

Statistical eigen-inference from large Wishart matrices

N. RAJ RAO JAMES MINGO ROLAND SPEICHER ALAN EDELMAN

Abstract

The asymptotic behavior of the eigenvalues of a sample covariance matrix is described when the observations are from a zero mean multivariate (real or complex) normal distribution whose covariance matrix has population eigenvalues of arbitrary multiplicity. In particular, the asymptotic normality of the fluctuation in the trace of powers of the sample covariance matrix from the limiting quantities is shown. Concrete algorithms for analytically computing the limiting quantities and the covariance of the fluctuations are presented. Tests of hypotheses for the population eigenvalues are developed and a technique for inferring the population eigenvalues (without requiring any assumptions on the population eigenvectors) is proposed that exploits this asymptotic normality of the trace of powers of the sample covariance matrix. Numerical simulations demonstrate the robustness of the proposed techniques in techniques in high-dimensional, (relatively) small sample size settings and the superiority over alternate procedures found in the literature for the special cases where a direct comparison can be made. The improved performance is a consequence of the fact that the proposed inference procedures are “global” (in a sense that we describe) and exploit “global” information thereby overcoming the inherent biases that cripple classical “local” inference procedures which rely on “local” information.

1. Introduction

Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ be a $p \times n$ data matrix where $\mathbf{x}_1, \dots, \mathbf{x}_n$, denote n independent measurements, where for each i , \mathbf{x}_i has an p -dimensional (real or complex) Gaussian distribution with mean zero, and positive definite covariance matrix $\mathbf{\Sigma}$. When the samples are complex, the real and imaginary components are assumed to be independent, identically distributed zero mean Gaussian vectors with a covariance of $\mathbf{\Sigma}/2$. The sample covariance matrix (SCM) when formed from these n samples as

$$\mathbf{S} := \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' = \frac{1}{n} \mathbf{X} \mathbf{X}', \quad (1.1)$$

has the (central) Wishart distribution [Wishart, 1928]. We focus on inference problems for parameterized covariance matrices modelled as $\mathbf{\Sigma}_{\boldsymbol{\theta}} = \mathbf{U} \mathbf{\Lambda}_{\boldsymbol{\theta}} \mathbf{U}'$ where

$$\mathbf{\Lambda}_{\boldsymbol{\theta}} = \begin{bmatrix} a_1 \mathbf{I}_{p_1} & & & & \\ & a_2 \mathbf{I}_{p_2} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & a_k \mathbf{I}_{p_k} \end{bmatrix}, \quad (1.2)$$

where $a_1 > \dots > a_k$ and $\sum_{j=1}^k p_j = p$. Defining $t_i = p_i/p$, allows us to conveniently express the $2k - 1$ dimensional parameter vector as $\boldsymbol{\theta} = (t_1, \dots, t_{k-1}, a_1, \dots, a_k)$ with the obvious non-negativity constraints on the elements.

Models of the form in (1.2) arise as a special case whenever the measurements are of the form

$$\mathbf{x}_i = \mathbf{A} \mathbf{s}_i + \mathbf{z}_i \quad \text{for } i = 1, \dots, n \quad (1.3)$$

where $\mathbf{z}_i \sim \mathcal{N}_p(0, \boldsymbol{\Sigma}_z)$, denotes an p -dimensional (real or complex) Gaussian noise vector with covariance $\boldsymbol{\Sigma}_z$, $\mathbf{s}_i \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}_s)$ denotes a k -dimensional zero mean (real or complex) Gaussian signal vector with covariance $\boldsymbol{\Sigma}_s$, and \mathbf{A} is a $p \times k$ unknown non-random matrix. In array processing applications, the j -th column of the matrix \mathbf{A} encodes the parameter vector associated with the j -th signal whose magnitude is described by the j -th element of \mathbf{s}_i .

Since the signal and noise vectors are independent of each other, the covariance matrix of \mathbf{x}_i can hence be decomposed as

$$\boldsymbol{\Sigma} = \boldsymbol{\Psi} + \boldsymbol{\Sigma}_z \quad (1.4)$$

where $\boldsymbol{\Sigma}_z$ is the covariance of \mathbf{z} and $\boldsymbol{\Psi} = \mathbf{A}\boldsymbol{\Sigma}_s\mathbf{A}'$ with $'$ denoting the conjugate transpose. One way of obtaining $\boldsymbol{\Sigma}$ with eigenvalues of the form in (1.2) is when $\boldsymbol{\Sigma}_z = \sigma^2 \mathbf{I}$ so that the $n - k$ smallest eigenvalues of $\boldsymbol{\Sigma}$ are equal to σ^2 . Then, if the matrix \mathbf{A} is of full column rank so and the covariance matrix of the signals $\boldsymbol{\Sigma}_s$ is nonsingular, the $p - k$ smallest eigenvalues of $\boldsymbol{\Psi}$ are equal to zero so that the eigenvalues of $\boldsymbol{\Sigma}$ will be of the form in (1.2). Alternately, if the eigenvalues of $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}_z$ have the identical subspace structure, *i.e.*, in (1.2), $t_i^\Psi = t_i^{\boldsymbol{\Sigma}_z}$ for all i , then whenever the eigenvectors associated with each of the subspaces of $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}_z$ align, the eigenvalues of $\boldsymbol{\Sigma}$ will have the subspace structure in (1.2).

Additionally, from an identifiability point of view, as shall be discussed in Section 6., if the practitioner has reason to believe that the population eigenvalues can be split into several clusters about $a_i \pm \sqrt{p/n}$, then the use of the model in (1.2) with a block subspace structure will also be justified.

1.1 Inferring the population eigenvalues from the sample eigenvalues

While inference problems for these models have been documented in texts such as [Muirhead, 1982], the inadequacies of classical algorithms in high-dimensional, (relatively) small sample size settings have not been adequately addressed. We highlight some of the prevalent issues in the context of statistical inference and hypothesis testing.

Anderson's landmark paper [Anderson, 1963] develops the theory that describes the (large sample) asymptotics of the sample eigenvalues (in the real valued case) for such models when the true covariance matrix has eigenvalues of arbitrary multiplicity. Indeed, for arbitrary covariance $\boldsymbol{\Sigma}$, the joint density function of the eigenvalues l_1, \dots, l_p of the SCM \mathbf{S} when $n > p + 1$ is shown to be given by

$$\tilde{Z}_{p,n}^\beta \sum_{i=1}^p l_i^{\beta(n-p+1)/2-1} \prod_{i < j} |l_i - l_j|^\beta \int_{\mathbf{Q}} \exp\left(-\frac{n\beta}{2} \text{Tr}(\boldsymbol{\Sigma}^{-1} \mathbf{Q} \mathbf{S} \mathbf{Q}')\right) d\mathbf{Q} \quad (1.5)$$

where $l_1 > \dots > l_p > 0$, $\tilde{Z}_{p,n}^\beta$ is a normalization constant, and $\beta = 1$ (or 2) when \mathbf{S} is real (resp. complex). In (1.5), $\mathbf{Q} \in \mathbf{O}(p)$ when $\beta = 1$ while $\mathbf{Q} \in \mathbf{U}(p)$ when $\beta = 2$ where $\mathbf{O}(p)$ and $\mathbf{U}(p)$ are, respectively, the set of $p \times p$ orthogonal and unitary matrices with Haar measure. Anderson notes that

If the characteristic roots of $\boldsymbol{\Sigma}$ are different, the deviations ... from the corresponding population quantities are asymptotically normally distributed. When some of the roots of $\boldsymbol{\Sigma}$ are equal, the asymptotic distribution cannot be described so simply.

Indeed, the difficulty alluded to, arises due to the presence of the integral over orthogonal (or unitary) group on the right hand side of (1.5). This problem is compounded in situations when some of the eigenvalues of $\boldsymbol{\Sigma}$ are equal as is the case for the model considered in (1.2). Nonetheless, Anderson is able to use the (large sample) asymptotics to derive the maximum likelihood estimate of the population eigenvalues, a_l , as

$$\hat{a}_l \approx \frac{1}{p_l} \sum_{j \in N_l} \hat{\lambda}_j \quad \text{for } l = 1, \dots, k, \quad (1.6)$$

where $\hat{\lambda}_j$ are the sample eigenvalues (arranged in descending order) and N_l is the set of integers $p_1 + \dots + p_{l-1} + 1, \dots, p_1 + \dots + p_l$. This is a reasonable estimator that works well in practice when $n \gg p$. The large

sample size asymptotics are, however, of limited utility because they ignore the (significant) effect of the dimensionality of the system on the behavior of the sample eigenvalues.

Consequently, (large sample size) asymptotic predictions, derived under the p fixed, $n \rightarrow \infty$ regime do not account for the additional complexities that arise in situations where the sample size n is large but the dimensionality p is of comparable order. Furthermore, the estimators developed using the classical large sample asymptotics invariably become degenerate whenever $p < n$, so that $p - n$ of the sample eigenvalues will identically equal to zero. For example, when $n = p/2$, and there are two distinct population eigenvalues each with multiplicity $p/2$ then the estimate of the smallest eigenvalue using (1.6) will be zero. Other such scenarios where the population eigenvalue estimates obtained using (1.6) are meaningless are easy to construct and are practically relevant in many applications such as radar and sonar signal processing [Trees, 2002, Smith, 2005], and many more.

There are, of course, other strategies one may employ for inferring the population eigenvalues. One might consider a maximum-likelihood technique based on maximizing the log-likelihood function of the observed data X which is given by (ignoring constants)

$$l(\mathbf{X}|\boldsymbol{\Sigma}) := -n(\text{tr } \mathbf{S}\boldsymbol{\Sigma}^{-1} + \log \det \boldsymbol{\Sigma}),$$

or, equivalently, when $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$, by minimizing the objective function

$$h(\mathbf{X}|\mathbf{U}, \boldsymbol{\Lambda}) = (\text{tr } \mathbf{S}\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}' + \log \det \boldsymbol{\Lambda}). \tag{1.7}$$

What should be apparent on inspecting (1.7) is that the maximum-likelihood estimation of the parameters of $\boldsymbol{\Lambda}$ of the form in (1.2) requires us to model the population eigenvectors \mathbf{U} as well (except when $k = 1$). If \mathbf{U} were known apriori, then an estimate of a_l obtained as

$$\hat{a}_l \approx \frac{1}{p_l} \sum_{j \in N_l} (\mathbf{U}'\mathbf{S}\mathbf{U})_{j,j} \quad \text{for } l = 1, \dots, k. \tag{1.8}$$

N_l is the set of integers $p_1 + \dots + p_{l-1} + 1, \dots, p_1 + \dots + p_l$ will provide a good estimate. In practical applications, the population eigenvectors might either be unknown or be misspecified leading to faulty inference. Hence it is important to have the ability to perform statistically sound, computationally feasible *eigen-inference* of the population eigenvalues, *i.e.*, from the sample eigenvalues alone, in a manner that is robust to high-dimensionality and sample size constraints.

We illustrate the difficulties encountered in high-dimensional settings with an example (summarized in Figure 1) of a SCM constructed from a covariance matrix modelled as $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$ with $p = 80$ and sample size $n = 160$. Half of the eigenvalues of $\boldsymbol{\Lambda}$ are of magnitude 3 while the remainder are of magnitude 1. The sample eigenvalues are significantly *blurred*, relative to the true eigenvalues as shown in Figure 1(a). Figures 1(b), and 1(d) plot the sample eigenvectors for the case when the true eigenvectors $\mathbf{U} = \mathbf{I}$, and an arbitrary \mathbf{U} , respectively. Figures 1(c) and 1(e) plot the diagonal elements $(\mathbf{S})_{j,j}$. Thus, if the true eigenvector was indeed $\mathbf{U} = \mathbf{I}$ then an estimate of the population eigenvalues formed as in (1.8) yields a good estimate; when $\mathbf{U} \neq \mathbf{I}$, however, the estimate is very poor.

1.2 Testing for equality of population eigenvalues

Similar difficulties are encountered in problems of testing as well. In such situations, Anderson proposes the likelihood ratio criterion for testing the hypothesis

$$\lambda_{p_1+\dots+p_{l-1}+1} = \lambda_{p_1+\dots+p_{l-1}+1, \dots, p_1+\dots+p_l}$$

given by

$$\left[\prod_{j \in N_l} \hat{\lambda}_j / (p_k^{-1} \sum_{j \in N_l} \hat{\lambda}_j)^{p_k} \right]^{\frac{1}{2}n} \quad \text{for } l = 1, \dots, k, \tag{1.9}$$

where $\widehat{\lambda}_j$ are the sample eigenvalues (arranged in descending order) and N_l is the set of integers $p_1 + \dots + p_{l-1} + 1, \dots, p_1 + \dots + p_l$. The test in (1.9) suffers from the same deficiency as the population eigenvalue estimator in (1.6) - it becomes degenerate when $p > n$. When the population eigenvectors \mathbf{U} are known, (1.9) may be modified by forming the criterion

$$\left[\prod_{j \in N_l} (\mathbf{U}' \mathbf{S} \mathbf{U})_{j,j} / (p_k^{-1} \sum_{j \in N_l} (\mathbf{U}' \mathbf{S} \mathbf{U})_{j,j})^{p_k} \right]^{\frac{1}{2}n} \quad \text{for } l = 1, \dots, k, \quad (1.10)$$

where N_l is the set of integers $p_1 + \dots + p_{l-1} + 1, \dots, p_1 + \dots + p_l$. When the eigenvectors are misspecified the inference provided will be faulty. For the earlier example, Figure 1(e) illustrates this for the case when it is assumed that the population eigenvectors are \mathbf{I} when they are really $\mathbf{U} \neq \mathbf{I}$. Testing the hypothesis $\Sigma = \Sigma_0$, reduces to testing the null hypothesis $\Sigma = \mathbf{I}$ when the transformation $\widetilde{\mathbf{x}}_i = \Sigma_0^{-1/2} \mathbf{x}_i$ is applied. The robustness of tests for sphericity in high dimensional settings has been extensively discussed in Ledoit and Wolf [2002] and is the focus of the work in Srivastava [2005, 2006].

1.3 Proposed statistical eigen-inference techniques

In this article our focus is on developing population eigenvalue estimation and testing algorithms for models of the form in (1.2) that are robust to high-dimensionality, sample size constraints and population eigenvector misspecification. We are able to develop such computationally feasible algorithms by exploiting the properties of the eigenvalues of large Wishart matrices. These results, analytically describe the non-random *blurring* of the sample eigenvalues, relative to the population eigenvalues, in the $p, n(p) \rightarrow \infty$ limit while compensating for the random *fluctuations* about the limiting behavior due to finite dimensionality effects. This allows us to handle the situation where the sample eigenvalues are blurred to the point that the block subspace structure of the population eigenvalues cannot be visually discerned, as in Figure 1(a), thereby extending the “signal” detection capability beyond the special cases tackled in Silverstein and Combettes [1992]. The nature of the mathematics being exploited makes them robust to the high-dimensionality and sample size constraints while the reliance on the sample eigenvalues alone makes them insensitive to any assumptions on the population eigenvectors. In such situations where the eigenvectors are accurately modelled, the practitioner may use the proposed methodologies to complement and “robustify” the inference provided by estimation and testing methodologies that exploit the eigenvector structure.

We consider testing the hypothesis for the equality of the population eigenvalues and statistical inference about the population eigenvalues. In other words, for some unknown \mathbf{U} , if $\Sigma_0 = \mathbf{U} \mathbf{\Lambda}_{\theta_0} \mathbf{U}'$ where $\mathbf{\Lambda}_{\theta}$ is modelled as in (1.2), techniques to 1) test if $\Sigma = \Sigma_0$, and 2) estimate θ_0 are summarized in Table 1. We note that inference on the population eigenvalues is performed using the *entire sample eigen-spectrum* unlike (1.6) and (1.9). This reflects the inherent non-linearities of the sample eigenvalue blurring induced by high-dimensionality and sample size constraints.

Table 2 compares the bias and mean square error of various techniques of estimating the non-unity population eigenvalue in Figure 1 when the block structure is known a priori, *i.e.*, when $t_1 = t_2 = 0.5$, and $a_2 = 1$ are known and $a := a_1$ is unknown. The first two columns refer to the procedure in (1.8) where the correct population eigenvectors $\mathbf{U} \neq \mathbf{I}$ are used, the third column refers to Anderson’s procedure in (1.6) while the fourth column refers to the procedure in (1.8) where $\mathbf{U} = \mathbf{I}$ is used instead of the population eigenvectors. The last two columns refer to the proposed statistical eigen-inference (SEI) technique in Table 1 with $\theta := a$, $v(\theta) = \text{Tr} S - p(0.5a + 0.5)$, and $Q_{\theta} = (1/2 a^2 + 1/2 a^2 c + ac + 1/2 + 1/2 c - a) c^2$ where $c = p/n$. Note that though the SEI techniques do not exploit any eigenvector information, its performance compares favorably to the maximum likelihood technique that does. As for the other techniques it is evident that the inherent biases in the problem cripple the estimators.

An important implication of this in practice is that in high dimensional, sample size starved settings, local inference, performed on a subset of sample eigenvalues alone, that fails to take into account the global structure (*i.e.*, by modelling the remaining eigenvalues) is likely to be inaccurate, or worse misleading. In

Testing:	$H_{\theta_0} : h(\boldsymbol{\theta}) := \mathbf{v}_{\boldsymbol{\theta}}^T \mathbf{Q}_{\boldsymbol{\theta}}^{-1} \mathbf{v}_{\boldsymbol{\theta}} \sim \chi_2^2 \geq \gamma,$	$q = \dim(\mathbf{v}_{\boldsymbol{\theta}}) = 2$
Estimation:	$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \{ \mathbf{v}_{\boldsymbol{\theta}}^T \mathbf{Q}_{\boldsymbol{\theta}}^{-1} \mathbf{v}_{\boldsymbol{\theta}} + \log \det \mathbf{Q}_{\boldsymbol{\theta}} \},$	$q = \dim(\mathbf{v}_{\boldsymbol{\theta}}) \geq \dim(\boldsymbol{\theta})$
Legend:	$(\mathbf{v}_{\boldsymbol{\theta}})_j = p \times \left(\frac{1}{p} \text{Tr} S^j - E \left[\frac{1}{p} \text{Tr} S^j \right] \right),$	$j = 1, \dots, q$
	$\mathbf{Q}_{\boldsymbol{\theta}} = \text{cov} [\mathbf{v}_{\boldsymbol{\theta}} \mathbf{v}_{\boldsymbol{\theta}}']$	

Table 1: Structure of proposed algorithms.

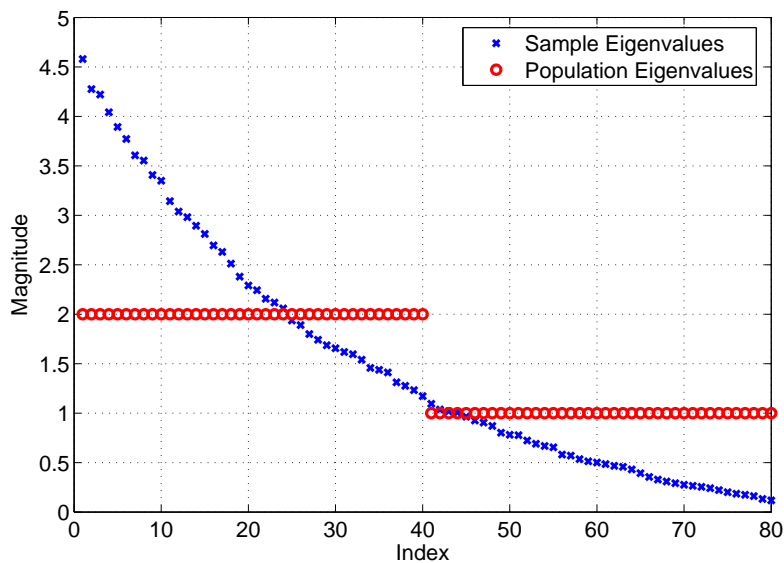
such settings, practitioners are advised to consider tests (such as the ones proposed) for the equality of the entire population eigen-spectrum instead of testing for the equality of individual population eigenvalues.

We view the inference techniques developed herein as the first step in the development of improved high-dimensional covariance matrix estimation algorithms. The issue of inverse covariance matrix estimation which Srivastava [2007] examines in the context of discriminant analysis is also related.

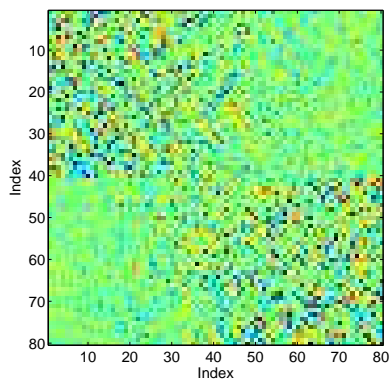
The approach we have in mind differs from the (sample eigenvalue) shrinkage based techniques in [Haff, 1980, Dey and Srinivasan, 1985] in a crucial regard. Our perspective is that the eigenvalues and the eigenvectors (or subspaces) of the sample covariance matrices are blurred relative to the population eigenvalues and eigenvectors (or subspaces), respectively. For the model considered in this article, the precise analytical characterization of the blurring of the eigenvalues (Theorem 2.7) allows us to formulate and solve the deblurring problem. The tools from free probability are applied in [Nadakuditi, 2007] to precisely describe the blurring of the population eigenvectors (or subspaces) as well. The answer is encoded in the form of a conditional eigenvector “distribution” that explicitly takes into account the dimensionality of the system and the sample size available - the conditioning is with respect to the population eigenvalues. The idea that the covariance matrix estimate thus constructed from the deblurred eigenvalues and eigenvectors should be significantly better has merit. The development of computationally realizable eigenvector deblurring algorithms is a significant obstacle to progress along this direction of research.

1.4 Related work

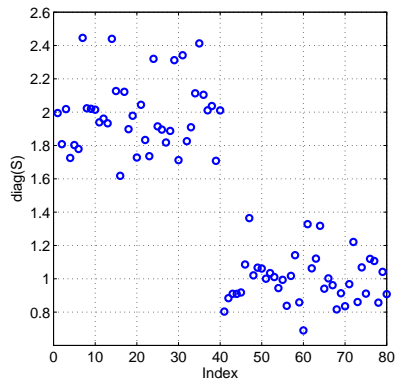
There are other alternatives found in the literature to the block subspace hypothesis testing problem considered in this article. [El Karoui, 2005] provides a test for the largest eigenvalue for a large class of complex Wishart matrices including those with a population covariance matrix of the form in (1.2). Though the results are stated for the case when $p < n$, simulations confirm the validity of the techniques to the general case when $p < n$ and for real Wishart matrices. El Karoui’s tests can be classified as a local test that utilizes global information, *i.e.*, information about the entire (assumed) population eigen-spectrum. Testing is performed by computing the largest eigenvalue of the sample covariance matrix, recentering, rescaling it and rejecting the hypothesis if it is too large. The recentering and rescaling parameters are determined by the a_i and t_i values in (1.2) while the threshold is determined by the quantiles of the appropriate (real or complex) Tracy-Widom distribution. A disadvantage of this procedure is the great likelihood whenever recentering by the false parameter pushes the test statistic towards the left tail of the distribution. Consequently, the identity covariance hypothesis will be accepted with great likelihood whenever the recentering and rescaling coefficients are calculated for the model in (1.2) with $a_i > 1$. The proposed global test based on global information avoids this pitfall and is based on distributional results for the traces of powers of Wishart matrices that also appear in Srivastava [2005]. The issue of whether a local test or a global test is more powerful is important and highlighted using simulations in the context of a joint estimation and testing problem in Section 6., its full resolution is beyond the scope of this article.



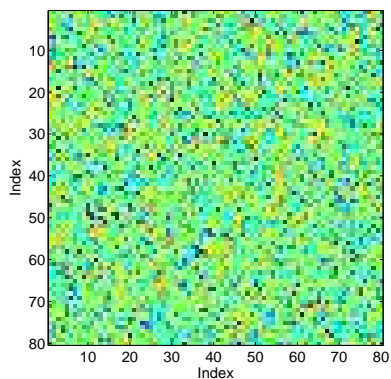
(a) Sample eigenvalues versus true eigenvalues ($p = 80, n = 160$).



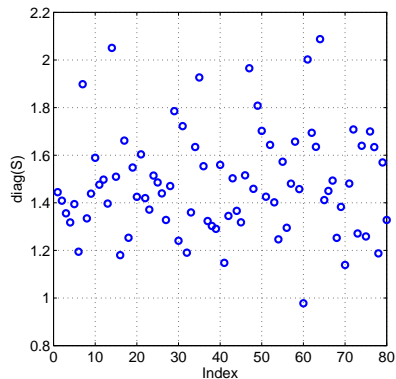
(b) Sample eigenvectors when $\mathbf{U} = \mathbf{I}$.



(c) Diagonal elements of \mathbf{S} when $\mathbf{U} = \mathbf{I}$.



(d) Sample eigenvectors for arbitrary \mathbf{U} .



(e) Diagonal elements of \mathbf{S} for arbitrary \mathbf{U} .

Figure 1: The challenge of estimating the population eigenvalues from the sample eigenvalues in high-dimensional settings.

(a) Bias.

p	n	Known \mathbf{U} Max Like.	Known \mathbf{U} Max Like. $\times p^2$	Anderson	Unknown \mathbf{U} Max Like.	SEI	SEI $\times p^2$
10	20	0.0117	0.1168	-1.9994	-0.5811	-0.0331	-0.3308
20	40	0.0000	0.0001	-1.9994	-0.5159	-0.0112	-0.2244
40	80	0.0008	0.0301	-1.9994	-0.5245	-0.0019	-0.0776
80	160	-0.0003	-0.0259	-1.9994	-0.4894	-0.0003	-0.0221
160	320	0.0000	0.0035	-1.9994	-0.4916	-0.0003	-0.0411
320	640	0.0001	0.0426	-1.9994	-0.5015	0.0001	0.0179

(b) MSE.

p	n	Known \mathbf{U} Max Like.	Known \mathbf{U} Max Like. $\times p^2$	Anderson	Unknown \mathbf{U} Max Like.	SEI	SEI $\times p^2$
10	20	0.0380	3.7976	3.9990	0.3595	0.0495	4.9463
20	40	0.0100	3.9908	3.9990	0.2722	0.0126	5.0256
40	80	0.0025	3.9256	3.9991	0.2765	0.0030	4.8483
80	160	0.0006	4.1118	3.9991	0.2399	0.0008	5.1794
160	320	0.0002	4.1022	3.9990	0.2417	0.0002	5.0480
320	640	0.0000	4.0104	3.9990	0.2515	0.0000	5.0210

Table 2: Comparison of performance of different techniques for estimating the non-unity population eigenvalue in Figure 1 when the block structure is known apriori.

Silverstein and Combettes [1992] consider the situation when the sample eigenvalues discernibly split into distinct clusters and suggest that the proportion of the eigenvalues in each cluster will provide a good estimate of the parameters a_i in (1.2). The nature of the distributional results in Bai and Silverstein [1998] imply that whenever the sample eigenvalues are thus clustered, then for large enough p , the estimate of a_i thus obtained will be exactly equal to true value. Such a procedure could not, however be applied for situations as such those depicted in Figure 1(a) where the sample eigenvalues do not separate into clusters. Silverstein and Combettes [1992] does not provide a strategy for computing the t_i in (1.2) once the a_i is computed - the proposed techniques fill the void.

A semi-parametric, grid-based technique for inferring the empirical distribution function of the population eigenvalues from the sample eigen-spectrum was proposed in El Karoui [2006]. The procedure described can be invaluable to the practitioner in the initial data exploration stage by providing a good estimate of the number of blocks in (1.2) and a less refined estimate of the underlying a_i and t_i associated with each block. Our techniques can then be used to improve or test the estimates.

1.5 Outline

The rest of this article is organized as follows. In Section 2. we introduce the necessary definitions and summarize the relevant theorems. The eigen-inference techniques are developed in Section 4.. Concrete algorithms for computing the analytic expectations that appear in the algorithms summarized in Table 1) are presented in Section 3.. The performance of the algorithms is illustrated using Monte-Carlo simulations in Section 5.. Some concluding remarks are presented in Section 7.. In Section 8. , we outline the relevance of free probability theory for the considered problem and provide a proof of Theorem 2.7.

2. Preliminaries

Definition 2.1. Let $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ be an $N \times N$ matrix with real eigenvalues. The j -th sample moment is defined as

$$\mathrm{tr}(\mathbf{A}^j) := \frac{1}{N} \mathrm{Tr}(\mathbf{A}^j).$$

where Tr is the usual un-normalized trace.

Definition 2.2. Let $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ be a sequence of self-adjoint $N \times N$ -random matrices. If the limit of all moments defined as

$$\alpha_j^A := \lim_{N \rightarrow \infty} \mathbb{E}[\mathrm{tr}(\mathbf{A}_N^j)] \quad (N \in \mathbb{N})$$

exists then we say that \mathbf{A} has a limit eigenvalue distribution.

Notation 2.3. For a random matrix \mathbf{A} with a limit eigenvalue distribution we denote by $M_A(x)$ the moment power series, which we define by

$$M_A(x) := 1 + \sum_{j \geq 1} \alpha_j^A x^j.$$

Notation 2.4. For a random matrix ensemble \mathbf{A} with limit eigenvalue distribution we denote by $g_A(x)$ the corresponding Cauchy-transform, which we define as formal power series by

$$g_A(x) := \lim_{N \rightarrow \infty} \mathbb{E}\left[\frac{1}{N} \mathrm{Tr}(x\mathbf{I}_N - \mathbf{A}_N)^{-1}\right] = \frac{1}{x} M_A(1/x).$$

Definition 2.5. Let $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ be a self-adjoint random matrix ensemble. We say that it has a second order limit distribution if for all $i, j \in \mathbb{N}$ the limits

$$\alpha_j^A := \lim_{N \rightarrow \infty} k_1(\mathrm{tr}(\mathbf{A}_N^j))$$

and

$$\alpha_{i,j}^A := \lim_{N \rightarrow \infty} k_2(\mathrm{Tr}(\mathbf{A}_N^i), \mathrm{Tr}(\mathbf{A}_N^j))$$

exist and if

$$\lim_{N \rightarrow \infty} k_r(\mathrm{Tr}(\mathbf{A}_N^{j(1)}), \dots, \mathrm{Tr}(\mathbf{A}_N^{j(r)})) = 0$$

for all $j \geq 3$ and all $j(1), \dots, j(r) \in \mathbb{N}$. In this definition, we denote the (classical) cumulants by k_n . Note that k_1 is just the expectation, and k_2 the covariance.

Notation 2.6. When $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ has a limit eigenvalue distribution, then the limits $\alpha_j^A := \lim_{N \rightarrow \infty} \mathbb{E}[\mathrm{tr}(\mathbf{A}_N^j)]$ exist. When \mathbf{A}_N has a second order limit distribution, the fluctuation

$$\mathrm{tr}(\mathbf{A}_N^j) - \alpha_j^A$$

is asymptotically Gaussian of order $1/N$. We consider the second order covariances defined as

$$\alpha_{i,j}^A := \lim_{N \rightarrow \infty} \mathrm{cov}(\mathrm{Tr}(\mathbf{A}_N^i), \mathrm{Tr}(\mathbf{A}_N^j)),$$

and denote by $\mathcal{M}_A(x, y)$ the second order moment power series, which we define by:

$$\mathcal{M}_A(x, y) := \sum_{i, j \geq 1} \alpha_{i,j}^A x^i y^j.$$

Theorem 2.7. *Assume that the $p \times p$ (non-random) covariance matrix $\Sigma = (\Sigma_p)_{p \in \mathbb{N}}$ has a limit eigenvalue distribution. Let \mathbf{S} be the (real or complex) sample covariance matrix formed from the n samples as in (1.1). Then for $p, n \rightarrow \infty$ with $p/n \rightarrow c \in (0, \infty)$, \mathbf{S} has both a limit eigenvalue distribution and a second order limit distribution. The Cauchy transform of the limit eigenvalue distribution, $g(x) \equiv g_S(x)$, satisfies the equation:*

$$g(x) = \frac{1}{1 - c + cxg(x)} g_\Sigma \left(\frac{x}{1 - c + cxg(x)} \right), \quad (2.1)$$

with the corresponding power series $M_S(x) = 1/x g_S(1/x)$. Define $\tilde{\mathbf{S}} = \frac{1}{n} \mathbf{X}' \mathbf{X}$ so that its moment power series is given by

$$M_{\tilde{\mathbf{S}}}(y) = c(M_S(z) - 1) + 1. \quad (2.2)$$

The second order moment generating series is given by

$$\mathcal{M}_S(x, y) = \mathcal{M}_{\tilde{\mathbf{S}}}(x, y) = \frac{2}{\beta} \mathcal{M}_{\tilde{\mathbf{S}}}^\infty(x, y) \quad (2.3a)$$

where

$$\mathcal{M}_{\tilde{\mathbf{S}}}^\infty(x, y) = xy \left(\frac{\frac{d}{dx}(xM_{\tilde{\mathbf{S}}}(x)) \cdot \frac{d}{dy}(yM_{\tilde{\mathbf{S}}}(y))}{(xM_{\tilde{\mathbf{S}}}(x) - yM_{\tilde{\mathbf{S}}}(y))^2} - \frac{1}{(x-y)^2} \right) \quad (2.3b)$$

where β equals 1 (or 2) when the elements of S are real (or complex).

PROOF. See Appendix 8.. \square

3. Computational aspects

Proposition 3.1. *For $\Sigma_\theta = \mathbf{U} \Lambda_\theta \mathbf{U}'$ as in (1.2), let $\theta = (t_1, \dots, t_{k-1}, a_1, \dots, a_k)$ where $t_i = p_i/p$. Then \mathbf{S} has a limit eigenvalue distribution as well as a second order limit distribution. The moments α_j^S , and hence $\alpha_{i,j}^S$, depend on θ and c . Let \mathbf{v}_θ be a q -by-1 vector whose j -th element is given by*

$$(\mathbf{v}_\theta)_j = \text{Tr} \mathbf{S}^j - p \alpha_j^S.$$

Then for large p and n ,

$$\mathbf{v}_\theta \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{Q}_\theta) \quad (3.1)$$

where $\boldsymbol{\mu}_\theta = 0$ if \mathbf{S} is complex and $(\mathbf{Q}_\theta)_{i,j} = \alpha_{i,j}^S$.

PROOF. This follows directly from Theorem 2.7. From (3.2) and (3.4), the moments α_k^S depend on α^Σ and $c = p/n$ and hence on the unknown parameter vector θ . The existence of the non-zero mean when \mathbf{S} is real follows from the statement in [Bai and Silverstein, 2004]. \square

3.1 Computation of moments of limiting eigenvalue distribution

Equation (2.1) expresses the relationship between the moment power series of Σ and that of \mathbf{S} via the limit of the ratio p/n . We can hence express the expected moments of \mathbf{S} in terms of the moments of Σ . The general form of the moments of $\tilde{\mathbf{S}}$, given by Corollary 9.12 in [?, pp.143], is

$$\alpha_j^{\tilde{\mathbf{S}}} = \sum_{\substack{i_j \geq 0 \\ 1i_1 + 2i_2 + 3i_3 + \dots + ji_j = j}} c^{i_1 + i_2 + \dots + i_j} (\alpha_1^\Sigma)^{i_1} (\alpha_2^\Sigma)^{i_2} \dots (\alpha_j^\Sigma)^{i_j} \cdot \gamma_{i_1, i_2, \dots, i_j}^{(j)}, \quad (3.2)$$

where $\gamma_{i_1, \dots, i_j}^{(j)}$ is the multinomial coefficient given by

$$\gamma_{i_1, i_2, \dots, i_j}^{(j)} = \frac{j!}{i_1! i_2! \dots i_j! (j + 1 - (i_1 + i_2 + \dots + i_j))!} \quad (3.3)$$

The multinomial coefficient in (3.3) has an interesting combinatorial interpretation. Let j a positive integer, and let $i_1, \dots, i_j \in \mathbb{N} \cup \{0\}$ be such that $i_1 + 2i_2 + \dots + ji_j = j$. The number of non-crossing partitions $\pi \in NC(j)$ which have i_1 blocks with 1 element, i_2 blocks with 2 elements, \dots , i_j blocks with j elements is given by the multinomial coefficient $\gamma_{i_1, \dots, i_j}^j$.

The moments of $\tilde{\mathbf{S}}$ are related to the moments of \mathbf{S} as

$$\alpha_j^{\tilde{\mathbf{S}}} = c \alpha_j^{\mathbf{S}} \quad \text{for } j = 1, 2, \dots \quad (3.4)$$

We can use (3.2) to compute the first few moments of \mathbf{S} in terms of the moments of Σ . This involves enumerating the partitions that appear in the computation of the multinomial coefficient in (3.3). For $j = 1$ only $i_1 = 1$ contributes with $\gamma_1^{(1)} = 1$, thus.

$$\alpha_1^{\tilde{\mathbf{S}}} = c \alpha_1^{\Sigma} \quad (3.5)$$

For $n = 2$ only $i_1 = 2, i_2 = 0$ and $i_1 = 0, i_2 = 1$ contribute with

$$\gamma_{2,0}^{(2)} = 1, \quad \gamma_{0,1}^{(2)} = 1,$$

and thus

$$\alpha_2^{\tilde{\mathbf{S}}} = c \alpha_2^{\Sigma} + c^2 (\alpha_1^{\Sigma})^2 \quad (3.6)$$

For $n = 3$ we have three possibilities for the indices, contributing with

$$\gamma_{3,0,0}^{(3)} = 1, \quad \gamma_{1,1,0}^{(3)} = 3, \quad \gamma_{0,0,1}^{(3)} = 1,$$

thus

$$\alpha_3^{\tilde{\mathbf{S}}} = c \alpha_3^{\Sigma} + 3c^2 \alpha_1^{\Sigma} \alpha_2^{\Sigma} + c^3 (\alpha_1^{\Sigma})^3 \quad (3.7)$$

For $n = 4$ we have five possibilities for the indices, contributing with

$$\gamma_{4,0,0,0}^{(4)} = 1, \quad \gamma_{2,1,0,0}^{(4)} = 6, \quad \gamma_{0,2,0,0}^{(4)} = 2, \quad \gamma_{1,0,1,0}^{(4)} = 4, \quad \gamma_{0,0,0,1}^{(4)} = 1$$

thus

$$\alpha_4^{\tilde{\mathbf{S}}} = c \alpha_4^{\Sigma} + 4c^2 \alpha_1^{\Sigma} \alpha_3^{\Sigma} + 2c^2 (\alpha_2^{\Sigma})^2 + 6c^3 (\alpha_1^{\Sigma})^2 \alpha_2^{\Sigma} + c^4 (\alpha_1^{\Sigma})^4. \quad (3.8)$$

For specific instances of Σ , we simply plug in the moments α_i^{Σ} into the above expressions to get the corresponding moments of \mathbf{S} . The general formula in (3.2) can be used to generate the expressions for higher order moments as well though such an explicit enumeration will be quite tedious even if symbolic software is used.

An alternate method is to use the software package RMTTool [Rao] based on the ‘‘polynomial method’’ developed in the second part of the first author’s dissertation [Nadakuditi, 2007]. The software enables the moments of \mathbf{S} to be enumerated rapidly whenever the moment power series of Σ is an algebraic power series, *i.e.*, it is the solution of an algebraic equation. This is always the case when Σ is of the form in (1.2). For example, if $\theta = (t_1, t_2, a_1, a_2, a_3)$ then we can obtain the moments of \mathbf{S} by typing in the following sequence of commands in MATLAB once RMTTool has been installed. This eliminates the need to obtain manually the expressions for the moments apriori.

```
>> startRMTTool
>> syms c t1 t2 a1 a2 a3
>> number_of_moments = 5;
>> LmzSigma = atomLmz([a1 a2 a3],[t1 t2 1-(t1+t2)]);
>> LmzS = AtimesWish(LmzSigma,c);
>> alpha_S = Lmz2MomF(LmzS,number_of_moments);
>> alpha_Stilde = c*alpha_S;
```

3.2 Computation of covariance moments of second order limit distribution

Equations (2.3) and (2.3b) express the relationship between the covariance of the second order limit distribution and the moments of \mathbf{S} . Let $M(x)$ denote a moment power series as in Notation 2.3 with coefficients α_j . Define the power series $H(x) = x M(x)$ and let

$$\mathcal{H}(x, y) := \left(\frac{\frac{d}{dx}(H(x)) \cdot \frac{d}{dy}(H(y))}{(H(x) - H(y))^2} - \frac{1}{(x - y)^2} \right) \quad (3.9)$$

so that $\mathcal{M}^\infty(x, y) := xy \mathcal{H}(x, y)$. The (i, j) -th coefficient of $\mathcal{M}^\infty(x, y)$ can then be extracted from a multivariate Taylor series expansion of $\mathcal{H}(x, y)$ about $x = 0, y = 0$. From (2.3), we then obtain the coefficients $\alpha_{i,j}^S = (2/\beta)\alpha_{i,j}^{M^\infty}$. This is best done using the MAPLE symbolic package where the following sequence of commands enumerates the coefficients $\alpha_{i,j}^S$ for $\beta = 1, 2$ and indices i and j such that $i+j \leq 2 \max_coeff$.

```
> with(numapprox):
> max_coeff := 5:
> H := x -> x*(1+sum(alpha[j]*x^2,j=1..2*max_coeff)):
> dHx := diff(H(x),x): dHy := diff(H(y),y):
> H2 := simplify(dHx*dHy/(H(x)-H(y))^2-1/(x-y)^2):
> H2series := mtaylor(H2, [x,y], 2*max_coeff):
> i:=5: j :=2:
> M2_infty_coeff[i,j] := simplify(coeff(coeff(H2series,x,i-1),y,j-1)):
> alphaS_second[i,j] := (2/beta)*M2_infty_coeff[i,j]:
```

Table 3 lists some of the coefficients of \mathcal{M}^∞ obtained using this procedure. When $\alpha_j = 1$ for all $j \in \mathbb{N}$, then $\alpha_{i,j} = 0$ as expected, since $\alpha_j = 1$ denotes the identity matrix. Note that the moments $\alpha_1, \dots, \alpha_{i+j}$ are needed to compute the second order covariance moments $\alpha_{i,j} = \alpha_{j,i}$.

The covariance matrix \mathbf{Q} with elements $\mathbf{Q}_{i,j} = \alpha_{i,j}$ gets increasingly ill-conditioned as $\dim(\mathbf{Q})$ increases; the growth in the magnitude of the diagonal entries $\alpha_{j,j}$ in Table 3 attests to this. This implies that the eigenvectors of \mathbf{Q} encode the information about the covariance of the second order limit distribution more efficiently than the matrix \mathbf{Q} itself. When $\Sigma = \mathbf{I}$ so that the SCM \mathbf{S} has the (null) Wishart distribution, the eigenvectors of \mathbf{Q} are the (appropriately normalized) Chebychev polynomials of the second kind [Mingo and Speicher, 2006]. The structure of the eigenvectors for arbitrary Σ is, as yet, unknown though research in that direction might yield additional insights.

4. Eigen-inference algorithms

4.1 Estimating θ for known model order

Estimating the unknown parameter vector θ follows from the asymptotic result in Proposition 3.1. For large p, n , since \mathbf{v}_θ is (approximately) normally distributed we can obtain the estimate θ by the principle of maximum-likelihood. When \mathbf{S} is real, Bai and Silverstein provide a formula, expressed as a difficult to compute contour integral, for the correction term μ_θ in (3.1). The log-likelihood of \mathbf{v}_θ is (ignoring constants and the correction term for the mean when \mathbf{S} is real) given by

$$\ell(\mathbf{v}_\theta | \theta) \approx -\mathbf{v}_\theta^T \mathbf{Q}_\theta^{-1} \mathbf{v}_\theta - \log \det \mathbf{Q}_\theta, \quad (4.1)$$

which allows us to obtain the maximum-likelihood estimate of θ as

$$\hat{\theta}_{(q)} = \arg \min_{\theta \in \Theta} \mathbf{v}_\theta^T \mathbf{Q}_\theta^{-1} \mathbf{v}_\theta + \log \det \mathbf{Q}_\theta \quad \text{for } q = \dim(\mathbf{v}_\theta) \geq \dim(\theta) \quad (4.2)$$

where Θ represents the parameter space for the elements of θ and \mathbf{v}_θ and \mathbf{Q}_θ are constructed as in Proposition 3.1.

Coefficient	Expression
$\alpha_{1,1}$	$\alpha_2 - \alpha_1^2$
$\alpha_{2,1}$	$-4\alpha_1\alpha_2 + 2\alpha_1^3 + 2\alpha_3$
$\alpha_{2,2}$	$16\alpha_1^2\alpha_2 - 6\alpha_2^2 - 6\alpha_1^4 - 8\alpha_1\alpha_3 + 4\alpha_4$
$\alpha_{3,1}$	$9\alpha_1^2\alpha_2 - 6\alpha_1\alpha_3 - 3\alpha_2^2 + 3\alpha_4 - 3\alpha_1^4$
$\alpha_{3,2}$	$6\alpha_5 + 30\alpha_1\alpha_2^2 - 42\alpha_1^3\alpha_2 - 18\alpha_2\alpha_3 + 12\alpha_1^5 + 24\alpha_1^2\alpha_3 - 12\alpha_1\alpha_4$
$\alpha_{3,3}$	$-18\alpha_3^2 - 27\alpha_2\alpha_4 + 9\alpha_6 - 30\alpha_1^6 + 21\alpha_2^3 + 36\alpha_1^2\alpha_4 - 72\alpha_1^3\alpha_3 + 126\alpha_1^4\alpha_2 - 135\alpha_1^2\alpha_2^2 + 108\alpha_1\alpha_2\alpha_3 - 18\alpha_1\alpha_5$
$\alpha_{4,1}$	$12\alpha_1\alpha_2^2 - 16\alpha_1^3\alpha_2 - 8\alpha_2\alpha_3 + 12\alpha_1^2\alpha_3 - 8\alpha_1\alpha_4 + 4\alpha_1^5 + 4\alpha_5$
$\alpha_{4,2}$	$-12\alpha_3^2 - 24\alpha_2\alpha_4 + 8\alpha_6 - 20\alpha_1^6 + 16\alpha_2^3 + 32\alpha_1^2\alpha_4 - 56\alpha_1^3\alpha_3 + 88\alpha_1^4\alpha_2 - 96\alpha_1^2\alpha_2^2 + 80\alpha_1\alpha_2\alpha_3 - 16\alpha_1\alpha_5$
$\alpha_{4,3}$	$96\alpha_2^2\alpha_3 + 60\alpha_1^7 + 84\alpha_1\alpha_3^2 + 432\alpha_1^3\alpha_2^2 + 180\alpha_1^4\alpha_3 - 48\alpha_3\alpha_4 + 12\alpha_7 - 36\alpha_2\alpha_5 - 24\alpha_1\alpha_6 + 144\alpha_1\alpha_2\alpha_4 + 48\alpha_1^2\alpha_5 - 96\alpha_1^3\alpha_4 - 156\alpha_1\alpha_2^3 - 300\alpha_1^5\alpha_2 - 396\alpha_1^2\alpha_2\alpha_3$
$\alpha_{4,4}$	$-140\alpha_1^8 - 76\alpha_2^4 - 48\alpha_6\alpha_2 + 256\alpha_3\alpha_4\alpha_1 - 40\alpha_4^2 + 16\alpha_8 - 64\alpha_3\alpha_5 - 32\alpha_1\alpha_7 + 1408\alpha_1^3\alpha_2\alpha_3 - 336\alpha_1^2\alpha_3^2 + 256\alpha_1^4\alpha_4 + 144\alpha_2^2\alpha_4 - 480\alpha_1^5\alpha_3 + 160\alpha_2\alpha_3^2 + 64\alpha_1^2\alpha_6 - 128\alpha_1^3\alpha_5 - 1440\alpha_1^4\alpha_2^2 + 832\alpha_1^2\alpha_2^3 + 800\alpha_1^6\alpha_2 - 768\alpha_1\alpha_2^2\alpha_3 - 576\alpha_1^2\alpha_2\alpha_4 + 192\alpha_1\alpha_2\alpha_5$
$\alpha_{5,1}$	$-5\alpha_3^2 - 10\alpha_2\alpha_4 + 5\alpha_6 - 5\alpha_1^6 + 5\alpha_2^3 + 15\alpha_1^2\alpha_4 - 20\alpha_1^3\alpha_3 + 25\alpha_1^4\alpha_2 - 30\alpha_1^2\alpha_2^2 + 30\alpha_1\alpha_2\alpha_3 - 10\alpha_1\alpha_5$
$\alpha_{5,2}$	$60\alpha_2^2\alpha_3 + 30\alpha_1^7 + 50\alpha_1\alpha_3^2 + 240\alpha_1^3\alpha_2^2 + 110\alpha_1^4\alpha_3 - 30\alpha_3\alpha_4 + 10\alpha_7 - 30\alpha_2\alpha_5 - 20\alpha_1\alpha_6 + 100\alpha_1\alpha_2\alpha_4 + 40\alpha_1^2\alpha_5 - 70\alpha_1^3\alpha_4 - 90\alpha_1\alpha_2^3 - 160\alpha_1^5\alpha_2 - 240\alpha_1^2\alpha_2\alpha_3$
$\alpha_{5,3}$	$-105\alpha_1^8 - 60\alpha_2^4 - 45\alpha_6\alpha_2 + 210\alpha_3\alpha_4\alpha_1 - 30\alpha_4^2 + 15\alpha_8 - 60\alpha_3\alpha_5 - 30\alpha_1\alpha_7 + 1140\alpha_1^3\alpha_2\alpha_3 - 270\alpha_1^2\alpha_3^2 + 225\alpha_1^4\alpha_4 + 120\alpha_2^2\alpha_4 - 390\alpha_1^5\alpha_3 + 135\alpha_2\alpha_3^2 + 60\alpha_1^2\alpha_6 - 120\alpha_1^3\alpha_5 - 1125\alpha_1^4\alpha_2^2 + 660\alpha_1^2\alpha_2^3 + 615\alpha_1^6\alpha_2 - 630\alpha_1\alpha_2^2\alpha_3 - 495\alpha_1^2\alpha_2\alpha_4 + 180\alpha_1\alpha_2\alpha_5$
$\alpha_{5,4}$	$-900\alpha_1^2\alpha_4\alpha_3 + 80\alpha_1^2\alpha_7 - 160\alpha_1^3\alpha_6 - 620\alpha_1^5\alpha_4 - 3200\alpha_1^3\alpha_2^3 + 700\alpha_1\alpha_2^4 + 3960\alpha_1^5\alpha_2^2 - 720\alpha_1^2\alpha_5\alpha_2 + 1840\alpha_1^3\alpha_4\alpha_2 - 4100\alpha_1^4\alpha_3\alpha_2 + 3600\alpha_1^2\alpha_2^2\alpha_3 - 1140\alpha_1\alpha_3^2\alpha_2 + 1040\alpha_1^3\alpha_3^2 - 440\alpha_2^3\alpha_3 + 440\alpha_3\alpha_4\alpha_2 + 240\alpha_1\alpha_6\alpha_2 + 320\alpha_1\alpha_5\alpha_3 - 1020\alpha_1\alpha_2^2\alpha_4 + 20\alpha_9 - 1820\alpha_1^7\alpha_2 + 180\alpha_2^2\alpha_5 + 320\alpha_1^4\alpha_5 + 180\alpha_1\alpha_4^2 + 1120\alpha_1^6\alpha_3 + 80\alpha_3^3 + 280\alpha_1^9 - 40\alpha_1\alpha_8 - 60\alpha_7\alpha_2 - 80\alpha_3\alpha_6 - 100\alpha_4\alpha_5$
$\alpha_{5,5}$	$2400\alpha_2\alpha_5\alpha_1^3 - 1350\alpha_2^2\alpha_5\alpha_1 + 600\alpha_3\alpha_5\alpha_2 + 300\alpha_1\alpha_7\alpha_2 - 900\alpha_6\alpha_2\alpha_1^2 - 1200\alpha_3\alpha_5\alpha_1^2 + 400\alpha_1\alpha_6\alpha_3 + 3000\alpha_3\alpha_4\alpha_1^3 + 5100\alpha_1^2\alpha_2^2\alpha_4 + 12300\alpha_1^5\alpha_2\alpha_3 + 5700\alpha_1^2\alpha_2\alpha_3^2 + 4400\alpha_1\alpha_2^3\alpha_3 + 400\alpha_1^4\alpha_6 - 15000\alpha_1^3\alpha_2^2\alpha_3 - 5750\alpha_1^4\alpha_2\alpha_4 - 200\alpha_1^3\alpha_7 + 500\alpha_1\alpha_4\alpha_5 + 225\alpha_6\alpha_2^2 - 675\alpha_4^2\alpha_1^2 - 3250\alpha_1^4\alpha_3^2 - 625\alpha_2^3\alpha_4 + 350\alpha_3^2\alpha_4 - 600\alpha_1\alpha_3^3 - 1050\alpha_2^2\alpha_3^2 - 2800\alpha_3\alpha_1^7 - 11550\alpha_1^6\alpha_2^2 - 3300\alpha_3\alpha_4\alpha_1\alpha_2 - 800\alpha_5\alpha_1^5 + 325\alpha_4^2\alpha_2 - 4375\alpha_1^2\alpha_2^4 - 630\alpha_1^{10} + 100\alpha_8\alpha_1^2 - 75\alpha_5^2 + 255\alpha_5^5 + 12000\alpha_1^4\alpha_2^3 + 4550\alpha_1^8\alpha_2 + 1550\alpha_1^6\alpha_4 + 25\alpha_{10} - 50\alpha_1\alpha_9 - 75\alpha_2\alpha_8 - 100\alpha_3\alpha_7 - 125\alpha_4\alpha_6$

Table 3: Relationship between the coefficients $\alpha_{i,j} = \alpha_{j,i}$ and α_i .

Canonically, the parameter vector $\boldsymbol{\theta}$ of models such as (1.2) is of length $2k-1$ so that $q = \dim(\mathbf{v}_\theta) \geq 2k-1$. In principle, estimation accuracy should increase with q since the covariance of \mathbf{v}_θ is explicitly accounted for via the weighting matrix \mathbf{Q}_θ .

Figure 2 compares the quantiles of the test statistic $\mathbf{v}_\theta' \mathbf{Q}_\theta \mathbf{v}_\theta$ for $\dim(\mathbf{v}_\theta) = q$ with the quantiles of the chi-square distribution with q degrees of freedom when $q = 2, 3$ for the model in (1.2) with $\boldsymbol{\theta} = (0.5, 2, 1)$, $n = p$ for $n = 40$ and $n = 320$. While there is good agreement with the theoretical distribution for large n, p , the deviation from the limiting result is not insignificant for moderate n, p . This justifies setting $q = 2$ for the testing procedures developed herein.

Hence, we suggest that for the estimation in (4.2), $q = \dim(\mathbf{v}_\theta) = \dim(\boldsymbol{\theta})$. This choice provide robustness in low to moderate dimensional settings where the deviations from the asymptotic result in Theorem 2.7 are not insignificant. Numerical simulations suggest that the resulting degradation in estimation accuracy in high dimensional settings, from such a choice, is relatively small. This loss in performance is offset by an increase in the speed of the underlying numerical optimization routine. This is the case because, though the dimensionality of $\boldsymbol{\theta}$ is the same, the matrix \mathbf{Q} gets increasingly ill-conditioned for higher values of q thereby reducing the efficiency of optimization methods .

4.2 Testing $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

Proposition 4.1. *Define the vector \mathbf{v}_θ and the covariance matrix \mathbf{Q}_θ as*

$$\mathbf{v}_\theta = \begin{bmatrix} \text{Tr } \mathbf{S} - p \alpha_1^\Sigma \\ \text{Tr } \mathbf{S}^2 - p \left(\alpha_2^\Sigma + \frac{p}{n} (\alpha_1^\Sigma)^2 \right) - \left(\frac{2}{\beta} - 1 \right) \alpha_2^\Sigma \frac{p}{n} \end{bmatrix} \quad (4.3a)$$

$$\mathbf{Q}_\theta = \frac{2}{\beta} \begin{bmatrix} \tilde{\alpha}_2 - \tilde{\alpha}_1^2 & 2\tilde{\alpha}_1^3 + 2\tilde{\alpha}_3 - 4\tilde{\alpha}_1\tilde{\alpha}_2 \\ 2\tilde{\alpha}_1^3 + 2\tilde{\alpha}_3 - 4\tilde{\alpha}_1\tilde{\alpha}_2 & 4\tilde{\alpha}_4 - 8\tilde{\alpha}_1\tilde{\alpha}_3 - 6\tilde{\alpha}_2^2 + 16\tilde{\alpha}_2\tilde{\alpha}_1^2 - 6\tilde{\alpha}_1^4 \end{bmatrix} \quad (4.3b)$$

with $\beta = 1$ (or 2) when S is real (or complex) and $\tilde{\alpha}_i \equiv \alpha_i^{\tilde{S}}$ given by

$$\tilde{\alpha}_1 = \frac{p}{n} \alpha_1^\Sigma \quad (4.4a)$$

$$\tilde{\alpha}_2 = \frac{p}{n} \alpha_2^\Sigma + \frac{p^2}{n^2} (\alpha_1^\Sigma)^2 \quad (4.4b)$$

$$\tilde{\alpha}_3 = \frac{p}{n} \alpha_3^\Sigma + 3 \frac{p^2}{n^2} \alpha_1^\Sigma \alpha_2^\Sigma + \frac{p^3}{n^3} (\alpha_1^\Sigma)^3 \quad (4.4c)$$

$$\tilde{\alpha}_4 = \frac{p}{n} \alpha_4^\Sigma + 4 \frac{p^2}{n^2} \alpha_1^\Sigma \alpha_3^\Sigma + 2 \frac{p^2}{n^2} (\alpha_2^\Sigma)^2 + 6 \frac{p^3}{n^3} (\alpha_1^\Sigma)^2 \alpha_2^\Sigma + \frac{p^4}{n^4} (\alpha_1^\Sigma)^4. \quad (4.4d)$$

and $\alpha_i^\Sigma = (1/p) \text{Tr } \boldsymbol{\Sigma}^i = \sum_{j=1}^k a_j t_j^i$. Thus, for large p and n , $\mathbf{v}_\theta \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\theta)$ so that

$$h(\boldsymbol{\theta}) := \mathbf{v}_\theta^T \mathbf{Q}_\theta^{-1} \mathbf{v}_\theta \sim \chi_2^2 \quad (4.5)$$

PROOF. This follows from Proposition 3.1. The correction term for the real case is discussed in a different context in Dumitriu and Edelman [2004]. A matrix theoretic derivation in the real case ($\beta = 1$) can be found in Srivastava [2005, Corollary 2.1, pp. 3]. \square

We test for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ by obtaining the test statistic

$$\boxed{H_{\boldsymbol{\theta}_0} : h(\boldsymbol{\theta}_0) = \mathbf{v}_{\boldsymbol{\theta}_0}^T \mathbf{Q}_{\boldsymbol{\theta}_0}^{-1} \mathbf{v}_{\boldsymbol{\theta}_0}} \quad (4.6)$$

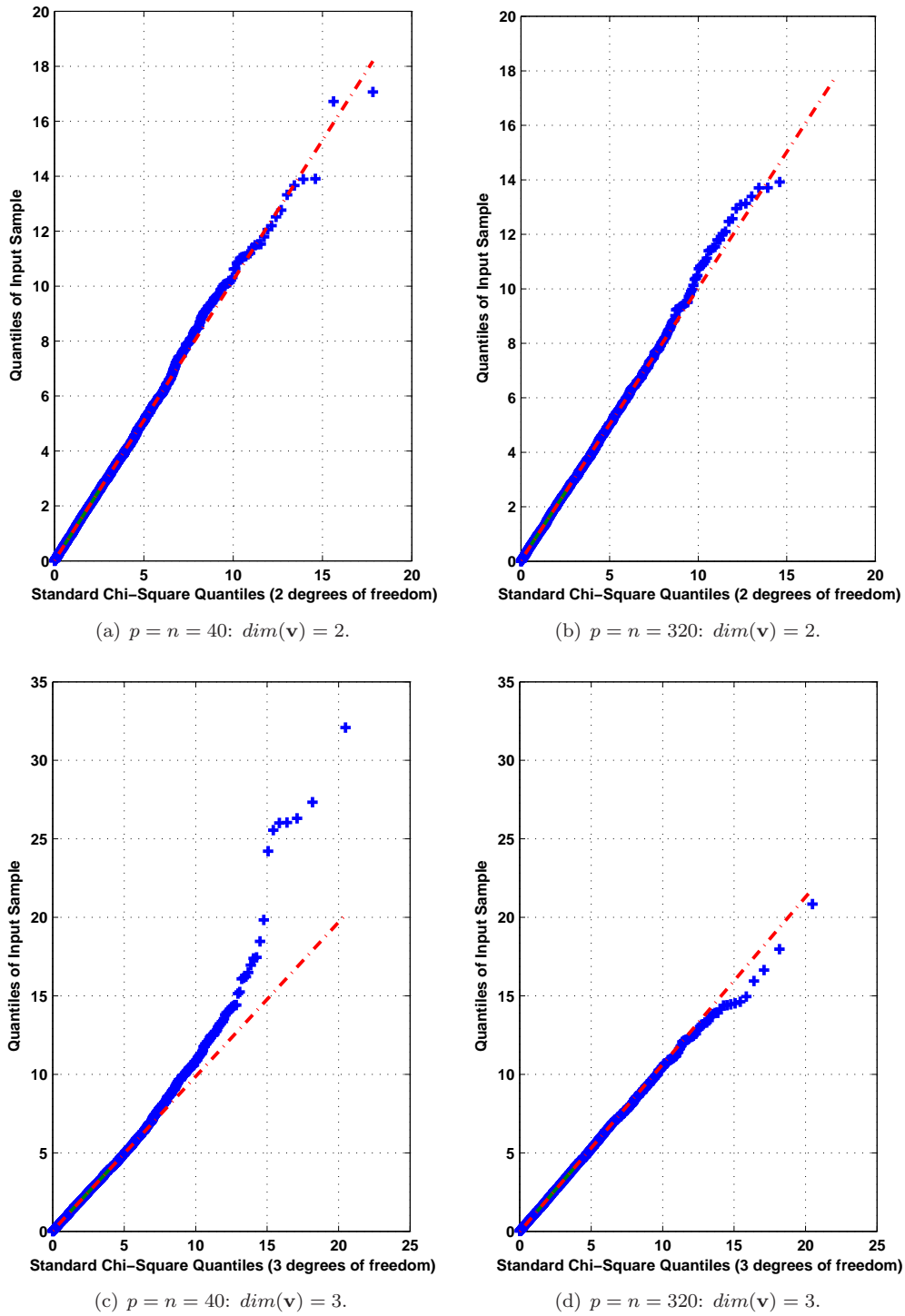


Figure 2: Numerical simulations (when \mathbf{S} is complex) illustrating the robustness of the distribution approximation for the test statistic in (4.2) formed with $\dim(\mathbf{v}) = 2$ to moderate dimensional settings.

where the \mathbf{v}_{θ_0} and \mathbf{Q}_{θ_0} are constructed as in (4.3a) and (4.3b), respectively. We reject the hypothesis for large values of H_{θ_0} . For a choice of threshold γ , the asymptotic convergence of the test statistic to the χ^2_2 distribution, implies that

$$\text{Prob.}(H_{\theta_0} = 1 | \boldsymbol{\theta} = \boldsymbol{\theta}_0) \approx F^{\chi^2_2}(\gamma). \quad (4.7)$$

Thus, for large p and n , when $\gamma = 5.9914$, $\text{Prob.}(H_{\theta_0} = 1 | \boldsymbol{\theta} = \boldsymbol{\theta}_0) \approx 0.95$.

4.3 Estimating $\boldsymbol{\theta}$ and testing $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$

When an $\hat{\boldsymbol{\theta}}$ is obtained using (4.2) then we may test for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ by forming the testing statistic

$$H_{\hat{\boldsymbol{\theta}}} : h(\hat{\boldsymbol{\theta}}) = \mathbf{u}_{\hat{\boldsymbol{\theta}}}^T \mathbf{W}_{\hat{\boldsymbol{\theta}}}^{-1} \mathbf{u}_{\hat{\boldsymbol{\theta}}} \quad (4.8)$$

where the $\mathbf{u}_{\hat{\boldsymbol{\theta}}}$, and $\mathbf{W}_{\hat{\boldsymbol{\theta}}}$ are constructed as in (4.3a) and (4.3b), respectively. However, the sample covariance matrix \mathbf{S} can no longer be used since the estimate $\hat{\boldsymbol{\theta}}$ was obtained from it. Instead, we form a test sample covariance matrix constructed from $\lceil (n/2) \rceil$ randomly chosen samples. Equivalently, since the samples are assumed to be mutually independent and identically distributed, we can form the test matrix from the first $\lceil (n/2) \rceil$ samples as

$$\bar{\mathbf{S}} = \frac{1}{\lceil \frac{n}{2} \rceil} \sum_{i=1}^{\lceil \frac{n}{2} \rceil} \mathbf{x}_i \mathbf{x}_i' \quad (4.9)$$

Note that $\alpha_k^{\bar{\mathbf{S}}}$ will have to be recomputed using $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ and $\bar{c} = p/\lceil (n/2) \rceil$. The hypothesis $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ is tested by rejecting values of the test statistic greater than a threshold γ . The threshold is selected using the approximation in (4.7). Alternately, the hypothesis can be rejected if the recentered and rescaled largest eigenvalue of \mathbf{S} is greater than the threshold γ . The threshold is selected using the quantiles of the (real or complex) Tracy-Widom distribution. The recentering and rescaling coefficients are obtained by the procedure described in [El Karoui, 2005].

4.4 Estimating $\boldsymbol{\theta}$ for unknown model order

Suppose we have a family of models parameterized by the vector $\boldsymbol{\theta}^{(\bar{k})}$. The elements of $\boldsymbol{\theta}^{(\bar{k})}$ are the free parameters of the model. For the model in (1.2), in the canonical case $\boldsymbol{\theta} = (t_1, \dots, t_{k-1}, a_1, \dots, a_k)$ since $t_1 + \dots + t_{k-1} + t_k = 1$ so that $\dim(\boldsymbol{\theta}^{(\bar{k})}) = 2k - 1$. If some of the parameters in (1.2) are known, then the parameter vector is modified accordingly.

When the model order is unknown, we select the model which has the minimum Akaike Information Criterion . For the situation at hand we propose that

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(\hat{k})} \text{ where } \hat{k} = \arg \min_{k \in \mathbb{N}} \left\{ \mathbf{u}_{\hat{\boldsymbol{\theta}}^{(k)}}^T \mathbf{W}_{\hat{\boldsymbol{\theta}}^{(k)}}^{-1} \mathbf{u}_{\hat{\boldsymbol{\theta}}^{(k)}} + \log \det \mathbf{W}_{\hat{\boldsymbol{\theta}}^{(k)}} \right\} + 2 \dim(\boldsymbol{\theta}^{(k)}) \quad (4.10)$$

where $\mathbf{u}_{\hat{\boldsymbol{\theta}}^{(k)}}$ and $\mathbf{W}_{\hat{\boldsymbol{\theta}}^{(k)}}$ are constructed as described in Section 4.3 using the test sample covariance matrix in (4.9). Alternately, a sequence of nested hypothesis tests using a largest eigenvalue based test as described in [El Karoui, 2005] can be used. Note that we think of the eigenvalues of the sample covariance matrix on which we are performing inference as a single sample and so applying the Bayesian Information Criterion to this problem is meaningless. It would be useful to compare the performance of the proposed and the nested hypothesis testing procedures in situations of practical interest. The full resolution of the model selection issues encountered is beyond the scope of this article though for a simple example we demonstrate the robustness of the proposed method in Section 5.1.

5. Numerical simulations

Let $\Sigma_{\bar{\theta}}$ be as in (1.2) with $\bar{\theta} = (t_1, a_1, a_2)$. When $t_1 = 0.5$, $a_1 = 2$ and $a_2 = 1$ then half of the population eigenvalues are of magnitude two while the remainder are of magnitude one. Let the unknown parameter vector $\theta = (t, a)$ where $t \equiv t_1$ and $a \equiv a_1$. Using the procedure described in Section 3.1, the first four moments can be obtained as (here $c = p/n$)

$$\alpha_1^S = 1 + t(a - 1) \quad (5.1a)$$

$$\alpha_2^S = (-2ac + a^2c + c)t^2 + (-1 + 2ac - 2c + a^2)t + 1 + c \quad (5.1b)$$

$$\begin{aligned} \alpha_3^S = & (-3c^2a^2 + a^3c^2 - c^2 + 3ac^2)t^3 + (3c^2 + 3c^2a^2 - 3ac - 6ac^2 - 3a^2c + 3a^3c + 3c)t^2 \\ & + (-3c^2 + a^3 - 1 - 6c + 3ac + 3a^2c + 3ac^2)t + 1 + c^2 + 3c \end{aligned} \quad (5.1c)$$

$$\begin{aligned} \alpha_4^S = & (6a^2c^3 + a^4c^3 - 4ac^3 - 4a^3c^3 + c^3)t^4 \\ & + (-6c^2 - 12a^3c^2 + 12ac^3 - 12a^2c^3 + 4a^3c^3 + 12ac^2 + 6a^4c^2 - 4c^3)t^3 \\ & + (-4a^2c - 4ac - 12ac^3 - 24ac^2 + 6a^4c + 6a^2c^3 + 12a^3c^2 + 6c - 6c^2a^2 + 6c^3 + 18c^2 - 4a^3c)t^2 \\ & + (-4c^3 + 4ac + 6c^2a^2 + 4ac^3 - 1 + 12ac^2 - 18c^2 + 4a^2c - 12c + 4a^3c + a^4)t \\ & + 1 + c^3 + 6c + 6c^2 \end{aligned} \quad (5.1d)$$

From the discussion in Section 3.2, we obtain the covariance of the second order limit distribution

$$\mathbf{Q}_{\theta} = \frac{2}{\beta} \begin{bmatrix} c^2(\alpha_2^S - \alpha_1^S) & c^3(2(\alpha_1^S)^3 + 2\alpha_3^S - 4\alpha_1^S\alpha_2^S) \\ c^3(2(\alpha_1^S)^3 + 2\alpha_3^S - 4\alpha_1^S\alpha_2^S) & c^4(4\alpha_4^S - 8\alpha_1^S\alpha_3^S - 6(\alpha_2^S)^2 + 16\alpha_2^S(\alpha_1^S)^2 - 6(\alpha_1^S)^4) \end{bmatrix}. \quad (5.2)$$

where $\beta = 1$ when S is real valued and $\beta = 2$ when \mathbf{S} is complex valued.

We then use (4.2) to estimate θ and hence the unknown parameters t and a . Table 4 and 5 compares the bias and mean squared error of the estimates for a and t respectively. Note the $1/p^2$ type decay in the mean squared error and how the real case has twice the variance as the complex case. As expected by the theory of maximum likelihood estimation, the estimates become increasingly normal for large p and n . This is evident from Figure 3. As expected, the performance improves as the dimensionality of the system increases.

5.1 Robustness to model overspecification

Consider the situation when the samples are complex valued and the true covariance matrix $\Sigma = 2\mathbf{I}$. We erroneously assume that there are two blocks for the model in (1.2) and that $a_1 = 1$ is known while $a := a_2$ and $t := t_1$ are unknown and have to be estimated. We estimate $\theta = (a, t)$ using (4.2) as before. The empirical cumulative distribution function (CDF) of \hat{t} over 4000 Monte-Carlo trials shown in Figure 4(d) shows that $\hat{t} \rightarrow 1$ as $p, n(p) \rightarrow \infty$. Figure 4(c) compares the quantiles of test statistic in (4.5) with that of the chi-squared distribution with two degrees of freedom. The excellent agreement for modest values of p and n validates the distributional approximation. Figures 4(a) and 4(b) plot the mean squared errors in estimating a and t , respectively. As before, the mean squared exhibits a $1/p^2$ behavior. Table 6 shows the $1/p$ decay in the bias of estimating these parameters.

For this same example, the seventh column and eighth column of Table 6 show the level at which a sphericity and the 2 block hypothesis are accepted when the procedure described in (4.2) is applied and a threshold is set at the 95% significance level. The ninth and tenth columns of Table 6 show the acceptance rate for the 2 block hypothesis when the largest eigenvalue test proposed in [El Karoui, 2005] is applied on a test sample covariance matrix formed using first $\lfloor n/2 \rfloor$ samples and the original sample covariance matrix, respectively. The largest eigenvalue value test has an acceptance rate closer to the 95% significance level

(a) $n = 0.5p$.

p	n	Bias	Complex Case MSE	MSE x $p^2/100$	Bias	Real Case MSE	MSE x $p^2/100$
20	10	0.0455	0.3658	1.4632	0.4862	1.2479	4.9915
40	20	-0.0046	0.1167	1.8671	0.2430	0.3205	5.1272
80	40	-0.0122	0.0337	2.1595	0.1137	0.08495	5.437
160	80	-0.0024	0.0083	2.1250	0.0598	0.02084	5.335
320	160	0.0008	0.0021	2.1790	0.0300	0.00528	5.406

(b) $n = p$.

p	n	Bias	Complex Case MSE	MSE x $p^2/100$	Bias	Real Case MSE	MSE x $p^2/100$
20	20	-0.0137	0.1299	0.5196	0.2243	0.3483	1.3932
40	40	-0.0052	0.0390	0.6233	0.1083	0.0901	1.4412
80	80	-0.0019	0.0093	0.5941	0.0605	0.0231	1.4787
160	160	-0.0005	0.0024	0.6127	0.0303	0.0055	1.4106
320	320	-0.0001	0.0006	0.6113	0.0162	0.0015	1.5155

(c) $n = 2p$.

p	n	Bias	Complex Case MSE	MSE x $p^2/100$	Bias	Real Case MSE	MSE x $p^2/100$
20	40	-0.0119	0.0420	0.1679	0.1085	0.1020	0.4081
40	80	-0.0017	0.0109	0.1740	0.0563	0.0255	0.4079
80	160	-0.0005	0.0028	0.1765	0.0290	0.0063	0.4056
160	320	-0.0004	0.0007	0.1828	0.0151	0.0016	0.4139
320	640	0.0001	0.0002	0.1752	0.0080	0.0004	0.4024

Table 4: Quality of estimation of $t = 0.5$ for different values of p (dimension of observation vector) and n (number of samples) – both real and complex case for the example in Section 5..

designed it was designed for. For all of the p and n values in Table 6, over the 4000 Monte-Carlo trials, applying the procedure described in Section 4.4 produced the correct estimate $\hat{k} = 1$ for the order of the model in (1.2) when $\Sigma = 2\mathbf{I}$.

(a) $n = 0.5p$.

p	n	Bias	Complex Case		Bias	Real Case	
			MSE	MSE $\times p^2/100$		MSE	MSE $\times p^2/100$
20	10	0.1278	0.1046	0.4185	0.00748	0.1024	0.4097
40	20	0.0674	0.0478	0.7647	-0.01835	0.04993	0.7989
80	40	0.0238	0.0111	0.7116	-0.02240	0.01800	1.1545
160	80	0.0055	0.0022	0.5639	-0.02146	0.00414	1.0563
320	160	0.0007	0.0005	0.5418	-0.01263	0.00112	1.1692

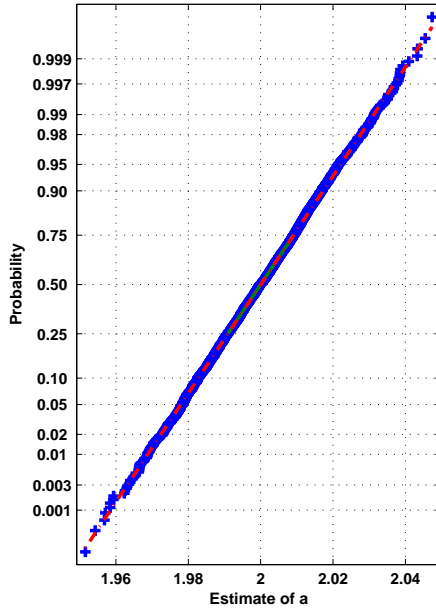
(b) $n = p$.

p	n	Bias	Complex Case		Bias	Real Case	
			MSE	MSE $\times p^2/100$		MSE	MSE $\times p^2/100$
20	20	0.0750	0.0525	0.2099	-0.0019	0.0577	0.2307
40	40	0.0227	0.0127	0.2028	-0.0206	0.0187	0.2992
80	80	0.0052	0.0024	0.1544	-0.0206	0.0047	0.3007
160	160	0.0014	0.0006	0.1499	-0.0126	0.0012	0.3065
320	320	0.0003	0.0001	0.1447	-0.0074	0.0003	0.3407

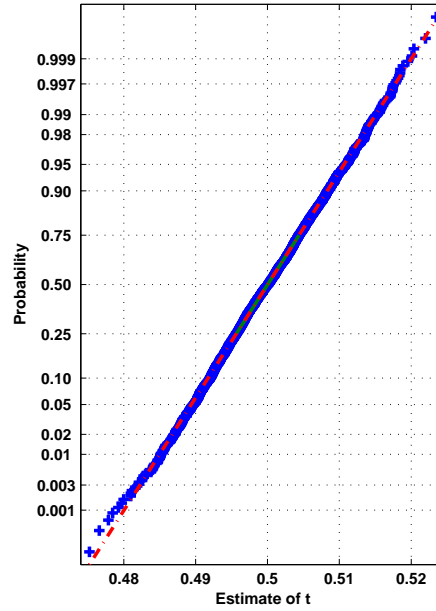
(c) $n = 2p$.

p	n	Bias	Complex Case		Bias	Real Case	
			MSE	MSE $\times p^2/100$		MSE	MSE $\times p^2/100$
20	40	0.0251	0.0134	0.0534	-0.0182	0.0205	0.0821
40	80	0.0049	0.0028	0.0447	-0.0175	0.0052	0.0834
80	160	0.0015	0.0007	0.0428	-0.0115	0.0014	0.0865
160	320	0.0004	0.0002	0.0434	-0.0067	0.0004	0.0920
320	640	0.0000	0.0000	0.0412	-0.0038	0.0001	0.0932

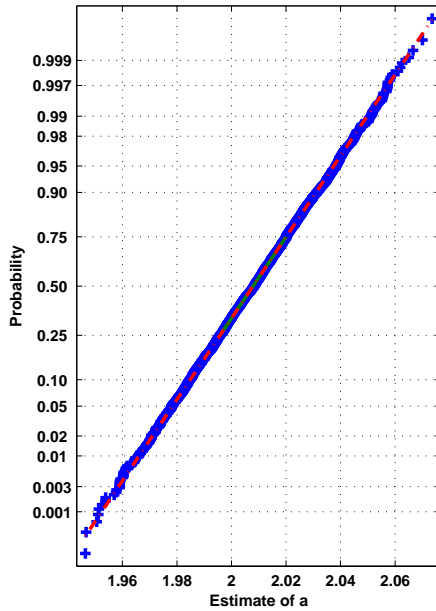
Table 5: Quality of estimation of $a = 2$ for different values of p (dimension of observation vector) and n (number of samples) – both real and complex case for the example in Section 5..



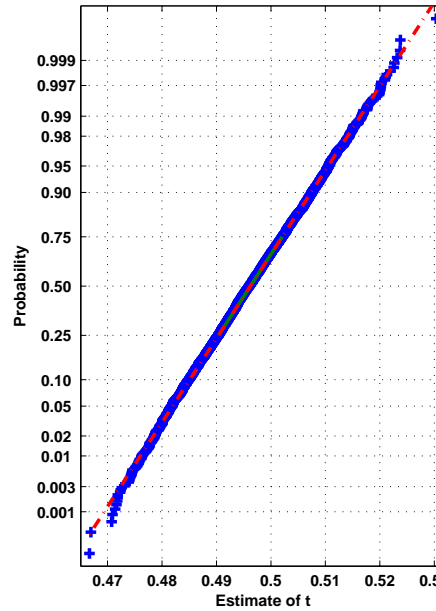
(a) \hat{a} : $p = 320, n = 640$.



(b) \hat{t} : $p = 320, p = 640$.

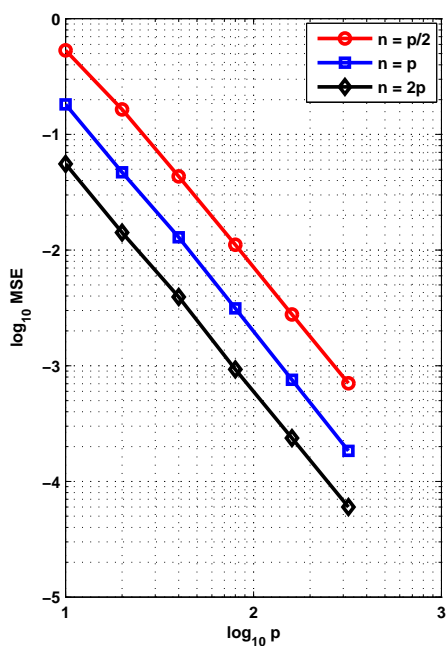


(c) \hat{a} : $p = 320, n = 640$. (Real valued)

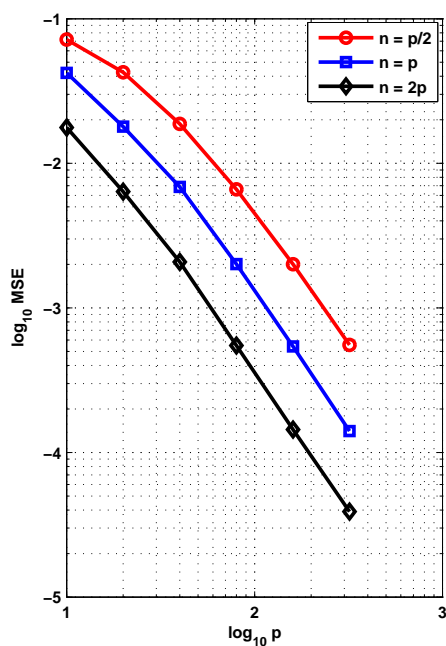


(d) \hat{t} : $p = 320, n = 640$. (Real valued)

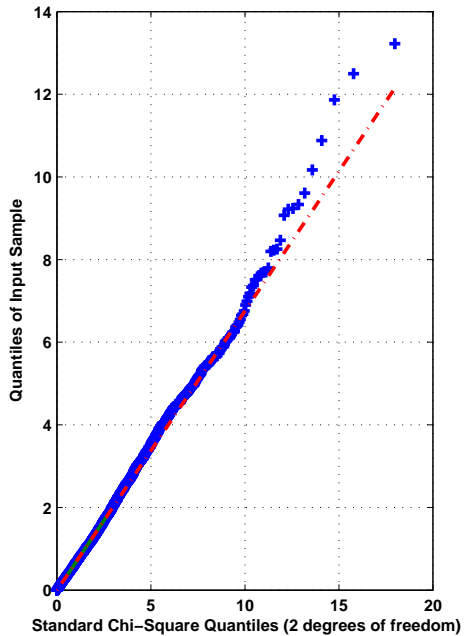
Figure 3: Normal probability plots of the estimates of a and t (true values: $a = 2, t = 0.5$) for the example in Section 5..



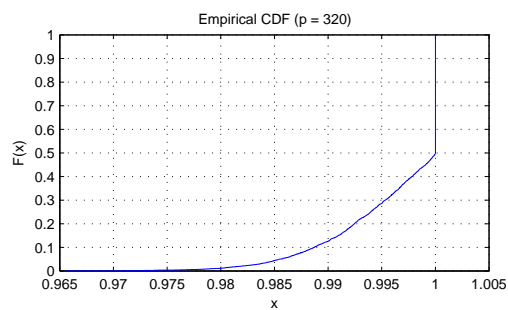
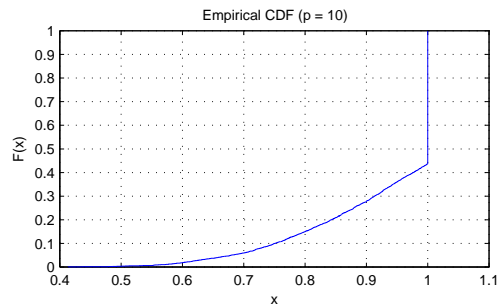
(a) MSE: \hat{a}



(b) MSE: \hat{t}



(c) QQ plot: Test statistic in (4.5) for $p = 320 = 2n$.



(d) Empirical CDF of \hat{t} : $n = p/2$.

Figure 4: Performance of estimation algorithm when model order has been overspecified and \mathbf{S} is complex. The population covariance matrix $\Sigma = 2\mathbf{I}$ while we assume that $a_1 = 1$ and estimate $a := a_2$ and $t := t_1$ in (1.2).

(a) $n = p/2$.

p	n	\hat{a} Bias	\hat{a} Bias x p	\hat{t} Bias	\hat{t} Bias x p	Sphericity Acceptance	2 block Acceptance	λ_{\max} test (full)	λ_{\max} test (half)
10	5	0.3523	3.5232	-0.1425	-1.4246	0.9820	0.9801	1.0000	0.9698
20	10	0.1997	3.9935	-0.1157	-2.3148	0.9783	0.9838	0.9998	0.9710
40	20	0.1078	4.3114	-0.0783	-3.1336	0.9795	0.9870	0.9958	0.9713
80	40	0.0545	4.3561	-0.0463	-3.7018	0.9765	0.9873	0.9838	0.9720
160	80	0.0272	4.3530	-0.0251	-4.0175	0.9743	0.9828	0.9763	0.9643
320	160	0.0141	4.5261	-0.0133	-4.2580	0.9805	0.9885	0.9753	0.9675

(b) $n = p$.

p	n	\hat{a} Bias	\hat{a} Bias x p	\hat{t} Bias	\hat{t} Bias x p	Sphericity Acceptance	2 block Acceptance	λ_{\max} test (full)	λ_{\max} test (half)
10	10	0.2087	2.0867	-0.1123	-1.1225	0.9793	0.9768	0.9998	0.9675
20	20	0.1050	2.0991	-0.0753	-1.5060	0.9773	0.9845	0.9965	0.9723
40	40	0.0558	2.2312	-0.0470	-1.8807	0.9850	0.9898	0.9898	0.9743
80	80	0.0283	2.2611	-0.0255	-2.0410	0.9813	0.9868	0.9773	0.9710
160	160	0.0137	2.1990	-0.0130	-2.0811	0.9805	0.9870	0.9790	0.9613
320	320	0.0067	2.1455	-0.0067	-2.1568	0.9775	0.9835	0.9608	0.9603

(c) $n = 2p$.

p	n	\hat{a} Bias	\hat{a} Bias x p	\hat{t} Bias	\hat{t} Bias x p	Sphericity Acceptance	2 block Acceptance	λ_{\max} test (full)	λ_{\max} test (half)
10	20	0.1067	1.0674	-0.0717	-0.7171	0.9790	0.9810	0.9993	0.9708
20	40	0.0541	1.0811	-0.0442	-0.8830	0.9753	0.9858	0.9890	0.9708
40	80	0.0290	1.1581	-0.0257	-1.0272	0.9743	0.9845	0.9830	0.9695
80	160	0.0140	1.1161	-0.0131	-1.0497	0.9763	0.9850	0.9743	0.9658
160	320	0.0071	1.1302	-0.0068	-1.0883	0.9778	0.9830	0.9703	0.9578
320	640	0.0036	1.1549	-0.0035	-1.1237	0.9758	0.9833	0.9598	0.9608

Table 6: Performance of estimation algorithm when model order has been overspecified and \mathbf{S} is complex. The population covariance matrix $\Sigma = 2\mathbf{I}$ while we assume that $a_1 = 1$ and estimate $a := a_2$ and $t := t_1$ in (1.2).

6. Inferential aspects of spiked covariance matrix models

Consider covariance matrix models whose eigenvalues are of the form $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > \lambda_{k+1} = \dots = \lambda_p = \lambda$. Such models arise when the signal occupies a k -dimensional subspace and the noise has covariance $\lambda \mathbf{I}$. Such models are referred to as *spiked covariance matrix models*. When $k \ll p$, then for large p , for \mathbf{v}_θ defined as in Proposition 3.1, the matrix \mathbf{Q}_θ may be constructed from the moments of the (null) Wishart distribution [Dumitriu and Rassart, 2003] instead, which are given by

$$\alpha_k^W = \lambda^k \sum_{j=0}^{k-1} c^j \frac{1}{j+1} \binom{k}{j} \binom{k-1}{j} \quad (6.1)$$

where $c = p/n$. Thus, for $q = 2$, \mathbf{Q}_θ is given by

$$\mathbf{Q}_\theta \equiv \mathbf{Q}_\lambda = \frac{2}{\beta} \begin{bmatrix} \lambda^2 c & 2\lambda^3(c+1)c \\ 2\lambda^3(c+1)c & 2\lambda^4(2c^2+5c+2)c \end{bmatrix}. \quad (6.2)$$

This substitution is motivated by Bai and Silverstein's analysis [Bai and Silverstein, 2004] where it is shown that when k is small relative to p , then the second order fluctuation distribution is asymptotically independent of the "spikes." When the multiplicities of the spike is known (say 1), then we let $t_i = 1/p$ and compute the moments α_j^S accordingly. The estimation problem thus reduces to

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \mathbf{v}_\theta^T \mathbf{Q}_\lambda^{-1} \mathbf{v}_\theta \quad \text{with } q = \dim(\mathbf{v}_\theta) = \dim(\theta) + 1 \quad (6.3)$$

where λ is an element of θ when it is unknown.

Consider the problem of estimating the magnitude of the spike for the model in (1.2) with $t_1 = 1/p$, and $a_2 = 1$ known and $a_1 = 10$ unknown so that $\theta = a \equiv a_1$. We obtain the estimate $\hat{\theta}$ from (6.3) with $\lambda = 1$ wherein the moments α_k^S given by

$$\alpha_1^S = \frac{-1 + a + p}{p} \quad (6.4a)$$

$$\alpha_2^S = \frac{a^2 p - 2pc + c - 2ac + cp^2 + p^2 - p + 2pac + a^2 c}{p^2} \quad (6.4b)$$

are obtained by plugging in $t = 1/p$ into (5.1).

Table 7 summarizes the estimation performance for this example. Note the $1/p$ scaling of the mean squared error and how the complex case has half the mean squared error. The estimates produced are asymptotically normal as seen in Figure 5.

6.1 Impact of the sample eigenvalue phase transition phenomenon

Consider testing for the hypothesis that $\Sigma = \mathbf{I}$. For the model in (1.2), which is equivalent to testing $\theta = (1, 1)$, from the discussion in Section 4.2, we form the test statistic

$$H_{\text{Sph.}} : h(\theta) = \mathbf{v}_\theta^T \mathbf{Q}_\theta^{-1} \mathbf{v}_\theta \quad (6.5)$$

where \mathbf{Q}_θ is given by (6.2) with $\lambda = 1$ and

$$\mathbf{v}_\theta = \begin{bmatrix} \text{Tr } \mathbf{S} - p \\ \text{Tr } \mathbf{S}^2 - p \left(1 + \frac{p}{n}\right) - \left(\frac{2}{\beta} - 1\right) \frac{p}{n} \end{bmatrix}$$

where $c = p/n$, as usual. We set a threshold $\gamma = 5.9914$ so that we accept the sphericity hypothesis whenever $h(\theta) \leq \gamma$. This corresponds to the 95-th percentile of the χ_2^2 distribution. Table 7(a) demonstrates how the

test is able to accept the hypothesis when $\Sigma = \mathbf{I}$ close to the 0.95 significance level it was designed for. Table 7(b) shows the acceptance of the sphericity hypothesis when $\Sigma = \Sigma = \text{diag}(10, 1, \dots, 1)$ instead.

Table 9 illustrates the performance of the sphericity test proposed by Ledoit and Wolf [2002] which consists of forming the test statistic

$$\text{LW}(\mathbf{S}) := \frac{np}{2} \left[\frac{1}{p} \text{Tr}[(\mathbf{S} - \mathbf{I})^2] - \frac{p}{n} \left[\frac{1}{p} \text{Tr} \mathbf{S} \right]^2 + \frac{p}{n} \right] \rightarrow \chi_{p(p+1)/2}^2 \quad (6.6)$$

and rejecting for large values above a threshold that is determined by using the asymptotic chi-squared approximation. Comparing Tables 9 and 8 reveals the (slightly) increased power of the test derived based on the framework developed in rejecting spiked models. Note how when p/n is large, both tests erroneously accept the null hypothesis an inordinate number of times. The faulty inference provided by the test based on the methodologies developed is not surprising given the phase transition phenomenon for the sample eigenvalues described by the following result due to Baik and Silverstein [2004], Paul [2005] and others [Baik et al., 2005].

Proposition 6.1. *Let \mathbf{S} denote a sample covariance matrix formed from an $p \times n$ matrix of Gaussian observations whose columns are independent of each other and identically distributed with mean $\mathbf{0}$ and covariance Σ . Denote the eigenvalues of Σ by $\lambda_1 \geq \lambda_2 > \dots \geq \lambda_k > \lambda_{k+1} = \dots = \lambda_p = \lambda$. Let l_j denote the j -th largest eigenvalue of $\widehat{\mathbf{R}}$. Then as $p, n \rightarrow \infty$ with $c_n = p/n \rightarrow c \in (0, \infty)$,*

$$l_j \rightarrow \begin{cases} \lambda_j \left(1 + \frac{\lambda c}{\lambda_j - \lambda} \right) & \text{if } \lambda_j > \lambda(1 + \sqrt{c}) \\ \lambda(1 + \sqrt{c})^2 & \text{if } \lambda_j \leq \lambda(1 + \sqrt{c}) \end{cases} \quad (6.7)$$

where the convergence is almost surely.

Since the inference methodologies we propose in this paper exploit the distributional properties of traces of powers of the sample covariance matrix, Proposition 6.1 pinpoints the fundamental inability of the sphericity test proposed to reject the hypothesis $\Sigma = \mathbf{I}$ whenever (for large p, n),

$$\lambda_i \leq 1 + \sqrt{\frac{p}{n}}$$

For the example considered, $\lambda_1 = 10$, so that the above condition is met whenever $p/n > c_t = 81$. For p/n on the order of c_t , the resulting inability to correctly reject the null hypothesis can be attributed to this phenomenon and the fluctuations of the largest eigenvalue.

Canonically speaking, eigen-inference methodologies which rely on traces of powers of the sample covariance matrix will be unable to differentiate between closely spaced population eigenvalues in high-dimensional, sample sized starved settings. This impacts the quality of the inference in a fundamental manner that is difficult to overcome. At the same time, however, the results in Baik and Silverstein [2004] suggest that if the practitioner has reason to believe that the population eigenvalues can be split into several clusters about $a_i \pm \sqrt{p/n}$, then the use of the model in (1.2) with a block subspace structure, where the individual blocks of sizes p_1, \dots, p_k are comparable to p , is justified. In such situations, the benefit of the proposed eigen-methodologies will be most apparent and might motivate experimental design that ensures that this condition is met.

(a) $n = p$.

p	n	Bias	Complex Case		Bias	Real Case	
			MSE	MSE x p		MSE	MSE x p
10	10	-0.5528	9.3312	93.3120	-0.5612	18.4181	184.1808
20	20	-0.2407	4.8444	96.8871	-0.2005	9.6207	192.4143
40	40	-0.1168	2.5352	101.4074	-0.0427	4.9949	199.7965
80	80	-0.0833	1.2419	99.3510	-0.03662	2.4994	199.9565
160	160	-0.0371	0.6318	101.0949	0.03751	1.2268	196.3018
320	320	-0.0125	0.3186	101.9388	0.04927	0.6420	204.4711

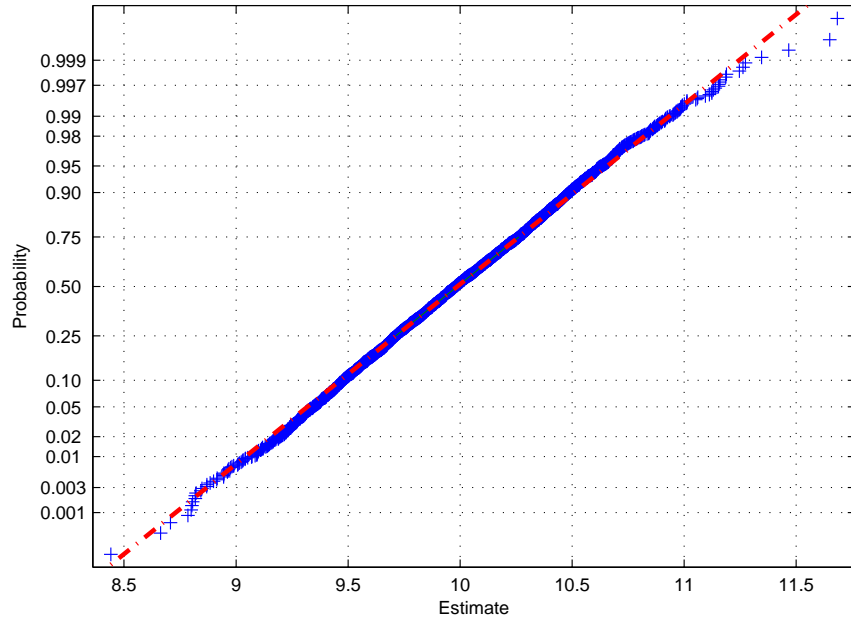
(b) $n = 1.5p$.

p	n	Bias	Complex Case		Bias	Real Case	
			MSE	MSE x p		MSE	MSE x p
10	15	-0.3343	6.6954	66.9537	-0.3168	12.7099	127.0991
20	30	-0.1781	3.2473	64.9454	-0.1454	6.4439	128.8798
40	60	-0.1126	1.6655	66.6186	-0.08347	3.2470	129.88188
80	120	-0.0565	0.8358	66.8600	-0.02661	1.6381	131.04739
160	240	-0.0287	0.4101	65.6120	0.02318	0.8534	136.5475
320	480	-0.0135	0.2083	66.6571	0.02168	0.4352	139.2527

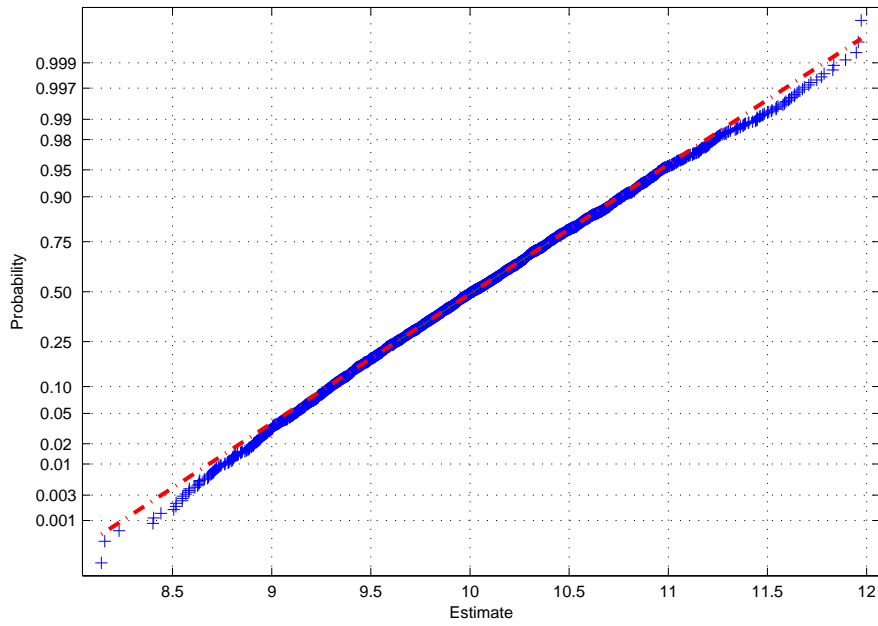
(c) $n = 2p$.

p	n	Bias	Complex Case		Bias	Real Case	
			MSE	MSE x p		MSE	MSE x p
10	20	-0.2319	4.9049	49.0494	-0.2764	9.6992	96.9922
20	40	-0.1500	2.5033	50.0666	-0.1657	4.6752	93.5043
40	80	-0.0687	1.2094	48.3761	-0.03922	2.5300	101.2007
80	160	-0.0482	0.6214	49.7090	-0.02426	1.2252	98.0234
160	320	-0.0111	0.3160	50.5613	0.01892	0.6273	100.3799
320	640	-0.0139	0.1580	50.5636	0.02748	0.3267	104.5465

Table 7: Algorithm performance for different values of p (dimension of observation vector) and n (number of samples) – both real and complex case.



(a) $p = 320, n = 640$ (Complex \mathbf{S}).



(b) $p = 320, n = 640$ (Real \mathbf{S}).

Figure 5: Normal probability plots of the spiked magnitude estimate (true value = 10).

(a) Empirical probability of accepting the null hypothesis when $\Sigma = \mathbf{I}$.

	n = 10	n = 20	n = 40	n = 80	n = 160	n = 320	n = 640
p = 10	0.9329	0.9396	0.9391	0.9411	0.9410	0.9464	0.9427
p = 20	0.9373	0.9414	0.9408	0.9448	0.9411	0.9475	0.9450
p = 40	0.9419	0.9482	0.9487	0.9465	0.9467	0.9451	0.9495
p = 80	0.9448	0.9444	0.9497	0.9496	0.9476	0.9494	0.9510
p = 160	0.9427	0.9413	0.9454	0.9505	0.9519	0.9473	0.9490
p = 320	0.9454	0.9468	0.9428	0.9451	0.9515	0.9499	0.9504

(b) Empirical probability of accepting the null hypothesis when $\Sigma = \text{diag}(10, 1, \dots, 1)$.

	n = 10	n = 20	n = 40	n = 80	n = 160	n = 320	n = 640
p = 10	0.0253	0.0003	-	-	-	-	-
p = 20	0.0531	0.0029	-	-	-	-	-
p = 40	0.1218	0.0093	-	-	-	-	-
p = 80	0.2458	0.0432	0.0080	-	-	-	-
p = 160	0.4263	0.1466	0.0002	-	-	-	-
p = 320	0.6288	0.3683	0.0858	0.0012	-	-	-

Table 8: The null hypothesis is accepted when the test statistic in (6.5) exceeds the 95% significance level for the χ^2 distribution with 2 degrees of freedom, *i.e.*, whenever $h(\theta) \leq 5.9914$.

(a) Empirical probability of accepting the null hypothesis when $\Sigma = \mathbf{I}$ (Ledoit-Wolf test).

	n = 10	n = 20	n = 40	n = 80	n = 160	n = 320	n = 640
p = 10	0.9483	0.9438	0.9520	0.9493	0.9510	0.9553	0.9465
p = 20	0.9498	0.9473	0.9510	0.9513	0.9498	0.9495	0.9423
p = 40	0.9428	0.9545	0.9468	0.9448	0.9488	0.9460	0.9478
p = 80	0.9413	0.9490	0.9513	0.9540	0.9480	0.9500	0.9460
p = 160	0.9438	0.9495	0.9475	0.9520	0.9508	0.9543	0.9448
p = 320	0.9445	0.9475	0.9493	0.9490	0.9485	0.9468	0.9453

(b) Empirical probability of the Ledoit-Wolf test accepting the null hypothesis when $\Sigma = \text{diag}(10, 1, \dots, 1)$.

	n = 10	n = 20	n = 40	n = 80	n = 160	n = 320	n = 640
p = 10	0.0345	0.0008	-	-	-	-	-
p = 20	0.0635	0.0028	-	-	-	-	-
p = 40	0.1283	0.0130	-	-	-	-	-
p = 80	0.2685	0.0450	0.0008	-	-	-	-
p = 160	0.4653	0.1575	0.0070	-	-	-	-
p = 320	0.6533	0.3700	0.0773	0.0010	-	-	-

Table 9: The null hypothesis is accepted when the test statistic in (6.6) exceeds the 95% significance level for the χ^2 distribution with $p(p + 1)/2$ degrees of freedom.

7. Extensions and lingering issues

In the development of the estimation procedures in this article, we ignored the correction term for the mean that appears in the real covariance matrix case (see Proposition 3.1). This was because Bai and Silverstein expressed it as a contour integral which appeared challenging to compute (see Eq. (1.6) in Bai and Silverstein [2004]). It is desirable to include this extra term in the estimation procedure if it can be computed efficiently using symbolic techniques. The recent work of Anderson and Zeitouni [Anderson and Zeitouni, 2006], despite its ambiguous title, represents a breakthrough on this and other fronts.

Anderson and Zeitouni encode the correction term in the coefficients of a power series that can be directly computed from the limiting moment series of the sample covariance matrix (see Theorem 3.4 [Anderson and Zeitouni, 2006]). Furthermore, they have expanded the range of the theory for the fluctuations of traces of powers of large Wishart-like sample covariance matrices, in the real sample covariance matrix case, to the situation when the entries are composed from a broad class of admissible non-Gaussian distributions. In such a scenario, the correction term takes into account the fourth moment of the distribution (see Eq. (5) and Theorems 3.3-3.4 in [Anderson and Zeitouni, 2006]). This latter development might be of use in some practical settings where the non-Gaussianity is well characterized. We have yet to translate their results into a computational recipe for determining the correction term though we intend to do so at a later date along with an Open Source distribution of a software implementation based on the principles outlined in this paper. The numerical results presented show the consistency of the proposed estimators; it would be of interest to establish this analytically and identify conditions in the real covariance matrix case, where ignoring the correction term in the mean can severely degrade the quality of estimation. The issue of how a local test that exploits global information, of the sort proposed by El Karoui [2005], compares to the global test developed in this article in terms of hypothesis discriminatory power is an unresolved question of great interest. A more systematic investigation is needed of the efficacy of various model order selection techniques for the problem considered.

Acknowledgements

We are grateful to Debashis Paul for his feedback on an early version of this paper. His comments improved the clarity of our exposition. The initial investigation comparing the power of local versus global tests was motivated by Patrick Perry's penetrating questions - only some of the issues he raised have been addressed in this paper while others remain outstanding. It has been a privilege to interact with Debashis Paul and Jack Silverstein on these and related matters during the authors' involvement in the SAMSI Program on High Dimensional Inference and Random Matrices. We thank Iain Johnstone for the opportunity to participate in this program and applaud James Berger, Nicole Scott, Margaret Polinkovsky, Sue McDonald and Donna Pike for creating the wonderful atmosphere at SAMSI. The first author thanks Arthur Baggeroer for encouragement during an early stage of this work.

R.R. and A.E. were supported by NSF Grant DMS-0411962. J.M. and R.S. acknowledge support by Discovery Grants and a Leadership Support Initiative Award from the Natural Sciences and Engineering Research Council of Canada. R.S. has also been supported by a Premier's Research Excellence Award from the Province of Ontario and a Killam Fellowship from the Canada Council for the Arts.

8. Appendix: Random Matrices and Free Probability Theory

8.1 Moments of random matrices and asymptotic freeness

Assume we know the eigenvalue distribution of two matrices \mathbf{A} and \mathbf{B} . What can we say about the eigenvalue distribution of the sum $\mathbf{A} + \mathbf{B}$ of the matrices? Of course, the latter is not just determined by the eigenvalues of \mathbf{A} and the eigenvalues of \mathbf{B} , but also by the relation between the eigenspaces of \mathbf{A} and of \mathbf{B} . Actually, it is a quite hard problem (Horn's conjecture) — which was only solved recently — to characterize

all possible eigenvalue distributions of $\mathbf{A} + \mathbf{B}$. However, if one is asking this question in the context of $N \times N$ -random matrices, then in many situations the answer becomes deterministic in the limit $N \rightarrow \infty$.

Definition 8.1. *Let $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ be a sequence of $N \times N$ -random matrices. We say that \mathbf{A} has a limit eigenvalue distribution if the limit of all moments*

$$\alpha_n := \lim_{N \rightarrow \infty} E[\text{tr}(\mathbf{A}_N^n)] \quad (n \in \mathbb{N})$$

exists, where E denotes the expectation and tr the normalized trace.

Using the language of limit eigenvalue distribution (see Def. 2.2), our question becomes: Given two random matrix ensembles of $N \times N$ -random matrices, $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ and $\mathbf{B} = (\mathbf{B}_N)_{N \in \mathbb{N}}$, with limit eigenvalue distribution, does also their sum $C = (C_N)_{N \in \mathbb{N}}$, with $C_N = \mathbf{A}_N + \mathbf{B}_N$, have a limit eigenvalue distribution, and furthermore, can we calculate the limit moments α_n^C of C out of the limit moments $(\alpha_k^A)_{k \geq 1}$ of \mathbf{A} and the limit moments $(\alpha_k^B)_{k \geq 1}$ of \mathbf{B} in a deterministic way. It turns out that this is the case if the two ensembles are in generic position, and then the rule for calculating the limit moments of C are given by Voiculescu's concept of "freeness".

Theorem 8.2 (Voiculescu [1991]). *Let \mathbf{A} and \mathbf{B} be two random matrix ensembles of $N \times N$ -random matrices, $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ and $\mathbf{B} = (\mathbf{B}_N)_{N \in \mathbb{N}}$, each of them with a limit eigenvalue distribution. Assume that \mathbf{A} and \mathbf{B} are independent (i.e., for each $N \in \mathbb{N}$, all entries of \mathbf{A}_N are independent from all entries of \mathbf{B}_N), and that at least one of them is unitarily invariant (i.e., for each N , the joint distribution of the entries does not change if we conjugate the random matrix with an arbitrary unitary $N \times N$ matrix). Then \mathbf{A} and \mathbf{B} are asymptotically free in the sense of the following definition.*

Definition 8.3 (Voiculescu [1985]). *Two random matrix ensembles $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ and $\mathbf{B} = (\mathbf{B}_N)_{N \in \mathbb{N}}$ with limit eigenvalue distributions are asymptotically free if we have for all $p \geq 1$ and all $n(1), m(1), \dots, n(p), m(p) \geq 1$ that*

$$\lim_{N \rightarrow \infty} E \left[\text{tr} \left\{ (\mathbf{A}_N^{n(1)} - \alpha_{n(1)}^A \mathbf{1}) \cdot (\mathbf{B}_N^{m(1)} - \alpha_{m(1)}^B \mathbf{1}) \cdots (\mathbf{A}_N^{n(p)} - \alpha_{n(p)}^A \mathbf{1}) \cdot (\mathbf{B}_N^{m(p)} - \alpha_{m(p)}^B \mathbf{1}) \right\} \right] = 0$$

One should realize that asymptotic freeness is actually a rule which allows to calculate all mixed moments in \mathbf{A} and \mathbf{B} , i.e. all expressions

$$\lim_{N \rightarrow \infty} E[\text{tr}(\mathbf{A}^{n(1)} \mathbf{B}^{m(1)} \mathbf{A}^{n(2)} \mathbf{B}^{m(2)} \dots \mathbf{A}^{n(p)} \mathbf{B}^{m(p)})]$$

out of the limit moments of \mathbf{A} and the limit moments of \mathbf{B} . In particular, this means that all limit moments of $\mathbf{A} + \mathbf{B}$ (which are sums of mixed moments) exist, thus $\mathbf{A} + \mathbf{B}$ has a limit distribution, and are actually determined in terms of the limit moments of \mathbf{A} and the limit moments of \mathbf{B} . The actual calculation rule is not directly clear from the above definition but a basic result of Voiculescu shows how this can be achieved by going over from the moments α_n to new quantities κ_n . In [Speicher, 1994], the combinatorial structure behind these κ_n was revealed and the name "free cumulants" was coined for them.

Definition 8.4 (Voiculescu [1986], Speicher [1994]). *Given the moments $(\alpha_n)_{n \geq 1}$ of some distribution (or limit moments of some random matrix ensemble), we define the corresponding free cumulants $(\kappa_n)_{n \geq 1}$ by the following relation between their generating power series: If we put*

$$M(x) := 1 + \sum_{n \geq 1} \alpha_n x^n \quad \text{and} \quad C(x) := 1 + \sum_{n \geq 1} \kappa_n x^n,$$

then we require as a relation between these formal power series that

$$C(xM(x)) = M(x).$$

Voiculescu actually formulated the relation above in a slightly different way using the so-called R -transform $\mathcal{R}(x)$, which is related to $C(x)$ by the relation

$$C(x) = 1 + z\mathcal{R}(x)$$

and in terms of the Cauchy transform $G(x)$ corresponding to a measure with moments α_n , which is related to $M(x)$ by

$$G(x) = \frac{M(\frac{1}{x})}{x}.$$

In these terms the equation $C(xM(x)) = M(x)$ says that

$$\frac{1}{G(x)} + \mathcal{R}(G(x)) = x, \tag{8.1}$$

i.e., that $G(x)$ and $K(x) := \frac{1}{x} + \mathcal{R}(x)$ are inverses of each other under composition.

One should also note that the relation $C(xM(x)) = M(x)$ determines the moments uniquely in terms of the cumulants and the other way around. The relevance of the κ_n and the R -transform for our problem comes from the following result of Voiculescu, which provides, together with (8.1), a very efficient way for calculating eigenvalue distributions of the sum of asymptotically free random matrices.

Theorem 8.5 (Voiculescu [1986]). *Let \mathbf{A} and \mathbf{B} be two random matrix ensembles which are asymptotically free. Denote by κ_n^A , κ_n^B , κ_n^{A+B} the free cumulants of \mathbf{A} , \mathbf{B} , $\mathbf{A} + \mathbf{B}$, respectively. Then one has for all $n \geq 1$ that*

$$\kappa_n^{A+B} = \kappa_n^A + \kappa_n^B.$$

Alternatively,

$$\mathcal{R}_{A+B}(x) = \mathcal{R}_A(x) + \mathcal{R}_B(x).$$

This theorem is one reason for calling the κ_n cumulants (as they linearize the "free convolution" in the same way as the usual convolution is linearized by classical cumulants), but there is also another justification for this, namely they are also the limit of classical cumulants of the entries of our random matrix, in the case that this is unitarily invariant.

Theorem 8.6 (Collins et al.). *Let $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ be a unitarily invariant random matrix ensemble of $N \times N$ random matrices \mathbf{A}_N whose limit eigenvalue distribution exists. Then the free cumulants of this matrix ensemble can also be expressed as the limit of special classical cumulants of the entries of the random matrices: If $\mathbf{A}_N = (a_{ij}^{(N)})_{i,j=1}^N$, then*

$$\kappa_n^A = \lim_{N \rightarrow \infty} N^{n-1} \cdot k_n \left(a_{i(1)i(2)}^{(N)}, a_{i(2)i(3)}^{(N)}, \dots, a_{i(n),i(1)}^{(N)} \right)$$

for any choice of distinct $i(1), \dots, i(n)$.

8.2 Fluctuations of random matrices and asymptotic second order freeness

There are many more refined questions about the limiting eigenvalue distribution of random matrices. In particular, questions around fluctuations have received a lot of interest in the last decade or so. The main motivation for introducing the concept of "second order freeness" was to understand the global fluctuations of the eigenvalues, which means that we look at the probabilistic behavior of traces of powers of our matrices. The limiting eigenvalue distribution, as considered in the last section, gives us the limit of the average of this traces. However, one can make more refined statements about their distributions. Consider a random matrix $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ and look on the normalized traces $\text{tr}(\mathbf{A}_N^l)$. Our assumption of a limit eigenvalue distribution means that the limits $\alpha_l := \lim_{N \rightarrow \infty} E[\text{tr}(\mathbf{A}_N^l)]$ exist. It turned out that in many cases the fluctuation around this limit,

$$\text{tr}(\mathbf{A}_N^l) - \alpha_l$$

is asymptotically Gaussian of order $1/N$; i.e., the random variable

$$N \cdot (\text{tr}(\mathbf{A}_N^l) - \alpha_l) = \text{Tr}(\mathbf{A}_N^l) - N\alpha_l = \text{Tr}(\mathbf{A}_N^l - \alpha_l \mathbf{1})$$

(where Tr denotes the unnormalized trace) converges for $N \rightarrow \infty$ to a normal variable. Actually, the whole family of centered unnormalized traces $(\text{Tr}(\mathbf{A}_N^l) - N\alpha_l)_{l \geq 1}$ converges to a centered Gaussian family. (One should note that we restrict all our considerations to complex random matrices; in the case of real random matrices there are additional complications, which will be addressed in some future investigations.) Thus the main information about fluctuations of our considered ensemble is contained in the covariance matrix of the limiting Gaussian family, i.e., in the quantities

$$\alpha_{m,n} := \lim_{N \rightarrow \infty} \text{cov}(\text{Tr}(\mathbf{A}_N^m), \text{Tr}(\mathbf{A}_N^n)).$$

Let us emphasize that the α_n and the $\alpha_{m,n}$ are actually limits of classical cumulants of traces; namely of the expectation as first and the variance as second cumulant. Nevertheless, the α 's will behave and will also be treated like moments; accordingly we will call the $\alpha_{m,n}$ 'fluctuation moments'. We will below define some other quantities $\kappa_{m,n}$, which take the role of cumulants in this context.

This kind of convergence to a Gaussian family was formalized in [Mingo and Speicher, 2006] by the notion of "second order limit distribution" (see our Def. 2.5).

Definition 8.7. Let $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ be an ensemble of $N \times N$ random matrices \mathbf{A}_N . We say that it has a second order limit distribution if for all $m, n \geq 1$ the limits

$$\alpha_n := \lim_{N \rightarrow \infty} k_1(\text{tr}(\mathbf{A}_N^n))$$

and

$$\alpha_{m,n} := \lim_{N \rightarrow \infty} k_2(\text{Tr}(\mathbf{A}_N^m), \text{Tr}(\mathbf{A}_N^n))$$

exist and if

$$\lim_{N \rightarrow \infty} k_r(\text{Tr}(\mathbf{A}_N^{n(1)}), \dots, \text{Tr}(\mathbf{A}_N^{n(r)})) = 0$$

for all $r \geq 3$ and all $n(1), \dots, n(r) \geq 1$.

We can now ask the same kind of question for the limit fluctuations as for the limit moments; namely, if we have two random matrix ensembles \mathbf{A} and \mathbf{B} and we know the second order limit distribution of \mathbf{A} and the second order limit distribution of \mathbf{B} , does this imply that we have a second order limit distribution for $\mathbf{A} + \mathbf{B}$, and, if so, is there an effective way for calculating it. Again, we can only hope for a positive solution to this if \mathbf{A} and \mathbf{B} are in a kind of generic position. As it turned out, the same requirements as before are sufficient for this. The rule for calculating mixed fluctuations constitutes the essence of the definition of the concept of second order freeness.

Theorem 8.8 (Mingo et al.). Let \mathbf{A} and \mathbf{B} be two random matrix ensembles of $N \times N$ -random matrices, $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ and $\mathbf{B} = (\mathbf{B}_N)_{N \in \mathbb{N}}$, each of them having a second order limit distribution. Assume that \mathbf{A} and \mathbf{B} are independent and that at least one of them is unitarily invariant. Then \mathbf{A} and \mathbf{B} are asymptotically free of second order in the sense of the following definition.

Definition 8.9 (Mingo and Speicher [2006]). Consider two random matrix ensembles $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ and $\mathbf{B} = (\mathbf{B}_N)_{N \in \mathbb{N}}$, each of them with a second order limit distribution. Denote by

$$Y_N(n(1), m(1), \dots, n(p), m(p))$$

the random variable

$$\text{Tr}((\mathbf{A}_N^{n(1)} - \alpha_{n(1)}^A \mathbf{1})(\mathbf{B}_N^{m(1)} - \alpha_{m(1)}^B \mathbf{1}) \cdots (\mathbf{A}_N^{n(p)} - \alpha_{n(p)}^A \mathbf{1})(\mathbf{B}_N^{m(p)} - \alpha_{m(p)}^B \mathbf{1})).$$

The random matrices $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ and $\mathbf{B} = (\mathbf{B}_N)_{N \in \mathbb{N}}$ are asymptotically free of second order if for all $n, m \geq 1$

$$\lim_{N \rightarrow \infty} k_2(\mathrm{Tr}(\mathbf{A}_N^n - \alpha_n^A \mathbf{1}), \mathrm{Tr}(\mathbf{B}_N^m - \alpha_m^B \mathbf{1})) = 0$$

and for all $p, q \geq 1$ and $n(1), \dots, n(p), m(1), \dots, m(p), \tilde{n}(1), \dots, \tilde{n}(q), \tilde{m}(1), \dots, \tilde{m}(q) \geq 1$ we have

$$\lim_{N \rightarrow \infty} k_2\left(Y_N(n(1), m(1), \dots, n(p), m(p)), Y_N(\tilde{n}(1), \tilde{m}(2), \dots, \tilde{n}(q), \tilde{m}(q))\right) = 0$$

if $p \neq q$, and otherwise (where we count modulo p for the arguments of the indices, i.e., $n(i+p) = n(i)$)

$$\begin{aligned} & \lim_{N \rightarrow \infty} k_2\left(Y_N(n(1), m(1), \dots, n(p), m(p)), Y_N(\tilde{n}(p), \tilde{m}(p), \dots, \tilde{n}(1), \tilde{m}(1))\right) \\ &= \sum_{l=1}^p \prod_{i=1}^p (\alpha_{n(i+l)+\tilde{n}(i)}^A - \alpha_{n(i+l)}^A \alpha_{\tilde{n}(i)}^A) (\alpha_{m(i+l)+\tilde{m}(i+1)}^B - \alpha_{m(i+l)}^B \alpha_{\tilde{m}(i+1)}^B). \end{aligned}$$

Again, it is crucial to realize that this definition allows one (albeit in a complicated way) to express every second order mixed moment, i.e., a limit of the form

$$\lim_{N \rightarrow \infty} k_2(\mathrm{Tr}(\mathbf{A}_N^{n(1)} \mathbf{B}_N^{m(1)} \dots \mathbf{A}_N^{n(p)} \mathbf{B}_N^{m(p)}), \mathrm{Tr}(\mathbf{A}_N^{\tilde{n}(1)} \mathbf{B}_N^{\tilde{m}(1)} \dots \mathbf{A}_N^{\tilde{n}(q)} \mathbf{B}_N^{\tilde{m}(q)}))$$

in terms of the second order limits of \mathbf{A} and the second order limits of \mathbf{B} . In particular, asymptotic freeness of second order also implies that the sum $\mathbf{A} + \mathbf{B}$ of our random matrix ensembles has a second order limit distribution and allows one to express them in principle in terms of the second order limit distribution of \mathbf{A} and the second order limit distribution of \mathbf{B} . As in the case of first order freeness, it is not clear at all how this calculation of the fluctuations of $\mathbf{A} + \mathbf{B}$ out of the fluctuations of \mathbf{A} and the fluctuations of \mathbf{B} can be performed effectively. In [Collins et al.] we were able to solve this problem by providing a second order cumulant machinery, similar to the first order case. Again, the idea is to go over to quantities which behave like cumulants in this setting. The actual description of those relies on combinatorial objects (annular non-crossing permutations), but as before this can be reformulated in terms of formal power series. Let us spell out the definition here in this form.

Definition 8.10 (Collins et al.). Let $(\alpha_n)_{n \geq 1}$ and $(\alpha_{m,n})_{m,n \geq 1}$ describe the first and second order limit moments of a random matrix ensemble. We define the corresponding first and second order free cumulants $(\kappa_n)_{n \geq 1}$ and $(\kappa_{m,n})_{m,n \geq 1}$ by the following requirement in terms of the corresponding generating power series. Put

$$C(x) := 1 + \sum_{n \geq 1} \kappa_n x^n, \quad C(x, y) := \sum_{m,n \geq 1} \kappa_{m,n} x^m y^n$$

and

$$M(x) := 1 + \sum_{n \geq 1} \alpha_n x^n, \quad M(x, y) := \sum_{m,n \geq 1} \alpha_{m,n} x^m y^n.$$

Then we require as relations between these formal power series that

$$C(xM(x)) = M(x) \tag{8.2}$$

and for the second order

$$M(x, y) = H(xM(x), yM(y)) \cdot \frac{\frac{d}{dx}(xM(x))}{M(x)} \cdot \frac{\frac{d}{dy}(yM(y))}{M(y)}, \tag{8.3}$$

where

$$H(x, y) := C(x, y) - xy \frac{\partial^2}{\partial x \partial y} \log\left(\frac{xM(y) - yM(x)}{x - y}\right), \tag{8.4}$$

or equivalently,

$$M(x, y) = C(xM(x), yM(y)) \cdot \frac{\frac{d}{dx}(xM(x))}{M(x)} \cdot \frac{\frac{d}{dy}(yM(y))}{M(y)} + xy \left(\frac{\frac{d}{dx}(xM(x)) \cdot \frac{d}{dy}(yM(y))}{(xM(x) - yM(y))^2} - \frac{1}{(x - y)^2} \right). \quad (8.5)$$

As in the first order case, instead of the moment power series $M(x, y)$ one can consider a kind of second order Cauchy transform, defined by

$$G(x, y) := \frac{M(\frac{1}{x}, \frac{1}{y})}{xy}.$$

If we also define a kind of second order R transform $\mathcal{R}(x, y)$ by

$$\mathcal{R}(x, y) := \frac{1}{xy} C(x, y),$$

then the formula (8.5) takes on a particularly nice form:

$$G(x, y) = G'(x)G'(y) \left\{ \mathcal{R}(G(x), G(y)) + \frac{1}{(G(x) - G(y))^2} \right\} - \frac{1}{(x - y)^2}. \quad (8.6)$$

$G(x)$ is here, as before, the first order Cauchy transform, $G(x) = \frac{1}{x} M(1/x)$.

The $\kappa_{m,n}$ defined above deserve the name ‘‘cumulants’’ as they linearize the problem of adding random matrices which are asymptotically free of second order. Namely, we have the following theorem, which provides, together with (8.6), an effective machinery for calculating the fluctuations of the sum of asymptotically free random matrices.

Theorem 8.11 (Collins et al.). *Let \mathbf{A} and \mathbf{B} be two random matrix ensembles which are asymptotically free. Then one has for all $m, n \geq 1$ that*

$$\kappa_n^{A+B} = \kappa_n^A + \kappa_n^B \quad \text{and} \quad \kappa_{m,n}^{A+B} = \kappa_{m,n}^A + \kappa_{m,n}^B.$$

Alternatively,

$$\mathcal{R}_{A+B}(x) = \mathcal{R}_A(x) + \mathcal{R}_B(x)$$

and

$$\mathcal{R}_{A+B}(x, y) = \mathcal{R}_A(x, y) + \mathcal{R}_B(x, y).$$

Again, one can express the second order cumulants as limits of classical cumulants of entries of a unitarily invariant matrix. In contrast to the first order case, we have now to run over two disjoint cycles in the indices of the matrix entries.

Theorem 8.12 (Collins et al.). *Let $\mathbf{A} = (\mathbf{A}_N)_{N \in \mathbb{N}}$ be a unitarily invariant random matrix ensemble which has a second order limit distribution. Then the second order free cumulants of this matrix ensemble can also be expressed as the limit of classical cumulants of the entries of the random matrices: If $\mathbf{A}_N = (a_{ij}^{(N)})_{i,j=1}^N$, then*

$$\kappa_{m,n}^A = \lim_{N \rightarrow \infty} N^{m+n} \cdot k_{m+n} \left(a_{i(1)i(2)}^{(N)}, a_{i(2)i(3)}^{(N)}, \dots, a_{i(m),i(1)}^{(N)}, a_{j(1)j(2)}^{(N)}, a_{j(2)j(3)}^{(N)}, \dots, a_{j(n),j(1)}^{(N)} \right)$$

for any choice of distinct $i(1), \dots, i(m), j(1), \dots, j(n)$.

8.3 Moments and Fluctuations of Wishart Matrices and Proof of Theorem 2.7

Wishart matrices, in the large size limit, fit quite well into the frame of free probability theory of first and second order (actually, they take over the role of Poisson distributions in that theory). In particular, their free cumulants of first and second order are quite easy to determine and are of a particularly nice form. We will use this to give a proof of Theorem 2.7. The statements in that theorem go back to the work of Bai and Silverstein [2004]. They give a more direct proof via analytic calculations of the Cauchy transforms. We prefer here, however, to show how Wishart matrices fit conceptually into the frame of free probability theory. At the end of the section we will also point out the generalization of formula (2.3b) to the case of random covariance matrices.

Let us remark that whereas the results around first order freeness are valid for complex as well as real random matrices, this is not the case any more for the second order; there are some complications to be dealt with in this case and at the moment the theory of second order freeness for real random matrices has not yet been developed. Thus our proof of the fluctuation formula (2.3b) will only cover the complex case. The fact that the real case differs from the complex case by a factor 2 can be found in the work of Bai and Silverstein [2004].

Instead of looking on the Wishart matrix $\mathbf{S} := \frac{1}{n}\mathbf{X}\mathbf{X}'$ from Equation (1.1) we will consider the closely related matrix

$$\tilde{\mathbf{S}} := \frac{1}{n}\mathbf{X}'\mathbf{X}.$$

Note that \mathbf{S} is a $p \times p$ -matrix, whereas $\tilde{\mathbf{S}}$ is an $n \times n$ matrix. The relation between the spectral behavior of those two matrices is quite straightforward, namely they have the same non-zero eigenvalues, which are filled up with additional zeros for the larger one. Thus the transition between these two matrices is very easy; their eigenvalue distributions are related by a rescaling (since the first order moments α_n go with the normalized trace) and their fluctuations are the same (since the second order moments $\alpha_{m,n}$ go with the unnormalized trace). The reason for considering $\tilde{\mathbf{S}}$ instead of \mathbf{S} is the following nice description of its first and second order distribution. In this theorem we will realize the Wishart matrix $\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}'$ with covariance matrix Σ in the form $\Sigma^{1/2}\mathbf{Y}\mathbf{Y}'\Sigma^{1/2}$ where \mathbf{Y} is a $p \times n$ Gaussian random matrix with independent entries of mean zero and variance $1/n$. The matrix $\tilde{\mathbf{S}}$ takes then on the form

$$\tilde{\mathbf{S}} = \mathbf{Y}'\Sigma\mathbf{Y}.$$

Note that we allow Σ to be itself random in the following theorem.

Theorem 8.13 (Nica and Speicher [1996], Mingo and Speicher [2006]). *Let $\Sigma = (\Sigma_p)_{p \in \mathbb{N}}$ be a random matrix ensemble of selfadjoint $p \times p$ -matrices and consider in addition a Gaussian ensemble $\mathbf{Y} = (\mathbf{Y}_p)_{p \in \mathbb{N}}$ of non-selfadjoint rectangular Gaussian $p \times n$ -random matrices (with mean zero and variance $1/n$ for the entries) such that \mathbf{Y} and Σ are independent. Put*

$$\tilde{\mathbf{S}} := (\mathbf{Y}'_p \Sigma_p \mathbf{Y}_p)_{p \in \mathbb{N}}.$$

In the following we consider the limit

$$p, n \rightarrow \infty \quad \text{such that} \quad \lim \frac{p}{n} =: c$$

for some fixed $c \in (0, \infty)$.

(1) *Assume that the limit eigenvalue distribution of $\Sigma = (\Sigma_p)_{p \in \mathbb{N}}$ exists for $p \rightarrow \infty$. Then $\tilde{\mathbf{S}}$, considered as an ensemble of $n \times n$ -random matrices $\mathbf{Y}'_p \Sigma_p \mathbf{Y}_p$, has a limit eigenvalue distribution. This limit eigenvalue distribution is determined by the fact that its free cumulants are given by the scaled corresponding limit moments of Σ , i.e., for all $n \geq 1$ we have*

$$\kappa_n^{\tilde{\mathbf{S}}} = c\alpha_n^\Sigma.$$

(2) Assume that we are in the complex case and that $\Sigma = (\Sigma_p)_{p \in \mathbb{N}}$ has a second order limit distribution for $p \rightarrow \infty$. Then $\tilde{\mathbf{S}}$ has a second order limit distribution, which is determined as follows: for all $m, n \geq 1$ we have

$$\kappa_n^{\tilde{\mathbf{S}}} = c\alpha_n^\Sigma \quad \text{and} \quad \kappa_{m,n}^{\tilde{\mathbf{S}}} = \alpha_{m,n}^\Sigma.$$

The first order statement of this theorem is due to Nica and Speicher, see [Nica and Speicher, 2006], the second order statement follows from the calculations in [Mingo and Speicher, 2006]. We will now use this theorem to prove our Theorem 2.7 in the complex case.

PROOF. If

$$M_\Sigma(x) = 1 + \sum_{n \geq 1} \alpha_n^\Sigma x^n$$

is the generating power series for the limit moments of Σ , then the above theorem says that the generating power series $C_{\tilde{\mathbf{S}}}(x)$ for the free cumulants of $\tilde{\mathbf{S}}$ is related with $M_\Sigma(x)$ by

$$\begin{aligned} C_{\tilde{\mathbf{S}}}(x) &= 1 + \sum_{n \geq 1} \kappa_n^{\tilde{\mathbf{S}}} x^n \\ &= 1 + c \sum_{n \geq 1} \alpha_n^\Sigma x^n \\ &= (1 - c) + cM_\Sigma(x). \end{aligned}$$

Thus, by the general relation $C_{\tilde{\mathbf{S}}}(xM_{\tilde{\mathbf{S}}}(x)) = M_{\tilde{\mathbf{S}}}(x)$, we get the generating power series $M_{\tilde{\mathbf{S}}}(x)$ for the limit moments of $\tilde{\mathbf{S}}$ as a solution to the equation

$$1 - c + cM_\Sigma[xM_{\tilde{\mathbf{S}}}(x)] = M_{\tilde{\mathbf{S}}}(x). \quad (8.7)$$

Let us now rewrite this for the Wishart matrix \mathbf{S} . Recall that the moments of \mathbf{S} and the moments of $\tilde{\mathbf{S}}$ are related by a simple scaling factor, resulting in a relation of the form

$$M_{\tilde{\mathbf{S}}}(x) = c(M_{\mathbf{S}}(x) - 1) + 1.$$

This gives

$$M_{\mathbf{S}}(x) = M_\Sigma[x(cM_{\mathbf{S}}(x) - c + 1)].$$

Rewriting this in terms of

$$g(x) := \frac{1}{x}M_{\mathbf{S}}(1/x) \quad \text{and} \quad g_\Sigma(x) := \frac{1}{x}M_\Sigma(1/x)$$

yields formula (2.1).

In order to get the result for second order only has to observe that the fluctuations of a non-random covariance matrix vanish identically, hence $C_{\tilde{\mathbf{S}}}(x, y) = C_{\mathbf{S}}(x, y) = 0$, and thus (8.5) reduces directly to (2.3). \square

If Σ is itself a random matrix with a second order limit distribution then we have (at least in the complex case) more general that

$$C_{\tilde{\mathbf{S}}}(x, y) = M_\Sigma(x, y)$$

and thus 8.5

$$\begin{aligned} M_{\mathbf{S}}(x, y) = M_{\tilde{\mathbf{S}}}(x, y) &= M_\Sigma(xM_{\tilde{\mathbf{S}}}(x), yM_{\tilde{\mathbf{S}}}(y)) \cdot \frac{\frac{d}{dx}(xM_{\tilde{\mathbf{S}}}(x))}{M_{\tilde{\mathbf{S}}}(x)} \cdot \frac{\frac{d}{dy}(yM_{\tilde{\mathbf{S}}}(y))}{M_{\tilde{\mathbf{S}}}(y)} \\ &\quad + xy \left(\frac{\frac{d}{dx}(xM_{\tilde{\mathbf{S}}}(x)) \cdot \frac{d}{dy}(yM_{\tilde{\mathbf{S}}}(y))}{(xM_{\tilde{\mathbf{S}}}(x) - yM_{\tilde{\mathbf{S}}}(y))^2} - \frac{1}{(x - y)^2} \right). \end{aligned} \quad (8.8)$$

Bibliography

- Greg W. Anderson and Ofer Zeitouni. A CLT for a band matrix model. *Probab. Theory Related Fields*, 134(2):283–338, 2006. ISSN 0178-8051.
- T. W. Anderson. Asymptotic theory of principal component analysis. *Annals of Math. Statistics*, 34:122–248, 1963.
- Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004. ISSN 0091-1798.
- Z.D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Ann. Probab.*, 26, No. 1:316–345, 1998.
- J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. <http://arxiv.org/math.ST/048165>, 2004.
- Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005. ISSN 0091-1798.
- B. Collins, J. Mingo, P. Śniady, and R. Speicher. Second order freeness and fluctuations of random matrices: III.higher order freeness and free cumulants. <http://arxiv.org/math.OA/0606431>.
- D. K. Dey and C. Srinivasan. Estimation of a covariance matrix under Stein’s loss. *Ann. Statist.*, 13:1581–1591, 1985.
- I. Dumitriu and A. Edelman. MOPS: Multivariate Orthogonal Polynomials (symbolically). <http://arxiv.org/math-ph/0409066>, 2004.
- Ioana Dumitriu and Etienne Rassart. Path counting and random matrix theory. *Electronic Journal of Combinatorics*, 7(7), 2003. R-43.
- N. El Karoui. Tracy-widom limit for the largest eigenvalue of a large class of complex wishart matrices. <http://arxiv.org/math.PR/0503109>, 2005.
- N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. <http://arxiv.org/math.ST/0609418>, 2006.
- L. R. Haff. Empirical bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.*, 8:586–597, 1980.
- O. Ledoit and M. Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30(4):1081–1102, 2002. ISSN 0090-5364.
- J. A. Mingo, P. Sniady, and R. Speicher. Second order freeness and fluctuations of random matrices: II. unitary random matrices. To appear in *Adv. in Math.*

- J. A. Mingo and R. Speicher. Second order freeness and fluctuations of random matrices: I. Gaussian and Wishart matrices and cyclic Fock spaces. *Journal of Functional Analysis*, 235:226–270, 2006.
- R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York, 1982. ISBN 0-471-09442-0. Wiley Series in Probability and Mathematical Statistics.
- R. R. Nadakuditi. *Applied Stochastic Eigen-Analysis*. PhD thesis, Massachusetts Institute of Technology, February 2007. Department of Electrical Engineering and Computer Science.
- A. Nica and R. Speicher. On the multiplication of free N -tuples of noncommutative random variables. *Amer. J. Math.*, 118(4):799–837, 1996. ISSN 0002-9327.
- A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. London Mathematical Society Lecture Note Series, New York: Cambridge University Press, 2006.
- D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Technical report, Stanford University, 2005. <http://anson.ucdavis.edu/~debashis/techrep/eigenlimit.pdf>.
- N. Raj Rao. RMTTool: A random matrix and free probability calculator in MATLAB. <http://www.mit.edu/~raj/rmtool/>.
- J. W. Silverstein and J. W. Combettes. Signal detection via spectral theory of large dimensional random matrices. *IEEE Trans. on Signal Process.*, 40(8):2100–2105, August 1992.
- S. T. Smith. Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.*, 53(5):1610–1630, 2005. ISSN 1053-587X.
- Roland Speicher. Multiplicative functions on the lattice of noncrossing partitions and free convolution. *Math. Ann.*, 298(4):611–628, 1994. ISSN 0025-5831.
- M. S. Srivastava. Some tests concerning the covariance matrix in high dimensional data. *Journal of Japan Sta. Soc.*, (35):251–272, 2005.
- M. S. Srivastava. Some tests criteria for the covariance matrix with fewer observations than the dimension. Technical report, University of Toronto, 2006. Technical Report No.0406.
- M. S. Srivastava. Multivariate theory for analyzing high dimensional data. *Journal of Japan Sta. Soc.*, 2007. To appear.
- H. L. Van Trees. *Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing*. John wiley and Sons, Inc., new York, 2002.
- D. Voiculescu. Symmetries of some reduced free product C^* -algebras. In *Operator algebras and their connections with topology and ergodic theory (Buşteni, 1983)*, volume 1132 of *Lecture Notes in Math.*, pages 556–588. Springer, Berlin, 1985.
- D. Voiculescu. Addition of certain noncommuting random variables. *J. Funct. Anal.*, 66(3):323–346, 1986. ISSN 0022-1236.
- D. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104:201–220, 1991.
- J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20 A:32–52, 1928.