Submitted to the Annals of Statistics
arXiv: math.PR/0000000

SIGNAL DETECTION IN HIGH DIMENSION: THE MULTISPIKED CASE

By Alexei Onatski[*], Marcelo J. Moreira[†] and Marc Hallin[§]

University of Cambridge, FGV/EPGE, and Université libre de Bruxelles and Princeton University.

This paper applies Le Cam's asymptotic theory of statistical experiments to the signal detection problem in high-dimension. We consider the problem of testing the null hypothesis of sphericity of a high-dimensional covariance matrix against an alternative of (unspecified) multiple symmetry-breaking directions (*multispiked* alternatives). Simple analytical expressions for the Gaussian asymptotic power envelope and the asymptotic powers of previously proposed tests are derived. Those asymptotic powers remain valid for non-Gaussian data satisfying mild moment restrictions. They appear to lie very substantially below the Gaussian power envelope, at least for small values of the number of symmetry-breaking directions. In contrast, the asymptotic power of Gaussian likelihood ratio tests based on the eigenvalues of the sample covariance matrix are shown to be very close to the envelope. Although based on Gaussian likelihoods, those tests remain valid under non-Gaussian densities satisfying mild moment conditions. The results of this paper extend to the case of multispiked alternatives and possibly non-Gaussian densities the findings of an earlier study (Onatski, Moreira and Hallin 2013a) of the single-spiked case. The methods we are using here, however, are entirely new, as the Laplace approximation methods considered in the single-spiked context do not extend to the multispiked case.

1. Introduction. In a recent paper, Onatski, Moreira and Hallin (2013a), hereafter OMH, analyze the asymptotic power of statistical tests for the detection of a signal in spherical real-valued Gaussian data as the sample size and the dimension of the observations increase at the same rate. This paper generalizes the OMH alternatives of a single symmetry-breaking direction

[*]Supported by the J.M. Keynes Fellowships Fund, University of Cambridge.

[†]Supported by CNPq and the NSF via grant number SES-0819761.

[§]Supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, a Discovery Project grant from Australian Research Council, and the 2012–2017 Belgian Science Policy Office Interuniversity Attraction Poles.

[§]ECARES, ECORE, Académie Royale de Belgique, and CenTER, Tilburg University.

AMS 2000 subject classifications: Primary 62H15, 62B15; secondary 41A60

Keywords and phrases: sphericity tests, large dimensionality, asymptotic power, spiked covariance, contiguity, power envelope

(*single-spiked* alternatives) to alternatives of multiple symmetry-breaking directions (*multispiked* alternatives), which is more relevant for applications.

Contemporary tests of sphericity in high dimension (see Ledoit and Wolf (2002), Srivastava (2005), Schott (2006), Bai et al. (2009), Chen et al. (2010), and Cai and Ma (2012)) consider general alternatives to the null of sphericity. Our interest in alternatives with only a few contaminating signals stems from the fact that in many applications (such as speech recognition, macroeconomics, finance, wireless communication, genetics, physics of mixture, and statistical learning), a few latent variables typically explain a large proportion of the variation in high-dimensional data; see Baik and Silverstein (2006) for references. As a possible explanation of this fact, Johnstone (2001) introduces the *spiked covariance model*, where all eigenvalues of the population covariance matrix of high-dimensional data are equal except for a small fixed number of distinct "spike eigenvalues." The alternative to the null of sphericity considered in this paper coincides with Johnstone's model.

The extension from the single-spiked alternatives of OMH to the multispiked alternatives considered here is not straightforward. The difficulty arises because the extension of the main technical tool in OMH (Lemma 2), which analyzes high-dimensional spherical integrals, to integrals over high-dimensional real Stiefel manifolds obtained in Onatski (2012) is not easily amenable to the Laplace approximation method used in OMH. Therefore, in this paper, we develop a completely different technique, inspired from the large deviation analysis of spherical integrals by Guionnet and Maïda (2005), hereafter GM.

Let us describe the setting and main results in more detail. Suppose that the data consist of n_p independent observations X_t, $t = 1, ..., n_p$ of a p-dimensional Gaussian vector with mean zero and positive definite covariance matrix Σ. Let $\Sigma = \sigma^2 (I_p + V \operatorname{diag}(h) V')$, where I_p is the p-dimensional identity matrix, σ is a scalar, $\operatorname{diag}(h)$ an $r \times r$ diagonal matrix with elements $h_j \geq 0$, $j = 1, ..., r$, along the diagonal, and V a $(p \times r)$-dimensional parameter normalized so that $V'V = I_r$. We are interested in the asymptotic power of tests of the null hypothesis $H_0 : h = 0$ against the alternative $H_1 : h \in (\mathbb{R}^+)^r \setminus \{0\}$, based on the eigenvalues of the sample covariance matrix when n_p and p both tend to infinity, in such a way that $p/n_p \to c$ with $0 < c < \infty$, an asymptotic regime which we abbreviate into $n_p, p \to_c \infty$. The matrix V is an unspecified nuisance parameter, the columns of which indicate the directions of the perturbations of sphericity. We consider the cases of known and unknown σ^2. For the sake of simplicity, this introduction only discusses the case of known $\sigma^2 = 1$.

Let $\lambda_p = (\lambda_{p1}, ..., \lambda_{pm})$, where λ_{pj} denotes the j-th largest sample covariance eigenvalue and $m = \min(n_p, p)$. We begin our analysis with a study of the asymptotic properties of the likelihood ratio process $\{L(\tau; \lambda_p); \tau \in [0, \bar{\tau}]^r\}$, where $\bar{\tau} \in [0, \sqrt{c})$, and $L(\tau; \lambda_p)$ is defined as the ratio of the density of λ_p under a point alternative hypothesis $h = \tau$ to that under the null hypothesis H_0, computed at λ_p. An exact formula for $L(\tau; \lambda_p)$ involves the integral $\int_{\mathcal{O}(p)} e^{\mathrm{tr}(AQBQ')} (dQ)$ over the orthogonal group $\mathcal{O}(p)$, where the $p \times p$ matrix A has a deficient rank r. In the single-spiked case ($r = 1$), OMH link this integral to the confluent form of the Lauricella function, and use this link to establish a representation of the integral in the form of a contour integral (Wang (2012) and Mo (2012) also obtain this contour integral representation for $r = 1$ using different derivations). The Laplace approximation to the contour integral is then used to study the asymptotic behavior of $\{L(\tau; \lambda_p); \tau \in [0, \bar{\tau}]^r\}$ under the null.

Onatski (2012) generalizes the contour integral representation of $L(\tau; \lambda_p)$ to the multispiked case ($r > 1$). Such a generalization allows extending the OMH results to the multi-spiked context for *complex-valued data*. Unfortunately, for *real-valued data*, this generalization is not straightforwardly amenable to the Laplace approximation method. Therefore, we consider a totally different approach. For the $r = 1$ case, GM use large deviation techniques to derive a second-order asymptotic expansion of $\int_{\mathcal{O}(p)} e^{\mathrm{tr}(AQBQ')} (dQ)$ as the non-zero eigenvalues of A diverge to infinity (see their Theorem 3). We extend GM's second-order expansion to the $r > 1$ case, and use that extension to derive the asymptotics of $\{L(\tau; \lambda_p); \tau \in [0, \bar{\tau}]^r\}$.

We show that, for any $\bar{\tau} \in [0, \sqrt{c})$, the sequence of log-likelihood ratio processes $\{\ln L(\tau; \lambda_p); \tau \in [0, \bar{\tau}]^r\}$ converges weakly, under the null hypothesis H_0 as $n_p, p \to_c \infty$, to a Gaussian process $\{\mathcal{L}_\lambda(\tau); \tau \in [0, \bar{\tau}]^r\}$. The limiting process has mean $\mathrm{E}[\mathcal{L}_\lambda(\tau)] = \sum_{i,j=1}^r \ln(1 - \tau_i\tau_j/c)/4$ and autocovariance function $\mathrm{Cov}(\mathcal{L}_\lambda(\tau), \mathcal{L}_\lambda(\tilde{\tau})) = -\sum_{i,j=1}^r \ln(1 - \tau_i\tilde{\tau}_j/c)/2$. That convergence entails the weak convergence of the τ-indexed statistical experiments $\mathcal{E}(\tau; \lambda_p)$ under which the eigenvalues $\lambda_{p1}, ..., \lambda_{pm}$ generated by the parameter value $h = \tau$ are observed, i.e. the statistical experiments with log-likelihood process $\{\ln L(\tau; \lambda_p); \tau \in [0, \bar{\tau}]^r\}$ (see p. 126 of van der Vaart (1998)). Although Gaussian, the limiting log-likelihood ratio process $\{\mathcal{L}_\lambda(\tau)\}$ is not that of a *Gaussian shift*, and the statistical experiments $\mathcal{E}(\tau; \lambda_p)$ under study are not *locally asymptotically normal* (LAN). As a consequence, the existence of asymptotically optimal procedures remains an open problem. Still, the asymptotic behavior of the log-likelihood process $\{\ln L(\tau; \lambda_p)\}$ has important implications:

(a) it follows from Le Cam's first lemma (see p. 88 of van der Vaart (1998)) that the sequences of joint distributions of the sample covariance eigenvalues under the null (h=0) and under alternatives ($h = \tau \in [0, \sqrt{c})^r$) are *mutually contiguous* as $n_p, p \to_c \infty$;

(b) as a consequence, although their existence can be detected, spiked eigenvalues, in this contiguity region, cannot be estimated consistently;

(c) the *asymptotic power envelope* for α-level λ-based tests for H_0 against H_1—namely, the mapping from $\tau \in [0, \bar{\tau}]^r$ to the maximum asymptotic power achievable, under Gaussian assumptions, at a point alternative of the form $h = \tau$—can be constructed by combining the Neyman–Pearson lemma and Le Cam's third lemma; this asymptotic power envelope constitutes, pointwise and under Gaussian assumptions, an upper bound for the asymptotic powers of all α-level λ-based tests, but also for all tests that are invariant under left orthogonal transformations of the observations;

(d) analytic expressions also can be obtained via Le Cam's third lemma for the asymptotic powers of the Gaussian likelihood ratio test and several existing tests of sphericity—we focus on the tests proposed by Ledoit and Wolf (2002), Bai et al (2009), and Cai and Ma (2012); we show that those expressions moreover remain valid under non-Gaussian densities satisfying mild moment restrictions.

These results are stronger than those that can be found in the literature. Baik et al (2005) and Féral and Péché (2009), for instance, provide results on the asymptotic behavior of the r largest empirical eigenvalues that preclude, below the phase transition, the existence of any consistent estimator based on these r leading eigenvalues. Instead, we analyze the log-likelihood processes and the convergence, in the Le Cam sense, of the statistical experiments in which *all* empirical eigenvalues are observed. Contiguity (a) does not just imply inconsistency of the leading sample eigenvalues; it entails (b) that, although their existence can be detected, no consistent estimation of the population spiked eigenvalues is possible below the phase transition threshold. That impossiblity property is also in agreement with more recent results by Cai et al (2013). They show how sparsity assumptions are restoring the consistency of the empirical eigenvalues. For the estimation of Σ, they obtain a minimax risk rate (under spectral norm loss function) as a function of a sparsity index k (see their Equation (8)); for $k \approx p$ (no sparsity), that minimax rate no longer goes to zero.

The asymptotic power results (d) allow for interesting performance comparisons. In particular, it appears that the asymptotic powers of the Ledoit and Wolf (2002), Bai et al (2009), and Cai and Ma (2012) tests are quite substantially lower than the corresponding (though unachievable) asymptotic power envelope values, whereas the asymptotic powers of the likelihood ratio tests are close to the same values—at least, for small values of r. While performance assessments involving the power envelope are valid under Gaussian assumptions, the power comparisons between likelihood ratio tests and other procedures are meaningful under the aforementioned milder moment assumptions.

The rest of the paper is organized as follows. Section 2 establishes the weak convergence of the log-likelihood ratio process to a Gaussian process. Section 3 analyzes the asymptotic powers of various sphericity tests, derives the asymptotic power envelope, and proves its validity for general invariant tests. Section 4 concludes. All proofs are given in the Appendix.

2. Asymptotics of likelihood ratio processes.

2.1. *Asymptotic representation of the log-likelihood process.* Let the data-generating process be

$$X = \sigma \left(I_p + V \operatorname{diag}(h) V'\right)^{1/2} \varepsilon, \tag{2.1}$$

where ε is a $p \times n_p$ matrix with i.i.d. entries with zero mean and unit variance. For now, we assume that the entries of ε are standard normal:

ASSUMPTION G: *ε has i.i.d. standard normal entries ε_{ij}.*

Later on, we shall relax that assumption. Denote by $\lambda_{p1} \geq ... \geq \lambda_{pp}$ the ordered eigenvalues of XX'/n_p, and write $\lambda_p = (\lambda_{p1}, ..., \lambda_{pm})$, where $m = \min\{n_p, p\}$. Similarly, let $\mu_{pi} = \lambda_{pi}/(\lambda_{p1} + ... + \lambda_{pp})$, $i = 1, ..., m$, and $\mu_p = (\mu_{p1}, ..., \mu_{p,m-1})$.

As explained in the introduction, our goal is to study the asymptotic power, as $n_p, p \to_c \infty$, of the eigenvalue-based tests of $H_0 : h = 0$ against $H_1 : h \in (\mathbb{R}^+)^r \setminus \{0\}$. If σ^2 is known, this testing problem is invariant with respect to left and right orthogonal transformations of X; sufficiency and invariance arguments (see Appendix A.10 for details) lead to considering tests based on λ_p only. If σ^2 is unknown, the same problem is invariant with respect to left and right orthogonal transformations of X and multiplications by non-zero scalars; sufficiency and invariance arguments (see Appendix A.10) lead to considering tests based on μ_p only. Note that the sufficiency and invariance arguments eliminate the nuisance parameter V. Moreover, the distribution of μ_p does not depend on σ^2, whereas, if σ^2 is specified, we can always normalize λ_p dividing it by σ^2. Therefore, without loss of generality, we henceforth assume that $\sigma^2 = 1$.

Let us denote the joint density of $\lambda_{p1}, ..., \lambda_{pm}$ at $\tilde{x} = (x_1, ..., x_m) \in (\mathbb{R}^+)^m$ as $f_{\lambda p}(\tilde{x}; h)$, and that of $\mu_{p1}, ..., \mu_{p,m-1}$ at $\tilde{y} = (y_1, ..., y_{m-1}) \in (\mathbb{R}^+)^{m-1}$

as $f_{\mu p}(\tilde{y};h)$. We then have

$$f_{\lambda p}(\tilde{x};h)=\gamma_p(\tilde{x})\prod_{j=1}^{r}(1+h_j)^{-n_p/2}\int_{\mathcal{O}(p)}e^{-\frac{n_p}{2}\operatorname{tr}(\Pi Q'\mathcal{X}Q)}(\mathrm{d}Q), \tag{2.2}$$

where $\gamma_p(\tilde{x})$ depends on n_p, p, and $\tilde{x}$, but not on h; Π and $\mathcal{X}$ are the $(p\times p)$ diagonal matrices

$$\operatorname{diag}((1+h_1)^{-1},...,(1+h_r)^{-1},1,...,1) \quad \text{and} \quad \operatorname{diag}(x_1,...,x_m,0,...,0),$$

respectively, $\mathcal{O}(p)$ is the set of all $p\times p$ orthogonal matrices, and $(\mathrm{d}Q)$ is the invariant measure on the orthogonal group $\mathcal{O}(p)$, normalized to make the total measure unity. Formula (2.2) is a special case of the density given in James (1964, p. 483) for $n_p\geq p$, and follows from Theorems 2 and 6 in Uhlig (1994) for $n_p<p$.

Let $x=x_1+...+x_m$ and $y_i=x_i/x$. Note that the Jacobian of the coordinate change from $(x_1,...,x_m)$ to $(y_1,...,y_{m-1},x)$ is x^{m-1}. Changing variables in (2.2) and integrating x out, we obtain

$$f_{\mu p}(\tilde{y};h)=\gamma_p(\tilde{y})\prod_{j=1}^{r}(1+h_j)^{-n_p/2}\int_0^\infty x^{\frac{n_p p}{2}-1}\int_{\mathcal{O}(p)}e^{-\frac{n_p}{2}x\operatorname{tr}(\Pi Q'\mathcal{Y}Q)}(\mathrm{d}Q)\,\mathrm{d}x, \tag{2.3}$$

where $\mathcal{Y}=\operatorname{diag}(y_1,...,y_m,0,...,0)$ is a $(p\times p)$ diagonal matrix.

Consider the Gaussian likelihood ratios $L_p(\tau;\lambda_p)=f_{\lambda p}(\lambda_p;\tau)/f_{\lambda p}(\lambda_p;0)$ and $L_p(\tau;\mu)=f_{\mu p}\left(\mu_p;\tau\right)/f_{\mu p}\left(\mu_p;0\right)$, where $\tau\in(\mathbb{R}^+)^r$. When ε is non-Gaussian, these ratios are to be interpreted as *pseudo-Gaussian likelihood ratios*. Formulae (2.2) and (2.3) imply the following.

PROPOSITION 1. *Define* $\Lambda_p=\operatorname{diag}(\lambda_{p1},...,\lambda_{pp})$, $S_p=\lambda_{p1}+...+\lambda_{pp}$, *and let* D_p *be the* $p\times p$ *diagonal matrix* $\operatorname{diag}\left(\frac{1}{2c_p}\frac{\tau_1}{1+\tau_1},...,\frac{1}{2c_p}\frac{\tau_r}{1+\tau_r},0,...,0\right)$, *where* $c_p=p/n_p$. *Then,*

$$L_p(\tau;\lambda_p)=\prod_{j=1}^{r}(1+\tau_j)^{-\frac{n_p}{2}}\int_{\mathcal{O}(p)}e^{p\operatorname{tr}(D_pQ'\Lambda_pQ)}(\mathrm{d}Q) \quad \textit{and} \tag{2.4}$$

$$L_p\left(\tau;\mu_p\right)=\prod_{j=1}^{r}(1+\tau_j)^{-\frac{n_p}{2}}\frac{\left(\frac{n_p}{2}\right)^{\frac{n_p p}{2}}}{\Gamma\left(\frac{n_p p}{2}\right)}\int_0^\infty x^{\frac{n_p p}{2}-1}e^{-\frac{n_p}{2}x}\int_{\mathcal{O}(p)}e^{p\frac{x}{S_p}\operatorname{tr}(D_pQ'\Lambda_pQ)}(\mathrm{d}Q)\mathrm{d}x. \tag{2.5}$$

Note that this proposition about Gaussian likelihood ratios is of a purely analytical nature, and does not require any distributional assumptions.

In the single-spiked case $(r=1)$, the rank of the matrix D_p is no larger than one, and the integrals over the orthogonal group in (2.4) and (2.5) can be rewritten as integrals over a p-dimensional sphere. OMH show how such spherical integrals can be represented as contour integrals, and apply Laplace approximation to these contour integrals to establish the asymptotic properties of $L_p\left(\tau;\lambda_p\right)$ and $L_p\left(\tau;\mu_p\right)$. In the multispiked case ($r>1$), the integrals in (2.4) and (2.5) can be rewritten as integrals over a Stiefel manifold, the set of all orthonormal r-*frames* in $\mathbb{R}^p$. Onatski (2012) obtains a generalization of the contour integral representation from spherical integrals to integrals over Stiefel manifolds. Unfortunately, the Laplace approximation method does not straightforwardly extend to that generalization, and we therefore propose an alternative method of analysis.

The second-order asymptotic behavior of integrals of the form $\int_{\mathcal{O}(p)} e^{p\operatorname{tr}(DQ'\Lambda Q)}\left(\mathrm{d}Q\right)$ as p goes to infinity is analyzed in GM (Theorem 3) for the particular case where D is a fixed matrix of rank one, Λ a deterministic matrix, and under the condition that the empirical distribution of Λ's eigenvalues converges to a distribution function with bounded support. Below, we extend GM's approach to cases where $D=D_p$ has rank larger than one, and to the stochastic setting of this paper. We then use such an extension to derive the asymptotic properties of $L_p\left(\tau;\lambda_p\right)$ and $L_p\left(\tau;\mu_p\right)$.

Let $\hat{F}_p^{\lambda}$ be the empirical distribution of $\lambda_{p1},\ldots,\lambda_{pp}$, and denote by F_p^{MP} the Marchenko-Pastur distribution function with density

$$f_p^{MP}\left(x\right)=\frac{1}{2\pi c_p x}\sqrt{\left(b_p-x\right)\left(x-a_p\right)}\mathbf{1}\left\{a_p\leq x\leq b_p\right\}, \tag{2.6}$$

where $a_p=\left(1-\sqrt{c_p}\right)^2$ and $b_p=\left(1+\sqrt{c_p}\right)^2$, and a mass of $\max\left(0,1-c_p^{-1}\right)$ at zero. Here and throughout the paper $\mathbf{1}\left\{\cdot\right\}$ denotes the indicator function. As follows from Theorem 1.1 of Silverstein and Bai (1995), if the entries of ε in (2.1) are (not necessarily Gaussian) i.i.d. with zero mean and variance one, then the difference between $\hat{F}_p^{\lambda}$ and F_p^{MP} weakly converges to zero a.s. as $p,n_p\to_c\infty$, irrespective of the true value of h. If, in addition, the entries of ε have finite moment of order four, then

$$\lambda_{p1}\overset{a.s}{\to}\left(1+\sqrt{c}\right)^2 \text{ and } \lambda_{pp}\overset{a.s}{\to}\left(1-\sqrt{c}\right)^2\mathbf{1}\left\{c<1\right\}$$

for any $h\in\left((1-\sqrt{c})\mathbf{1}\left\{c<1\right\}-1,\sqrt{c}\right)^r$ (see Baik and Silverstein (2006)).

Consider the Hilbert transform $H_p^{MP}(x)=\int\left(x-\lambda\right)^{-1}\mathrm{d}F_p^{MP}\left(\lambda\right)$ of F_p^{MP}. That transform is well defined for real x outside the support of F_p^{MP}, that is, on the set $\mathbb{R}\backslash\operatorname{supp}\left(F_p^{MP}\right)$. Using (2.6), we get

$$H_p^{MP}\left(x\right)=\frac{x+c_p-1-\sqrt{\left(x-c_p-1\right)^2-4c_p}}{2c_p x}, \tag{2.7}$$

where the sign of the square root is chosen to be the sign of $(x - c_p - 1)$. It is not hard to see that $H_p^{MP}(x)$ is strictly decreasing on $\mathbb{R}\backslash \operatorname{supp}\left(F_p^{MP}\right)$. Thus, on $H_p^{MP}\left(\mathbb{R}\backslash \operatorname{supp}\left(F_p^{MP}\right)\right)$, we can define an inverse function K_p^{MP}, with values

$$K_p^{MP}(x) = 1/x + 1/(1 - c_p x). \tag{2.8}$$

The so-called R-transform R_p^{MP} of F_p^{MP} takes the form

$$R_p^{MP}(x) = K_p^{MP}(x) - 1/x = 1/(1 - c_p x).$$

For $\omega > 0$ and $\eta > 0$ sufficiently small, consider the subset of $\mathbb{R}$

$$\Omega_{\omega\eta} = \begin{cases} \left[-\eta^{-1}, 0\right) \cup \left(0, \frac{1}{\sqrt{c}(1+\sqrt{c})} - \omega\right] & \text{for } c \geq 1, \\ \left[-\frac{1}{\sqrt{c}(1-\sqrt{c})} + \omega, 0\right) \cup \left(0, \frac{1}{\sqrt{c}(1+\sqrt{c})} - \omega\right] & \text{for } c < 1. \end{cases}$$

It is straightforward to verify that $\Omega_{\omega\eta} \subset H_p^{MP}\left(\mathbb{R}\backslash \operatorname{supp}\left(F_p^{MP}\right)\right)$ for sufficiently large p as $n_p, p \to_c \infty$, and hence, $K_p^{MP}(x)$ and $R_p^{MP}(x)$ are well defined for $x \in \Omega_{\omega\eta}$.

In what follows, we shall consider possibly non-Gaussian ε's in (2.1). More specifically, we refer to the following distributional assumptions.

Assumption nG: *ε has i.i.d. entries ε_{ij} with $E\varepsilon_{ij} = 0$, $E\varepsilon_{ij}^2 = 1$, and $E\varepsilon_{ij}^4 < \infty$.*

Assumption nG*: *ε has i.i.d. entries ε_{ij} with $E\varepsilon_{ij} = 0$, $E\varepsilon_{ij}^2 = 1$, and $E\varepsilon_{ij}^4 = 3$.*

Clearly, Assumption G implies Assumption nG*, which in turn implies Assumption nG. The following result holds under Assumption nG.

Proposition 2. *Let $\{\Theta_p\}$ be a sequence of deterministic $p \times p$ diagonal matrices $\operatorname{diag}(\theta_{p1}, ..., \theta_{pr}, 0, ..., 0)$ such that, for some $\omega > 0$ and $\eta > 0$, $2\theta_{pj} \in \Omega_{\omega\eta}$ for all $j = 1, ..., r$ and sufficiently large p as $n_p, p \to_c \infty$. Let $v_{pj} = R_p^{MP}(2\theta_{pj})$. Then, for any $h \in ((1 - \sqrt{c})\mathbf{1}\{c < 1\} - 1, \sqrt{c})^r$, under Assumption nG, uniformly over all sequences $\{\Theta_p\}$ satisfying the above requirements, letting $\Lambda_p = \operatorname{diag}(\lambda_{p1}, ..., \lambda_{pp})$,*

$$\begin{aligned} \int_{\mathcal{O}(p)} e^{p \operatorname{tr}(\Theta_p Q' \Lambda_p Q)} (\mathrm{d}Q) &= e^{p \sum_{j=1}^r \left[\theta_{pj} v_{pj} - \frac{1}{2p} \sum_{i=1}^p \ln(1 + 2\theta_{pj} v_{pj} - 2\theta_{pj} \lambda_{pi})\right]} \\ &\quad \times \prod_{j=1}^r \prod_{s=1}^j \sqrt{1 - 4(\theta_{pj} v_{pj})(\theta_{ps} v_{ps}) c_p}\,(1 + o_{\mathrm{P}}(1)). \end{aligned} \tag{2.9}$$

This proposition extends Theorem 3 of GM to cases when $\operatorname{rank}(\Theta_p) > 1$, the θ_{pj}'s depend on p, and Λ_p is random. When $r = 1$ and $2\theta_{p1} = 2\theta \in \Omega_{\omega\eta}$, it is straightforward to verify that

$$\sqrt{1 - 4\theta_{p1}^2 v_{p1}^2 c_p} = \sqrt{4\theta^2}/\sqrt{Z}, \text{ where } Z = \int \left(K_p^{MP}(2\theta) - \lambda\right)^{-2} \mathrm{d}F_p^{MP}(\lambda).$$

In GM's Theorem 3, the expression $\sqrt{4\theta^2}/\sqrt{Z}$ should be used instead of $\sqrt{Z - 4\theta^2}/\theta\sqrt{Z}$, which is a typographical error.

Setting $r = 1$ and $\theta_{p1} = \tau/(2c_p(1+\tau))$ in Proposition 2, and using formula (2.4) from Proposition 1 yields an expression for $L_p(\tau; \lambda_p)$ which is equivalent to formula (4.1) in Theorem 7 of OMH. Theorem 3 below uses Proposition 2 to generalize Theorem 7 of OMH to the multispiked case ($r > 1$).

Let $\theta_{pj} = \tau_j/(2c_p(1+\tau_j))$ and

$$H_\delta = \begin{cases} [-1+\delta, 0) \cup (0, \sqrt{c} - \delta] & \text{for } c > 1, \\ [-\sqrt{c}+\delta, 0) \cup (0, \sqrt{c} - \delta] & \text{for } c \le 1. \end{cases} \tag{2.10}$$

The condition $\tau_j \in H_\delta$ for some $\delta > 0$ implies that $2\theta_{pj} \in \Omega_{\omega\eta}$ for some $\omega > 0$, $\eta > 0$ and p sufficiently large. Below, we are only interested in non-negative values of τ_j, and assume that $\tau_j \in (0, \sqrt{c} - \delta]$. The corresponding θ_{pj}, thus, is positive.

With the above setting for θ_{pj}, we have $v_{pj} = R_p^{MP}(2\theta_{pj}) = 1 + \tau_j$ and $K_p^{MP}(2\theta_{pj}) = (c_p + \tau_j)(1+\tau_j)/\tau_j$. As in Theorem 7 of OMH, we denote the latter expression by $z_j(\tau)$. Define

$$\Delta_p(z_j(\tau)) = \sum_{i=1}^{p} \ln(z_j(\tau) - \lambda_{pi}) - p \int \ln(z_j(\tau) - \lambda) \mathrm{d}F_p^{MP}(\lambda). \tag{2.11}$$

We then have the following asymptotic representation.

THEOREM 3. *Let Assumption nG hold, and let δ be a fixed number such that $0 < \delta < \sqrt{c}$. Then, for any $h \in [0, \sqrt{c} - \delta]^r$, we have*

$$L_p(\tau; \lambda_p) = \prod_{j=1}^{r} \exp\left\{ -\frac{1}{2}\Delta_p(z_j(\tau)) + \frac{1}{2}\sum_{s=1}^{j} \ln\left(1 - \frac{\tau_j \tau_s}{c_p}\right)\right\}(1 + o_{\mathrm{P}}(1)) \tag{2.12}$$

and

$$L_p\left(\tau; \mu_p\right) = L_p(\tau; \lambda_p) \exp\left\{ \frac{1}{4c_p}\left(\sum_{j=1}^{r}\tau_j\right)^2 - \frac{S_p - p}{2c_p}\sum_{j=1}^{r}\tau_j \right\}(1 + o_{\mathrm{P}}(1)), \tag{2.13}$$

where $o_{\mathrm{P}}(1) \to 0$ in probability, uniformly in $\tau \in [0, \sqrt{c} - \delta]^r$ as $n_p, p \to_c \infty$.

2.2. *Weak convergence of the log-likelihood process.* Theorem 3 approximates the pseudo-likelihood ratios by functions of the *linear spectral statistics* $\Delta_p(z_j(\tau))$, $j = 1, ..., r$, and S_p. Such an approximation allows us to use Bai and Silverstein's (2004) Central Limit Theorem (CLT) to study the asymptotics of $L_p(\tau; \lambda_p)$ and $L_p\left(\tau; \mu_p\right)$ both under the null H_0 and under point alternatives associated with $h \in (0, \sqrt{c} - \delta]^r$. Let $C[0, \sqrt{c} - \delta]^r$, where $\delta \in (0, \sqrt{c})$, denote the space of real-valued continuous functions on $[0, \sqrt{c} - \delta]^r$ equipped with the supremum norm. The Bai–Silverstein CLT requires that the elements of ε have zero excess kurtosis. Hence we replace Assumption nG by Assumption nG* in the next proposition.

PROPOSITION 4. *Suppose that Assumption nG* holds. Then, under $h = 0$, $\ln L_p(\tau; \lambda_p)$ and $\ln L_p\left(\tau; \mu_p\right)$, viewed as random elements of $C[0, \sqrt{c} - \delta]^r$, converge weakly to $\mathcal{L}_\lambda(\tau)$ and $\mathcal{L}_\mu(\tau)$ with Gaussian finite-dimensional distributions such that, for any $\tau, \tilde{\tau} \in [0, \sqrt{c} - \delta]^r$,*

$$\mathrm{E}\left(\mathcal{L}_\lambda(\tau)\right) = -\frac{1}{2}\mathrm{Var}\left(\mathcal{L}_\lambda(\tau)\right), \quad \mathrm{E}\left(\mathcal{L}_\mu(\tau)\right) = -\frac{1}{2}\mathrm{Var}\left(\mathcal{L}_\mu(\tau)\right),$$

$$\mathrm{Cov}\left(\mathcal{L}_\lambda(\tau), \mathcal{L}_\lambda(\tilde{\tau})\right) = -\frac{1}{2}\sum_{i,j=1}^{r}\ln\left(1 - \frac{\tau_i\tilde{\tau}_j}{c}\right), \text{ and} \tag{2.14}$$

$$\mathrm{Cov}\left(\mathcal{L}_\mu(\tau), \mathcal{L}_\mu(\tilde{\tau})\right) = -\frac{1}{2}\sum_{i,j=1}^{r}\left(\ln\left(1 - \frac{\tau_i\tilde{\tau}_j}{c}\right) + \frac{\tau_i\tilde{\tau}_j}{c}\right). \tag{2.15}$$

Under Assumption G (Gaussian ε's), $L_p(\tau; \lambda_p)$ and $L_p(\tau; \mu_p)$ are the actual likelihood (as opposed to pseudo-likelihood) ratios; Proposition 4 and Le Cam's first lemma (van der Vaart (1998), p.88) then imply that the joint distributions of $\lambda_{p1}, ..., \lambda_{pm}$ (as well as those of $\mu_{p1}, ..., \mu_{p,m-1}$) under the null and under any alternative of the form $h = \tau \in (0, \sqrt{c})^r$ are *mutually contiguous*. By applying Le Cam's third lemma (van der Vaart (1998), p.90), we can study the asymptotic powers of tests detecting signals in noise.

3. Asymptotic power analysis.

3.1. *Gaussian power envelope.* Denote by $LR_{\lambda,\tau}$ and $LR_{\mu,\tau}$, respectively, the most powerful, under Assumption G, α-level λ- and μ-based tests of $H_0 : h = 0$ against the point alternative $h = \tau$, where $\tau \in [0, \sqrt{c} - \delta]^r$. Formally, each test is a statistic ϕ with values in $[0, 1]$; it follows from the Neyman-Pearson lemma that

$$\begin{aligned} LR_{\lambda,\tau}(\lambda_p) &= \mathbf{1}\{\ln L_p(\tau; \lambda_p) > c_{\lambda,\tau}\} \text{ and} \\ LR_{\mu,\tau}(\mu_p) &= \mathbf{1}\{\ln L_p(\tau; \mu_p) > c_{\mu,\tau}\}, \end{aligned}$$

where $c_{\lambda,\tau}$ and $c_{\mu,\tau}$ are the $1-\alpha$ quantiles of the null distributions of the log-likelihood ratios $\ln L_p(\tau;\lambda_p)$ and $\ln L_p(\tau;\mu_p)$, respectively. Let

$$\beta_\lambda(\tau) = \lim_{n_p,p\to_c\infty} \mathrm{E}_\tau[LR_{\lambda,\tau}(\lambda_p)] \text{ and } \beta_\mu(\tau) = \lim_{n_p,p\to_c\infty} \mathrm{E}_\tau[LR_{\mu,\tau}(\mu_p)],$$

where expectations are taken under Assumption G and the alternative $h=\tau$. The functions $\tau \mapsto \beta_\lambda(\tau)$ and $\tau \mapsto \beta_\mu(\tau)$ are called the (Gaussian) *asymptotic power envelopes* at level α. Clearly, $\beta_\lambda(\tau)$ and $\beta_\mu(\tau)$ are upper bounds, under Assumption G, for the asymptotic power at $h=\tau$ of any λ- or μ-based test of H_0.

PROPOSITION 5. *Denoting by Φ the standard normal distribution function,*

$$\beta_\lambda(\tau) = 1-\Phi\left[\Phi^{-1}(1-\alpha) - \sqrt{-\frac{1}{2}\sum_{i,j=1}^{r}\ln\left(1-\frac{\tau_i\tau_j}{c}\right)}\right] \tag{3.1}$$

and

$$\beta_\mu(\tau) = 1-\Phi\left[\Phi^{-1}(1-\alpha) - \sqrt{-\frac{1}{2}\sum_{i,j=1}^{r}\left(\ln\left(1-\frac{\tau_i\tau_j}{c}\right)+\frac{\tau_i\tau_j}{c}\right)}\right]. \tag{3.2}$$

Figure 1 shows the Gaussian asymptotic power envelopes β_λ and β_μ as functions of $\tau_1/\sqrt{c}$ and $\tau_2/\sqrt{c}$ in the bivariate case $\tau=(\tau_1,\tau_2)$.

It is important to realize that the asymptotic power envelopes derived in Proposition 5 are valid—that is, provide valid upper bounds for asymptotic powers—not only for λ- and μ-based tests but also for any test invariant under left orthogonal transformations of the observations ($X \mapsto QX$, where Q is a $p\times p$ orthogonal matrix), and for any test invariant under multiplication by any non-zero constant and left orthogonal transformations of the observations ($X \mapsto aQX$, where $a \in \mathbb{R}_0^+$ and Q is a $p\times p$ orthogonal matrix), respectively. Let $\|A\|_F = (\mathrm{tr}\,(A'A))^{1/2}$ and $\|A\|_2 = \lambda_1^{1/2}(A'A)$ denote the Frobenius norm and the spectral norm, respectively, of a matrix A. Let H_0 be the null hypothesis $h=0$, and let H_1 be any of the following alternatives: $H_1 : h \in (\mathbb{R}^+)^r \setminus \{0\}$, or $H_1 : \Sigma \neq \sigma^2 I_p$, or $H_1 : \{\Sigma : \|\Sigma-\sigma^2 I_p\|_F > \omega_{n,p}\}$, or $H_1 : \{\Sigma : \|\Sigma-\sigma^2 I_p\|_2 > \omega_{n,p}\}$, where $\omega_{n,p}$ is a positive constant that may depend on n and p.

PROPOSITION 6. *For specified σ^2, consider tests of H_0 against H_1 that are invariant with respect to left orthogonal transformations of the data $X=[X_1,...,X_n]$. For any such test, there exists a λ-measurable test with the same* *size and power function. Similarly, for unspecified σ^2, consider tests that, in addition, are invariant with respect to multiplication of the data X by non-zero constants. For any such test, there exists a μ-measurable test with the same size and power function.*

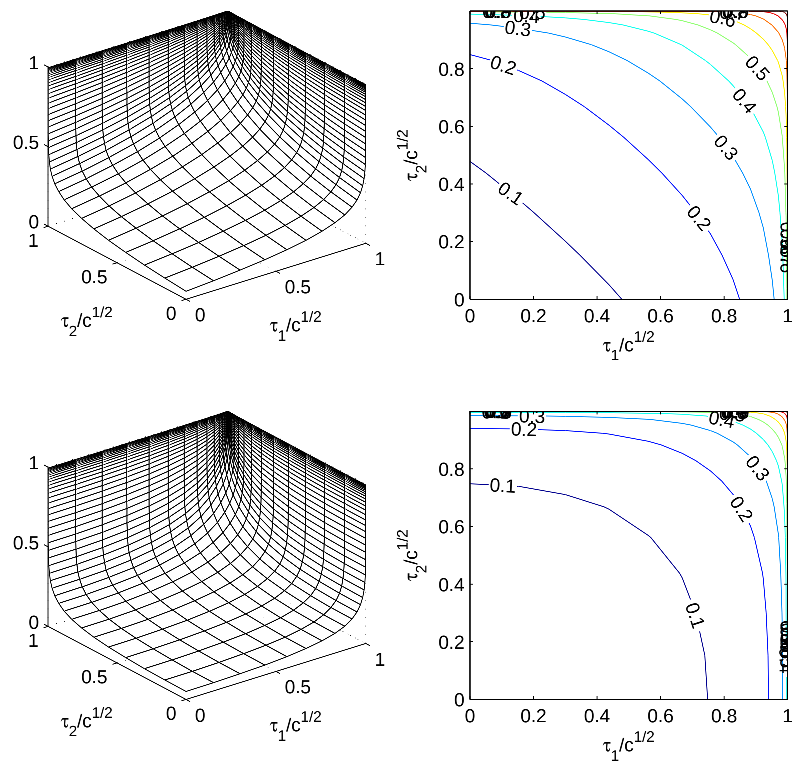

FIG 1. *The Gaussian power envelopes $\beta_\lambda(\tau)$ (upper panel) and $\beta_\mu(\tau)$ (lower panel) for $\alpha = 0.05$, as functions of $\tau/\sqrt{c} = (\tau_1, \tau_2)/\sqrt{c}$.*

Examples of tests that are invariant in the sense of Proposition 6 without being λ- or μ-measurable are the tests proposed (for specified and/or unspecified σ^2) by Chen et al (2010) and Cai and Ma (2012). It follows from Proposition 6 that their asymptotic powers, under Assumption G, are uniformly bounded by the power envelopes β_λ (specified σ^2) or β_μ (unspecified σ^2).

3.2. *Likelihood ratio tests.* We now consider λ- and μ-based α-level Gaussian *likelihood ratio* (LR) *tests* for $H_0 : h = 0$ against alternatives of the form $H_1 : h \in \Upsilon$, where $\Upsilon \subseteq (\mathbb{R}^+)^r \setminus \{0\}$. Those tests are defined as

$$\begin{aligned} LR_{\lambda,\Upsilon}(\lambda_p) &= \mathbf{1}\Big\{ \sup{}_{t\in\Upsilon} \ln L_p \Big(t;\lambda_p\Big) > c_{\lambda,\Upsilon} \Big\} \text{ and} \\ LR_{\mu,\Upsilon}(\mu_p) &= \mathbf{1}\Big\{ \sup{}_{t\in\Upsilon} \ln L_p \Big(t;\mu_p\Big) > c_{\mu,\Upsilon} \Big\}, \end{aligned}$$

where $c_{\lambda,\Upsilon}$ and $c_{\mu,\Upsilon}$ are the $(1-\alpha)$ quantiles of the (exact or asymptotic) null distributions of $\sup_{t\in\Upsilon} \ln L_p(t;\lambda_p)$ and $\sup_{t\in\Upsilon} \ln L_p \Big(t;\mu_p\Big)$, respectively. In case ε is not Gaussian, $LR_{\lambda,\Upsilon}$ and $LR_{\mu,\Upsilon}$ are to be interpreted as *pseudo-Gaussian likelihood ratio tests*.

PROPOSITION 7. *Let $\Upsilon = (0,\bar{\tau}]^r$, where $0 < \bar{\tau} < \sqrt{c}$. Then, under Assumption nG*,*

(i) the asymptotic sizes (as $n_p, p \to_c \infty$) of $LR_{\lambda,\Upsilon}$ and $LR_{\mu,\Upsilon}$ are α;
(ii) the asymptotic powers (as $n_p, p \to_c \infty$) of $LR_{\lambda,\Upsilon}$ and $LR_{\mu,\Upsilon}$ at $h = \tau \in [0,\bar{\tau}]^r$ are

$$\mathrm{P}\Big[\sup_{t\in\Upsilon}\{\mathcal{L}_\lambda(t) + \mathrm{Cov}(\mathcal{L}_\lambda(t), \mathcal{L}_\lambda(\tau))\} > c_{\lambda,\Upsilon}\Big] \tag{3.3}$$

and

$$\mathrm{P}\Big[\sup_{t\in\Upsilon}\{\mathcal{L}_\mu(t) + \mathrm{Cov}(\mathcal{L}_\mu(t), \mathcal{L}_\mu(\tau))\} > c_{\mu,\Upsilon}\Big], \tag{3.4}$$

respectively.

Note that Proposition 7 does not require ε to be Gaussian: (pseudo)-Gaussian LR tests are asymptotically valid, and their asymptotic powers remain the same, under any ε with zero excess kurtosis.

The asymptotic powers (3.3) and (3.4) depend on the distribution of the supremum over $t \in (0,\bar{\tau}]^r$ of a Gaussian process indexed by t. In principle, the distribution function of such suprema can be represented in the form of converging Rice series (related with the factorial moments of the number of upcrossings of the Gaussian process with a particular level); see Theorem 2.1 of Azais and Wschebor (2002). This may lead to analytic expressions, for the asymptotic powers of our tests. These expressions, however, still would involve Rice series, which somewhat restricts their practical value, and we rather rely here on numerical evaluations. To compute the critical value corresponding to $\alpha = 0.05$, we simulate $\mathcal{L}_\lambda(t)$ on a grid over $t \in \Upsilon = (0,\bar{\tau}]^r$, and save its maximum. We choose the critical value $c_{\lambda,\Upsilon}$ as the 95% quantile over 100,000 replications. To compute the asymptotic power at $h = \tau$, we similarly simulate $\mathcal{L}_\lambda(t) + \mathrm{Cov}(\mathcal{L}_\lambda(t), \mathcal{L}_\lambda(\tau))$ and record the proportion of replications for which the maximum of the simulated process lies above $c_{\lambda,\Upsilon}$.

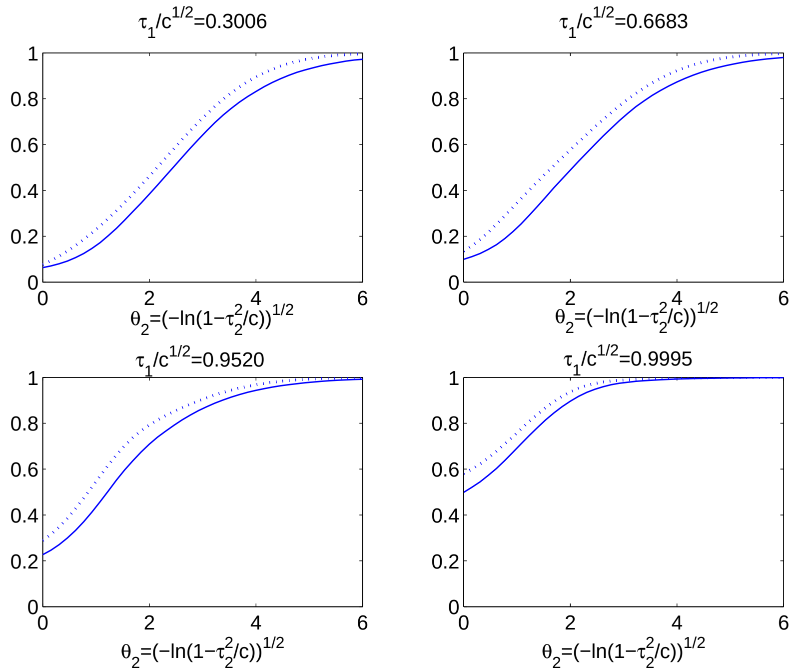

FIG 2. *Profiles of the asymptotic power (under Assumption nG*) of the* λ*-based LR test (solid lines) relative to the Gaussian asymptotic power envelope (dotted lines) for several values of* τ_1 *under the alternative* $h = \tau$*;* $\alpha = 0.05$.

The asymptotic power of the μ-based LR test is computed similarly. Unfortunately, implementing this procedure becomes increasingly cumbersome as r grows, as we need to simulate an r-dimensional Gaussian random field.

For $r = 2$, Figure 2 shows profiles of the asymptotic power of $LR_{\lambda,\Upsilon}$ (with $\Upsilon = (0, \bar{\tau}]^2$) corresponding to alternatives $(h_1, h_2) = (\tau_1, \tau_2) \in (0, \bar{\tau}]^2$ with four different values of τ_1. We set $\bar{\tau}$ to $\sqrt{c\,(1 - e^{-36})}$, which is very close to the boundary $\sqrt{c}$ of the contiguity region $[0, \sqrt{c})$. Following OMH, and in order to enhance readability of the figures, we use a different parametrization $\tau_j \mapsto \theta_j = [-\ln(1 - \tau_j^2/c)]^{1/2}$ of the values of h_j under various point alternatives. The asymptotic power profiles are superimposed with those of the Gaussian asymptotic power envelope (dotted lines). We see that the asymptotic power of $LR_{\lambda,\Upsilon}$ is close to the envelope. Figure 3 shows the same plots for the $LR_{\mu,\Upsilon}$ test.

Figure 4 further explores the relationship between the asymptotic powers of the λ- and μ-based LR tests and the corresponding Gaussian asymptotic power envelopes when $r = 2$. Select all alternatives $(h_1, h_2) = (\tau_1, \tau_2)$ with $\tau_1 \geq \tau_2$ such that the Gaussian asymptotic power envelope for λ-based tests is exactly 25, 50, 75, and 90%. We compute and plot the corresponding

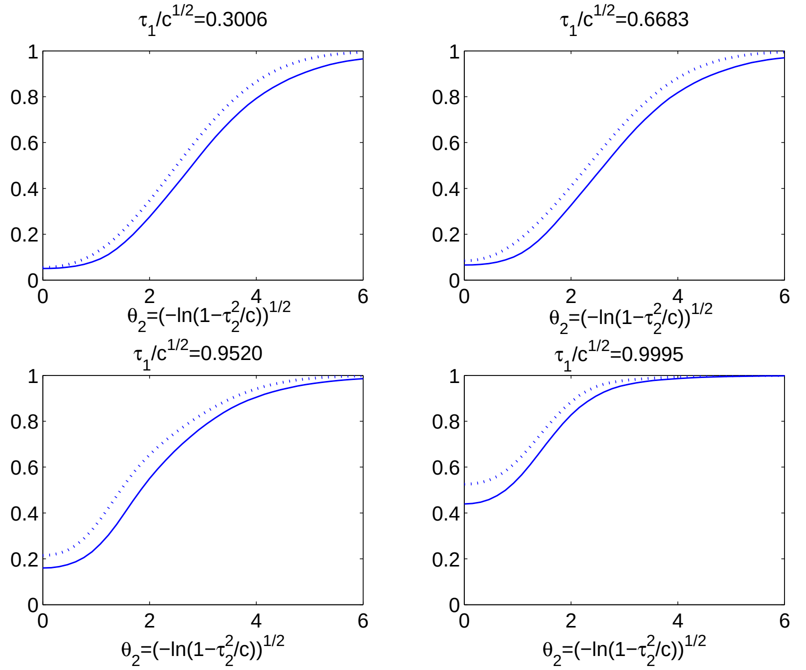

FIG 3. *Profiles of the asymptotic power (under Assumption nG*) of the μ-based LR test (solid lines) relative to the Gaussian asymptotic power envelope (dotted lines) for different values of τ_1 under the alternative $h = \tau$; $\alpha = 0.05$.*

power of $LR_{\lambda,\Upsilon}$ (solid lines) as a function of the ratio τ_2/τ_1. The dashed lines show similar graphs for $LR_{\mu,\Upsilon}$. The value $\tau_2/\tau_1 = 0$ corresponds to single-spiked alternatives $(h_1, h_2) = (\tau_1, 0)$ with $\tau_1 > 0$, the value $\tau_2/\tau_1 = 1$ to equispiked alternatives $(h_1, h_2) = (\tau, \tau)$ with $\tau > 0$. The intermediate values of τ_2/τ_1 link the two extreme cases. We do not consider values of τ_2/τ_1 larger than one, as the power function is symmetric about the 45-degree line in the (τ_1, τ_2) space.

Somewhat surprisingly, the asymptotic power of the LR test along the set of alternatives $(h_1, h_2) = (\tau_1, \tau_2)$ corresponding to the same values of the Gaussian asymptotic power envelope is not a monotone function of τ_2/τ_1. Equispiked alternatives typically seem harder to detect by the LR tests. However, for the set of alternatives corresponding to a Gaussian asymptotic power envelope value of 90%, single-spiked alternatives are even harder.

A natural question is: how do the asymptotic powers of the LR tests depend on the choice of r, that is, how do those tests perform when the actual r does not coincide with the value the test statistic is based on? For example, a natural way to proceed in signal detection practice is to start with a LR test of the null hypothesis against single-spiked alternatives ($r = 1$). If the null is rejected, one then moves to $r = 2$, $r = 3$ etc. How do the asymptotic powers of such tests compare? Figure 5 reports the asymptotic powers of the λ- and μ-based LR tests designed against single- and double-spiked ($r = 1$, dashed line; $r = 2$, solid line) alternatives computed at equispiked alternatives of the form $(h_1, h_2) = (\tau, \tau)$. As in Figures 2 and 3, we use the parametrization $\theta = [-\ln(1 - \tau^2/c)]^{1/2}$. The asymptotic power of the test incorrectly specifying $r = 1$ is slightly smaller than that of the test with correct specification $r = 2$ for most values of θ (and τ).

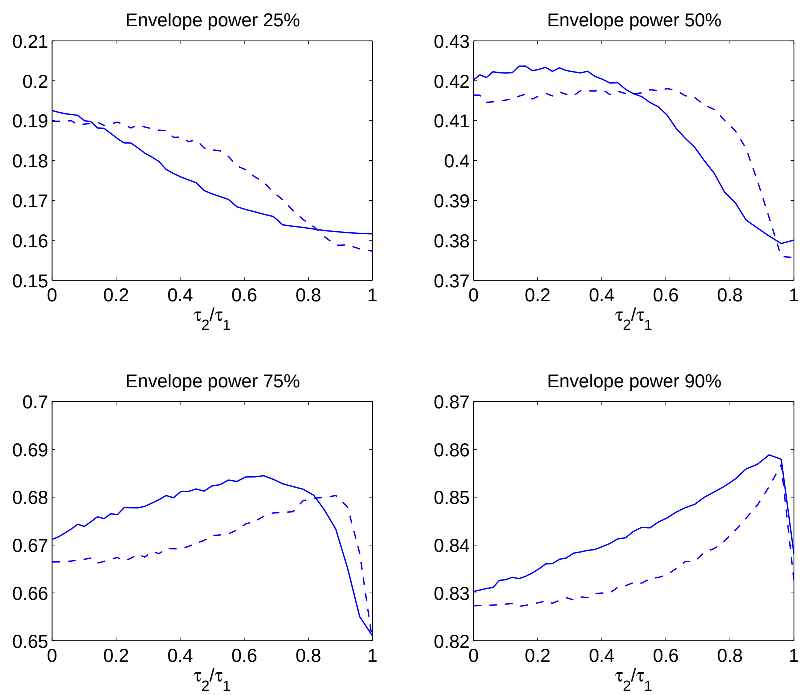

Fig 4. *Power of λ-based (solid lines) and μ-based (dashed lines) LR tests plotted against the ratio τ_2/τ_1, where (τ_1, τ_2) are such that the respective asymptotic power envelopes $\beta_\lambda(\tau)$ and $\beta_\mu(\tau)$ equal 25, 50, 75 and 90%.*

3.3. *Asymptotic power of related tests.* The same results on the likelihood process as above allow for computing the asymptotic powers of several tests available in the literature.

Example 1 (John's (1971) test of sphericity – $H_0 : \Sigma = \sigma^2 I_p$). *John (1971)*

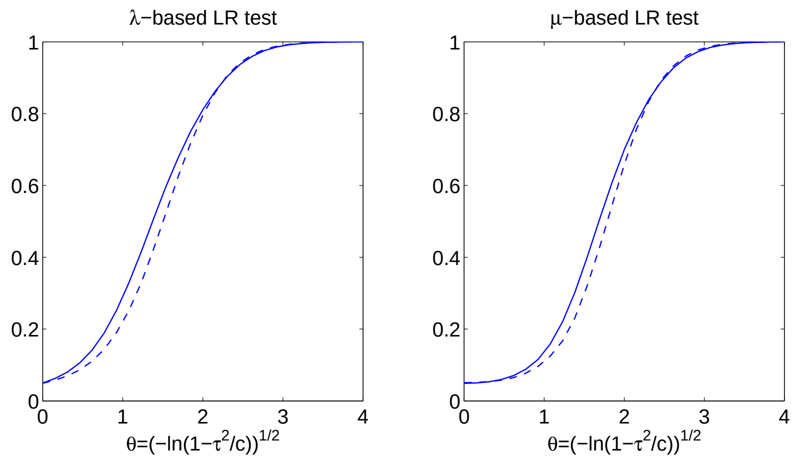

FIG 5. *Asymptotic power of the λ-based (left panel) and μ-based (right panel) LR tests. Solid line: power against equispiked alternatives $(h_1, h_2) = (\tau, \tau)$ when $r = 2$ is correctly assumed. Dashed line: power when $r = 1$ is incorrectly assumed.*

proposes testing sphericity against general alternatives via the test statistic

$$U = \frac{1}{p} \operatorname{tr} \left[\left(\frac{\hat{\Sigma}}{(1/p) \operatorname{tr} \left(\hat{\Sigma} \right)} - I_p \right)^2 \right], \tag{3.5}$$

where $\hat{\Sigma}$ is the sample covariance matrix. He shows that, when $n > p$, such a test is locally most powerful invariant. Ledoit and Wolf (2002) study John's test when $n_p, p \to_c \infty$. They prove that, for Gaussian data, under the null, $nU - p \xrightarrow{d} \mathcal{N}(1,4)$. Hence, the test with asymptotic size α (as $n_p, p \to_c \infty$) rejects the null hypothesis of sphericity whenever $\frac{1}{2}(nU - p - 1) > \Phi^{-1}(1-\alpha)$.

EXAMPLE 2 (The Ledoit-Wolf (2002) test – $H_0 : \Sigma = I_p$). *Ledoit and Wolf (2002) propose the test statistic*

$$W = \frac{1}{p} \operatorname{tr} \left[\left(\hat{\Sigma} - I_p \right)^2 \right] - \frac{p}{n} \left[\frac{1}{p} tr \hat{\Sigma} \right]^2 + \frac{p}{n} \tag{3.6}$$

They show that, for Gaussian data, under the null, $nW - p \xrightarrow{d} \mathcal{N}(1,4)$ as $n_p, p \to_c \infty$. As in the previous example, H_0 is rejected at asymptotic size α whenever $\frac{1}{2}(nW - p - 1) > \Phi^{-1}(1-\alpha)$.

EXAMPLE 3 (The Bai et al. (2009) "corrected" LRT – $H_0 : \Sigma = I_p$).

When $n > p$, Bai et al. (2009) propose to use a corrected version

$$CLR = \operatorname{tr}\hat{\Sigma} - \ln\det\hat{\Sigma} - p - p\left(1 - \left(1 - \frac{n}{p}\right)\ln\left(1 - \frac{p}{n}\right)\right)$$

of the likelihood ratio statistic to test $H_0 : \Sigma = I_p$ against general alternatives. Under the null, when the data have zero excess kurtosis (Assumption nG), $CLR \xrightarrow{d} \mathcal{N}\left(-\frac{1}{2}\ln(1-c), -2\ln(1-c)-2c\right)$ as $n_p, p \rightarrow_c \infty$. The null hypothesis is rejected at asymptotic level α whenever $CLR + \frac{1}{2}\ln(1-c)$ is larger than $(-2\ln(1-c) - 2c)^{1/2}\,\Phi^{-1}(1-\alpha)$.*

Example 4 (The Cai–Ma (2012) minimax test – $H_0 : \Sigma = I_p$). *Cai and Ma (2012) propose the U-statistic*

$$T_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \ell(X_i, X_j),$$

where $\ell(X_1, X_2) = (X_1' X_2)^2 - (X_1' X_1 + X_2' X_2) + p$, to test the hypothesis that the population covariance matrix is the unit matrix. For Gaussian data, under the null, as $n_p, p \rightarrow_c \infty$, $T_n \xrightarrow{d} \mathcal{N}(0, 4c^2)$. The null hypothesis is rejected at asymptotic level α whenever T_n exceeds $2\sqrt{p(p+1)/n(n-1)}\,\Phi^{-1}(1-\alpha)$. Cai and Ma (2012) show that this test is rate-optimal against general alternatives from a minimax point of view.

Example 5 (Tracy–Widom-type tests – $H_0 : \Sigma = I_p$). *Let $\varphi(\lambda_1, ..., \lambda_r)$ be any function of the r largest eigenvalues increasing in all its arguments. The asymptotic distribution of $\varphi(\lambda_1, ..., \lambda_r)$ under the null, assuming that the distribution of ε_{ij} is symmetric and has sub-Gaussian moments, is determined by the functional form of $\varphi(\cdot)$ and the fact (Péché (2009)) that*

$$(\sigma_{n,c}(\lambda_1 - \nu_c), ..., \sigma_{n,c}(\lambda_r - \nu_c)) \xrightarrow{d} TW(r), \tag{3.7}$$

where $TW(r)$ denotes the r-dimensional Tracy–Widom law of the first kind, $\sigma_{n,c} = n^{2/3}c^{1/6}(1+\sqrt{c})^{-4/3}$ and $\nu_c = (1+\sqrt{c})^2$. Call Tracy–Widom-type tests all tests that reject the null whenever $\varphi(\lambda_1, ..., \lambda_r)$ is larger than the corresponding asymptotic critical value obtained from (3.7).

Consider the tests described in Examples 1, 2, 3, 4 and 5, and denote by $\beta_J(\tau), \beta_{LW}(\tau), \beta_{CLR}(\tau), \beta_{CM}(\tau)$ and $\beta_{TW}(\tau)$ their respective asymptotic powers against the point alternative $h = \tau$, at asymptotic level α.

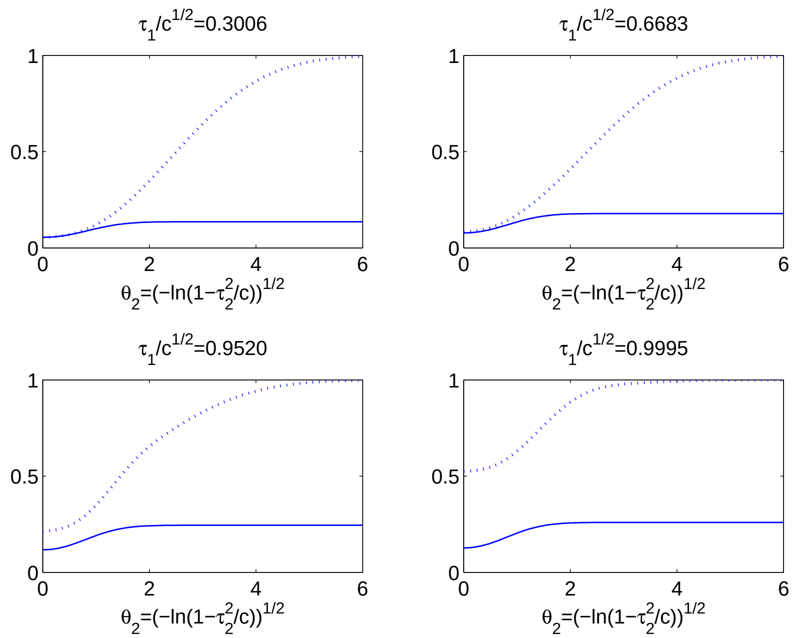

Fig 6. *Profiles of the asymptotic power of John's test (solid lines) relative to the asymptotic power envelope β_μ (dotted lines) for different values of τ_1 under the alternative $h = \tau$; $\alpha = 0.05$.*

Proposition 8. *The following statements hold for any $\tau \in [0, \sqrt{c})^r$.*

(i) Suppose that Assumption nG holds; then,*

$$\beta_J(\tau) = \beta_{LW}(\tau) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{1}{2}\sum_{j=1}^{r}\frac{\tau_j^2}{c}\right) \tag{3.8}$$

and

$$\beta_{CLR}(\tau) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \sum_{j=1}^{r}\frac{\tau_j - \ln(1+\tau_j)}{\sqrt{-2\ln(1-c) - 2c}}\right). \tag{3.9}$$

(ii) If Assumption nG is strengthened into Assumption G,*

$$\beta_{CM}(\tau) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{1}{2}\sum_{j=1}^{r}\frac{\tau_j^2}{c}\right). \tag{3.10}$$

(iii) Let ε in (2.1) be i.i.d. with symmetric, not necessarily Gaussian, distribution such that $\mathrm{E}\varepsilon_{ij}^2 = 1$, with sub-Gaussian moments—that is, such that, for some $\delta > 0$ and all positive integers k, $\mathrm{E}\varepsilon_{ij}^{2k} \leq (\delta k)^k$; then,

$$\beta_{TW}(\tau) = \alpha. \tag{3.11}$$

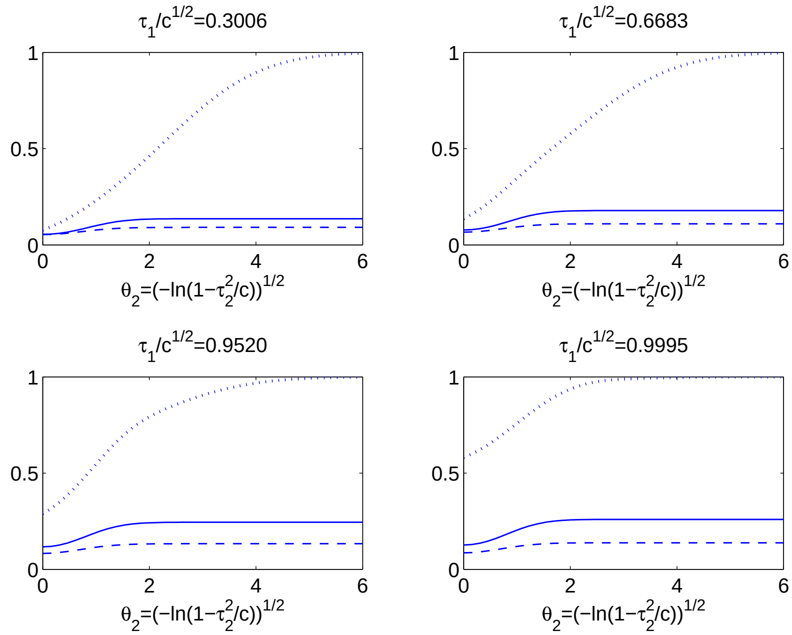

FIG 7. *Ledoit–Wolf and Cai–Ma tests (solid lines) and the CLR test (dashed lines, for* $c = 0.5$*) relative to the asymptotic power envelope* β_λ *(dotted lines) for different values of* τ_1 *under the alternative* $h = \tau$*;* $\alpha = 0.05$.

Tracy-Widom-type tests based on $\mu_1, \ldots, \mu_r$ for the hypothesis of sphericity $H_0 : \Sigma = \sigma^2 I_p$ (σ^2 unspecified) could be considered as well; mutatis mutandis, part (iii) of Proposition 8 below similarly holds for such tests. Details are skipped.

To establish (3.8) and (3.9), we use Bai and Silverstein's (2004) CLT that holds for ε with zero excess kurtosis. This explains why Assumption nG* is needed in part (i) of Proposition 8. In contrast to the John, Ledoit–Wolf, and "corrected" likelihood ratio statistics, the Cai–Ma test statistic is not asymptotically equivalent to a linear spectral statistic. Hence, in part (ii) of the proposition, we cannot use the Bai–Silverstein CLT, and make a stronger assumption of Gaussianity to obtain (3.10). The moment assumptions of part (iii) (which clearly imply Assumption nG) mimic assumptions $H_1 - H_3$ of Féral and Péché (2009).

The asymptotic power functions of the John, Ledoit–Wolf, "corrected" likelihood ratio, and Cai–Ma tests are non-trivial. Figures 6 and 7 compare these power functions to the corresponding power envelopes for $r = 2$. Since John's test is invariant with respect to orthogonal transformations and scalings of the data, Figure 6 compares $\beta_J(\tau)$ (solid line) to the Gaussian asymptotic power envelope $\beta_\mu(\tau)$ (dotted line). The Ledoit–Wolf test, the "corrected" likelihood ratio test, and the Cai-Ma test are invariant only with respect to orthogonal transformations of the data, and Figure 7 thus compares the asymptotic power functions $\beta_{LW}(\tau) = \beta_{CM}(\tau)$ and $\beta_{CLR}(\tau)$ (solid and dashed lines, respectively) to the Gaussian asymptotic power envelope $\beta_\lambda(\tau)$ (dotted line). Note that $\beta_{CLR}(\tau)$ depends on c. As c converges to one, $\beta_{CLR}(\tau)$ converges to α, which corresponds to the case of trivial power. As c converges to zero, $\beta_{CLR}(\tau)$ converges to $\beta_{LW}(\tau) = \beta_{CM}(\tau)$. In Figure 7, we provide the plots of $\beta_{CLR}(\tau)$ that correspond to $c = 0.5$.

These comparisons show that, contrary to our LR tests (see Figures 2 and 3), all those tests either have trivial power α (the Tracy–Widom ones), or power functions that increase very slowly with τ_1 and τ_2, and lie very far below the corresponding Gaussian power envelope.

4. Conclusion. This paper extends OMH's study of the power of high-dimensional sphericity tests to the case of multi-spiked alternatives. We derive the asymptotic distribution of the log-likelihood ratio process and use it to obtain simple analytical expressions for the Gaussian maximal asymptotic power envelope and for the asymptotic powers of several commonly used tests. These asymptotic powers turn out to be very substantially below the envelope. We propose the Gaussian likelihood ratio tests based on the data reduced to the eigenvalues of the sample covariance matrix. We show that those tests remain valid under mild moment assumptions. Our computations show that their asymptotic power is close to the envelope.

APPENDIX A

This appendix contains the proofs of some of the main results of this paper. A more complete version can be found in Onatski et al (2013b), where we refer to for further details. For the sake of readability and easy reference, though, we are using the same numberings here as in the complete version, which explains the gaps, for instance, between Equations (A.14) and (A.36), etc.

A.1. Proof of Proposition 2. Let $I_p(\Theta_p, \Lambda_p)$ stand for the integral $\int_{\mathcal{O}(p)} e^{p\,\mathrm{tr}(\Theta_p Q' \Lambda_p Q)}(\mathrm{d}Q)$. As explained in GM, p.454, we can write

$$I_p(\Theta_p, \Lambda_p) = \mathbb{E}_{\Lambda_p} \exp\left\{ p \sum_{j=1}^{r} \theta_{pj} \frac{\tilde{g}^{(j)\prime} \Lambda_p \tilde{g}^{(j)}}{\tilde{g}^{(j)\prime} \tilde{g}^{(j)}} \right\}, \tag{A.1}$$

where $\mathbb{E}_{\Lambda_p}$ denotes expectation conditional on Λ_p, and the p-dimensional vectors $(\tilde{g}^{(1)}, ..., \tilde{g}^{(r)})$ are obtained from standard Gaussian p-dimensional vectors $(g^{(1)},...,g^{(r)})$, independent from Λ_p, by a Schmidt orthogonalization procedure. More precisely, we have $\tilde{g}^{(j)}=\sum_{k=1}^{j}A_{jk}g^{(k)}$, where $A_{jj}=1$ and

$$\sum_{k=1}^{j-1}A_{jk}g^{(k)\prime}g^{(t)}=-g^{(j)\prime}g^{(t)}\text{ for }t=1,...,j-1. \tag{A.2}$$

In the spirit of the proof of GM's Theorem 3, define

$$\gamma_{p1}^{(j,s)}=\sqrt{p}\left(\frac{1}{p}g^{(j)\prime}g^{(s)}-\delta_{js}\right)\text{ and }\gamma_{p2}^{(j,s)}=\sqrt{p}\left(\frac{1}{p}g^{(j)\prime}\Lambda_p g^{(s)}-v_{pj}\delta_{js}\right), \tag{A.3}$$

where $\delta_{js}=\mathbf{1}\{j=s\}$ stands for the Kronecker symbol. As will be shown below, $\gamma_{p1}^{(j,s)}$ and $\gamma_{p2}^{(j,s)}$, after an appropriate change of measure, are asymptotically centered Gaussian. Expressing the exponent in (A.1) as a function of $\gamma_{p1}^{(j,s)}$ and $\gamma_{p2}^{(j,s)}$, changing the measure of integration, and using the asymptotic Gaussianity will establish the proposition.

Let $\gamma_p=\left(\gamma_p^{(1,1)},...,\gamma_p^{(r,1)},\gamma_p^{(2,2)},...,\gamma_p^{(r,2)},...,\gamma_p^{(r,r)}\right)'$, where $\gamma_p^{(j,s)}=\left(\gamma_{p1}^{(j,s)},\gamma_{p2}^{(j,s)}\right)$. With this notation, using (A.1), (A.2), and (A.3), we obtain

(A.4)

$$I_p(\Theta_p,\Lambda_p)=\int f_{p,\theta}\left(\gamma_p\right)e^{\sqrt{p}\sum_{j=1}^{r}\theta_{pj}\left(\sqrt{p}v_{pj}+\gamma_{p2}^{(j,j)}-v_{pj}\gamma_{p1}^{(j,j)}\right)}\prod_{j=1}^{r}\prod_{i=1}^{p}\mathrm{d}\mathbb{P}\left(g_i^{(j)}\right),$$

where $\mathbb{P}$ is the standard Gaussian probability measure, and

$$f_{p,\theta}\left(\gamma_p\right)=\exp\left\{\sum_{j=1}^{r}\theta_{pj}\frac{N_{1j}+...+N_{6j}}{D_j}\right\} \tag{A.5}$$

with

$$\begin{aligned}
N_{1j} &= -\gamma_{p1}^{(j,j)}\left(\gamma_{p2}^{(j,j)}-v_{pj}\gamma_{p1}^{(j,j)}\right),\\
N_{2j} &= \gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)}+I\right)^{-1}\left(G_{p2}^{(j)}+W_{pj}\right)\left(G_{p1}^{(j)}+I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},\\
N_{3j} &= -2\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)}+I\right)^{-1}\gamma_{p2}^{(j,1:j-1)},\\
N_{4j} &= v_{pj}\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)}+I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},\\
N_{5j} &= p^{-1/2}\gamma_{p2}^{(j,j)}\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)}+I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},\\
N_{6j} &= -p^{-1/2}v_{pj}\gamma_{p1}^{(1:j-1,j)\prime}\left(G_{p1}^{(j)}+I\right)^{-1}\gamma_{p1}^{(1:j-1,j)}\gamma_{p1}^{(j,j)},\text{ and}\\
D_j &= 1+p^{-1/2}\gamma_{p1}^{(j,j)}-p^{-1}\gamma_{p1}^{(j,1:j-1)\prime}\left(G_{p1}^{(j)}+I\right)^{-1}\gamma_{p1}^{(j,1:j-1)},
\end{aligned}$$

where $G_{pi}^{(j)}$ is a $(j-1)\times(j-1)$ matrix with (k,s)-th element $p^{-1/2}\gamma_{pi}^{(k,s)}$, $W_{pj}=\operatorname{diag}(v_{p1},...,v_{p,j-1})$, and $\gamma_{pi}^{(j,1:j-1)}=\left(\gamma_{pi}^{(j,1)},...,\gamma_{pi}^{(j,j-1)}\right)'$.

Next, define the event

$$B_{M,M'}=\left\{\left|\gamma_{p1}^{(j,s)}\right|\le M \text{ and } \left|\gamma_{p2}^{(j,s)}\right|\le M' \text{ for all } j,s=1,...,r\right\},$$

where M and M' are positive parameters to be specified later. With a slight abuse of notation, we shall also refer to $B_{M,M'}$ as a rectangular region in $\mathbb{R}^{r^2+r}$ that consists of vectors with odd coordinates in $(-M,M)$ and even coordinates in $(-M',M')$. Let

$$I_p^{M,M'}(\Theta_p,\Lambda_p)=\int \mathbf{1}\{B_{M,M'}\}f_{p,\theta}\left(\gamma_p\right)$$
$$\times e^{\sqrt{p}\sum_{j=1}^r\theta_{pj}\left(\sqrt{p}v_{pj}+\gamma_{p2}^{(j,j)}-v_{pj}\gamma_{p1}^{(j,j)}\right)}\prod_{j=1}^r\prod_{i=1}^p \mathrm{d}\mathbb{P}\left(g_i^{(j)}\right).$$

Below, we establish the asymptotic behavior of $I_p^{M,M'}(\Theta_p,\Lambda_p)$ as first p, and then M and M', diverge to infinity. We then show that the asymptotics of $I_p^{M,M'}(\Theta_p,\Lambda_p)$ and $I_p(\Theta_p,\Lambda_p)$ coincide.

Consider infinite arrays $\left\{\mathbb{P}_{pi}^{(j)},p=1,2,...;i=1,...,p\right\}$, $j=1,...,r$, of random centered Gaussian measures

$$\mathrm{d}\mathbb{P}_{pi}^{(j)}(x)=\sqrt{\frac{1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{pi}}{2\pi}}e^{-\frac{1}{2}(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{pi})x^2}\mathrm{d}x. \tag{A.6}$$

Since $v_{pj}=R_p^{MP}(2\theta_{pj})=1/(1-2\theta_{pj}c_p)$ and $2\theta_{pj}\in\Omega_{\omega\eta}$, there exists $\hat{\omega}>0$ such that, for sufficiently large p,

$$\begin{array}{lcll} v_{pj}+1/(2\theta_{pj}) & > & (1+\sqrt{c})^2+\hat{\omega} & \text{when } \theta_{pj}>0 \quad \text{and} \\ v_{pj}+1/(2\theta_{pj}) & < & (1-\sqrt{c})^2\,\mathbf{1}\{c<1\}-\hat{\omega} & \text{when } \theta_{pj}<0. \end{array}$$

Recall that $\lambda_{pp}\to(1-\sqrt{c})^2\,\mathbf{1}\{c<1\}$ and $\lambda_{p1}\to(1+\sqrt{c})^2$ a.s. as $n_p,p\to_c\infty$ (Baik and Silverstein (2006)). Therefore, $v_{pj}+1/(2\theta_{pj})>\lambda_{p1}$ when $\theta_{pj}>0$, and $v_{pj}+1/(2\theta_{pj})<\lambda_{pp}$ when $\theta_{pj}<0$ a.s., for sufficiently large p. Hence, the measures $\mathbb{P}_{pi}^{(j)}$ are a.s. well defined for sufficiently large p. Whenever $\mathbb{P}_{pi}^{(j)}$ is not well defined, we re-define it arbitrarily.

We have

$$I_p^{M,M'}(\Theta_p,\Lambda_p)=e^{p\sum_{j=1}^r\left[\theta_{pj}v_{pj}-\frac{1}{2p}\sum_{i=1}^p\ln(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{pi})\right]}J_p^{M,M'}, \tag{A.7}$$

where

$$J_p^{M,M'} = \int \mathbf{1}\left\{B_{M,M'}\right\} f_{p,\theta}\left(\gamma_p\right) \prod_{j=1}^{r} \prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right). \tag{A.8}$$

Section A.2 of the complete version of this Appendix contains a proof of the following lemma.

LEMMA 9. *For any $a > 0$, there exist M_a and M_a' such that, for any $M > M_a$ and $M' > M_a'$,*

$$\left| J_p^{M,M'} - \prod_{j=1}^{r} \prod_{s=1}^{j} \sqrt{1-4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p} \right| < a, \tag{A.9}$$

uniformly over $\{2\theta_{pk} \in \Omega_{\omega\eta}, k \le r\}$, with probability arbitrarily close to one, for sufficiently large p.

Lemma 9 and equation (A.7) imply that $I_p^{M,M'}(\Theta_p, \Lambda_p)$ behaves as the right-hand side of equation (2.9), as first p, and then M and M', diverge to infinity. Now, let us show that the asymptotics of $I_p^{M,M'}(\Theta_p, \Lambda_p)$ and $I_p(\Theta_p, \Lambda_p)$ coincide. Let B_M be the event $\left\{\left|\gamma_{p1}^{(j,s)}\right| \le M \text{ for all } j, s \le r\right\}$, and define

$$I_p^M(\Theta_p, \Lambda_p) = \mathbb{E}_{\Lambda_p}\left(\mathbf{1}\left\{B_M\right\} \exp\left\{p \sum_{j=1}^{r} \theta_{pj} \frac{\tilde{g}^{(j)\prime}\Lambda_p \tilde{g}^{(j)}}{\tilde{g}^{(j)\prime}\tilde{g}^{(j)}}\right\}\right).$$

The following lemma is established in Section A.3 of the complete version of this Appendix.

LEMMA 10. *Under the assumptions of Proposition 2,*

$$I_p(\Theta_p, \Lambda_p) \ge I_p^M(\Theta_p, \Lambda_p) \ge \left(1 - 2r^2 e^{-M^2/16}\right) I_p(\Theta_p, \Lambda_p) \tag{A.10}$$

for sufficiently large p, uniformly over $\{2\theta_{pk} \in \Omega_{\omega\eta}, k \le r\}$.

Similarly to $I_p^{M,M'}(\Theta_p, \Lambda_p)$, $I_p^M(\Theta_p, \Lambda_p)$ can be represented in the form

$$I_p^M(\Theta_p, \Lambda_p) = e^{p\sum_{j=1}^{r}\left[\theta_{pj}v_{pj} - \frac{1}{2p}\sum_{i=1}^{p}\ln(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{pi})\right]} J_p^M, \tag{A.11}$$

where

$$J_p^M = \int \mathbf{1}\left\{B_M\right\} f_{p,\theta}\left(\gamma_p\right) \prod_{j=1}^{r} \prod_{i=1}^{p} \mathrm{d}\mathbb{P}_{pi}^{(j)}\left(g_i^{(j)}\right).$$

The following lemma shows that the difference $J_p^{M,M',\infty} = J_p^M - J_p^{M,M'}$ is small. It is proven in Section A.4 of the complete version of this Appendix.

LEMMA 11. *Under the assumptions of Proposition 2, there exist positive constants β_0 and β_1 such that, for any positive M and M' that satisfy inequality $M'/(4\beta_0^2) > Mr^2\beta_1$,*

$$J_p^{M,M',\infty} \leq 4r^2 e^{-(M')^2/(16\beta_0^2)+\beta_1 r^2 MM'} \tag{A.12}$$

for sufficiently large p, uniformly over $\{2\theta_{pk} \in \Omega_{\omega\eta}, k \leq r\}$.

Combining (A.10), (A.11), and (A.12), we obtain

$$J_p^{M,M'} \leq J_p \leq \frac{J_p^{M,M'} + 4r^2 e^{-M'^2/16\beta_0^2+\beta_1 r^2 MM'}}{1 - 2r^2 e^{-M^2/16}}. \tag{A.13}$$

Let $\varphi > 0$ be an arbitrarily small number. Let us choose $M > M_{\varphi/4}$ and $M' > M'_{\varphi/4}$ (where M_a and M'_a are as in Lemma 9) so that

$$\left(1 - 2r^2 e^{-M^2/16}\right)^{-1} < 2,$$

$$\left(1 - 2r^2 e^{-M^2/16}\right)^{-1} 4r^2 e^{-(M')^2/(16\beta_0^2)+\beta_1 r^2 MM'} < \varphi/4,$$

and

$$\left[\left(1 - 2r^2 e^{-M^2/16}\right)^{-1} - 1\right] \times \sup_{\{2\theta_{pk}\in\Omega_{\omega\eta}, k\leq r\}} \prod_{j=1}^{r} \prod_{s=1}^{j} \sqrt{1 - 4\left(\theta_{pj} \upsilon_{pj}\right)\left(\theta_{ps} \upsilon_{ps}\right) c_p} < \varphi/4$$

for all sufficiently large p, a.s. Then, (A.13) implies that

$$\left| J_p - \prod_{j=1}^{r} \prod_{s=1}^{j} \sqrt{1 - 4\left(\theta_{pj} \upsilon_{pj}\right)\left(\theta_{ps} \upsilon_{ps}\right) c_p} \right| < \varphi \tag{A.14}$$

with probability arbitrarily close to one, for all sufficiently large p, uniformly over $\{2\theta_{pk} \in \Omega_{\omega\eta}, k \leq r\}$. Since φ can be chosen arbitrarily, we have, from (A.11) and (A.14),

$$\begin{aligned} I_p\left(\Theta_p, \Lambda_p\right) &= e^{p\sum_{j=1}^{r}\left[\theta_{pj}\upsilon_{pj} - \frac{1}{2p}\sum_{i=1}^{p}\ln(1+2\theta_{pj}\upsilon_{pj}-2\theta_{pj}\lambda_{pi})\right]} \\ &\quad \times \left(\prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1 - 4\left(\theta_{pj}\upsilon_{pj}\right)\left(\theta_{ps}\upsilon_{ps}\right)c_p} + o_{\mathrm{P}}(1)\right), \end{aligned}$$

where the $o_{\mathrm{P}}(1)$ term is uniform, as $n_p, p \to_c \infty$, in $\{2\theta_{pk} \in \Omega_{\omega\eta}, k \leq r\}$. Proposition 2 follows from this, and the fact that $1 - 4\theta_{pj}\upsilon_{pj}\theta_{ps}\upsilon_{ps}c_p$ is bounded away from zero for sufficiently large p, uniformly in $\{2\theta_{pk} \in \Omega_{\omega\eta}, k \leq r\}$. □

A.2. – A.4. Proofs of Lemmas 9, 10, and 11. See Supplementary Material (Onatski et al 2013b).

A.5. Proof of Theorem 3. First, we prove equation (2.12). For $\theta_{pj} = \frac{1}{2c_p}\frac{\tau_j}{1+\tau_j}$, we have $v_{pj} = 1+\tau_j$, $\theta_{pj}v_{pj} = \tau_j/(2c_p)$, and

$$\ln\left(1 + 2\theta_{pj}v_{pj} - 2\theta_{pj}\lambda_{pi}\right) = \ln\left(\frac{1}{c_p}\frac{\tau_j}{1+\tau_j}\right) + \ln\left(z_j\left(\tau\right) - \lambda_{pi}\right).$$

Further, by Lemma 11 and formula (3.3) of OMH,

$$\int \ln\left(z_j\left(\tau\right) - \lambda\right) \mathrm{d}F_p^{MP}\left(\lambda\right) = \frac{\tau_j}{c_p} - \frac{1}{c_p}\ln\left(1+\tau_j\right) + \ln\frac{\left(1+\tau_j\right)c_p}{\tau_j}$$

a.s. for sufficiently large p. With these auxiliary results, equation (2.12) is a straightforward consequence of Equation (2.4) and Proposition 2.

Turning to the proof of (2.13), consider the integrals

$$\mathcal{I}\left(k_1,k_2\right) = \int_{k_1}^{k_2} x^{\frac{n_p p}{2}-1} e^{-\frac{n_p}{2}x} \int_{\mathcal{O}(p)} e^{p\frac{x}{S_p}\operatorname{tr}\left(D_p Q' \Lambda_p Q\right)} \left(\mathrm{d}Q\right) \mathrm{d}x,\ \ k_1 < k_2 \in \mathbb{R}.$$

In what follows, we omit the subscript p in n_p to simplify notation. Note that $\mathcal{I}\left(0,\infty\right)$ is the integral appearing in expression (2.5) for $L_p\left(\tau;\mu_p\right)$. Section A.6 of the complete version of this Appendix contains a proof of the following lemma.

LEMMA 12. *There exists a constant $\alpha > 0$, such that*

$$\mathcal{I}\left(0,\infty\right) = \mathcal{I}\left(p-\alpha\sqrt{p}, p+\alpha\sqrt{p}\right)\left(1+o_{\mathrm{P}}\left(1\right)\right), \tag{A.36}$$

where $o_{\mathrm{P}}(1)$ is uniform in $\tau \in [0,\sqrt{c}-\delta]^r$.

Now, letting $\tilde{\theta}_{pj} = \frac{x}{S_p}\theta_{pj} = \frac{x}{S_p}\frac{1}{2c_p}\frac{\tau_j}{1+\tau_j}$, note that there exist $\omega > 0$ and $\eta > 0$ such that $\left\{2\tilde{\theta}_{pj} : \tau_j \in [0,\sqrt{c}-\delta] \text{ and } x \in \left[p-\alpha\sqrt{p}, p+\alpha\sqrt{p}\right]\right\} \subseteq \Omega_{\omega\eta}$ with probability arbitrarily close to one for sufficiently large p. Hence, by (A.36), and Proposition 2,

$$\begin{aligned}\mathcal{I}\left(0,\infty\right) &= \int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1} e^{-\frac{n}{2}x} e^{p\sum_{j=1}^r \left[\tilde{\theta}_{pj}\tilde{v}_{pj} - \frac{1}{2p}\sum_{i=1}^p \ln\left(1+2\tilde{\theta}_{pj}\tilde{v}_{pj} - 2\tilde{\theta}_{pj}\lambda_{pi}\right)\right]} \\ &\quad \times \left(\prod_{j=1}^r\prod_{s=1}^j \sqrt{1 - 4\left(\tilde{\theta}_{pj}\tilde{v}_{pj}\right)\left(\tilde{\theta}_{ps}\tilde{v}_{ps}\right)c_p} + o_{\mathrm{P}}(1)\right)\mathrm{d}x,\end{aligned} \tag{A.37}$$

where $\tilde{v}_{pj} = \left(1 - 2\tilde{\theta}_{pj} c_p\right)^{-1}$ and the $o_{\mathrm{P}}(1)$ term is uniform in $h \in [0, \sqrt{c} - \delta]^r$ and $x \in [p - \alpha\sqrt{p}, p + \alpha\sqrt{p}]$.

Expanding $\tilde{\theta}_{pj}\tilde{v}_{pj} - \frac{1}{2p}\sum_{i=1}^{p} \ln\left(1 + 2\tilde{\theta}_{pj}\tilde{v}_{pj} - 2\tilde{\theta}_{pj}\lambda_{pi}\right)$ and $\left(\tilde{\theta}_{pj}\tilde{v}_{pj}\right)\left(\tilde{\theta}_{ps}\tilde{v}_{ps}\right)$ into power series of $x/p - 1$, we get

$$
\begin{aligned}
\mathcal{I}(0, \infty) &= \int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1} e^{-\frac{n}{2}x} e^{p\left(B_0 + B_1(x/p-1) + B_2(x/p-1)^2\right)} \\
&\quad \times \left(\prod_{j=1}^{r}\prod_{s=1}^{j} \sqrt{1 - 4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p} + o_{\mathrm{P}}(1)\right) \mathrm{d}x, \qquad \text{(A.38)}
\end{aligned}
$$

where B_0, B_1 and B_2 are $O_{\mathrm{P}}(1)$ uniformly in $\tau \in [0, \sqrt{c} - \delta]^r$. The following lemma simplifies the above expression. Its proof is given in Section A.7 of the complete version of this Appendix.

LEMMA 13. *The quadratic term $B_2(x/p + 1)^2$ can be omitted from the exponent in the right-hand side of (A.38) without affecting (A.38)'s validity. That is,*

$$
\begin{aligned}
\mathcal{I}(0, \infty) &= \int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1} e^{-\frac{n}{2}x} e^{p\left(B_0 + B_1\left(\frac{x}{p}-1\right)\right)} \\
&\quad \times \left(\prod_{j=1}^{r}\prod_{s=1}^{j} \sqrt{1 - 4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p} + o_{\mathrm{P}}(1)\right) \mathrm{d}x.
\end{aligned}
$$

Lemma 13 shows that only the constant and linear terms in the expansion of $\tilde{\theta}_{pj}\tilde{v}_{pj} - \frac{1}{2p}\sum_{i=1}^{p} \ln\left(1 + 2\tilde{\theta}_{pj}\tilde{v}_{pj} - 2\tilde{\theta}_{pj}\lambda_{pi}\right)$ into power series of $x/p - 1$ matter for the evaluation of $\mathcal{I}(0, \infty)$. Let us find these terms.

As in the proof of Lemma 9, let $\hat{F}_p^{\lambda\varepsilon}(\lambda)$ be the empirical distribution of the eigenvalues of $\sigma^2\varepsilon\varepsilon'/n$, and let $[x_{1p}, x_{2p}]$ be the smallest interval that includes both the support of $\hat{F}_p^{\lambda}$ and the support of $\hat{F}_p^{\lambda\varepsilon}$. By Theorem 1.1 of Bai and Silverstein (2004), $p\int \lambda \mathrm{d}\hat{F}_p^{\lambda\varepsilon}(\lambda) - p = O_{\mathrm{P}}(1)$. On the other hand,

$$
\begin{aligned}
\left|S_p - p\int \lambda \mathrm{d}\hat{F}_p^{\lambda\varepsilon}(\lambda)\right| &= p\left|\int \lambda \mathrm{d}\left(\hat{F}_p^{\lambda}(\lambda) - \hat{F}_p^{\lambda\varepsilon}(\lambda)\right)\right| \\
&= p\left|\int \left(\hat{F}_p^{\lambda}(\lambda) - \hat{F}_p^{\lambda\varepsilon}(\lambda)\right) \mathrm{d}\lambda\right| \leq r\left(x_{2p} - x_{1p}\right),
\end{aligned}
$$

where the last inequality follows from the fact, established in the proof of Lemma 9, that $\sup_\lambda \left|\hat{F}_p^{\lambda}(\lambda) - \hat{F}_p^{\lambda\varepsilon}(\lambda)\right| \leq r/p$. Since $x_{2p} - x_{1p} = O_{\mathrm{P}}(1)$ (Baik and Silverstein (2006)), $\left|S_p - p\int \lambda \mathrm{d}\hat{F}_p^{\lambda\varepsilon}(\lambda)\right| = O_{\mathrm{P}}(1)$ and

$$S_p - p = S_p - p\int \lambda \mathrm{d}\hat{F}_p^{\lambda\varepsilon}(\lambda) + p\int \lambda \mathrm{d}\hat{F}_p^{\lambda\varepsilon}(\lambda) - p = O_{\mathrm{P}}(1).$$

The latter equality implies that $x/S_p - 1 = x/p - S_p/p + O_{\mathrm{P}}(p^{-1})$ uniformly in $x \in [p - \alpha\sqrt{p}, p + \alpha\sqrt{p}]$. Using this fact, we obtain

$$\begin{aligned}
\tilde{\theta}_{pj}\tilde{v}_{pj} &= \theta_{pj}v_{pj} + \theta_{pj}v_{pj}^2\left(x/p - S_p/p\right) + O_{\mathrm{P}}\left(\left(x/p-1\right)^2\right), \\
\ln\left(2\tilde{\theta}_{pj}\right) &= \ln\left(2\theta_{pj}\right) + \left(x/p - S_p/p\right) + O_{\mathrm{P}}\left(\left(x/p-1\right)^2\right),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{p} \ln\left(K_p^{MP}\left(2\tilde{\theta}_{pj}\right) - \lambda_{pi}\right) &= \sum_{i=1}^{p} \ln\left(K_p^{MP}\left(2\theta_{pj}\right) - \lambda_{pi}\right) \\
&\quad - p\left(1 - 4c_p\theta_{pj}^2 v_{pj}^2\right)\left(x/p - S_p/p\right) + O_{\mathrm{P}}\left(\left(x/p-1\right)^2\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathcal{I}(0,\infty) &= \int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1} e^{-\frac{n}{2}x} e^{p\sum_{j=1}^{r}\left[\theta_{pj}v_{pj} - \frac{1}{2p}\sum_{i=1}^{p}\ln(1+2\theta_{pj}v_{pj}-2\theta_{pj}\lambda_{pi})\right]} \\
&\quad \times e^{\sum_{j=1}^{r}\theta_{pj}v_{pj}(x-S_p)}\left(\prod_{j=1}^{r}\prod_{s=1}^{j}\sqrt{1 - 4\left(\theta_{pj}v_{pj}\right)\left(\theta_{ps}v_{ps}\right)c_p} + o_{\mathrm{P}}(1)\right)\mathrm{d}x.
\end{aligned}$$

This equality, together with (2.4) and Proposition 2, implies that

$$\begin{aligned}
\mathcal{I}(0,\infty) &= \prod_{j=1}^{r}\left(1+\tau_j\right)^{\frac{np}{2}} L_p\left(\tau;\lambda_p\right) \\
&\quad \times \int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1} e^{-\frac{n}{2}x} e^{\sum_{j=1}^{r}\theta_{pj}v_{pj}(x-S_p)}\mathrm{d}x\left(1+o_{\mathrm{P}}(1)\right).
\end{aligned} \tag{A.39}$$

Equations (A.39), (2.5) and the fact that

$$\begin{aligned}
\int_{p-\alpha\sqrt{p}}^{p+\alpha\sqrt{p}} x^{\frac{np}{2}-1} e^{-\frac{n}{2}x} e^{\sum_{j=1}^{r}\theta_{pj}v_{pj}(x-S_p)}\mathrm{d}x &= e^{\sum_{j=1}^{r} -S_p\tau_j/(2c_p)} \\
&\quad \times\left(n/2 - \sum_{j=1}^{r}\tau_j/(2c_p)\right)^{-\frac{np}{2}}\Gamma\left(np/2\right)\left(1+o(1)\right)
\end{aligned}$$

entail

$$\begin{aligned}
L_p\left(\tau;\mu_p\right) &= L_p\left(\tau;\lambda_p\right) e^{\sum_{j=1}^{r} -\frac{\tau_j}{2c_p}S_p}\left(1 - \sum_{j=1}^{r}\tau_j/nc_p\right)^{-\frac{np}{2}}\left(1+o_{\mathrm{P}}(1)\right) \\
&= L_p\left(\tau;\lambda_p\right) e^{-\frac{S_p-p}{2c_p}\sum_{j=1}^{r}\tau_j + \frac{1}{4c_p}\left(\sum_{j=1}^{r}\tau_j\right)^2}\left(1+o_{\mathrm{P}}(1)\right),
\end{aligned}$$

which establishes (2.13). □

A.6. – A.7. Proofs of Lemmas 12 and 13. See Supplementary Material (Onatski et al 2013b).

A.8. Proof of Proposition 4. Proposition 4 follows from Theorem 3. Together with Lemma 12 of OMH, equations (2.12) and (2.13) imply the convergence of the finite-dimensional distributions of

$$\left\{\ln L_p\left(\tau;\lambda_p\right);\tau\in\left[0,\sqrt{c}-\delta\right]^r\right\} \text{ and } \left\{\ln L_p\left(\tau;\mu_p\right);\tau\in\left[0,\sqrt{c}-\delta\right]^r\right\} \tag{A.43}$$

to those of $\{\mathcal{L}_\lambda(\tau);\tau\in[0,\sqrt{c}-\delta]^r\}$ and $\{\mathcal{L}_\mu(\tau);\tau\in[0,\sqrt{c}-\delta]^r\}$, respectively. Note that Lemma 12 of OMH is derived as a corollary to Theorem 1.1 of Bai and Silverstein (2004). Since all statements there hold for non-Gaussian ε with i.i.d. standardized entries having zero excess kurtosis, Lemma 12 of OMH holds under that condition too. Finally, the weak convergence in $C[0,\sqrt{c}-\delta]^r$ follows from the tightness of the sequences of processes (A.43), which is implied by (2.12) and (2.13), and by the fact that $\Delta_p(z_j(\tau))$, $j=1,...,r$ are $O_{\mathrm{P}}(1)$ uniformly in $\tau\in[0,\sqrt{c}-\delta]^r$. $\square$

A.9. Proof of Proposition 5. To save space, we only derive the Gaussian asymptotic power envelope for the relatively more difficult case of real-valued data and μ-based tests. It follows from Proposition 4 that the point-optimal test $LR_{\mu,\tau}(\mu_p)=\mathbf{1}\left\{\ln L_p(\tau;\mu_p)>c_{\mu,\tau}\right\}$ has asymptotic size α if and only if

$$c_{\mu,\tau}=\sqrt{W(\tau)}\Phi^{-1}(1-\alpha)+m(\tau), \tag{A.44}$$

where

$$\begin{aligned} m(\tau) &= \frac{1}{4}\sum_{i,j=1}^{r}\left(\ln\left(1-\frac{\tau_i\tau_j}{c}\right)+\frac{\tau_i\tau_j}{c}\right) \qquad \text{and} \\ W(\tau) &= -\frac{1}{2}\sum_{i,j=1}^{r}\left(\ln\left(1-\frac{\tau_i\tau_j}{c}\right)+\frac{\tau_i\tau_j}{c}\right). \end{aligned}$$

Now, Le Cam's third lemma and Proposition 4 entail that, under $h=\tau$, $\ln L_p(\tau;\mu_p)\xrightarrow{d}\mathcal{N}(m(\tau)+W(\tau),W(\tau))$. Proposition 5 follows. $\square$

A.10. Invariant tests. See Supplementary Material (Onatski et al 2013b).

A.11. – A.13. Proofs of Propositions 6, 7, and 8. See Supplementary Material (Onatski et al 2013b).

SUPPLEMENTARY MATERIAL

Supplement : Appendix to "Signal Detection in High Dimension: The Multispiked Case"
(doi: COMPLETED BY THE TYPESETTER; .pdf). This supplement (Onatski et al 2013b) provides an extended version of the mathematical appendix above, including Sections A2-A4, A6-A7, and A10-A13.

REFERENCES

[1] Azais, J-M. and Wschebor, M. (2002). The distribution of the maximum of a Gaussian process: Rice method revisited, in V. Sidoravicius ed. *In and Out of Equilibrium: Probability with a Physical Flavour, Progress in Probability*. Birkhäuser, Basel.

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

[3] Bai, Z.D., Jiang, D., Yao, J.F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Annals of Statistics* **37** 3822–3840.

[4] Baik, J., Ben Arous, G., and Péché, S. (2005). Phase transition of the largest eigenvalue for non-null complex covariance matrices. *Ann. Probab.* textbf33 1643–1697.

[5] Baik, J. and Silverstein, J.W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis* textbf97 1382–1408.

[6] Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*. John Wiley & Sons, New York.

[8] Cai, T.T. and Ma, Z. (2012). Optimal hypothesis testing for high-dimensional covariance matrices. *Bernoulli*, to appear.

[8] Cai, T.T., Ma, Z. and Wu, Y. (2013). Optimal estimation and rank detection for sparse spiked covariance matrices. arXiv: 1305.3235.

[9] Chen, S.X., Zhang, L.X. and Zhong, P.S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association* **105** 810-819.

[10] Féral, D. and Péché, S. (2009). The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case. *Journal of Mathematical Physics* **50** 073302.

[11] Guionnet, A. and Maïda, M. (2005). A Fourier view on the R-transform and related asymptotics of spherical integrals. *Journal of Functional Analysis* **222** 435–490.

[12] James, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics* **35** 475-501.

[13] John, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127.

[14] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics* **29** 295–327.

[15] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Annals of Statistics* **30** 1081-1102.

[16] Lehmann, E.L. and Romano, J. P. (2005). *Testing Statistical Hypotheses.* Third Edition, Springer, New York.

[17] Mo, M.Y. (2012). The rank 1 real Wishart spiked model. *Communications on Pure and Applied Mathematics* **65** 1528–1638.

[18] Olver, F.W.J. (1997). *Asymptotics and Special Functions.* A K Peters, Natick, MA.

[21] ONATSKI, A. (2012). Detection of weak signals in high-dimensional complex-valued data. arXiv:1207.7098.

[20] ONATSKI, A., MOREIRA, M. J., and HALLIN, M. (2013a). Asymptotic power of sphericity tests for high-dimensional data. *Annals of Statistics* **41**, 1204-1231.

[21] ONATSKI, A., MOREIRA, M. J., and HALLIN, M. (2013b). Appendix to "Signal Detection in High Dimension: The Multispiked Case". DOI ... to be supplied

[22] PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probability Theory and Related Fields* **143** 481–516.

[23] SCHOTT, J. (2006). A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix. *Journal of Multivariate Analysis* **97** 827-843.

[24] SILVERSTEIN, J.W. and BAI, Z.D. (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis* **54** 175-192.

[25] SRIVASTAVA, M.S. (2005). Some tests concerning the covariance matrix in high-dimensional data. *Journal of the Japan Statistical Society* **35** 251-272.

[26] UHLIG, H. (1994). On singular Wishart and singular multivariate Beta distributions. *Annals of Statistics* **22** 395-405.

[27] VAN DER VAART, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press.

[28] WANG, D. (2012). The largest eigenvalue of real symmetric, Hermitian and Hermitian self-dual random matrix models with rank one external source, part I. *Journal of Statistical Physics* **146** 719-761.

ADDRESS OF THE FIRST AUTHOR
ALEXEI ONATSKI
FACULTY OF ECONOMICS
UNIVERSITY OF CAMBRIDGE
SIDGWICK AVENUE
CAMBRIDGE, CB3 9DD, UK.
E-MAIL: ao319@cam.ac.uk

ADDRESS OF THE SECOND AUTHOR
MARCELO J. MOREIRA
ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS
FUNDAÇÃO GETULIO VARGAS (FGV/EPGE)
PRAIA DE BOTAFOGO, 190 - SALA 1100
RIO DE JANEIRO-RJ 22250-900, BRAZIL.
E-MAIL: mjmoreira@fgv.br

ADDRESS OF THE THIRD AUTHOR
MARC HALLIN
ECARES
UNIVERSITÉ LIBRE DE BRUXELLES CP 114/04
50, AVENUE F.D. ROOSEVELT
B-1050 BRUXELLES, BELGIUM
E-MAIL: mhallin@ulb.ac.be