

# TESTING IN HIGH-DIMENSIONAL SPIKED MODELS

BY IAIN M. JOHNSTONE<sup>1</sup> AND ALEXEI ONATSKI<sup>2</sup>

<sup>1</sup>*Department of Statistics, Stanford University, [imj@stanford.edu](mailto:imj@stanford.edu)*

<sup>2</sup>*Faculty of Economics, University of Cambridge, [ao319@cam.ac.uk](mailto:ao319@cam.ac.uk)*

We consider the five classes of multivariate statistical problems identified by James (*Ann. Math. Stat.* **35** (1964) 475–501), which together cover much of classical multivariate analysis, plus a simpler limiting case, symmetric matrix denoising. Each of James' problems involves the eigenvalues of  $E^{-1}H$  where  $H$  and  $E$  are proportional to high-dimensional Wishart matrices. Under the null hypothesis, both Wisharts are central with identity covariance. Under the alternative, the noncentrality or the covariance parameter of  $H$  has a single eigenvalue, a spike, that stands alone. When the spike is smaller than a case-specific phase transition threshold, none of the sample eigenvalues separate from the bulk, making the testing problem challenging. Using a unified strategy for the six cases, we show that the log likelihood ratio processes parameterized by the value of the subcritical spike converge to Gaussian processes with logarithmic correlation. We then derive asymptotic power envelopes for tests for the presence of a spike.

**1. Introduction.** High-dimensional multivariate models and methods, such as regression, principal components and canonical correlation analysis, repay study in frameworks where the dimensionality diverges to infinity together with the sample size. “Spiked” models that deviate from a reference model along a small fixed number of unknown directions have proven to be a fruitful abstraction and research tool in this context. A basic statistical question that arises in the analysis of such models is how to test for the presence of spikes in the data.

James [18] arranges multivariate statistical problems in five different groups with broadly similar features. His remarkable classification, recalled in Table 1, relies on the five most common hypergeometric functions  ${}_pF_q$ . In this paper we describe rank-one spiked models that represent each of James' classes in a high-dimensional setting. We derive the asymptotic behavior of the corresponding likelihood ratios in a regime where the dimensionality  $p$  of the data and the degrees of freedom  $n_1, n_2$  increase proportionally. Specifically, we study the ratios of the joint densities of the relevant data under the alternative hypothesis, which assumes the presence of a spike, to that under the null of no spike. In each case, the relevant data consist of the maximal invariant statistic represented by eigenvalues of a large random matrix.

We find that the joint distributions of the eigenvalues under the alternative hypothesis and under the null are mutually contiguous when the values of the spike is below a phase transition threshold. The value of the threshold depends on the problem type. Furthermore, we find that the log likelihood ratio processes, parametrized by the value of the spike, are asymptotically Gaussian with logarithmic mean and autocovariance functions. These findings allow us to compute the asymptotic power envelopes for the tests for the presence of spikes in five multivariate models representing each of James' classes.

Our analysis is based on classical results that assume Gaussian data. All the likelihood ratios that we study correspond to the joint densities of the solutions to the basic equation of

---

Received February 2017; revised January 2018.

*MSC2010 subject classifications.* Primary 62E20; secondary 62H15.

*Key words and phrases.* Likelihood ratio test, hypergeometric function, principal components analysis, canonical correlations, matrix denoising, multiple response regression.

TABLE 1  
*The five cases of James [18]*

	Statistical method	$n_1 H$	$n_2 E$
${}_0 F_0$	PCA Principal components analysis [latent roots of covariance matrix]	$W_p(n_1, \Sigma + \Phi)$	$n_2 \Sigma$
${}_1 F_0$	SigD Signal Detection [equality of covariance matrices]	$W_p(n_1, \Sigma + \Phi)$	$W_p(n_2, \Sigma)$
${}_0 F_1$	REG <sub>0</sub> Multivariate regression, known error covariance [noncentral means]	$W_p(n_1, \Sigma, n_1 \Phi)$	$n_2 \Sigma$
${}_1 F_1$	REG Multivariate regression, unknown error covariance [noncentral latent roots]	$W_p(n_1, \Sigma, n_1 \Phi)$	$W_p(n_2, \Sigma)$
${}_2 F_1$	CCA Canonical correlation analysis	$W_p(n_1, \Sigma, \Phi(Y))$	$W_p(n_2, \Sigma)$

James' names for the cases, when different from ours, are shown in brackets. The final two columns interpret  $H$  and  $E$  of (1) for Gaussian data, so that  $W_p$  denotes a  $p$ -variate central or noncentral Wishart distribution; see Definitions. Matrix  $\Phi$  has low rank, equal to one in this paper. For CCA,  $\Phi(Y)$  is a random noncentrality matrix; see Section 3.2 in the Supplementary Material [19] for definition. In cases 1 and 3,  $E$  is deterministic,  $\Sigma$  is known and  $n_2$  disappears. Otherwise,  $E$  is assumed independent of  $H$ .

classical multivariate statistics,

$$(1) \quad \det(H - \lambda E) = 0,$$

where the hypothesis  $H$  and error sums of squares  $E$  are proportional to Wishart matrices, as summarized for the various cases in Table 1. The five cases can be linked via sufficiency and invariance arguments to the statistical problems listed in the table. We briefly discuss these links in the next section.

James' classification suggests common features that call for a systematic approach. Thus the main steps of our asymptotic analysis are the same for all the five cases. The likelihood ratios have explicit forms that involve hypergeometric functions of two high-dimensional matrix arguments. However, one of the arguments has low rank under our spiked model alternatives. Indeed, for tractability we focus on the rank one setting. We can then represent the hypergeometric function of two high-dimensional matrix arguments in the form of a contour integral that involves a *scalar* hypergeometric function of the same type, Lemma 1, using the recent result of [12]. Then, we deform the contour of integration so that the integral becomes amenable to Laplace approximation analysis, extending [28], Chapter 4.

Using the Laplace approximation technique, we show that the log likelihood ratios are asymptotically equivalent to simpler random functions of the spike parameters, Theorems 10 and 11. The randomness enters via a linear spectral statistic of a large random matrix of either sample covariance or  $F$ -ratio type. Using central limit theorems for the two cases, due to [5] and [39], respectively, we derive the asymptotic Gaussianity and obtain the mean and the autocovariance functions of the log likelihood ratio processes, Theorem 12.

These asymptotics of the log likelihood processes show that the corresponding statistical experiments do not converge to Gaussian shift models. In other words, the experiments that consist of observing the solutions to equation (1), parameterized by the value of the spike under the alternative hypothesis, are not of Locally Asymptotically Normal (LAN) type. This implies that there are no ready-to-use optimality results associated with LAN experiments that can be applied in our setting. However at the fundamental level the derived asymptotics of the log likelihood ratio processes is all that is needed for the asymptotic analysis of the risk of the corresponding statistical decisions.

In this paper we use the derived asymptotics together with the Neyman–Pearson lemma and Le Cam's third lemma to find simple analytic expressions for the asymptotic power envelopes for the statistical tests of the null hypothesis of no spike in the data, Theorem 13. The

form of the envelope depends only on whether both  $H$  and  $E$  in equation (1) are Wisharts or only  $H$  is Wishart whereas  $E$  is deterministic.

For most of the cases, as the value of the spike under the alternative increases, the envelope, at first, rises very slowly. Then, as the spike approaches the phase transition, the rise quickly accelerates and the envelope “hits” unity at the threshold. However, in cases of two Wisharts and when the dimensionality is not much smaller than the degrees of freedom of  $E$ , the envelope rises more rapidly. In such cases the information in all the eigenvalues of  $E^{-1}H$  might be useful for detecting population spikes which lie far below the phase transition threshold.

A type of the analysis performed in this paper has been previously implemented in the study of the principal components case by [31]. Our work here identifies common features in James’ classification of multivariate statistical problems and uses them to extend the analysis to the full system. One of the hardest challenges in such an extension is the rigorous implementation of the Laplace approximation step. With this goal in mind, we have developed asymptotic approximations to the hypergeometric functions  ${}_1F_1$  and  ${}_2F_1$  which are uniform in certain domains of the complex plane, Lemma 3.

The simple observation that the solutions to equation (1) can be interpreted as the eigenvalues of random matrix  $E^{-1}H$  relates our work to the vast literature on the spectrum of large random matrices. Three extensively studied classical ensembles of random matrices are the Gaussian, Laguerre and Jacobi ensembles, for example, [23]. However, only the Laguerre and Jacobi ensembles appear in high-dimensional analysis of James’ five-fold classification. This prompts us to look for a “missing” class in James’ system that could be linked to the Gaussian ensemble.

Such a class is easy to obtain by taking the limit of  $\sqrt{n_1}(H - \Sigma)$  with  $\Sigma = I_p$  as  $n_1 \rightarrow \infty$ , for  $p$  fixed. We call the corresponding statistical problem “symmetric matrix denoising” (SMD). Under the null hypothesis, the observations are given by a  $p \times p$  matrix  $Z/\sqrt{p}$  with  $Z$  from the Gaussian Orthogonal Ensemble. Under the alternative, the observations are given by  $Z/\sqrt{p} + \Phi$ , where  $\Phi$  is a deterministic symmetric matrix of low rank, again of rank one for this paper. We add this “case zero” to James’ classification and derive the asymptotics of the corresponding log likelihood ratio and power envelope.

To summarize, the contributions of this paper are as follows:

- We revisit James’ classification, which covers a large part of classical multivariate analysis, now in the setting of high-dimensional data and show that the classification accommodates low-rank structures as departures from the classical null hypotheses.
- We show that in such high-dimensional settings with rank-one structure, random matrix theory allows tractable approximations to the joint eigenvalue density functions, in place of slowly converging zonal polynomial series.
- We show that the log likelihood ratio processes, when parametrized by spike magnitude, converge to Gaussian process limits in the subcritical interval.
- Hence, we show that informative tests are possible based on *all* of the eigenvalues whereas tests based on the largest eigenvalue alone are uninformative.
- As a tool, we develop new uniform approximations to certain hypergeometric functions.
- We identify symmetric matrix denoising as a limiting case of each of James’ models. It is the simplest model displaying all the phenomena seen in the paper. We clarify the manner in which the simpler cases are limits of the more complex ones.

The rest of the paper fleshes out this program and its conclusions. The proofs are largely deferred to the extensive Supplementary Material [19]. They reflect substantial effort to identify and exploit common structure in the six cases. Indeed, some of this common structure appears remarkable and not yet fully explained.

*Definitions and global assumptions.* Let  $Z$  be an  $n \times p$  data matrix with rows drawn i.i.d. from  $N_p(0, \Sigma)$ , a  $p$ -dimensional normal distribution with mean 0 and covariance  $\Sigma$ . Suppose that  $M$  is also  $n \times p$ , but deterministic. If  $Y = M + Z$ , then  $H = Y'Y$  has a  $p$ -dimensional Wishart distribution  $W_p(n, \Sigma, \Psi)$  with  $n$  degrees of freedom, covariance matrix  $\Sigma$  and noncentrality matrix  $\Psi = \Sigma^{-1}M'M$ . The central Wishart distribution, corresponding to  $M = 0$ , is denoted  $W_p(n, \Sigma)$ .

Throughout the paper we shall assume that

$$p \leq \min\{n_1, n_2\},$$

where  $p$  is the dimensionality of matrices  $H$  and  $E$  and  $n_1, n_2$  are the degrees of freedom of the corresponding Wishart distributions, as summarized in Table 1. The assumption  $p \leq n_2$  ensures almost sure invertibility of matrix  $E$  in (1), whereas the assumption  $p \leq n_1$ , while not essential, is made for brevity, as it reduces the number of various situations which need to be considered.

**2. Links to statistical problems.** We briefly review examples of statistical problems, old and new, that lead to each of James' five cases, plus symmetric matrix denoising, and explain our choice of labels for those cases.

*PCA.* In the first case  $n_1$  i.i.d.  $N_p(0, \Omega)$  observations are used to test the null hypothesis that the population covariance  $\Omega$  equals a given matrix  $\Sigma$ . The alternative of interest is

$$\Omega = \Sigma + \Phi \quad \text{with } \Phi = \theta \psi \psi',$$

where  $\theta > 0$  and  $\psi$  are unknown and  $\psi$  is normalized so that  $\|\Sigma^{-1/2}\psi\| = 1$ .

Without loss of generality (wlog), we may assume that  $\Sigma = I_p$ . Then, under the null, the data are isotropic noise, whereas, under the alternative, the first principal component explains a larger portion of the variation than the other principal components.

The null and the alternative hypotheses can be formulated in terms of the spectral "spike" parameter  $\theta$  as

$$(2) \quad H_0 : \theta_0 = 0 \quad \text{and} \quad H_1 : \theta_0 = \theta > 0,$$

where  $\theta_0$  is the true value of the "spike." This testing problem remains invariant under the multiplication of the  $p \times n_1$  data matrix, from the left and from the right by orthogonal matrices, and under the corresponding transformation in the parameter space. A maximal invariant statistic consists of the solutions  $\lambda_1 \geq \dots \geq \lambda_p$  of equation (1) with  $n_1 H$  equal to the sample covariance matrix and  $E = \Sigma$ . We restrict attention to the invariant tests. Therefore, the relevant data are summarized by  $\lambda_1, \dots, \lambda_p$ . For convenience, details of the invariance and sufficiency arguments for all cases are in Section 2.1 in the Supplementary Material [19].

*SigD.* Consider testing the equality of covariance matrices,  $\Omega$  and  $\Sigma$ , corresponding to two independent  $p$ -dimensional zero-mean Gaussian samples of sizes  $n_1$  and  $n_2$ . The alternative hypothesis is the same as for case PCA. Invariance considerations lead to tests based on the eigenvalues of the  $F$ -ratio of the sample covariance matrices. Matrix  $H$  from (1) equals the sample covariance corresponding to the observations that might contain a "signal" responsible for the covariance spike, whereas matrix  $E$  equals the other "noise" sample covariance matrix. We again can assume that the population covariance of the "noise"  $\Sigma = I_p$ , although this time it is unknown to the statistician (Section 2.1 in the Supplementary Material [19] explains why such an assumption involves no loss of generality). Here, we find it more convenient to work with the  $p$  solutions to the equation

$$(3) \quad \det\left(H - \lambda\left(E + \frac{n_1}{n_2}H\right)\right) = 0,$$

which we also denote  $\lambda_1 \geq \dots \geq \lambda_p$  to make the notations as uniform across the different cases as possible. Note that as the second sample size  $n_2 \rightarrow \infty$ , while  $n_1$  and  $p$  are held constant, equation (3) reduces to equation (1),  $E$  converges to  $\Sigma$  and SigD reduces to PCA.

*REG<sub>0</sub>, REG.* Now, consider linear regression with multivariate response

$$Y = X\beta + \varepsilon$$

when the goal is to test linear restrictions on the matrix of coefficients  $\beta$ . In case  $REG_0$  the covariance matrix  $\Sigma$  of the i.i.d. Gaussian rows of the error matrix  $\varepsilon$  is assumed known. REG corresponds to unknown  $\Sigma$ .

As explained in [25], pp. 433–434, the problem of testing linear restrictions on  $\beta$  can be cast in the canonical form, where the matrix of transformed response variables is split into three parts,  $Y_1, Y_2$  and  $Y_3$ . Matrix  $Y_1$  is  $n_1 \times p$ , where  $p$  is the number of response variables and  $n_1$  is the number of linear restrictions (per each of the  $p$  columns of matrix  $\beta$ ). Under the null hypothesis,  $\mathbb{E}Y_1 = 0$ , whereas under the alternative,

$$(4) \quad \mathbb{E}Y_1 = \sqrt{n_1}\theta\varphi\psi',$$

where  $\theta > 0$ ,  $\|\Sigma^{-1/2}\psi\| = 1$  and  $\|\varphi\| = 1$ . Matrices  $Y_2$  and  $Y_3$  are  $(q - n_1) \times p$  and  $(T - q) \times p$ , respectively, where  $q$  is the number of regressors and  $T$  is the number of observations. These matrices have, respectively, unrestricted and zero means under both the null and the alternative. Section 2.1 in the Supplementary Material [19] contains a discussion of the relationship between alternative (4) and a corresponding constraint on the coefficients of the untransformed regression model.

In the important example of comparison of  $q$  group means, that is, one-way MANOVA, the null hypothesis imposes equality of all means, while a rank one alternative would posit that the  $q$  mean vectors lie along a line, for example  $\mu_k = \mu_1 + s_k\psi$  for scalar  $s_k, k = 2, \dots, q$  and  $\psi \in \mathbb{R}^p$ . This will be a plausible reduction of a global alternative hypothesis in some applications.

For  $REG_0$ , sufficiency and invariance arguments lead to tests based on the solutions  $\lambda_1, \dots, \lambda_p$  of (1) with

$$H = Y_1'Y_1/n_1 \quad \text{and} \quad E = \Sigma.$$

These solutions represent a multivariate analog of the difference between the sum of squared residuals in the restricted and unrestricted regressions. Under the null hypothesis,  $n_1H$  is distributed as  $W_p(n_1, \Sigma)$  whereas, under the alternative, it is distributed as  $W_p(n_1, \Sigma, n_1\Phi)$ , where  $\Phi = \theta\Sigma^{-1}\psi\psi'$ . Without loss of generality, we may assume that  $\Sigma = I_p$ .

The canonical form of  $REG_0$  is essentially equivalent to the recently studied setting of *matrix denoising*

$$Y_1 = M + Z.$$

References, which point to a variety of applications, include [11, 15, 26, 36]. Often  $M$  is assumed to have low rank and the matrix valued noise  $Z$  to have i.i.d. Gaussian entries. Here, we test  $M = 0$  vs. a rank one alternative.

For REG, similar arguments lead to tests based on the  $p$  solutions  $\lambda_1, \dots, \lambda_p$  of (3) with

$$H = Y_1'Y_1/n_1 \quad \text{and} \quad E = Y_3'Y_3/n_2,$$

where the error d.f.  $n_2 = T - q$ . These solutions represent a multivariate analog of the  $F$  ratio: the difference between the sum of squared residuals in the restricted and unrestricted regressions to the sum of squared residuals in the restricted regression. Again, we may assume wlog that  $\Sigma$ , although unknown to the statistician, equals  $I_p$ . Note that, as  $n_2 \rightarrow \infty$  while  $n_1$  and  $p$  are held constant, REG reduces to  $REG_0$ .

CCA. Consider testing for independence between Gaussian vectors  $x_t \in \mathbb{R}^p$  and  $y_t \in \mathbb{R}^{n_1}$ , given zero mean observations with  $t = 1, \dots, n_1 + n_2$ . Partition the population and sample covariance matrices of the observations  $(x'_t, y'_t)'$  into

$$\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix},$$

respectively. Under  $H_0 : \Sigma_{xy} = 0$ , the alternative of interest is

$$(5) \quad \Sigma_{xy} = \sqrt{\frac{n_1\theta}{n_1\theta + n_1 + n_2}} \psi \varphi',$$

where the vectors of nuisance parameters  $\psi \in \mathbb{R}^p$  and  $\varphi \in \mathbb{R}^{n_1}$  are normalized so that

$$\|\Sigma_{xx}^{-1/2}\psi\| = \|\Sigma_{yy}^{-1/2}\varphi\| = 1.$$

The peculiar parameterizations of the alternative  $\theta \neq 0$  in (4) and (5) are chosen to allow unified treatments of PCA and REG<sub>0</sub> and of SigD, REG and CCA in our main results, Theorems 11 and 12 below.

The test can be based on the squared sample canonical correlations  $\lambda_1, \dots, \lambda_p$ , which are solutions to (1) with

$$H = S_{xy}S_{yy}^{-1}S_{yx} \quad \text{and} \quad E = S_{xx}.$$

Remarkably, the squared sample canonical correlations also solve (3) with different  $H$  and  $E$ , such that  $E$  is a central Wishart matrix and  $H$  is a noncentral Wishart matrix conditionally on a random noncentrality parameter (see Section 3.2 in the Supplementary Material [19]).

SMD. We observe a  $p \times p$  matrix  $X = \Phi + Z/\sqrt{p}$ , where  $Z$  is a noise matrix from the Gaussian Orthogonal Ensemble (GOE), that is, it is symmetric and

$$Z_{ii} \sim N(0, 2) \quad \text{and} \quad Z_{ij} \sim N(0, 1) \quad \text{if } i > j.$$

We seek to make inference about a symmetric rank-one “signal” matrix  $\Phi = \theta\psi\psi'$ . The null and the alternative hypotheses are given by (2). The nuisance vector  $\psi \in \mathbb{R}^p$  is normalized so that  $\|\psi\| = 1$ . The problem remains invariant under the multiplication of  $X$  from the left by an orthogonal matrix and from the right by its transpose. A maximal invariant statistic consists of the solutions  $\lambda_1, \dots, \lambda_p$  to (1) with  $H = X$  and  $E = I_p$ . We consider tests based on  $\lambda_1, \dots, \lambda_p$ .

The SMD case can be viewed as a degenerate version of each of the above cases. For example, consider PCA with  $p$  held fixed and  $n_1 \rightarrow \infty$ . Take  $\Sigma = I_p$  for convenience, and set  $\Omega = I_p + \sqrt{p/n_1}\Phi$  with  $\Phi = \theta\psi\psi'$ , so that the original value of the spike is rescaled to be a local perturbation. Now, write  $H$  in the form  $\Omega^{1/2}\check{H}\Omega^{1/2}$  where  $\check{H} \sim W_p(n_1, I_p)$ . A standard matrix central limit theorem for  $p$  fixed, for example, [16], Theorem 2.5.1, says that

$$\check{H} = I_p + Z/\sqrt{n_1} + o_p(n_1^{-1/2}),$$

where  $Z$  belongs to GOE. Writing  $\Omega^{1/2} = I_p + \frac{1}{2}\sqrt{p/n_1}\Phi + o(n_1^{-1/2})$  and introducing  $\mu = \sqrt{n_1/p}(\lambda - 1)$ , we can rewrite

$$\det(H - \lambda I_p) = (p/n_1)^{p/2} \det[\Phi + Z/\sqrt{p} - \mu I_p + o_p(1)],$$

so that PCA degenerates to SMD. Compare also [7].

Indeed, each of the cases eventually degenerate to SMD via sequential asymptotic links (Section 2.2 in the Supplementary Material [19] has details). For convenience, we summarize links between the different cases and the definitions of the corresponding matrices  $H$  and

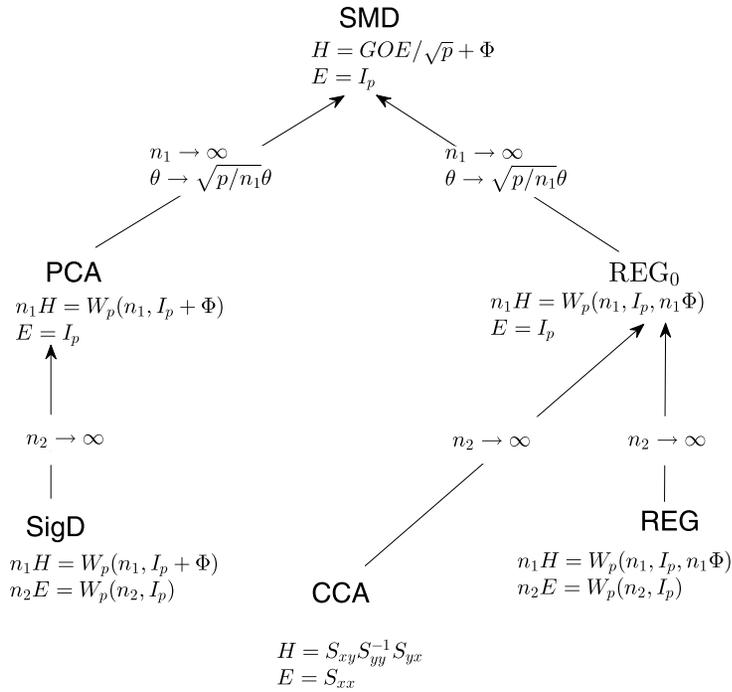


FIG. 1. Matrices  $H$  and  $E$  and links between the different cases. Without loss of generality, matrix  $E$  or, in SigD, REG and CCA cases, its population counterpart  $\Sigma$  is assumed to be equal to  $I_p$ . Matrix  $\Phi$  has the form  $\theta\psi\psi'$  with  $\theta \geq 0$  and  $\|\psi\| = 1$ .

$E$  in Figure 1. We note that the SMD model has been studied recently, for example, [9, 21] and references therein, though not with our techniques.

Cases SMD, PCA and  $REG_0$ , forming the upper half of the diagram, correspond to random  $H$  and deterministic  $E$ . The cases in the lower half of the diagram correspond to both  $H$  and  $E$  being random. Cases PCA and SigD are “parallel” to cases  $REG_0$  and REG in the sense that the alternative hypothesis is characterized by a rank one perturbation of the covariance and of the noncentrality parameter of  $H$  for the former and for the latter two cases, respectively. Case CCA “stands alone” because of the different structure of  $H$  and  $E$ . As discussed above, CCA can be reinterpreted in terms of  $H$  and  $E$  such that  $E$  is Wishart, but  $H$  is a noncentral Wishart only after conditioning on a random noncentrality parameter.

**3. The likelihood ratios.** Our goal is to study the asymptotic behavior of likelihood ratios based on the observed eigenvalues

$$\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_p\}.$$

Let  $p(\Lambda; \theta)$  be the joint density of the eigenvalues under the alternative and  $p(\Lambda; 0)$  the corresponding density under the null. James’ formulas for these joint densities lead to our starting point, which is a unified form for the likelihood ratio

$$(6) \quad L(\theta; \Lambda) = \frac{p(\Lambda; \theta)}{p(\Lambda; 0)} = \alpha(\theta)_p F_q(a, b; \Psi, \Lambda),$$

where  $\Psi = \Psi(\theta)$  is a  $p$ -dimensional matrix  $\text{diag}\{\Psi_{11}, 0, \dots, 0\}$  and the values of  $\Psi_{11}$ ,  $\alpha(\theta)$ ,  $p$ ,  $q$ ,  $a$  and  $b$  are as given in Table 2.

For SMD, we prove that  $L(\theta; \Lambda)$  is as in (6) in Section 3.1 in the Supplementary Material [19]. For PCA, the explicit form of the likelihood ratio is derived in [31]. For SigD,  $REG_0$  and REG, the expressions (6) follow, respectively, from equations (65), (68) and (73) of [18].

TABLE 2  
Parameters of the explicit expression (6) for the likelihood ratios. Here,  $n \equiv n_1 + n_2$

Case	${}_pF_q$	$\alpha(\theta)$	$a$	$b$	$\Psi_{11}$
SMD	${}_0F_0$	$\exp(-p\theta^2/4)$	–	–	$\theta p/2$
PCA	${}_0F_0$	$(1 + \theta)^{-n_1/2}$	–	–	$\theta n_1/(2(1 + \theta))$
SigD	${}_1F_0$	$(1 + \theta)^{-n_1/2}$	$n/2$	–	$\theta n_1/(n_2(1 + \theta))$
REG <sub>0</sub>	${}_0F_1$	$\exp(-n_1\theta/2)$	–	$n_1/2$	$\theta n_1^2/4$
REG	${}_1F_1$	$\exp(-n_1\theta/2)$	$n/2$	$n_1/2$	$\theta n_1^2/(2n_2)$
CCA	${}_2F_1$	$(1 + n_1\theta/n)^{-n/2}$	$(n/2, n/2)$	$n_1/2$	$\theta n_1^2/(n_2^2 + n_2 n_1(1 + \theta))$

For CCA, the expression is a corollary of [25], Theorem 11.3.2. Further details appear in Section 3.2 in the Supplementary Material [19].

Recall that hypergeometric functions of two matrix arguments  $\Psi$  and  $\Lambda$  are defined as

$${}_pF_q(a, b; \Psi, \Lambda) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\kappa \vdash k} \frac{(a_1)_\kappa \cdots (a_p)_\kappa}{(b_1)_\kappa \cdots (b_q)_\kappa} \frac{C_\kappa(\Psi)C_\kappa(\Lambda)}{C_\kappa(I_p)},$$

where  $a = (a_1, \dots, a_p)$  and  $b = (b_1, \dots, b_q)$  are parameters,  $\kappa$  are partitions of the integer  $k$ ,  $(a_j)_\kappa$  and  $(b_i)_\kappa$  are the generalized Pochhammer symbols and  $C_\kappa$  are the zonal polynomials, for example, [25], Definition 7.3.2. Note that some links between the cases illustrated in Figure 1 can also be established via asymptotic relations between the hypergeometric functions. For example, the confluence relations

$${}_0F_0(\Psi, \Lambda) = \lim_{a \rightarrow \infty} {}_1F_0(a; a^{-1}\Psi, \Lambda) \quad \text{and}$$

$${}_0F_1(b; \Psi, \Lambda) = \lim_{a \rightarrow \infty} {}_1F_1(a, b; a^{-1}\Psi, \Lambda),$$

for example, [29], equation 35.8.9, imply the links SigD  $\mapsto$  PCA and REG  $\mapsto$  REG<sub>0</sub> as  $n_2 \rightarrow \infty$  for  $p$  and  $n_1$  held constant.

In the next section we shall study the asymptotic behavior of the likelihood ratios (6) as  $n_1, n_2$  and  $p$  go to infinity so that

$$(7) \quad c_1 \equiv p/n_1 \rightarrow \gamma_1 \in (0, 1) \quad \text{and} \quad c_2 \equiv p/n_2 \rightarrow \gamma_2 \in (0, 1].$$

We denote this asymptotic regime by  $\mathbf{n}, p \rightarrow_{\boldsymbol{\gamma}} \infty$ , where  $\mathbf{n} = \{n_1, n_2\}$  and  $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2\}$ . To make our exposition as uniform as possible, we use this notation for all the cases, even though the simpler ones, such as SMD, do not refer to  $\mathbf{n}$ . We briefly discuss possible extensions of our analysis to the situations with  $\gamma_1 \geq 1$  in Section 7.

We are interested in the asymptotics of the likelihood ratios under the null hypothesis, that is, when the true value of the spike,  $\theta_0$ , equals zero. First, some background on the eigenvalues. Under the null,  $\lambda_1, \dots, \lambda_p$  are the eigenvalues of  $\text{GOE}/\sqrt{p}$  in the SMD case, of  $W_p(n_1, I_p)/n_1$  for PCA and REG<sub>0</sub> and of a  $p$ -dimensional multivariate beta matrix, for example, [24], p. 110, with parameters  $n_1/2$  and  $n_2/2$  and, here, scaled by a factor of  $n_2/n_1$ , in the SigD, REG and CCA cases. The empirical distribution of  $\lambda_1, \dots, \lambda_p$

$$\hat{F} = \frac{1}{p} \sum_{j=1}^p I\{\lambda_j \leq \lambda\}$$

is well known, [3], to converge weakly almost surely (a.s.) in each case

$$\hat{F} \Rightarrow F_{\boldsymbol{\gamma}} = \begin{cases} F^{\text{SC}} & \text{for SMD,} \\ F^{\text{MP}} & \text{for PCA, REG}_0, \\ F^{\text{W}} & \text{for SigD, REG, CCA,} \end{cases}$$

TABLE 3  
*Semicircle, Marchenko–Pastur and scaled Wachter distributions*

Case	$F_{\mathbf{y}}^{\text{lim}}$	Density, $\lambda \in [\beta_-, \beta_+]$	$\beta_{\pm}$	Threshold $\bar{\theta}$
SMD	SC	$\frac{R(\lambda)}{2\pi}$	$\pm 2$	1
PCA REG <sub>0</sub>	MP	$\frac{R(\lambda)}{2\pi\gamma_1\lambda}$	$(1 \pm \sqrt{\gamma_1})^2$	$\sqrt{\gamma_1}$
SigD REG CCA	W	$\frac{(\gamma_1 + \gamma_2)R(\lambda)}{2\pi\gamma_1\lambda(\gamma_1 - \gamma_2\lambda)}$	$\gamma_1 \left(\frac{\rho \pm 1}{\rho \pm \gamma_2}\right)^2$	$\frac{\rho + \gamma_2}{1 - \gamma_2}$
		$R(\lambda) = \sqrt{(\beta_+ - \lambda)(\lambda - \beta_-)}$	$\rho = \sqrt{\gamma_1 + \gamma_2 - \gamma_1\gamma_2}$	

the semicircle, Marchenko–Pastur and (scaled) Wachter distributions, respectively. Table 3 recalls the explicit forms of these limiting distributions. The cumulative distribution functions  $F_{\mathbf{y}}^{\text{lim}}(\lambda)$  are linked in the sense that

$$F_{\mathbf{y}}^{\text{W}}(\lambda) \rightarrow F_{\gamma_1}^{\text{MP}}(\lambda) \quad \text{as } \gamma_2 \rightarrow 0,$$

$$F_{\gamma_1}^{\text{MP}}(\sqrt{\gamma_1}\lambda + 1) \rightarrow F^{\text{SC}}(\lambda) \quad \text{as } \gamma_1 \rightarrow 0.$$

If  $\varphi$  is a “well-behaved” function, the centered *linear spectral statistic*

$$(8) \quad \sum_{j=1}^p \varphi(\lambda_j) - p \int \varphi(\lambda) dF_{\mathbf{c}}^{\text{lim}}(\lambda)$$

converges in distribution to a Gaussian random variable in each of the semicircle [6], Marchenko–Pastur [5] and Wachter [39] cases. Note that the centering constant is defined in terms of  $F_{\mathbf{c}}$ , where  $\mathbf{c} = \{c_1, c_2\}$ . That is, the “correct centering” can be computed using the densities from Table 3, where  $\gamma_1$  and  $\gamma_2$  are replaced by  $c_1 \equiv p/n_1$  and  $c_2 \equiv p/n_2$ , respectively.

Finally, let us recall the behavior of the largest eigenvalue  $\lambda_1$  under the alternative hypothesis. As long as  $\theta \leq \bar{\theta}$ , the phase transition threshold reported in Table 3, the top eigenvalue  $\lambda_1 \rightarrow \beta_+$ , the upper boundary of support of  $F_{\mathbf{y}}$ , almost surely. When  $\theta > \bar{\theta}$ ,  $\lambda_1$  separates from “the bulk” of the other eigenvalues and a.s. converges to a point strictly above  $\beta_+$ . For details, we refer to [8, 10, 12, 22, 27, 30] for the respective cases SMD, PCA, SigD, REG<sub>0</sub>, REG and CCA.

The fact that  $\lambda_1$  converges to different limits under the null and, under the alternative hypothesis, sheds light on the behavior of the likelihood ratio when  $\theta$  is above the phase transition threshold  $\bar{\theta}$ . In such *super-critical* cases, the likelihood ratio degenerates. The sequences of measures corresponding to the distributions of  $\Lambda$  under the null and under super-critical alternatives are asymptotically mutually singular as  $\mathbf{n}, p \rightarrow_{\mathbf{y}} \infty$ , as shown in [22] and [31] for SMD and PCA, respectively. In contrast, as we show below, the sequences of measures corresponding to the distributions of  $\Lambda$  under the null and under *subcritical* alternatives  $\theta < \bar{\theta}$  are mutually contiguous, and the likelihood ratio converges to a Gaussian process. In the super-critical setting, an analysis of the likelihood ratios under local alternatives appears in [13].

**4. Contour integral representation.** The asymptotic behavior of the likelihood ratios (6) depends on that of  ${}_pF_q(a, b; \Psi, \Lambda)$ . When the dimension of the matrix arguments remains fixed, there is a large and well-established literature on the asymptotics of  ${}_pF_q(a, b; \Psi, \Lambda)$  for

large parameters and norm of the matrix arguments; see [24] for a review. In contrast, relatively little is known about when the dimensionality of the matrix arguments  $\Psi, \Lambda$  diverge to infinity. It is this regime we study in this paper, noting that, in single-spiked models, the matrix argument  $\Psi$  has rank one. This allows us to represent  ${}_pF_q(a, b; \Psi, \Lambda)$  in the form of a contour integral of a hypergeometric function with a single scalar argument. Such a representation implies contour integral representations for the corresponding likelihood ratios.

LEMMA 1. Assume that  $p \leq \min\{n_1, n_2\}$ . Let  $\mathcal{K}$  be a contour in the complex plane  $\mathbb{C}$  that starts at  $-\infty$ , encircles 0 and  $\lambda_1, \dots, \lambda_p$  counterclockwise and returns to  $-\infty$ . Then,

$$(9) \quad L(\theta; \Lambda) = \frac{\Gamma(s+1)\alpha(\theta)q_s}{\Psi_{11}^s 2\pi i} \int_{\mathcal{K}} {}_pF_q(a-s, b-s; \Psi_{11}z) \prod_{j=1}^p (z-\lambda_j)^{-1/2} dz,$$

where  $s = p/2 - 1$ , the values of  $\alpha(\theta), \Psi_{11}, a, b, p$ , and  $q$  for the different cases are given in Table 2;  $a-s$  and  $b-s$  denote vectors with elements  $a_j - s$  and  $b_j - s$ , respectively, and

$$q_s = \prod_{j=1}^p \frac{\Gamma(a_j - s)}{\Gamma(a_j)} \prod_{i=1}^q \frac{\Gamma(b_i)}{\Gamma(b_i - s)}.$$

In cases SigD and CCA, we require, in addition, that the contour  $\mathcal{K}$  does not intersect  $[\Psi_{11}^{-1}, \infty)$  which ensures the analyticity of the integrand in an open subset of  $\mathbb{C}$  that includes  $\mathcal{K}$ .

The statement of the lemma immediately follows from [12], Proposition 1, and from equation (6). Our next step is to apply the Laplace approximation to integrals (9). To this end, we shall transform the right-hand side of (9) so that it has a ‘‘Laplace form’’

$$(10) \quad L(\theta; \Lambda) = \sqrt{\pi p} \frac{1}{2\pi i} \int_{\mathcal{K}} \exp\{-(p/2)f(z; \theta)\} g(z; \theta) dz.$$

The dependence on  $\theta$  will usually not be shown explicitly. Leaving  $\sqrt{\pi p}/(2\pi i)$  separate from  $g(z)$  allows us to choose  $f(z)$  and  $g(z)$  that are bounded in probability and makes some of the expressions below more compact. In order to apply the Laplace approximation, we shall deform the contour of integration so that it passes through a critical point  $z_0$  of  $f(z)$  and is such that  $\text{Re } f(z)$  is strictly increasing as  $z$  moves away from  $z_0$  along the contour, at least in a vicinity of  $z_0$ .

4.1. *The Laplace form.* We shall transform (9) to (10) in three steps. As a result, functions  $f$  and  $g$  will have the forms of a sum and a product,

$$(11) \quad \begin{aligned} f(z) &= f_c + f_e(z) + f_h(z) \quad \text{and} \\ g(z) &= g_c \times g_e(z) \times g_h(z), \end{aligned}$$

where  $f_c$  and  $g_c$  do not depend on  $z$ . The subscripts (c,e,h) are mnemonic for ‘‘coefficient’’, ‘‘eigenvalues’’ and ‘‘hypergeometric.’’

First, using the definitions of  $\alpha(\theta), q_s, \Psi_{11}$  and employing Stirling’s approximation, we obtain a decomposition

$$(12) \quad \frac{\Gamma(s+1)\alpha(\theta)q_s}{\sqrt{\pi p}\Psi_{11}^s} = \exp\{-(p/2)f_c\} g_c,$$

where  $g_c$  remains bounded as  $\mathbf{n}, p \rightarrow \infty$ . The values of  $f_c$  and  $g_c$  are given in Table 4. Details of the derivation are given in Section 4.1 in the Supplementary Material [19].

TABLE 4

Values of  $f_c$  and  $\check{g}_c = g_c/(1 + o(1))$  for the different cases. The terms  $o(1)$  do not depend on  $\theta$  and converge to zero as  $\mathbf{n}, p \rightarrow_{\mathcal{Y}} \infty$ . In the table,  $l(\theta) = 1 + (1 + \theta)c_2/c_1$  and  $r^2 = c_1 + c_2 - c_1c_2$

Case	$f_c$	$\check{g}_c = g_c/(1 + o(1))$
SMD	$1 + \theta^2/2 + \log \theta$	$\theta$
PCA	$1 + \frac{1-c_1}{c_1} \log(1 + \theta) + \log \frac{\theta}{c_1}$	$\theta(1 + \theta)^{-1}c_1^{-1}$
SigD	$f_c^{\text{PCA}} + f_{10}$	$\check{g}_c^{\text{PCA}} \check{g}_{10}$
REG <sub>0</sub>	$1 + \frac{\theta+c_1}{c_1} + \log \frac{\theta}{c_1} + \frac{1-c_1}{c_1} \log(1 - c_1)$	$\theta c_1^{-1}(1 - c_1)^{-1/2}$
REG	$f_c^{\text{REG}_0} + f_{10}$	$\check{g}_c^{\text{REG}_0} \check{g}_{10}$
CCA	$f_c^{\text{REG}} + f_{21}$	$\check{g}_c^{\text{REG}} \check{g}_{10}/l(\theta)$
	$f_{10} = -1 - \frac{r^2}{c_1c_2} \log \frac{r^2}{c_1+c_2} + \log \frac{c_1+c_2}{c_1}$	$\check{g}_{10} = c_1^{-1}r(c_1 + c_2)^{1/2}$
	$f_{21} = -1 - \frac{\theta}{c_1} - \frac{r^2}{c_1c_2} \log \frac{r^2}{c_1l(\theta)}$	

Second, we consider the decomposition

$$(13) \quad \prod_{j=1}^p (z - \lambda_j)^{-1/2} = \exp\{-(p/2) f_e(z)\} g_e(z),$$

where

$$(14) \quad f_e(z) = \int \ln(z - \lambda) dF_c(\lambda)$$

and

$$(15) \quad g_e(z) = \exp\left\{-(p/2) \int \ln(z - \lambda) d(\hat{F}(\lambda) - F_c(\lambda))\right\}.$$

For  $f_e(z)$  and  $g_e(z)$  to be well defined, we need  $z$  not to belong to the support of  $F_c$ , which we assume. In addition,  $z \notin \text{supp}(\hat{F})$  since by definition contour  $\mathcal{K}$  encircles it. Note that  $g_e(z)$  is the exponent of a linear spectral statistic which converges to a Gaussian random variable as  $\mathbf{n}, p \rightarrow_{\mathcal{Y}} \infty$  under the null hypothesis.

Third and finally, we describe a decomposition

$$(16) \quad {}_pF_q(a - s, b - s; \Psi_{11z}) = \exp\{-(p/2) f_h(z)\} g_h(z).$$

For the  $q = 0$  cases, the corresponding  ${}_pF_q$  can be expressed in terms of elementary functions. Indeed,  ${}_0F_0(z) = e^z$  and  ${}_1F_0(a; z) = (1 - z)^{-a}$ . We set

$$(17) \quad f_h(z) = \begin{cases} -z\theta & \text{for SMD,} \\ -z\theta/(c_1(1 + \theta)) & \text{for PCA,} \\ \ln[1 - c_2z\theta/\{c_1(1 + \theta)\}]r^2/(c_1c_2) & \text{for SigD} \end{cases}$$

and

$$(18) \quad g_h(z) = \begin{cases} 1 & \text{for SMD and PCA,} \\ [1 - c_2z\theta/\{c_1(1 + \theta)\}]^{-1} & \text{for SigD.} \end{cases}$$

Unfortunately, for the  $q = 1$  cases, the corresponding  ${}_pF_q$  do not admit exact representations in terms of elementary functions. Therefore, we shall consider their asymptotic approximations instead. Let

$$m = (n_1 - p)/2 \quad \text{and} \quad \kappa = (n - p)/(n_1 - p).$$

Further, let

$$(19) \quad \eta_j = \begin{cases} z\theta/(1 - c_1)^2 & \text{for } j = 0, \\ z\theta c_2/[c_1(1 - c_1)] & \text{for } j = 1, \\ z\theta c_2^2/[c_1^2 l(\theta)] & \text{for } j = 2, \end{cases}$$

where

$$(20) \quad l(\theta) = 1 + (1 + \theta)c_2/c_1.$$

With this notation, we have

$$(21) \quad {}_pF_q = \begin{cases} {}_0F_1(m + 1; m^2\eta_0) \equiv F_0 & \text{for REG}_0, \\ {}_1F_1(m\kappa + 1; m + 1; m\eta_1) \equiv F_1 & \text{for REG}, \\ {}_2F_1(m\kappa + 1, m\kappa + 1; m + 1; \eta_2) \equiv F_2 & \text{for CCA}. \end{cases}$$

The function  $F_0(z)$  can be expressed in terms of the modified Bessel function of the first kind  $I_m(\cdot)$ ; see [1], equation 9.6.47, as

$$(22) \quad F_0 = \Gamma(m + 1)(m^2\eta_0)^{-m/2} I_m(2m\eta_0^{1/2}).$$

This representation allows us to use a known uniform asymptotic approximation of the Bessel function [1], equation 9.7.7, to obtain Lemma 2, proven in Section 4.2 in the Supplementary Material [19]. To state it, let

$$(23) \quad \varphi_0(t) = \ln t - t - \eta_0/t + 1 \quad \text{and} \quad t_0 = (1 + \sqrt{1 + 4\eta_0})/2.$$

Further, for any  $\delta > 0$ , let  $\Omega_{0\delta}$  be the set of  $\eta_0 \in \mathbb{C}$  such that

$$|\arg \eta_0| \leq \pi - \delta \quad \text{and} \quad \eta_0 \neq 0.$$

LEMMA 2. *As  $m \rightarrow \infty$ , we have*

$$(24) \quad F_0 = (1 + 4\eta_0)^{-1/4} \exp\{-m\varphi_0(t_0)\}(1 + o(1)).$$

*The convergence  $o(1) \rightarrow 0$  holds uniformly with respect to  $\eta_0 \in \Omega_{0\delta}$  for any  $\delta > 0$ .*

To foreshadow our results for  $F_1(z)$  and  $F_2(z)$ , we note that the right-hand side of (24) can be formally linked, via (22), to the saddle-point approximation of the integral representation; see [38], p. 181,

$$I_m(2m\eta_0^{1/2}) = \frac{\eta_0^{m/2} e^m}{2\pi i} \int_{-\infty}^{(0+)} \exp\{-m\varphi_0(t)\} t^{-1} dt.$$

Point  $t_0$  can be interpreted as a saddle point of  $\varphi_0(t)$ , and the term  $(1 + 4\eta_0)^{-1/4}$  in (24) can be interpreted as a factor of  $(\varphi_0''(t_0))^{-1/2}$ .

Turning now to functions  $F_1(z)$  and  $F_2(z)$ , to obtain uniform asymptotic approximations, we use the contour integral representations; see [29], equations 13.4.9 and 15.6.2,

$$(25) \quad F_j = \frac{C_m}{2\pi i} \int_0^{(1+)} \exp\{-m\varphi_j(t)\} \psi_j(t) dt,$$

where

$$(26) \quad C_m = \frac{\Gamma(m + 1)\Gamma(m(\kappa - 1) + 1)}{\Gamma(m\kappa + 1)},$$

$$(27) \quad \varphi_j(t) = \begin{cases} -\eta_j t - \kappa \ln t + (\kappa - 1) \ln(t - 1) & \text{for } j = 1, \\ -\kappa \ln(t/(1 - \eta_j t)) + (\kappa - 1) \ln(t - 1) & \text{for } j = 2 \end{cases}$$

and

$$(28) \quad \psi_j(t) = \begin{cases} (t - 1)^{-1} & \text{for } j = 1, \\ (t - 1)^{-1}(1 - \eta_j t)^{-1} & \text{for } j = 2. \end{cases}$$

For  $j = 2$ , the contour does not encircle  $1/\eta_2$ , and the representation is valid for  $\eta_2$  such that  $|\arg(1 - \eta_2)| < \pi$ . We derive a saddle-point approximation to the integral in (25) to be summarized in Lemma 3 below. The relevant saddle points are

$$(29) \quad t_j = \begin{cases} \frac{1}{2\eta_j} \{ \eta_j - 1 + \sqrt{(\eta_j - 1)^2 + 4\kappa\eta_j} \} & \text{for } j = 1, \\ \frac{1}{2\eta_j(\kappa - 1)} \{ -1 + \sqrt{1 + 4\kappa(\kappa - 1)\eta_j} \} & \text{for } j = 2. \end{cases}$$

We shall need the following additional notation. Let

$$(30) \quad \omega_j = \arg \varphi_j''(t_j) + \pi \quad \text{and} \quad \omega_{0j} = \arg(t_j - 1),$$

where the branches of  $\arg(\cdot)$  are chosen so that  $|\omega_j + 2\omega_{0j}| \leq \pi/2$ .

LEMMA 3. *As  $m \rightarrow \infty$ , we have for  $j = 1, 2$*

$$(31) \quad F_j = C_m \psi_j(t_j) e^{-i\omega_j/2} |2\pi m \varphi_j''(t_j)|^{-1/2} \exp\{-m\varphi_j(t_j)\} (1 + o(1)).$$

The convergence  $o(1) \rightarrow 0$  holds uniformly with respect to  $(\kappa, \eta) \in \Omega_{j\delta}$  for any  $\delta > 0$ , where  $\Omega_{j\delta}$  are as defined in Table 5.

Pointwise asymptotic approximation (31) was established in [35] for  $j = 1$ , and in [33, 34] for  $j = 2$ . However, those papers do not study the uniformity of the approximation error which is important for our analysis. Lemma 3 is proved at length in Section 4.3 in the Supplementary Material [19]. It is fair to say that the corresponding derivations constitute the technically most challenging part of our analysis. This further highlights the technical difficulties that occur when going from SMD, PCA and SigD cases to REG<sub>0</sub>, REG and CCA.

Using Lemmas 2 and 3 and Stirling’s approximation

$$(32) \quad C_m = \frac{\sqrt{\pi p(1 - c_1)}}{r} \exp\{m(\kappa - 1) \ln(\kappa - 1) - m\kappa \ln \kappa\} (1 + o(1)),$$

we set the components of the “Laplace form” (16) of  ${}_pF_q$  for the  $q = 1$  cases as follows:

$$(33) \quad f_h(z) = \begin{cases} \frac{1 - c_1}{c_1} \varphi_0(t_0) & \text{REG}_0, \\ \frac{1 - c_1}{c_1} (\varphi_j(t_j) + \kappa \ln \kappa - (\kappa - 1) \ln(\kappa - 1)) & \text{REG, CCA} \end{cases}$$

TABLE 5  
Definition of  $\Omega_{j\delta}$  from Lemma 3

$\Omega_{j\delta} = \Omega_\delta \cap \hat{\Omega}_{j\delta}$ with the following $\Omega_\delta$ and $\hat{\Omega}_{j\delta}$	
Set	Definition: pairs $(x, z) \in \mathbb{R} \times \mathbb{C}$ s.t.
$\Omega_\delta$	$\delta \leq x - 1 \leq 1/\delta,  z  \leq 1/\delta, \text{ and } \inf_{y \in \mathbb{R} \setminus [0, \infty)}  z - y  \geq \delta$
$\hat{\Omega}_{1\delta}$	$\text{Re } z \geq -2x + 1$
$\hat{\Omega}_{2\delta}$	$\inf_{y \in \mathbb{R} \setminus (-\infty, 1]}  z - y  \geq \delta$ and $x$ is unconstrained

and

$$(34) \quad g_h(z) = \begin{cases} (1 + 4\eta_0)^{-1/4}(1 + o(1)) & \text{REG}_0, \\ \sqrt{c_1/r^2} e^{-i\omega_j/2} |\phi_j''(t_j)|^{-1/2} \psi_j(t_j)(1 + o(1)) & \text{REG, CCA.} \end{cases}$$

To express  $t_j$  and  $\eta_j$  in terms of  $z$ , one should use (29) and (19). We do not need to know how exactly the  $o(1)$  in (34) depend on  $z$ . For our purposes, the knowledge of the fact that  $o(1)$  are analytic functions of  $\eta_j$  that converge to zero uniformly with respect to  $(\kappa, \eta_j) \in \Omega_{j\delta}$  is sufficient. The analyticity of  $o(1)$  follows from the analyticity of the functions on the left-hand sides and of the factors of  $1 + o(1)$  on the right-hand sides of the equations (24) and (31).

*Confluences of functions  $f$ .* As  $c_2 \rightarrow 0$  with  $c_1$  held fixed, we have

$$(35) \quad \begin{aligned} f^{\text{SigD}}(z) &\rightarrow f^{\text{PCA}}(z), \\ f^{\text{REG}}(z), f^{\text{CCA}}(z) &\rightarrow f^{\text{REG}_0}(z). \end{aligned}$$

Also, as  $c_1 \rightarrow 0$ ,

$$(36) \quad f^{\text{PCA}}(z), f^{\text{REG}_0}(z) \rightarrow f^{\text{SMD}}(z),$$

after making the substitutions  $\theta \rightarrow \sqrt{c_1}\theta$  and  $z \rightarrow \sqrt{c_1}z + 1$  on the left-hand side. Some details appear in Section 4.4 in the Supplementary Material [19].

4.2. *Saddle points and contours of steep descent.* We shall now show how to deform contours  $\mathcal{K}$  in (10) into the contours of steep descent. First, we find saddle points of functions  $f(z)$  for each of the six cases. Note that

$$-df_e(z)/dz = \int (\lambda - z)^{-1} dF_c(\lambda) = m_c(z),$$

the Stieltjes transform of  $F_c$ . Although the Stieltjes transform is formally defined on  $\mathbb{C}^+$ , the definition remains valid on the part of the real line outside the support  $[b_-, b_+]$  of  $F_c$ . Since we assume that  $p \leq n_1$ ,  $F_c$  does not have any nontrivial mass at 0.

To find saddle points  $z_0$  of  $f(z)$ , we therefore solve the equation

$$(37) \quad m_c(z) = df_h(z)/dz.$$

A proof of the following lemma appears in Section 4.5 in the Supplementary Material [19].

LEMMA 4. *The saddle points  $z_0(\theta, \mathbf{c})$  of  $f(z)$  satisfy*

$$(38) \quad z_0(\theta, \mathbf{c}) = \begin{cases} \theta + 1/\theta & \text{for SMD,} \\ (1 + \theta)(\theta + c_1)/\theta & \text{for PCA and REG}_0, \\ (1 + \theta)(\theta + c_1)/[\theta l(\theta)] & \text{for SigD, REG, and CCA.} \end{cases}$$

For  $\theta \in (0, \bar{\theta}_c)$ ,  $z_0 > b_+$ , where  $\bar{\theta}_c$  is the threshold corresponding to  $F_c$  which is an analogue of the threshold  $\bar{\theta}_\gamma \equiv \bar{\theta}$  corresponding to  $F_\gamma$  given in Table 3.

As  $c_2 \rightarrow 0$  while  $c_1$  stays constant, the value of  $z_0$  for SigD, REG and CCA converges to that for PCA and  $\text{REG}_0$ . The latter value in its turn converges to the value of  $z_0$  for SMD when  $c_1 \rightarrow 0$ , after the transformations  $\theta \mapsto \sqrt{c_1}\theta$  and  $z_0 \mapsto \sqrt{c_1}z_0 + 1$ . Precisely, solving equation

$$\sqrt{c_1}z_0 + 1 = (1 + \sqrt{c_1}\theta)(\sqrt{c_1}\theta + c_1)/(\sqrt{c_1}\theta)$$

for  $z_0$  and taking limit as  $c_1 \rightarrow 0$  yields  $z_0 = \theta + 1/\theta$ .

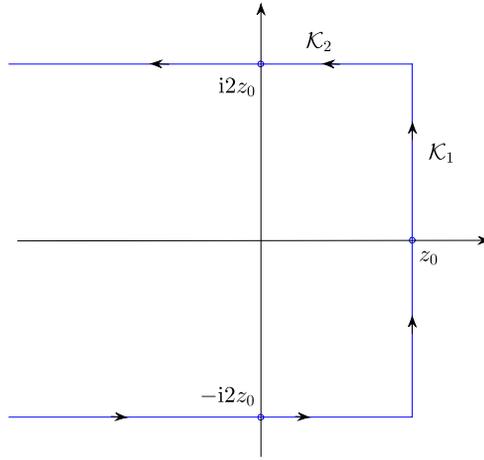


FIG. 2. Deformed contour  $\mathcal{K}$  for SMD, PCA and SigD.

REMARK 5. For all of the six cases that we study,  $f(z_0)$  equals zero. Section 4.6 in the Supplementary Material [19] has a verification of this important fact.

REMARK 6. As  $\mathbf{n}, p \rightarrow_{\mathcal{Y}} \infty, z_0(\theta, \mathbf{c}) \rightarrow z_0(\theta, \mathcal{Y}) > \beta_+$ , where the latter inequality holds for any  $\theta \in (0, \bar{\theta})$ . Since  $\lambda_1 \xrightarrow{\text{a.s.}} \beta_+$ , the inequality  $z_0(\theta, \mathbf{c}) > \lambda_1$  must hold with probability approaching one as  $\mathbf{n}, p \rightarrow_{\mathcal{Y}} \infty$ .

For the rest of the paper, assume that  $\theta \in (0, \bar{\theta})$ . We deform contour  $\mathcal{K}$  in (10) so that it passes through the saddle point  $z_0$  as follows. Let  $\mathcal{K} = \mathcal{K}_+ \cup \mathcal{K}_-$ , where  $\mathcal{K}_-$  is the complex conjugate of  $\mathcal{K}_+$  and  $\mathcal{K}_+ = \mathcal{K}_1 \cup \mathcal{K}_2$ . For SMD, PCA and SigD, let

$$(39) \quad \mathcal{K}_1 = \{z_0 + it : 0 \leq t \leq 2z_0\} \quad \text{and}$$

$$(40) \quad \mathcal{K}_2 = \{x + i2z_0 : -\infty < x \leq z_0\}.$$

The deformed contour is shown on Figure 2.

Note that the singularities of the integrand in (10) are situated at  $z = \lambda_j$  (plus an additional singularity at  $z = c_1(1 + \theta)/(\theta c_2) < z_0$  for SigD). Since  $z_0 > \lambda_1$  holds with probability approaching one as  $\mathbf{n}, p \rightarrow_{\mathcal{Y}} \infty$ , Cauchy’s theorem ensures that the deformation of the contour does not change the value of  $L(\theta; \Lambda)$  with probability approaching one as  $\mathbf{n}, p \rightarrow_{\mathcal{Y}} \infty$ .

Strictly speaking, the deformation of the contour is not continuous because  $\mathcal{K}_+$  does not approach  $\mathcal{K}_-$  at  $-\infty$ . In particular, in contrast to the original contour, the deformed one is not “closed” at  $-\infty$ . Nevertheless, such an “opening up” at  $-\infty$  does not lead to the change of the value of the integral because the integrand converges fast to zero in absolute value as  $\text{Re } z \rightarrow -\infty$ .

REMARK 7. In the event of asymptotically negligible probability that the deformed contour  $\mathcal{K}$  does not encircle all  $\lambda_j$ , we not only lose the equality (10) but also face the difficulty that function  $g(z)$  ceases to be well defined as the definition of  $g_e(z)$  contains a logarithm of a nonpositive number. To eliminate any ambiguity, if such an event holds, we shall redefine  $g_e(z)$  as unity.

For  $\text{REG}_0$  and CCA, let

$$z_1 = \begin{cases} -(1 - c_1)^2/[4\theta] & \text{for } \text{REG}_0, \\ -c_1(1 - c_1)^2 l(\theta)/[4\theta r^2] & \text{for CCA,} \end{cases}$$

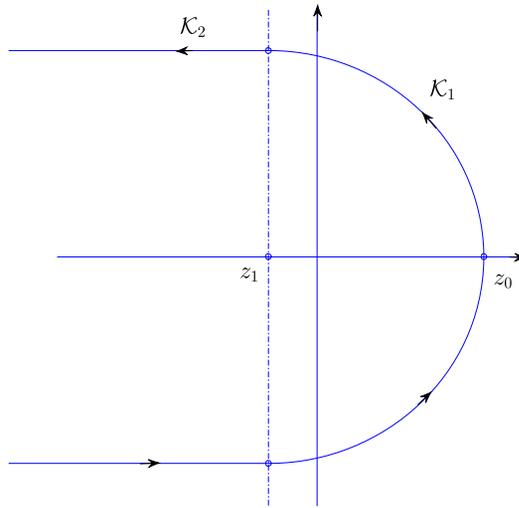


FIG. 3. Deformed contour  $\mathcal{K}$  for  $REG_0$  and CCA.

and let

$$\begin{aligned} \mathcal{K}_1 &= \{z_1 + |z_0 - z_1| \exp\{i\gamma\} : \gamma \in [0, \pi/2]\} \quad \text{and} \\ \mathcal{K}_2 &= \{z_1 - x + |z_0 - z_1| \exp\{i\pi/2\} : x \geq 0\}. \end{aligned}$$

The corresponding contour  $\mathcal{K}$  is shown on Figure 3. Similarly to the SMD, PCA and SigD cases, the deformation of the contour in (10) to  $\mathcal{K}$  does not change the value of  $L(\theta; \Lambda)$  with probability approaching one as  $\mathbf{n}, p \rightarrow \gamma \infty$ .

For REG, deformed contour  $\mathcal{K}$  in  $z$ -plane is simpler to describe as an image of a contour  $\mathcal{C}$  in  $\tau$ -plane, where  $\tau = \eta_1 t_1$  with

$$(41) \quad \eta_1 = z\theta c_2 / [c_1(1 - c_1)]$$

and  $t_1$  as defined in (29). Let  $\mathcal{C} = \mathcal{C}_+ \cup \mathcal{C}_-$ , where  $\mathcal{C}_-$  is the complex conjugate of  $\mathcal{C}_+$  and  $\mathcal{C}_+ = \mathcal{C}_1 \cup \mathcal{C}_2$ , and let

$$\begin{aligned} \mathcal{C}_1 &= \{-\kappa + |\tau_0 + \kappa| \exp\{i\gamma\} : \gamma \in [0, \pi/2]\} \quad \text{and} \\ \mathcal{C}_2 &= \{-\kappa - x + |\tau_0 + \kappa| \exp\{i\pi/2\} : x \geq 0\}, \end{aligned}$$

where  $\tau_0 = (\theta + c_1)/(1 - c_1)$ .

Using (41) and the identity

$$(42) \quad \eta_1 = \tau(\tau + 1)/(\tau + \kappa),$$

we obtain

$$(43) \quad z = \frac{c_1(1 - c_1)}{\theta c_2} \frac{\tau(\tau + 1)}{\tau + \kappa}.$$

We define the deformed contour  $\mathcal{K}$  in  $z$ -plane as the image of  $\mathcal{C}$  under the transformation  $\tau \rightarrow z$  given by (43). The parts  $\mathcal{K}_+, \mathcal{K}_-, \mathcal{K}_1$  and  $\mathcal{K}_2$  of  $\mathcal{K}$  are defined as the images of the corresponding parts of  $\mathcal{C}$ . Note that  $\tau_0$  is transformed to  $z_0$  so that  $\mathcal{K}$  passes through the saddle point  $z_0$ .

The next lemma, proven in Section 4.7 in the Supplementary Material [19], shows that  $\mathcal{K}_1$  are contours of steep descent of  $-\text{Re } f(z)$  for all the six cases, SMD, PCA, SigD,  $REG_0$ , REG and CCA.

LEMMA 8. For any of the six cases that we study, as  $z$  moves along the corresponding  $\mathcal{K}_1$  away from  $z_0$ ,  $-\text{Re } f(z)$  is strictly decreasing.

**5. Laplace approximation.** The goal of this section is to derive Laplace approximations to the integral (9) for the six cases that we study. First, consider a general integral

$$I_{p,\omega} = \int_{\mathcal{K}_{p,\omega}} e^{-p\phi_{p,\omega}(z)} \chi_{p,\omega}(z) dz,$$

where  $p$  is large,  $\omega \in \Omega \subset \mathbb{R}^k$  is a  $k$ -dimensional parameter and  $\mathcal{K}_{p,\omega}$  is a path in  $\mathbb{C}$  that starts at  $a_{p,\omega}$  and ends at  $b_{p,\omega}$ . We allow  $\chi_{p,\omega}(z)$  to be a random element of the normed space of continuous functions on  $\mathcal{K}_{p,\omega}$  with the supremum norm. Assume that there is a domain  $T_{p,\omega} \supset \mathcal{K}_{p,\omega}$  on which for sufficiently large  $p$ ,  $\phi_{p,\omega}(z)$  and  $\chi_{p,\omega}(z)$  are single-valued holomorphic functions of  $z$ , in the case of  $\chi_{p,\omega}$  with probability increasing to 1.

We describe an extension of the Laplace approximation detailed by Olver [28], p. 127, to a situation in which functions  $\phi$ ,  $\chi$  and contour  $\mathcal{K}$  depend on  $p$  and  $\omega$  and, in addition,  $\chi$  is random. In Olver’s original theorem, both functions and contour are fixed. In what follows, however, we omit subscripts  $p$  and  $\omega$  from  $\phi_{p,\omega}$ ,  $\chi_{p,\omega}$ ,  $\mathcal{K}_{p,\omega}$ , etc. to lighten notation.

Suppose that  $\phi'(z) = 0$  at  $z_0$ , which is an interior point of  $\mathcal{K}$ , and suppose that  $\text{Re } \phi(z)$  is strictly increasing as  $z$  moves away from  $z_0$  along the path. In other words, the path  $\mathcal{K}$  is a contour of steep descent of  $-\text{Re } \phi(z)$ . Denote a closed segment of  $\mathcal{K}$  contained between  $z_1$  and  $z_2$  as  $[z_1, z_2]_{\mathcal{K}}$ . Similarly, denote the segments that exclude one or both endpoints as  $(z_1, z_2]_{\mathcal{K}}$ ,  $(z_1, z_2)_{\mathcal{K}}$  and  $(z_1, z_2)_{\mathcal{K}}$ . Let  $\beta$  be the limiting value of  $\arg(z - z_0)$  on the principal branch as  $z \rightarrow z_0$  along  $(z_0, b)_{\mathcal{K}}$ . Finally, let  $\phi_s$  and  $\chi_s$  be the coefficients in the power series representations

$$(44) \quad \phi(z) = \sum_{s=0}^{\infty} \phi_s (z - z_0)^s, \quad \chi(z) = \sum_{s=0}^{\infty} \chi_s (z - z_0)^s.$$

We assume that there exist positive constants  $C_1, \dots, C_4$  that do not depend on  $p$  or  $\omega$ , such that for all  $\omega \in \Omega$ , for sufficiently large  $p$ :

A0. The length of the path  $\mathcal{K}$  is bounded, uniformly over  $\omega \in \Omega$  and all sufficiently large  $p$ . Furthermore,

$$\sup_{z \in (z_0, b)_{\mathcal{K}}} |z - z_0| > C_1 \quad \text{and} \quad \sup_{z \in (a, z_0)_{\mathcal{K}}} |z - z_0| > C_1,$$

- A1. Functions  $\phi(z)$  and  $\chi(z)$  are holomorphic in the ball  $|z - z_0| \leq C_1$ ,
- A2. The coefficient  $\phi_2$  satisfies  $C_2 \leq |\phi_2| \leq C_3$ ,
- A3. The third derivative of  $\phi(z)$  satisfies inequality

$$\sup_{|z - z_0| \leq C_1} |d^3 \phi(z) / dz^3| \leq C_4,$$

A4. For any positive  $\varepsilon < C_1$ , which does not depend on  $p$  and  $\omega$ , and for all  $z_1 \in \mathcal{K}$  such that  $|z_1 - z_0| = \varepsilon$ , there exist positive constants  $C_5, C_6$ , such that

$$\text{Re}(\phi(z_1) - \phi_0) > C_5 \quad \text{and} \quad |\text{Im}(\phi(z_1) - \phi_0)| < C_6,$$

A5. For a subset  $\Theta$  of  $\mathbb{C}$  that consists of all points whose Euclidean distance from  $\mathcal{K}$  is no larger than  $C_1$ ,

$$\sup_{z \in \Theta} |\chi(z)| = O_P(1)$$

as  $p \rightarrow \infty$ , where  $O_P(1)$  is uniform in  $\omega \in \Omega$ .

Assumptions A0–A5 ensure that Olver’s proof of the Laplace approximation theorem (Theorem 7.1 on p. 127 of Olver [28]) can be extended to cases where functions  $\phi(z)$  and  $\chi(z)$ , as well as the contour  $\mathcal{K}$ , depend on  $p$  and  $\omega$ . Note that in Olver’s notations,  $\phi(z)$ ,  $\chi(z)$  and  $p$  are, respectively,  $p(t)$ ,  $q(t)$  and  $z$ .

The first part of A0, which requires the boundedness of  $|\mathcal{K}|$ , taken together with A5, and the assumption that  $\mathcal{K}$  is a contour of steep descent guarantee the absolute convergence of the integral  $\int_{\mathcal{K}} e^{-p(\phi(z)-\phi_0)} \chi(z) dz$ , in probability. The second part of A0 ensures that as  $p \rightarrow \infty$ ,  $\mathcal{K}$  does not collapse to a point.

Assumption A1 excludes situations where  $z_0$  approaches singular points of  $\phi(z)$  or  $\chi(z)$  as  $p \rightarrow \infty$ . Assumption A2 guarantees that the second derivative of  $\phi(z)$  at  $z_0$  does not degenerate to 0 or infinity as  $p \rightarrow \infty$ . Assumption A3 implies that  $|\phi(z) - \phi(z_0)|$  can be bounded from below by a fixed quadratic function of  $z$  in a vicinity of  $z_0$  as  $p \rightarrow \infty$ . This ensures a regular behavior of function  $(\phi(z) - \phi(z_0))^{1/2}$ . Assumption A4 implies that  $|\arg(\phi(z) - \phi(z_0))| < \pi/2$  is some neighborhood of  $z_0$  as  $p \rightarrow \infty$ . We need this condition to be able to use an asymptotic expansion of an incomplete Gamma function in our proofs (Section 5.1 of the Supplementary Material [19]). Assumption A5 ensures that  $|\chi(z)|$  remains bounded in probability as  $p \rightarrow \infty$ .

LEMMA 9. *Under assumptions A0–A5, for any positive integer  $k$ , as  $p \rightarrow \infty$ , we have*

$$I_{p,\omega} = 2e^{-p\phi_0} \left[ \sum_{s=0}^{k-1} \Gamma\left(s + \frac{1}{2}\right) \frac{a_{2s}}{p^{s+1/2}} + \frac{O_{\mathbb{P}}(1)}{p^{k+1/2}} \right],$$

where  $O_{\mathbb{P}}(1)$  is uniform in  $\omega \in \Omega$  and the coefficients  $a_{2s}$  can be expressed through  $\phi_s$  and  $\chi_s$  defined above. In particular, we have  $a_0 = \chi_0/[2\phi_2^{1/2}]$ , where  $\phi_2^{1/2} = \exp\{(\log |\phi_2| + i \arg \phi_2)/2\}$  with the branch of  $\arg \phi_2$  chosen so that  $|\arg \phi_2 + 2\beta| \leq \pi/2$ .

Lemma 9 is proved in Section 5.1 in the Supplementary Material [19]. We use it to obtain the Laplace approximation to

$$(45) \quad L_1(\theta; \Lambda) = \sqrt{\pi p} \frac{1}{2\pi i} \int_{\mathcal{K}_1 \cup \bar{\mathcal{K}}_1} e^{-(p/2)f(z)} g(z) dz.$$

Then, we show that  $L_1(\theta; \Lambda)$  asymptotically dominates the “residual”  $L(\theta; \Lambda) - L_1(\theta; \Lambda)$ . For this analysis, it is important to know the values of  $f(z_0)$  and  $d^2 f(z_0)/dz^2$ . As was mentioned in Remark 5,  $f(z_0) = 0$  for all the six cases that we study. The values of  $d^2 f(z_0)/dz^2$  are derived in Section 5.2 in the Supplementary Material [19]. All of them are negative. The explicit form of  $D_2 \equiv \theta^2(-d^2 f(z_0)/dz^2)^{-1}$ , which is somewhat shorter than that for  $d^2 f(z_0)/dz^2$ , is reported in Table 6. We formulate the main result of this section in the following theorem, proven in Section 5.3 in the Supplementary Material [19].

TABLE 6  
The values of  $D_2 \equiv \theta^2(-d^2 f(z_0)/dz^2)^{-1}$  for the different cases

Case	Value of $D_2$	Case	Value of $D_2$
SMD	$1 - \theta^2$	REG <sub>0</sub>	$c_1(1 + c_1 + 2\theta)(c_1 - \theta^2)$
PCA	$c_1(c_1 - \theta^2)(1 + \theta)^2$	REG	$c_1 h(c_1 + \theta + (1 + \theta)l)/l^4$
SigD	$r^2 h(1 + \theta)^2/l^4$	CCA	$c_1^2 h(2(c_1 + \theta) + l(1 - c_1))/(l^3(c_1 + c_2))$
	$l \equiv l(\theta) = 1 + (1 + \theta)c_2/c_1$		$h \equiv h(\theta) = c_1 + c_2(1 + \theta)^2 - \theta^2$

**THEOREM 10.** *Suppose that the null hypothesis holds, that is,  $\theta_0 = 0$ . Let  $\bar{\theta}$  be the threshold corresponding to  $F_{\mathbf{y}}$ , as given in Table 3, and let  $\varepsilon$  be an arbitrarily small fixed positive number. Then, for any  $\theta \in (0, \bar{\theta} - \varepsilon]$ , as  $\mathbf{n}, p \rightarrow_{\mathbf{y}} \infty$ , we have*

$$(46) \quad L(\theta; \Lambda) = \frac{g(z_0)}{\sqrt{-d^2 f(z_0)/dz^2}} + O_{\mathbf{P}}(p^{-1}),$$

where  $O_{\mathbf{P}}(p^{-1})$  is uniform in  $\theta \in (0, \bar{\theta} - \varepsilon]$  and the principal branch of the square root is taken.

**6. Asymptotics of LR.** Combining the results of Theorem 10 with the definitions of  $g(z)$  and the values of  $-d^2 f(z_0)/dz^2$ , given in Table 6, it is straightforward to establish the following theorem, details in Section 6.1 in the Supplementary Material [19]. Let

$$\Delta_p(\theta) = p \int \ln(z_0(\theta) - \lambda) d(\hat{F}(\lambda) - F_{\mathbf{c}}(\lambda)).$$

In accordance with Remark 7, we define  $\Delta_p(\theta)$  as zero in the event of asymptotically negligible probability that  $z_0 \leq \lambda_1$ .

**THEOREM 11.** *Suppose that the null hypothesis holds, that is,  $\theta_0 = 0$ . Let  $\bar{\theta}$  be the threshold corresponding to  $F_{\mathbf{y}}$ , as given in Table 3, and let  $\varepsilon$  be an arbitrarily small fixed positive number. Then, for any  $\theta \in (0, \bar{\theta} - \varepsilon]$ , as  $\mathbf{n}, p \rightarrow_{\mathbf{y}} \infty$ , we have*

$$L(\theta; \Lambda) = \exp\left\{-\frac{1}{2}\Delta_p(\theta) + \frac{1}{2}\ln(1 - [\delta_p(\theta)]^2)\right\}(1 + o_{\mathbf{P}}(1)),$$

where

$$\delta_p(\theta) = \begin{cases} \theta & \text{for SMD,} \\ \theta/\sqrt{c_1} & \text{for PCA and REG}_0, \\ \theta r/(c_1 l(\theta)) & \text{for SigD, REG, and CCA,} \end{cases}$$

$$r^2 = c_1 + c_2 - c_1 c_2 \text{ and } o_{\mathbf{P}}(1) \text{ is uniform in } \theta \in (0, \bar{\theta} - \varepsilon].$$

Statistic  $\Delta_p(\theta)$  is a linear spectral statistic. As follows from the CLT for such statistics derived by [5, 6] and [39] for the Semicircle, Marchenko–Pastur and Wachter limiting distributions  $F_{\mathbf{c}}$ , respectively, statistic  $\Delta_p(\theta)$  weakly converges to a Gaussian process indexed by  $\theta \in (0, \bar{\theta} - \varepsilon]$ . The explicit form of the mean and the covariance structure can be obtained from the general formulae for the asymptotic mean and covariance of linear spectral statistics given in [6], Theorem 1.1, for SMD, in [5], Theorem 1.1, for PCA and  $\text{REG}_0$ , and in [39], Theorem 4.1 and Example 4.1, for the remaining cases. Section 6.2 in the Supplementary Material [19] provides details on the use of [5, 6, 39] to establish convergence of  $\Delta_p(\theta)$ , and the use of Theorem 11 to obtain the following theorem:

**THEOREM 12.** *Suppose that the null hypothesis holds, that is,  $\theta_0 = 0$ . Let  $\bar{\theta}$  be the threshold corresponding to  $F_{\mathbf{y}}$ , as given in Table 3, and let  $\varepsilon$  be an arbitrarily small fixed positive number. Further, let  $C[0, \bar{\theta} - \varepsilon]$  be the space of continuous functions on  $[0, \bar{\theta} - \varepsilon]$  equipped with the supremum norm. Then,  $\ln L(\theta; \Lambda)$  viewed as random elements of  $C[0, \bar{\theta} - \varepsilon]$  converge weakly to  $\mathcal{L}(\theta)$  with Gaussian finite dimensional distributions such that*

$$\mathbb{E}\mathcal{L}(\theta) = \frac{1}{4}\ln(1 - \delta^2(\theta))$$

and

$$\text{Cov}(\mathcal{L}(\theta_1), \mathcal{L}(\theta_2)) = -\frac{1}{2} \ln(1 - \delta(\theta_1)\delta(\theta_2))$$

with

$$\delta(\theta) = \begin{cases} \theta & \text{for SMD,} \\ \theta/\sqrt{\gamma_1} & \text{for PCA and REG}_0, \\ \theta\rho/(\gamma_1 + \gamma_2 + \theta\gamma_2) & \text{for SigD, REG, and CCA.} \end{cases}$$

Here,  $\rho, \gamma_1, \gamma_2$  are the limits of  $r, c_1, c_2$  as  $\mathbf{n}, p \rightarrow_{\mathbf{y}} \infty$ .

Let  $\{\mathbb{P}_{p,\theta}\}$  and  $\{\mathbb{P}_{p,0}\}$  be the sequences of measures corresponding to the joint distributions of  $\lambda_1, \dots, \lambda_p$  when  $\theta_0 = \theta$  and when  $\theta_0 = 0$ , respectively. Then, Theorem 12 implies, via Le Cam’s first lemma, the mutual contiguity of  $\{\mathbb{P}_{p,\theta}\}$  and  $\{\mathbb{P}_{p,0}\}$  as  $\mathbf{n}, p \rightarrow_{\mathbf{y}} \infty$ , for each  $\theta < \bar{\theta}$ . This reveals the statistical meaning of the phase transition thresholds as the upper boundaries of the contiguity regions for spiked models.

The precise form of the autocovariance of  $\mathcal{L}(\theta)$  shows that,<sup>1</sup> although the experiment of observing  $\lambda_1, \dots, \lambda_p$  is asymptotically normal, it does not converge to a Gaussian shift experiment. In particular, the optimality results available for Gaussian shifts cannot be used in our framework. To analyze asymptotic risks of various statistical problems related to the experiment of observing  $\lambda_1, \dots, \lambda_p$ , one should directly use Theorem 12.

Here, we use it to derive the asymptotic power envelopes for tests of the null hypothesis  $\theta_0 = 0$  against the point alternative  $\theta_0 > 0$ . By the Neyman–Pearson lemma, the most powerful test would reject the null when  $\ln L(\theta; \Lambda)$  is above a critical value. By Theorem 12 and Le Cam’s third lemma (see [37], Chapter 6),

$$\ln L(\theta; \Lambda) \xrightarrow{d} N\left(\pm\frac{1}{4} \ln(1 - \delta^2(\theta)), -\frac{1}{2} \ln(1 - \delta^2(\theta))\right)$$

with the plus sign holding under the null and the minus under the alternative. This implies the following theorem.

**THEOREM 13.** *Let  $\bar{\theta}$  be the threshold corresponding to  $F_{\mathbf{y}}$ , as given in Table 3. For any  $\theta \in [0, \bar{\theta})$ , the value of the asymptotic power envelope for the tests of the null  $\theta_0 = 0$  against the alternative  $\theta_0 > 0$ , which are based on  $\lambda_1, \dots, \lambda_p$  and have asymptotic size  $\alpha$ , is given by*

$$\text{PE}(\theta) = 1 - \Phi[\Phi^{-1}(1 - \alpha) - \sigma(\theta)], \quad \sigma(\theta) = \sqrt{-\frac{1}{2} \ln(1 - \delta^2(\theta))}.$$

Here,  $\Phi$  denotes the standard normal cumulative distribution function. For  $\theta \geq \bar{\theta}$ , the value of the asymptotic power envelope equals one.

The envelopes differ only for cases with different limiting spectral distributions: Semicircle, Marchenko–Pastur and Wachter, denoted  $\text{PE}^{\text{SC}}(\theta)$ ,  $\text{PE}^{\text{MP}}(\theta, \gamma_1)$  and  $\text{PE}^{\text{W}}(\theta, \mathbf{y})$ , respectively. Figure 4 shows the graphs of the envelopes for  $\alpha = 0.05$  and  $\gamma_1 = \gamma_2 = 0.9$ . Such large values of  $\gamma_1$  and  $\gamma_2$  correspond to situations where the dimensionality  $p$  is not very different from the degrees of freedom  $n_1$  and  $n_2$ .

---

<sup>1</sup>Fyodorov, Khoruzhenko and Simm [17] have an interesting discussion of ubiquity of random processes with logarithmic covariance structure in physics and engineering applications. In that paper such processes appear as limiting objects related to the behavior of the characteristic polynomials of large matrices from Gaussian Unitary Ensemble.

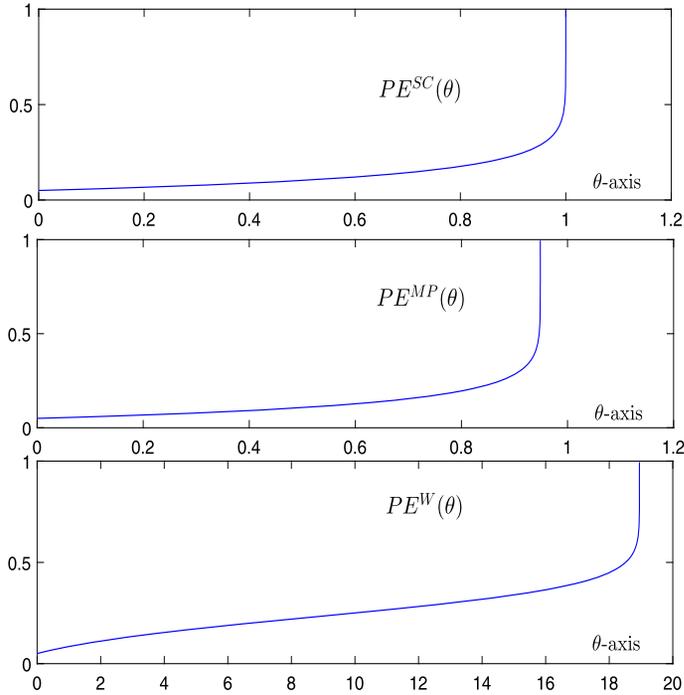


FIG. 4. The asymptotic power envelopes  $PE^{SC}(\theta)$ ,  $PE^{MP}(\theta, \gamma_1)$  and  $PE^W(\theta, \boldsymbol{\gamma})$  for  $\alpha = 0.05$ ,  $\gamma_1 = \gamma_2 = 0.9$ .

Envelope  $PE^{MP}(\theta, \gamma_1)$  can be obtained from  $PE^W(\theta, \boldsymbol{\gamma})$  by sending  $\gamma_2$  to zero. Further,  $PE^{SC}(\theta)$  can be obtained from  $PE^{MP}(\theta, \gamma_1)$  by transformation  $\theta \mapsto \sqrt{\gamma_1}\theta$ . Further, note the difference in the horizontal scale of the bottom panel of Figure 4 relative to the two other panels. For  $\gamma_1 = \gamma_2 = 0.9$ , the phase transition threshold corresponding to the Wachter distribution is relatively large. It equals  $(\gamma_2 + \rho)/(1 - \gamma_2) \approx 18.9$ . Moreover, the value of  $PE^W(\theta)$  becomes substantially larger than the nominal size  $\alpha = 0.05$  for  $\theta$  that are situated far below this threshold. This suggests that the information in all the eigenvalues  $\lambda_1, \dots, \lambda_p$  might be effectively used to detect spikes that are small relative to the phase transition threshold in two sample problems. We leave a confirmation or rejection of this speculation for future research.

**7. Concluding remarks.** Note that Theorem 12 establishes the weak convergence of the log likelihood ratio viewed as a random element of the space of continuous functions. This is much stronger than simply the convergence of the finite dimensional distributions of the log likelihood process. In particular, the theorem can be used to find the asymptotic distribution of the supremum of the likelihood ratio and, thus, to find the asymptotic critical values of the likelihood ratio test. It also can be used to construct asymptotic confidence intervals for a subcritical spike as well as to describe the asymptotic properties of its maximum likelihood estimator. We do not pursue this line of research here but provide a general outline.

Consider the log likelihood ratio  $\ln L(\theta; \Lambda) - \ln L(\theta_0; \Lambda)$ . According to Theorem 12, this ratio converges to  $X(\theta) \equiv \mathcal{L}(\theta) - \mathcal{L}(\theta_0)$ . By Le Cam’s third lemma, under the null hypothesis that the true value of the spike equals  $\theta_0$ ,  $X(\theta)$  is a Gaussian process with mean

$$\mathbb{E}X(\theta) = \frac{1}{4} \ln \frac{(1 - \delta^2(\theta))(1 - \delta^2(\theta_0))}{(1 - \delta(\theta)\delta(\theta_0))^2}$$

and covariance function

$$\text{Cov}(X(\theta_1), X(\theta_2)) = -\frac{1}{2} \ln \frac{(1 - \delta(\theta_1)\delta(\theta_2))(1 - \delta(\theta_0)\delta(\theta_0))}{(1 - \delta(\theta_1)\delta(\theta_0))(1 - \delta(\theta_2)\delta(\theta_0))}.$$

An approximation to the distribution of the supremum of such a process over  $\theta \in [0, \bar{\theta} - \varepsilon]$  can be obtained via simulation. Alternatively, it might be expressed analytically in the form of converging Rice series (see, e.g., [2]). Quantiles of the distribution can be used as asymptotic critical values for the likelihood ratio test of the hypothesis  $\theta = \theta_0$ . Inverting the test, we obtain asymptotic confidence intervals for the true value of a subcritical spike.

The maximum likelihood estimator for the spike,  $\hat{\theta}_{ML}$ , equals the arg max of  $\ln L(\theta; \Lambda) - \ln L(\theta_0; \Lambda)$  over  $\theta \in [0, \bar{\theta} - \varepsilon]$ . By Lemma 2.6 of [20], the limiting process  $X(\theta)$  achieves maximum at a unique point with probability one. Therefore, by the argmax continuous mapping theorem,  $\hat{\theta}_{ML}$  converges in distribution to the arg max of  $X(\theta)$ . The distribution of such an arg max can be approximated using simulations.

Unfortunately, the quality of the estimator  $\hat{\theta}_{ML}$  cannot be “good.” For PCA we were able to prove that no estimator of  $\theta$  has root mean squared error better than the order of magnitude of the subcritical parameter  $\theta$ . This result will appear in another work.

Our asymptotic discussion of James’ framework can likely be extended to a fixed number of subcritical spikes. Such an extension would require developing Laplace approximations to multiple contour integrals and uniform approximations to hypergeometric functions of two matrix arguments in terms of elementary functions. Alternatively, one may employ large deviation analysis of spherical integrals, as in [32], which covers the PCA case. As this paper is already long, the extension will appear separately.

Addressing the case of slowly increasing number of spikes may require new techniques, perhaps, similar to those developed in [14]. In such a case, relatively little is known even about the phase transition phenomenon. For sample covariance matrices, Theorem 1.1 of [4] can be used to show that the phase transition still happens at the usual threshold  $\bar{\theta} = \sqrt{\gamma_1}$ . However, it is not clear whether the experiments of observing sample covariance eigenvalues corresponding to the null case and an alternative with a growing number of subcritical eigenvalues remain mutually contiguous.

Note that, intuitively, the asymptotic power of the likelihood ratio test of the null hypothesis of no spikes against the alternative of one spike should not decrease, if the rank-one assumption on the alternative is wrong and there are additional spikes. In Section 7.1 in the Supplementary Material [19], we confirm this intuition for SMD and PCA cases. A confirmation or refutation of the intuition for the other James’ cases requires further analysis and is left for future research.

In this paper we make the assumption that  $n_2 \geq p$  to ensure the invertibility of matrix  $E$  in (1) with probability one. However, we also make the assumption  $n_1 \geq p$ , mostly to simplify our exposition. It can probably be lifted without a substantial reformulation of the problem. Indeed, for SMD the assumption is irrelevant. For PCA, the case  $p > n_1$  was explicitly covered in [31]. For REG<sub>0</sub>, the assumption can be relaxed using the symmetry of the problem. Specifically, the canonical REG<sub>0</sub> problem tests restriction  $M = 0$  in the model  $Y = M + \varepsilon$ , where each matrix is  $n_1 \times p$  and  $\varepsilon$  has i.i.d. standard normal components. Clearly, interchanging roles of  $n_1$  and  $p$  yield essentially the same problem.

For CCA, the sample canonical correlations are not well defined for  $p > n_1$ . For SigD, our derivations (not reported here) show that the equivalent of (6) for  $p > n_1$  involves the hypergeometric function  ${}_2F_1$ . Therefore, SigD with  $p > n_1$  represents the fifth, rather than the second, group of multivariate statistical problems according to James’ [18] classification. For REG, an equivalent of (6) for  $p > n_1$  can be obtained using [18], equation (74). However, further analysis of SigD and REG in the situation where  $p > n_1$  needs more substantial changes to our derivations. We leave such an analysis for future research.

Finally, many existing results in the random matrix literature do not require that the data are Gaussian. This suggests that some results about tests for the presence of the spikes in the data may remain valid without the Gaussian assumption. We hope that the results of this paper might provide a benchmark for such future studies.

**Acknowledgements.** We thank two anonymous referees, an Associate Editor and the Editor for careful reading of the manuscript and helpful comments.

I. M. Johnstone is supported in part by NSF DMS 0906812 and 1407813.

A. Onatski is supported by the J.M. Keynes Fellowship Fund, University of Cambridge.

## SUPPLEMENTARY MATERIAL

**Supplement to “Testing in high-dimensional spiked models”** (DOI: [10.1214/18-AOS1697SUPP](https://doi.org/10.1214/18-AOS1697SUPP); .pdf). The supplement has proofs for all results in the paper, organized by section for easier cross-reference.

## REFERENCES

- [1] ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55*. U.S. Government Printing Office, Washington, DC. [MR0167642](https://doi.org/10.1214/18-AOS1697SUPP)
- [2] AZAÏS, J.-M. and WSCHBOR, M. (2002). The distribution of the maximum of a Gaussian process: Rice method revisited. In *In and Out of Equilibrium (Mambucaba, 2000)*. *Progress in Probability* **51** 321–348. Birkhäuser, Boston, MA. [MR1901961](https://doi.org/10.1214/18-AOS1697SUPP)
- [3] BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. [MR1711663](https://doi.org/10.1214/18-AOS1697SUPP)
- [4] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. [MR1617051](https://doi.org/10.1214/aop/1022855421) <https://doi.org/10.1214/aop/1022855421>
- [5] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. [MR2040792](https://doi.org/10.1214/aop/1078415845) <https://doi.org/10.1214/aop/1078415845>
- [6] BAI, Z. D. and YAO, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** 1059–1092. [MR2189081](https://doi.org/10.3150/bj/1137421640) <https://doi.org/10.3150/bj/1137421640>
- [7] BAI, Z. D. and YIN, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. [MR0929083](https://doi.org/10.1214/18-AOS1697SUPP)
- [8] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. [MR2279680](https://doi.org/10.1016/j.jmva.2005.08.003) <https://doi.org/10.1016/j.jmva.2005.08.003>
- [9] BAIK, J. and WANG, D. (2013). On the largest eigenvalue of a Hermitian random matrix model with spiked external source II: Higher rank cases. *Int. Math. Res. Not. IMRN* **2013** 3304–3370. [MR3085761](https://doi.org/10.1093/imrn/rns136) <https://doi.org/10.1093/imrn/rns136>
- [10] BAO, Z., HU, J., PAN, G. and ZHOU, W. (2014). Canonical correlation coefficients of high-dimensional normal vectors: Finite rank case. Available at [arXiv:1407.7194](https://arxiv.org/abs/1407.7194).
- [11] CAI, J.-F., CANDÈS, E. J. and SHEN, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20** 1956–1982. [MR2600248](https://doi.org/10.1137/080738970) <https://doi.org/10.1137/080738970>
- [12] DHARMAWANSA, P. and JOHNSTONE, I. M. (2014). Joint density of eigenvalues in spiked multivariate models. *Stat* **3** 240–249. [MR4027340](https://doi.org/10.1002/sta.4.58) <https://doi.org/10.1002/sta.4.58>
- [13] DHARMAWANSA, P., JOHNSTONE, I. M. and ONATSKI, A. (2014). Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios. Available at [arXiv:1411.3875](https://arxiv.org/abs/1411.3875).
- [14] DOBRIBAN, E. (2017). Sharp detection in PCA under correlations: All eigenvalues matter. *Ann. Statist.* **45** 1810–1833. [MR3670197](https://doi.org/10.1214/16-AOS1514) <https://doi.org/10.1214/16-AOS1514>
- [15] DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. [MR3269984](https://doi.org/10.1214/14-AOS1257) <https://doi.org/10.1214/14-AOS1257>
- [16] FUJIKOSHI, Y., ULYANOV, V. V. and SHIMIZU, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. [MR2640807](https://doi.org/10.1002/9780470539873) <https://doi.org/10.1002/9780470539873>
- [17] FYODOROV, Y. V., KHORUZHENKO, B. A. and SIMM, N. J. (2016). Fractional Brownian motion with Hurst index  $H = 0$  and the Gaussian unitary ensemble. *Ann. Probab.* **44** 2980–3031. [MR3531684](https://doi.org/10.1214/15-AOP1039) <https://doi.org/10.1214/15-AOP1039>
- [18] JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.* **35** 475–501. [MR0181057](https://doi.org/10.1214/aoms/1177703550) <https://doi.org/10.1214/aoms/1177703550>
- [19] JOHNSTONE, I. M. and ONATSKI, A. (2020). Supplement to “Testing in high-dimensional spiked models.” <https://doi.org/10.1214/18-AOS1697SUPP>

- [20] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 <https://doi.org/10.1214/aos/1176347498>
- [21] LELARGE, M. and MIOLANE, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields* **173** 859–929. MR3936148 <https://doi.org/10.1007/s00440-018-0845-x>
- [22] MAÏDA, M. (2007). Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.* **12** 1131–1150. MR2336602 <https://doi.org/10.1214/EJP.v12-438>
- [23] MEHTA, M. L. (2004). *Random Matrices*, 3rd ed. *Pure and Applied Mathematics (Amsterdam)* **142**. Elsevier/Academic Press, Amsterdam. MR2129906
- [24] MUIRHEAD, R. J. (1978). Latent roots and matrix variates: A review of some asymptotic results. *Ann. Statist.* **6** 5–33. MR0458719
- [25] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0652932
- [26] NADAKUDITI, R. R. (2014). OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Trans. Inf. Theory* **60** 3002–3018. MR3200641 <https://doi.org/10.1109/TIT.2014.2311661>
- [27] NADAKUDITI, R. R. and SILVERSTEIN, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE J. Sel. Top. Signal Process.* **4** 468–480.
- [28] OLVER, F. W. J. (1997). *Asymptotics and Special Functions*. A K Peters, Natick, MA.
- [29] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W., eds. (2010). *NIST Handbook of Mathematical Functions*. Cambridge Univ. Press, Cambridge. MR2723248
- [30] ONATSKI, A. (2007). Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise. Manuscript, University of Cambridge.
- [31] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 <https://doi.org/10.1214/13-AOS1100>
- [32] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multi-spiked case. *Ann. Statist.* **42** 225–254. MR3189485 <https://doi.org/10.1214/13-AOS1181>
- [33] PARIS, R. B. (2013). Asymptotics of the Gauss hypergeometric function with large parameters, I. *J. Class. Anal.* **2** 183–203. MR3322252 <https://doi.org/10.7153/jca-02-15>
- [34] PARIS, R. B. (2013). Asymptotics of the Gauss hypergeometric function with large parameters, II. *J. Class. Anal.* **3** 1–15. MR3322253 <https://doi.org/10.7153/jca-03-01>
- [35] PASSEMIER, D., MCKAY, M. R. and CHEN, Y. (2015). Asymptotic linear spectral statistics for spiked Hermitian random matrices. *J. Stat. Phys.* **160** 120–150. MR3357970 <https://doi.org/10.1007/s10955-015-1233-x>
- [36] SHABALIN, A. A. and NOBEL, A. B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *J. Multivariate Anal.* **118** 67–76. MR3054091 <https://doi.org/10.1016/j.jmva.2013.03.005>
- [37] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 <https://doi.org/10.1017/CBO9780511802256>
- [38] WATSON, G. N. (1944). *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press, Cambridge; The Macmillan Company, New York. MR0010746
- [39] ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional  $F$ -matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263 <https://doi.org/10.1214/11-AIHP414>