Y_1) are the same as the spectral properties of $\Gamma_p^{1/2}(X-\bar{X})'(X-\bar{X})\Gamma_p^{1/2}$ $= (Y-\bar{Y})'(Y-\bar{Y})$.*

In particular, the Stieltjes transform of the limiting spectral distribution of corr(Y_1) satisfies the Marčenko-Pastur equation, with parameter the spectral distribution of Γ_p: namely, if H_p, the spectral distribution of Γ_p has a.s a limit H, if p/n has a finite limit ρ, and if m_n is the Stieltjes transform of corr(Y_1), we have, calling $w_n = -(1-p/n)/z + (p/n)m_n(z)$,

$$w_n(z) \to w(z) \ \text{a.s} \ , \ \text{which satisfies} \ -\frac{1}{w(z)} = z - \rho \int \frac{\lambda dH(\lambda)}{1+\lambda w(z)} \ ,$$

and w is the unique function mapping $\mathbb{C}^+$ into $\mathbb{C}^+$ to satisfy this equation.

Also, if the norm of $\Gamma_p^{1/2}(X-\bar{X})'(X-\bar{X})\Gamma_p^{1/2}$ has a limit, the norm of corr(Y) has the same limit.

This theorem is related to that of Jiang (2004), which was concerned with $\Gamma_p = \mathrm{Id}_p$, which would amount to doing multivariate analysis with i.i.d variables, an assumption that for obvious statistical reasons, practitioners are not willing to make. By contrast, here we are able to handle general covariance structures, assuming that the spectral norm of Γ_p is bounded. However, Jiang (2004) required only 4 moments and we require a bit more. We explain in subsubsection 2.2.2 why it is so.

We note that the proof can actually handle cases where $|||\Gamma_p|||_2$ grows slowly with p. We refer the reader to Silverstein (1995) for more information on the Marčenko-Pastur equation in the case of covariance matrices, and an important strengthening of the result of Marčenko and Pastur (1967).

Recent progress has led to fairly explicit characterization of the norm of large dimensional sample covariance matrices - a fact that makes these results potentially useful in, among other fields, statistics. In particular, the following consequence for the norm of the correlation matrix can be drawn from the recent El Karoui (2007), specifically Fact 2 there (which is partly a consequence of a deep result in Bai and Silverstein (1998)):

Corollary 1. *Under the assumptions of Theorem 1, we have: if $\lambda_1(\Gamma_p)$ tends to the endpoint of the support of H, and the model $\{\Gamma_p, n, p\}$ is in the class $\mathcal{G}$ defined in El Karoui (2007), then*

$$|||corr(Y_1)|||_2 - \mu_{n,p} \to 0 \ \ a.s \ \ ,$$

where

$$\begin{aligned}\mu_{n,p} &= \frac{1}{c_0}\left(1 + \frac{p}{n}\int \frac{\lambda c_0}{1-\lambda c_0} dH(\lambda)\right)\\ \frac{n}{p} &= \int \left(\frac{\lambda c_0}{1-\lambda c_0}\right)^2 dH(\lambda)\ , c_0 \in [0, 1/\lambda_1(\Gamma_p))\ .\end{aligned}$$

2.2 Proof of Theorem 1

The proof is in three steps. The first one consists in showing that we need to focus only on the matrix $Y'Y$ (or $(Y-\bar{Y})'(Y-\bar{Y})$). Then we need a truncation and centralization step for the entries of X. And finally, we use concentration of measures result to show that indeed the diagonal of the corresponding covariance matrix converges in operator norm to the identity.

2.2.1 Replacing Σ_p by Γ_p

Since the correlation coefficient is invariant under shifting and scaling random variables, we see that for any diagonal matrix D,

$$\text{corr}Y = \text{corr}YD\ ,$$

since $(YD)_{i,j} = Y_{i,j}d_{jj}$. In particular, for D we can use $(\text{diag}(\Sigma_p))^{-1/2}$. After this adjustment, our data matrix takes the form XG, where $G = \Sigma_p^{1/2}(\text{diag}(\Sigma_p))^{-1/2} = \Sigma_p^{1/2}D$, and $G'G = \Gamma_p$. Note in particular that since Γ_p is a correlation matrix, its diagonal is full of 1-s. Because G is not symmetric, it is in general not equal to $\sqrt{\Gamma_p}$. However, G is similar to $D^{1/2}\Sigma_p^{1/2}D^{1/2}$, so all its eigenvalues are real and non-negative. Further, because $G'G = \Gamma_p$, the eigenvalues of G are equal to the square root of the eigenvalues of Γ_p.

Because $\Sigma_p^{1/2}$ and D are invertible, so is $\Sigma_p^{1/2}D$. Therefore the spectrum of the matrix of interest is the same as the spectrum of $X'X\Sigma_p^{1/2}D^2\Sigma_p^{1/2}$. Even though in general $\Sigma_p^{1/2}D^2\Sigma_p^{1/2} \neq \Gamma_p$, they have the same eigenvalues. Because the Marčenko-Pastur equation involves only the eigenvalues of the deterministic matrix at stake, the limiting spectral distribution of $\text{corr}(Y)$ is the same as the limiting spectral distribution of $\Gamma_p^{1/2}X'X\Gamma_p^{1/2}$. A similar conclusion applies to individual eigenvalues.

2.2.2 Truncation and centralization step

In this part, we show that we can truncate the entries of X at level $\sqrt{n}/(\log n)^{(1+\epsilon)/2} = \sqrt{n}/\delta_n$ and not change the value of $\text{corr}Y$, at least for p large enough. The same holds when the truncated values are then recentered. The conclusion of this subsection is that it is enough to study matrices X whose entries are i.i.d mean 0 and are bounded in absolute value by $C\sqrt{n}/(\log n)^{(1+\epsilon)/2}$.

The proof is similar to the argument given for the proof of Lemma 2.2 in Yin et al. (1988). However, because the term $1/(\log n)^{(1+\epsilon)/2}$ is crucial in our later arguments, and the author of Yin et al. (1988) gloss over the details of their choice of δ_n, we feel a full argument is needed to give a convincing proof, though we do not claim the arguments are new. This is where we need a slightly stronger assumption that just the finite 4-th moment assumption made in Yin et al. (1988). (Our problem is with Remark 1 in Yin et al. (1988), which is not properly justified. There also appears to be counter-examples to this claim. However, it does not seem that (the full-strength of) this remark is ever really used in that paper, and the rest of the arguments are clear.)

We have the following lemma, which follows closely Lemma 2.2 in Yin et al. (1988).

Lemma 2 (Truncation). *Let X be an infinite double array of identically distributed (i.d) random variables. Suppose X_n is an $n \times p$ matrix of i.d random variables, with mean 0, variance 1, whose entries satisfy $\mathbf{E}\left(|X_{i,j}|^4(\log(|X_{i,j}|))^{2+2\epsilon}\right) < \infty$, corresponding to the upper left corner of X. Suppose that p/n has a finite limit ρ. Call T_n the matrix with (i,j)-th entry $x_{i,j}1_{|x_{i,j}|<\sqrt{n}/(\log n)^{(1+\epsilon)/2}}$. Then*

$$P(X_n \neq T_n \ i.o\ ) = 0 \ .$$

Proof. Because of the moment assumption made on $X_n(i,j)$, we have, if we call $f_\epsilon(x) = x^4(\log x)^{2(1+\epsilon)}$,

$$\int_0^\infty f'_\epsilon(y)P(Y>y)dy = \sum_{m=0}^{\infty}\int_{u_m}^{u_{m+1}} f'_\epsilon(y)P(Y>y)dy < \infty \ ,$$

for any increasing sequence $\{u_m\}_{m=0}^\infty$, with $u_0 = 0$ and $u_m \to \infty$ as $m \to \infty$. Now, when y is large enough, $f'_\epsilon(y) \geq 0$, so

$$\int_{u_m}^{u_{m+1}} f'_\epsilon(y)P(Y>y)dy \geq P(Y > u_{m+1})(f_\epsilon(u_{m+1}) - f_\epsilon(u_m)) \ .$$

Call $\gamma_m = 2^m$ and $u_m = \sqrt{\gamma_m/(\log\gamma_m)^{1+\epsilon}}$. Note that u_m is increasing for sufficiently large. Elementary computations show that, as m tends to ∞, $u_m^4(\log u_m)^{2+2\epsilon} \sim 2^{2m-(2+2\epsilon)}$. Consequently, $f_\epsilon(u_{m+1}) - f_\epsilon(u_m) \sim 3\times 2^{2(m-1)}$. Now note that our moments requirement therefore imply that

$$\sum_{m=1}^{\infty} 2^{2m}P(Y > u_m) < \infty \ .$$

Now, for n satisfying $\gamma_{m-1} \leq n < \gamma_m$, we threshold $X_n(i,j)$ at level u_{m-1}. (In what follows, $2\rho\gamma_m$ should be replaced by the smallest integer greater than this number, but to avoid cumbersome notations, we do not stress this particular fact.)

$$\begin{aligned}
P(X_n \neq T_n \text{ i.o }) &\leq \sum_{m=k}^{\infty} P\Big(\bigcup_{\gamma_{m-1}2^{m-1}\leq n<\gamma_m 2^m} \bigcup_{i=1}^{n}\bigcup_{j=1}^{p}(|X_n(i,j)| > u_{m-1})\Big)\\
&\leq \sum_{m=k}^{\infty} P\Big(\bigcup_{\gamma_{m-1}2^{m-1}\leq n<\gamma_m 2^m} \bigcup_{i=1}^{\gamma_m 2^m}\bigcup_{j=1}^{2\rho\gamma_m 2^m}(|X_n(i,j)| > u_{m-1})\Big)\\
&= \sum_{m=k}^{\infty} P\Big(\bigcup_{i=1}^{\gamma_m 2^m}\bigcup_{j=1}^{2\rho\gamma_m 2^m}(|X_n(i,j)| > u_{m-1})\Big)\\
&\leq 2\rho\sum_{m=k}^{\infty} 2^{2m}P(Y > u_{m-1}) = \frac{\rho}{2}\sum_{m=k}^{\infty} 2^{2(m-1)}P(Y > u_{m-1}) \ .
\end{aligned}$$

The right hand side tends to 0 when k tends to infinity and the left hand side is independent of k. We conclude that

$$P(X_n \neq T_n \text{ i.o }) = 0 \ .$$

□

Lemma 3 (Centralization). *Call TC_n the matrix with entries $TC_n(i,j) = T_n(i,j) - ET_n(i,j)$. Then*

$$\frac{1}{n}|||T'_nT_n - TC'_nTC_n|||_2 \to 0 \ \ a.s \ \ .$$

Proof. The proof would be a simple repetition of the arguments in Lemma 2.3 of Yin et al. (1988), with $r = 1/2$ and $\delta = (\log n)^{-(1+\epsilon)/2}$ in the notation of their papers, so we skip it. Note that their proof finds a bound on the spectral norm of $T_n' T_n - TC_n' TC_n$. $\square$

We note that since we are dealing with correlation matrices, and those are invariant under recentering of the columns of Y, we can assume without loss of generality that we work with TC_n instead of T_n; i.e we are working with centered random variables that are bounded by $C\sqrt{n/(\log n)^{1+\epsilon}}$. Note also that their variances tend to 1.

What the centralization lemma 3 guarantees is that the spectral characteristics of $\Gamma^{1/2} TC_n' TC_n \Gamma^{1/2}$ are asymptotically the same as those of $\Gamma^{1/2} T_n' T_n \Gamma^{1/2}$, and hence those of $\Gamma^{1/2} X_n' X_n \Gamma^{1/2}$ by the truncation lemma.

2.2.3 Controlling the diagonal in operator norm

Now that we have seen that a.s we can assume that the entries of X are bounded by $C\sqrt{n}/(\log n)^{(1+\epsilon)/2}$, we turn our attention to showing that the diagonal of $G'X'XG$ is close to 1. We assume wlog that $C \leq 2$, which is true since $E(T_n(i,i)) \to E(X_n(i,i)) = 0$.

Lemma 4 (Mean 0 Gaussian MLE situation). *Here we focus on $S_p = \frac{1}{n} Y'Y = \frac{1}{n} G'X'XG$, a quantity often studied in random matrix theory.*

When $p \asymp n$, we have

$$\max_i \left| \sqrt{S_p(i,i)} - 1 \right| \to 0 \ a.s$$

Proof. We call v_i the ith column of G. Denoting $M = X'X/n$, we note that

$$S_p(i,i) = v_i' M v_i = \|X v_i / \sqrt{n}\|_2^2 \ .$$

Now consider the function f from $\mathbb{R}^{np}$ to $\mathbb{R}$ defined by turning the vector x into the matrix X, by filling first the rows of X and then computing the Euclidian norm of Xv_i. In other words,

$$f(x) = \|X v_i\|_2 \ .$$

This function is clearly convex, and 1-Lipschitz with respect to Euclidian norm. As a matter of fact, for $\theta \in [0,1]$, and $x, z \in \mathbb{R}^{np}$,

$$f(\theta x + (1-\theta) z) = \|(\theta X + (1-\theta) Z) v_i\|_2 \leq \|\theta X v_i\|_2 + \|(1-\theta) Z v_i\|_2 = \theta f(x) + (1-\theta) f(z) \ .$$

Similarly,

$$|f(x) - f(z)| = |\|X v_i\|_2 - \|Z v_i\|_2| \leq \|(X - Z) v_i\|_2 \leq \|X - Z\|_F \|v_i\|_2 = \|x - z\|_2 \ ,$$

using the Cauchy-Schwarz's inequality and the fact that $\|v_i\|_2 = (G'G)(i,i) = \Gamma(i,i) = 1$. Because the X_{ij} are i.i.d we can apply recent results concerning concentration of measure of convex Lipschitz functions. In particular, from Corollary 4.10 in Ledoux (2001) (a consequence of Talagrand's inequality, Theorem 4.6 in Ledoux (2001)) we see that for any $r > 0$, we have

$$P(|f(X) - m_f| \geq r) \leq 4 \exp(-r^2 / (16 C^2 n / (\log n)^{(1+\epsilon)})) \ ,$$

where m_f is a median of $f(X)$. In particular, since $\sqrt{S_p(i,i)} = f(X)/\sqrt{n}$, we see that

$$P(|\sqrt{S_p(i,i)} - m_{i,i}| \geq r) \leq 4 \exp(-r^2 (\log n)^{(1+\epsilon)} / 16 C^2) \ .$$

Finally,

$$P([\max_i |\sqrt{S_p(i,i)} - m_{i,i}|] \geq r) \leq 4p \exp(-r^2 (\log n)^{(1+\epsilon)} / (16 C^2)) \ ,$$

so, since $p \asymp n$, using the first Borel-Cantelli lemma (see Durrett (1996), p. 47), we see that

$$\max_i |\sqrt{S_p(i,i)} - m_{i,i}| \to 0 \text{ a.s}$$

All we have to do now is show that the $m_{i,i}$ are all close to 1. We call $\upsilon_n = \text{var}\,(TC_n(i,j))$. Note that υ_n is independent of i, j and $\upsilon_n \to 1$, as $n \to \infty$.

Since we have Gaussian concentration, using Proposition 1.9 in Ledoux (2001), we have

$$|E(\sqrt{S_p(i,i)} - m_{i,i}| \leq 8C\sqrt{\pi}(\log n)^{-(1+\epsilon)/2}$$

and, since $E(S_p(i,i)) = \|v_i\|_2^2 \upsilon_n = \Gamma(i,i)\upsilon_n = \upsilon_n$,

$$0 \leq \upsilon_n - E(\sqrt{S_p(i,i)})^2 \leq \frac{64C^2}{(\log n)^{(1+\epsilon)}}$$

Consequently,

$$-\frac{8C\sqrt{\pi}}{(\log n)^{(1+\epsilon)/2}} + \sqrt{\upsilon_n - \frac{64C^2}{(\log n)^{(1+\epsilon)}}} \leq m_{i,i} \leq \upsilon_n + \frac{8C\sqrt{\pi}}{(\log n)^{(1+\epsilon)/2}} .$$

Therefore, $\max|m_{i,i} - 1| = \mathrm{O}(\max(|1-\upsilon_n|, (\log n)^{-(1+\epsilon)/2}))$ and we have

$$\max_i |\sqrt{S_p(i,i)} - 1| \to 0 \text{ a.s}$$

□

We now turn to the more interesting situation of a covariance matrix.

Lemma 5 (Covariance matrix)**.** *We now focus on the matrix*

$$S_p = \frac{1}{n-1}(Y - \bar{Y})'(Y - \bar{Y}) .$$

For this matrix, we also have

$$\max_{1 \leq i \leq p} |\sqrt{S_p(i,i)} - 1| \to 0 \quad a.s \quad .$$

Proof. Note that $Y - \bar{Y} = (\text{Id}_n - \frac{1}{n}11')Y = (\text{Id}_n - \frac{1}{n}11')XG$. Now $S_p(i,i) = v_i'X'(\text{Id}_n - \frac{1}{n}11')Xv_i/(n-1)$, so the same strategy as above can be employed, with f now defined as

$$f(x) = f(X) = \|(\text{Id}_n - \frac{1}{n}11')Xv_i\|_2 .$$

This function is again a convex 1-Lipschitz function of x. Convexity is a simple consequence of the fact that norms are convex; the Lipschitz coefficient is equal to $\|v_i\|_2 |||\text{Id}_n - \frac{1}{n}11'|||_2$. The eigenvalues of this matrix are $(n-1)$ 1s and one zero. So its operator norm is 1. We therefore have Gaussian concentration. All we need to check to conclude is that $E(S_p(i,i)) \to 1$. By renormalizing by $1/\sqrt{n-1}$, we ensure that $E(S_p(i,i)) = \upsilon_n$, and so we can conclude as before. □

2.2.4 A remark on $|||(X - \bar{X})'(X - \bar{X})/n - 1|||_2$

Finally, we note that in the literature, most results concerning the spectral norm of covariance matrices are dealing only with the case of the mean 0 Gaussian MLE, namely the matrix $X'X/n$. Since in practice, $(X - \bar{X})'(X - \bar{X})/(n-1)$ is almost always used, it is of interest to know what happens for this matrix. Note that the spectral norm of the difference between these two matrices goes to ρ, as n and p goes to infinity, so coarse bounding of this type will not be enough to find the limiting behavior of the quantity we are interested in.

However, calling $H_n = \mathrm{Id}_n - \frac{1}{n}11'$, we see that

$$X - \bar{X} = H_n X \ .$$

Therefore, since σ_1, the largest singular value, is matrix norm, we have

$$\sigma_1(X - \bar{X})/\sqrt{n} \leq \sigma_1(H_n)\sigma_1(X/\sqrt{n}) = \sigma_1(X/\sqrt{n}) \ ,$$

since H_n is a symmetric matrix with $(n-1)$ eigenvalues equal to 1 and 1 eigenvalue equal to 0.

Now because $X'X/n$ and $(X-\bar{X})'(X-\bar{X})/n-1$ have asymptotically the same spectral distribution, calling l_1 the right endpoint of this limit (if it exists), we conclude that

$$\liminf \sigma_1^2((X-\bar{X})/\sqrt{n-1}) \geq l_1 \ .$$

Hence, when $|||X'X/n|||_2 \to l_1$, we also have

$$|||(X-\bar{X})'(X-\bar{X})/(n-1)|||_2 \to l_1 \ .$$

This justifies the assertion made in Corollary 1, and more generally, the fact that when the norm of a non-recentered sample covariance matrix convergence to the right endpoint of the support of its limiting spectral distribution, so does the norm of the centered sample covariance matrix.

3 Elliptically distributed data

We now turn our attention to the problem of finding a Marčenko-Pastur -type system of equations for the limiting spectral distribution of sample covariance matrices computed, first, from elliptically distributed data, and then from more general distributions. Our aim in doing so is to explain the lack of robustness in high-dimension of this estimate of scatter, and to explain some of the numerical findings highlighted in Frahm and Jaekel (2005). We refer to this paper and to the book McNeil et al. (2005) for interesting discussions of the potential relevance of elliptical distributions to problems arising in the analysis of financial data. However, let us mention at least two properties that make them appealing. The first is the tail-dependence properties that they induce between components of our data vector, a property that in practice is found in financial data and cannot be accounted for by say multivariate Gaussian data. Second, at least some of these distributions allow for a certain amount of heavy-tailed observations. This is often mentioned as an important feature in financial data modeling. By contrast, it is sometimes advocated in the random matrix community that matrices with say i.i.d heavy-tailed entries should be studied as models for those financial data. We find that these models suffer at least from one deep flaw: in the case of a crash, many companies or stocks suffer on the same day, and a model of i.i.d heavy-tailed entries does not account for this. Besides the particulars of different models, what is also important to notice is that the limiting spectra will be drastically different and the behavior of extreme eigenvalues is also very likely to be so. Before we return to our study, we refer the reader to Anderson (2003) and Fang et al. (1990) for thorough introductions to elliptically distributed data.

We will assume that we observe n i.i.d observations of an elliptically distributed vector v in $\mathbb{R}^d$. Specifically, v can be written as

$$v = \mu + \lambda \Gamma r \ ,$$

where μ is a deterministic d-dimensional vector, λ is a real-valued random variable, r is uniformly distributed on the unit sphere in $\mathbb{R}^p$ (i.e $\|r\|_2 = 1$) and Γ is a $d \times p$ matrix. We call $\Sigma = \Gamma\Gamma'$. Here Σ, a $d \times d$ matrix, is assumed to be deterministic, and λ and r are independent. We call the corresponding data matrix X, which is $n \times d$, i.e the vectors of observations are stacked horizontally in this matrix. We will assume below that n/p and d/p have finite limits.

Note that without loss of generality, we can assume that $\mu = 0$, and deal with the corresponding $X'X$ matrix, because $(X-\bar{X})'(X-\bar{X})$ or $(X-\mu)'(X-\mu)$ have asymptotically the same spectral distribution as $X'X$, because the difference between the matrices is a rank 1 matrix. In what follows, we will therefore assume that

$$v_i = \lambda_i \Gamma r_i \ , i = 1, \ldots, n \ .$$

As is now classical, we will obtain our main result on this question (Theorem 2) by making use of Stieltjes transform arguments. If needed, we refer the reader to Geronimo and Hill (2003) for background on the connection between weak convergence of distributions and pointwise convergence of Stieltjes transforms.

We note that our model basically falls into the class of covariance matrices of the type $T_p^{1/2} X_{n,p}^* L_n X_{n,p} T_p^{1/2}$, where $X_{n,p}$ is a random matrix, independent of the square matrices T_p ($p \times p$) and L_n ($n \times n$), which can also be assumed to be random, as long as their spectral distribution converge to a limit. These matrices have been the subject of investigations already: see Tulino and Verdú (2004) Theorem 2.43, which refers to Boutet de Monvel et al. (1996), Li et al. (2004) and Girko (1990) and the very recent Paul and Silverstein (2007), which refers to Burda et al. (2005) and to Zhang (2006) for systems of equations involving Stieltjes transforms similar to the one we will derive. We note that under some restrictions, methods of free probability using the S-transform (see Voiculescu (2000)) could be used to derive a characterization of the limit.

However in all these papers, the entries of $X_{n,p}$ are assumed to be independent. Naturally, this is not the case in the situation we are considering, since the vectors r_i all have norm 1. (We note that the original Marčenko and Pastur (1967) allowed for dependence, too, and one of our questions was to know if one could recover (and generalize) those results from a different angle than the one taken in Marčenko and Pastur (1967).) Also, our matrix Γ is $d \times p$, and usually only square matrices are considered. Our aim here is to show that independence in $X_{n,p}$ is not the key element, rather we will rely on the fact the rows of $X_{n,p}$ are independent, and that the distribution of the corresponding vectors satisfy certain concentration properties. As our proof will make clear, using the "rank 1 perturbation" method originally proposed in Silverstein and Bai (1995) and Silverstein (1995), proofs of convergence of spectra of random matrices basically boil down to concentration of certain quadratic forms and concentration of Stieltjes transforms, the latter being easily achieved using Azuma's inequality. We discuss this in more detail in subsection 3.3, and propose some extensions of Theorem 2 there, in particular in situations where the random vectors of interest cannot be broken into independent parts. As far as we know, some of these results are new and cannot be achieved with other methods involving (one way or another) moment computations.

Finally, we outline in Remark 4 a possible strategy for deriving Theorem 2 from known results, using certain properties of vectors sampled uniformly at random on the 1-sphere. While that would give us the result we want for elliptically distributed data, it would not be as generalizable and reach as wide results as our approach will. Also, one of our points is really that the importance of concentration inequalities in this context appears to not have been realized and they permit generalizations of random matrix results to problems that look intractable by other methods.

3.1 A preliminary lemma of independent interest

We show a result of independent interest, namely the fact that the Stieltjes transform of a matrix which is the sum of n independent rank 1 matrices is asymptotically equivalent to a deterministic function. We have a bit more than this: we show concentration around its mean, which also gives us immediately some lower bounds on the rate of convergence.

Lemma 6 (Concentration of Stieltjes transforms)**.** *Suppose M is a $p \times p$ matrix, with*

$$M = \sum_{i=1}^{n} r_i r_i^* \ ,$$

where r_i are independent random vectors in $\mathbb{R}^p$. Call

$$m_p(z) = \frac{1}{p} trace\left((M - z\mathrm{Id}_p)^{-1}\right) \ .$$

Then, if $Im[z] = v$,

$$P(|m_p(z) - \mathbf{E}\left(m_p(z)\right)| > r) \leq 4\exp(-r^2 p^2 v^2/(16n)) \ .$$

Note that the lemma makes no assumptions whatsoever about the structure of the vectors $\{r_i\}_{i=1}^n$, beside the fact that they are independent.

Proof. We call $M_k = M - r_k r_k^*$. We call $\mathcal{F}_i$ the filtration generated by $\{r_l\}_{l=1}^i$. The first classical remark (see Bai (1999) p. 649, but note that the equation after (3.16) there contains a spurious $1/n$, which is problematic for the rest of the argument) is to write the random variable of interest as sum of martingale differences:

$$m_p(z) - \mathbf{E}\left(m_p(z)\right) = \sum_{k=1}^{n} \mathbf{E}\left(m_p(z)|\mathcal{F}_k\right) - \mathbf{E}\left(m_p(z)|\mathcal{F}_{k-1}\right) \ .$$

Now we note that $\mathbf{E}\left(\text{trace}\left((M_k - z\text{Id}_p)^{-1}\right)|\mathcal{F}_k\right) = \mathbf{E}\left(\text{trace}\left((M_k - z\text{Id}_p)^{-1}\right)|\mathcal{F}_{k-1}\right)$. So

$$\begin{aligned}
|\mathbf{E}\left(m_p(z)|\mathcal{F}_k\right) - \mathbf{E}\left(m_p(z)|\mathcal{F}_{k-1}\right)| &= \Bigg|\mathbf{E}\left(m_p(z)|\mathcal{F}_k\right) - \mathbf{E}\left(\frac{1}{p}\text{trace}\left((M_k - z\text{Id}_p)^{-1}\right)|\mathcal{F}_k\right) \\
&\quad + \mathbf{E}\left(\frac{1}{p}\text{trace}\left((M_k - z\text{Id}_p)^{-1}\right)|\mathcal{F}_{k-1}\right) - \mathbf{E}\left(m_p(z)|\mathcal{F}_{k-1}\right)\Bigg| \\
&\leq \left|\mathbf{E}\left(m_p(z) - \frac{1}{p}\text{trace}\left((M_k - z\text{Id}_p)^{-1}\right)|\mathcal{F}_k\right)\right| \\
&\quad + \left|\mathbf{E}\left(m_p(z) - \frac{1}{p}\text{trace}\left((M_k - z\text{Id}_p)^{-1}\right)|\mathcal{F}_{k-1}\right)\right| \\
&\leq \frac{2}{pv} \ ,
\end{aligned}$$

the last inequality following from Silverstein and Bai (1995), Lemma 2.6. So $m_p(z) - \mathbf{E}\left(m_p(z)\right)$ is a sum of bounded martingale differences. Note that the same would be true for its real and imaginary parts. For both of them we can apply Azuma's inequality (see Ledoux (2001), Lemma 4.1), to get that

$$P(|\text{Re}\left[m_p(z) - \mathbf{E}\left(m_p(z)\right)\right]| > r) \leq 2\exp(-r^2 p^2 v^2/(8n)) \ ,$$

and similarly for its imaginary part. We therefore conclude that

$$\begin{aligned}
P(|m_p(z) - \mathbf{E}\left(m_p(z)\right)| > r) &\leq P(|\text{Re}\left[m_p(z) - \mathbf{E}\left(m_p(z)\right)\right]| > r/\sqrt{2}) + P(|\text{Im}\left[m_p(z) - \mathbf{E}\left(m_p(z)\right)\right]| > r/\sqrt{2}) \\
&\leq 4\exp(-r^2 p^2 v^2/(16n)) \ .
\end{aligned}$$

□

We have the following immediate corollary.

Corollary 2. *Suppose we consider the following sequence of random matrices: for each p, pick n independent p dimensional vectors. Call $M = \sum_{i=1}^n r_i r_i^*$. Assume that p/n remains bounded away from 0. Then*

$$\forall\, z \in \mathbb{C}^+ \ , m_p(z) - \mathbf{E}\left(m_p(z)\right) \ \to 0 \quad a.s \ \ ,$$

and also

$$\forall\, z \in \mathbb{C}^+ \ , \frac{\sqrt{p}}{(\log p)^{(1+\alpha)/2}}|m_p(z) - \mathbf{E}\left(m_p(z)\right)| \ \to 0 \quad a.s \ \ , \ \textit{for } \alpha > 0 \ .$$

In other words, $m_p(z)$ is asymptotically deterministic.

Proof. The proof is an immediate consequence of the first Borel-Cantelli lemma. □

Remark 1. We note that if Σ is a matrix independent of the r_i, similar results would apply to

$$\frac{1}{p}\text{trace}\left((M - z\text{Id}_p)^{-1}\Sigma^l\right) \ ,$$

after we replace v by $v/|||\Sigma|||_2^l$. In particular, if $|||\Sigma|||_2 \leq C(\log p)^m$, for some m, we have

$$\frac{1}{p}\text{trace}\left((M - z\text{Id}_p)^{-1}\Sigma^l\right) - \mathbf{E}\left(\frac{1}{p}\text{trace}\left((M - z\text{Id}_p)^{-1}\Sigma^l\right)\right) \to 0 \text{ a.s} \ \ .$$

However, the rate in the second part of the previous corollary needs to be adjusted.

Remark 2. We notice that the rate given by Azuma's inequality does not match the rate that appears in results concerning fluctuation behavior of linear spectral statistics, which is n and not $\sqrt{n}$. Of course, our result encompasses many situations that are not covered by the currently available results on linear spectral statistics, which might contribute to explain this discrepancy. The "correct" rate can be recovered using ideas similar to the ones discussed in Guionnet and Zeitouni (2000) and Ledoux (2001), Chapter 8, Section 5. As a matter of fact, if we consider the Stieltjes transform of the measure that puts mass $1/p$ at each of the singular values of $M = X^*X/n$, it is an easy exercise to see that this function (of X) is $1/(\sqrt{np}v^2)$-Lipschitz with respect to Euclidian (or Frobenius) norm. Hence if the np dimensional vector made up of the entries of X has a distribution that satisfies a dimension free concentration property with respect to Euclidian norm, we find that the fluctuations of the Stieltjes transform at z are of order $\sqrt{np}$, which corresponds to the "correct" rate found in the analysis of these models. (Note however, that results have been shown beyond the case of distributions with dimension-free concentration.)

The conclusion of this discussion is that since the spectral distribution of random matrices is characterized by their Stieltjes transforms, it is not surprising that they are asymptotically non-random, for a very wide class of data matrices of covariance type. We now turn to examining a case of particular interest, the one where the data are elliptically distributed.

3.2 Marčenko-Pastur -type system for covariance matrices computed from elliptically distributed data

We refer the reader to the discussion introducing Section 3 for a review of related literature. In what follows, we assume that we have a triangular "array" of random variables, where the n-th line contains n i.i.d λ_i's and n i.i.d r_i's uniformly distributed on the unit sphere in $\mathbb{R}^p$. In what follows, we allow $\Sigma = \Gamma\Gamma'$ to be random, as long as it is independent of the vectors r_i's. For all practical matters, however, Σ can be considered deterministic.

Theorem 2. *Let $\{\{v_i\}_{i=1}^n\}_{n=1}^\infty$ form a triangular array of independent random vectors, elliptically distributed as described above. In particular, recall that they are in $\mathbb{R}^d$. Call $\theta_n = d/p$, $\rho_n = p/n$, $\xi_n = d^2/np = \theta_n^2\rho_n$. Call H_d the spectral distribution of $\Gamma\Gamma' = \Sigma$ (which is $d \times d$), and ν_n the spectral distribution of the diagonal matrix containing the λ_i. Assume that H_d converges weakly a.s to a probability distribution $H \neq 0$. Assume that ν_n converges weakly a.s to a probability distribution $\nu \neq 0$. Assume further that $\int \tau dH_d(\tau)$ remains bounded.*

Call X the $n \times d$ data matrix whose i-th row is v_i. Consider the matrix

$$B_n = \frac{d}{n}X'X = \frac{d}{n}\sum_{i=1}^n v_i v_i' \triangleq \sum_{i=1}^n u_i u_i' \ .$$

If ρ_n has a finite non-zero limit, ρ, and θ_n has a finite non-zero limit θ, then ξ_n obviously has a finite non-zero limit ξ and the Stieltjes transform of B_n, m_n, has a deterministic limit m satisfying the equations:

$$m(z) = \int \frac{dH(\tau)}{\tau \int \frac{\theta\lambda^2}{1+\xi\lambda^2 w(z)} d\nu(\lambda) - z} \quad \text{and}$$
$$w(z) = \int \frac{\tau dH(\tau)}{\tau \int \frac{\theta\lambda^2}{1+\xi\lambda^2 w(z)} d\nu(\lambda) - z} \ .$$

w is the unique solution of this equation mapping $\mathbb{C}^+$ into $\mathbb{C}^+$. (The intuitive meaning of w is explained just below.) Let us remind the reader that m uniquely characterizes the limiting spectral distribution of B_n.

We note further that we have

$$1 + zm(z) = w(z) \int \frac{\theta\lambda^2}{1+\xi\lambda^2 w(z)} d\nu(\lambda) \ .$$

The same results hold for the scaled sample covariance matrix $d/n(X-\bar{X})'(X-\bar{X})$, since it is a finite rank perturbation of B_n.

The conclusion is that the limiting spectral distribution of B_n is non-random and is characterized by the previous system of two equations.

In the proof we actually do not need the λ_i's to be independent of each other. We only need them to be independent of the r's and we need their empirical distribution to converge a.s to a deterministic limit, ν. In the case of i.i.d λ_i's, we note that ν_n has an almost sure limit ν by the Glivenko-Cantelli Theorem (van der Vaart (1998), Theorem 19.1) for triangular arrays. (A simple modification to the proof given in van der Vaart (1998), which is not for triangular arrays, can be obtained using Hoeffding's inequality for the variables $1_{\lambda_i \leq t}$, which guarantees that the result is true for triangular arrays.)

We note that, maybe interestingly, the proof could be adapted to show that quantities of the type $\mathrm{trace}\left(\Sigma^k (B_n - z\mathrm{Id}_p)^{-1}\right)/d$ satisfy the same equation as w, with τ raised to the power k at the numerator and the same denominator involving w, provided the H_d's have enough moments. (Note that this is the case for m, with k=0 and w which basically corresponds to $k = 1$.)

Finally let us say that we explain in Subsection 3.3 how our approach can be generalized beyond elliptically distributed data to include vectors with more general dependence structure. This allows us to derive other new results, in particular when the dependence structure between the entries of the data vectors is complicated.

3.2.1 On quadratic forms involving vectors sampled uniformly from the 1-sphere in $\mathbb{R}^p$

We consider the concentration properties of quantities of the type

$$r'(M - z\mathrm{Id}_p)^{-1} r \, ,$$

where r is sampled uniformly at random from the 1 sphere in $\mathbb{R}^p$, and M is a symmetric matrix, independent of r. For the sake of simplicity, let us assume for a moment that M is deterministic. We also note that M could be assumed to be Hermitian.

We have the following lemma.

Lemma 7. *Let r be a random vector uniformly distributed on the 1 sphere in $\mathbb{R}^p$. Let M be a deterministic complex matrix. Assume that $|||M|||_2 \leq K$. Then we have,*

$$P(|r'Mr - \frac{1}{p}\mathit{trace}\,(M)\,| > t) \leq 4\exp(-(p-1)(t-c_p)^2/16K^2) \, , t > 0 \, ,$$

with $c_p = \sqrt{\frac{8\pi K^2}{(p-1)}}$.

Proof. Let us write $M = RM + iIM$, where RM and IM are real matrices. We note that

$$|||RM|||_2 \leq K \quad \text{ and } |||IM|||_2 \leq K \, ,$$

by simply writing $RM = (M + \bar{M})/2$.

Now, because we are on the unit sphere, we see that

$$|r_1' RM r_1 - r_2' RM r_2| = |r_1' RM(r_1 - r_2) + (r_1 - r_2)' RM r_2| \leq \|r_1 - r_2\|_2 |||RM|||_2 (\|r_1\|_2 + \|r_2\|_2) = 2|||RM|||_2 \|r_1 - r_2\|_2 \, .$$

So the map $r \to r'RMr$ is $2K$-Lipschitz on the unit sphere, equipped with the geodesic distance, since $d(r_1, r_2) \geq \|r_1 - r_2\|_2$.

We can therefore use well-known concentration results on the unit sphere (see Ledoux (2001), Theorem 2.3) to conclude that, if m_{RM} is a median of $r'RMr$,

$$P(|r'RMr - m_{RM}| > t) \leq 2\exp(-(p-1)t^2/8K^2) \, .$$

Similarly,

$$P(|r'IMr - m_{IM}| > t) \leq 2\exp(-(p-1)t^2/8K^2) \, .$$

Now using Lemma 1.9 in Ledoux (2001), and the fact that $\mathbf{E}\,(rr') = \mathrm{Id}_p/p$, (see Anderson (2003), p.49), we have, because RM and IM are deterministic so independent of r,

$$\left|\mathbf{E}\,\left(r'RMr\right) - m_{RM}\right| = \left|\frac{1}{p}\mathrm{trace}\,(RM) - m_{RM}\right| \leq \sqrt{\frac{8\pi K^2}{(p-1)}} \, , \text{ and}$$

$$\left|\frac{1}{p}\mathrm{trace}\,(IM) - m_{IM}\right| \leq \sqrt{\frac{8\pi K^2}{(p-1)}} \, .$$

Therefore,

$$P(|r'Mr - \frac{1}{p}\text{trace}\,(M)\,| > t) \leq 4\exp(-(p-1)(t-c_p)^2/16K^2)\ .$$

$\square$

Remark 3. We note that similar concentration arguments to those developed here can be derived for other types of correlated random vectors and can also be used to strengthen the results of Lemma 3.1 in Silverstein and Bai (1995). The key here is really the phenomenon of dimension-free concentration, which induces strong concentration of quadratic forms around their mean, for vectors distributed according to measures having the dimension-free concentration property. We develop this remark in subsection 3.3. The technical gist of the remark lies in the fact that if M is a positive semidefinite matrix, whose largest eigenvalue is λ_1, then $g(y) = \sqrt{y'My} = \|M^{1/2}y\|$ is a convex $\sqrt{\lambda_1}$-Lipschitz function of y, and we can apply known results on concentration of convex 1-Lipschitz functions. (See for instance Corollary 4.10 in Ledoux (2001), or Theorem 2.7 where the assumptions of convexity is not needed.) To give the reader a flavor of such results, let us just say that in the case of i.i.d entries for y, belonging to $[a,b]$, we have, if m_g denotes a median of g we have $P(|g-m_g|>t) \leq 4\exp(-t^2/(4(a-b)^2\lambda_1))$. We note further that with the help of Proposition 1.9 in Ledoux (2001), we can also control the distance of any median to the mean μ_g of g, as well as the distance of μ_g^2 to $\mathbf{E}\,(y'My)$, which here would just be trace$\,(M)$, because the covariance of y is Id_p, if $y \in \mathbb{R}^p$, and its entries are independent.

Corollary 3. *Suppose r_i are independent random vectors uniformly distributed on the unit sphere and M_i are random matrices, M_i being independent of r_i, with $|||M_i|||_2 \leq K$, where K is non random, and having the property that, for some matrix $\mathcal{M}$, and some K_p, with $K_p = \mathrm{O}(K/p)$ and $K_p \to 0$,*

$$\forall i\ , \left|\frac{1}{p}\mathit{trace}\,(M_i) - \frac{1}{p}\mathit{trace}\,(\mathcal{M})\right| \leq K_p\ .$$

Then

$$\frac{\sqrt{p}}{(\log p)^{(1+\alpha)/2}K}\max_i \left|r_i'M_ir_i - \frac{1}{p}\mathit{trace}\,(\mathcal{M})\right| \to 0 \quad a.s\ \ . \tag{1}$$

Proof. From the previous lemma, we have

$$P(\max_i |r_i'M_ir_i - \frac{1}{p}\text{trace}\,(M_i)\,| > t) \leq \sum_{i=1}^{p} P(|r_i'M_ir_i - \frac{1}{p}\text{trace}\,(M_i)\,| > t) \leq 4p\exp(-(p-1)(t-c_p)^2/16K^2)\ ,$$

by conditioning on M_i to compute each probability in the sum. Therefore, using the first Borel-Cantelli lemma, we have

$$\frac{\sqrt{p}}{(\log p)^{(1+\alpha)/2}K}\max_i |r_i'M_ir_i - \frac{1}{p}\text{trace}\,(M_i)\,| \to 0 \ \ \text{a.s}\ \ .$$

And because $|\frac{1}{p}\text{trace}\,(M_i) - \frac{1}{p}\text{trace}\,(\mathcal{M})\,| \leq K_p$, we conclude that

$$\frac{\sqrt{p}}{(\log p)^{(1+\alpha)/2}K}\max_i |r_i'M_ir_i - \frac{1}{p}\text{trace}\,(\mathcal{M})\,| \to 0 \ \ \text{a.s}\ \ .$$

$\square$

3.2.2 Preliminaries

We note that the matrix we are considering is of the form $\Gamma X'DX\Gamma'$, where D is a diagonal matrix, containing the λ_k^2. We call τ_i the eigenvalues of $\Sigma = \Gamma\Gamma'$. We call the entries of D λ_i^2.

If we denote by $\|F\|$ the value $\sup_x |F(x)|$, and by F^M the cdf of the spectral distribution of the matrix M. We see using Lemma 2.5 in Silverstein and Bai (1995) that

$$\|F^{Q^*TQ} - F^{\tilde{Q}^*\tilde{T}\tilde{Q}}\| \leq \frac{1}{p}\left(\text{rank}\left(T-\tilde{T}\right) + 2\text{rank}\left(\tilde{Q}-Q\right)\right)\ .$$

In our situation, we have $Q = X\Gamma'$ and $T = D$, so using the fact that rank $(AB) \leq \min(\text{rank}\,(A)\,, \text{rank}\,(B))$, we conclude that

$$\|F^{Q^*TQ} - F^{\tilde{Q}^*\tilde{T}\tilde{Q}}\| \leq \frac{1}{p}\left(\text{rank}\left(D - \tilde{D}\right) + 2\text{rank}\left(\tilde{\Gamma}' - \Gamma'\right)\right) \ .$$

Now let us choose for $\tilde{D}$ the diagonal matrix with entries $\lambda_i^2 1_{|\lambda_i| \leq \alpha_p}$, which we abbreviate by $D1_{|D| \leq \alpha_p}$, and for $\tilde{\Gamma'} = \Gamma' 1_{|\Sigma| \leq \beta_p}$ (this is understood using the singular value decomposition of Γ', where we keep the singular values that are less than $\sqrt{\beta_p}$ and replace the others by 0).

We see that rank $\left(D - \tilde{D}\right) = \sum_{i=1}^n 1_{|\lambda_i| > \alpha_p}$, and similarly, $0 \leq$ rank $\left(\Gamma' - \tilde{\Gamma'}\right) \leq \sum_{i=1}^d 1_{|\tau_i| > \beta_p}$. Because we assumed that H_d converges weakly a.s to H, and ν_n converges weakly a.s to ν, we conclude that for $\alpha_p = \beta_p = \log p$, rank $\left(\Gamma' - \tilde{\Gamma'}\right)/p \to 0$ a.s and rank $\left(D - \tilde{D}\right)/p \to 0$ a.s. Here it is important that d/p and p/n have finite non-zero limits.

So to prove the theorem, it is sufficient to prove it for D and Σ bounded in operator norm by, for instance, $\log p$, since we just showed that Q^*TQ and $\tilde{Q}^*\tilde{T}\tilde{Q}$ will have the same limiting spectral measure, provided it exists.

3.2.3 Proof of Theorem 2

As explained above, we now assume that all the eigenvalues of $\Sigma = \Gamma\Gamma'$ are less than $\log p$ and similarly, we assume that $|\lambda_i| < \sqrt{\log p}$. We call the corresponding spectral measures $\widetilde{H}_d$ and $\widetilde{\nu}_n$, to keep track of the modifications we have induced by truncation. However, to avoid cumbersome notations, we use Σ and Γ to refer to the matrices we deal with. ($\widetilde{\Sigma}$ might have been more appropriate but the notation would be too cumbersome.) The approach we use follows the "rank-1 perturbation" approach developed in Silverstein and Bai (1995) and Silverstein (1995).

Recall that $u_k = \sqrt{d/n}\lambda_k \Gamma r_k$. We call $B_{(k)} = B_n - u_k u_k'$, $M_k = (B_n - u_k u_k' - z\text{Id}_d)^{-1}$, $\mathcal{M}_n = (B_n - z\text{Id}_d)^{-1}$, and

$$\beta(z) = \frac{1}{n}\sum_{k=1}^n \frac{\theta_n \lambda_k^2}{1 + u_k' M_k u_k} \ .$$

We note that B_n is $d \times d$ and so are all the other matrices involved here. Using the first resolvent identity $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, and the fact that (see Silverstein (1995))

$$B_n(B_n - z\text{Id}_d)^{-1} = \text{Id}_d + z(B_n - z\text{Id}_d)^{-1} = \sum_{k=1}^n \frac{u_k u_k' M_k}{1 + u_k' M_k u_k} \ , \tag{2}$$

we have

$$(\beta(z)\Sigma - z\text{Id}_d)^{-1} - (B_n - z\text{Id}_d)^{-1} = (\beta(z)\Sigma - z\text{Id}_d)^{-1}\left[\sum_{k=1}^n \frac{u_k u_k' M_k}{1 + u_k' M_k u_k} - \beta(z)\Sigma(B_n - z\text{Id}_d)^{-1}\right] \ ,$$

and hence

$$(\beta(z)\Sigma - z\text{Id}_d)^{-1} - (B_n - z\text{Id}_d)^{-1} =$$
$$\sum_{k=1}^n \frac{1}{1 + u_k' M_k u_k}\left[(\beta(z)\Sigma - z\text{Id}_d)^{-1} u_k u_k' M_k - \frac{\theta_n}{n}\lambda_k^2 (\beta(z)\Sigma - z\text{Id}_d)^{-1}\Sigma(B_n - z\text{Id}_d)^{-1}\right]$$

Taking traces and dividing by d, we get

$$\int \frac{d\widetilde{H}_d(\tau)}{\beta(z)\tau - z} - m_n(z) =$$
$$\frac{1}{d}\sum_{k=1}^n \frac{1}{1 + u_k' M_k u_k}\left[u_k' M_k (\beta(z)\Sigma - z\text{Id}_d)^{-1} u_k - \frac{\theta_n}{n}\lambda_k^2 \text{trace}\left((\beta(z)\Sigma - z\text{Id}_d)^{-1}\Sigma\mathcal{M}_n\right)\right]$$

Now using for instance Equation (2.3) in Silverstein (1995), we have easily

$$\frac{1}{1+u_k'M_ku_k}\leq\frac{|z|}{v}\;.$$

On the other hand, it is clear that $\mathrm{Im}\left[\beta(z)\right]\leq 0$. As a matter of fact, the eigenvalues of M_k all have a positive imaginary part (if $z=u+iv$, they are $1/(\lambda_j(B_{(k)})-u-iv)$), so the imaginary part of $1+u_k'M_ku_k$ is positive, and the imaginary part of $1/(1+u_k'M_ku_k)$ is negative. Hence the imaginary part of the eigenvalues of $\beta(z)\Sigma-z\mathrm{Id}_d$ is smaller than $-v$ (Σ is positive semidefinite), and their module is greater than v. Therefore

$$|||\mathrm{Re}\left[(\beta(z)\Sigma-z\mathrm{Id}_d)^{-1}\right]|||_2\leq\frac{1}{v}\;\text{ and }|||\mathrm{Im}\left[(\beta(z)\Sigma-z\mathrm{Id}_d)^{-1}\right]|||_2\leq\frac{1}{v}\;.$$

Now $\beta(z)$ depends on all the u_k's in a non-trivial fashion, so we cannot apply our concentration results directly. However, we note that if $b(z)$ is another complex number we have, if $\Sigma=\sum_{i=1}^p\tau_ie_ie_i'$,

$$(\beta(z)\Sigma-z\mathrm{Id}_d)^{-1}-(b(z)\Sigma-z\mathrm{Id}_d)^{-1}=\sum\frac{\tau_i(b(z)-\beta(z))}{(\tau_ib(z)-z)(\tau_ib(z)-z)}e_ie_i'\;\text{ and}$$
$$\Sigma^m\left[(\beta(z)\Sigma-z\mathrm{Id}_d)^{-1}-(b(z)\Sigma-z\mathrm{Id}_d)^{-1}\right]\Sigma^l=\sum\frac{\tau_i^{l+m+1}(b(z)-\beta(z))}{(\tau_ib(z)-z)(\tau_ib(z)-z)}e_ie_i'$$

Therefore, if $b(z)$ is such that $|\beta(z)-b(z)|\leq\epsilon$, and $\mathrm{Im}\left[b(z)\right]\leq 0$, we have,

$$|||(\beta(z)\Sigma-z\mathrm{Id}_d)^{-1}-(b(z)\Sigma-z\mathrm{Id}_d)^{-1}|||_2\leq\frac{\epsilon|||\Sigma|||_2}{v^2}$$
$$|u_k'M_k(\beta(z)\Sigma-z\mathrm{Id}_d)^{-1}u_k-u_k'M_k(b(z)\Sigma-z\mathrm{Id}_d)^{-1}u_k|\leq\frac{1}{v^3}\epsilon|||\Sigma|||_2\|u_k\|_2^2\;\text{ and}$$
$$\left|\frac{1}{p}\mathrm{trace}\left(\Sigma^lM_k\left[(\beta(z)\Sigma-z\mathrm{Id}_d)^{-1}-(b(z)\Sigma-z\mathrm{Id}_d)^{-1}\right]\right)\right|\leq\frac{4|||\Sigma|||_2^l\epsilon}{v^3}\;,$$

by decomposing the matrices appearing in the trace into real and imaginary parts, which are both symmetric in this instance, and using well-known result (see e.g Anderson (2003), Theorem A.4.7) on bounds of the trace of a product of symmetric matrices.

Consider

$$b_n(z)=\frac{\theta_n}{n}\sum_{k=1}^n\frac{\lambda_k^2}{1+\xi_n\lambda_k^2\mathbf{E}\left(\Omega_1(z)\right)}\;,\;\text{ with }\Omega_1(z)=\frac{1}{d}\mathrm{trace}\left(\Sigma(B_n-z\mathrm{Id}_p)^{-1}\right)\;.$$

Our Corollary 3 on concentration implies that $\max_i|r_i'\Gamma'M_i\Gamma r_i-\mathbf{E}\left(\Omega_1(z)\right)|$ is less $\epsilon(\log p)^{(1+\alpha)/2}K/\sqrt{p}\triangleq\epsilon\gamma_p$ a.s. We note that here, K is of the order of $|||\Sigma|||_2/v$, so $(\log p)/v$, because we are thresholding Σ at this level.

When this happens, we have, if we call $\alpha_k=r_k'\Gamma'M_k\Gamma r_k=u_k'M_ku_k/(\xi_n\lambda_k^2)$, and $\alpha=\mathbf{E}\left(\Omega_1(z)\right)$,

$$|\beta(z)-b_n(z)|=\left|\frac{\theta_n}{n}\sum_{k=1}^n\left(\frac{\lambda_k^2}{1+\xi_n\lambda_k^2\alpha_k}-\frac{\lambda_k^2}{1+\xi_n\lambda_k^2\alpha}\right)\right|\;,$$
$$\leq\frac{\xi_n\theta_n}{n}\sum_{k=1}^n\frac{\lambda_k^4\epsilon\gamma_p}{|(1+\xi_n\lambda_k^2\alpha_k)(1+\xi_n\lambda_k^2\alpha)|}\leq\frac{\xi_n\theta_n\epsilon|z|^2\gamma_p}{nv^2}\sum_{k=1}^n\lambda_k^4\;.$$

Using our concentration bounds from Corollary 3 applied to

$$\left[u_k'M_k\left(b_n(z)\Sigma-z\mathrm{Id}_d\right)^{-1}u_k/\lambda_k^2-\frac{\theta_n}{n}\mathrm{trace}\left(\left(b_n(z)\Sigma-z\mathrm{Id}_d\right)^{-1}\Sigma\mathcal{M}_n\right)\right]\;,$$

we see that, we have a.s

$$max_{1\leq k\leq p}\left|\left[u_k'M_k\left(b_n(z)\Sigma-z\mathrm{Id}_d\right)^{-1}u_k/\lambda_k^2-\frac{\theta_n}{n}\mathrm{trace}\left(\left(b_n(z)\Sigma-z\mathrm{Id}_d\right)^{-1}\Sigma\mathcal{M}_n\right)\right]\right|<\frac{\epsilon\gamma_p}{v}\;,$$

and therefore

$$\left|\frac{1}{d}\sum_{k=1}^{n}\frac{1}{1+u_k' M_k u_k}\left[u_k' M_k\left(b_n(z)\Sigma - z\mathrm{Id}_d\right)^{-1}u_k - \frac{\theta_n}{n}\lambda_k^2 \mathrm{trace}\left(\left(b_n(z)\Sigma - z\mathrm{Id}_d\right)^{-1}\Sigma\mathcal{M}_n\right)\right]\right| \leq C\epsilon\gamma_p\frac{|z|}{v^2}\frac{1}{d}\sum_{k=1}^{n}\lambda_k^2 .$$

We conclude that a.s,

$$\left|\int \frac{d\widetilde{H}_d(\tau)}{b_n(z)\tau - z} - m_n(z)\right| \leq C(z)\gamma_p\epsilon\frac{1}{n}\sum_{k=1}^{n}\left(\lambda_k^2+\lambda_k^4\right) \leq C(z)\epsilon\frac{2(\log p)^5}{\sqrt{p}}$$

Because of our assumptions, we finally get

$$\int \frac{d\widetilde{H}_d(\tau)}{b_n(z)\tau - z} - m_n(z) \to 0 \text{ a.s} \quad .$$

This corresponds to the first part of the theorem. Now note that $\mathrm{Im}\left[b_n(z)\right] \leq 0$, and therefore $|1/(b_n(z)\tau - z)| \leq 1/v$. Because $\int |d\widetilde{H}_d(\tau) - dH_d(\tau)| \to 0$, we conclude that

$$\int \frac{dH_d(\tau)}{b_n(z)\tau - z} - m_n(z) \to 0 \text{ a.s} \quad .$$

To get to the second part of the theorem, we consider instead

$$\Sigma\left(\beta(z)\Sigma - z\mathrm{Id}_d\right)^{-1} - \Sigma(B_n - z\mathrm{Id}_d)^{-1} .$$

Taking traces and dividing by d, we get

$$\int \frac{\tau d\widetilde{H}_d(\tau)}{\tau\beta(z) - z} - \frac{1}{d}\mathrm{trace}\left(\Sigma(B_n - z\mathrm{Id}_d)^{-1}\right) \quad .$$

To control this quantity, we can use the same expansions we used before, replacing everywhere $(\beta(z)\Sigma - z\mathrm{Id}_d)^{-1}$ by $\Sigma\left(\beta(z)\Sigma - z\mathrm{Id}_d\right)^{-1}$. This has the effect of multiplying the upper bounds by $|||\Sigma|||_2$ and dividing the terms appearing in the exponential by $|||\Sigma|||_2^2$. So we conclude that

$$\int \frac{\tau d\widetilde{H}_d(\tau)}{\tau b_n(z) - z} - \Omega_1(z) \to 0 \text{ a.s} \quad .$$

Now the result we got using Azuma's inequality shows clearly (see Remark 1) that

$$\Omega_1(z) - \mathbf{E}\left(\Omega_1(z)\right) \to 0 \text{ a.s} \quad .$$

Calling $w_n(z) = \mathbf{E}\left(\Omega_1(z)\right)$, we have shown that

$$\begin{cases} \int \frac{\tau d\widetilde{H}_d(\tau)}{\tau\int\frac{\theta_n\lambda^2 d\widetilde{\nu}_n(\lambda)}{1+\xi_n\lambda^2 w_n(z)} - z} - w_n(z) & \to 0 \text{ a.s} \ , \text{ and} \\ \int \frac{d\widetilde{H}_d(\tau)}{\tau\int\frac{\theta_n\lambda^2 d\widetilde{\nu}_n(\lambda)}{1+\xi_n\lambda^2 w_n(z)} - z} - m_n(z) & \to 0 \ \text{ a.s} . \end{cases} \tag{3}$$

• **Subsequence argument to reach the conclusion of Theorem 2**

We now need to turn to technical arguments to go from the statement of Equation 3 to that of Theorem 2. Because of our assumption that $\int \tau dH_d(\tau) < K$, for all d (or p, which is equivalent), with K fixed and independent of d, we see that $|w_n(z)| \leq \mathrm{trace}\left(\Sigma\right)/(dv) < K/v$. So, at z fixed, $w_n(z)$ is bounded. From this sequence, let us extract a convergent subsequence $w_{m(n)}(z)$, or w_m for short, that converges to w. Through tightness arguments (see below), we see that $w \in \mathbb{C}^+$. We will now show that $w(z)$ satisfies

$$\int \frac{\tau dH(\tau)}{\tau\int\frac{\theta\lambda^2 d\nu(\lambda)}{1+\xi\lambda^2 w(z)} - z} - w(z) = 0$$

and that there is a unique solution to this equation in $\mathbb{C}^+$. Let us call $b_m(z) = \int \theta_m \lambda^2 d\tilde{\nu}_m(\lambda)/(1 + \xi_m \lambda^2 w_m(z))$. We first show that $b_m \to b = \int \theta \lambda^2 d\nu(\lambda)/(1+\xi\lambda^2 w(z))$. To do so, note that $\lambda^2/(1+a_m\lambda^2) - \lambda^2/(1+a\lambda^2) = (a-a_m)\lambda^4/[(1+a\lambda^2)(1+a_m\lambda^2)]$. Note because $a_m \to a \in \mathbb{C}^+$, their imaginary parts are uniformly bounded below by δ, from which we conclude that, if $a_m \to a \in \mathbb{C}^+$,

$$\int \frac{\lambda^2 d\tilde{\nu}_m}{1+a_m\lambda^2} - \int \frac{\lambda^2 d\tilde{\nu}_m}{1+a\lambda^2} \to 0$$

On the other hand, for $a \in \mathbb{C}^+$, $\lambda^2/(1+a\lambda^2)$ is a bounded continuous function of λ. Since $\nu_m \Rightarrow \nu$, and therefore $\tilde{\nu}_m \Rightarrow \nu$, we conclude that

$$\int \frac{\lambda^2 d\tilde{\nu}_m}{1+a\lambda^2} \to \int \frac{\lambda^2 d\nu}{1+a\lambda^2} \; .$$

Therefore, since $\theta_m \to \theta$, $b_m(z) \to b(z)$. Because we have assumed that $\nu \neq 0$, we have $b(z) \in \mathbb{C}^-$. By essentially the same arguments, using the fact that $|\mathrm{Im}\,[b_m(z)]\,|$ is bounded below by δ and $b(z) \in \mathbb{C}^-$, we conclude that

$$\int \frac{\tau d\tilde{H}_{d(m(n))}(\tau)}{\tau b_{m(n)}(z) - z} - \int \frac{\tau dH(\tau)}{\tau b(z) - z} \to 0 \; .$$

In other words,

$$\int \frac{\tau dH(\tau)}{\tau b(z) - z} - w(z) = 0 \; ,$$

where

$$b(z) = \int \frac{\theta \lambda^2 d\nu(\lambda)}{1+\xi\lambda^2 w(z)} \; .$$

Similarly, we can show that along this subsequence,

$$\int \frac{dH_d(\tau)}{\tau b_m(z) - z} \to \int \frac{dH(\tau)}{\tau b(z) - z} \; ,$$

and so we also get the first equation in Theorem 2.

• **Uniqueness of possible limit**

We now prove that there is a unique solution in $\mathbb{C}^+$ to the equation characterizing w, the only question remaining to tackle being uniqueness. To do so, we employ an argument similar to that given in Silverstein and Bai (1995), though the details are slightly different.

Suppose we have two solutions in $\mathbb{C}^+$ to the equation characterizing $w(z)$. Let us call them w_1 and w_2 and b_1 and b_2 are the corresponding b's. We have

$$\begin{aligned} w_1 - w_2 &= \int \left(\frac{\tau}{\tau b_1 - z} - \frac{\tau}{\tau b_2 - z} \right) dH(\tau) \\ &= (b_2 - b_1) \int \frac{\tau^2}{(\tau b_1 - z)(\tau b_2 - z)} dH(\tau) \\ &= \theta(w_1 - w_2) \int \frac{\lambda^4 \xi d\nu(\lambda)}{(1+\xi\lambda^2 w_1(z))(1+\xi\lambda^2 w_2(z))} \int \frac{\tau^2}{(\tau b_1 - z)(\tau b_2 - z)} dH(\tau) \end{aligned}$$

Let us call f the quantity multiplying $w_1 - w_2$ in the previous equation. We want to show that $|f| < 1$. As in Silverstein and Bai (1995), using Holder's inequality, we have, given that $\theta > 0$,

$$|f| \leq \left(\theta \int \frac{\lambda^4 \xi d\nu(\lambda)}{|1+\xi\lambda^2 w_1(z)|^2} \int \frac{\tau^2}{|\tau b_1 - z|^2} dH(\tau) \right)^{1/2} \left(\theta \int \frac{\lambda^4 \xi d\nu(\lambda)}{|1+\xi\lambda^2 w_2(z)|^2} \int \frac{\tau^2}{|\tau b_2 - z|^2} dH(\tau) \right)^{1/2}$$

Let us write $w_1 = a + ic$, $z = u + iv$, and $b_1 = \alpha - i\gamma$. By writing the definition of b_1 in terms of w_1 we see immediately that

$$\gamma = c \int \frac{\theta \xi \lambda^4}{|1+\xi\lambda^2 w_1|^2} d\nu(\lambda) \; ,$$

so $\int \frac{\theta\xi\lambda^4}{|1+\xi\lambda^2 w_1|^2} d\nu(\lambda) = -\mathrm{Im}\,[b_1]\,/\mathrm{Im}\,[w_1]$. Since $\nu \neq 0$ by our assumptions, we see that $\gamma > 0$. On the other hand, using the definition of w_1 in terms of b_1, we see that

$$\mathrm{Im}\,[w_1] = \int -\mathrm{Im}\,[b_1] \frac{\tau^2}{|\tau b_1 - z|^2} dH(\tau) + \mathrm{Im}\,[z] \int \frac{\tau}{|\tau b_1 - z|^2} dH(\tau) ,$$

and therefore $\mathrm{Im}\,[w_1] > -\mathrm{Im}\,[b_1] \int \frac{\tau^2}{|\tau b_1 - z|^2} dH(\tau)$, since $H \neq 0$.

Hence,

$$\left(\int \frac{\theta\lambda^4 \xi d\nu(\lambda)}{|1+\xi\lambda^2 w_1(z)|^2} \int \frac{\tau^2}{|\tau b_1 - z|^2} dH(\tau) \right)^{1/2} < 1 ,$$

and $|f| < 1$. We conclude that $w_2 = w_1$, so there is at most one solution to the equation characterizing w.

• **Tightness of B_n and consequences for w**

Finally, we need to show that the spectral distribution F^{B_n} is tight a.s and draw consequences for w. It is shown - through Lemma 2.3 - in Silverstein and Bai (1995), that if $B_n = T_n^{1/2} Y_n^* Y_n T_n^{1/2}$, the spectral distributions of the T_n's form a tight sequence and so do the spectral distributions of the $Y_n^* Y_n$'s, then F^{B_n} form a tight sequence. We note that in our case $B_n = \Gamma R_n^* D_n^2 R_n \Gamma'$, which up to a number of zeros has the same eigenvalues as $\Sigma^{1/2} R_n^* D_n^2 R_n \Sigma^{1/2}$; we temporarily call R_n the matrix containing our random vectors uniformly distributed on the sphere, to insist on this property which will play a crucial role shortly. So all we have to show is that $F^{R_n^* D_n^2 R_n}$ forms a tight sequence. Note that our assumption on the convergence of the spectral distribution of the λ's implies that the spectral distribution of D_n^2 form a tight sequence. So all we have to do to be able to conclude is to show that so does $F^{R_n^* R_n}$. Note that $\mathrm{trace}\,(R_n^* R_n)\,/p = \sum_{i=1}^n \mathrm{trace}\,(r_i r_i^*)\,/p = n/p$. Because n/p is uniformly bounded, we conclude that $F^{R_n^* R_n}$ forms a tight sequence. So F^{B_n} forms a tight sequence, a.s. Note also that $F^{R_n^* R_n}([M,\infty)) \leq n/(pM)$. So for any ϵ, we can find M_ϵ such that $F^{B_n}[M_\epsilon, \infty) < \epsilon$, a.s. Using the second inequality in Lemma 2.3 in Silverstein and Bai (1995) and the fact that H and ν are deterministic, as well as the fact that if $X_n \Rightarrow X$ and F is closed, $\limsup P(X_n \in F) \leq P(X \in F)$, we see that M_ϵ can be chosen uniformly in ω.

We now want to show that $w \in \mathbb{C}^+$; to do so, we will show that a.s, $\mathrm{Im}\,[w_n]$ is bounded away from zero. Note that $\mathrm{Im}\,\left[(B_n - z\mathrm{Id})^{-1}\right]$ is a symmetric matrix. Its eigenvalues, which we denote by a_k, are, if l_k denote the eigenvalues of B_n, $v/((l_k-u)^2+v^2) \geq v/(2(l_k^2+u^2)+v^2$. Assume $a_1 \geq a_2 \geq \ldots \geq a_d$. Using Theorem A.4.7 in Anderson (2003), we see that, if we call τ_i's the decreasingly ordered eigenvalues of Σ,

$$\mathrm{Im}\left[\frac{1}{d}\mathrm{trace}\left(\Sigma(B_n - z\mathrm{Id}_p)^{-1}\right)\right] \geq \frac{1}{d}\sum_{i=1}^d \tau_i a_{d-i+1} .$$

Now all we need to show is that a.s a fixed non-zero proportion of $\tau_i a_{d-i}$ stay bounded away from 0. Because $H \neq 0$, we can find η such that $H(\eta,\infty) > \epsilon$, for some $\epsilon > 0$. Let us pick such an $\epsilon \neq 0$. In particular, the proportion of indices for which $\tau_i > \eta$ is a.s greater than ϵ, because $\liminf H_d(\eta,\infty) \geq H(\eta,\infty)$, a.s. For this ϵ, we can find $m_\epsilon < \infty$, such that $F^{B_n}[0, m_\epsilon] \geq 1 - \epsilon/2$, a.s from our arguments above. So the proportion of i's such that $a_{d-i+1} \geq v/(2(m_\epsilon^2 + u^2) + v^2)$ is greater than $1-\epsilon/2$. So the proportion of i's for which both $\tau_i > \eta$ and $a_{d-i+1} \geq v/(2(m_\epsilon^2 + u^2) + v^2)$ must be greater than $\epsilon/2$, a.s. Hence, $\mathrm{Im}\,[w_n(z)] \geq \delta > 0$, a.s.

Remark 4. We also note that another approach to the proof of the theorem is possible by starting with Gaussian random vectors. We give a rough sketch here. A point distributed uniformly at random of the unit sphere in $\mathbb{R}^p$ can be obtained by generating a Gaussian random vector in $\mathbb{R}^p$, with identity covariance and dividing each entry of the vector by the Euclidian norm of the vector. Hence our covariance matrix for elliptically distributed data is of the form (say if $d = p$) $T_p^{1/2} X_{n,p}^* L_n D_n X_{n,p} T_p^{1/2}$. Now D_n is diagonal and $D_n(i,i) = 1/\|X_i\|_2^2$, where X_i is the i-th row of $X_{n,p}$. X_i is standard Gaussian in $\mathbb{R}^p$. The problem now is that we have dependence between $L_n D_n$ and $X_{n,p}$. So the known results do not apply directly. However, since our B_n is p times the standard estimate of covariance, we see that B_n is a standard covariance of the above type, if we replace D_n by pD_n. Now standard results in extreme value theory give that $|||pD_n - \mathrm{Id}_n|||_2$ is of order $\sqrt{\log(p)/p}$ because n and p are of the same order of magnitude. What we then need to do is truncate L_n and T_p and remove the eigenvalues that are say larger than $\log(p)$. This does not change the results on the convergence of the spectral distribution (see subsubsection 3.2.2). After this truncation is done, using coarse bounds on $|||\cdot|||_2$, and known results on the largest eigenvalue of sample covariance matrices computed from i.i.d Gaussian data, we see that

$$|||B_n - \frac{1}{n} T_p^{1/2} X_{n,p}^* L_n X_{n,p} T_p^{1/2}|||_2 = \mathrm{O}\left(\frac{(\log p)^3}{\sqrt{p}}\right) ,$$

and the two matrices have asymptotically the same spectrum.

3.3 Possible generalization and comments on the proof

The reader will have noted that the crux of our argument relies on the fact that certain quadratic forms are concentrated around their mean, and the rank 1 developments explained in Silverstein and Bai (1995) and Silverstein (1995). As far as concentration is concerned, we essentially used so-called dimension free Gaussian concentration results. In our case, we used the fact that this applied to random vectors uniformly distributed on the sphere of radius 1 to get around the difficulty that the dependence of the coordinates creates.

We note that this approach is not limited to such vectors. In fact, it would work as soon as we were working with independent random vectors that have the property that convex Lipschitz functionals of those vectors are themselves concentrated, in an "almost dimension free" fashion and Gaussian manner. As a matter of fact, we have the following result, which is a small generalization of our main idea and was implicit in the study of elliptically distributed data.

Lemma 8 (Role of Gaussian Concentration). *Suppose that the random vector $r \in \mathbb{R}^p$ has the property that for any convex 1-Lipschitz (with respect to Euclidian norm) functional F, we have*

$$P(|F(r) - m_F| > t) \leq C \exp(-c(p)t^2) ,$$

where C and $c(p)$ are independent of F and C is independent of p. We allow $c(p)$ to be a constant or to go to zero with p like $p^{-\alpha}$, $0 \leq \alpha < 1$. Suppose further that $E(rr^*) = \Sigma$, with $|||\Sigma|||_2 \leq \log(p)$.

If M is a complex deterministic matrix, with $|||M|||_2 \leq \xi$, where ξ is independent of p,

$$\frac{1}{p} r'Mr \text{ is strongly concentrated around its mean, } \frac{1}{p} \mathrm{trace}\,(M\Sigma) .$$

The same is true if one works with $\Sigma^{1/2} r$ instead of r, when r has identity covariance.

The statement might seem a bit vague, but what we mean by strong concentration here is the fact the probabilities of deviations are exponentially small in p.

Proof. In what follows, K denotes a generic constant, that may change from display to display, but is independent of p. First, as seen above, we can rewrite M as $M = RM + iIM$ where RM and IM are real matrices. Further the spectral norm of those matrices is less than ξ.

Now strong concentration for $r'RMr/p$ and $r'IMr/p$ will imply strong concentration for the sum of those two terms. We note that, since $r'RMr$ is a real, $r'RMr = (r'RMr)'$ and

$$r'RMr = r'\left(\frac{RM + RM'}{2}\right) r .$$

Hence instead of working on RM we can work on its symmetrized version.

Now let us decompose $(RM + RM')/2$ into $RM_+ + RM_-$, where RM_+ is positive semi-definite and $-RM_-$ is positive definite (or 0 if $(RM + RM')/2$ itself is positive semi-definite). This is possible because $(RM+RM')/2$ is real symmetric and we do this decomposition by just following its spectral decomposition. Note that both matrices have spectral norm less than ξ. Now the map $\phi : r \to \sqrt{r'RM_+r}$ is $\sqrt{\xi}$-Lipschitz (with respect to Euclidian norm) and convex, which is easily seen after one notices that $\sqrt{r'RM_+r} = \|RM_+^{1/2} r\|_2$. This guarantees by our assumption that

$$P(|\sqrt{r'RM_+r} - m_\phi| > t) \leq C \exp(-c(p)t^2/\xi) , \text{ where } m_\phi \text{ is a median of } \phi .$$

Using the same type of arguments as in the end of Lemma 4, we see that the variance of $\sqrt{r'RM_+r/p}$ can be explicitly bounded, as well as the deviation of its mean from its median which guarantees that $E(r'RM_+r/p) = \text{trace}\,(RM_+\Sigma/p)$ cannot be far from m_ϕ^2. Again, as in Lemma 4, we can get a bound of the type

$$P(|\sqrt{r'RM_+r/p}-\sqrt{\text{trace}\,(RM_+\Sigma/p)}| > t) \leq C\exp(-pc(p)(t-\kappa_p)^2/\xi) \text{ , for some } K \text{ and } \kappa_p \to 0 \text{ as } p \to \infty \; .$$

The fact that $\kappa_p \to 0$ is consequence of the fact that $pc(p) \to \infty$. Let us denote by

$$\begin{aligned}\zeta_p &= \text{trace}\,(RM_+\Sigma/p)\\ A &= \{|r'RM_+r/p - \zeta_p| > t\}\\ B &= \{\sqrt{r'RM_+r/p} \leq \sqrt{\zeta_p}+1\}\end{aligned}$$

Our aim is to show that the probability of A is exponentially small in p. Of course, we have $P(A) \leq P(A\cap B) + P(B^c)$. We note that $P(B^c)$ is exponentially small in p from our previous arguments. Now, note that

$$A\cap B \subseteq D = \{|\sqrt{r'RM_+r/p} - \sqrt{\zeta_p}| > \frac{t}{2\sqrt{\zeta_p}+1}\} \; .$$

To see this, note simply that for positive reals, $|x-y| = |\sqrt{x}-\sqrt{y}|(\sqrt{x}+\sqrt{y})$. Finally, because of our bounds on the norm of Σ and the fact that $|||RM_+|||_2 \leq \xi$, we see that $\text{trace}\,(RM_+\Sigma/p) = \zeta_p \leq \log(p)\xi$. Hence, $P(D) \leq C\exp(-Kc(p)(t-\kappa_p)^2p/[(\log p)^2\xi])$, for some K independent of p, and hence, we have

$$P(A) \leq 2C\exp(-Kc(p)(t-\kappa_p)^2p/[(\log p)^2\xi]) \; .$$

Similarly, we can obtain the same type of bounds for $\sqrt{-r'RM_-r/p}$. From those we conclude that

$$P(\left|r'RMr/p - \text{trace}\,(RM\Sigma_p)\,/p\right| > t) \leq 4C\exp(-Kc(p)v(t/2-\kappa_p)^2p/[(\log p)^2\xi]) \; .$$

And finally,

$$P(\left|r'Mr/p - \text{trace}\,(M\Sigma_p)\,/p\right| > t) \leq \tilde{C}\exp(-Kc(p)v(t/2\sqrt{2}-\kappa_p)^2p/[(\log p)^2\xi]) \; .$$

This guarantees the same convergence results as in Lemma 7. And hence a result similar to Corollary 3 holds. $\square$

Along the same lines, we also have:

Lemma 9 (Beyond Gaussian Concentration). *Suppose that the random vector $r \in \mathbb{R}^p$ has the property that for any convex 1-Lipschitz (with respect to Euclidian norm) functional F, we have*

$$P(|F(r) - m_F| > t) \leq C\exp(-c(p)t^b) \; ,$$

where C and $c(p)$ are independent of F and C is independent of p. We allow $c(p)$ to be a constant or to go to zero with p like $p^{-\alpha}$, $0 \leq \alpha < b/2$. Suppose further that $E(rr^*) = \Sigma$, with $|||\Sigma|||_2 \leq \log(p)$.

If M is a complex deterministic matrix, with $|||M|||_2 \leq \xi$, where ξ is independent of p,

$$\frac{1}{p}r'Mr \textit{ is strongly concentrated around its mean, } \frac{1}{p}\textit{trace}\,(M\Sigma) \; .$$

The same is true if one works with $\Sigma^{1/2}r$ instead of r, when r has identity covariance.

Proof. We only give a sketch of the proof. The ideas are exactly the same as above. However, when studying the concentration of $\sqrt{r'RM_+r/p}$, the exponent of the exponential is to leading order $p^{b/2}c(p)(t-\kappa_p)^b$. We note that κ_p will be a bit different in its form than it was in the Gaussian concentration case. This comes from the fact, following the analysis in Proposition 1.9 of Ledoux (2001), the inequalities we now have, if μ_F denotes the mean of F are:

$$|\mu_F - m_F| \leq \frac{C}{bc^{1/b}}\Gamma(\frac{1}{b}) \quad \text{and } \text{var}\,(F) \leq \frac{2C}{bc^{2/b}}\Gamma(\frac{2}{b}) \; ,$$

where Γ denotes the Gamma function. After this adjustment the previous proof goes through. $\square$

Corollary 4. *When the two above lemmas apply, and trace* $(\Sigma_p)/p$ *is bounded, the spectral distribution of* $\sum r_i r_i^*/n$ *is a.s tight.*

Proof. As we did above, we consider the first moment of the spectral distribution of $R_n^* R_n$, which is equal to M_1, with $M_1 = 1/n\text{trace}\left(\sum r_i r_i^*/p\right)$. Its mean is $\text{trace}\left(\Sigma\right)/p$. As we just saw, $r_i^* r_i/p$ is strongly concentrated around $\text{trace}\left(\Sigma\right)/p(= \mathbf{E}\left(M_1\right))$ and this property transfers to M_1 using the fact that $P(|M_1 - \mathbf{E}\left(M_1\right)| > t) \leq nP(|r_i^* r_i/p - \mathbf{E}\left(r_i^* r_i/p\right)| > t)$. A.s tightness of the sequence of measure follows easily. $\square$

We note that these last inequalities are the equivalent of the result of Lemma 7, which is the key to the rest of the analysis in the case of vectors uniformly distributed on the sphere. The same subsequent analysis can therefore be carried out for all family of random vectors that satisfy the conditions of the preceding Lemmas. Note however that we need the spectral distribution of Σ to be asymptotically non-degenerate and have a finite first moment. In particular, when these conditions are satisfied here are a few examples, to which Theorem 2 applies, with the modification that

$$B_n = \frac{d}{p}\frac{1}{n}\sum \lambda_i^2 v_i v_i^*$$

Example of distributions for v_i for which Theorem 2 applies

- Gaussian random variables, with covariance having uniformly bounded 1st moment.
- Vectors of the type $\sqrt{p}r$ where r is uniformly distributed on the 1-sphere is dimension p, i.e vectors uniformly distributed on the p-sphere in $\mathbb{R}^p$.
- Vectors $\Gamma\sqrt{p}r$, with r uniformly distributed on the 1-sphere in $\mathbb{R}^p$ and with Σ having the characteristics explained in Theorem 2. This is actually Theorem 2
- Vectors of the type $p^{1/b}r$, $1 \leq b \leq 2$, where r is uniformly distributed in the 1-ℓ^b ball or sphere in $\mathbb{R}^p$. (See Ledoux (2001), Theorem 4.21, which refers to Schechtman and Zinn (2000) as the source of the theorem.)
- Vectors with log-concave density of the type $\mathrm{e}^{-U(x)}$, with the Hessian of U satisfying, for all x, $\text{Hess}(U) \geq c\text{Id}_p$, where $c > 0$ has the characteristics of $c(p)$ in the previous two lemmas: see Ledoux (2001), Theorem 2.7. Here we might also need $|||\Sigma|||_2$ to not grow too fast with d since we cannot use truncation arguments.
- Vectors with i.i.d entries with a second moment. We note that it is essentially enough in this case to deal with entries bounded by $\log(p)$ as seen in Silverstein and Bai (1995) and Silverstein (1995). See Corollary 4.10 in Ledoux (2001) for the concentration part. In this case, the analysis of concentration of quadratic forms above can be carried out for vectors of the type $\Sigma^{1/2}r$. Hence we obtain a strengthening of Lemma 3.1 in Silverstein and Bai (1995). We note that the fact the functional is convex is crucial here, whereas it would not matter for the two previous examples. In particular, a look at the proof in Silverstein and Bai (1995) shows that the strong result there of convergence of spectral distribution under the existence of only 2 moments for the random variables of interest is derived after thresholding the entries of the vectors at level $\log(p)$ and giving an argument justifying the fact that this did not change anything as far as limiting spectral distributions were concerned. If the entries are bounded by $\log(p)$, Corollary 4.10 in Ledoux (2001) gives $c(p) = K/\log(p)^2$, and hence we still have strong concentration.

Theorem 2 extends to all these distributions and our approach is one answer to the question of knowing how to handle dependence within the vectors v (see also Pajor and Pastur (2007) for related questions). We also note that using Theorem 2.4 and 3.1 in Ledoux (2001), we could, with a bit of geometric work, extend our analysis of random vectors uniformly distributed on the sphere in $\mathbb{R}^p$ to certain more general smooth Riemannian submanifolds of $\mathbb{R}^p$, answering a question which is sometimes of interest in Statistics.

Finally, let us note that the previous list contains most known results in the literature on models of the type $T_p^{1/2} X^* L_n X T_p^{1/2}$, except that we limit ourselves here to the case of diagonal L_n, and it appears that the case of L_n non-diagonal, but requiring X to have i.i.d entries is known.

4 Conclusion

We have seen that the concentration of measure phenomenon can be seen as an essential tool in the understanding of the behavior of the limiting spectral distributions of a number of random matrix models.

Motivated by applications, we have used one flavor of it to deduce from spectral properties of sample covariance matrices the corresponding properties for sample correlation matrices. On the other hand, for more complicated models, we have generalized known results about random covariance-type matrices to sample covariance matrices computed from elliptically distributed data, a type of assumptions that is popular in financial modeling. We have done it from first principles highlighting the role of concentration properties in this specific example. We have also explained that the same computations allow us to recover pretty much all known results and to obtain new results for data coming from distributions for which the dependence between entries cannot be broken up.

Very strikingly, in all the models considered the results tell us that only the covariance or the correlation between the entries of the data vector matter, and the more complicated dependence or moment structure is irrelevant as far as limiting distributions of eigenvalues are concerned.

References

Anderson, G. W. and Zeitouni, O. (2006). A CLT for a band matrix model. *Probab. Theory Related Fields* **134**, 283–338.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition.

Bai, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9**, 611–677. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

Bai, Z. D. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26**, 316–345.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32**, 553–605.

Baik, J., Ben Arous, G., and Péché, S. (2005). Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. *Ann. Probab.* **33**, 1643–1697.

Bhatia, R. (1997). *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

Bickel, P. J. and Levina, E. (2007). Regularized estimation of large covariance matrices. *The Annals of Statistics* To Appear.

Boutet de Monvel, A., Khorunzhy, A., and Vasilchuk, V. (1996). Limiting eigenvalue distribution of random matrices with correlated entries. *Markov Process. Related Fields* **2**, 607–636.

Burda, Z., Görlich, A., Jarosz, A., and Jurkiewicz, J. (2004). Signal and noise in correlation matrix. *Physica A* **343**, 295–310.

Burda, Z., Jurkiewicz, J., and Wacław, B. (2005). Spectral moments of correlated Wishart matrices. *Phys. Rev. E* **71**.

Durrett, R. (1996). *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition.

El Karoui, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and $p/n \to \infty$. *arXiv:math.ST/0309355* .

El Karoui, N. (2006). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Submitted* See `math.ST/0609418`.

El Karoui, N. (2007). Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability* **35**, 663–714.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.

Forrester, P. J. (1993). The spectrum edge of random matrix ensembles. *Nuclear Phys. B* **402**, 709–728.

Frahm, G. and Jaekel, U. (2005). Random matrix theory and robust covariance matrix estimation for financial data. *arXiv:physics/0503007* .

Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8**, 252–261.

Geronimo, J. S. and Hill, T. P. (2003). Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *J. Approx. Theory* **121**, 54–60.

Girko, V. L. (1990). *Theory of random determinants*, volume 45 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht. Translated from the Russian.

Gray, R. M. (2002). Toeplitz and circulant matrices: A review. Available at `http://ee.stanford.edu/~gray/toeplitz.pdf`.

Grenander, U. and Szegö, G. (1958). *Toeplitz forms and their applications*. California Monographs in Mathematical Sciences. University of California Press, Berkeley.

Guionnet, A. and Zeitouni, O. (2000). Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.* **5**, 119–136 (electronic).

Jiang, T. (2004). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66**, 35–48.

Johansson, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209**, 437–476.

Johnstone, I. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.* **29**, 295–327.

Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12**, 1–38.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88**, 365–411.

Ledoux, M. (2001). *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Li, L., Tulino, A. M., and Verdú, S. (2004). Design of reduced-rank MMSE multiuser detectors using random matrix methods. *IEEE Trans. Inform. Theory* **50**, 986–1008.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72 (114)**, 507–536.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ. Concepts, techniques and tools.

Pajor, A. and Pastur, L. (2007). On the limiting empirical measure of the sum of rank one matrices with log-concave distribution Available at `http://www.arxiv.org/abs/0710.1346`.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica* **17**. Available at `http://anson.ucdavis.edu/~debashis/techrep/techrep.html`.

Paul, D. and Silverstein, J. (2007). No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix Available at `http://www4.ncsu.edu/~jack/pub.html`.

Rao, N. R., Mingo, J., Speicher, R., and Edelman, A. (2007). Statistical eigen-inference from large Wishart matrices Available at `arXiv:math/0701314`.

Schechtman, G. and Zinn, J. (2000). Concentration on the l_p^n ball. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pp. 245–256. Springer, Berlin.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55**, 331–339.

Silverstein, J. W. and Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54**, 175–192.

Tulino, A. and Verdú, S. (2004). *Random matrix theory and wireless communications*, volume 1 of *Foundations and Trends in Communications and Information Theory*. Now Publishers.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Voiculescu, D. (2000). Lectures on free probability theory. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, volume 1738 of *Lecture Notes in Math.*, pp. 279–349. Springer, Berlin.

Yin, Y. Q., Bai, Z. D., and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78**, 509–521.

Zhang, L. (2006). *Spectral Analysis of Large Dimensional Random Matrices*. Ph.D. thesis, National University of Singapore.